Segmentation-based optimized tone mapping for high dynamic range image and video coding

Paul Lauga, Alper Koz, Giuseppe Valenzise, Frédéric Dufaux

Segmentation-based optimized tone mapping for high dynamic range image and video coding

Paul Lauga, Alper Koz, Giuseppe Valenzise and Frédéric Dufaux
Institut Mines-Télécom ; Télécom ParisTech ; CNRS LTCI
{firstname.lastname}@telecom-paristech.fr

***Abstract*—A core part of the state-of-the art high dynamic range (HDR) image and video compression methods is the tone mapping operation to convert the visible luminance range into the finite bit depths that can be supported by the current video codecs. These conversions are until now optimized to provide backward compatibility to the existing low dynamic range (LDR) displays. However, a direct application of these methods for the emerging HDR displays can result in a loss of details in the bright and dark regions of the HDR content. In this paper, we overcome this limitation by designing a tone mapping operation which handles the bright and dark regions separately. The proposed method first finds the optimal segmentation of the HDR image into two parts, namely dark and bright regions, and then designs the optimal tone mapping for each region in terms of the mean square error between the logarithm of the luminance values of the original and reconstructed HDR content (HDR-MSE). The results indicate the superiority of the proposed method over the state-of-the art HDR coding methods.**

I. Introduction

High dynamic range (HDR) imaging plays a key role in making the fruition of the video content more appealing and immersive, thanks to the possibility to display a huge range of luminance values in scene with high contrast. This fact, along with the availability on the market of HDR-ready displays, have recently motivated a research effort towards the compression of this kind of content [1][2]. Differently from conventional low dynamic range (LDR) imagery, where 8 bits/pixel/channel are generally sufficient to represent the limited contrast reproducible on LDR screens, HDR content can be encoded with up to 32 bits/pixel for the luminance channel, and several HDR formats have been proposed for this purpose such EXR [3] and RGBE. As a result, conventional coding schemes cannot be directly applied to HDR content, and new and efficient coding schemes have to be pursued.

HDR video coding has been mainly dealt with in two ways. In the first case, backward compatibility to 8 bits display is tried to be assured by using scalable video coding approach. An LDR base layer, typically formed of 8 bits, is obtained by applying a tone mapping operator (TMO) to the HDR content. The enhancement layer, which enables to reconstruct the full HDR video, consists in the residuals between the HDR image and the prediction obtained by inverse tone mapping the base layer [4][5][2]. According to whether the users display supports HDR, either the LDR layer or both the LDR and residual layer are transmitted. Due to the backward compatibility requirement, the output of tone mapping should be a credible and accurate LDR representation of the original HDR content.At the same time, the LDR layer should be optimized in such a way to minimize the energy of inter-layer residuals. The recent work [4] derives the optimum prediction with respect to the mean square error (MSE) of the logarithm of luminance values by designing an optimal global tone mapping. The obtained TMO is especially interesting since: i) it can be computed in closed form from the original HDR image; and ii) its optimality still holds for LDR bit depths other than 8 (i.e., also for the backward incompatible scenario). Given the relevance of this approach to our proposed method, we review it in detail in Section II.

The scalable approach may be inefficient when backward compatibility to 8-bit displays is not demanded. Therefore, a second class of techniques employ existing state-of-the-art video codecs to encode the HDR signals with a bit depth larger than the conventional 8 bits [6][7]. For instance, the high profile of H.264/AVC [8] or HEVC [9] codec support inputs up to 14 bits. Thus, the coding problem turns into representing the HDR signal with this bit depth. As in the backward compatible case, this is achieved through a TMO whose output LDR signal, however, is in a bit range larger than 8 bits. At the decoder, the HDR content is then estimated by inverse tone mapping. In [10], the TMO corresponds to the modified version of LogLuv transform [11], which maps the luminance range of each HDR frame to the range $[0, 2^{14} - 1]$. In a similar work [12], such a mapping is enhanced by smoothing the HDR frames with respect to the Human Visual System (HVS) sensitivities in different wavelet resolutions. These methods can achieve higher bitrate efficiency for the HDR stream than the back-compatible approach. However, they first guarantee the optimality of the adopted TMO in terms of the HDR quality at the decoder. Second, current TMOs tend to reproduce as close as possible the range of the luminance where most of the signal is concentrated; however, outside of this range, data would be saturated. The resulting loss of information can produce annoying distortions when the inverse tone-mapped HDR image is viewed on an HDR display.

