A SINGLE-LAYER HDR VIDEO CODING FRAMEWORK WITH SDR COMPATIBILITY

Edouard François[1], Leon van de Kerkhof[2]

[1]Technicolor, France, [2]Philips, Netherlands

ABSTRACT

The migration from High Definition (HD) TV to Ultra High Definition (UHD) is already underway. In addition to an increase of picture spatial resolution, UHD potentially provides more colour by introducing a wider colour gamut (WCG), and better contrast by moving from Standard Dynamic Range (SDR) to High Dynamic Range (HDR). The transition from SDR to HDR will require distribution solutions supporting some level of SDR backward compatibility. This paper presents the HDR content distribution scheme jointly developed by Technicolor and Philips. The solution is based on a single layer codec design and provides SDR compatibility, thanks to a pre-processing step applied prior to the encoding. The resulting SDR video can be compressed and distributed then decoded using standard-compliant decoders (e.g. HEVC Main 10 compliant) and directly rendered on SDR displays. Dynamic metadata of limited size are used to reconstruct the HDR signal from the decoded SDR video, using a post-processing that is the functional inverse of the pre-processing. Both HDR quality and artistic intent are preserved. Pre- and post-processing are applied independently per picture, do not involve any inter-pixel dependency, and are codec agnostic.

INTRODUCTION

The arrival of the High Efficiency Video Coding (HEVC) standard enables the deployment of new video services with enhanced viewing experience, such as Ultra HD broadcast services. In addition to an increased spatial resolution, Ultra HD can bring a wider colour gamut (WCG) and a higher dynamic range (HDR) than the Standard dynamic range (SDR) HD-TV currently deployed. Different solutions for the representation and coding of HDR/WCG video have been proposed [1, 2, 3, 4]. As stated in [5, 6, 7, 8], SDR backward compatibility with decoding and rendering devices is an important feature in some video distribution systems, such as broadcasting or multicasting systems. Dual-layer coding is one solution to support this feature. However, due to its multi-layer design, this solution is not adapted to all distribution workflows. An alternative is to transmit HDR content and to apply at the receiving device an HDR-to-SDR adaptation process (tone mapping). One issue in this scenario is that the tone mapped content may be out of control of the content provider or creator. Another issue is that a new HDR-capable receiving device is needed to apply this tone mapping for existing SDR displays. Alternatively, the Hybrid Log Gamma (HLG) transfer function [2] has been designed as a straightforward solution to address the SDR backward compatibility, that is, an HDR video graded on a display using the HLG transfer function can be in principle directly displayed on an SDR display (using the BT.1886 transfer function [9]) without any adaptation. However, this solution may result in colour shifting when the HLG-graded video is displayed on an SDR rendering device, especially when dealing with content with high dynamic range and peak luminance [10, 11, 12]. Also, there is no way to optimize the brightness and contrast of the SDR image.

Technicolor and Philips have jointly developed a new Single Layer HDR distribution solution, SL-HDR1 [14], aiming at addressing these issues. The solution is SDR compatible and leverages SDR distribution networks and services already in place. It enables both high quality HDR rendering on HDR-enabled CE devices, while also offering high quality SDR rendering on SDR CE devices.

The main features of the HDR distribution system are as follows:

- Single layer with metadata: the HDR system is based on a single layer coding process, with side metadata (of a few bytes per video frame or scene) that can be used as a post-decoding processing stage to reconstruct the HDR signal.
- Distribution codec agnostic: the HDR system is codec independent (a 10 bits codec is recommended).
- Direct SDR compatibility: a decoded bitstream can be directly displayed on an SDR display. A post-processing is applied to convert the decoded SDR picture to HDR, thanks to the metadata, with preservation of the artistic intent.
- Preserved quality of HDR content: there is no visible impairment due to the SDR compatibility feature in comparison with coding of the HDR10 signal.
- Limited complexity: the post-processing of limited complexity can be implemented in a low-cost CE device. The involved operations are strictly pixel-based.
- Independent from the HDR transfer function: the HDR system is independent from the HDR video signal transfer function that is input at the pre-processing stage.
- No dedicated production metadata required: in case of live broadcasting, the SDR signal can be automatically derived by the pre-processing of the distribution encoder.

