



HAL
open science

Optimized Tone Mapping with Perceptually Uniform Luminance Values for Backward-Compatible High Dynamic Range Video Compression

Alper Koz, Frederic Dufaux

► To cite this version:

Alper Koz, Frederic Dufaux. Optimized Tone Mapping with Perceptually Uniform Luminance Values for Backward-Compatible High Dynamic Range Video Compression. IEEE Proc. Visual Communications and Image Processing (VCIP'2012), IEEE, Nov 2012, San Diego, United States. <hal-01436377>

HAL Id: hal-01436377

<https://hal.science/hal-01436377v1>

Submitted on 10 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

OPTIMIZED TONE MAPPING WITH PERCEPTUALLY UNIFORM LUMINANCE VALUES FOR BACKWARD-COMPATIBLE HIGH DYNAMIC RANGE VIDEO COMPRESSION

Alper Koz and Frederic Dufaux

Télécom ParisTech, Dept. Traitement du Signal et des Images, 75014 Paris, France.
{alper.koz, frederic.dufaux}@telecom-paristech.fr

ABSTRACT

Backward compatibility for high dynamic range image and video compression forms one of the essential requirements in the transition phase from low dynamic range (LDR) displays to high dynamic range (HDR) displays. In a recent work [1], an optimized solution for tone mapping and inverse tone mapping of HDR images is achieved in terms of mean square error (MSE) of the logarithm of luminance values of HDR image pixels for backward-compatible compression. A disadvantage of this approach was to use non uniform luminance values according to Human perception for minimization, which causes quite non-natural over-illumination in the produced LDR images. In this paper, we propose to use perceptually uniform luminance values as an alternative for the optimization of tone mapping curve. The results indicate that the proposed approach gives better performance (0.5-1 dB gains) in terms of *Perceptually Uniform Peak Signal to Noise Ratio* (PU-PSNR) and produces more realistic LDR images.

Index Terms— High Dynamic Range Video Compression, Tone Mapping, Perceptually Uniform Encoding, Backward Compatible

1. INTRODUCTION

Video content represented by eight bits per pixel has been widely accepted in many applications such as internet/video streaming, DVD, HDTV, camcorders, personal computers, televisions and other consumer electronics products. The reasons of such a wide acceptance of eight-bit representation can be argued as the compact representation of one pixel value as a *byte* for the storage in memory chips and the sufficiency of 256 levels to cover the luminance range of commonly used low dynamic range (LDR) displays (0.1-80 cd/m²) [2]-[4].

However, image capturing and display technologies have improved to span the wider dynamic range and true color representations in the last decade. LCD displays and plasma displays with a peak luminance of 500-1000 cd/m² have dominated the market. HDR prototype displays are now available with a dynamic range of 1,000,000:1 and a peak luminance of 4000 cd/m². Accordingly, such a limited representation has become insufficient for these new technologies and related applications in digital cinema, medical imaging, and post-production [2]-[4].

HDR image and video formats are proposed to overcome the limitations of standard eight-bit representation. The main goal of these formats is to encode the entire luminance range of real world scenes ranging from extreme darkness (10⁻⁶ cd/ m²) to bright sunshine (10⁸ cd/ m²), and hence to generate and store a scene referred image independent of the display technology [2]-[4]. The proposed formats for such a goal obviously require higher bit

depths than standard eight bits. Current formats such as Radiance RGBE (*.hdr*), OpenEXR (*.exr*), and LogLuv TIFF (*.tiff*) use 16-bit floating point, 32-bit floating point or 32-bit integer pixels [1].

A major challenge for such a high-bit depth representation is an efficient compression, in particular for HDR video which requires significantly more storage size and transmission bandwidth than standard 8-bit LDR video. The previous work has mainly handled this challenge in two ways. The first approach [5]-[7] is to take the advantage of high profile of existing state-of-the-art H.264/AVC (Advanced Video Coding) codec. As the maximum bit-depth supported in high-profile mode is 14 bits, the core part of this approach was to convert 16 bits HDR video into 14 bits. In [5], this conversion is performed by applying the modified versions of LogLuv transform [8], which maps the luminance range of each HDR frame to the range $[0, 2^{14}-1]$ with respect to the max. and min. luminance values of the frame. In the extended version of this work [6], temporal motion compensation problems generated by individual mapping of frames is solved by weighted prediction (WP) tool of H.264/AVC [9]. In a similar work [7], such a mapping is enhanced by smoothing the HDR frames with respect to the Human Visual System (HVS) sensitivities in different wavelet resolutions. The advantage of these methods was their conceptual simplicity. On the down side, they do not support the backward compatibility with 8-bit displays and are far from satisfying the real time requirement as the encoding complexity of the H.264/AVC high profile is still high.