In this paper, we address this second limitation in a scenario where backward compatibility is not required. Specifically, we extend the technique proposed in [4] for the MSE-optimal tone mapping in order to handle with the extreme regions of the HDR luminance values in the following way. We first segment the HDR image into dark and bright regions based on the image histogram. We show how to find this optimal segmentation and how to signal it to the decoder with little overhead by transmitting a binary mask. Next, for each region, we design and apply the optimal TMO following [4]. We produce an LDR image (with bit depth, in our work, fixed to 12 bits as in [10]) by applying the two designed TMOs independently on each region. Notice that, differently from the backward compatibility setup, we do not care of the visual quality in the so-obtained LDR image (which looks indeed unnatural), as far as this enables a higher fidelity reconstruction of the HDR content after inverse tone mapping.

The rest of the paper is organized as follows. In Section II we review the optimal TMO derived in [4] for minimizing the log MSE between inverse tone-mapped image and the original HDR. We use this closed-form result to show how to optimally segment an HDR image for compression in Section III, where we also describe our coding scheme. Section IV demonstrates experimentally the performance of the proposed scheme, while Section V draws conclusions.

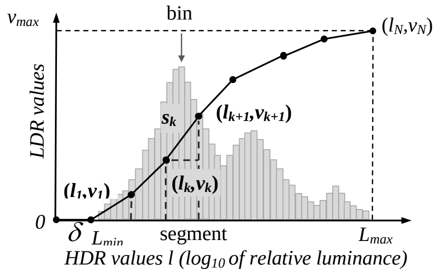

Fig. 1: piecewise parameterization of the optimal tone mapping curve of *Mai et al.*

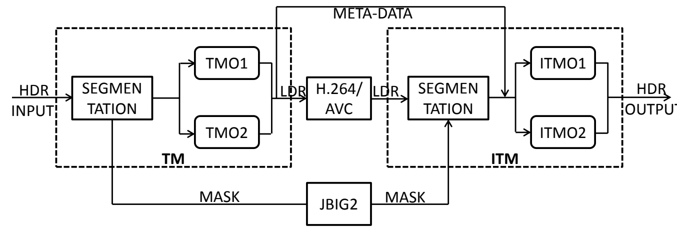

Fig. 2: general structure of our proposed method

II. Optimum tone mapping to minimize reconstructed HDR-MSE

Mai et al. [4] have recently developed an optimum tone mapping and inverse tone mapping for backward compatible HDR coding, which minimizes the MSE 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 the tone mapping curve 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 [4] is first parameterized as a piece-wise linear function with the nodes (l_k, v_k) as shown in figure 1. 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. Using such a parameterization and the statistical modeling of the H.264/AVC coding error, the optimized tone mapping problem is given as

$$\arg\min_{S_1 \dots S_N} \sum_{k=1}^{n} \frac{p_k}{s_k^2}$$
$$\text{subject to } \sum_{k=1}^{n} s_k = \frac{v_{max}}{\delta}, \quad (2)$$

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. By computing the first order Karush-Kahn-Tucker (KKT) optimality conditions of corresponding Lagrangian, their ultimate closed form solution is derived as

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

A tone mapping characterized by the slopes in (3) minimizes the MSE between the original and reconstructed HDR frames. While originally this technique has been developed for backward compatible HDR coding (i.e., the LDR image was 8-bit), the very same principles, derivation and results apply as well to any other LDR image (and particularly, with bit depth larger than 8). Figure 3a and 3e shows the tone mapping curve and the tone mapped image for *Memorial*. We build on this fact to extend the work in [4] by a segmentation-based procedure in the next section.

