|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  16th Meeting: San José, US, 9–17 Jan. 2014 | Document: JCTVC-P0159r1  m32076 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **High Dynamic Range video coding** | | |
| *Status:* | Input Document to JCT-VC | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Sébastien Lasserre, Fabrice Le Léannec, Patrick Lopez, Yannick Olivier, David Touzé, Edouard François  975 avenue des champs blancs CS 17616, 35576, Cesson-Sévigné Cedex, France | Tel: Email: | +33 6.16.44.53.07  [sebastien.lasserre@technicolor.com](mailto:sebastien.lasserre@technicolor.com) [fabrice.leleannec@technicolor.com](mailto:fabrice.leleannec@technicolor.com)  [edouard.francois@technicolor.com](mailto:edouard.francois@technicolor.com) |
| *Source:* | Technicolor | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

This contribution presents an HDR video coding scheme. The proposed solution aims at offering high compression performance while enabling re-using legacy low bit-depth decoders (e.g. AVC 8-bit, HEVC 8- or 10-bit). The contribution gives first an overview of the solution. It is reported that the solution, while being fully generic, can be configured to guarantee backward compatibility with LDR devices (decoders and displays) and to be directly adapted to existing LCD-LED HDR display technologies. This adaptation is also presented. Finally preliminary coding results are provided, comparing performance of the proposed scheme to a coding chain based on prior global quantization of the HDR signal.

# Introduction

HDR and Wide Color Gamut are two major features 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 applications (e.g. storage media such as bluray) to medium-to-low quality applications (satellite, cable or terrestrial broadcast to low bit-rate video streaming over IP), [1].

Considering this wide range of applications and related operational points, it is important to develop video compression technologies that guarantee good compression performance and implementation complexity trade-off.

In this contribution, an HDR video coding scheme is presented. The proposed solution has the following main characteristics:

* Re-use of a legacy low bit-depth decoder (e.g. AVC 8-bit, HEVC 8- or 10-bit)
* Usage of HDR-related side information
* Potential backward compatibility with LDR devices (decoders and displays)
* Generic technology, that can at the same time be directly adapted to existing LCD-LED HDR displays
* High compression performance

The document gives an overview of this generic solution. A specific implementation targeting the backward compatibility with LDR devices is also described. Compression performance of this latest approach is presented on a set of several HDR test sequences. It also includes some comparisons with a PQ EOTF-based coding chain.

# Overview of the coding scheme

## Generic description

The Figure 1 and Figure 2 provide a simplified synoptic of the proposed HDR encoder and decoder schemes. Both the encoder and decoder are made of two main parts:

* the HDR signal decomposition/recomposing, which decomposes at the encoder side the input HDR signal into two signals of low dynamic range and recomposes at the decoder side the HDR signal from two decoded signals of low dynamic range;
* the encoding/decoding, which aims 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, and a residual signal which corresponds to the locally LDR signal made of the remaining high frequency 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.

* The luminance component of the input HDR signal is first extracted.
* The luminance component is then 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.
* The residual signal is then generated as the remaining high frequency component of the HDR signal once the low frequency signal has been extracted. This extraction is based on a demodulation of the HDR signal by the low frequency signal.
* A step of perceptual color transform is then applied in order to quantize the signal while perceptually preserving the signal characteristics and variations. In particular, low values are more finely quantized than high values; in addition high values are quantized in order to control luminance and color saturations. The 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 also be encoded using an existing limited bit-depth encoder. This signal is by nature of very low entropy. Its resolution can be reduced and its coding cost is small compared to the residual signal.



Figure 1: simplified encoding scheme

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 possibly uses the low frequency signal as input, in order 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.



Figure 2: simplified decoding scheme

## Main characteristics of the proposed scheme

The solution has the following main assets:

* **Backward compatibility with LDR decoders**: it enables re-using a legacy low bit-depth decoder (e.g. AVC 8-bit, HEVC 8-bit, or HEVC 10-bit); it offers backward compatibility with LDR decoders;
* **Potential backward compatibility with LDR displays**: it offers a potential backward compatibility with LDR displays; indeed 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;
* **High compression performance**: 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.
* **Potential backward compatibility with current generation of HDR displays**: the solution is generic since it is able to generate a reconstructed HDR signal; but it can also be directly adapted to existing LCD HDR displays based on LED backlighting, which operate by signal modulation; the scheme can be configured so that the low frequency signal corresponds to the backlight signal of the display, while the residual signal corresponds to the LCD picture.

