HDRi2014 - Second International Conference and SME Workshop on HDR imaging (2014)
Bouatouch K. and Chalmers A. (Editors)

HDR Video Coding based on Local LDR Quantization

David Touzé Yannick Olivier Sébastien Lasserre Fabrice Le Léannec Ronan Boitard Edouard François

Technicolor, 975 Av. des Champs Blancs, 35576 Cesson-Sevigne, France;

Abstract

High Dynamic Range (HDR) is considered as a major feature that will bring to the end-user a significantly improved experience, closer to the human perception. It is even considered as possibly more important than increased spatial or temporal resolutions. Its deployment in a mid-term future would preferably require coding solutions able to re-use existing or in-coming video codec. This paper presents an HDR video coding scheme aiming at offering high compression performance while re-using legacy low bit-depth decoders (e.g. AVC 8-bit, HEVC 8- or 10-bit). To this end, we propose a fully generic solution that can be configured to guarantee backward compatibility with LDR devices (decoders and displays). The proposed technique transforms HDR frames into LDR-like ones and uses side- and meta-data to help the decoder achieve a good HDR reconstruction. Results show that this solution performs well for bit-rate typically targeted in broadcast or internet streaming video applications.

Categories and Subject Descriptors (according to ACM CCS): I.4.2 [IMAGE PROCESSING AND COMPUTER VISION]: Compression (Coding)—

1. INTRODUCTION

HDR and Wide Color Gamut (WCG) are two major technologies that will bring to the end-user a significantly improved experience, possibly more than increased spatial or temporal resolutions. Deployment of HDR/WCG video services can be envisioned in a wide variety of applications, from very high quality (e.g. storage media such as Blu-ray) to medium-to-low quality applications (satellite, cable, terrestrial broadcast or low bit-rate video streaming over IP). Considering this wide range of applications and related operational points, it is important to develop video compression technologies that guarantee a good trade-off between compression efficiency and implementation complexity.

Much work has been performed on the HDR video compression problem and most of it id using state-of-the-art video codecs, such as ITU-T H.264 / MPEG-4 Part 10 Advanced Video Codec (AVC) codec [WSBL03], further referred as H.264/AVC, or as the most recent MPEG/ITU-T video coding standard HEVC [BHO*13]. However, most of the existing methods either propose a proprietary (codec with more than 8-bits) or a multi-instance implementation (more than one codec at a time). In this paper, we present an HDR video coding scheme that allows to re-use legacy low bit-depth decoder (e.g. AVC 8-bit, HEVC 8- or 10-bit). Our method only uses one codec instance as HDR-related side information can be compressed using a simple encoding scheme and enclosed as metadata in the main bitstream. Our solution is generic, and can be configured to have a backward compatibility with LDR devices (decoders and displays).

This paper provides an overview of our generic solution along with a specific implementation targeting the backward compatibility with LDR devices. The compression performance of our implementation is reported on a set of HDR test sequences.

2. Related Work

Previous works, based on video codecs such as H.264/AVC, can be divided in at least three main techniques. Most of these techniques require a tone mapping operation to transform HDR floating-point data into LDR 8-bits data. This operation can be performed either off-line in post-production by colorists or during the encoding stage.

The first main technique [LK08, WS04] uses a ratio-based method and requires two codec instances: one to encode an LDR tone mapped version while the second encodes a quantized ratio between the original HDR content and its associated tone mapped LDR content (from 8 to 14 bits) . At the decoder side, the decoded LDR content is multiplied by the decoded ratio content to reconstruct the HDR content.

The second main technique [MMM*11, KD12, MEMS06] relies on an inverse tone mapping operator (operation that transforms LDR 8-bits data into floating-point HDR data). This approach is scalable and requires two codec instances: one to compress a tone mapped LDR version of the original HDR content and a second that encodes only the residuals between the original HDR content and an inverse tone mapped version of the decoded LDR tone mapped content.

The third main technique [MKMS04, ZRB11, MT10] is based on a perceptual curve that transforms HDR content (floating-point data) into a LDR content of high bit-depth (integer of more than 8 bits). The generated LDR content is directly compressed using a single high-bit depth codec. At the decoder side, the decoded LDR content is computed by the inverse of the perceptual curve to reconstruct the HDR content.