The second approach [1],[3],[10] aims to provide the backward compatibility with existing standard 8-bit displays in order to enable a successful transition to HDR technology. The basic strategy for such an aim is to include an inter-layer prediction stage in order to estimate the HDR video from LDR video. The residue of the estimation and the prediction coefficients are then encoded and written to the output bit-stream with the independently encoded LDR video. In [10], a block wise approach is proposed for interlayer prediction. A linear multiplication and an addition of offset are applied to each block of LDR video frame to estimate the corresponding block of HDR video frame. This method is improved in [3] by differentially encoding the offsets with respect to the previously encoded left and upper blocks. In a recent work [1], block wise prediction is replaced by a global (inverse) tone mapping operation, and the optimum prediction with respect to the MSE of the logarithm of luminance values is formulated in a detailed and complete manner.

However, a disadvantage of this sophisticated work was to assume that the classical *Weber* law [11] is valid for the whole luminance range and the logarithm operation applied on the luminance values produce uniform values with respect to the HVS sensitivity. In [4], Aydin *et al.* has indicated that the luminance sensitivities of HVS cannot be just modeled with *Weber* law for the

whole luminance range and has instead developed a perceptual mapping which converts the luminance values into perceptually uniform encoded values.

In this paper, we propose to use these perceptually uniform encoded values rather than the absolute luminance values for the optimization of tone mapping curve. We have compared both approaches and showed that perceptually uniform approach gives better results in terms of perceptually uniform peak signal to noise ratio (PU-PSNR). In addition, the quality of the LDR images produced with tone mapping is examined for different HDR images for both methods. The next section gives an overview of perceptually uniform encoding [4] of luminance values. Section 3 presents the details of the optimization procedure of tone mapping developed in [1]. Section 4 explains the proposed tone mapping method and section 5 gives the experimental results and comparisons. Finally, the conclusions are given in section 6.

2. PERCEPTUALLY UNIFORM ENCODING OF LUMINANCE VALUES

Commonly used quality metrics such as PSNR and SSIM assumes that pixel intensities are perceptually uniform regardless of their luminance values. Although such an assumption applies well for CRT displays with the luminance range of 0.1-80 cd/m^2 , it does not exactly fit to much brighter LDR (500 cd/m^2), plasma (1000 cd/m^2) and HDR (4000 cd/m^2) displays as the distortions in image pixels become more visible with higher luminance. Considering this fact, Aydin *et al.* [4] has proposed a mapping which converts luminance values into approximately perceptually uniform values in order to enable an extension of the well-known PSNR metric to all luminance levels visible to HVS.

Their motivation to this end is mainly based on two observations. First, classical *Weber* law [11] stating that the ratio of the perceivable luminance difference to the background luminance (i.e. *contrast threshold*) is constant is only valid for the background luminance greater than 500 cd/m^2 , as indicated with the contrast versus intensity (*cvi*) curve in Figure 1. It is therefore not a very accurate model for the range of luminance shown in typical displays (0.1 cd/m^2 to 100-1000 cd/m^2). Second, the compressive transform function of sRGB color space used in typical displays to convert luma values (pixel intensities) to luminance values is not very appropriate for brighter displays. This is verified in Figure 1 by showing the strong deviations in the slopes of the quantization error between the luminance values corresponding to consecutive luma values from *cvi* curve for brighter displays.

In order to derive a perceptually uniform (PU) encoding that is applicable to whole luminance range, contrast thresholds should be properly estimated as a function of pixel luminance. In [4], this is performed by modeling the effects of different aspects of complex images such as spatial frequency, orientation, and adapting luminance to contrast thresholds of HVS. To ensure that the estimated threshold is always conservative, the value that corresponds to the maximum sensitivity is chosen for each aspect of HVS utilized in the model. The estimated contrast thresholds are then input to a recursive formula to find the optimum mapping from luminance values to perceptually uniform values [4]. The resulting PU encoding is shown in Figure 2. While there is a closer match to sRGB color transform for the range of 1-100 cd/m^2 , the difference for the higher luminance is quite significant.