The remaining of the paper is structured as follows. The solution overview is presented in the next section. Then the HDR decomposition and reconstruction processes are detailed in the two following sections. The next section relates to the metadata signalling. Finally, complexity and performance are commented, and closing remarks are made.

HDR SYSTEM OVERVIEW

Figure 1 shows an end-to-end workflow supporting content production and delivery to HDR and SDR displays. The core of the HDR distribution solution is indicated in yellow and green boxes. It involves a single-layer SDR/HDR encoding-decoding, with side metadata. At the distribution stage, an incoming HDR signal is decomposed in an SDR signal and content-dependent dynamic metadata. The SDR signal is encoded with any distribution codec (e.g. HEVC) and carried throughout the existing SDR distribution network with accompanying metadata conveyed on a specific channel or embedded in the SDR bitstream. The dynamic metadata are typically carried in an SEI message when used in conjunction with an HEVC codec. The post-processing stage is functionally the inverse of the pre-processing and performs the HDR reconstruction. It occurs just after SDR bitstream decoding. The post-processing takes as input an SDR video frame and associated dynamic metadata in order to reconstruct an HDR picture. Single-layer encoding/decoding requires only one encoder instance at HDR encoding side, and one decoder instance at player/display side. It supports the real-time workflow requirements of broadcast applications. The dynamic metadata are produced by the HDR decomposition process and remain internal to the distribution process. They do not need to be conveyed to the rendering device. Additional metadata, originated from the production/post-production, can optionally be distributed and conveyed to the rendering device.

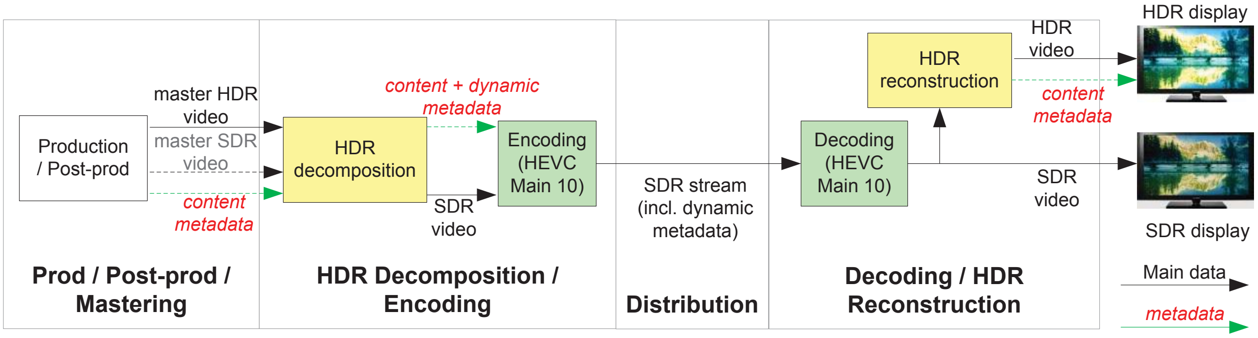

Figure 1 – Example of HDR end-to-end system

The block diagram in Figure 2 depicts in more details the HDR decomposition and reconstruction processes. The centre block included in dash-red box corresponds to the distribution encoding and decoding stages. The left and right grey-coloured boxes respectively enable format adaptation to the input video signal of the HDR system and to the targeted system (e.g. a STB, a connected TV). The yellow boxes show the HDR specific processing. The core component of the HDR decomposition stage is the HDR-to-SDR decomposition that generates an SDR video from the HDR signal. Optionally, gamut mapping may be used when the input HDR and output SDR signals are represented in different colour spaces. The decoder side implements the inverse processes, in particular the SDR-to-HDR reconstruction step that goes back to HDR from the SDR video provided by the decoder.