The main advantage of the first two methods is the ability to provide two contents: an HDR one and a LDR one, the latter being compatible with conventional displays when the tone mapping is performed in Post-Production. However, their main drawback is the use of two codec instances, with one (The HDR codec) that may require more than 8-bit for a good HDR reconstruction.

On the contrary, the third technique only needs one codec, however, as it relies on a perceptual curve, the HDR content needs to be scene-referred (say each HDR pixel's value represents the luminance value of the real scene). Note if the HDR content is not scene-referred, one can apply a log-transform that is a coarse approximation of a perceptual curve over the whole range. In addition, the used codec usually requires a bit-depth from 10 to 14 bits for a good HDR reconstruction. To solve the aforementioned issue, we propose a compression scheme based on the product of a low resolution illumination image with a full resolution detail image.

3. Overview of the coding scheme

3.1. Generic description

The Figures 1 and 2 provide a simplified synoptic of the proposed HDR encoder and decoder schemes. Both the encoder and decoder are made of two main parts. Part 1 corresponds to the HDR signal decomposition/recomposition. At the encoder side, the input HDR signal is decomposed into two signals of low dynamic range. At the decoder side, the HDR signal is recomposed from two the decoded signals of low dynamic range. Part 2 corresponds to the encoding/decoding processes, which aim at re-using limited bit-depth schemes such as AVC 8-bit, HEVC 8- or 10bit to encode/decode the LDR signals.

The proposed solution exploits the locally low dynamic range property of the HDR signal (LDR localization). Based on this property, the approach consists in splitting the input HDR signal into two integer signals of low dynamic range and limited bit-depth (e.g. 8 or 10 bits):

- a low frequency signal which corresponds to the local luminance signal mean,
- a residual signal between the locally LDR signal and HDR signal.

The signal decomposition enables keeping a very high signal precision and finely adapting the quantization to the local signal characteristics.

At the encoder side (cf. Figure 1), the HDR signal decomposition works as follows. First the luminance component of the input HDR signal is computed and processed to generate a low dynamic range low frequency signal. Thanks to its low frequency property, the spatial resolution of the low frequency signal can be significantly lower than the input signal resolution. A residual signal is then computed as the remaining between HDR signal and the resulting low frequency signal. This corresponds to a demodulation of the HDR signal by the low frequency signal. A perceptual color transform is then applied in order to quantize the signal while preserving perceptually its characteristics and variations. In particular, low values are more finely quantized than high values; note that high values are quantized in order to control luminance and color saturations. This process can also take into account local signal properties. Metadata are also associated to the HDR signal decomposition process.

The resulting signals are then encoded. The residual signal, of same resolution as the native HDR signal, can be encoded using an existing limited bit-depth encoder, such as AVC 8-bit, HEVC 8- or 10-bit. The low frequency signal can either be encoded using an existing limited bit-depth encoder (in this case two codec

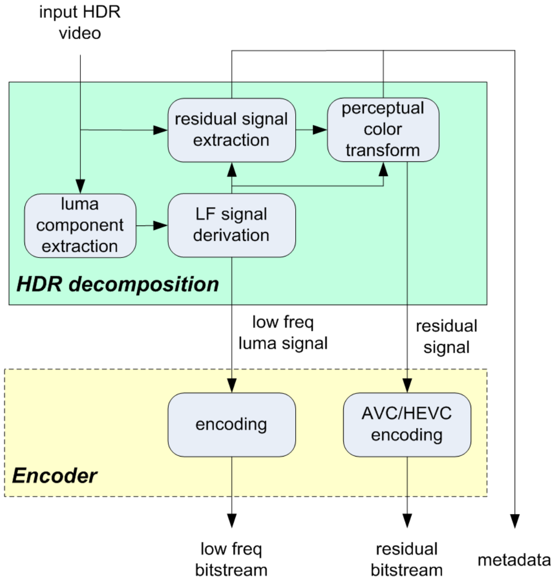

Figure 1: simplified encoding scheme.