III. Proposed method based on segmentation

The tone mapping operation described in the previous section is designed with the aim of producing a LDR image which is supposed to be displayed on LDR screens. Therefore, the optimal TMO has to respect some constraints to guarantee a meaningful display of the generated LDR picture, e.g. the curve has to be monotonically increasing. In this paper, instead, we are not concerned with the display of LDR content, and we consider TMO and ITMO as a practical channel to transmit HDR information by means of existing LDR codecs. even if the tone-mapped LDR image is no longer visually pleasing nor realistic.

Figure 2 shows the general structure of our proposal. There are three steps in the structure, the compression part, the encoding and decoding part and the decompression part. In the compression part, first, the input RGB image is put in Yuv format in the same way as Adaptive LogLuv describes in [10]. We will work on L the logarithm of Y, $L = log_{10}Y$, as it is the case in [4] by letting u and v unchanged. From L and its histogram, we find the optimal threshold for the image to make the binary mask. As this binary mask is needed to reconstruct the HDR image, we encode and decode it in a lossless way using JBIG2 [15]. After finding the threshold and spliting the image in two regions, we apply the optimal tone mapping of *Mai et al.* [4] independently on the two differents regions. The tone mapping curves are performed using several parameters which have to be sent as meta-data to inverse these curves in the decompression part. Then we encode and decode the created LDR image in Yuv format using H.264/AVC codec in high profile [8]. The binary mask is used to split the LDR image and the meta data are used to inverse the tone mapping curves which will be applied on the two regions of the LDR image. Then we pass from Yuv format to RGB format.

Section III-A derives how to optimally split the image based on the histogram, then section III-B describes the HDR coding scheme based on this segmentation.

A. Optimal Segmentation and tone mapping of the HDR content

The pixels x_i of the image are segmented into two sets of pixels, $I_1(T) = \{x_i : 0 \leq L(x_i) \leq T\}$ and $I_2(T) = \{x_i : T < L(x_i) \leq N\}$. This section describes how to find the optimal threshold value T.

For one fixed T, we can estimate the MSE between the logarithm of the original and reconstruct HDR luminance with the help of (2) as follows:

$$MSE(T) = \sum_{k=1}^{N} \frac{p_k}{s_k^2} = \sum_{k=1}^{T} \frac{p_k}{s_k^2} + \sum_{k=T+1}^{N} \frac{p_k}{s_k^2}, \quad (4)$$

where p_k is the normalized histogram value for the k-th bin of the logarithm of HDR image, and s_k is the optimal slopes of the tone mapping described in (3). So we can compute the reconstruction error

