Luma HDRv: an open source high dynamic range video codec optimized by large-scale testing

Supplementary material

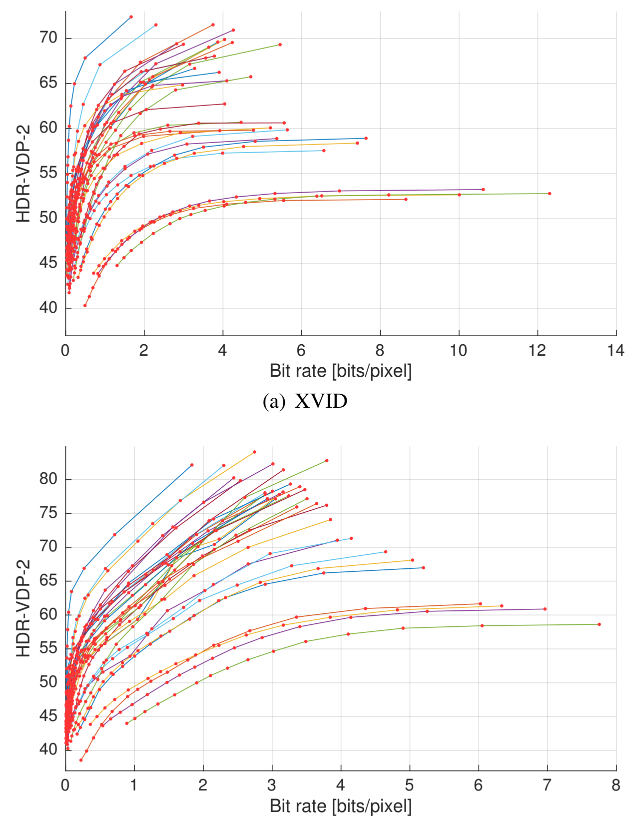
Gabriel Eilertsen[1] Rafał K. Mantiuk[2] Jonas Unger[1]
[1]Linköping University, Sweden [2]Computer Laboratory, University of Cambridge, UK

1 Introduction

This document presents supplemental material to the Siggraph 2016 talk titled "Luma HDRv: an open source high dynamic range video codec optimized by large-scale testing". It provides some additional information and figures, related to the evaluation of HDR video encoding techniques, that did not fit in the abstract.

2 Measured data

The set of HDR video sequences used in the objective comparison experiment encompasses a large variety of scenes and transitions. Thus, the compression performance varies considerably between the sequences. Figure 1 shows bit rate curves for all 33 HDR videos included in the comparisons, to demonstrate the wide variations in performance due to scene content. The curves are plotted for one setting of the HDR video encoder, but using the XVID video codec (a), and the VP9 video codec (b). The quality is measured in terms of the HDR-VDP-2 (v2.2) quality predictor "Q". For each curve, 15 quality levels are used, where each level is averaged over all frames in the corresponding sequence. To provide representative bit rate plots for the different conditions (Figure 3-5, and Figure 1 in the abstract), 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.



(a) XVID

(b) VP9

Figure 1: *Per sequence bit rate curves for two settings of the HDR video encoder: (a) using the XVID codec, (b) using the VP9 codec. Data points are averaged over all frames in respective sequence.*

3 Luminance transfer functions

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 Figure 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 [Mantiuk et al. 2004] derives a luminance encoding that ensures that the quantization errors are below the visibility threshold, $tvi(L)$. This leads to an integral equation:

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

The equation can be solved, given the boundary conditions, using the *shooting method*. The threshold-versus-intensity function, $tvi(L)$, is taken from [Ferwerda et al. 1996], and is based on psychophysical measurements of the threshold visibility.