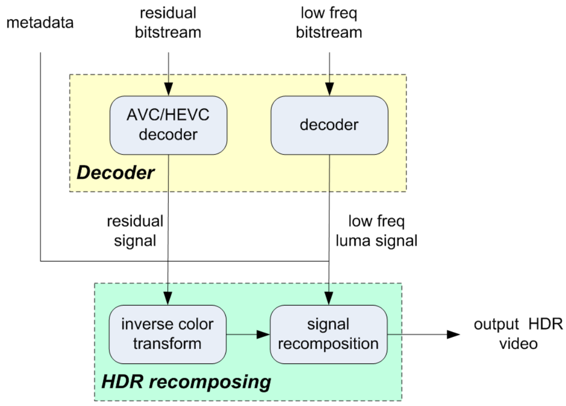

Figure 2: Simplified decoding scheme.

instances are used) or using more simple scheme. As this signal is by nature of very low entropy, its resolution can be reduced resulting in a small coding cost compared to the residual signal. Among those more simple scheme, we can cite the codec's entropy coder. In this case, the coding cost of those weights were below 100 kbps for HD 1080p resolution (around 2000 coefficients). Using those more simple cases allow the use of only one codec instance.

The decoder operates as follows (cf. Figure 2). First the LDR low frequency and residual signals are decoded. A legacy decoder (AVC 8-bit, HEVC 8- or 10-bit) can be used. The decoded signals are then post-processed to generate the decoded HDR signal. An inverse color transform is applied to the decoded residual signal. This step can use the low frequency signal as input to locally adapt the transform process to the local low frequency luminance. Then both signals are combined to generate the reconstructed HDR signal. The combination is similar to a modulation process of the residual signal by the low frequency signal.

3.2. Main characteristics of the proposed scheme

The solution has the following main assets: it enables to re-use a legacy low bit-depth decoder (e.g. AVC 8-bit, HEVC 8-bit, or HEVC 10-bit) and it offers backward compatibility with LDR decoders. It also provides a potential backward compatibility with LDR displays as it can be configured to generate a viewable residual signal corresponding to a tone-mapped version of the HDR signal as illustrated in the next section. Moreover our method can be designed to require only one bitstream and codec instance to transmit HDR data. The exploitation of the locally LDR property of the HDR signal, as well as the perceptually based color adaptation of the residual signal, allow for a fine quantization of the LDR signal resulting in a better coding efficiency than using a global HDR signal quantization.

4. Implementation based on LDR and Illuminant Map

In this section a specific implementation of the generic solution is described. This implementation aims at directly exploiting the design of existing HDR displays which are based on the modulation of a LDR signal by an illuminant map. Furthermore, the residual signal can be directly viewed on a LDR display with a rendering consistent with the original HDR scene.

The low frequency signal derivation consists in modeling the illuminant map as a linear combination of overlapping shape functions. The array of weights $a_{i,j}$ of this linear combination is the low frequency signal LF. The shape functions are pure mathematical construction in order to fit the illuminant map at best. Shape functions can be defined by default, or signaled as metadata.

Let W and H be the picture resolution while W/m and H/n are the illuminant map resolution. Let $\Phi_{i,j}$ be a shape function with weight $a_{i,j}$ at position (i,j) in the illuminant map. The full resolution illuminant map LF for any pixel (k,l) is simply the linear com-

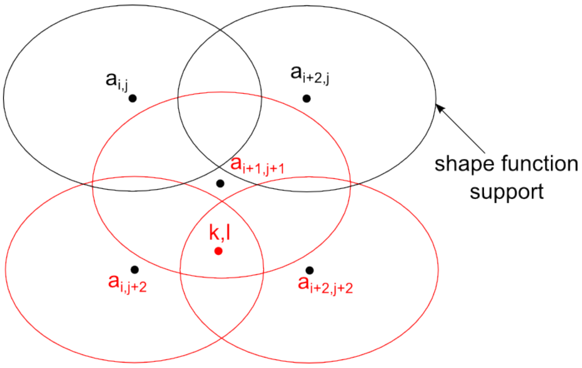

Figure 3: Example of shape function topology. The value of the pixel at location (k,l) will depend on the three coefficient ($a_{i+1,j+1}$, $a_{i,j+2}$ and $a_{i+2,j+2}$) associated to shape functions.

