Visual Color Difference Evaluation of Standard Color Pixel Representations for High Dynamic Range Video Compression

Maryam Azimi, Ronan Boitard, Panos Nasiopoulos
Electrical and Computer Engineering Department,
University of British Columbia,
Vancouver, Canada

Mahsa. T. Pourazad
TELUS communications Inc.,
Vancouver, Canada

***Abstract*— With the recent introduction of High Dynamic Range (HDR) and Wide Color Gamut (WCG) technologies, viewers' quality of experience is enriched with higher brightness and wider color range. To distribute HDR videos over a transmission pipeline, color pixels need to be quantized into integer code-words. Linear quantization is not optimal since the Human Visual System (HVS) do not perceive light in a linear fashion. Thus, perceptual transfer functions (PTFs) and color pixel representations are used to convert linear light and color values into a non-linear domain, corresponding more closely to the response of the human eye. In this work, we measure the visual color differences caused by different PTFs and color representation for 10-bit quantization. Our study encompasses all the visible colors of the BT.2020 gamut for different luminance levels. Visual color differences are predicted using a perceptual color error metric (CIE ΔE2000). Results show that visible distortion can already occur before any type of video compression is performed on the signal and that choosing the right PTF and color representation can greatly reduce these distortions.**

Keywords*— *HDR, Color difference, Perceptual transfer function, Color pixel representation, Quantization.

I. INTRODUCTION

The emerging High Dynamic Range (HDR) technology has enormously increased the viewers' quality of experience by enriching video content with higher brightness and wider color range. HDR technology's broad range of brightness is represented by floating-point values. To compress and transmit HDR content, pixel values need to be first transformed from floating point values to integer-coded ones through perceptual transfer functions (PTF) and bit-depth quantization, as the current video transmission pipeline is based on integer inputs. If this lossy transformation is not perceptually optimized, that is to say taking advantage of the Human Visual System (HVS) limitations, it will produce visible artifacts to the signal, even before video compression.

An efficient PTF quantizes the physical luminance information of a captured scene, such that only information invisible to the human eye is excluded. Previously, 8-bit quantization was deemed sufficient for the brightness range supported by Standard Dynamic Range (SDR) technology (i.e., 0.1 to 100 cd/m^2). However, for representing HDR's wider range of luminance, i.e., 0.005 to 10,000 cd/m^2 [1], the minimum bit-depth requirement needed to be increased to avoid compromising visual quality. In cases where no limitation on bit-depth value is imposed, a point of no visible error can be reached [1] [2], however current infrastructures of video transmission only support 8 and/or 10-bit signals.

Towards the standardization of an HDR video distribution pipeline, the ITU-R BT.2100 [3] recommends two PTFs, namely the Perpetual Quantizer (PQ) and Hybrid Log-Gamma (HLG), as well as two color pixel representations, namely YC_bC_r and IC_tC_p. The BT.2100 standard recommends 10 or 12-bit (for future pipeline) quantization. Currently, 10-bit quantization is the bit-depth defined in HDR10, a profile recommended by the Motion Picture Experts Group (MPEG) for HDR video compression.

While the final quality of the video transmitted through the recommended pipeline has been evaluated comprehensively in MPEG [4][5], the effect of suggested PTFs and color pixel representations by BT.2100, on the perceptual color quality of the 10-bit encoded signal has not been studied in depth. By 'encoded' here and for the rest of the paper we refer to the signal that has been transformed through PTF and not the video compressed signal.

In this work, we evaluate the perceptual color difference of the non-linear quantized color signal across different PTFs and color pixel representations compared to the original linear ones. We sample visible colors based on their luminance levels and across the u'v' chromaticity plane but only consider colors lying in the BT.2020 gamut [6]. In order to isolate the error caused solely by quantization, we do not apply compression or chroma subsampling on the signal. To evaluate the color errors, we rely on a perceptual objective color difference metric, which is based on Human Visual System (HVS) characteristics. Figure 1 shows the general workflow of the evaluation process, while this figure is discussed in more detail in Section III.

