CHROMA SCALING FOR HIGH DYNAMIC RANGE VIDEO COMPRESSION

Ronan Boitard[⋆] *Mahsa T. Pourazad*[†‡] *Panos Nasiopoulos*[⋆‡]

[⋆] Elect. and Computer Eng. Department, University of British Columbia, Vancouver, Canada
[†]TELUS Communications Inc., Vancouver, Canada
[‡]The Institute for Computing, Information and Cognitive Systems (ICICS), Vancouver, Canada

ABSTRACT

Color pixel encoding optimizes the conversion of linear physical values of light into integer values. The efficiency of such encoding methods depends on a trade-off between the bit-depth used and the visible distortion introduced by quantization. This efficiency for different color pixel encoding approaches has been evaluated in literature, without considering the fact that before transmission to the end-user, color encoded content needs to be compressed using a video codec. Thus, to be efficient, a color pixel encoding scheme needs not only to achieve the lowest bit-depth, but also to allow for efficient video compression ratio. Yet, when compressing dark video sequences, the most efficient color pixel encoding scheme known as Y'$D_u D_v$ requires much higher bit-rates, hence negating its high encoding efficiency. In this article, we propose a chroma scaling technique that adaptively restricts the bit-depth of the chroma channels for optimized encoding of High Dynamic Range content. Results show that the proposed scaling reduces Y'$D_u D_v$ bit-rate requirements for dark content while preserving its high color accuracy.

***Index Terms*—** HDR, Pixel Encoding, Compression

1. INTRODUCTION

High Dynamic Range (HDR) content and displays are gaining momentum in commercial shows and the broadcasting industry sector. Several international standard organizations, such as MPEG (Motion Picture Expert Group) and SMPTE (Society of Motion Picture Television Engineers), are already working on standardizing techniques to allow this new revolution in digital media to reach the consumer market. However, this requires major effort, since HDR pixel representation is fundamentally different from that of Low Dynamic Range (LDR).

Indeed, HDR pixels correspond to absolute light intensity (measured in cd/m^2) represented by floating point values, while LDR pixels are integer code values whose light intensity is related to the capabilities of the used display [1]. Since the entire distribution pipeline has been devised for integer code values, HDR content needs to be adapted to fit the restricted bit-depth that a codec (coder-decoder) can handle. This is done through color pixel encoding. Optimum encoding should require the smallest number of bits per pixel while minimizing the visibility of contouring artifacts due to quantization into integer values. A color pixel encoding scheme is considered as perceptually lossless over a defined color gamut and dynamic range, if at a targeted bit-depth, no human observer can detect a difference between the encoded and original content.

A recent study [2] has evaluated several color pixel encoding methods with respect to the minimum bit-depth required to encode color patches without visual loss. The results of this study show that content, encoded in Y'$D_u D_v$ on a single bit-depth, oversamples the chroma channels, such that it includes information that is invisible to the human eye. To remove this information (visual noise), we propose a chroma scaling technique that exploits the experimental results reported in [2] to adaptively restrict the bit-depth of the chroma channels for optimized encoding.

The rest of this paper is organized as follows. Section 2 presents the different color pixel encoding schemes considered in this paper along with the results reported in [2]. Then, a chroma scaling technique agnostic to any color pixel encoding scheme is introduced in Section 3. Section 4 evaluates the compression efficiency of different color pixel encoding schemes and assesses the gain brought by the proposed scaling technique. Finally, Section 5 concludes this article and provides some future research opportunities.

2. COLOR PIXEL ENCODING SCHEMES FOR HIGH DYNAMIC RANGE CONTENT

HDR technology, through the use of full gamut color space (e.g., CIE XYZ [3]) in floating point values, matches and can even surpass the human vision system capabilities [1, 4]. However, such a representation requires large amounts of data and thus causes challenges in terms of storage capacity, computational complexity and throughput. Furthermore, image and video processing are devised to process integer valued images. These challenges and issues become a barrier for entering HDR to the consumer market. To this end, color pixel encoding is used to convert floating-point physical values to

