



HAL
open science

A comparative survey on high dynamic range video compression

Alper Koz, Frederic Dufaux

► **To cite this version:**

Alper Koz, Frederic Dufaux. A comparative survey on high dynamic range video compression. Applications of Digital Image Processing XXXV, SPIE, Aug 2012, San Diego, United States. hal-01436381

HAL Id: hal-01436381

<https://hal.science/hal-01436381>

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.

A comparative survey on 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

High dynamic range (HDR) video compression has until now been approached by using the high profile of existing state-of-the-art H.264/AVC (Advanced Video Coding) codec or by separately encoding low dynamic range (LDR) video and the residue resulted from the estimation of HDR video from LDR video. Although the latter approach has a distinctive advantage of providing backward compatibility to 8-bit LDR displays, the superiority of one approach to the other in terms of the rate distortion trade-off has not been verified yet. In this paper, we first give a detailed overview of the methods in these two approaches. Then, we experimentally compare two approaches with respect to different objective and perceptual metrics, such as HDR mean square error (HDR MSE), perceptually uniform peak signal to noise ratio (PU PSNR) and HDR visible difference predictor (HDR VDP). We first conclude that the optimized methods for backward compatibility to 8-bit LDR displays are superior to the method designed for high profile encoder both for 8-bit and 12-bit mappings in terms of all metrics. Second, using higher bit-depths with a high profile encoder is giving better rate-distortion performances than employing an 8-bit mapping with an 8-bit encoder for the same method, in particular when the dynamic range of the video sequence is high. Third, rather than encoding of the residue signal in backward compatible methods, changing the quantization step size of the LDR layer encoder would be sufficient to achieve a required quality. In other words, the quality of tone mapping is more important than residue encoding for the performance of HDR image and video coding.

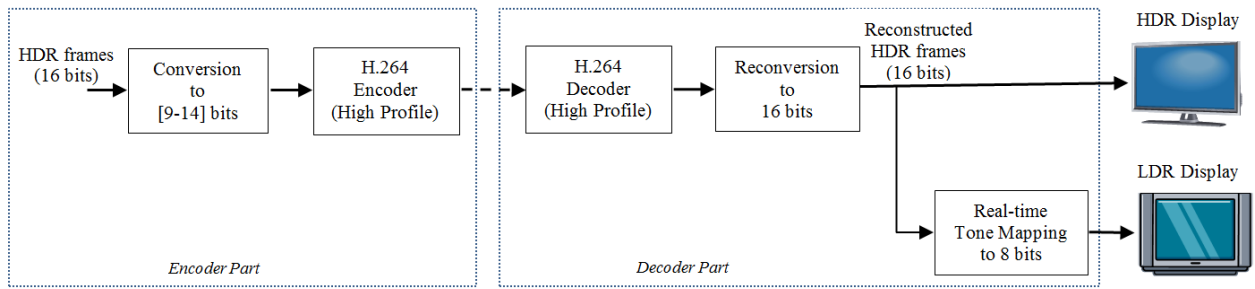
Keywords: High Dynamic Range Video Compression, Tone Mapping, Perceptually Uniform Luminance Values, Backward Compatibility, H.264/AVC

1. INTRODUCTION

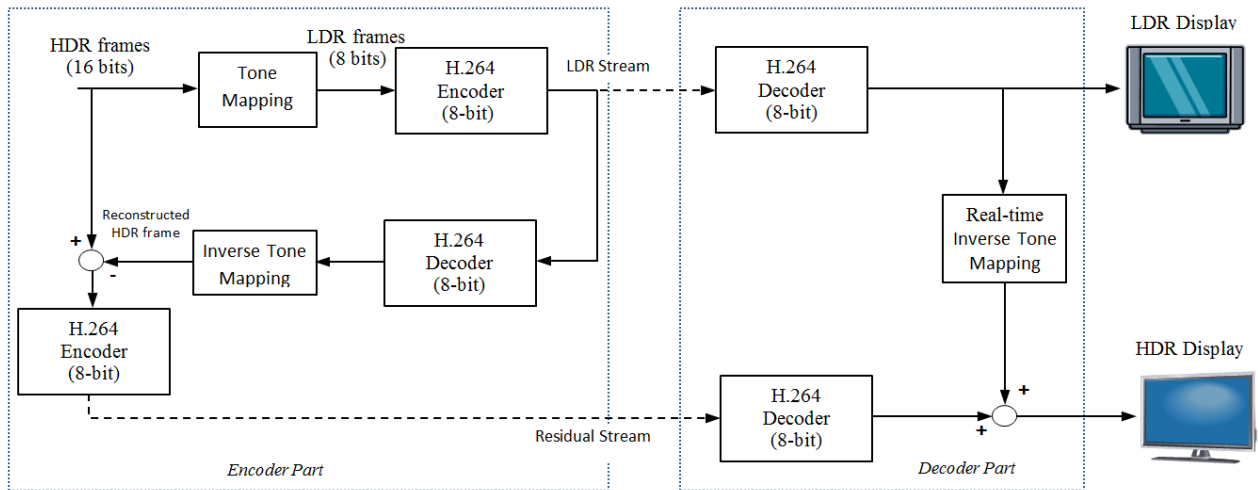
Video content represented by eight bits per pixel has been widely accepted in many applications such as internet/video streaming, DVD, HDTV, camcorders, personal computers, televisions and other consumer electronics products. The reasons of such a wide acceptance of eight-bit representation can be argued 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 ($0.1\text{-}80\text{ cd/m}^2$) [1]-[6]. 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 $500\text{-}1000\text{ cd/m}^2$ 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/m^2 . Accordingly, such a limited representation has become insufficient for these new technologies and related applications in digital cinema, medical imaging, and post-production [1]-[6].

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/m^2) to bright sunshine (10^8 cd/m^2), and hence to generate and store a scene referred image independent of the display technology [1]-[6]. 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 (Figure 1). The first approach [7]-[9] is to take the advantage of high profile of existing state-of-the-art H.264/AVC (Advanced Video Coding) codec, which can support the bit-depths



(a)



(b)

Figure 1. A general scheme for the two approaches in HDR video compression, (a) using high profile H.264/AVC codec, and (b) using 8-bit H.264/AVC codec for backward compatibility to LDR displays.

more than 8 bits up to 14 bits. These methods first convert the real luminance values (cd/m^2) of the HDR video pixels into a bit-depth between [9-14] bits and apply the H.264/AVC high profile encoder. After the decoding and reversion to luminance values (cd/m^2), the reconstructed HDR video is shown on a HDR display or on a LDR display by performing a real-time tone mapping operation, as illustrated in Figure 1-a.

The second approach [1], [2], [4], [5] aims to provide the backward compatibility with existing standard 8-bit displays in order to enable a successful transition to HDR technology (Figure 1-b). The methods in this approach first map the HDR video into 8-bit to produce an LDR video. The LDR video goes through the standard video encoding and decoding process by using the 8-bit H.264/AVC encoder. Then, the high dynamic range video is reconstructed from LDR video by applying a real-time inverse tone mapping operation. The parameters regarding the tone mapping and inverse tone mapping can be sent as a look-up table [1] or can be encoded using supplemental enhancement information [10]. The residue of the reconstructed HDR video can also be further encoded as an enhancement layer in the bit-stream. As this approach includes the LDR stream separately in the resulting bit-stream, it provides a direct compatibility to LDR displays. However, it requires real-time inverse-tone mapping operation to show the video content on HDR displays.

Although the backward compatible methods has a distinctive advantage of providing a better transition to HDR technology, the following different aspects of these two approaches have not been experimentally studied until now to the knowledge of the authors. First, the comparison of two approaches with respect to the rate-distortion trade-off has not been performed. Second, the benefit of using a bit-depth higher than 8-bits for tone mapping with a high profile H.264/AVC encoder over employing 8-bit tone mapping and 8-bit encoder has not been verified. Third, the implementation of the residue coding for backward compatible methods has not been realized yet and therefore, its contribution in terms of rate-distortion gain has also not been confirmed.