The main aim of PU encoding was to define the *perceptually uniform peak-signal to noise ratio* (PU-PSNR), which takes the

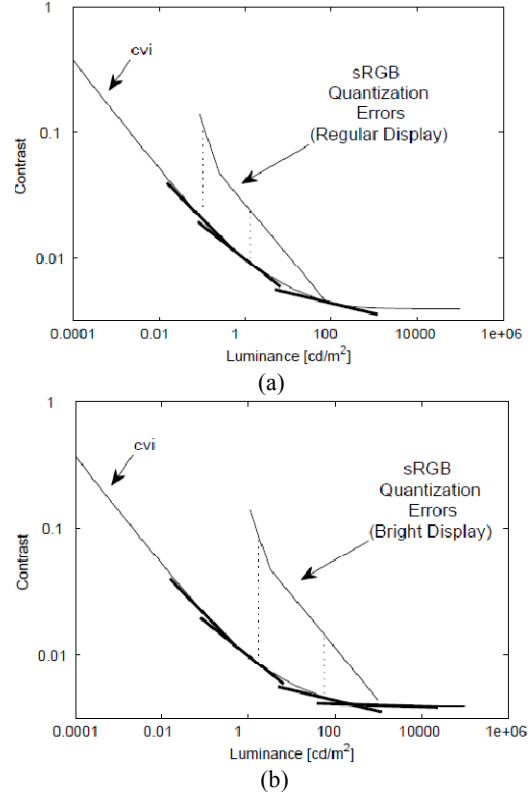


Fig. 1 Contrast versus Intensity (*cvi*) function and quantization errors due to sRGB encoding, (a) in a regular display (80 cd/m^2) and, (b) in a bright display (1000 cd/m^2) [4].

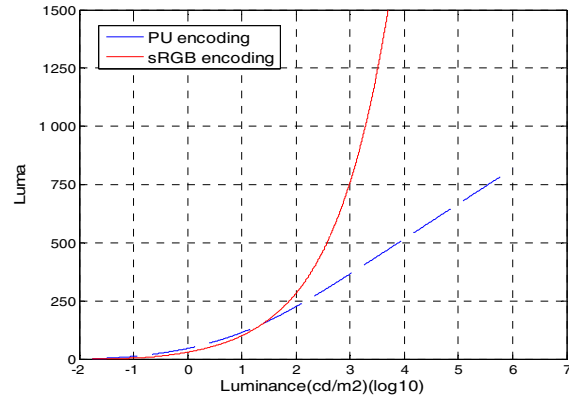


Fig. 2 Perceptually Uniform Encoding and sRGB encoding as a function of luminance (cd/m^2).

perceptually uniform encoded values as the input to estimate the quality of HDR images and the quality of the LDR images shown in bright displays. Our target in this paper would be instead to use the perceptually uniform encoded values to find the optimum tone mapping for backward-compatible HDR image and video compression.

3. OPTIMUM TONE MAPPING FOR BACKWARD COMPATIBLE HDR VIDEO COMPRESSION

A general scheme for backward compatible HDR video compression is illustrated in Figure 3 [1]. The HDR video is first

mapped into 8-bit by using a tone mapping operator to produce an LDR video. The LDR video goes through the standard video encoding and decoding process (H.264/AVC). Then, a high dynamic range video is reconstructed from LDR video by applying an inverse tone mapping operator. The parameters regarding the tone mapping and inverse tone mapping can be stored as a look-up table [1] or can be encoded using supplemental enhancement information [12]. The residue of the reconstructed HDR video can also be further encoded as an enhancement layer in the bitstream.

Mai *et al.* [1] in their recent work have developed the optimum tone mapping and inverse tone mapping which minimizes the mean square error between the logarithmic (\log_{10}) values of the luminance of original HDR content and its reconstructed version. In order to find a computationally efficient solution, they have estimated the distortion in the chain of tone-mapping, encoding, decoding, and inverse-tone mapping with a statistical distortion model and found a closed-form solution for tone mapping based on the luminance histogram of the HDR video frames.

Given the notations l and v corresponding to the logarithm of the luminance of HDR frame and the pixel values of tone mapped LDR version respectively, tone mapping curve in their solution [1] is first parameterized as a piece-wise linear function with the nodes (l_k, v_k) as shown in Figure 4. Each segment k between two nodes (l_k, v_k) and (l_{k+1}, v_{k+1}) has a constant width in HDR values equal to δ (selected as 0.1). The tone mapping operation is then characterized by a set of slopes

$$s_k = \frac{v_{k+1} - v_k}{\delta} \quad (1)$$

which forms a vector of tone-mapping parameters θ . Using such a parameterization and the statistical modeling of the H.264/AVC coding error produced in LDR video, their ultimate closed form solution is derived as

$$s_k = \frac{v_{\max} \cdot p_k^{1/3}}{\delta \cdot \sum_{k=1}^N p_k^{1/3}} \quad (2)$$

where p_k is the summation of the normalized histogram of luminance values for the k 'th bin, N is the total number of bins in the histogram, and v_{\max} is the maximum LDR value [1].