978-1-4799-9988-0/16/$31.00 ©2016 IEEE

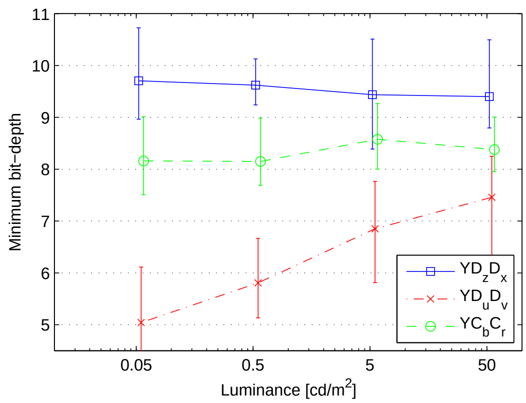

Fig. 1: Minimum bit-depth required to encode chroma channels using $\mathrm{Y'D}_z\mathrm{D}_x$, $\mathrm{Y'C}_b\mathrm{C}_r$ and $\mathrm{Y'D}_u\mathrm{D}_v$.

integer (perceptually encoded) values. Perceptually encoded luminance channel (Y) is traditionally denoted as luma (Y') while chrominance channels are known as chroma.

A recent evaluation compared HDR color pixel encoding with respect to the minimum bit-depth required to avoid quantization artifacts. Investigated color pixel encoding schemes were $\mathrm{Y'C}_b\mathrm{C}_r$, $\mathrm{Y'D}_u\mathrm{D}_v$ and $\mathrm{Y'D}_z\mathrm{D}_x$. The $\mathrm{Y'C}_b\mathrm{C}_r$ approach cannot represent the full visible gamut and is obtained by converting RGB tri-stimulus values using the transformation matrix described in the BT.2020 recommendation [5] (the BT.709 recommendation [6] was not considered). The $\mathrm{Y'D}_u\mathrm{D}_v$ scheme is based on the CIE Lu′v′ color space [7] and is approximately perceptually uniform over the full visible gamut. Finally $\mathrm{Y'D}_z\mathrm{D}_x$ converts pixels represented in the CIE 1931 XYZ color space [3] using a standardized SMPTE transformation matrix [8]. Note that all three encodings rely on the SMPTE ST 2084 [9] luminance encoding, also known as the Perceptual Quantizer (PQ) [10]. PQ has shown to be the most efficient encoding approach, requiring no more than 11 bits to represent gray patches without any visual loss [2] (for natural images it requires no more than 10 bits [10]).

Fig. 1 reports the results of an experiment, conducted in [2], in which observers determined the minimum number of bits required for encoding chroma channels using the three aforementioned encodings. Note that the range of tested luminance varied from 0.05 to 50 cd/m^2 as it is assumed that below 0.05 cd/m^2, few color information can be perceived by the HVS (Human Visual System). Actual measurements above 50 cd/m^2 were not made because current commercial displays are limited in term of reproducing saturated colors at high luminance value. As it is observed in Fig. 1, the $\mathrm{Y'C}_b\mathrm{C}_r$ scheme requires slightly more than 9 bits to encode chroma channels while $\mathrm{Y'D}_u\mathrm{D}_v$ requires from 6 to 9 bits depending on the luminance. As the $\mathrm{Y'D}_z\mathrm{D}_x$ scheme provides the worst results and thus is no longer considered by the SMPTE, we will not include it in this article.

From the above discussion and Fig 1, two observations can be made. First, luma and chroma channels require different bit-depths. Second, for the $\mathrm{Y'D}_u\mathrm{D}_v$ encoding, the optimum bit-depth of the chroma channels varies with the luminance. However, video compression standard such as the ITU-T H.265/MPEG-H Part 2 High Efficiency Video Codec (HEVC) [11] rely on input content quantized on a fixed bit-depth (e.g. 8, 10 or 12 bits usually). Thus when quantizing any content, using a single bit-depth for all channels results in oversampling the chroma information.

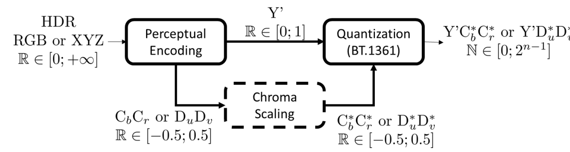

Fig. 2: Workflow of HDR perceptual encoding technique with the proposed chroma scaling method.

