IMPROVED TONE MAPPING OPERATOR FOR HDR CODING OPTIMIZING THE DISTORTION/SPATIAL COMPLEXITY TRADE-OFF

Paul Lauga, Giuseppe Valenzise, Giovanni Chierchia, Frédéric Dufaux

Institut Mines-Télécom; Télécom ParisTech; CNRS LTCI

ABSTRACT

A common paradigm to code high dynamic range (HDR) image/video content is based on tone-mapping HDR pictures to low dynamic range (LDR), in order to obtain backward compatibility and use existing coding tools, and then use inverse tone mapping at the decoder to predict the original HDR signal. Clearly, the choice of a proper tone mapping is essential in order to achieve good coding performance. The state-of-the-art to design the optimal tone mapping operator (TMO) minimizes the mean-square-error distortion between the original and the predicted HDR image. In this paper, we argue that this is suboptimal in rate-distortion sense, and we propose a more effective TMO design strategy that takes into account also the spatial complexity (which is a proxy for the bitrate) of the coded LDR image. Our results show that the proposed optimization approach enables to obtain substantial coding gain with respect to the minimum-MSE TMO.

Index Terms— High dynamic range; coding; convex optimization; spatial complexity.

1. INTRODUCTION

High dynamic range (HDR) imaging enables to capture and display the huge range of luminance values of a real-world scene, and is therefore extremely important in making video user experience more immersive and realistic. In order to represent scene-referred luminance values, HDR pictures use more than the 8 bits/pixel/channel employed for display-referred low dynamic range (LDR) pictures [1]. As a result, HDR content entails much more information to be stored and transmitted, which has recently motivated research towards new coding schemes for HDR pictures and video [2] [3] [4].

HDR content can be coded using a scalable approach. First, an LDR base layer is obtained by applying a tone mapping operator (TMO) to the HDR picture. Then, an enhancement layer is computed as the residual between the HDR image and the prediction obtained by inverse tone mapping the base layer [5] [6] [3]. This approach enables to ensure backward compatibility with legacy LDR displays, and to provide at the same time full HDR experience for users equipped with HDR displays. Mai et al. [5] derived a closed-form expression for the TMO that minimizes the energy of the inter-layer prediction residuals, i.e., the mean square error (MSE) between the original HDR image and the one obtained by inverse tone mapping the base layer. This is shown to be beneficial in terms of coding efficiency, since a lower residual energy enables to have a better prediction quality and to reduce the bitrate of the enhancement layer. The results in [5] demonstrate that the minimum-MSE TMO achieves higher rate-distortion performance[1] than other commonly used TMOs designed to enhance LDR visual quality.

However, minimizing MSE distortion is not necessarily optimal in terms of coding performance. For instance, LDR images obtained with the method in [5] preserve more details of the original HDR, but they could be harder to code due to an increased spatial complexity [7]. A better approach would be instead to find a TMO that is optimal in *rate-distortion* (RD) sense. In this paper, we reformulate the problem of finding an optimal TMO for HDR compression by taking explicitly into account the spatial complexity of the resulting LDR base layer. To this end, we modify the cost function of [5] by adding an extra term which describes the amount of spatial information in the LDR image, in terms of the total variation measure [8]. We show that the so-obtained problem is convex and can be solved efficiently through a proximal optimization method [9]. We plot rate-distortion curves for HDR images and demonstrate that our approach brings higher RD gains with respect to the minimum-MSE TMO of [5].

The rest of the paper is organized as follows. In Section 2 we review the optimal TMO derived in [5] for minimizing the log MSE between inverse tone-mapped image and the original HDR. In Section 3 we describe the proposed optimization scheme which takes into consideration the spatial complexity of the tone-mapped image in the base layer. Section 4 demonstrates experimentally the performance of the proposed method, while Section 5 draws conclusions.

2. OPTIMUM TONE MAPPING TO MINIMIZE RECONSTRUCTED HDR-MSE

Mai et al. [5] have proposed an efficient HDR compression method optimizing the MSE between the logarithm values of

[1] Notice that in the results in [5], the distortion is measured between the original and the reconstructed HDR image, while the bitrate is that of transmitting the base-layer only. Coding of the enhancement layer is not directly taken into account.