# Specific implementation based on LDR and backlighting

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 an LDR signal by an LED backlighting. In addition the residual signal can be directly viewed with an LDR display, and can be considered as a toned-mapped version of the HDR image which renders consistently compared to the original HDR scene.

## Encoder

### Low frequency signal derivation

The low frequency signal derivation consists in modeling the luminance map as a linear combination of overlapping shape functions. The array of weights *ai,j* of this linear combination is the low frequency signal which corresponds to the backlight signal *Ba*. The shape functions may be the true physical response of a display backlight (made of LED’s for instance, the shape function corresponds to the response of one LED) or may be a pure mathematical construction in order to fit the luminance map at best. Shape functions can be defined by default, or signaled as metadata.

Let *W* and *H* be the picture resolution, and *W*/*m* and *H*/*n* be the backlight picture resolution. Let *ψi,j* be a shape function with weight *ai,j* at position (*i,j*) in the backlight picture. The full resolution backlight *Ba* for any pixel (*k,l*) is simply the linear combination of the shape functions:

with *K* = I [ *k* / *m* ] and *L* = I [ *l* / *n* ], I [ *x* ] being the nearest integer from *x*, *m* and *n* being the horizontal and vertical sizes of the shape function, and *VK,L* being a neighborhood of the pixel (*K*,*L*). For instance, *VK,L* contains the 8 pixels (*K*-1...*K*+1, *L*-1...*L*+1), as illustrated in Figure 3, where the backlight array is made of shape functions placed according to a quincunx structure.

The coefficients *ai,j* can be identified using a least mean square method to minimize the mean square error between the backlight and the luminance map (fast algorithms using downsampled versions of the HDR signal and shape functions can be used without lack of accuracy).



Figure 3: example of shape functions topology.

The backlight signal *Ba* is then re-normalized into *Ba’* 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. 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 can be transmitted as metadata.

### Residual signal extraction

The re-normalized backlight signal *Ba’* is used to demodulate the HDR signal as follows:

*Xres*[*k,l*] = *X*[*k,l*] / *Ba’* [*k,l*]

*Yres*[*k,l*] = *Y*[*k,l*] / *Ba’* [*k,l*]

*Zres*[*k,l*] = *Z*[*k,l*] / *Ba’* [*k,l*]

The same process applies to each component in case of RGB format.

### Perceptual color transform

As the backlight signal 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 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 built as follows:

* a gamma curve, with power γ, below 1,
* an S-log curve above 1 : S-log(x) = a ln(x+b) + c

The parameters of the S-log curves are determined such that 0 and 1 are invariant (S-log(0) = 0 and S-log(1) = 1) and the derivative is continuous in 1 (S-log’(1) = γ). This leads to the parameters shown in the table below.

Table 1: S-log parameters for different gamma values.

|  |  |  |  |
| --- | --- | --- | --- |
| **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 |

In Figure 4, the standard gamma curve with γ=1/2.4 and several gamma-Slog (also noted GSlog) curves are plotted for 1/2.0, 1/2.4 and 1./2.8. It can be observed that high lights are lowered much more aggressively with a gamma-Slog curve than with a simple gamma curve.

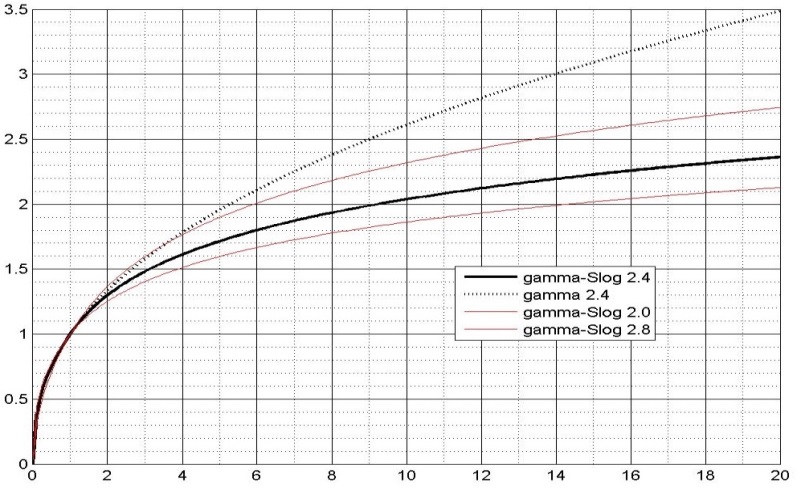


Figure 4: gamma-Slog curves

Experimentally a value γ=1/2.8 is generally needed to catch all the dynamic of the residual. Of course, this value depends on the way the backlight is determined, and may be adjusted depending on the sequence and encoded as accompanying metadata.