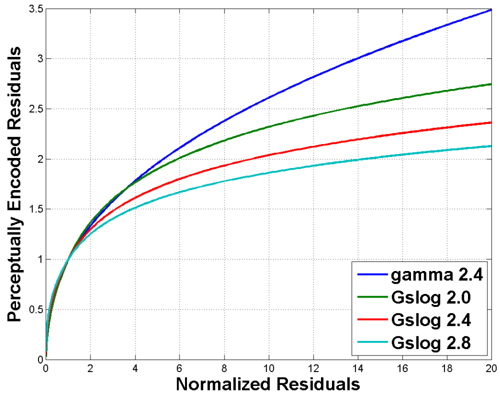

Figure 4: Example of gamma (with γ=1/2.4) and Gslog curves for 1/2.0, 1/2.4 and 1./2.8.

bination of the shape functions:

$$LF[k,l] = \sum_{(i,j)\in V_{K,L}}^{max} a_{i,j}\Phi_{i,j}[k - m \cdot i, l - n \cdot j] \quad (1)$$

with $K = I[k/m]$ and $L = I[l/n]$, $I[x]$ being the nearest integer from x, m and n the horizontal and vertical sizes of the shape function and $V_{K,L}$ a neighborhood of the pixel (K, L). Figure 3 illustrates how the illuminant map (weight $a_{i,j}$) are used to compute the value at pixel position (k,l) using shape functions.

The coefficients $a_{i,j}$ can be identified using a least mean square method to minimize the mean square error between the illuminant map and the luminance (fast algorithms using downsampled versions of the HDR signal and shape functions can be used without lack of accuracy). Furthermore, a temporal stabilization term is included to preserve the temporal consistency of the illuminant map:

$$a_{opt} = argmin\left(||LF - L_w||_2^2 + \lambda^2||a - aprev||_2^2\right) \quad (2)$$

The illuminant map LF is then re-normalized into LF' in order to keep the brightness consistency of the residual signal with the original HDR signal, such that bright scenes look bright, dark scenes look dark, and mid-grey scenes look mid-grey on the LDR residual video sequence.

$$LF' = LF\frac{C_m L_{w,mean}^{\alpha}}{LF_{mean}} \quad (3)$$

with C_m a parameter to adjust the overall brightness and L, w the HDR luminance. As multiplying directly the illumant map by the geometric mean ($L_{w,mean}$) would result in too much dynamic in the residuals, α is used to compress $L_{w,mean}$ (value of 1/3 is usually considered as it allows to map 24 f-stops on 8 f-stops).

The resulting coefficients $a'_{i,j}$ are then quantized into N bits (typically 8 or 10) using a configurable a_{max} value. We have found experimentally a very consistent value of 2.5 for a various set of HDR sequences. This value is transmitted as metadata. The re-normalized illuminant map LF' is used to compute the HDR residual as follows (a similar process would apply to RGB components):

$$XYZ_{res}[k,l] = XYZ_{HDR}[k,l]/LF'[k,l] \quad (4)$$

As the illuminant map is re-normalized around mid-grey levels, the residual signal is centered around value 1 for mid-bright HDR images and may still be of wide dynamic. In particular, very bright pixels may still be present because of specular light or very bright small objects. It is known that a gamma correction does not flatten high lights fast enough to avoid burning of bright pixels after clipping. Therefore a combination of Gamma correction and S-log correction (noted Gslog hereafter) is used in order to finely quantize the dark ranges (thanks to the gamma function) while avoiding too harsh high light saturations (thanks to the S-log function). The actual inverse transfer function is therefore made of two parts and is defined as:

$$Gslog(x) = \begin{cases} a \cdot ln(x+b) + c & \text{if } x \geq 1 \\ x^{1/\gamma} & \text{if } x < 1 \end{cases} \quad (5)$$

Table 1 provides the different values of the parameter a, b and c with the corresponding γ that are plotted in Figure 4. It can be observed that high lights are lowered much more aggressively with a Gslog curve than with a simple gamma curve. Experimentally a value γ=1/2.8 is generally needed to catch all the dynamic of the residual.