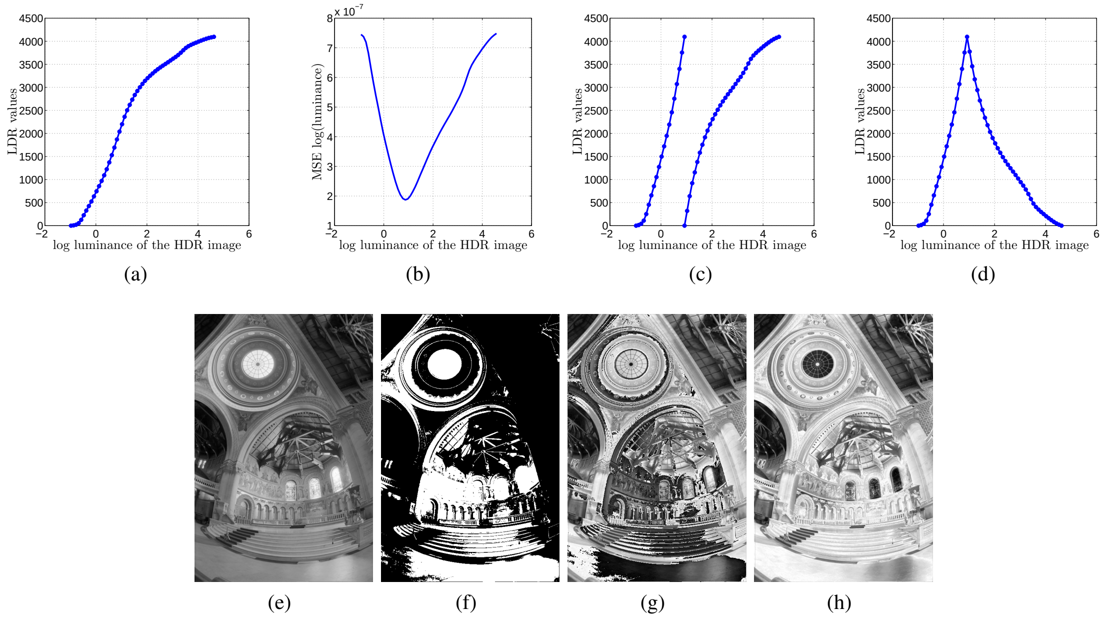

Fig. 3: (a) *Mai et al.* tone map curve for *Memorial*, (b) MSE result for different threshold for *Memorial*, (c) proposed tone map curve for *Memorial*, (d) proposed tone map curve for *Memorial* after inverse the second curve, (e) Tone mapped image of *Memorial* using *Mai et al.* tone map curve, (f) binary mask of *Memorial* using the optimal threshold, (g) tone mapped image of *Memorial* using our proposed tone map curve, (h) tone mapped image of *Memorial* using our proposed tone map curve after inverse the second tone map curve.

as follows:

$$MSE(T) = \sum_{k=1}^{T} \frac{\delta^2 . p_k^{1/3} . (\sum_{l=1}^{T} p_l^{1/3})^2}{v_{max}^2} + \sum_{k=T+1}^{N} \frac{\delta^2 . p_k^{1/3} . (\sum_{l=T+1}^{N} p_l^{1/3})^2}{v_{max}^2} \quad (5)$$

Figure 3b shows the MSE curve as a function of T for *Memorial*. The convex characteristics and the optimal threshold value minimizing the MSE are quite observable on the graph. This luminance value is used to split the HDR image into two parts, namely the bright and dark parts, which would be handled separately. At the same time, a binary mask which shows the part each pixel belongs to, is also computed to reconstruct the HDR image at the decoder.

B. The proposed system with optimal segmentation

After finding the optimal threshold, we apply the tone mapping on the two parts according to the optimal tone mapping curves obtained with (3). The tone mapping curve and the tone mapped image obtained for *Memorial* are illustrated in figure 3c and 3g, respectively. We can see that a direct implementation of the tone mapping obtained from (3) to each part can generate strong edges on previously smooth areas, which would cause an increase in the coding bitrate of H.264/AVC. To avoid such an effect, we inverse the tone mapping curve corresponding to the second part of the HDR image. The resulting curve and the image are seen on figure 3d and 3h, respectively. Compared to the previous case, the resulting image involve much smoother transition between the regions.

We can notice from the tone mapping curve image that two same value in the tone mapped image can describe two different luminance values, it is why the binary mask is needed to reconstruct the HDR image.

As it is the case in [4], our tone mapping curve will be made by a piece-wise linear function between the nodes (l_k, v_k) with all the l_k spaced by the same value δ in the logarithm domain (see section II for more details) and is performed independently on the two regions of the image.