The rest of the paper is organized as follows. Section II provides details on standard HDR PTFs and color pixels representations for distribution. Section III includes details on

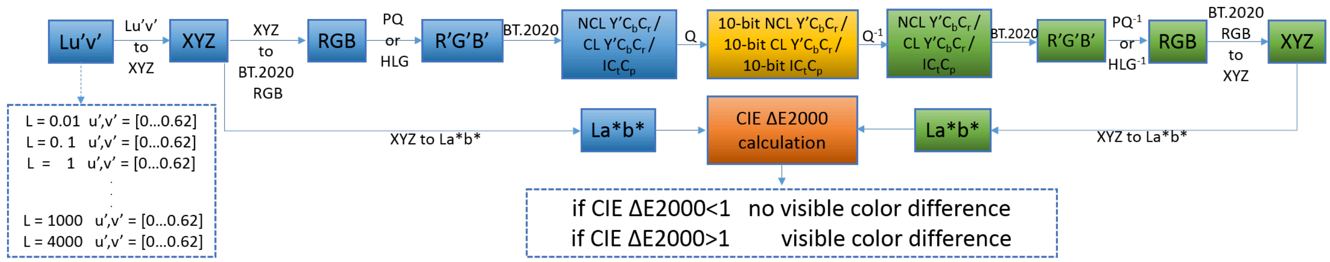

Figure 1. Color difference experiment workflow

our setup and discusses the reported evaluation results. Conclusions are drawn in Section IV.

II. BACKGROUND

A. HDR Perceptual Transfer Functions

Considering that the HVS does not perceive and interpret light in a linear way, perceptual (and hence non-linear) transfer functions are used to transfer the physical linear light values of a scene to values that coincide with the HVS perceptual characteristics. The conventional SDR perceptual transfer function is gamma encoding, standardized as ITU-R BT.1886 [7]. Gamma encoding is not efficient for HDR, since it was designed specifically for the SDR luminance range (i.e., 0.1 to 100 cd/m^2). In addition, the HVS response diverts from a gamma function behavior at higher luminance levels covered by HDR technology. To overcome the gamma function limitations for HDR content, a Hybrid-Log-Gamma function was introduced in [8], and later standardized by the Association of Radio Industries and Businesses (ARIB) as ARIB STD-B67 [9]. This function combines a conventional gamma function for dark areas, and a logarithmic function for bright areas.

In [10], another perceptual transfer function was derived using the peak sensitivities of the Barten Contrast Sensitivity Function (CSF) model [11]. This transfer function is usually referred to as Perceptual Quantizer (PQ) and it has been standardized by the Society of Motion Pictures and Television Engineers (SMPTE) as SMPTE ST 2084 [12].

PQ is designed for luminance values range from 0.005 to 10,000 cd/m^2 and its code-words allocation does not change according to the peak luminance of the content, as long as the content peak luminance falls in this range. However, HLG is mainly designed for the range of luminance supported by current reference grading displays, mainly 0.01 to 1000 cd/m^2 (or 4,000 cd/m^2). Therefore, its code-words allocation varies depending on the maximum peak luminance of the graded content. It is also worth mentioning that PQ is designed to better address HDR displays while HLG's conventional gamma function gives more emphasis on SDR legacy displays.

B. HDR Color Pixel Representations

Since the human eye is more sensitive to luminance than chrominance, it is common practice to de-correlate chroma from luminance in order to compress the chroma channels much more than the luminance without having a huge impact on the overall quality. Presently, video distribution pipelines convert RGB color channels to YC_bC_r, with Y being the luminance channel, and C_b and C_r being respectively blue and red difference channels. YC_bC_r color representation is used in all video compression standards including HEVC [13].