The Gslog curve is applied to the three components

Table 1: Parameters for the Gslog curves for different γ.

γ	a	b	c
1/2.0	0.6275	0.2550	0.8575
1/2.4	0.4742	0.1382	0.9386
1/2.8	0.3861	0.0811	0.9699

X, Y and Z of XYZ_{res}. The resulting perceptually encoded residuals (XYZ_{perc}) are then converted to a LDR signal by the following operation:

$$XYZ_{LDR} = max(2^N - 1, S_Q \cdot XYZ_{perc}) \quad (6)$$

The scaling (S_Q) and clipping are applied component by component on either mapped XYZ or RGB. The scaling parameter S_Q is also encoded as a metadata. Ideally, it should map 1 around the neutral gray 2^{N-1}. For a LDR video with a standard number of bits N=8, we have found experimentally a very consistent value of 120. A summary of the whole process is illustrated in Figure 5. Finally the LDR 4:4:4 signal is converted to the YUV 4:2:0 color space before encoding. More characteristics of the Gslog curve can be found in [LLLF13].

At the decoder side, the inverse operations are applied once the residual signal (YUV_{dec}) and the low frequency signal $a_{i,j}^{dec}$ have been decoded. As we use a lossless coding for the weights $a_{i,j}$, the weights are the same at the encoder and decoder stage.

5. Experiments

This section reports results using the proposed coding scheme. Four HDR 1080p test sequences are used. All sequences are in BT.709 color space. They correspond to a variety of ranges (15 to 21 f-stops), illumination conditions, color ranges, scene and motion complexity, and frame rate. More details can be found in [LLLF13]. The encoding/decoding are achieved using the HM, with a bit-depth of 10 bits (Main10 profile). The quantization steps are adaptively chosen to reach bitrates typically targeted in consumer-oriented video distribution applications such as broadcast.

The objective performance is measured using the ΔE_{2000} measure. ΔE_{2000} has been specified by the CIE to measure the perceptual distance or difference between two colors. It may be used to measure the shift in color due to quantization of the color image signals. It is often considered that a difference of $\Delta E_{2000} = 1$ is the visibility threshold of a color difference. Instead of using CIE Lab color space that is designed exclusively for LDR imagery, we used the HDR Lab space [Fai]. We chose the D65 illuminant as white point. Results, summarized in Table 1, show ΔE_{2000} values ranging from 0.4 to 4.7, which corresponds visually to limited perceptual degradations. Note that the use of objective metrics is still an open question for the HDR field. These results have been confirmed by visual evaluation made in our lab using a SIM2 display.

Table 2: Coding results for the considered test sequences (mean ΔE_{2000}).

Sequence	kbps	ΔE_{2000}
Market3	15041	4.09
	10464	4.40
	5249	5.18
	3378	5.72
Balloon	13158	2.43
	8148	2.48
	5582	2.53
	3270	2.62
Fire-eater	15949	0.42
	9072	0.45
	5302	0.48
	3384	0.51
Seine	14639	1.80
	10404	1.96
	5514	2.33
	3137	2.66

6. Conclusion

A new solution for HDR video coding has been presented. Its main asset is the use of legacy encoding/decoding devices with potential backward compatibility with existing LDR displays. This method can be implemented using either one or two codec instances and bitstreams. As the low frequency signal is of low entropy, the use of a second codec is too complex especially since encoding the weights using a codec's entropy coder is sufficient. Performance at typical broadcast bitrates are good which show that HDR content could be deployed on existing distribution infrastructure without significant impact in terms of equipment and used bitrates.

References

[BHO*13] Bross B., Han W.-J., Ohm J.-R., Sullivan G. J., Wang Y.-K., Wiegand T.: High efficiency video coding (hevc) text specification draft 10 (for fdis & consent). ISO/IEC JTC1/SC29/WG11 JCTVC-L1003, Jan 2013.

[Fai] Fairchild M. D.: Brightness, Lightness, and Specifying Color in High-Dynamic-Range Scenes and Images, .