In this paper, we experimentally investigate these three mentioned aspects of the two main HDR video coding approaches. In particular, we compare the rate-distortion performances of two main approaches in two ways, first by reducing the methods designed for high-profile encoder into 8-bits and second, by adapting the methods designed for 8-bit encoder to higher bit-depths. For this purpose, the parameters of the tone mapping operations are changed appropriately in both approaches during comparisons. In addition to the rate-distortion curves, we also examine the quality of the produced LDR images and video sequences for different HDR image and video sequences. The next section overviews the HDR video coding methods and gives the details of tone mapping operations. Section 3 gives the experimental results and comparisons for all the mentioned aspects in detail. Finally, the conclusions are given in section 4.

2. HDR VIDEO COMPRESSION CLASSIFICATION

The methods on HDR video compression can be classified in two categories based on the utilization of high profile H.264/AVC encoder or 8-bit standard H.264/AVC encoder.

2.1 Methods using the high profile H.264/AVC encoder and decoder

The core part of the methods in this approach is to convert the real luminance values of HDR video pixels into the bit-depths that the high profile can support. The pioneer work by Thoma *et al.* [7] performs this conversion by applying the modified versions of LogLuv transform [11], which maps the luminance range of each HDR frame to the range $[0, 2^n-1]$ with respect to the maximum and minimum luminance values of the frame, where n is the bit-depth of the resulting video. As this pioneer work forms one of the base methods in our comparisons, we would give the descriptions in more detail.

The LogLuv transform [11] is originally proposed by Larson to provide a compact encoding suitable for the transfer, manipulation, and storage of high dynamic range color images. This format is a replacement for conventional RGB images and encodes color pixels as log luminance values and CIE (u, v) chromaticity components [11] as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.497 & 0.339 & 0.164 \\ 0.256 & 0.678 & 0.066 \\ 0.023 & 0.113 & 0.864 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

$$x = \frac{X}{X + Y + Z} \quad (2)$$

$$y = \frac{Y}{X + Y + Z} \quad (3)$$

$$u_e = \left[410 \frac{4x}{-2x + 12y + 3} \right] \quad (4)$$

$$v_e = \left[410 \frac{4y}{-2x + 12y + 3} \right] \quad (5)$$

$$L = \left[256(\log_2 Y + 64) \right] \quad (6)$$

Such a transform maps the real luminance values (Y) in the interval $[5.44 \times 10^{-20}, 1.84 \times 10^{19}]$ to 15 bit integer luma values (L) in the range $[0, 2^{15}-1]$ and covers over 38 orders of magnitude in 0.27 % relative step size. As this step size is much below the visible quantization threshold of 1 % [12], Thoma *et al.* [7] modify the LogLuv mapping to exploit the full range of luma code values for a given bit depth:

$$L_n = \left[410 \frac{2^n - 1}{\log_2 \left(\frac{Y_{\max}}{Y_{\min}} \right)} (\log_2(Y) - \log_2(Y_{\min})) \right] \quad (7)$$

$$Y = 2^{(L_n + 0.5) \left(\frac{\log_2(Y_{\max} / Y_{\min})}{2^n - 1} \right) + \log_2(Y_{\min})} \quad (8)$$

where Y_{\max} and Y_{\min} are the maximum and minimum luminance values of a HDR video frame. The transformation is applied to each frame of the HDR video, and Y_{\max} and Y_{\min} for each frame are also transferred with the bit-stream to perform the reconversion (Figure 1-a).

In the extended version of this work [8], temporal motion compensation problems generated by individual mapping of frames is solved by weighted prediction (WP) tool of H.264/AVC [13]. In a similar work [9], 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.

2.2 Methods backward compatible to 8-bit LDR Displays

The methods in this approach [1], [2], [4], [5] are proposed to provide the backward compatibility with existing standard 8-bit displays. 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 [4], 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 [5] 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 in terms of mean square error of logarithm of luminance values of the original and reconstructed HDR frames (HDR MSE) is formulated in a detailed and complete manner. In order to find a computationally efficient solution, Mai *et al.* [1] 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. We extend this work in our recent study [2] by means of using perceptually uniform luminance values instead of the logarithmic values for optimization and show the superior results in terms of the perceptual metric, PU PSNR. As these methods form the latest state of the art for backward compatible HDR video compression, we would give the descriptions in more detail.

Optimum tone mapping for backward compatible HDR video compression

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 the proposed solution [1] is first parameterized as a piece-wise linear function with the nodes (l_k, v_k) as shown in Figure 2. 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

$$s_k = \frac{v_{k+1} - v_k}{\delta} \quad (9)$$

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

$$s_k = \frac{v_{\max} \cdot p_k^{1/3}}{\delta \cdot \sum_{k=1}^N p_k^{1/3}} \quad (10)$$

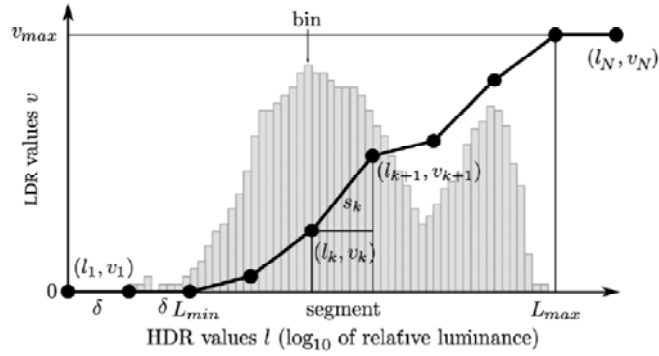


Figure 2. Parameterization of the tone-mapping curve in [1]. The bar-plot is the luminance histogram of the HDR frame.

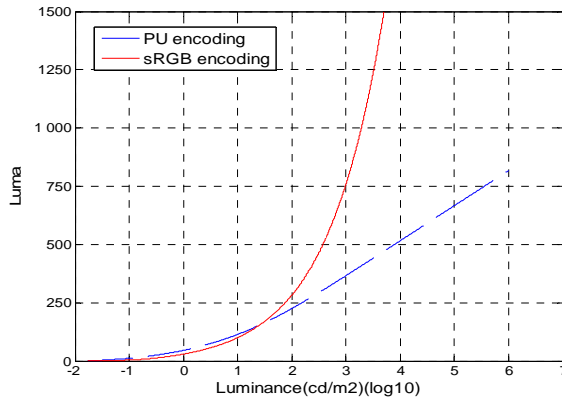


Figure 3. Perceptually Uniform Encoding and sRGB encoding as a function of luminance (cd/m^2) [6].

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].

Optimum tone mapping with perceptually uniform encoded luminance values

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 [12]. However, *Weber* law is not an exact model of HVS contrast sensitivity for the whole luminance range [6]. 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.

Considering the insufficiency of the *Weber* law, Aydin *et al.* in [6] has developed a perceptual mapping (encoding) which converts the luminance values into perceptually uniform encoded values for the whole visible luminance range. Such an encoding 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 [6]. The resulting PU encoding is shown in Figure 3. While there is a closer match to sRGB color transform used in typical displays for the range of 1-100 cd/m^2 , the difference for the higher luminance is quite significant.

In a recent study [2], we propose to use *perceptually uniform encoding* [6] instead of the logarithm operation, and to minimize the mean square error between the perceptually uniform coded values of original and reconstructed frames. Figure 4 illustrates the main stages of the proposed method [2] in parallel with Figure 2. First, the luminance values of

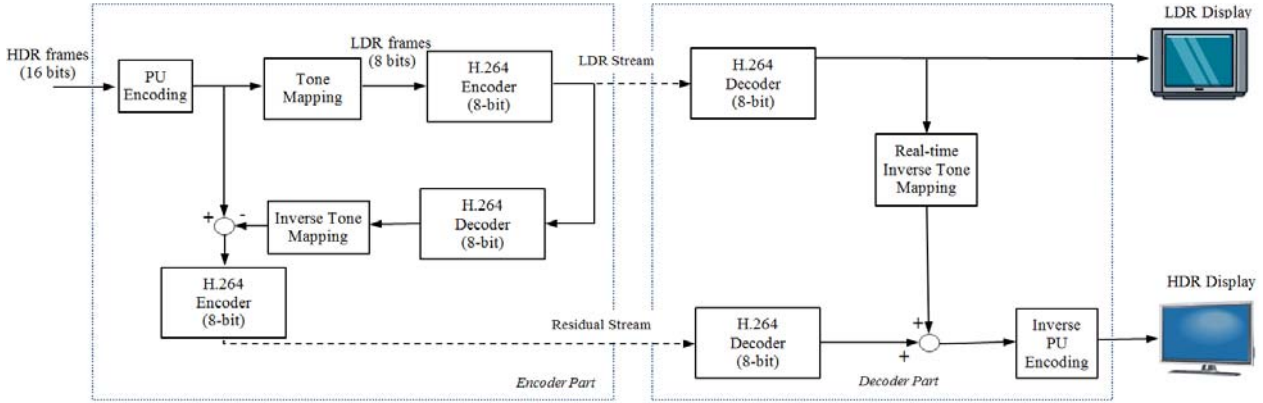


Figure 4. Proposed scheme in [2] for backward compatible HDR video compression with PU encoding.