There are two versions of YC_bC_r based on how a PTF is applied on the original linear RGB signal to obtain a 10-bit $Y'C_bC_r$ (the prime represents that the channel has been encoded using a PTF and no longer correspond to linear light values): Non-constant Luminance (NCL) YC_bC_r and Constant Luminance (CL) YC_bC_r. The former applies the PTF on each linear RGB channels to obtain R'G'B and derive Y' from these encoded R'G'B'. The latter one relies on linear RGB values to derive Y and then applies the PTF on Y to obtain encoded Y'.

NCL is the conventional approach that is widely adopted for video distribution pipeline to derive $Y'C_bC_r$. However, it has been shown in [14] that the NCL approach will cause visible artifacts on the encoded and transmitted HDR, which could have been avoided with the CL approach.

Although YC_bC_r de-correlates luminance from chroma, its Y channel is still correlated with C_b and C_r [15]. That means that any changes in Y will eventually affect the color, resulting in color shift between the original and the decoded signals. The IC_tC_p color space, proposed first in [15], is a color pixel representation for HDR, which claims to achieve better de-correlation between intensity and chroma information, closely matching the HVS perceptual mechanism.

III. COLOR DIFFERENCE EVALUATION EXPERIMENTS

In this work, we investigate how the PTFs and color pixel representations recommended in ITU-R BT.2100 [3] alter each color perceptually. The PTFs used are PQ and HLG while the color pixel representations are NCL $Y'C_bC_r$, CL $Y'C_bC_r$, and IC_tC_p. Since neither compression nor chroma sub-sampling is applied on the signals, the generated errors are due to quantization only (see Figure 1). Please note that in this work we only consider signal transmission application therefore, 10-bit BT.2020 colors. The 10-bit quantization performed throughout this test follows the restricted range quantization as described in BT.2100.

Our test encompasses all visible colors representable with BT. 2020 and for luminance levels ranging from 0.01 to 1,000 cd/m^2, and 4,000 cd/m^2. To construct these colors we start with CIE 1976 Lu'v' color space due to its perceptual uniformity. For each luminance level, while L is constant, the u' and v' values are increased from 0 to 0.62 with step size of 0.001. According to [16], chromaticity changes lower than 0.45/410 ~= 0.001 are imperceptible to the human eye. The tested PTFs and color pixel representations are applied on the constructed colors, followed by 10-bit quantization. Please see Figure 1 for the complete workflow. The reason for choosing two maximum luminance values of 1,000 and 4,000 cd/m^2 is that these values correspond to the peak luminance of currently available reference displays.

To evaluate the color deviations from the original signal (blue boxes in Figure 1) and the tested signal (green boxes in Figure 1), we employ the perceptual objective metric of CIE ΔE2000 [17]. This metric is designed to work on CIE 1976 L*a*b* color space (CIELAB) values. For this reason, the original and the encoded signals are transformed to this color space for comparison (see Figure 1). The Just Noticeable Difference (JND) threshold in terms of CIE ΔE2000 is one. In other words, any color difference less than 1 is not perceptible by human eyes. Moreover, the larger the value of the CIE ΔE2000 metric is, the more different the tested colors are perceptually.

Figures 2 and 3 show errors generated due to 10-bit NCL Y'C_bC_r and 10-bit CL Y'C_bC_r color encoding, respectively, at luminance levels of 0.01, 0.1, 1, 10, 100, 500 and 1,000 cd/m^2, with PQ as the PTF. We demonstrate the CIE ΔE2000 values using a color error bar system where dark blue corresponds to values less than a JND (below 1). Therefore, as soon as a light blue is shown it represents a visible color distortion. We have not included all the tested luminance levels in this manuscript, but the complete set of results are publically available at [8]. Please note the loss of colors at luminance level of 0.01, and 1,000 cd/m^2 are due to the clipping enforced by luminance level, which is 10,000 in case of PQ (refer to Y derivation formula in BT. 2020 for more details [6]). As it can be observed the color errors are mainly around the white point. It is well known that HVS is more sensitive to changes in brightness. As the colors around the white point are brighter, any change due to quantization is more visible (and hence larger CIE ΔE2000 value). This observation is consistent throughout our experiment when color error is measured in the Y'C_bC_r color space.

