RATE DISTORTION OPTIMIZED TONE CURVE FOR HIGH DYNAMIC RANGE COMPRESSION

Mikaël Le Pendu [1,2], *Christine Guillemot* [1] *and Dominique Thoreau* [2]

[1] *INRIA*
Campus de Beaulieu, 35042 Rennes Cedex France
firstname.lastname@irisa.fr

[2] *Technicolor*
Av. des Champs Blancs, 35576 Cesson-Sévigné France
firstname.lastname@technicolor.com

ABSTRACT

In this paper, we define a reversible tone mapping-operator (TMO) for efficient compression of High Dynamic Range (HDR) images using a Low Dynamic Range (LDR) encoder. In our compression scheme, the HDR image is tone mapped and encoded. The inverse tone curve is also encoded, so that the decoder can reconstruct the HDR image from the LDR version. Based on a statistical model of the encoder error and assumptions on the rate of the encoded LDR image, we find a closed form solution for the optimal tone curve with respect to the rate and the mean square error (MSE) of the reconstructed HDR image. It is shown that the proposed method gives superior compression performance compared to existing tone mapping operators.

***Index Terms*—** High Dynamic Range (HDR), Tone Mapping, Companding, Gaussian Mixture Model (GMM), HEVC

1. INTRODUCTION

With the development of high dynamic range (HDR) imaging, the compression of high bit depth images is a subject of increasing interest. Standard images have a bit depth of 8 bits per color channel which is not sufficient to represent with precision the range of luminance that the human eye can perceive. Existing codecs are designed for relatively low bit depth. For example, in the recent compression standard HEVC [1], 8 to 10 bits input data is supported. Although extended versions can support higher bit depths, their use can be restricted because of the increased implementation and computational cost.

An alternative solution for HDR Compression is to apply a tone mapping operator (TMO) on the image to reduce the bit-depth. A low bit-depth encoder can then be used to compress the Low Dynamic Range (LDR) version of the image. In this approach, side information needs to be encoded in order to perform the inverse operation in the decoder. In [2], [3] and [4], local TMOs which are not invertible, are used. In this case, a second layer (e.g. residual or ratio image) is necessary to reconstruct the HDR image from the decoded LDR version.

Other attempts have been made at reducing the bit-depth using a global TMO which consists in applying a non linear curve (called compressor), followed by uniform quantization. Applying the inverse quantization and the inverse curve (called expander) to the encoded and decoded LDR image expands the data to its original dynamic. This approach, often referred to as companding is equivalent to non-uniform quantization. As an example, the method in [5] iteratively optimizes the parameters of the photographic TMO from [6] in order to minimize the HDR reconstruction error. In [7], an approximation of the data distribution based on Gaussian mixture models (GMM) is used to build compressor and expander curves that approaches the results obtained by the Lloyd-Max algorithm [8]. The latter algorithm is used, for example in [9], in the context of HDR compression. It aims at finding the optimal quantizer in terms of distortion, but does not consider further encoding of the quantized image. In [10], *Mai et al.* define a segment based curve that minimizes the data loss caused by both the tone mapping and the encoder error. However, all these methods only focus on the distortion without taking into account the rate of the encoded LDR image.

In our method, both rate and distortion are optimized. The image is first tone mapped with a global invertible TMO and then encoded with a complex encoder such as HEVC. Based on a statistical model of the complete HDR compression scheme and assumptions on the rate of the encoded LDR image, a closed form solution is found for the optimal tone curve in the sense of rate distortion performance.

The rest of the paper is organized as follows. In section 2, the statistical model of the complete compression scheme is presented and the problem to solve is posed. A closed form solution is given in section 3. Then, the implementation of our optimized tone mapping is described in section 4.

2. STATISTICAL MODEL

The problem consists in minimizing an objective function of the form $D + \lambda_0.R$, where D is the total distortion after reconstruction of the HDR image, R is the rate of the encoded LDR image and λ_0 is a Lagrangian multiplier that is adjusted to obtain the optimal rate distortion performance.