4. OPTIMUM TONE MAPPING WITH PERCEPTUALLY UNIFORM ENCODING

The proposed tone mapping method by Mai *et al.* [1] operates on the logarithm of luminance values to comply with the sensitivity of HVS to different luminance levels in accordance with *Weber* law [11]. However, *Weber* law is not an exact model of HVS contrast sensitivity for the whole luminance range as revealed in Section 2. Therefore, such an optimization procedure is not optimum in the perceptual sense although it minimizes the mean square error of logarithm of luminance values between the original and reconstructed HDR frames.

Our proposal in this paper is to use perceptually uniform encoding instead of the logarithm operation, and to minimize the mean square error between the perceptually uniform encoded values of original and reconstructed frames. Figure 5 illustrates the main stages of the proposed method in parallel with Figure 3. First, the luminance values of HDR frames is converted into perceptually uniform values by PU encoding (Fig. 2). The resulting values are represented as l_u . Then, the optimized tone mapping is applied to l_u

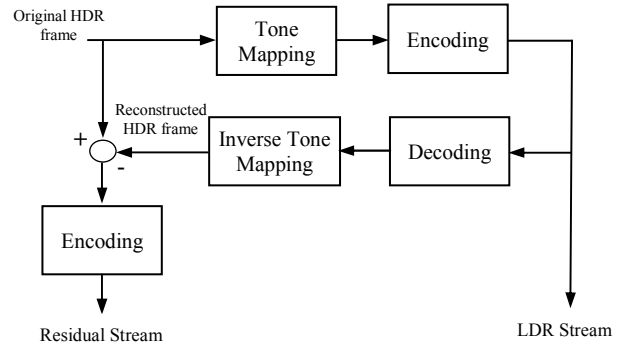


Fig. 3 General scheme for backward compatible HDR video compression.

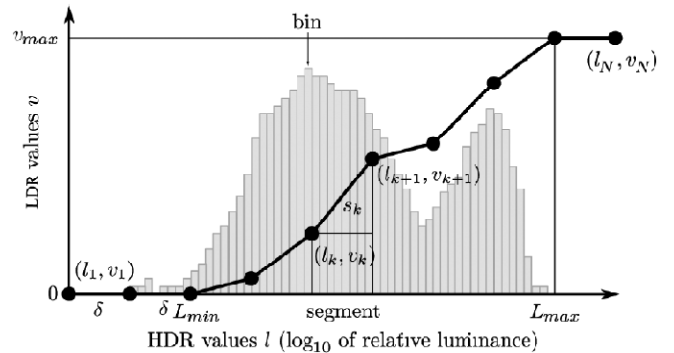


Fig. 4 Parameterization of the tone-mapping curve in [1]. The bar-plot is the luminance histogram of the HDR frame.

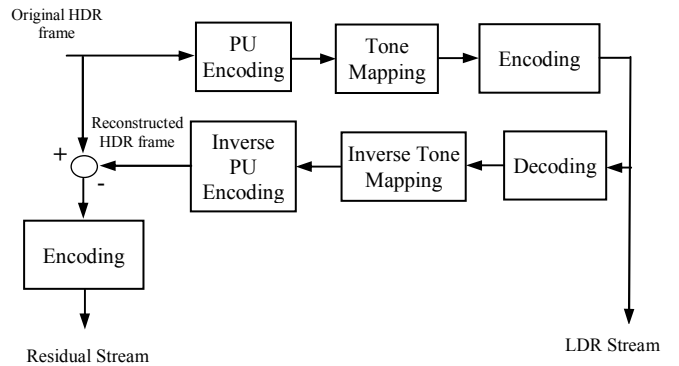


Fig. 5 Proposed scheme for backward compatible HDR video compression with PU encoding.

values to obtain the LDR video. After the LDR video goes through the coding and decoding, inverse tone mapping is applied to the resulting sequence to obtain the perceptually uniform values. Finally, the actual luminance values are found by using the inverse PU encoding.

As the solution derived in [1] is only based on the histogram of the input image and the statistical modeling of H.264/AVC coding error, such a formulation for the optimum solution can also be applied to perceptually uniform values without loss of generality. For the given frame, the histogram over l_u values is computed and the slopes of the tone mapping operator is found similar to Eq. (2). Then, the resulting tone mapping is applied to the frame.

5. EXPERIMENTAL RESULTS AND COMPARISONS

We have utilized a look-up table for PU encoding and inverse PU encoding, which gives the absolute luminance value and its corresponding PU encoded value as a pair, rather than fitting a model, during the experiments. A look-up table offers better accuracy and is usually faster to compute than power or logarithmic functions approximating such encodings [4]. The size of the table for the whole luminance range from 10^{-2} to 10^7 cd/m² is about 96 kB, which is quite small for storage both at the encoder and decoder.