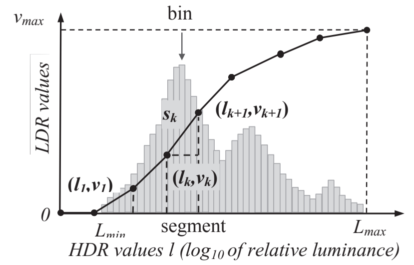


Fig. 1: Piecewise parameterization of the optimal tone mapping curve as in [5]

the luminance of original HDR content and the reconstructed version. The distortion in the process of tone mapping, encoding, decoding and inverse tone mapping have been estimated by a statistical distortion model and they use a closed-form solution to tone map the HDR image based on the luminance histogram of the HDR image.

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, the tone mapping curve 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

$$s_k = \frac{v_{k+1} - v_k}{\delta}, \tag{1}$$

which forms a vector of tone-mapping parameters. Using such a parameterization, the optimized tone mapping problem is given as

$$\underset{s \in]0,+\infty[^N}{\text{minimize}} \quad \sum_{k=1}^{N} p_k \, s_k^{-2} \quad \text{s.t.} \quad \sum_{k=1}^{N} s_k = \frac{v_{max}}{\delta}, \tag{2}$$

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

$$s_k = \frac{v_{max} \, p_k^{1/3}}{\delta \sum_{k=1}^{N} p_k^{1/3}}. \tag{3}$$

A tone mapping characterized by the slopes s_k reads

$$v = \begin{cases} (l - l_1) \cdot s_1 + v_1, & \text{if } l \in [l_1, l_2], \\ \dots \\ (l - l_N) \cdot s_N + v_N, & \text{if } l \in [l_N, l_{N+1}], \end{cases} \tag{4}$$

where l_k and v_k are represented on figure 1. The tone mapping operator represented by equation (4) used with the s_k defined in equation (3) minimizes the MSE between the original and reconstructed HDR frames.

3. PROPOSED METHOD

The method presented in Section 2 allows one to minimize the distortion by means of a tone-mapping compression. However, as we can see in Equation (3), this method neglects the obvious spatial dependencies than any real-world image exhibits. From a rate-distortion standpoint, it is certainly preferable to take into account both the distortion and a suitable spatial regularization. In this work, we modify problem (2) by adding a regularity term C, leading to

$$\underset{s \in]0,+\infty[^N}{\text{minimize}} \quad \sum_{k=1}^{N} p_k \, s_k^{-2} + \lambda \, C(Ts) \quad \text{s.t.} \quad \sum_{k=1}^{N} s_k = \frac{v_{max}}{\delta}, \tag{5}$$

where λ is a positive constant, T denotes the (linear) operator that performs the tone-mapping in (4),[2] and C denotes a real-valued convex function that models the spatial regularization.

In [10] the authors also proposed to minimize the bit rate of the LDR layer by modeling the smoothness of the tone-mapped image. However, their model approximates spatial regularity based on conditional pixel value probabilities, while in our case we optimize explicitly the spatial dependencies at the pixel level.

3.1. TV regularization

The quality of the results obtained through a variational approach strongly depends on the ability of the function C to model the regularity present in images. Since natural images usually exhibit a smooth spatial behaviour, except around some locations (e.g. object edges) where discontinuities arise, popular regularization models tend to penalize the image gradient. In this context, Total Variation (TV) [8] has emerged as a simple, yet successful, convex optimization tool. This regularization term can be expressed as

$$C(v) = \|\nabla v\|_{1,2} = \sum_{i \in \Omega} \|(\nabla v)_i\|_2, \tag{6}$$

where Ω is the rectangular lattice over which the image v is defined, and $(\nabla v)_i$ is the 2-element vector denoting the gradient of v at site i.

Note that an alternative regularization approach consists in replacing the gradient operator with a frame representation which yields a more suitable sparse representation of the image [11]. The connections between these two different approaches have been studied in [12]. In this work, we focus our attention on TV regularization, although our proposed algorithm is quite general and it can also be adapted to frame-based approaches.

[2] Note that $v = Ts$ is a linear operation, as the HDR image l is regarded as a constant in the minimization problem.

