A HIGH DYNAMIC RANGE VIDEO CODEC OPTIMIZED BY LARGE-SCALE TESTING

Gabriel Eilertsen[⋆] *Rafał K. Mantiuk*[†] *Jonas Unger*[⋆]

[⋆] Media and Information Technology, Linköping University, Sweden
[†] Computer Laboratory, University of Cambridge, UK

ABSTRACT

While a number of existing high-bit depth video compression methods can potentially encode high dynamic range (HDR) video, few of them provide this capability. In this paper, we investigate techniques for adapting HDR video for this purpose. In a large-scale test on 33 HDR video sequences, we compare 2 video codecs, 4 luminance encoding techniques (transfer functions) and 3 color encoding methods, measuring quality in terms of two objective metrics, PU-MSSIM and HDR-VDP-2. From the results we design an open source HDR video encoder, optimized for the best compression performance given the techniques examined.

***Index Terms*—** High dynamic range (HDR) video, HDR video coding, perceptual image metrics

1. INTRODUCTION

High dynamic range (HDR) video constitutes a key component for next generation imaging technologies. However, efficient encoding of high dynamic range video is still an open problem. While existing video codecs that support bit depths between 10 and 12 bits have the potential to store the HDR frames in a single stream, this requires transforming luminance and colors to a format suited for encoding. Despite active development, different HDR luminance and color encoding techniques lack a comprehensive comparison.

In this work we perform a thorough comparative assessment of the steps involved in accommodating HDR video to be compressed in a single stream using existing codecs. The steps are 1) the codec used for compressing the final bit stream, 2) the transformation from linear luminances to values for encoding, and 3) the space used for representing colors and luminances. In total we consider 33 different video sequences, and 2 video codecs, 4 luminance transformations and in 3 color spaces. We then perform a large-scale testing using two objective metrics: HDR-VDP-2 and PU-MSSIM.

The contribution of the paper is twofold. Firstly, we provide a large-scale comparison of techniques for transforming HDR video to a format suited for existing video codecs. Secondly, we report on a new open source software, which has been created by combining the best performing techniques. The software can be found at: `http://lumahdrv.org`.

2. BACKGROUND

With recent advances in HDR imaging, we are now at a stage where high-quality videos can be captured with a dynamic range of up to 24 stops [1, 2]. However, storage and distribution of HDR video still mostly rely on formats for static images, neglecting the inter-frame relations that could be explored. Although the first steps are being taken in standardizing HDR video encoding, where MPEG recently released a call for evidence for HDR video coding, most existing solutions are proprietary and HDR video compression software is not available on Open Source terms.

The challenge in HDR video encoding, as compared to standard video, is in the wide range of possible pixel values and their linear floating point representation. Existing video encoding systems are very efficient, and some provide profiles that support high bit depths. While these could potentially encode high dynamic range data, the codecs do not consider the problem of adapting HDR pixels so that they could be represented as integer numbers suitable for video coding. For this purpose, the linear luminances of the HDR data need to be transformed and quantized before the encoding. The transformation can be a normalization to the range provided by the video codec [3], where the codec itself can be modified to better handle the differently distributed HDR values [4]. However, a perceptually motivated luminance transformation can more optimally account for the non-linear response of the human visual system [5].

An alternative approach to the single-stream encoding is to split the HDR video into two (8-bit) streams: one containing a (backward-compatible) tone-mapped low dynamic range (LDR) video, and the other containing a residual required to reconstruct HDR frames from the first LDR stream [6]. Although this method allows for backward-compatibility, the single-stream technique has been shown to yield better compression performance [7]. This is also the approach we consider here, where we attend all the constituting steps of the method presented by Mantiuk *et al.* [5]. We compare the original techniques to standard techniques, and consider possible new candidates. Then, optimizing for the best performance in a large-scale evaluation, we design an HDR video encoding configuration from the top performing algorithms in each step.

3. EVALUATION

To evaluate the configuration of the perceptually motivated single-stream HDR video encoding, [5], we consider its three primary components: the video codec, the luminance transfer function, and the color representation. After first describing the experimental setup, we treat each of these components separately and compare their performance using objective quality metrics.