Comparing the results in Figures 2 and 3, we observe that by simply changing from NCL to CL Y'C_bC_r, color errors are reduced and are less noticeable. This reduction in color errors is more evident with red and blue combinations. That is because the CL Y' is more de-correlated from the C_b and C_r (red and blue difference from Y' [6]) compared to NCL Y'. As a result, changing NCL Y' to CL Y' makes the reconstruction of blue and red channels more error-resilient.

Figures 4 and 5 are also showing 10-bit NCL Y'C_bC_r and 10-bit CL Y'C_bC_r color pixel representation, respectively, with HLG as the PTF where the reference display peak luminance is assumed to be equal to 4,000 cd/m^2. Figures 6 and 7 are similar to Figures 4 and 5 with the exception that the reference display peak luminance of 1,000 cd/m^2 is assumed to be 4,000 cd/m^2. The errors that are generated with HLG at high luminance levels (L = 500 and 1,000 for Figure 4 and 5 and L = 100 and, 500 and 1,000 for Figures 6 and 7) are due to the clipping enforced by reference display luminance level. Note that the same errors will also happen with PQ if it is assumed that the content was mastered on a grading display before encoding.

Comparing the results of CL and NCL (compare Figure 4 with Figure 5 and Figure 6 with Figure 7), we found that the color errors are reduced and are less noticeable in the case of CL. This observation is consistent with the one derived when PQ PTF is used (comparing Figure 2 and Figure 3). The rest of the errors present in Y'C_bC_r encoding, even when using CL method at the different luminance levels (See Figure 5 and 7), are due to quantization and the correlation of Y' with C_b and C_r. Quantization errors are the result of limited number of code words assigned to each luminance levels. By comparing HLG and PQ at each luminance level (compare Figure 2 with Figures 4 and 6, and Figure 3 with Figures 4 and 5), it can be observed that PQ outperforms HLG at dark luminance levels (up to 100 cd/m^2) in the Y'C_bC_r color space. This behavior can be explained by the fact that HLG consists of a gamma