For the backward transformation, as we get the slope as meta-data, we know exactly the tone mapping algorithm used. Let us define

$$l = \begin{cases} f_{dark}(L) \text{ if L} \leq L_{threshold} \\ f_{bright}(L) \text{ otherwise} \end{cases} \quad (6)$$

These two functions are known and easily invertible. So, our backward transformation is defined as

$$L = \begin{cases} f_{dark}^{-1}(l) \text{ if Mask = 0} \\ f_{bright}^{-1}(l) \text{ otherwise} \end{cases} \quad (7)$$

and

$$Y = 10^L. \quad (8)$$

Finally, the reconstructed RGB values are obtained from Y, u and v in the same way as Adaptive LogLuv in [10].

IV. Experimental Results

In the experiments, we use the publicly avaliable HDR images *Memorial*, *Atrium Night*, *Coby* and *Bristol Bridge* [13]. We code the image using H264/AVC high profile (JM 17.2) [14]. JBIG2 [15] is used to code the binary mask in a lossless way. The encoding format chosen is YUV 4:2:0, and bit-depth of tone mapping for luminance is selected as 12 bits as in the setup in [10], whereas it is 8 bits for chrominance components, u and v.

We compare our proposed method with the optimized tone mapping developed by *Mai et al.* [4]. The number of the histogram bins in the proposed method is held equal to the one in the compared method for a fair comparison. δ is fixed to 0.1 as in [4]. We also

Image	*Memorial*	*Atrium Night*	*Coby*	*Bristol Bridge*
Bitrate [bpp]	0.1739	0.1412	0.0883	0.0280

TABLE I: Bitrate of the binary mask for four HDR images.

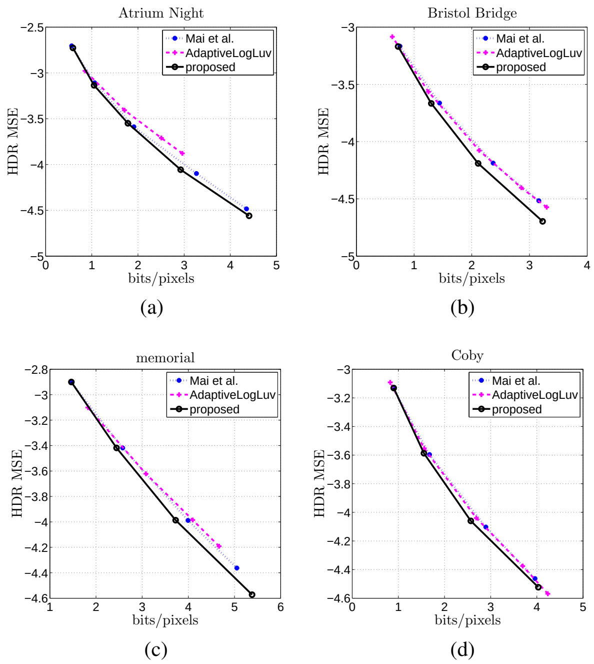

Fig. 4: Rate-distortion performance of the proposed method for 4 HDR images.

compare our proposed method to Adaptive LogLuv method [10], a pioneer method to compress HDR contents using the high profile of H.264/AVC.

The supplemental information regarding the tone mapping parameterization is about 50-100 coefficients representing the slopes of the TMO which is in fact very negligible compared to the coding bitrate. Moreover, this information is also in the same size for both the methods of *Mai et al.* and our proposed method.

As the coding bitrate for the mask is particular to our proposed method, we add the bitrate for the mask to the overall bitrate. The binary mask, which is originally one bit per pixel, is compressed using JBIG2 to further decrease the bitrate. Table I shows the cost of this mask after being coded with JBIG2 for several images. The difference in the bitrate of the mask can be explained by the amount of high frequency content. While the mask for *Memorial* has the most high frequency content, it is followed by *Atrium Night*, *Coby* and *Bristol Bridge* in descending order.