HDR frames are converted into perceptually uniform values by PU encoding (Fig. 3). The resulting values are represented as l_u . Then, the optimized tone mapping is applied to l_u 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. (10). Then, the resulting tone mapping is applied to the frame.

3. COMPARISONS OF TWO COMPRESSION CLASSES

We compare the two backward compatible methods which are designed to minimize the HDR MSE and PU-PSNR respectively [1] [2], and the adaptive LogLuv transform based method using high profile H.264/AVC encoder [7]. The comparisons are performed with respect to the quality of the produced LDR images and video sequences, and with respect to the quality of reconstructed HDR image and video sequences after the coding of LDR versions at different bitrates. We have used three different HDR quality metrics, namely, HDR MSE, PU-PSNR and the HDR visible difference predictor (HDR VDP) [14]. H.264/AVC reference software JM 17.2 [15] is used for the encoding of LDR image and video sequences. Four HDR test images (*Atrium Night*, *Memorial*, *Bristol Bridge*, and *Forest Path* [16] [17]) and two HDR test sequences (first 30 frames *Sun* and *Tunnel* [18]) are utilized for the rate-distortion curves.

The following four cases are considered for the comparisons.

- (i) All the methods are compared for 8 bit tone mapping. The parameters n (Eq. 7) and v_{max} (Eq. 10) are fixed as 8 and 255, respectively (no residue coding).
- (ii) All the methods are compared for 12 bit mapping. The parameters n (Eq. 7) and v_{max} (Eq. 10) are fixed as 12 and 4095, respectively (no residue coding).
- (iii) Rate-distortion curves for 8-bit mapping and 12-bit mapping are compared to investigate the benefit of high profile encoder (no residue coding).
- (iv) Rate-distortion curves with and without residue coding are compared to investigate the benefit of residue coding.

3.1 Comparison with respect to 8-bit mapping

Figure 5 illustrates the resulting tone mapping curves for all methods and the luminance histograms for the four HDR test images, and the first frames of *Sun* and *Tunnel* sequences. The numbers of the histogram bins in the compared methods are held equal for a fair comparison. The resulting tone mapping curves for the first three images are different for all approaches. Tone mapping curves for the methods [1] and [2] for *Forest path* are similar as the luminance range of *Forest Path* is mostly higher than the 100 cd/m^2 , where PU encoding is mostly linear with respect to the logarithm of

the luminance (Figure 3). Similarly, tone mapping curves for the three methods are very close for *Tunnel* sequence due to the luminance distribution.

Figure 6 shows the produced LDR images for all 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 in [2] 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 [2], 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 for the methods [1] and [2] as the tone mapping curves are very close. For the case of [7], the linear characteristics of the tone mapping curve with respect to logarithmic values is producing much dimmer and duller LDR images, as such a mapping does not fit well to HVS characteristics.

Figure 7 shows the rate distortion curves in terms of coding bit rate of LDR image vs. quality of the reconstructed HDR image. As expected, the compared method [1] gives lower MSE results than the proposed method in [2], as this method is optimally designed to give minimum MSE for the logarithm of luminance values. However, the proposed method [2] gives better PU-PSNR results as it minimizes the MSE between the *perceptually uniform* values. As an exception, the performance of the methods [1] and [2] are very close for *Forest Path* due to very close tone mapping curves. For the case of HDR VDP, while the method [2] gives better HDR VDP results for *Atrium Night* and *Bristol Bridge*, the compared method [1] is better for *Memorial*. The leading method in terms of HDR VDP shows variations with respect to the HDR test image. The method in [7] shows the minimum achievement for all the images and metrics, which clearly indicates that the adaptive LogLuv transform is not suitable for 8-bit backward compatible compression.

Figure 8 shows the rate distortion curves for Sun and Tunnel sequences. The same conclusions are very apparent for Sun sequence. However, the contrast is getting lower for Tunnel due to the very close tone mapping curves for this sequence (Fig. 5-f).

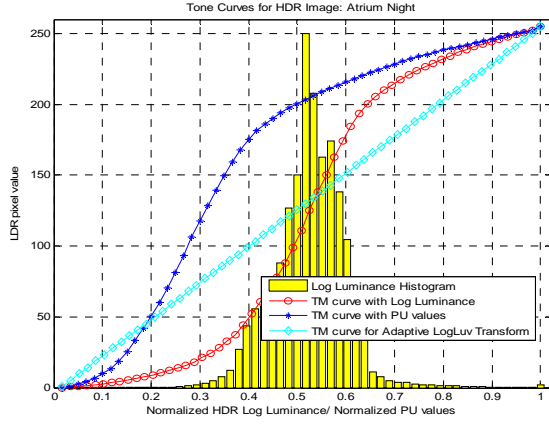
3.2 Comparison with respect to 12-bit mapping

Figure 9 and 10 shows the rate distortion curves for 12 bit mapping using high profile H.264/AVC encoder for the test images and video sequences, respectively. It is important to observe that the methods optimally designed for backward compatibility to 8-bit displays [1]-[2] works also well for higher bit-depths than 8-bits. Although the method in [7] is the first method introducing the use of high profile H.264/AVC encoder for HDR video coding, its adaptive LogLuv transform based mapping is far from being optimal as indicated by its lower performance. The same conclusions regarding the specific images and video sequences for 8-bit mapping are also valid for 12-bit mapping due to the similar behaviors in the graphs.

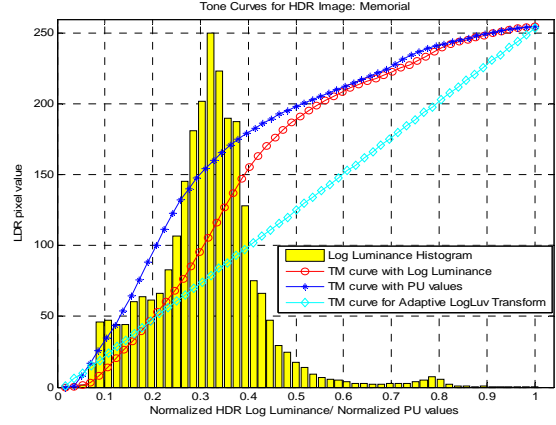
3.3 Comparison of the 8-bit and 12-bit (high profile) H.264/AVC Encoding

In order to judge the benefit of a high profile encoder with a high bit-depth with respect to an 8-bit encoder, we give the rate distortion curves for 8-bit and 12-bit tone mapping together in Figure 11 for all the methods. For Sun sequence, the lower distortion in HDR MSE and HDR VDP for 12-bit mapping compared to 8-bit mapping is very apparent for the adaptive LogLuv transform based method [7] (Figure 11-c and f). However, the difference is relatively smaller for the methods minimizing MSE [1] and PU-PSNR [2]. Based on the luminance distribution of the HDR frames, it might be expected that increasing the bit depth would not have extra gain in the optimized methods. The rate-distortion curves for *Tunnel* sequence are very close to each other. As the *Tunnel* sequence (for the utilized first 30 frames) has a lower dynamic range (i.e. the maximum luminance is about 130-140 cd/m²) compared to the Sun sequence, using higher bit-depths is not offering extra gain.