3.1. Experimental configuration

To extensively evaluate the performance of the HDR video encoding, we make use of 33 HDR video sequences at 1080p resolution, containing approximately 100 frames per sequence. In total we use 9 conditions at 15 quality settings, which makes for 442 395 images to compare. To make this feasible, or even possible, we employed a large-scale computer cluster, where all calculations could be performed in a matter of a few days.

Test material: The videos are taken from the database by Froehlich *et al.* [8], which is the most comprehensive collection of HDR videos to date. The sequences exhibit a dynamic range of up to 18 stops and contain a wide variety of footage from different scenes. Since the employed metrics are sensitive to parameters of the display on which video is viewed, the material was tone-mapped for an HDR display with a peak luminance of 10 000 cd/m^2 using the display adaptive tone-mapping [9].

Metric: To provide a reliable estimate of the perceptual differences between the methods tested, we use two objective metrics: the visual difference predictor for HDR images (HDR-VDP-2, v2.2) [10], and the multi-scale SSIM [11] applied after perceptual linearization [12] (PU-MSSIM). HDR-VDP-2 is a full-reference metric using an HVS model to assess visual differences, and it has been demonstrated to correlate well with subjective studies [13, 14, 15, 16]. PU-MSSIM was shown to have the second best correlation (after HDR-VDP-2) with subjective mean-opinion-scores in case of JPEG XT HDR image compression [16].

Another possible metric, not considered in our evaluation, is HDR-VQM [17], which is especially designed for HDR video. The metric was shown to correlate better than HDR-VDP-2 with a subjective experiment, for some of the sequences used. However, in other cases HDR-VDP-2 had better correlation, and in [15] HDR-VDP-2 was better in differentiating between different HDR video encoding solutions.

Interpretation of results: Since the set of HDR video sequences encompasses a large variety of scenes and transitions, the performance varies considerably between sequences. To provide representative bit rate plots for the different conditions in Fig. 1, 3 and 4, the data was sampled at a selected number of bit rates. At these points data were averaged over equal bit rates, and linearly interpolated across different bit rates. The errorbars represent standard errors.

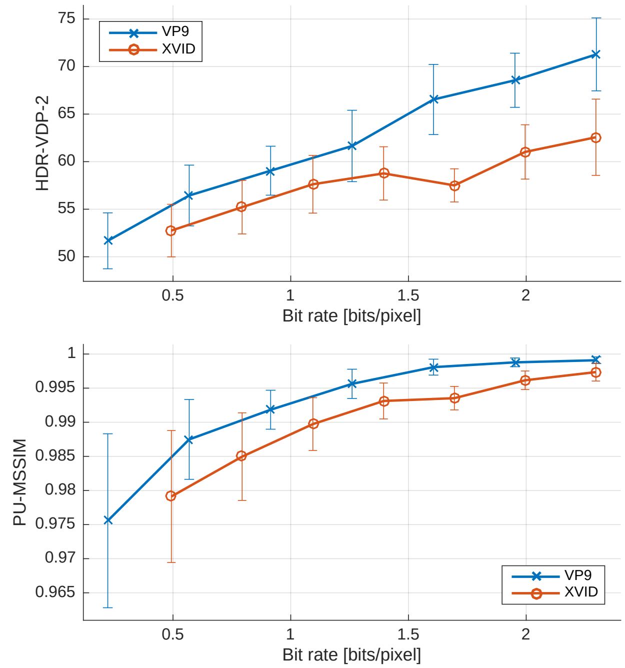

Fig. 1. Quality at different bit rates, in terms of HDR-VDP-2 and PU-MSSIM, encoding HDR video with different codecs.

3.2. Video codec

To encode HDR video without visible quantization artifacts, a minimum of 11 bits are needed [5]. A number of existing video codecs provide this precision, with high bit depth profiles for encoding at up to 12 bits. Since our goal is to release the codec on open source terms, we selected Google's VP9 codec as its license permits such use. The performance of the codec is on par with the widely used H.264 standard [18].