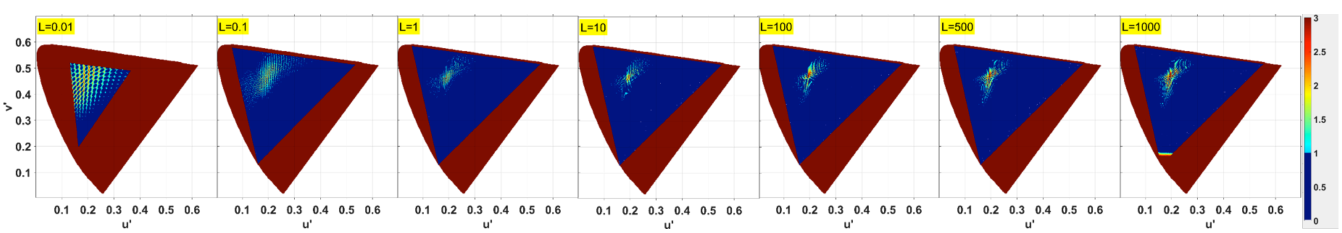

Figure 2. 10-bit NCL Y'C_bC_r with PQ

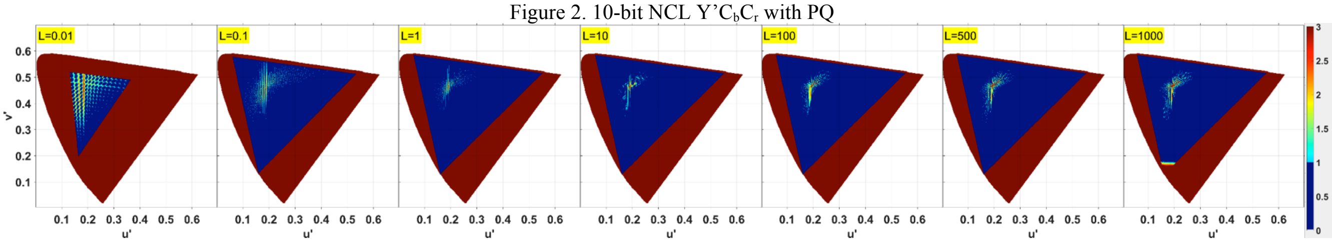

Figure 3. 10-bit CL Y'C_bC_r with PQ

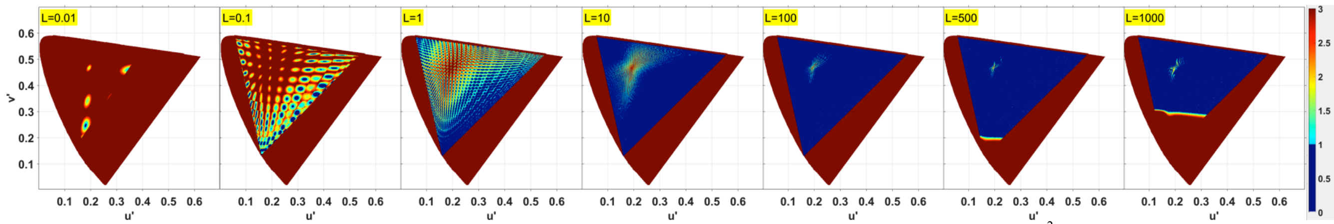

Figure 4. 10-bit NCL $Y'C_bC_r$ with HLG (reference display peak luminance of 4000 cd/m^2)

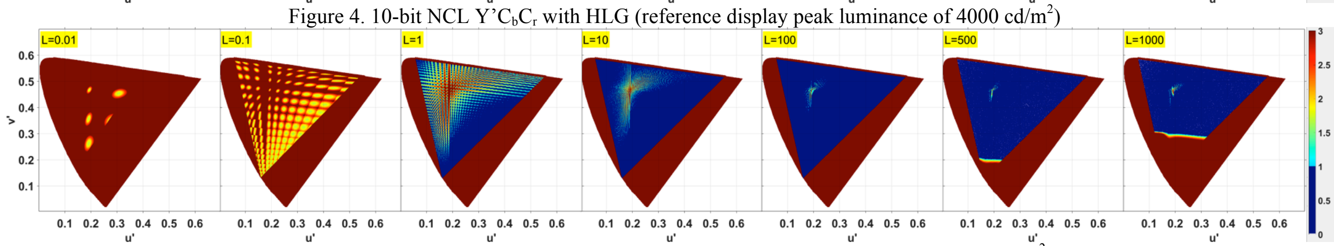

Figure 5. 10-bit CL $Y'C_bC_r$ with HLG (reference display peak luminance of 4000 cd/m^2)

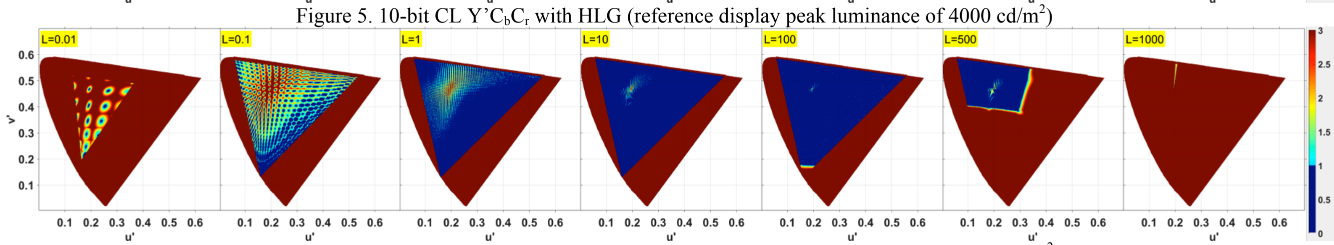

Figure 6. 10-bit NCL $Y'C_bC_r$ with HLG (reference display peak luminance of 1000 cd/m^2)