3. CHROMA SCALING FOR HIGH DYNAMIC RANGE VIDEO COMPRESSION

In the previous section, we reported the results of an experiment detailing the minimum bit-depth to encode color patches without introducing quantization artifacts. We stated that when quantizing content in $\mathrm{Y'C}_b\mathrm{C}_r$ or $\mathrm{Y'D}_u\mathrm{D}_v$, the chroma channels are usually oversampled compared to the luma. Thus we propose to adjust the number of code values required to quantize chroma channels so that it matches the minimum bit-depth reported in [2].

The workflow of the modified HDR perceptual encoding is depicted in Fig. 2. Note that our technique aims at reducing the number of code values, thus the scaling is always smaller or equal to 1. The performed scaling can be expressed in a generic fashion by:

$$\mathbf{C^*} = min(1, 2^{\mathcal{F}(\mathbf{Y})-n}\mathbf{C}), \tag{1}$$

where $\mathbf{C}$ and $\mathbf{C^*}$ are the original and scaled chroma channels, respectively, while n is the targeted bit-depth. $\mathcal{F}(\mathbf{Y})$ is a scaling function depending on the chosen encoding. Note that if $\mathcal{F}(\mathbf{Y}) >= n$), no scaling is performed.

Fig.1 shows that for the $\mathrm{Y'C}_b\mathrm{C}_r$ encoding, the bit-depth is independent of the luminance value. The scaling of the chroma channel is then simply:

$$\mathcal{F}(\mathbf{Y}) = p, \tag{2}$$

where p is the bit-depth corresponding to the restricting range for the chroma value taken from Fig. 1. For example, a conservative approach would be to choose $p = 10$ while a more aggressive one would be $p = 9$. Recall that p is always smaller or equal to the targeted bit-depth n.

For the $\mathrm{Y'D}_u\mathrm{D}_v$ encoding, Fig.1 shows that the bit-depth

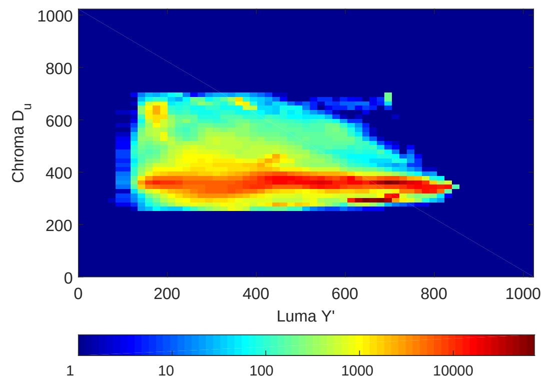

Fig. 3: Luma Y' versus chroma D_u for the first frame of the Market3 sequence. Channels are quantized on 10 bits.

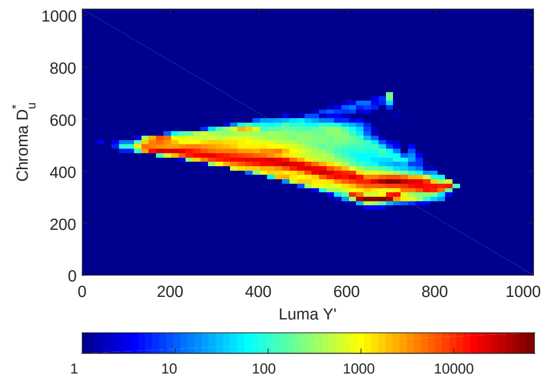

Fig. 4: Luma Y' versus scaled chroma D_u^* for the first frame of the Market3 sequence. Channels are quantized on 10 bits.

should vary linearly in the log domain from 0.05 to 50 cd/m^2:

$$\mathcal{F}(\mathbf{Y}) = a \cdot log(\mathbf{Y}) + b = \frac{p(50) - p(0.05)}{log(50/0.05)} log(\mathbf{Y}) + b, \quad (3)$$