In order to find a closed form solution to this problem, we define a statistical model of our compression scheme. It

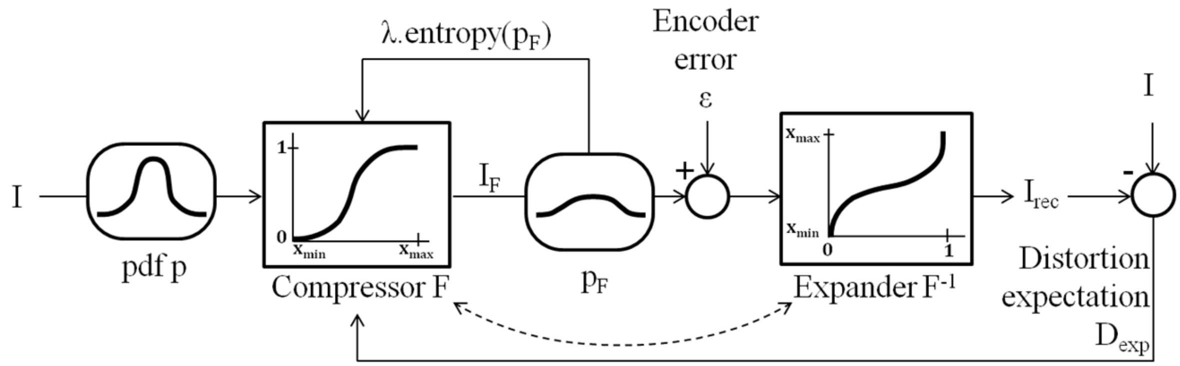


Fig. 1. Statistical model of the HDR compression scheme.

is illustrated in figure 1. In this model, we consider that the pixels have real values (not necessarily integers). The input image I has a probability density function (p.d.f.) p and its minimum and maximum pixel values are respectively x_{min} and x_{max}.

First, a function F that we call compressor function is applied to the pixel values. F is defined on the interval $[x_{min}, x_{max}]$ such that $F(x_{min}) = 0$ and $F(x_{max}) = 1$. We assume that F is a continuous and strictly monotonous function. These properties ensure that F has an inverse function F^{-1} (called expander). After applying the function F, no data is lost since F is mathematically invertible. We call I_F the obtained image and p_F its probability density function. Note that in a real implementation, the operation of tone mapping with a curve is equivalent to applying a compressor function such as F followed by uniform quantization.

Then a random variable ($\mathcal{E}$) is added to each pixel value. It models both the quantization error from the tone mapping and the encoder error. Here we suppose that the distribution of these random variables does not depend on the position or the value of the pixel. It has zero mean, and a variance σ^2.

Finally the expander F^{-1} is applied to reconstruct the image I_{rec}. Based on this model, the total mean square error (MSE) of the reconstructed image can be estimated by its statistical expectation value D_{exp}. Given an image I and an encoder with fixed quality settings (e.g. fixed QP), we make the assumption that the rate R of the encoded image is proportional to the entropy of I_F. Thus, minimizing $D_{exp} + \lambda_0.R$ is equivalent to minimizing $D_{exp} + \lambda.entropy(I_F)$, where λ is another Lagrangian multiplier.

3. CLOSED FORM SOLUTION

Considering the mean square error as distortion metric, for an input value x in the original image, the distortion $D(x)$ is given by :

$$D(x) = (x - F^{-1}(F(x) + \mathcal{E}))^2$$

For small values of $\mathcal{E}$, the following approximation can be done :

$$D(x) \approx (x - F^{-1}(F(x)) + \mathcal{E} \cdot F^{-1\prime}(F(x)))^2$$
$$D(x) \approx (\mathcal{E} \cdot F^{-1\prime}(F(x)))^2$$
$$D(x) \approx \frac{\mathcal{E}^2}{F'(x)^2}$$

Thus, the expected distortion for the value x is :

$$E(D(x)) = \frac{var(\mathcal{E})}{F'(x)^2} = \frac{\sigma^2}{F'(x)^2}$$

And the total mean distortion is :

$$D_{exp} = \int_{x_{min}}^{x_{max}} p(x) \cdot E(D(x)) \, \mathrm{d}x = \int_{x_{min}}^{x_{max}} p(x) \cdot \frac{\sigma^2}{F'(x)^2} \, \mathrm{d}x \quad (1)$$

The entropy H_F of the new probability density function p_F after applying the compressor function F is :

$$H_F = -\int_0^1 p_F(y) \cdot log_2(p_F(y)) \, \mathrm{d}y$$
$$H_F = -\int_0^1 p(F^{-1}(y)) \cdot log_2(p(F^{-1}(y))) \, \mathrm{d}y$$

Applying the substitution $y = F(x)$, we obtain :

$$H_F = -\int_{x_{min}}^{x_{max}} p(x) \cdot log_2(p(x)) \cdot F'(x) \, \mathrm{d}x \quad (2)$$