3.2. Convex optimization

In this paper, we propose to find the vector s that achieves the optimal balance between the distortion and TV regularization. Hence, we need to solve the following problem

$$\underset{s\in]0,+\infty[^N}{\text{minimize}} \quad \sum_{k=1}^{N} p_k\, s_k^{-2} + \lambda\, \|\nabla T s\|_{1,2} \quad \text{s.t.} \quad \sum_{k=1}^{N} s_k = \frac{v_{max}}{\delta}. \tag{7}$$

To gain some insight into the solution of Problem (7), let us define the (non-smooth) convex functions

$$f(s) = \sum_{k=1}^{N} f_k(s_k), \text{ with } f_k(s_k) = \begin{cases} +\infty & \text{if } s_k \le 0, \\ p_k\, s_k^{-2} & \text{if } s_k > 0, \end{cases} \tag{8}$$

$$g(y) = \lambda \|y\|_{1,2}, \tag{9}$$

and the closed convex set

$$C = \{s \in \mathbb{R}^N \mid \sum_{k=1}^{N} s_k = \frac{v_{max}}{\delta}\}. \tag{10}$$

Therefore, Problem (7) can be more conveniently rewritten as

$$\underset{s\in C}{\text{minimize}}\ f(s) + g(\nabla T s), \tag{11}$$

which consists in minimizing the sum of two convex functions, one of which composed by a linear operator, over a closed convex set.

The solution of Problem (11) requires an efficient algorithm for dealing with non-smooth functions. Among the many approaches proposed to solve convex optimization problems, we resort here to proximal algorithms [9], since they provide a unifying framework that allows one to address both non-smooth functions and hard constraints. Within the large panel of existing proximal algorithms [13–16], we consider the primal-dual M+LFBF algorithm recently proposed in [15], which is able to address general convex optimization problems involving non-smooth functions and linear operators without requiring any matrix inversion.

4. EXPERIMENTAL RESULTS

We test our method on several images taken from the HDR Photographic Survey dataset[3]. The images have a maximum resolution of 4,288×2,848 and span a wide range of luminance conditions and contrast ratios. We cut a square portion of 1024×1024 pixels in each image to apply our algorithm and test coding performance. A subset of these images used in our tests is shown in Figure 2. We encode images using JPEG and JPEG2000 on image luminance tone mapped to 8 bits values using the piecewise linear tone mapping obtained by solving (11). We compare this method with the tone mapping in [5], which corresponds to solving problem (2) or to setting $\lambda = 0$ in our problem. In both cases, we consider luminance only as in the original setup [5], based on the same consideration that one could easily extend tone mapping to red, green and blue color channel independently while preserving color appearance [17]. The cost of transmitting supplemental information to reconstruct the tone mapping curve (i.e., the vector of s_k), which is present both in our proposal and in the baseline method [5], depends on the choice of the histogram step δ (fixed here to 0.1) and is in general negligible with respect to the total bitrate.

We use the standard JPEG implementation available in Matlab through the `imwrite` command, while for JPEG2000 we employ the Kakadu implementation available at [18]. The quality rate used for JPEG are 20 to 90 with a step of 10. As for JPEG2000, we use 5-level, 9/7 Daubechies wavelet decomposition, and we fix the rate from 0.1 to 1 bits per pixel with a step of 0.1. In order to evaluate the distortion between the original and the reconstructed HDR image after coding, we use the HDR-MSE metric as also done in [5]. The HDR-MSE is defined as the logarithm of the MSE between the logarithm of the luminance chanel of the reconstructed HDR image and the reference one. A popular perceptually-based image difference predictor is the HDR-VDP2 [19]. We compared the reconstructed HDR images using our TMO with those produced using minimum-MSE, with a publicly available implementation of HDR-VDP2 [4], and found that the results are inconsistent with the compression quality used for the LDR layer, e.g., higher LDR rates could randomly produce worse HDR-VDP2 results. Therefore, we do not include performance evaluation relative to HDR-VDP2 in this paper.

Figure 3 shows rate-distortion results for minimum-MSE TMO [5] and our proposal for three HDR pictures. The optimal λ in (11) has been found empirically. The experiments show that a λ between 0.001 and 0.004 tends to yield important coding gains with respect to minimum-MSE TMO. Higher values of λ may produce washed-out LDR images, which are certainly smoother but poorer in details. On the other hand, very low values of λ do not improve significantly the performance of the minimum-MSE method. On the results we present, we choose a value of 0.004 for "AmikeusBeaverDamPM1" and "Flamingo" and a value of 0.002 for "GoldenGate(2)".