The performance of the proposed tone mapping based on perceptually uniform encoded values is compared with the tone mapping based on luminance values [1]. The comparisons are performed with respect to the quality of the produced LDR images and with respect to the quality of reconstructed HDR image after the coding of LDR images at different bitrates.

Figure 6 illustrates the resulting tone mapping curves and the histograms for l and l_u , respectively, for the HDR test images *Atrium Night*, *Memorial*, *Bristol Bridge* and *Forest Path* [13], [14]. The number of the histogram bins in the proposed method is held equal to the one in the compared method for a fair comparison. The resulting tone mapping curves for the first three images are different for both approaches. Tone mapping curves for *Forest path* are similar as the luminance range of *Forest Path* is mostly higher than the 100 cd/m², where PU encoding is mostly linear with respect to the logarithm of the luminance (Fig. 2).

Figure 7 and 8 shows the produced LDR images for both approaches¹. Some over-illumination spread over the whole image is observable for the compared method [1] in particular for *Atrium Night* and *Bristol Bridge*. Such an affect can also be seen in the high luminance regions (e.g. the windows and the dome) of *Memorial*. The proposed method eliminates these effects by using the perceptually uniform luminance values and produces more natural LDR images. A drawback is seen however for the darker regions of *Memorial*. Although the high luminance regions are more realistic for the proposed approach, the details in the darker regions (e.g. the ceiling in the middle-left of *Memorial*) are less visible in contrast to the compared method [1]. The produced LDR images for *Forest Path* are similar as the tone mapping curves are very close.

Figure 9 shows the rate distortion curves in terms of coding bit rate of LDR image vs. quality of the reconstructed HDR image. We have used three different quality metrics for the comparisons, namely, HDR MSE (i.e. mean square error between the logarithm of luminance values of original and reconstructed HDR images), PU-PSNR [4] and the HDR visible difference predictor (HDR VDP) [15]. H.264/AVC reference software JM 17.2 [16] is used for the encoding of LDR image. As expected, the compared method [1] gives lower MSE results than the proposed method, as this method is optimally designed to give minimum MSE for the logarithm of luminance values. However, the proposed method gives better PU-PSNR results as it minimizes the MSE between the perceptually uniform values. While the proposed method gives better HDR VDP results for *Atrium Night* and *Bristol Bridge*, the compared method is better for *Memorial*. The leading method in terms of HDR VDP shows variations with respect to the HDR test image.

6. CONCLUSIONS

We have revealed that an optimized tone mapping [1] which is designed to minimize the MSE between the logarithms of luminance values is not optimum in perceptual sense in two aspects. First, the produced LDR images can show some over illumination in the bright regions. Second, the quality of the reconstructed images can be further improved with respect to the perceptual metrics. We have proposed a method based on the perceptually uniform encoding of luminance values, which eliminates such a disadvantage and gives better results in terms of the perceptual metric, PU-PSNR. We have also shown some limitations of the proposed method in revealing the contrast in dark regions in contrast to the compared method [1].

ACKNOWLEDGMENTS

This work was partly funded by the *Région Ile de France* in the framework of the *NEVEx* project.