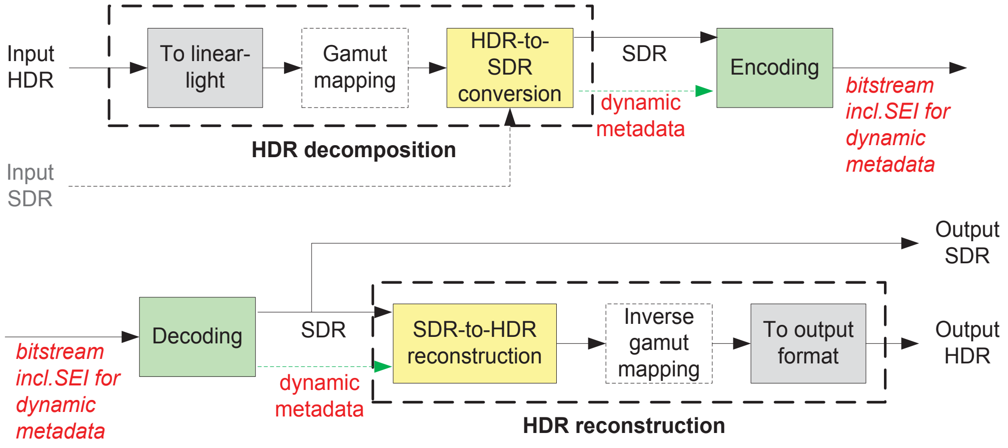

Figure 2 – HDR system architecture overview

HDR-TO-SDR DECOMPOSITION PROCESS

The HDR-to-SDR decomposition process aims at converting the input linear-light 4:4:4 RGB HDR signal to an SDR compatible version. The process uses side information such as the colour primaries and gamut of the container of the HDR and SDR pictures. The process operates without colour gamut change: the HDR and SDR pictures are defined in the same colour gamut. If needed, a gamut mapping pre-processing may be applied to convert the HDR picture from its native colour gamut to the target SDR colour gamut.

The process is depicted in Error! Reference source not found.. It is primarily based on the analysis of the HDR content (picture per picture) in order to derive a set of mapping parameters that will be further used to convert the HDR signal into SDR (step 1 of Figure 3). Once the mapping parameters have been derived, a luminance mapping function, noted *TM,* is obtained, as explained in the next sub-section. The next steps can be summarized as follows.

In step 2, the luminance *L*, derived from the HDR linear-light RGB signal, is mapped to an SDR luma signal using the luminance mapping function *TM* (step 2 of Figure 3):

$$L = A_1 \times \begin{bmatrix} R \\ G \\ B \end{bmatrix} \qquad \text{(eq.1)}$$

$$Y_{tmp} = TM[L] \qquad \text{(eq.2)}$$

Where $A = [A_1 A_2 A_3]^T$ is the conventional 3x3 R'G'B'-to-Y'CbCr conversion matrix (e.g. BT.2020 or BT.709 depending on the colour space), $\mathrm{A}_1, \mathrm{A}_2, \mathrm{A}_3$ being 1x3 matrices.

The chroma components are then derived as follows (step 3 in Figure 3). First the R, G, B values are scaled by the ratio (Y_{tmp} / L), which results in a linear-light SDR version of RGB. Then a square-root is applied, to reproduce a transfer function close to the BT.709 OETF (the usage of a square root guarantees the reversibility of the process). The resulting R, G, B signal is converted to chroma components U_{tmp}, V_{tmp}:

$$\begin{bmatrix} U_{tmp} \\ V_{tmp} \end{bmatrix} = \sqrt{\frac{Y_{tmp}}{L}} \times \begin{bmatrix} A_2 \\ A_3 \end{bmatrix} \times \begin{bmatrix} \sqrt{R} \\ \sqrt{G} \\ \sqrt{B} \end{bmatrix} \qquad \text{(eq.3)}$$

A final colour correction is applied in order to match the SDR colours to the input HDR signal colours (step 4 in Figure 3). First the chroma components are adjusted by a scaling factor $1/\beta(Y_{tmp})$, where $\beta(Y_{tmp})$ is a function that enables to control the colour saturation and hue.

$$\begin{bmatrix} U_{sdr} \\ V_{sdr} \end{bmatrix} = \frac{1}{\beta(Y_{tmp})} \times \begin{bmatrix} U_{tmp} \\ V_{tmp} \end{bmatrix} \qquad \text{(eq.4)}$$

Then the luma component is adjusted to further control the perceived saturation, as follows:

$$Y_{sdr} = Y_{tmp} - Max(0, a \times U_{sdr} + b \times V_{sdr}) \qquad \text{(eq.5)}$$

where *a* and *b* are two control parameters. As demonstrated in [13], this step is fundamental to control the SDR colours and to guarantee their matching to the HDR colours. This is in general not possible when using a fixed transfer function.