PQ-HDR-VDP: HDR-VDP-2 [Mantiuk et al. 2011] 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 [Miller et al. 2013] is termed the perceptual quantizer (PQ). However, since

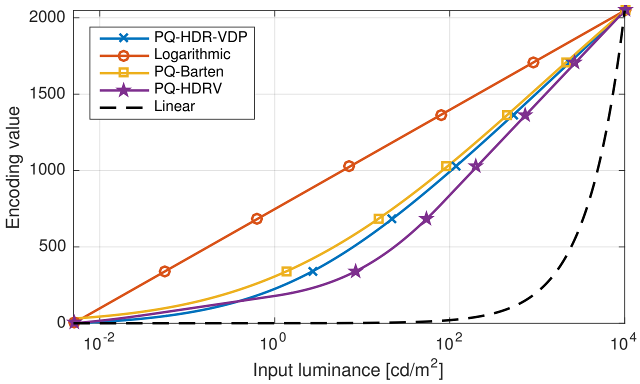


Figure 2: *Luminance transfer functions, mapping physical units to integer values for encoding. Here, luminances in the range* $[0.005, 10^4]$ *cd/m*^{2} *are mapped to 11 bits,* $[0, 2047]$.

we also include other perceptual quantization schemes above, we distinguish this as PQ-Barten. It is based on the CSF model in [Barten 2004], derived in a similar fashion as the luminance encoding in [Mantiuk et al. 2004]. The result is fitted to a cone response model, and the final transformation is formulated according to:

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

4 Comparisons

Figures 3-5 present results of comparing codecs, luminance encodings, and color encodings, respectively. The data has been sampled and averaged from per-sequence data (as the ones in Figure 1), as described in Section 2. Here, in addition to HDR-VDP-2, the comparisons are complemented with PU-MSSIM, PU-PSNR and PU-SSIM. PU-SSIM and PU-MSSIM are SSIM and the multi-scale SSIM, respectively [Wang et al. 2004], applied after perceptual linearization [Aydın et al. 2008]. PU-PSNR is the PSNR after the same linearization. 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 [Artusi et al. 2015].

5 Color encoding example

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 [Larson 1998]. For the comparison experiment (Figure 5), YC_bC_r and $Lu'v'$ were included to measure the improvement of using a perceptually motivated approach for encoding of the HDR colors. Also, RGB was included for reference, and looking at the results it is clearly inefficient.

The visual differences in using the $u'v'$ and C_bC_r color encodings are illustrated in Figure 6, where an HDR video has been encoded with VP9 at approximately the same bit rate using $Lu'v'$ and YC_bC_r. The figure shows one frame from the sequence, and also illustrates the incapabilities of PSNR in perceptual comparisons. Although the images show about the same quality in terms of PSNR, HDR-VDP-2 shows a large difference in favor for $Lu'v'$. Inspecting the images, it is clear that there are perceptual differences in quality, where YC_bC_r suffers from color blocking artifacts.

References

ARTUSI, A., MANTIUK, R. K., RICHTER, T., HANHART, P., KORSHUNOV, P., AGOSTINELLI, M., TEN, A., AND EBRAHIMI, T. 2015. Overview and evaluation of the JPEG XT HDR image compression standard. *Journal of Real-Time Image Processing*.

AYDIN, T. O., MANTIUK, R. K., AND SEIDEL, H.-P. 2008. Extending quality metrics to full dynamic range images. In *Proc. SPIE, Human Vision and Electronic Imaging XIII*.

BARTEN, P. G. J. 2004. Formula for the contrast sensitivity of the human eye. In *Proc. SPIE 5294, Image Quality and System Performance*, Y. Miyake and D. R. Rasmussen, Eds., 231–238.

FERWERDA, J. A., PATTANAIK, S. N., SHIRLEY, P., AND GREENBERG, D. P. 1996. A model of visual adaptation for realistic image synthesis. In *ACM SIGGRAPH*.

FROEHLICH, J., GRANDINETTI, S., EBERHARDT, B., WALTER, S., SCHILLING, A., AND BRENDEL, H. 2014. Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR-displays. In *Proc. SPIE, Digital Photography X*.

LARSON, G. W. 1998. LogLuv encoding for full-gamut, high-dynamic range images. *Journal of Graphic Tools 3*, 1 (Mar.).

MANTIUK, R. K., KRAWCZYK, G., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2004. Perception-motivated high dynamic range video encoding. *ACM Trans. on Graphics 23*, 3.