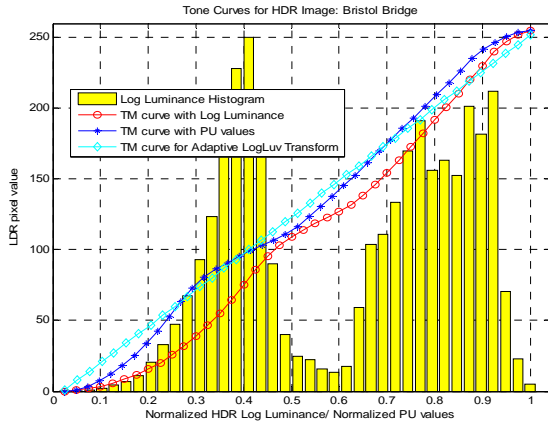
¹ The resulting LDR images and video sequences, and the utilized configuration files for JM 17.2 are available at <http://perso.telecom-paristech.fr/~dufaux/ADIP2012/>



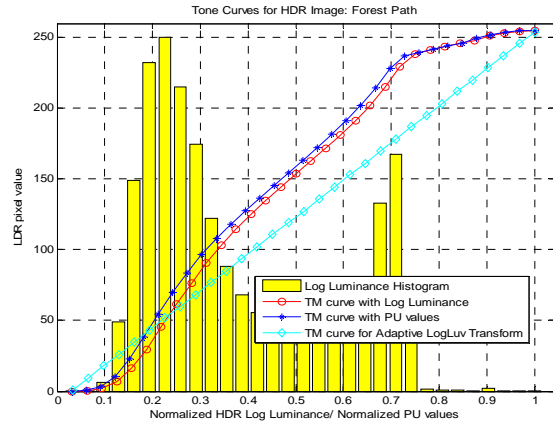
(a) Atrium night



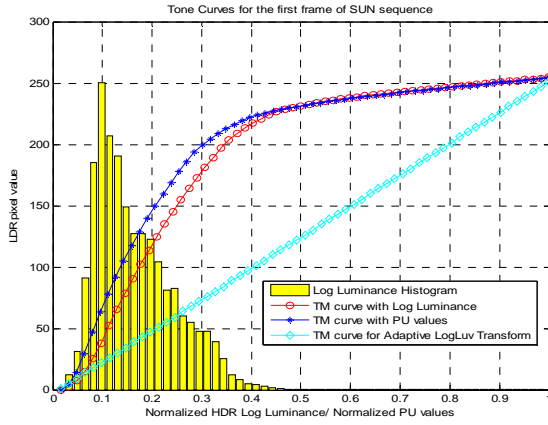
(b) Memorial



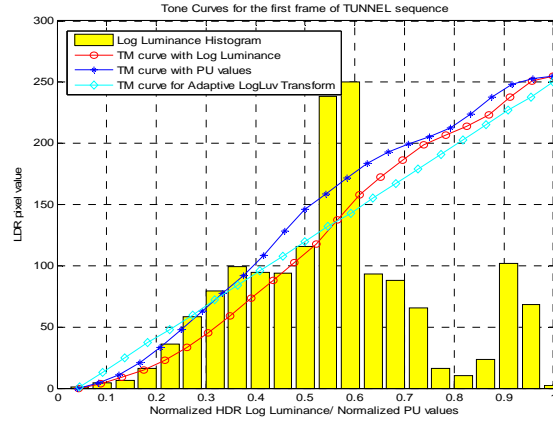
(c) Bristol Bridge



(d) Forest Path



(e) Sun sequence (first frame)



(f) Tunnel sequence (first frame)

Figure 5. Tone curves for log. luminance [1], PU encoding [2], and adaptive loglucv transform based methods [7]. (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*, $l_{max} = 4.38$; $l_{min} = 1.36$; $l_{u,max} = 572.81$; $l_{u,min} = 149.04$. (e) *Sun sequence (first frame)*, $l_{max} = 6.62$; $l_{min} = 0.54$; $l_{u,max} = 910.98$; $l_{u,min} = 77.16$, (f) *Tunnel sequence (first frame)*, $l_{max} = 2.13$; $l_{min} = -0.01$; $l_{u,max} = 241.34$; $l_{u,min} = 45.10$. (Horizontal axes of the graphs are normalized with respect to the max. and min. values of log. luminance and PU values).

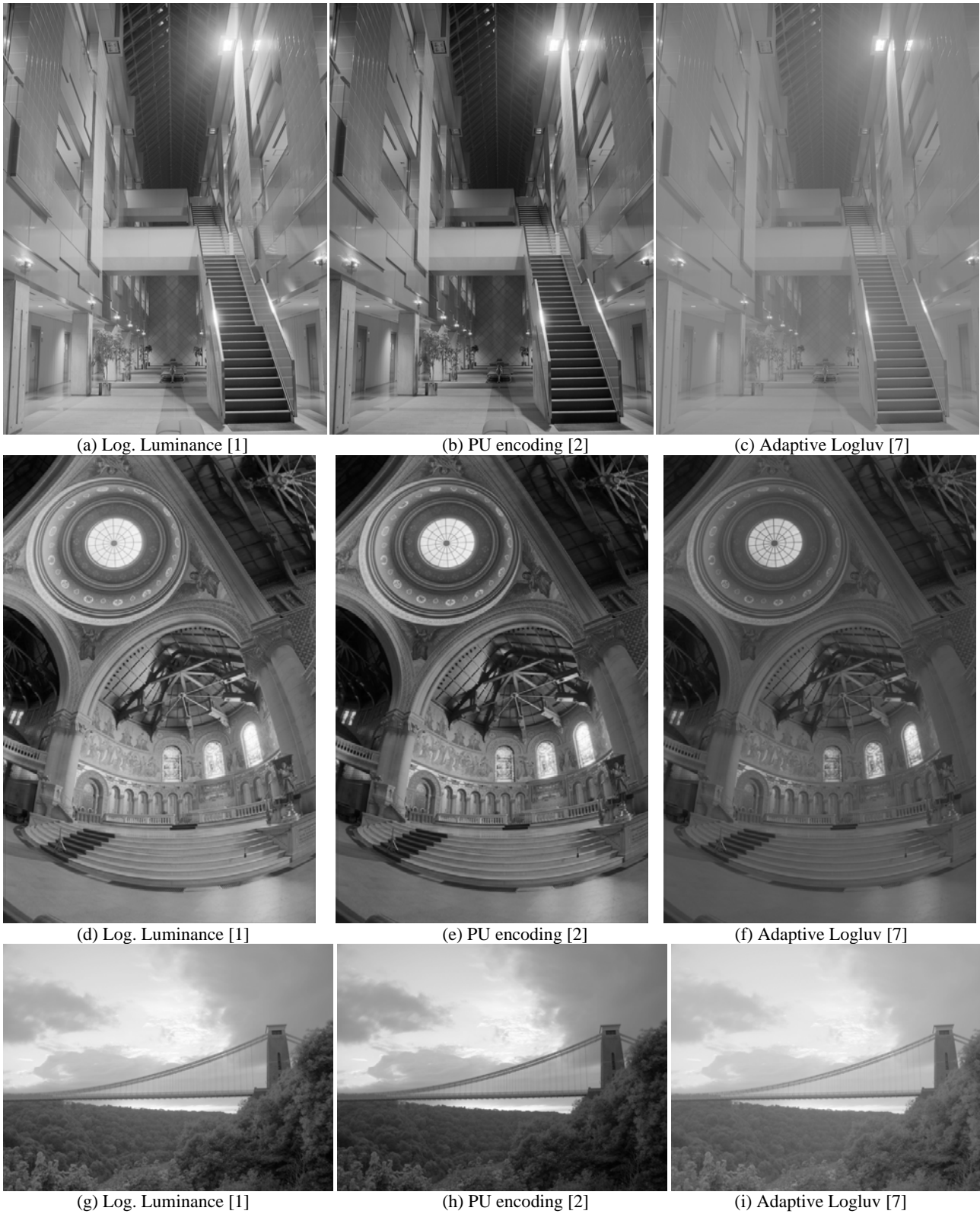


Figure 6. LDR images produced with log. luminance [1], PU encoding [2], and adaptive logluv transform based methods [7]. (a), (b), and (c): *Atrium Night*; (d), (e) and (f): *Memorial*; (h), (i), and (j): *Bristol Bridge*.