REFERENCES

- [1] Z. Mai, H. Mansour, R. Mantiuk, P. Nasiopoulos, R. Ward, and W. Heidrich, "Optimizing a Tone Curve for Backward-Compatible High Dynamic Range Image and Video Compression", *IEEE Transactions on Image Processing*, pp. 1558-1571, Vol. 20, No. 6, June 2011.
- [2] R. Mantiuk, G. Krawczyk, R. Mantiuk and H.-P. Seidel, "High dynamic range imaging pipeline: perception-motivated representation of visual content", *Proc. SPIE 6492*, article no, 649212, 2007.
- [3] S. Liu, W.-S. Kim, and A. Vetro, "Bit-depth scalable coding for high dynamic range video," in *Proc. SPIE Vis. Commun. Image Process.*, vol. 6822, 2008.
- [4] T. O. Aydin, R. Mantiuk, and H.-P. Seidel, "Extending Quality Metrics to Full Luminance Range Images", *Human Vision and Electronic Imaging XIII (Proceedings of SPIE)*, pp. 6806-10, Jan. 2008.
- [5] A. Motra and H. Thoma, "An Adaptive LogLuv Transform for High Dynamic Range Video Compression", pp. 2061-2064, in *Proc. IEEE ICIP*, 2010.
- [6] J.-U. Garbas and H. Tahoma, "Temporally Coherent Luminance-to-Luma Mapping for High Dynamic Range Video Coding with H.264/AVC", pp. 829 – 832, in *Proc. IEEE ICASSP*, 2011.
- [7] Y. Zhang, E. Reinhard, and D. Bull "Perception-based High Dynamic Range Video Compression with Optimal Bit Depth Transformation", pp. 1321-1324, in *Proc. IEEE ICIP*, 2011.
- [8] G. W. Larson, "Logluv encoding for full-gamut, high-dynamic range images," *J. Graph. Tools*, vol. 3, no. 1, pp. 15–31, 1998.
- [9] J.M. Boyce, "Weighted prediction in the H.264/MPEG AVC video coding standard," pp. 789-792, in *Proc. Int. Symposium on Circuits and Systems (ISCAS)*, May 2004.
- [10] A. Segall, "Scalable Coding of High Dynamic Range Video", vol. 1, in *Proc. IEEE ICIP*, 2007.
- [11] S. Hecht, "The visual discrimination of intensity and the weber-fechner law," *J. General Physiol.*, vol. 7, no. 2, pp. 235–267, Nov. 1924.
- [12] A. Segall, L. Kerofsky, and S. Lei, "Tone Mapping SEI Message," Jul. 2006, *Tech. Rep.*, ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-T060.
- [13] HDR image Gallery [online]. Available: <http://www.mpi-inf.mpg.de/resources/hdr/gallery.html>
- [14] HDR image Gallery [online]. Available: <http://www.anyhere.com/gward/hdrenc/pages/originals.html>
- [15] R. Mantiuk, S. J. Daly, K. Myszkowski, and H.-P. Seidel, "Predicting visible differences in high dynamic range images – model and its calibration", in *SPIE Human Vision and Electronic Imaging X*, vol. 5666, pp. 204-214, San Jose, USA, 2005.
- [16] H.264/AVC JM17.2 Reference Software [Online]. Available: <http://iphome.hhi.de/suehring/tml/>

¹ The LDR images and the configuration files for JM 17.2 are available at <http://perso.telecom-paristech.fr/~dufaux/VCIP2012/>