where $p(50)$ and $p(0.05)$ correspond to the chosen bit-depth at 50 and 0.05 cd/m^2, respectively, and b is computed so as to ensure that $\mathcal{F}(50) = p(50)$. A conservative approach would be to choose $p(50) = 9$, $p(0.05) = 6$ and $b = 7.301$. Using these values, we plot in Figs. 3 and 4, the distribution of the D_u channel depending of the luma value before and after performing the scaling. We observe that the number of used code values is reduced for the scaled chroma (Fig. 4). It is evident that chroma scaling efficiently restricts the range of the code values used to better match the limitations of the HVS. The resulted perceptually encoded content is still represented with the same number of bits, as the minimum bit-depth is based on the luma channel requirements. Note that reducing the range of code value increases the quantization loss, even though this loss corresponds to visual noise. Consequently, using chroma scaling has higher distortion compared to the original content when distortion is measured using metrics such as the MSE (Mean Square of Error). The assumption behind perceptual encoding though is that, although the distortion is higher, the visual quality is the same since only visual noise has been removed. Furthermore, the compression of processed content, using a codec, should be more efficient since code values are easier to predict and a smaller amount of invisible high frequencies is present in the videos. This assumption is similar to the work in [12], where the dynamic range of HDR content is restricted to preserve temporal coherency and hence increase temporal prediction between frames. The next section assesses the gain that the chroma scaling technique can bring in term of compression efficiency.

4. COMPRESSION EFFICIENCY

To evaluate the compression efficiency, we encoded 4 HDR video sequences considered by the MPEG Call for Evidence (CfE) for HDR and Wide Color Gamut (WCG) Video Coding [13] using HEVC (HM 16.6). We tested four encoding schemes: Y'C_bC_r (BT.2020), Y'$C_b^*C_r^*$ (BT.2020 with $p = 9$), Y'D_uD_v and Y'$D_u^*D_v^*$ (with $p(50) = 9$, $p(0.05) = 6$ and $b = 7.301$). As Y'C_bC_r is known to have color artifacts when using 4:2:0 chroma sampling [14], all videos were compressed using full chroma sampling 4:4:4. Each compressed content was perceptually decoded (inverse of perceptual encoding in Fig. 2) and converted to the CIE XYZ color space [3] before running any metrics. As there is no consensus on which metric should be used to test HDR video compression, we chose, for the luminance channel, to compute the HDR-VDP 2.2 [15], as this metric takes into account most of the HVS limitations. HDR-VDP 2.2 attempt to predict the Mean Opinion Score (MOS) of subjective tests. However, HDR-VDP does not take into account colors, hence we computed the ΔE_{00} metric which reports distortion between two different colors [16]. It is usually considered that a distortion is visible if the $\Delta E_{00} > 1$.

Fig. 5 plots the compression results for a nighttime sequence (FireEater2). As it oversamples the chroma channels in low luminance values (use of 10 bits when only 6 are necessary), Y'D_uD_v requires much more bandwidth compared to Y'C_bC_r. Using Y'$D_u^*D_v^*$ allows to bring back the bandwidth to similar levels with Y'C_bC_r while achieving the same HDR-VDP MOS and color distortion ($\Delta E_{00} < 1$). Note that the use of Y'$C_b^*C_r^*$ does not affect the compression efficiency.

Fig. 6 plots the results for a broad daylight video sequence (Market3). Y'D_uD_v achieves a lower color distortion (≈ 0.5 reduction in ΔE_{00}), however it reduces the HDR-VDP MOS score. Y'$D_u^*D_v^*$ is in between the Y'C_bC_r and Y'D_uD_v for both color and luma metrics. Note that encoding using Y'$C_b^*C_r^*$ results in a great increase in the color distortion, showing that more than 9 bits are required to encode chroma channels in Y'C_bC_r. As the results for the daylight scene BalloonFestival are similar to Market3 and since those from the computer generated Tibul2 exhibited no difference between the different encodings, they were not plotted.