Figure 4 shows the rate-distortion (RD) performance of the proposed method, compared to the Adaptive LogLuv method [10] and *Mai et al.* method [4] for the four HDR images, in terms of coding bit rate of LDR image vs. the HDR-MSE. HDR-MSE is defined as the mean square error between the logarithm of luminance values of original and reconstructed HDR image. For each image, it can be observed that our method outperforms the two other methods. In particular, the performance of the proposed method is high for Bristol Bridge. This can be explained by the low bitrate to encode the binary mask (Table I). For *Coby*, the curves appear to be closer to each other than the other HDR images. If we look at the input HDR histogram, we can notice that *Coby*'s histogram is much more stretched and ”uniform” on its luminance range than the other histograms. Our proposal is more effective when most of the energy of the signal is concentrated on smaller luminance ranges.

Finally, we observe that our method outperforms the two methods in [10] and [4] at higher bitrates, while the RD performance are almost the same at lower bitrates. Indeed, at low bitrates the RD improvement brought by the optimized tone mapping of *Mai et al.* [4] becomes negligible as the statistical error model becomes distorted at high quantization levels. Thus our approach will not sensibly improve the RD performance. Nonetheless, our gains at higher bitrates are especially interesting for the practical scenario where HDR content has to be transmitted by a broadcaster for high-end applications.

V. Conclusion

In this paper we proposed a new tone mapping for HDR video content compression which conserves more information in the HDR image. After segmenting into low and high luminance regions the input HDR image, we optimize tone mapping independently on each part. The resulting image is compressed using H.264/AVC codec in high profile. Our simulations show that splitting the image into two parts is more efficient than applying a global optimized TMO on the whole image.

References

[1] R.Mantiuk, G.Krawcyk, K. Myszkowski, and H.-P. Seidel, “Perception-motivated high dynamic range video encoding ”, *ACM Trans. Graph. (Proc. SIGGRAPH)*, vol. 23, no. 3, pp. 730-738, 2004.

[2] A. Segall, “Scalable Coding of High Dynamic Range Video”, vol. 1, in Proc. IEEE ICIP, 2007.

[3] F. Kainz, R. Bogart, “Technical Introduction to OpenEXR”, *Industrial Light & Magic*, 18 February 2009.

[4] 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.

[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] J.-U. Garbas and H. Tahoma, “Temporally Coherent Luminance-to-Luma Mapping for High Dynamic Range Video Coding with H.264/AVC”, *in Proc. IEEE ICASSP*, pp. 829 832, 2011.

[7] Y. Zhang, E. Reinhard, and D. Bull, “Perception-based High Dynamic Range Video Compression with Optimal Bit Depth Transformation”, *in Proc. IEEE ICIP*, pp. 1321-1324, 2011.

[8] T. Wiegand, S. Gordon, and A. Luthra, “H. 264/AVC High Profile”, *DVB*, pp. 1, 2004.

[9] GJ. Sullivan, and JR. Ohm, “Recent developments in standardization of high efficiency video coding (HEVC)”, *Proc. SPIE*, vol. 7798, pp. 77980V, 2010.

[10] A. Motra and H. Thoma, “An Adaptive LogLuv Transform for High Dynamic Range Video Compression”, *in Proc. IEEE ICIP*, pp. 2061-2064, June 2010.

[11] G. W. Larson, “Logluv encoding for full-gamut, high-dynamic range images”, *J. Graph. Tools*, vol. 3, no. 1, pp. 1531, 1998.

[12] Y. Zhang, E. Reinhard, and D. Bull, “Perception-based High Dynamic Range Video Compression with Optimal Bit Depth Transformation”, *in Proc. IEEE ICIP*, pp. 1321-1324, 2011.

[13] HDR image Gallery [online].
Available: http://www.mpi-inf.mpg.de/resources/hdr/gallery.html
http://www.anyhere.com/gward/hdrenc/pages/originals.html

[14] H.264/AVC JM17.2 Reference Software [Online]. Available: http://iphome.hhi.de/suehring/tml/

[15] JBIG2 Reference Software [Online]. Available: http://blog.rubypdf.com/2012/07/25/open-source-jbig2-encoder-and-decoder/

ACKNOWLEDGEMENT

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