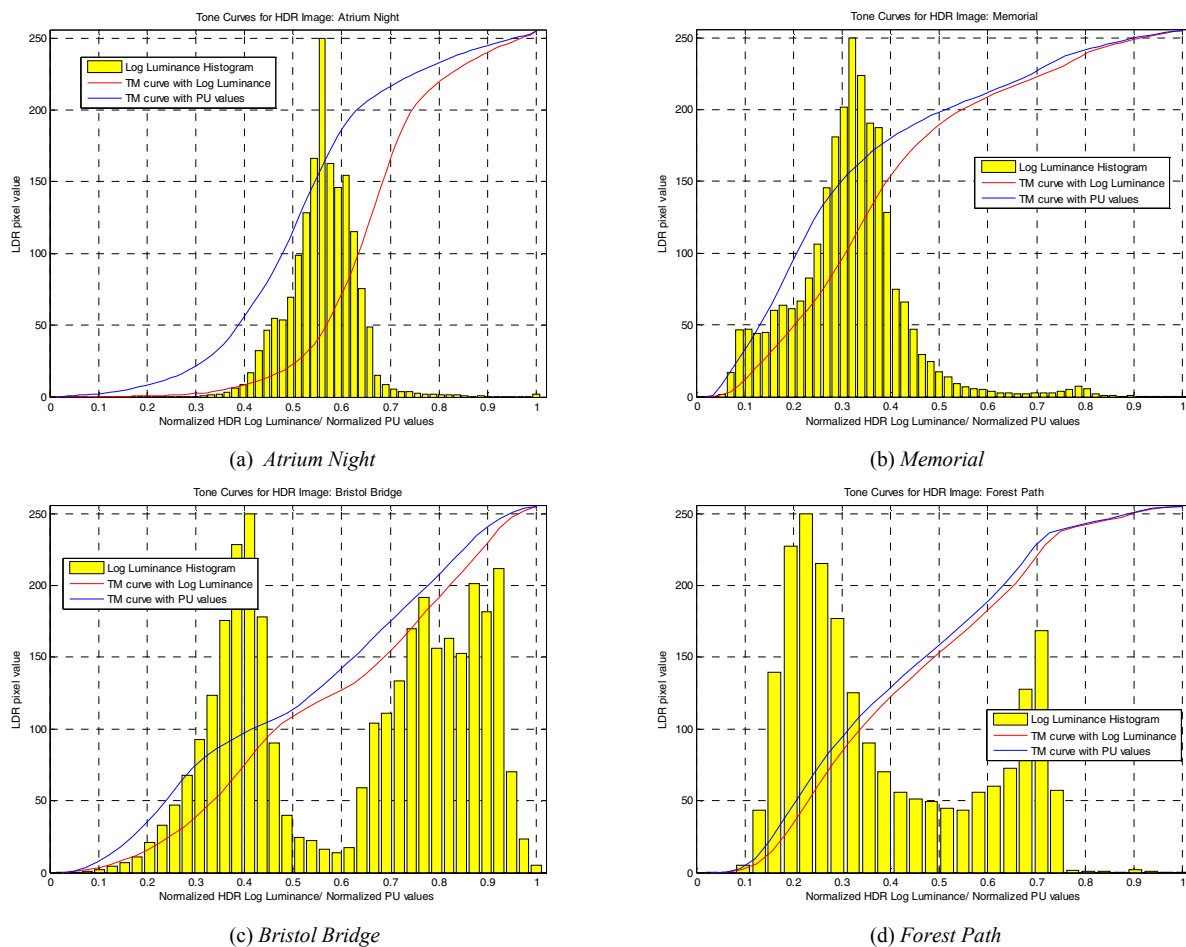


Fig. 6 Tone curves with log. luminance [1] and with the proposed PU encoding. (a) *Atrium Night*, $l_{max} = 3.95$; $l_{min} = -1.77$; $l_{u, max} = 506.62$; $l_{u, min} = 0.32$, (b) *Memorial*, $l_{max} = 5.05$; $l_{min} = -0.48$; $l_{u, max} = 673.89$; $l_{u, min} = 25.88$, (c) *Bristol Bridge*, $l_{max} = 3.51$; $l_{min} = -0.34$; $l_{u, max} = 441.58$; $l_{u, min} = 30.93$, (d) *Forest Path*, $l_{max} = 4.38$; $l_{min} = 1.36$; $l_{u, max} = 572.81$; $l_{u, min} = 149.04$. (Horizontal axes of the graphs are normalized with respect to the max. and min. values of log. luminance and PU values.)

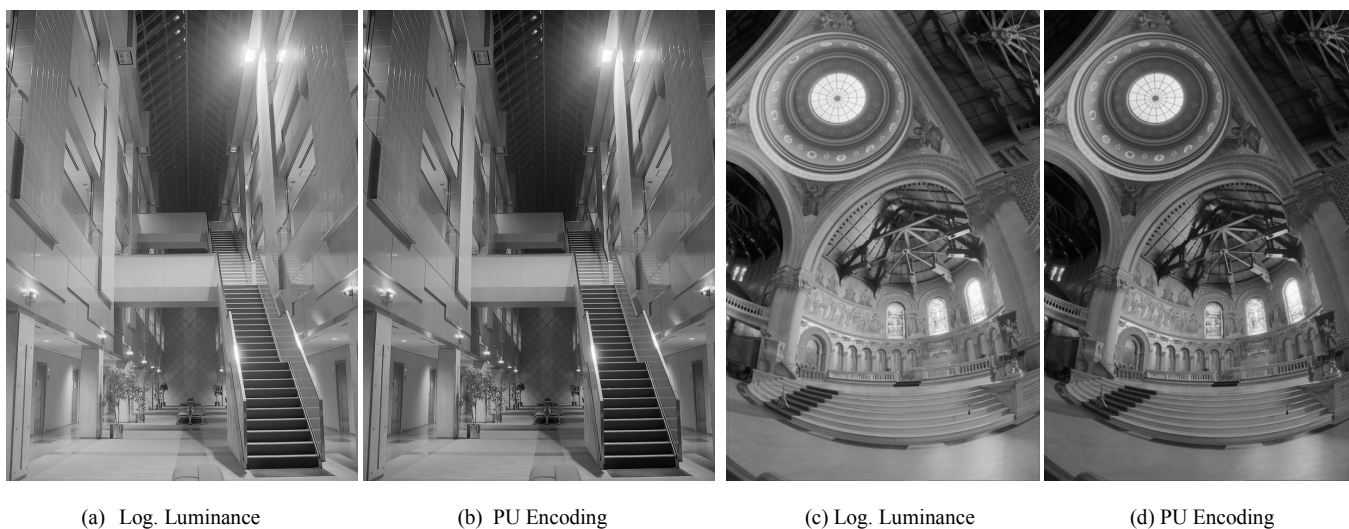
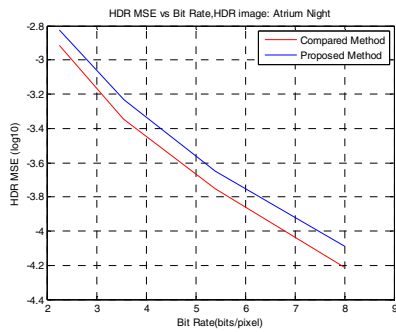


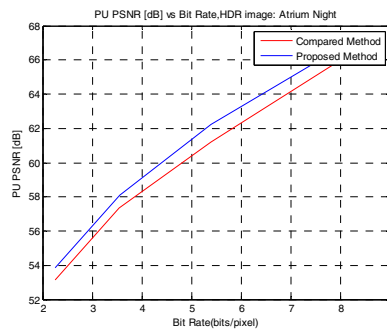
Fig. 7 LDR images produced with log. luminance [1] and with the proposed PU encoding. (a) and (b): *Atrium Night*, (c) and (d): *Memorial*.