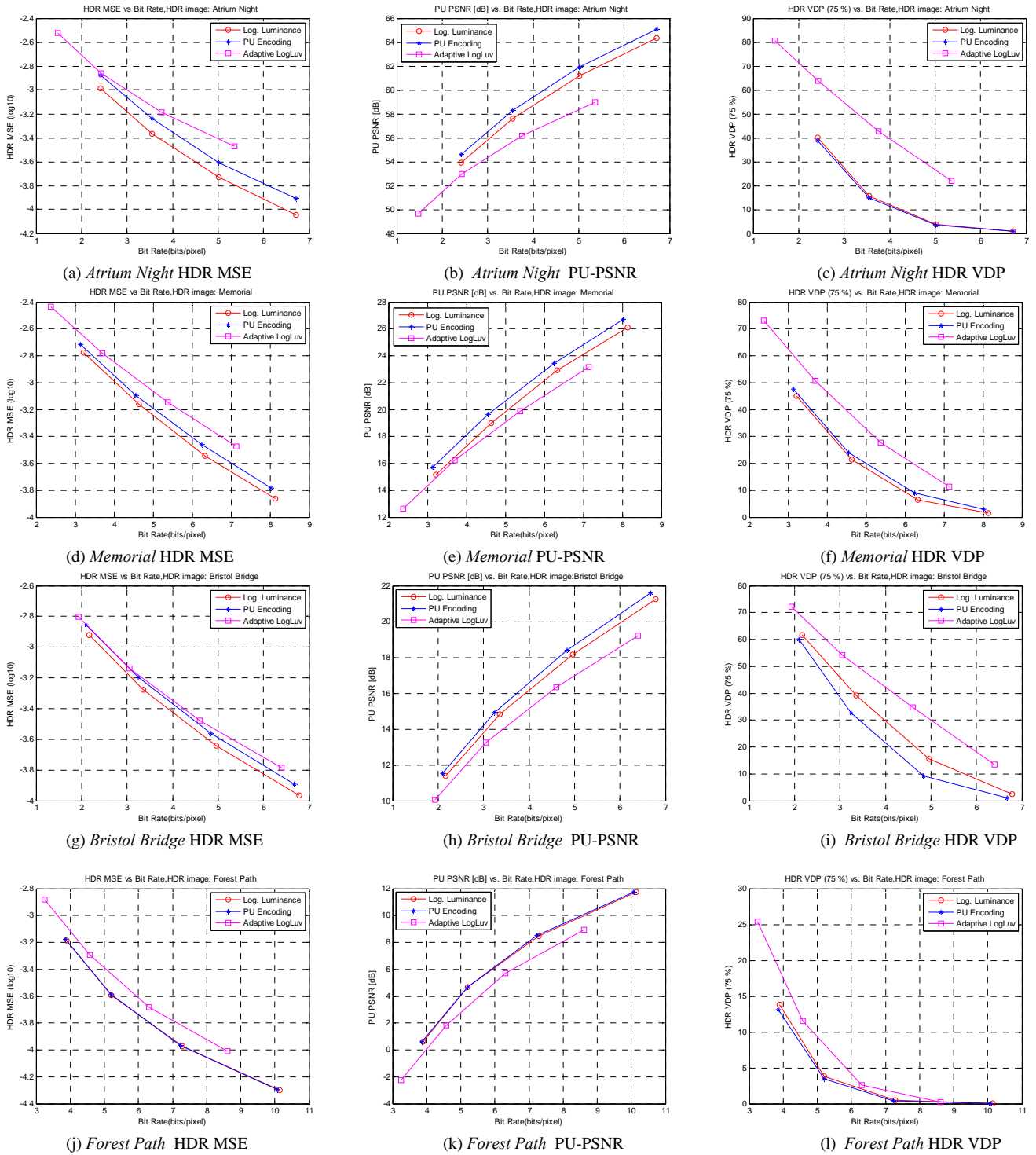


Figure 7. HDR MSE (log10), PU-PSNR (dB) and HDR VDP (75 %) results for *Atrium Night*, *Memorial*, *Bristol Bridge* and *Forest Path* for the log. luminance [1], PU Encoding [2] and Adaptive LogLuv transform [7] based methods for the case of 8-bit mapping. Some encoding parameters for the utilized H.264/AVC reference software JM 17.2 are as follows: ProfileIDC = 244; LevelIDC = 40; YUVformat=3 (i.e. 4:4:4); SourceBitDepthLuma = 8; SourceBitDepthChroma = 8.

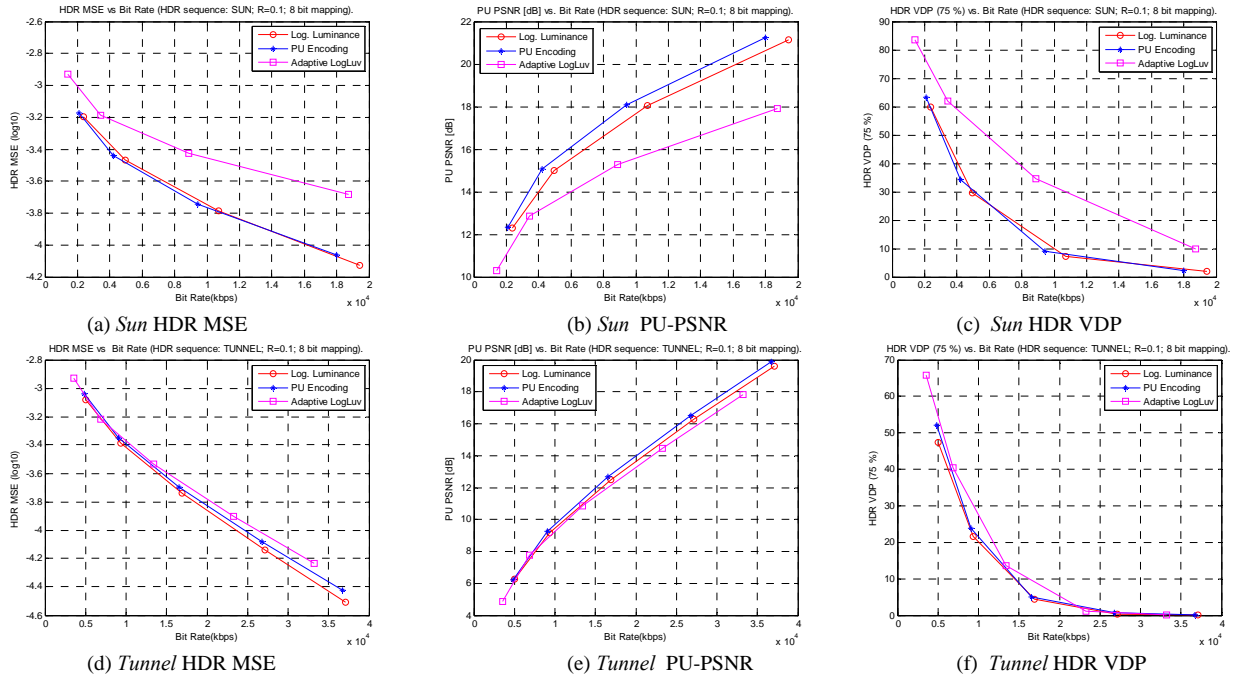


Figure 8. HDR MSE (log10), PU-PSNR (dB) and HDR VDP (75 %) results for Sun and Tunnel HDR Sequences for the log. luminance [1], PU Encoding [2] and Adaptive LogLuv transform [7] based methods for the case of 8-bit mapping. Some encoding parameters for the utilized H.264/AVC reference software JM 17.2 are as follows: ProfileIDC = 244; LevelIDC = 40; YUVformat=1 (i.e. 4:2:0); SourceBitDepthLuma = 8; SourceBitDepthChroma = 8.