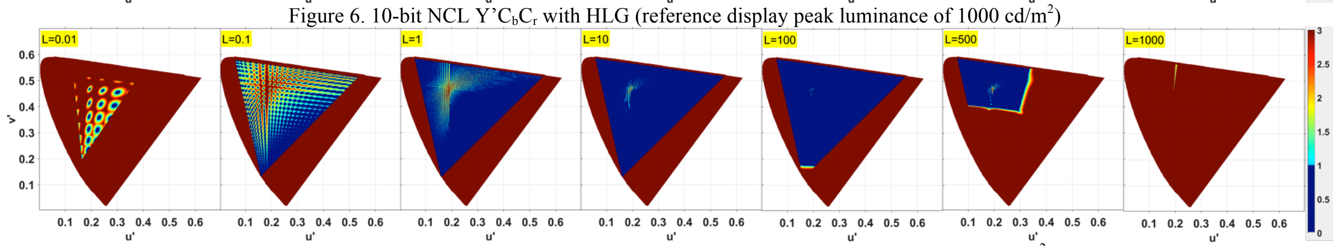

Figure 7. 10-bit CL $Y'C_bC_r$ with HLG (reference display peak luminance of 1000 cd/m^2)

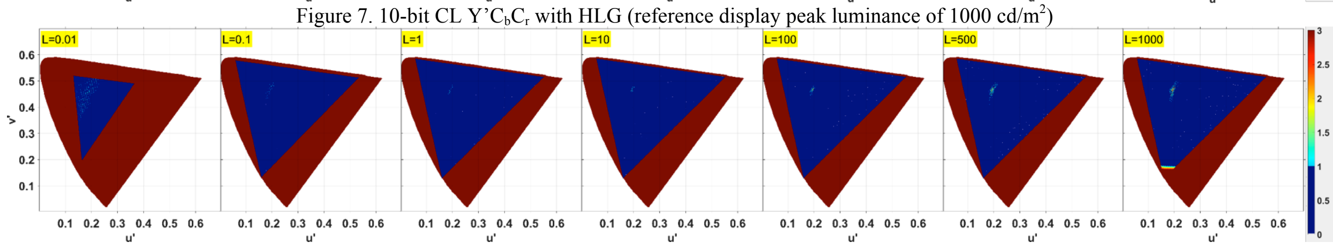

Figure 8. 10-bit IC_TC_P with PQ

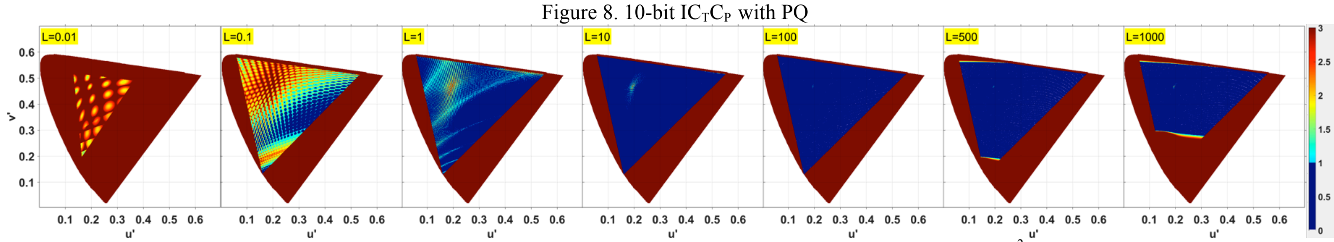

Figure 9. 10-bit IC_TC_P with HLG (reference display peak luminance of 4000 cd/m^2)