To confirm that VP9 outperforms XVID's MPEG-4 Part 2 compression scheme used in the original single stream HDR video encoding [5], Fig. 1 shows a comparison of the codecs in terms of HDR-VDP-2 and PU-MSSIM. Although the data shows high variance at some points, due to the wide variety of sequences, on average VP9 clearly performs superior to XVID, with about half the bit rate for the same quality.

3.3. Luminance encoding

Similarly as a "gamma" transfer function is required for standard dynamic range content, linear HDR pixel values need to be compressed with an appropriate transfer function before encoding. The simplest choice, justified by the approximately logarithmic response of the eye (according to the

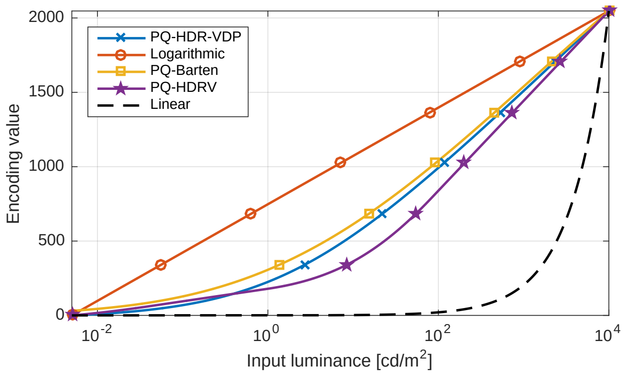

Fig. 2. Luminance transfer functions, mapping physical units to integer values for encoding. Here, luminances in the range $[0.005, 10^4]$ cd/m^2 are mapped to 11 bits, $[0, 2047]$. Markers are used to differentiate the plots in a b/w print.

Weber–Fechner law), is to encode values in the logarithmic domain. However, more sophisticated methods have been derived using perceptual measurements. These are referred to as perceptual transfer functions (PTFs) or electro-optical transfer functions (EOTFs), and their purpose is to translate linear floating point luminances to the screen-referred integer representation of an encoding system, with quantization errors that are perceptually equally distributed across all luminances.

Given a luminance $L \in [L_{min}, L_{max}]$, the mapping function $V(L)$ should transform it to the range of the codec. To simplify notation, we specify a target range of $V(L) \in [0, 1]$. This transformed value, or luma, should then be scaled by $2^b - 1$ before quantization at a target bit depth b. The PTFs considered are illustrated in Fig. 2, and are as follows:

Logarithmic: The most straightforward solution for the transformation, is to scale and normalize in the log domain. This serves as a reference, when comparing more sophisticated formulations of the luminance transformation:

$$V(L) = \frac{log_{10}(L) - log_{10}(L_{min})}{log_{10}(L_{max}) - log_{10}(L_{min})} \tag{1}$$

PQ-HDRV: The method proposed in [5] derives a luminance encoding that ensures that the quantization errors are below the visibility threshold, $tvi(L)$. This leads to an integral equation:

$$\begin{aligned} V(L) &= \frac{1}{2} \int \frac{k}{tvi(L)} dL, \\ V(L_{min}) &= 0, \quad V(L_{max}) = 1 \end{aligned} \tag{2}$$

The equation can be solved, given the boundary conditions, using the *shooting method*. The threshold-versus-intensity function, $tvi(L)$, is taken from [19], and is based on psychophysical measurements of the threshold visibility.