In order to gain a better insight of how the proposed method improves RD performance, we show as an example in Figure 4 two tone mapping curves obtained from the image "AmikeusBeaverDamPM1": the minimum-MSE TMO, and our proposal based on TV regularization ($\lambda = 0.004$). Minimum-MSE TMO allocates a larger interval of LDR values to mid luminance range, and compresses instead the extreme regions of the original HDR histogram. This enables to produce well contrasted LDR pictures which preserve most part of HDR high-frequency content and details. Conversely,

[3] Available at http://www.cis.rit.edu/fairchild/HDR.html

[4] Available at: http://sourceforge.net/apps/mediawiki/hdrvdp/

(a) AmikeusBeaverDamPM1 (b) Flamingo (c) GoldenGate(2)

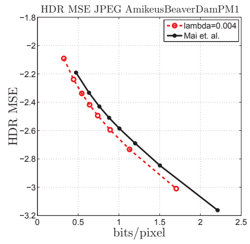
Fig. 2: Some HDR images from the HDR Photographic Survey dataset.

TV regularization generates curves which are closer to linear tone mapping. As a result, the LDR image has lower global contrast with respect to [5]. However, local image structures such as edges are well preserved thanks to the definition of the TV measure (5). We give an example in Figure 5, which shows the tone-mapped LDR image and the reconstructed HDR content for a detail of the image "GoldenDate(2)". The two decoded HDR images details have similar HDR-MSE values of -3.52 and -3.64 for [5] and the proposed method, respectively. However, it is apparent that the LDR image obtained with our TMO is smoother than the minimum-MSE one, thus it has lower spatial complexity and is easier to code, which explains the RD gain in Figure 3. For our matlab optimization implemention (wich is not optimized for speed), we put a large number of iteration to assure the convergence. however, the process stops hitself before the end of the iterations when the difference between two following iteration results is to small. The number of needed iteration depends on the image and on λ. In general, 5000 iterations are enough to assure the convergence and the process took on average 5 minutes to find the s_k for an 1024×1024 image.

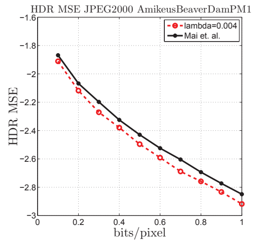
5. CONCLUSION

In the context of backward compatible HDR image compression, the HDR content is mapped to LDR to be compressed with available coding tools and be used as predictor at the decoder. In this paper we have presented a tone mapping operator that minimizes the prediction error *and* the spatial complexity of the LDR image, modeled through the total variation measure. The resulting problem is convex and can be solved efficiently. While we do not claim that this is optimal from the rate-distortion point of view, we show that it achieves significant RD gains with respect to the state of the art technique that minimizes mean-square-error distortion between original and reconstructed HDR [5].

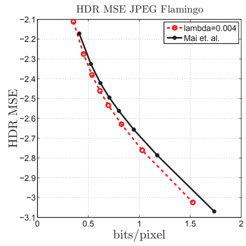
We are currently working to extend this model to video, by modeling temporal complexity as well. In addition, the simple quadratic MSE term is know to be poorly related to perception [20], which motivates for using more advanced,



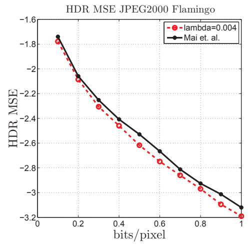
(a)



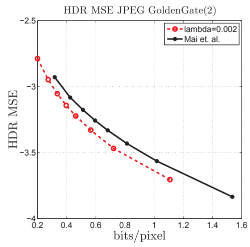
(b)



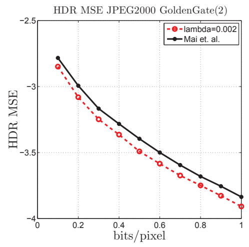
(c)



(d)



(e)



(f)

Fig. 3: Rate-distortion performance of the proposed method compared to the minimum-MSE TMO [5] with both JPEG and JPEG2000, for the images in Figure 2.

yet feasible to optimize, fidelity measures for HDR content.