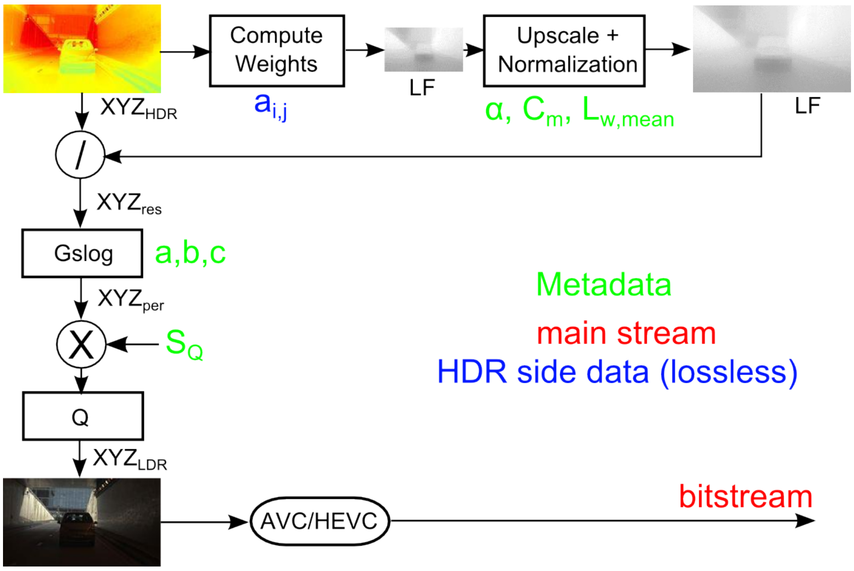

Figure 5: Workflow summarizing the decomposition of the HDR signal into two LDR ones.

[KD12] Koz A., Dufaux F.: Optimized Tone Mapping with Perceptually Uniform Luminance Values for Backward-Compatible High Dynamic Range Video Compression. *IEEE Proc. Visual Communications and Image Processing (VCIP'2012)* (2012).

[LK08] Lee C., Kim C.-s.: Rate-Distorsion Optimized Compression of High Dynamic Range Videos. *16th European Signal Processing Conference (EUSIPCO)*, Eusipco (2008).

[LLLF13] Lasserre S., Le Léannec F., Francois E.: Description of HDR sequences proposed by Technicolor. In *ISO/IEC JTC1/SC29/WG11 JCTVC-P0228* (San Jose, USA, 2013), IEEE.

[MEMS06] Mantiuk R., Efremov A., Myszkowski K., Seidel H.-p.: Backward compatible high dynamic range MPEG video compression. In *ACM SIGGRAPH 2006 Papers on - SIGGRAPH '06* (New York, New York, USA, 2006), vol. 1, ACM Press, p. 713.

[MKMS04] Mantiuk R., Krawczyk G., Myszkowski K., Seidel H.-P.: Perception-motivated high dynamic range video encoding. *ACM Transactions on Graphics (Proc. of SIGGRAPH) 23*, 3 (Aug. 2004), 733.

[MMM*11] Mai Z., Mansour H., Mantiuk R., Nasiopoulos P., Ward R., Heidrich W.: Optimizing a tone curve for backward-compatible high dynamic range image and video compression. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society 20*, 6 (June 2011), 1558–71.

[MT10] Motra A., Thoma H.: An adaptive Logluv transform for High Dynamic Range video compression. In *2010 IEEE International Conference on Image Processing* (Sept. 2010), IEEE, pp. 2061–2064.

[WS04] Ward G., Simmons M.: Subband encoding of high dynamic range imagery. In *Proceedings of the 1st Symposium on Applied perception in graphics and visualization - APGV '04* (New York, New York, USA, 2004), ACM Press, p. 83.

[WSBL03] Wiegand T., Sullivan G., Bjontegaard G., Luthra A.: Overview of the H.264/AVC video coding standard. *IEEE Transactions on Circuits and Systems for Video Technology 13*, 7 (July 2003), 560–576.

[ZRB11] Zhang Y., Reinhard E., Bull D.: Perception-based high dynamic range video compression with optimal bit-depth transformation. In *2011 18th IEEE International Conference on Image Processing* (Sept. 2011), IEEE, pp. 1321–1324.