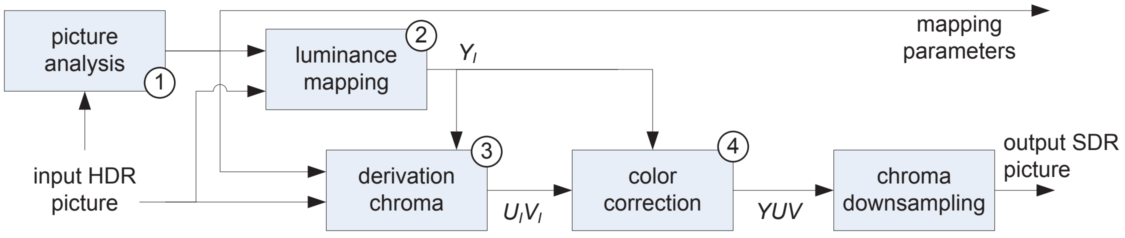

Figure 3 – synoptic of HDR-to-SDR decomposition process

Luminance mapping

The luminance mapping aims at converting the input linear-light luminance signal into an SDR luma signal. It builds the mapping function or look-up-table *TM*. The process is based on a perceptual transfer function, and uses a limited set of control parameters, that have to be further conveyed to the post-processing in order to be able to invert the luminance mapping process. The process works as follows.

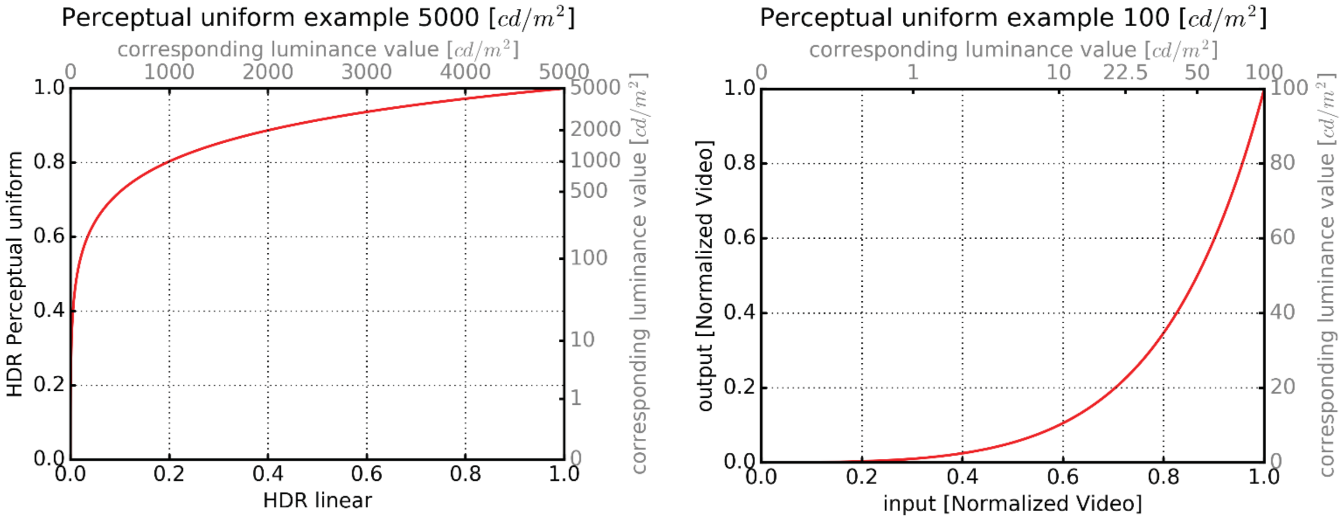

Figure 4 – Example conversion curves for converting from linear light to perceptual domain (left, with peak luminance 5000 cd/m$^{2)}$ and back to SDR linear light (right)

The input linear-light luminance signal L is first converted to the perceptually-uniform domain based on the mastering display peak luminance, using a perceptual transfer function illustrated in left picture of Figure 4. This process is controlled by the mastering display peak luminance parameter. To better control the black and white levels, a signal stretching between content-dependent black and white levels (parameters *blackLevelOffset* and *whiteLevelOffset*) is applied. Then the signal is tone mapped using a piece-wise curve constructed out of three parts, as illustrated in Figure 5. The lower and upper sections are linear, the steepness being determined by the *shadowGain* and *highlightGain* parameters. The mid-section is a parabola providing a smooth bridge between the two linear sections. The width of the cross-over is determined by the *midToneWidthAdjFactor* parameter. The curve can be further fine-tuned using a piece-wise linear corrective function. Then the signal is converted back to the linear light domain based on the targeted SDR display maximum luminance of 100 cd/m^2, as illustrated in the right picture of Figure 4. The resulting signal is the SDR luma.