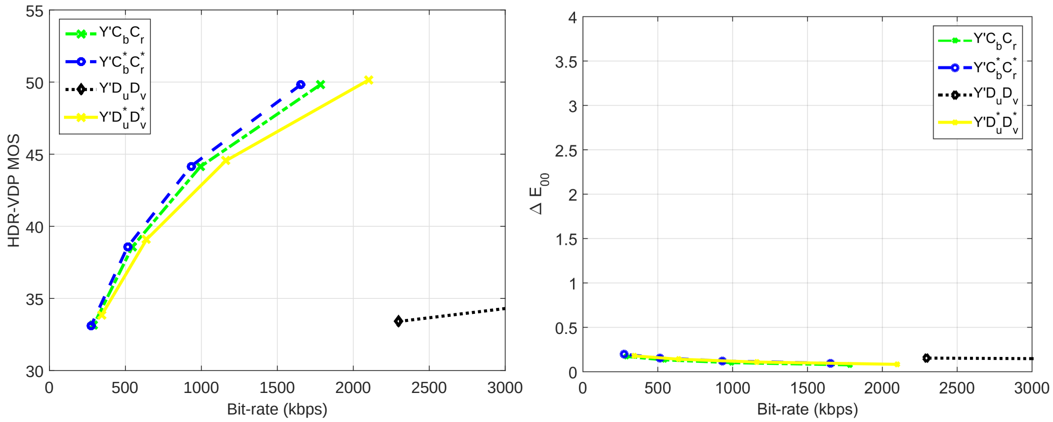

Fig. 5: HDR-VDP MOS and ΔE_{00} for FireEater2 [17]. Higher bit-rates of Y'D$_u$D$_v$ were not plotted for better readability.

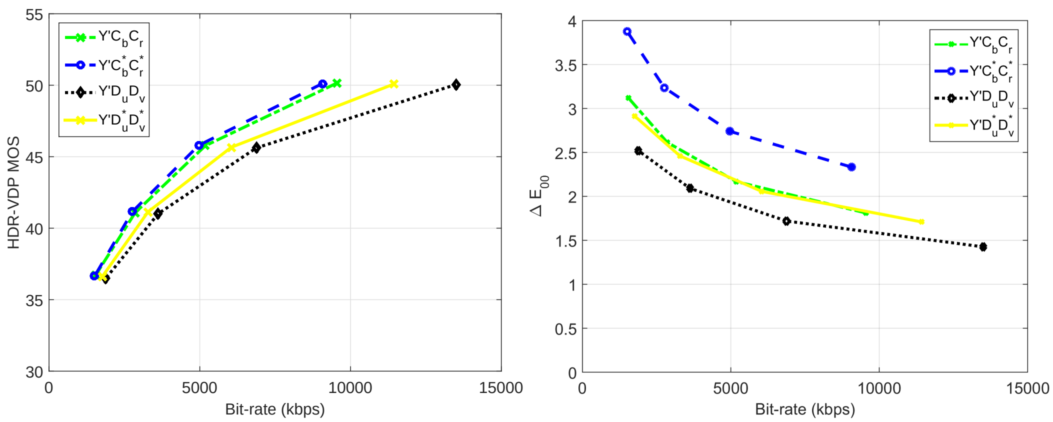

Fig. 6: HDR-VDP MOS and ΔE_{00} for Market3 [17].

These results show that the shortcomings in nighttime scenes encoded with Y'D$_u$D$_v$ can be solved through the use of our proposed scaling, while preserving a good efficiency for daytime scenes. Regarding Y'C$_b^*$C$_r^*$, the results indicate that scaling with $p = 9$ decreases the compression efficiency, especially for color reproduction. However, we chose $p = 9$ because videos were encoded on 10 bits. We expect Y'C$_b^*$C$_r^*$ to bring higher compression efficiency if the videos were to be encoded on 12 bits and $p = 10$ was used.

5. CONCLUSION

In this article, we proposed to scale color pixel encoding schemes in order to remove visual noise and thus improve their compression efficiency. A special focus was put on Y'D$_u$D$_v$, since it is based on the CIE Lu′v′ color space [7], which exhibits many advantages for video processing, such as perceptual uniformity and low bit-depth requirements. The proposed scaling reduces Y'D$_u$D$_v$'s main drawback, which is its high bandwidth requirements for nighttime scenes due to the oversampling of chroma channels.

The results of scaling Y'C$_b$C$_r$ show that using fewer than 10 bits, decreases the content color reproduction quality. However, for content encoded on more than 10 bits (i.e., 12 or 14 bits), the chroma scaling proposed in this article could greatly improve compression efficiency. Further tests with high-bit depth should be conducted to assess the potential gain. Note that albeit Y'C$_b$C$_r$ overall provides better results than Y'D$_u$D$_v$, Y'C$_b$C$_r$ is known to produce artifacts when performing chroma subsampling with HDR content, which are not present when using Y'D$_u$D$_v$. Thus using Y'D$_u$D$_v$ with chroma scaling is a simple alternative solution for preventing/reducing chroma artifacts.