The Gamma-Slog curve is applied to the three components *XresYresZres* (or *RresGresBres*).

This mapped residual is then converted to the LDR signal by the following operation:

*LDR* = max(2*N*-1, *cscal*\**res*).

The scaling and the clipping are applied component by component on either mapped XYZ or RGB. The scaling parameter 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.

Finally the LDR 4:4:4 signal is converted to the YUV 4:2:0 color space before encoding.

## Decoder

The inverse operations are applied at the decoder side once the residual signal *YdecUdecVdec* and the low frequency signal have been decoded.

### Inverse color adaptation

The following steps are applied:

* Color space conversion: *YdecUdecVdec* to *XLDRYLDRZLDR*
* Inverse scaling : *Xiscal* = *XLDR* / *cscal Yiscal* = *YLDR* / *cscal Ziscal* = *ZLDR* / *cscal*
* Inverse Gamma-Slog function: *Xtf* = GSlog-1(*Xiscal*) *Ytf* = GSlog-1(*Yiscal*) *Ztf* = GSlog-1(*Ziscal*)

### Signal recomposing

The signal recomposing first reconstructs the backlight picture from the decoded low frequency signal () and from the shape functions .

Finally the reconstructed XYZ signal is modulated to generate the HDR signal:

*XHDR* = *Xtf* \* *Barec**YHDR* = *Ytf* \* *Barec*ZHDR = *Ztf* \* *Barec*

# Experiments

This section reports results using the proposed coding scheme. A comparison with an implementation of the PQ EOTF approach is also made.

## Test conditions

Four HDR test sequences are used (see Table 2 and [2] for more details). All sequences are in BT.709 color space. They correspond to a variety of illumination conditions, color ranges, scene and motion complexity, and frame rate.

The quantization steps are adaptively chosen to reach bitrates ranges typically used in consumer oriented video distribution applications such as broadcast.

The objective performance is measured using the *E2000* PSNR. *E2000* has been specified by the CIE to measure the perceptual distance or difference between two colors. Additional explanations about *E2000* can be found in [3]. The PSNR*E* is computed for each picture as follows:

with *Emean* being the average *E2000* over the picture. 65504 is the peak value of the half-float representation of the HDR sequences. The PSNR is averaged over the entire sequence.

Table 2: list of test sequences.

|  |  |  |  |
| --- | --- | --- | --- |
| Seine  C:\Users\lasserres\Documents\MATLAB\img_exr\Peniches sat=1,3\colorgraded 709\PENICHES_FUSION_01500sat=1,3_contrast_tonemapped.jpg | 1080p 25 fps | 21 f-stops | Mostly dark with some bright spots  Limited colors  Static scene with slow motion |
| Balloon  \\RENNDW7RDP0230\public\catherine\montgolfiere_hd_fusion_clip7-1_sat14gamma135_contrast00500_tonemapped_bis.jpg | 1080p 30 fps | 16 f-stops | Medium illumination  Medium color spectrum  Slow global and local motions |
| Market3  \\RENNDW7RDP0230\public\catherine\marche1_fusion_sat14gamma16_contrast00300_tonemapped.jpg | 1080p 50 fps | 15 f-stops | High illumination  Wide color spectrum  Static scene with slow motion |
| Fire-eater2  C:\Users\lasserres\Documents\MATLAB\img_exr\cracheur2\colorgraded 709\cracheur2_fusion_00495_sat11gamma12_contrast_tonemapped.jpg | 1080p 25 fps | 18 f-stops | Mostly dark with high luminance fires  Limited colors  Static scene with complex random motions |

## PQ EOTF/ Ydzdx chain implementation

An implementation of the PQ EOTF approach has been made for comparisons. The coding chain has been implemented according to the SMPTE draft recommendations WD ST 2084 EOTF and SMPTE 2085 ‘YDzDx Color-Difference Encoding for XYZ Signals’, as illustrated in Figure 5. In order to keep the maximum accuracy, the conversion to 12-bit integer is achieved in the ultimate step prior to the HEVC encoding.