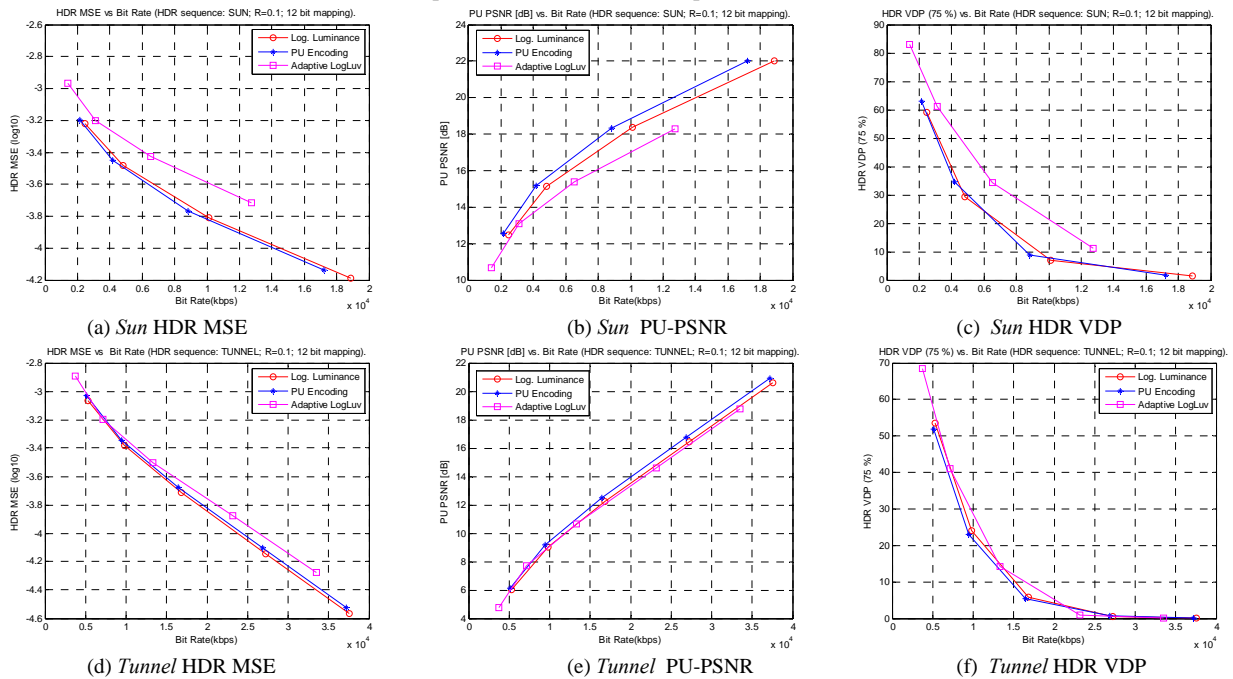
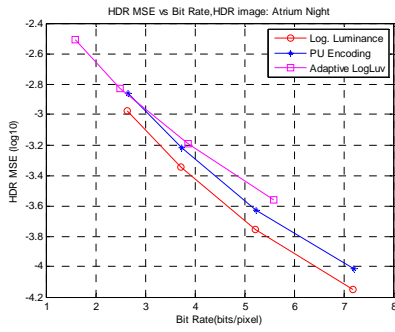
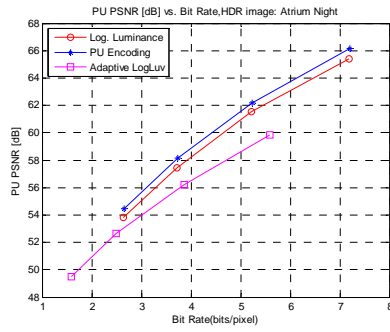


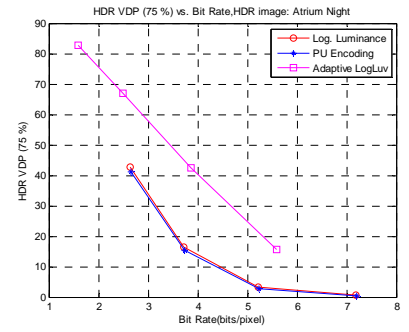
Figure 10. HDR MSE (log10), PU-PSNR (dB) and HDR VDP (75 %) results for Sun and Tunnel HDR Sequences for the log. luminance [1], PU Encoding [2] and Adaptive LogLuv transform [7] based methods for the case of 12-bit mapping. Some encoding parameters for the utilized H.264/AVC reference software JM 17.2 are as follows: ProfileIDC = 244; LevelIDC = 40; YUVformat=1 (i.e. 4:2:0); SourceBitDepthLuma = 12; SourceBitDepthChroma = 8.