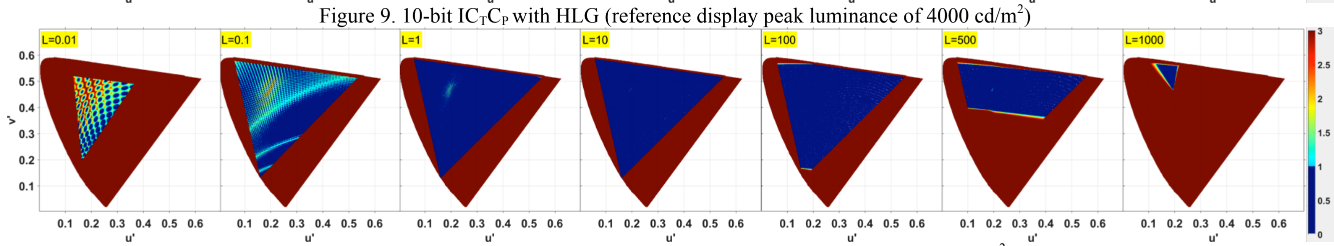

Figure 10. 10-bit IC_TC_P HLG(reference display peak luminance of 1000 cd/m^2)

function for dark areas and a logarithmic one for bright areas. This results in fewer code-words for the dark areas compared to the bright areas. This also explains why HLG is producing fewer errors at high luminance levels compared to PQ (compare luminance levels of 100 cd/m^2 in Figures 2 and 3 with Figures 4, 5, 6 and 7).

Another note-worthy observation is how HLG preforms based on the peak luminance of the display by comparing Figure 6 with 8 in CL case (or 5 and 7 in the NCL case). With HLG at reference display peak luminance of 1,000, more code words are allocated to dark areas, as the content range is normalized to a smaller value compared to the case of reference display peak luminance of 4,000 cd/m^2. This behavior-change of HLG at different peak luminance levels does not happen with PQ, as the latter always assumes a peak luminance of 10,000 cd/m^2.

Please note that in BT.2100, it is suggested to apply clipping on HLG signals that are out of [0, 1] range, at the display side. However, since addressing the display is out of the scope of this paper, we did not clip the encoded signal to [0, 1] range.

Finally, Figures 8, 9 and 10 are showing color errors generated by the IC_tC_p color encoding paired with PQ with peak luminance 4,000 cd/m^2 and HLG with peak luminance of 1,000 cd/m^2. As it can be observed, IC_tC_p with PQ can represent most of the colors without any visible error at the majority of the luminance levels. As Figure 9 shows, since IC_tC_p de-correlates the chrominance channels from luminance channels quite well (see [15]), when using PQ the errors are mainly due to the quantization and are centered at the white point. When HLG is used with IC_tC_p, it is shown that colors at darker luminance levels are represented with more errors compared to the color at higher luminance levels. The loss of colors due to the clipping enforced by luminance levels (10,000 for Figure 8, 4,000 for Figure 9 and 1,000 for Figure 10) is also visible in Figures 8, 9, and 10. Please note how color errors with IC_tC_p are not only towards red and blue channels as compared to YC_bC_r. This can be explained by the de-correlation of the intensity (I) channel from C_t and C_p.

We conclude that based on the presented results, IC_tC_p with PQ yields better performance in terms of preserving HDR colors over the tested luminance levels when only quantization error are taken into account. These results can be explained by the fact that IC_tC_p was designed to better de-correlates intensity from chroma channels. HLG can be beneficial due to its backward-compatibility characteristics, since it also represents HDR colors in bright areas with minimal errors.