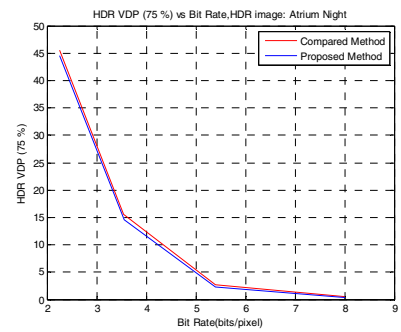
Fig. 8 LDR images produced with log. luminance [1], and with the proposed PU encoding for *Bristol Bridge*.



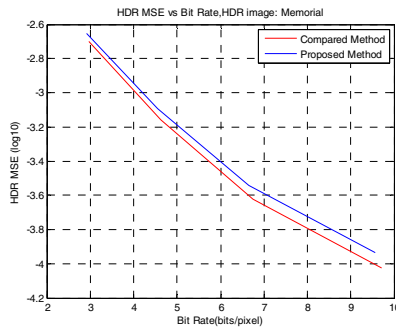
(a) *Atrium Night* HDR MSE



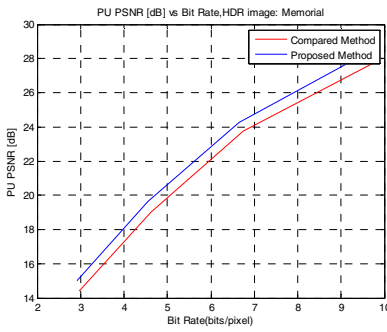
(b) *Atrium Night* PU-PSNR



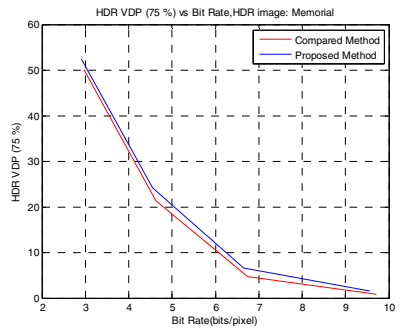
(c) *Atrium Night* HDR VDP



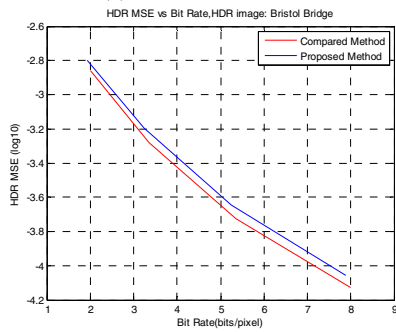
(d) *Memorial* HDR MSE



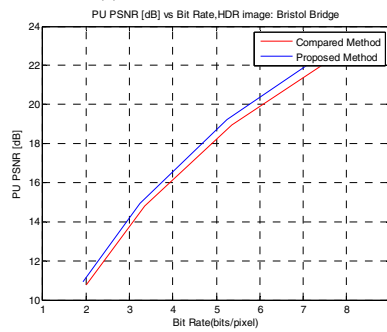
(e) *Memorial* PU-PSNR



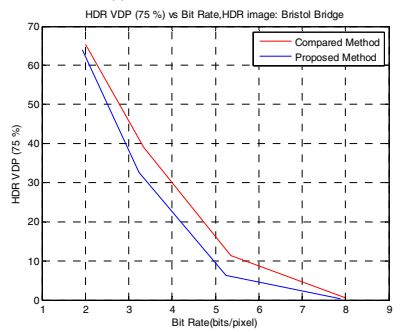
(f) *Memorial* HDR VDP



(g) *Bristol Bridge* HDR MSE



(h) *Bristol Bridge* PU-PSNR



(i) *Bristol Bridge* HDR VDP

Fig. 9 HDR MSE (log10), PU-PSNR (dB) and HDR VDP (75 %) results for *Atrium Night*, *Memorial* and *Bristol Bridge* for the proposed method and compared method [1]. Some encoding parameters for the utilized H.264/AVC reference software JM 17.2 are as follows: ProfileIDC = 244; LevelIDC = 40; YUVformat=3 (i.e. 4:4:4)