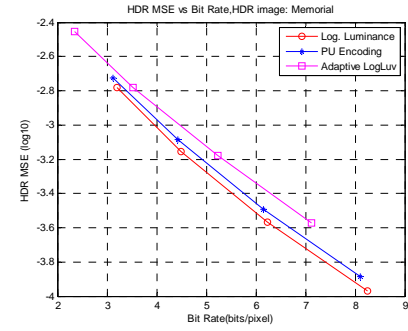
(a) *Atrium Night* HDR MSE



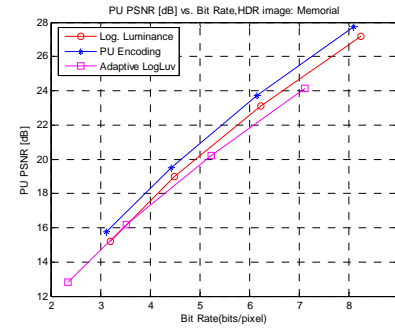
(b) *Atrium Night* PU-PSNR



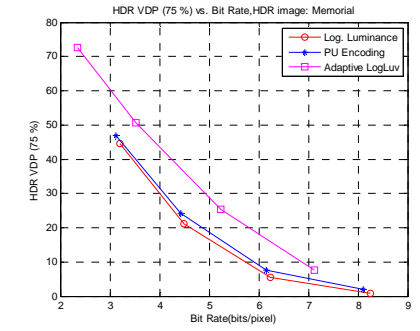
(c) *Atrium Night* HDR VDP



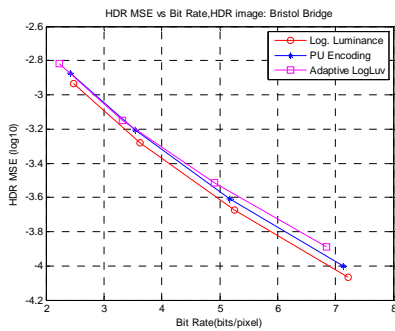
(d) *Memorial* HDR MSE



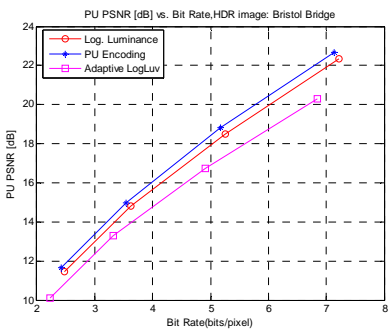
(e) *Memorial* PU-PSNR



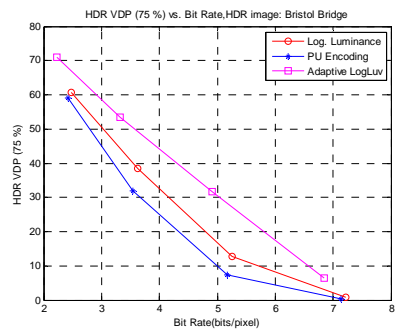
(f) *Memorial* HDR VDP



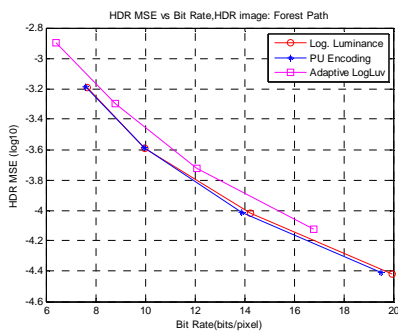
(g) *Bristol Bridge* HDR MSE



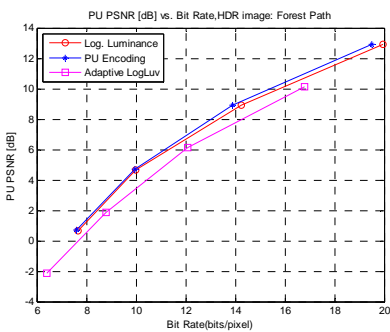
(h) *Bristol Bridge* PU-PSNR



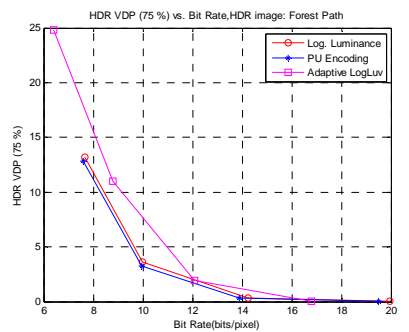
(i) *Bristol Bridge* HDR VDP



(j) *Forest Path* HDR MSE

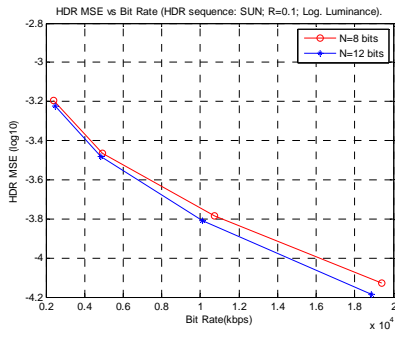


(k) *Forest Path* PU-PSNR

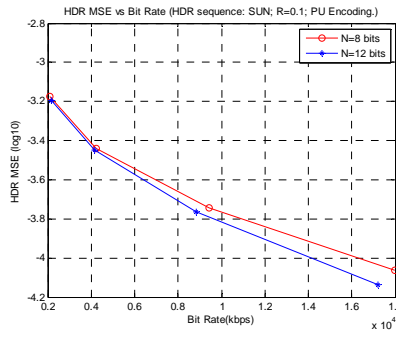


(l) *Forest Path* HDR VDP

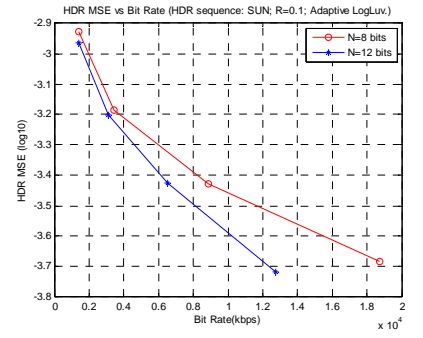
Figure 9. HDR MSE (log10), PU-PSNR (dB) and HDR VDP (75 %) results for *Atrium Night*, *Memorial*, *Bristol Bridge* and *Forest Path* for the log. luminance [1], PU Encoding [2] and Adaptive LogLuv transform [7] based methods for the case of 12-bit mapping. Some encoding parameters for the utilized H.264/AVC reference software JM 17.2 are as follows: ProfileIDC = 244; LevelIDC = 40; YUVformat=3 (i.e. 4:4:4); SourceBitDepthLuma = 12; SourceBitDepthChroma = 8.



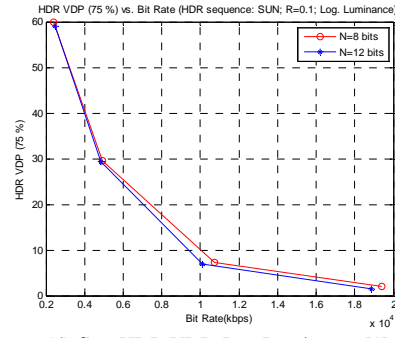
(a) *Sun*, HDR MSE, Log Luminance [1]



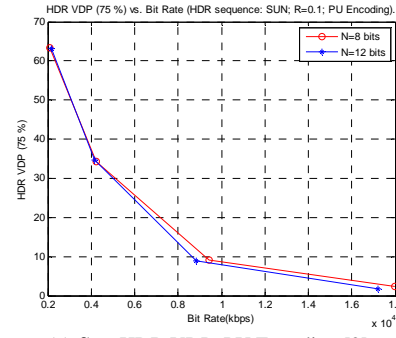
(b) *Sun*, HDR MSE, PU Encoding [2]



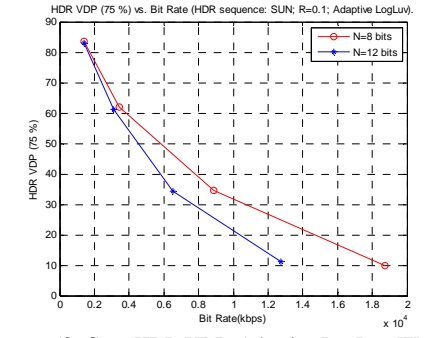
(c) *Sun*, HDR MSE, Adaptive LogLuv [7]



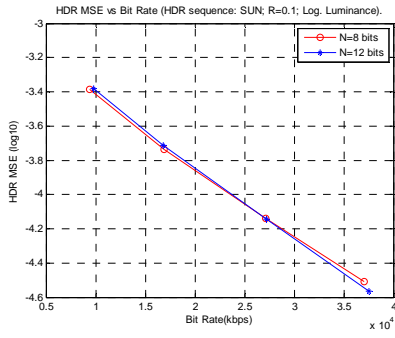
(d) *Sun*, HDR VDP, Log Luminance [1]



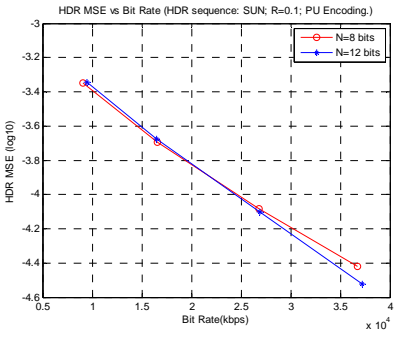
(e) *Sun*, HDR VDP, PU Encoding [2]



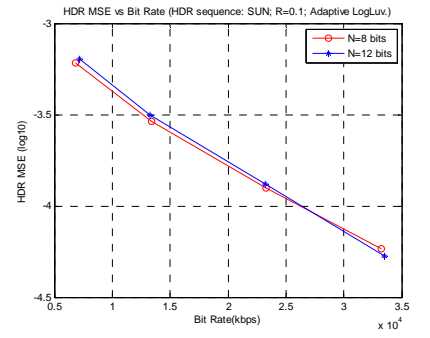
(f) *Sun*, HDR VDP, Adaptive LogLuv [7]



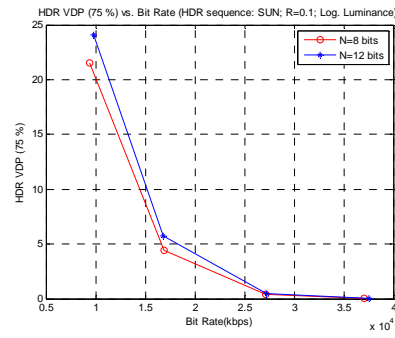
(g) *Tunnel*, HDR MSE, Log Luminance [1]



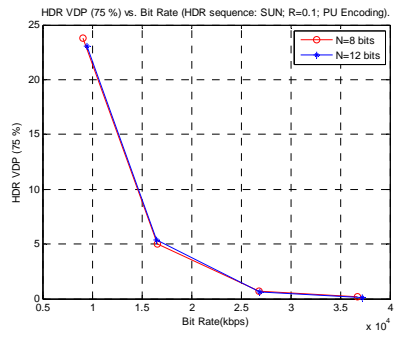
(h) *Sun*, HDR MSE, PU Encoding [2]



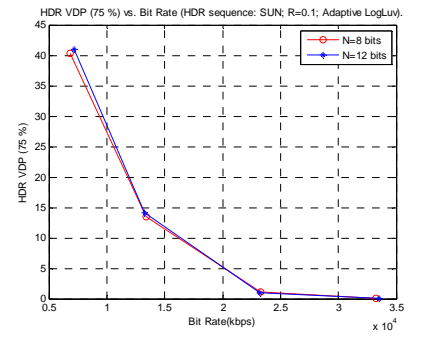
(i) *Sun*, HDR MSE, Adaptive LogLuv [7]



(j) *Sun*, HDR VDP, Log Luminance [1]



(k) *Sun*, HDR VDP, PU Encoding [2]



(l) *Sun*, HDR VDP, Adaptive LogLuv [7]

Figure 11. HDR MSE (log₁₀ and HDR VDP (75 %) results for Sun and Tunnel HDR Sequences for N=8 and N=12 bits for the log. luminance [1], PU Encoding [2] and Adaptive LogLuv transform [7] based methods. Some encoding parameters for the utilized H.264/AVC reference software JM 17.2 are as follows: ProfileIDC = 244; LevelIDC = 40; YUVformat=1 (i.e. 4:2:0).