PQ-HDR-VDP: HDR-VDP-2 [10] uses a measured contrast sensitivity function (CSF) for prediction of the visual differences at different luminances. While the threshold-versus-intensity function is measured for a fixed pattern, the CSF is measured for different spatial frequencies ρ (*e.g.* sinusoidal patterns or Gabor patches). In order to use this in Equation 2, a conservative choice is to always sample it at the frequency where the sensitivity is the highest:

$$tvi(L) = \frac{L}{max_\rho CSF(\rho, L)} \tag{3}$$

PQ-Barten: The luminance transformation presented in [20] is termed the perceptual quantizer (PQ). However, since we also include other perceptual quantization schemes above, we distinguish this as PQ-Barten. It is based on the CSF model in [21], derived in a similar fashion as the luminance encoding in [5]. The result is fitted to a cone response model, and the final transformation is formulated according to:

$$\begin{aligned} V(L) &= \left(\frac{0.8359 + 18.8516\ L_p}{1 + 18.6875\ L_p} \right)^{78.8438}, \\ L_p &= \left(\frac{L}{L_{max}} \right)^{0.1593} \end{aligned} \tag{4}$$

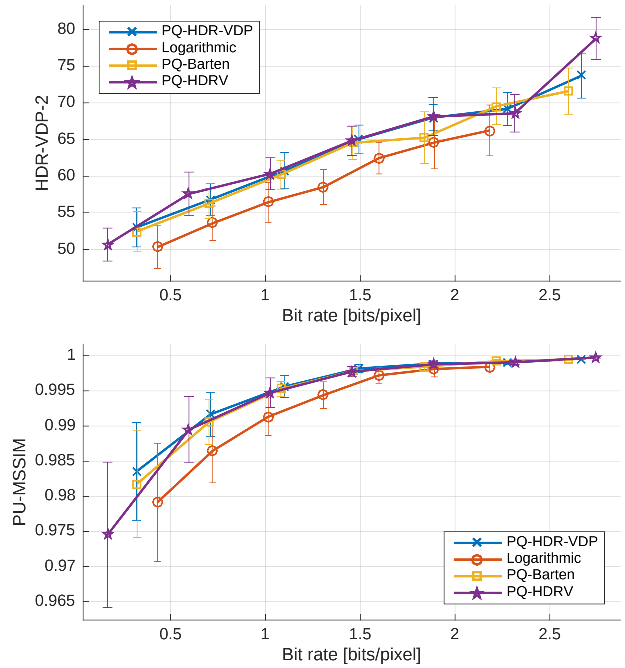

Fig. 3. Comparing different luminance transformation functions for encoding, in terms of HDR-VDP-2 and PU-MSSIM.

Fig. 3 shows a performance assessment of the luminance encodings. Since PQ-Barten only considers luminances below $10\,000\,\text{cd/m}^2$, the material used has been re-targeted as described in Section 3.1. As expected, the perceptual transfer functions show a substantial improvement over the

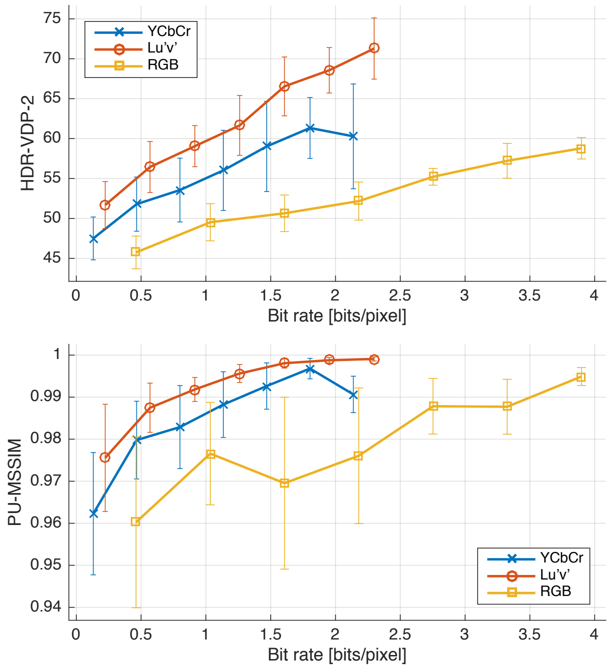

Fig. 4. Comparing different colorspaces for encoding, in terms of HDR-VDP-2 and PU-MSSIM.

logarithmic scaling. Differentiating between these, however, is difficult considering the variance of the measurements. Even though there are no clear evidence in favor of any of the perceptual encodings, we employ PQ-Barten as default for our open source codec. This is based on observations that PQ-Barten distributes distortions more uniformly across luminances as compared to the other methods [22].