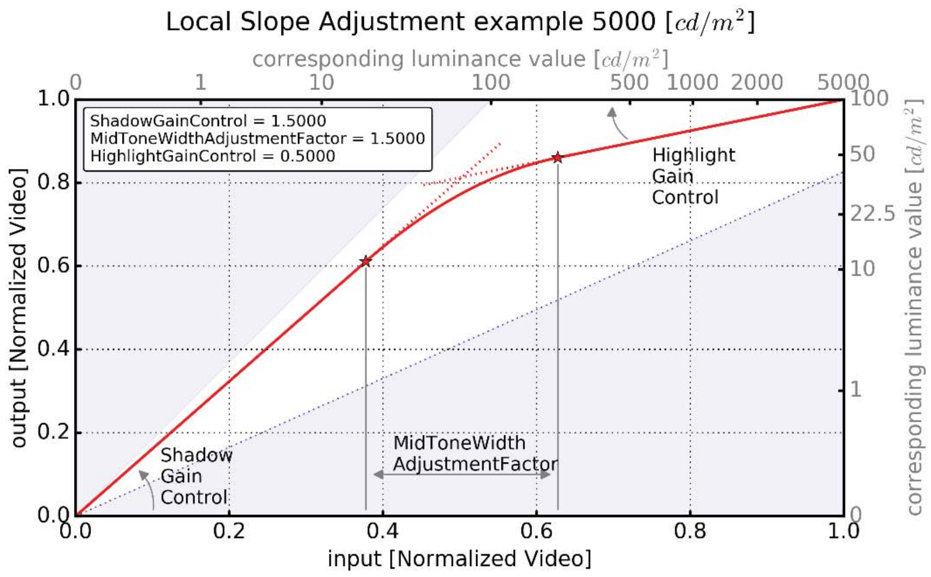

Figure 5 – Tone mapping curve shape example

HDR RECONSTRUCTION

The HDR reconstruction process is depicted in Figure 6. From the input dynamic metadata (detailed in next sub-section) a luma-related look-up table, *lutMapY*, and a colour correction look-up table, *lutCC*, are derived. The next step consists in applying the SDR-to-HDR reconstruction from the input SDR picture, the derived luma-related look-up table and colour correction look-up table. This process produces an output linear-light HDR picture. An optional gamut mapping can be applied when the colour spaces of the SDR picture and of the HDR picture are different.

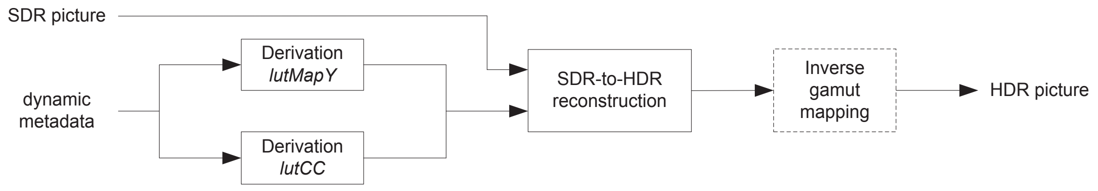

Figure 6 – Overview of the HDR reconstruction process

The SDR-to-HDR reconstruction process is the functional inverse of the decomposition process. However, for implementation complexity reasons, some operations are concatenated or applied in a different order. The LUT *lutMapY* corresponds to the inverse of the square-root of the mapping LUT *TM*. The post-processing colour correction LUT *lutCC* is actually linked to the pre-processing colour correction LUT β and the tone mapping LUT *lutMapY* by the following equation:

$$\beta[Y] = 2^B \times \ lutMapY[Y] \times lutCC[Y] \qquad \text{(eq.6)}$$

where *B* is the bit-depth of the luma signal.