Finally, the results presented in this paper are based on objective metrics. As the chroma scaling technique aims at exploiting the limitations of the HVS, using a distortion metrics such as the ΔE_{00} most likely penalize the proposed method. Indeed, the proposed scaling removes invisible information but increases the quantization loss which in turn increases the distortion metrics. A subjective evaluation conducted with human observers could possibly reveal higher compression gain for the proposed scaling and a more pertinent score for the color accuracy.

6. REFERENCES

[1] Ronan Boitard, Mahsa T. Pourazad, Panos Nasiopoulos, and Jim Slevinsky, "Demystifying high-dynamic-range technology: A new evolution in digital media.," *IEEE Consumer Electronics Magazine*, vol. 4, no. 4, pp. 72–86, oct 2015.

[2] Ronan Boitard, Rafał K. Mantiuk, and Tania Pouli, "Evaluation of color encodings for high dynamic range pixels," in *Proc. SPIE 9394, Human Vision and Electronic Imaging XX*, 2015, pp. 1–9.

[3] T Smith and J Guild, "The c.i.e. colorimetric standards and their use," *Transactions of the Optical Society*, vol. 33, no. 3, pp. 73, 1931.

[4] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumant Pattanaik, Greg Ward, and Karol Myszkowski, *High Dynamic Range Imaging, 2nd Edition: Acquisition, Display, and Image-Based Lighting*, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 2010.

[5] ITU, "Recommendation itu-r bt.2020: Parameter values for ultra-high definition television systems for production and international programme exchange," International Telecommunications Union, 2012.

[6] ITU, "Recommendation itu-r bt.709-3: Parameter values for the hdtv standards for production and international programme exchange," International Telecommunications Union, 1998.

[7] Greg Ward Larson, "Logluv encoding for full-gamut, high-dynamic range images," *Journal of Graphics Tools*, vol. 3, no. 1, pp. 815–30, 1998.

[8] SMPTE, "Ydzdx color-difference encoding for xyz signals," Society of Motion Picture and Television Engineers SMPTE ST 2085, 2014.

[9] Society of Motion Picture & Television Engineers, "High dynamic range electro- optical transfer function of mastering reference displays table," in *SMPTE ST 2084*. 2014, pp. 1–14, SMPTE.

[10] Scott Miller, Mahdi Nezamabadi, and Scott Daly, "Perceptual signal coding for more efficient usage of bit codes," *SMPTE Motion Imaging Journal*, vol. 122, no. 4, pp. 52–59, May 2013.

[11] Mahsa Pourazad, Colin Doutre, Maryam Azimi, and Panos Nasiopoulos, "Hevc: The new gold standard for video compression: How does hevc compare with h.264/avc?," *IEEE Consumer Electronics Magazine*, vol. 1, no. June, pp. 36–46, 2012.

[12] Ronan Boitard, Dominique Thoreau, Remi Cozot, and Kadi Bouatouch, "Impact of temporal coherence-based tone mapping on video compression," in *Proceedings of the 21st European Signal Processing Conference (EUSIPCO)*, Marrakech, Morocco, 2013.

[13] Ajay Luthra, Edouard Francois, and Walt Husak, "Call for evidence (cfe) for hdr and wcg video coding," in *ISO/IEC JTC1/SC29/WG11 MPEG2014/N15083*, Geneva, Switzerland, 2015.

[14] Jacob Strom, Jonatan Samuelsson, Martin Pettersson, Kenneth Andersson, Per Wennersten, and Rickard Sjöberg, "Ericssons response to cfe for hdr and wcg," in *ISO/IEC JTC1/SC29/WG11 MPEG2015/M36184*, 2015.

[15] Manish Narwaria, Rafał K. Mantiuk, Matthieu Perreira Da Silva, and Patrick Le Callet, "Hdr-vdp-2 . 2 : a calibrated method for objective quality prediction of high-dynamic range and standard images," *Journal of Electronic Imaging*, vol. 24, no. 1, 2015.

[16] Gaurav Sharma, Wencheng Wu, and Edul N. Dalal, "The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations," *Color Research and Application*, vol. 30, no. 1, pp. 21–30, 2005.

[17] Sébastien Lasserre, Fabrice LeLéannec, and Edouard Francois, "Description of hdr sequences proposed by technicolor," in *ISO/IEC JTC1/SC29/WG11 JCTVC-P0228*, San Jose, USA, 2013, IEEE.