3.4. Color encoding

Encoding of color information is typically done by separating luminance and chrominance, substantially lowering inter-channel correlations as compared to for example RGB. In the case of standard video, this is generally done using YC_bC_r, *e.g.* according to the most recent recommendation ITU-R BT.2020. However, this standard is incapable of encoding the full visible color gamut in HDR images, for which the perceptually linear $Lu'v'$ color space can be used [23].

Fig. 4 compares the performance of encoding the HDR video sequences in RGB, YC_bC_r and $Lu'v'$. For YC_bC_r and $Lu'v'$, the chroma channels have been sub-sampled to half the image width and height (4:2:2 sampling), while RGB uses the full-sized channels. To encode luminances, the PQ-HDRV transformation has been used on the luminance channels, and on RGB channels separately. As expected, RGB is clearly inefficient. Also, although variance is high, comparing average performance $Lu'v'$ shows a great improvement over YC_bC_r, with about half the bit rate for the same quality.

4. OPEN SOURCE CODEC

Guided by the results of the tests in the previous section, we selected the best combination of codec, transfer function and color coding and implemented them in a new open source video compression software, called *Luma HDRv*[1]. The software has been released under the BSD license. *Luma HDRv* provides libraries for including the HDR video encoding and decoding functionalities in software development, as well as applications to perform encoding and decoding with a number of different settings. Furthermore, we also provide an HDR video player for real-time decoding, playback and tone-mapping of encoded HDR videos.

Luma HDRv uses VP9 for encoding, and the default settings are according to the results discussed above, with PQ-Barten for luminance quantization at 11 bits, and encoding in the $Lu'v'$ colorspace. The encoded HDR videos are stored using the Matroska container[2], for flexibility and easy integration into existing software.

5. CONCLUSION

In this work we compared different techniques for the components of a single stream HDR video encoder that uses existing video codecs. The evaluation was made on a large range of different HDR video sequences, in terms of objective metrics that have been shown to correlate well with subjective assessments. From the results we clearly see how important perceptual methods are when encoding HDR as compared to standard video. This goes both for the quantization transformation, as well as the colorspace utilized. Also, we have demonstrated the large difference in encoding performance for HDR video, when comparing a modern codec to older alternatives.

The results demonstrated were sub-sequently used to construct *Luma HDRv* – an optimized HDR video codec solution which is released as an open source project.

For future work, additional support and confirmation of the results could be gained through a subjective evaluation experiment on a subset of the sequences.

6. ACKNOWLEDGEMENTS

This project was funded by the Swedish Foundation for Strategic Research (SSF) through grant IIS11-0081, Linköping University Center for Industrial Information Technology (CENIIT), the Swedish Research Council through the Linnaeus Environment CADICS, and through COST Action IC1005. The objective quality evaluation was possible thanks to High Performance Computing Wales, Wales national supercomputing service (hpcwales.co.uk).

[1] `http://lumahdrv.org`
[2] `http://www.matroska.org`

7. REFERENCES

[1] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen, "A versatile HDR video production system," *ACM Trans. Graph.*, vol. 30, no. 4, July 2011.

[2] J. Kronander, S. Gustavson, G. Bonnet, and J. Unger, "Unified HDR reconstruction from raw CFA data," in *Proc. of the IEEE International Conference on Computational Photography*, 2013.

[3] A. Banitalebi-Dehkordi, M. Azimi, M. T. Pourazad, and P. Nasiopoulos, "Compression of high dynamic range video using the HEVC and H. 264/AVC standards," in *IEEE Heterogeneous Networking for Quality, Reliability, Security and Robustness (QShine)*, 2014.

[4] M. Azimi, M. T. Pourazad, and P. Nasiopoulos, "Rate-distortion optimization for high dynamic range video coding," in *IEEE Communications, Control and Signal Processing (ISCCSP)*, 2014.