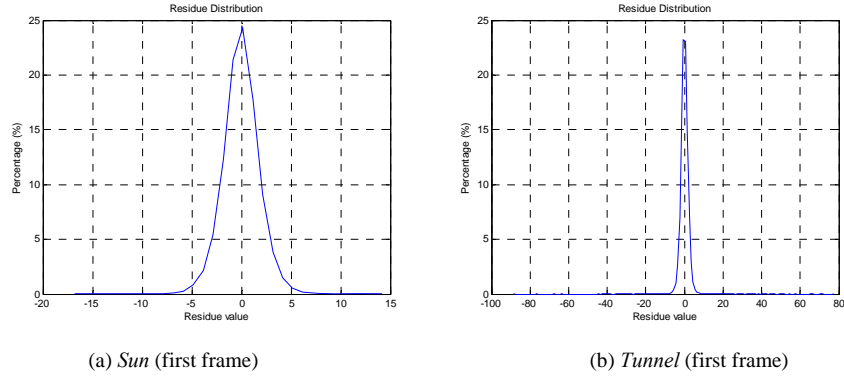


Figure 12. Distribution of the residue for the first frames of *Sun* and *Tunnel* sequences after encoding the tone mapped sequences with the quantization step sizes 20 and 24 respectively for the PU encoding based method [1].

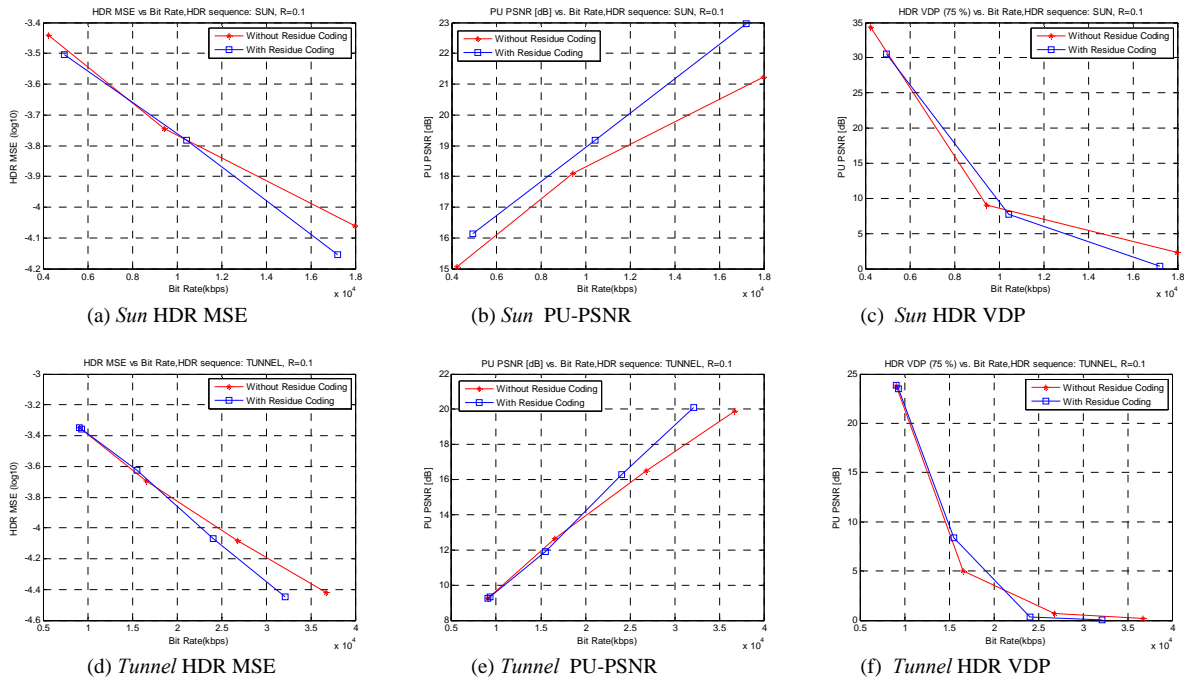


Figure 13. *HDR MSE* (\log_{10}), *PU-PSNR* (dB) and *HDR VDP* (75 %) results with and without *residue coding* for *Sun* and *Tunnel* HDR sequences for the PU Encoding based method [1].

3.4 Comparison with residue coding

For the completeness of the comparisons, we also test the effect of the residue coding to rate-distortion trade-off (Figure 4). Since the residue contains negative values as indicated in Figure 12, we shift all the residue values in a group of pictures (GOP) with the minimum residue over all the frames of GOP and round them to an integer. As a very direct solution, such a shift makes all the residue values positives. Then, we input the residue to the 8-bit encoder.

We apply different quantization step sizes for the residue encoding for the fixed quantization size of the LDR layer encoder. Note that just applying the same quantization size to residue and to the tone mapped LDR frame is not very convenient due to the scale and unit differences. In order to observe the changes in a large distortion interval, we select the quantization size of the LDR layer encoder as 20 and 24 for *Sun* and *Tunnel*, which correspond to 24 % and 34 % HDR VDP values, respectively (Figure 8). Then, the quantization sizes for the residue coding are changed from 8 to 20 and from 8 to 24, respectively. Figure 14 shows the rate-distortion curves with and without residue coding together.

There is not a significant difference in particular for HDR VDP. We conclude that rather than encoding of the residue signal, changing the quantization step size of the LDR layer encoder would be sufficient to achieve a required quality.

4. CONCLUSIONS

We have experimentally compared the two approaches of HDR video compression. We have first indicated that the optimized methods that are designed for backward compatibility to LDR displays are better than the methods designed for high profile of H.264/AVC coder both for 8-bit mapping and 12-bit mapping. Second, using higher bit-depths and high profile encoder is showing a better performance than employing an 8-bit mapping and 8-bit encoder for the same method, in particular when the dynamic range of the video sequences is high as in *Sun* sequence. Third, residue encoding does not offer a significant gain in our tests in terms of rate-distortion trade-off. We conclude that rather than encoding of the residue signal, changing the quantization size of the LDR layer encoder would be sufficient to achieve a required quality. This also shows that the quality of tone mapping is playing the crucial role in HDR image and video coding.

ACKNOWLEDGEMENTS

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] A. Koz and F. Dufaux, "Optimized Tone Mapping with Perceptually Uniform Luminance Values for backward-compatible High Dynamic Range Compression", submitted to *IEEE Visual Communication and Image Processing Conference (VCIP)*, 2012.
- [3] 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.
- [4] A. Segall, "Scalable Coding of High Dynamic Range Video", vol. 1, in *Proc. IEEE ICIP*, 2007.
- [5] 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.
- [6] 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.
- [7] A. Motra and H. Thoma, "An Adaptive LogLuv Transform for High Dynamic Range Video Compression", pp. 2061-2064, in *Proc. IEEE ICIP*, 2010.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] G. W. Larson, "Logluv encoding for full-gamut, high-dynamic range images," *J. Graph. Tools*, vol. 3, no. 1, pp. 15–31, 1998.
- [12] S. Hecht, "The visual discrimination of intensity and the weber-fechner law," *J. General Physiol.*, vol. 7, no. 2, pp. 235–267, Nov. 1924.
- [13] 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.
- [14] 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.
- [15] H.264/AVC JM17.2 Reference Software [Online]. Available: <http://iphome.hhi.de/suehring/tml/>
- [16] HDR image Gallery [online]. Available: <http://www.mpi-inf.mpg.de/resources/hdr/gallery.html>
- [17] HDR image Gallery [online]. Available: <http://www.anyhere.com/gward/hdrenc/pages/originals.html>
- [18] HDR Video Samples [Online]. Available: <http://www.mpi-inf.mpg.de/resources/hdr/video/>