REFERENCES

[1] F. kainz and R. Bogart, "Technical introduction to OpenEXR," *Industrial Light & Magic*, February 2009.

[2] R. Mantiuk, G. Krawcyk, K. Myszkowski, and H.-P. Seidel, "Perception-motivated high dynamic range video encoding," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 730–738, 2004.

[3] A. Segall, "Scalable coding of high dynamic range video," in *Proc. IEEE Int. Conf. on Image Processing*, vol. 1, pp. I–1–I–4. 2007.

[4] T. Richter, "On the standardization of the JPEG XT image compression," in *Proc. of the Picture Coding Symposium*, San Jose, CA, Dec 2013, pp. 37–40.

[5] Z. Mai, H. Mansour, R. Mantiuk, P. Nasiopoulos, and W. Ward, R. Heidrich, "Optimizing a tone curve for

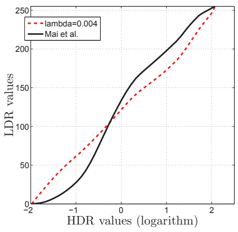


Fig. 4: Tone mapping curve for AmikeusBeaverDamPM1 using *Mai et al.* method (in black) and our propose with λ=0.004 (in red).

backward-compatible high dynamic range image and video compression," *IEEE Trans. on Image Processing*, vol. 20, no. 6, pp. 1558–1571, June 2011.

[6] S. Liu, W.-S. Kim, and A. Vetro, "Bit-depth scalable coding for high dynamic range video," *SPIE Visual Communications and Image Processing*, vol. 6822, 2008.

[7] H. Yu and S. Winkler, "Image complexity and spatial information," in *Proc. 5th Int. Workshop on Quality of Multimedia Experience*, July 2013, pp. 12–17.

[8] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992.

[9] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., pp. 185–212. Springer-Verlag, New York, 2011.

[10] Z. Mai, H. Mansour, P. Nasiopoulos, and R. Ward, "Visually-favorable tone-mapping with high compression performance," *17th IEEE International Conference on Image Processing (ICIP)*, pp. 1285–1288, Sept. 2010.

[11] S. Mallat, *A wavelet tour of signal processing*, Academic Press, San Diego, USA, 1997.

[12] J.-F. Cai, B. Dong, S. Osher, and Z. Shen, "Image restoration: Total variation, wavelet frames and beyond," *Journal of the American Mathematical Society*, vol. 25, no. 4, pp. 1033–1089, 2012.

[13] J.-C. Pesquet and N. Pustelnik, "A parallel inertial proximal optimization method," *Pacific Journal of Optimization*, vol. 8, no. 2, pp. 273–305, Apr. 2012.

[14] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[15] P. L. Combettes and J.-C. Pesquet, "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, lipschitzian, and parallel-sum type monotone operators," *Set-Valued and Variational Analysis*, vol. 20, no. 2, pp. 307–330, Jun. 2012.

(a) *Mai et al.* tone mapping

(b) Our proposal using λ=0.002

(c) HDR reconstruct with *Mai et al.* using 0.8 bits per pixel

(d) HDR reconstruct with our proposal using λ=0.002 and 0.7 bits per pixel

Fig. 5: the first row shows detail on GoldenGate(2) after tone map with *Mai et al.* and our propose with λ=0.002. The second row presents tone mapping of the detail on Golden Gate(2) after reconstruction of the HDR image.

[16] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu, "Epigraphical splitting for solving constrained convex formulations of inverse problems with proximal tools," 2012, Preprint: http://arxiv.org/pdf/1210.5844.

[17] R. Mantiuk, A. Tomaszewska, and W. Heidrich, "Color correction for tone mapping," in *Computer Graphics Forum*. Wiley Online Library, 2009, vol. 28, pp. 193–202.

[18] "Kakadu software," available at: `http://www.kakadusoftware.com/`.

[19] R. Mantiuk, K. Kim, A.G. Rempel, and W. Heidrich, "HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions," in *ACM Trans. on Graphics*. ACM, 2011, vol. 30, p. 40.

[20] Z. Wang and A.C. Bovik, "Mean squared error: love it or leave it? a new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.