[5] R. K. Mantiuk, G. Krawczyk, K. Myszkowski, and H.-P. Seidel, "Perception-motivated high dynamic range video encoding," *ACM Trans. on Graphics*, vol. 23, no. 3, 2004.

[6] R. K. Mantiuk, A. Efremov, K. Myszkowski, and H.-P. Seidel, "Backward compatible high dynamic range MPEG video compression," *ACM Trans. on Graphics*, vol. 25, no. 3, July 2006.

[7] M. Azimi, R. Boitard, B. Oztas, S. Ploumis, H. R. Tohidypour, M. T. Pourazad, and P. Nasiopoulos, "Compression efficiency of HDR/LDR content," in *IEEE Quality of Multimedia Experience (QoMEX)*, 2015.

[8] J. Froehlich, S. Grandinetti, B. Eberhardt, S. Walter, A. Schilling, and H. Brendel, "Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR-displays," in *Proc. SPIE, Digital Photography X*, 2014.

[9] R. K. Mantiuk, S. Daly, and L. Kerofsky, "Display adaptive tone mapping," *ACM Trans. Graphics*, vol. 27, no. 3, 2008.

[10] R. K. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graphics*, vol. 30, no. 4, 2011.

[11] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. on Image Processing*, vol. 13, no. 4, 2004.

[12] T. O. Aydın, R. K. Mantiuk, and H.-P. Seidel, "Extending quality metrics to full dynamic range images," in *Proc. SPIE, Human Vision and Electronic Imaging XIII*, San Jose, USA, 2008.

[13] P. Hanhart, M. V. Bernardo, P. Korshunov, M. Pereira, A. M. G. Pinheiro, and T. Ebrahimi, "HDR image compression: a new challenge for objective quality metrics," in *IEEE Quality of Multimedia Experience*, 2014.

[14] M. Azimi, A. Banitalebi-Dehkordi, Y. Dong, M. T. Pourazad, and P. Nasiopoulos, "Evaluating the performance of existing full-reference quality metrics on high dynamic range (HDR) video content," in *International Conference on Multimedia Signal Processing (ICMSP 2014)*, 2014.

[15] P. Hanhart, M. Řeřábek, and T. Ebrahimi, "Towards high dynamic range extensions of HEVC: subjective evaluation of potential coding technologies," in *SPIE Optical Engineering + Applications*, 2015.

[16] A. Artusi, R. K. Mantiuk, T. Richter, P. Hanhart, P. Korshunov, M. Agostinelli, A. Ten, and T. Ebrahimi, "Overview and evaluation of the JPEG XT HDR image compression standard," *Journal of Real-Time Image Processing*, 2015.

[17] M. Narwaria, M. P. Da Silva, and P. Le Callet, "HDR-VQM: an objective quality measure for high dynamic range video," *Signal Processing: Image Communication*, vol. 35, 2015.

[18] M. Řeřábek and T. Ebrahimi, "Comparison of compression efficiency between HEVC/H. 265 and VP9 based on subjective assessments," in *SPIE Optical Engineering + Applications*, 2014.

[19] J. A. Ferwerda, S. N. Pattanaik, P. Shirley, and D. P. Greenberg, "A model of visual adaptation for realistic image synthesis," in *ACM SIGGRAPH*, 1996.

[20] S. Miller, M. Nezamabadi, and S. Daly, "Perceptual signal coding for more efficient usage of bit codes," *SMPTE Motion Imaging Journal*, vol. 122, no. 4, 2013.

[21] Peter G. J. Barten, "Formula for the contrast sensitivity of the human eye," in *Proc. SPIE 5294, Image Quality and System Performance*, Yoichi Miyake and D. Rene Rasmussen, Eds., dec 2004, pp. 231–238.

[22] R. Boitard, R. K. Mantiuk, and T. Pouli, "Evaluation of color encodings for high dynamic range pixels," in *Proc. SPIE, Human Vision and Electronic Imaging XX*, 2015.

[23] G. Ward Larson, "LogLuv encoding for full-gamut, high-dynamic range images," *Journal of Graphic Tools*, vol. 3, no. 1, Mar. 1998.