MANTIUK, R. K., KIM, K. J., REMPEL, A. G., AND HEIDRICH, W. 2011. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graphics 30*, 4.

MILLER, S., NEZAMABADI, M., AND DALY, S. 2013. Perceptual signal coding for more efficient usage of bit codes. *SMPTE Motion Imaging Journal 122*, 4.

WANG, Z., BOVIK, A. C., SHEIKH, H. R., AND SIMONCELLI, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. on Image Processing 13*, 4.

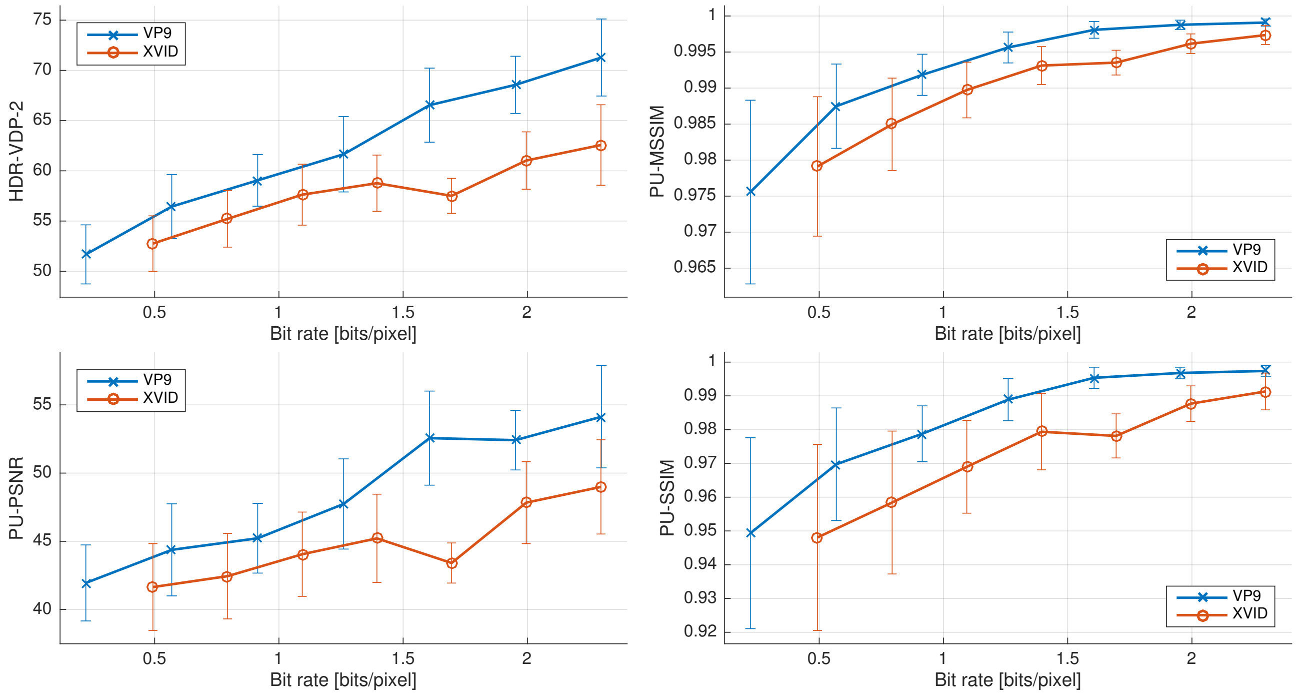


Figure 3: *Codec comparison.*

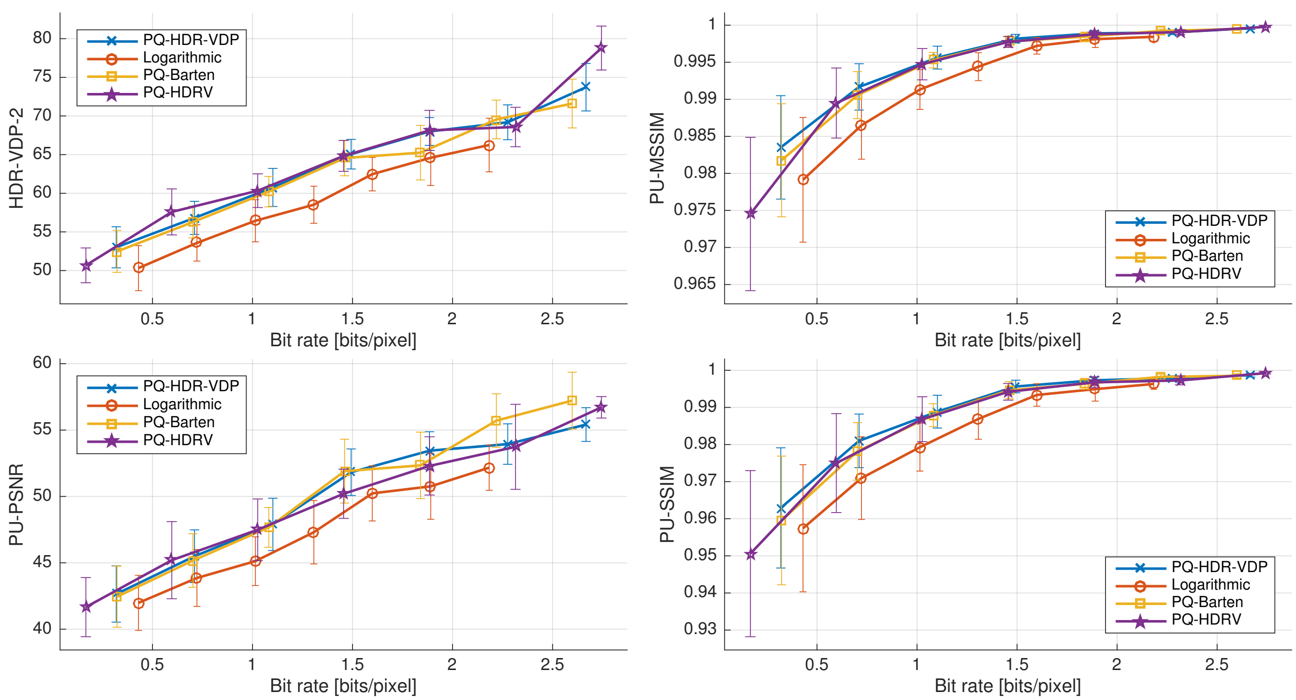


Figure 4: *Luminance encoding comparison.*

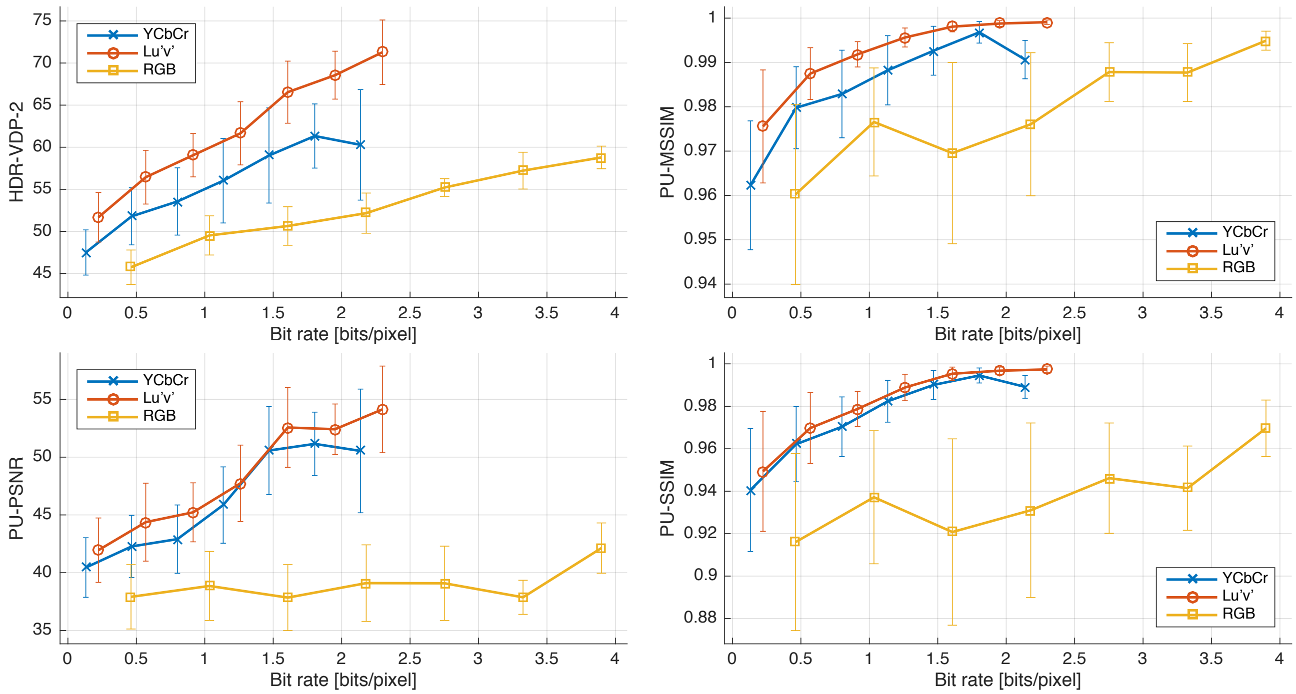
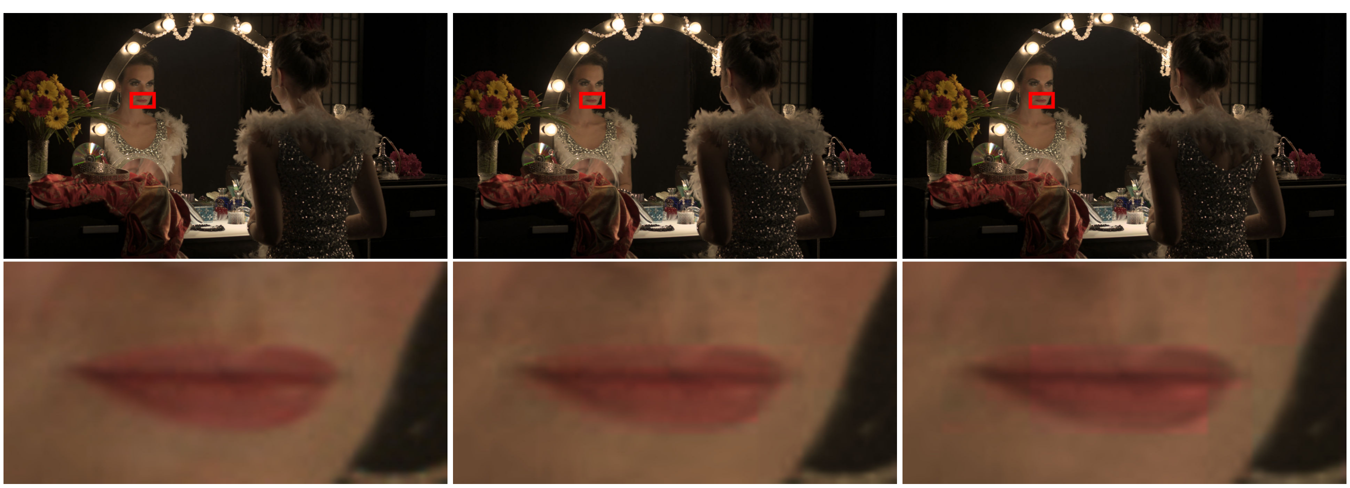


Figure 5: *Color encoding comparison.*



(a) Input (b) u′v′, PSNR=48.45, HDR-VDP-2=64.29 (c) C_bC_r, PSNR=48.95, HDR-VDP-2=57.83

Figure 6: *Comparing $u'v'$ and C_bC_r for encoding of colors. Even though C_bC_r gives slightly higher PSNR, $u'v'$ performs significantly better in terms of HDR-VDP-2 quality prediction. As shown in the enlargements, the perceptual metric corresponds better to how the differences are perceived, where the C_bC_r encoding shows apparent color blocking artifacts. The examples are encoded at approximately the same bit rate. Video frame from [Froehlich et al. 2014].*