IV. CONCLUSIONS

In this work, the visual color difference caused by different PTFs and color representations followed by 10-bit quantization was evaluated. It is shown that even before compression, choice of PTF and color pixel representation will affect the visual color perception. Particularly, it was shown in the case of YC_bC_r that PQ performs better than HLG in dark luminance levels while HLG performs as well as PQ at bright luminance levels. The performance of HLG according to its reference display peak luminance also showed that the higher this value is, the better HLG performs at both dark and bright luminance levels. It is also shown that 10-bit IC_tC_p outperforms 10-bit YC_bC_r both with CL and NCL derivation in representing color due to its better de-correlation of luminance and chrominance. Although IC_tC_p with PQ represent colors throughout most of the tested luminance levels with minimal errors, there are still large errors in bright areas around the white point due to 10-bit quantization.

REFERENCES

[1] D. G. Brooks, "The art of better pixels." SMPTE Motion Imaging Journal 124.4, 42-48, 2015.

[2] R. Boitard, R. K. Mantiuk, and T. Pouli. "Evaluation of color encodings for high dynamic range pixels." SPIE/IS&T Electronic Imaging. International Society for Optics and Photonics, 2015.

[3] ITU, "Recommendation ITU-R BT.2100-0: Image parameters values for high dynamic range television for use in production and international programme exchange," International Telecommunications Union, 2016.

[4] P. Hanhart, M. Řeřábek and T. Ebrahimi, "Subjective and objective evaluation of HDR video coding technologies," 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), pp. 1-6, Lisbon, 2016.

[5] J. Ström, K. Andersson, M. Pettersson, P. Hermansson, J. Samuelsson, A. Segall, J. Zhao, S. Kim, K. Misra, A. Tourapis, Y. Su, and D. Singer, "High Quality HDR Video Compression using HEVC Main 10 Profile," in Proceedings of the IEEE Picture Coding Symposium (PCS), Nuremberg, December 2016.

[6] ITU, "Recommendation ITU-R BT.2020: Parameter values for ultrahigh definition television systems for production and international programme exchange," International Telecommunications Union, 2012.

[7] ITU, "Recommendation ITU-R BT.1886: Reference electro-optical transfer function for flat panel displays used in HDTV studio production," International Telecommunications Union, 2011.

[8] T. Borer, A. Cotton, "Display Independent High Dynamic Range Television System" in International Broadcasting Convention, 2015.

[9] ARIB, "ARIB STD-B67: for the Extended Image Dynamic Range Television (EIDRTV) System for Programme Production," in Association Radio Industries and Business, 2015.

[10] Miller, M. Nezamabadi, and S. 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] P. G. J Barten,, "Physical model for the contrast sensitivity of the human eye," in SPIE 1666, Human Vision, Visual Processing, and Digital Display III, Rogowitz, B. E., ed., 57-72, Aug. 1992.

[12] SMPTE, "SMPTE standard ST 2086, Mastering Display Color Volume Metadata Supporting High Luminance and Wide Color Gamut Images," Society of Motion Pictures Television Engineers, 2014.

[13] M. T. Pourazad, C. Doutre, M. Azimi and P. Nasiopoulos, "HEVC: The New Gold Standard for Video Compression: How Does HEVC Compare with H.264/AVC?" Consumer Electronics Magazine, IEEE, vol. 1, pp. 36-46, 2012.

[14] E. Francois, MPEG HDR AhG: about using a BT.2020 container for BT.709 content," 110th MPEG meeting in Strasbourg Compression Conference, Strasbourg, France, October 2014.

[15] T. Lu, F. Pu, P. Yin, T. Chen, W. Husak, J. Pytlarz, R. Atkins, J. Fr-hlich, and G. Su, "ITP Colour Space and Its Compression Performance for High Dynamic Range and Wide Colour Gamut Video Distribution," in ZTE Communications, February 2016 Vol.14 No.1

[16] G.W. Larson, "LogLuv encoding for full-gamut, high-dynamic range images," Journal of Graphics Tools, 3(1), pp. 15-31, 1998.

[17] G. Sharma, W. Wu, and E. N. Dalal, "The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations," Color Research & Application, vol. 30, no. 1, pp. 21-30, 2005.

[18] http://dml.ece.ubc.ca/data/HDR-Color-Evaluation