The process performs the following successive steps for each sample *Y*, *U* (Cb component), *V* (Cr component), of the SDR picture. First *U* and *V* are centred (by subtracting the chroma offset, e.g. 512 for a 10 bits signal). Then the variable Y_{post} U_{post} and V_{post} are derived as:

$$Y_{post} = Clamp(0, 2^B - 1, Y + Max(0, a \times U + b \times V) \qquad \text{(eq.7)}$$

$$\begin{bmatrix} U_{post} \\ V_{p} \end{bmatrix} \text{[illegible]} \qquad \text{(eq.8)}$$

Equations don’t copy over - need to add in - saved as screenshot called equations

The reconstruction of the HD[illegible] made of the following steps. A parameter *T* is first compute[illegible]

$$T = k0 \times U_{post} \times V_{p} \text{[illegible]} {}_{post} \times V_{post} \qquad \text{(eq.9)}$$

where $k0, k1, k2$ are predefine[illegible] the coefficients of the R'G'B'-to-Y'CbCr conversion matrix *A* [illegible] R_{im}, G_{im}, B_{im} are derived as follows:

$$\begin{bmatrix} R_{im} \\ G_{im} \\ B_{im} \end{bmatrix} = A^{-1} \times \begin{bmatrix} \sqrt{1 - T} \\ U_{post} \\ V_{post} \end{bmatrix} \qquad \text{(eq.10)}$$

Then, *R*, *G*, *B* are updated by the following equation:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \left(lutMapY[Y_{post}]\right)^2 \times \begin{bmatrix} R_{im}^2 \\ G_{im}^2 \\ B_{im}^2 \end{bmatrix} \qquad \text{(eq.11)}$$

Finally, a clamping is done to *0*, L_{HDR}, where L_{HDR} is the HDR mastering display peak luminance.

It can be demonstrated that equations (eq.9) to (eq.11) invert the pre-processing operation of (eq.1) to (eq.3), that is, the conversion of the HDR version *R*, *G*, *B* into chroma components. When *T* is larger than 1 (which is in principle not possible, but may happen because of quantization and compression), U_{post} and V_{post} are scaled by $1/\sqrt{T}$, the resulting *T* then becoming 1.

METADATA DESCRIPTION

The post-processing uses the LUTs *lutMapY* and *lutCC*, and the parameters *a* and *b*, as dynamic data. These data enable to finely control the texture and colours of the SDR version, and to ensure a good fitting to the HDR intent. The LUTs *lutMapY* and *lutCC* are conveyed either using a limited set of parameters (parameter-based mode), or explicitly coded (table-based mode). In both cases, the metadata payload corresponds to a few bytes per video frame or scene. The parameter-based mode may be of interest for distribution workflows which primary goal is to provide direct SDR backward compatible services with very low additional payload or bandwidth usage for carrying the dynamic metadata. The table-based mode may be of interest for workflows equipped with low-end terminals or when a higher level of adaptation is required for representing properly both HDR and SDR streams.

In the parameter-based mode, the metadata for reconstructing *lutMapY* consist of the parameters mentioned in the section “Luminance mapping”. For reconstructing *lutCC*, a default pre-defined LUT is used at the post-processing side, and a piece-wise linear table made of at most 6 points is used as a scaling function to adjust the default table. These parameters are conveyed using the *Colour Volume Reconstruction Information* (CVRI), based on the SMPTE ST 2094-20 specification. Typical payload is about 25 bytes per scene.

In the table-based mode, *lutMapY* and *lutCC* are explicitly coded using the *Colour Remapping Information* (CRI) standardized in the HEVC and SMPTE ST 2094-30 specifications. Typical payload is about 160 bytes per scene.

In both cases, the metadata are limited to the codec space. They do not come from the production side, and do not need to be conveyed outside the decoding platform. They are conveyed using standardized metadata containers.

Next to the dynamic metadata, the system uses the *Mastering Display Colour Volume* (MDCV), standardized in AVC, HEVC and SMPTE ST 2086 specifications. This is static information (typically fixed per program) required by the post-processing. It comprises the colour gamut of the SDR/HDR signal and the mastering display peak luminance.