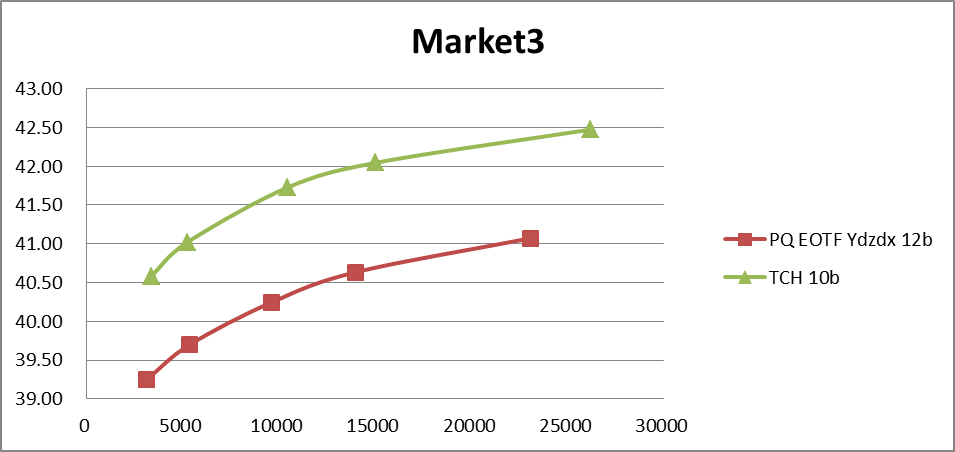


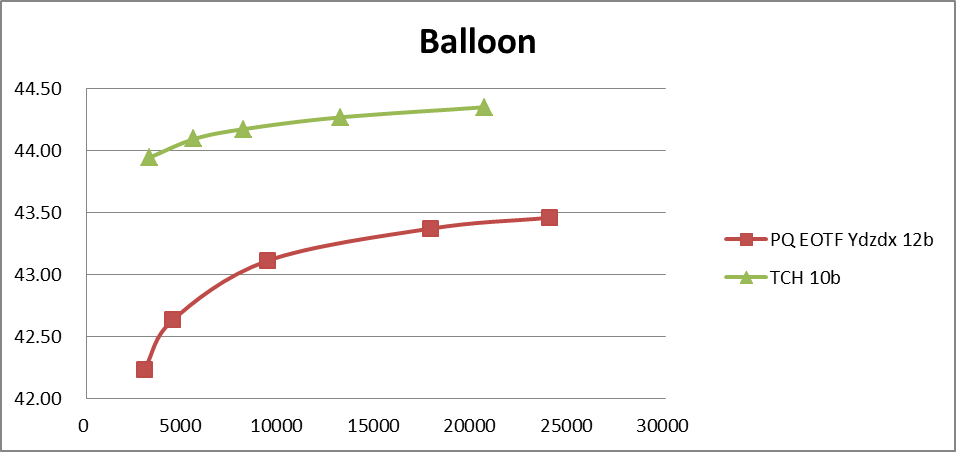


Figure 5: implementation of the PQ EOTF / Ydzdx coding chain.

## Results

The results are reported in the following plots and can also be found in the attached xls file.









# References

[1] E. François, S. Lasserre, F. Le Léannec, Use cases and requirements for HDR and WCG video coding, ISO/IEC JTC1/SC29/WG11 MPEG2014/ m31956, Oct. 2014, San Jose, USA

[2] S. Lasserre, F. Le Léannec, E. François, Description of HDR sequences proposed by Technicolor, ISO/IEC JTC1/SC29/WG11 MPEG2014/ m31957, Oct. 2014, San Jose, USA

[3] S. Lasserre, F. Le Léannec, E. François, Quantitative quality evaluation of images for MPEG XYZ, ISO/IEC JTC1/SC29/WG11 MPEG2014/ m31959, Oct. 2014, San Jose, USA

# Patent rights declaration

**Technicolor may have current or pending patent rights relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation | ISO/IEC International Standard (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).**