The usage of dynamic metadata allows a fine control of the SDR texture (using the tone mapping LUT *lutMapY*) and of colours (using the colour correction LUT *lutCC* and the parameters *a* and *b*). This guarantees the preservation of the HDR texture and colours intent in the SDR version, as illustrated in Figure 7 (the HDR picture has of course been tone mapped to be displayed in the paper). High SDR and HDR video quality is obtained, without any strong limitation of the dynamic range and peak luminance (no limitation to peak luminance of around 1000-1500 nits). This also gives high flexibility which enables to easily adapt the system (for instance thanks to the easy control of the dynamic metadata payload) to the distribution workflow.

Figure 7 – Illustration of the impact of the colour correction process shown on 06_EBU_ZurichAthletics2014_HD_100p_HDR test content (left: HDR picture, middle: SDR picture without colour correction, right: SDR picture with colour correction)

COMMENTS AND CONCLUSIONS

The solution developed by Technicolor and Philips has been designed with a particular focus on low complexity and high performance. The pre- and post-processing are of very low added complexity. The involved operations are pixel-based, without inter-sample or temporal dependency. Actual set-top-boxes implementations have been demonstrated and leading TV and Set Top Box (STB) System-on-Chip (SoC) manufacturers are integrating the technology [14]. The solution has been standardized as ETSI TS 103 433.

REFERENCES

1. SMPTE 2014, High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays. SMPTE ST 2084, 2014.

2. Borer T. and Cotton A. 2015, A display independent high dynamic range television system, International Broadcasting Convention (IBC), September, 2015. http://www.bbc.co.uk/rd/publications/whitepaper309

3. EBU 2016. High-Performance Single Layer Directly Standard Dynamic Range (SDR) Compatible High Dynamic Range (HDR) System for use in Consumer Electronics devices (SL-HDR1). ETSI technical specification ETSI TS 103 433. May 2016.

4. Diaz, R., Blinstein, S. and Qu, S. Integrating HEVC Video Compression with a High Dynamic Range Video Pipeline. SMPTE Motion Imaging Journal, Vol. 125, Issue 1. Feb, 2016, pp 14-21.

5. Luthra A., François E. and Husak W. 2015, Call for Evidence (CfE) for HDR and WCG Video Coding. ISO/IEC JTC1/SC29/WG11 MPEG2015/N15083, October, 2015.

6. Luthra A., François E. and Husak W. 2015, Requirements and Use Cases for HDR and WCG Content Coding. ISO/IEC JTC1/SC29/WG11 MPEG2015/N15084, October, 2015.

7. François E., Fogg, C., He, Y., Li, X., Luthra, A. and Segall, A., High Dynamic Range and Wide Color Gamut Video Coding in HEVC: Status and Potential Future Enhancements. IEEE TCSVT, Vol. 26, Issue 1, July, 2015, pp 63-75.

8. François, E. 2015. Standardizing the landscape, High Dynamic Range & Wide Color Gamut Video. DVB scene. March, 2015, pp. 7.

9. ITU-R 2011, Reference electro-optical transfer function for flat panel displays used in HDTV studio production. Recommendation ITU-R BT.1886. March, 2011.

10. Luthra, A., François, E and van de Kerkhof, L. 2016, Report of HDR Core Experiment 7: On the visual quality of HLG generated HDR and SDR video. Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCTVC-W0027. February, 2016.

11. Pindoria, M., Naccari, M., Borer T. Cotton, A. 2016, Some considerations on hue shifts observed in HLG backward compatible video. Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCTVC-W0119. February, 2016.

12. Holm, J. 2016, Information and Comments on Hybrid Log Gamma. Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCTVC-W0132. February, 2016.

13. Lasserre, S., Le Léannec, F., Poirier, T. and Galpin, F. 2016, Backward Compatible HDR Video Compression System. In Proc. Data Compression Conference. March, 2016, pp 309-318.

14. Technicolor High Dynamic Range Delivery Solution Receives Support From Leading System-On-Chip Manufacturers, Jan, 2016. http://www.technicolor.com/en/who-we-are/press-news-center/press-releases/technicolor-high-dynamic-range-delivery-solution-receives-support-leading-system-chip-manufacturers

ACKNOWLEDGEMENTS

The authors would like to thank their colleagues for their contributions to this work.