The expression of the total cost to minimize is then :

$$Cost = \int_{x_{min}}^{x_{max}} \frac{\sigma^2 \cdot p(x)}{F'(x)^2} - \lambda \cdot p(x) \cdot log_2(p(x)) \cdot F'(x) \, \mathrm{d}x \quad (3)$$

Applying the Euler Lagrange equation gives the following condition for the optimal function F^* with respect to the cost:

$$\frac{-2 \cdot \sigma^2 \cdot p(x)}{F'(x)^3} - \lambda \cdot p(x) \cdot log_2(p(x)) = c,$$

$$F^{*\prime}(x) = \sqrt[3]{\frac{-2 \cdot \sigma^2 \cdot p(x)}{c + \lambda \cdot p(x) \cdot log_2(p(x))}}$$

Thus,

$$F^*(x) = \int_{x_{min}}^{x} \sqrt[3]{\frac{-2 \cdot \sigma^2 \cdot p(t)}{c + \lambda \cdot p(t) \cdot log_2(p(t))}}\, \mathrm{d}t \qquad (4)$$

where c is a constant that must be adjusted so that $F(x_{max}) = 1$.

However, we do not have an analytical solution to determine the value of c given λ and σ. Moreover, in practice, the value of σ is not known. A model of the actual encoder used would be necessary to find σ knowing the encoding parameters (e.g. QP in HEVC, bitdepth of the LDR image).

To solve this problem, we define the function :

$$S(x, \lambda_i) = \int_{x_{min}}^{x} \sqrt[3]{\frac{-2 \cdot \sigma_0^2 \cdot p(t)}{c_0 + \lambda_i \cdot p(t) \cdot log_2(p(t))}}\, \mathrm{d}t \qquad (5)$$

where c_0 and σ_0 are fixed arbitrarily ($c_0 < 0$ and $\sigma_0 > 0$).

It can be shown that for any positive value λ and σ ($\sigma \neq 0$), there exists a value $\lambda_i \in \mathbb{R}$ such that :

$$\forall x \in [x_{min}, x_{max}],\ F^*(x) = \frac{S(x, \lambda_i)}{S(x_{max}, \lambda_i)} \qquad (6)$$

The consequence is that only one parameter λ_i is necessary to adjust the shape of the tone curve. In practice, we have to compute $S(x, \lambda_i)$ by numerical integration and divide the result by $S(x_{max}, \lambda_i)$.

4. IMPLEMENTATION

In this article, we consider HDR images which are originally in an integer format. In the case of floating point images representing physical luminance values, a log or gamma encoding must be applied first to obtain a perceptually uniform integer image. The method is described for grayscale images. For color images, the same procedure can be applied independently to the chroma components.

4.1. Model of the probability density function

As shown in equation (4), the optimal tone mapping curve depends on the probability density function p of the pixel values in the original image. It can be easily estimated by computing the histogram of the image. However, since the decoder must be able to compute the inverse tone mapping curve, a better solution is to parameterize the histogram. This way, only a few parameters must be encoded. For that purpose, a Gaussian mixture model (GMM) of the p.d.f. is found using the Expectation Maximization algorithm (EM) from [11]. A GMM is a weighted sum of several Gaussians. The model parameters are the variance v_j, the mean value μ_j and the weight α_j of each Gaussian j in the mixture model. The p.d.f. is thus given by :

$$p(x) = \sum_{j=1}^{m} \frac{\alpha_j}{\sqrt{2\pi \cdot v_j}} \cdot exp\left(\frac{(x - \mu_j)^2}{2v_j}\right) \qquad (7)$$

where m is the number of Gaussians used in the model.

4.2. Computation of a Lookup Table

As it is explained in section 3, the compressor function F^* can be computed from equations (5) and (6) using arbitrary values for σ_0 and c_0. In our implementation we take $c_0 = -1$ and $\sigma_0 = 1$. Given a value of λ_i, we first compute the derivative of the function S with respect to x:

$$\frac{\partial S}{\partial x}(x, \lambda_i) = \sqrt[3]{\frac{-2 \cdot p(x)}{-1 + \lambda_i \cdot p(x) \cdot log_2(p(x))}} \qquad (8)$$

This function is tabulated at every integer value x from x_{min} to x_{max}. Then, $S(x, \lambda_i)$ can be approximated by numerical integration. The final tone curve is then computed as a Lookup Table (LUT) from the following equation :

$$LUT(x) = \left[(2^n - 1) \cdot \frac{S(x, \lambda_i)}{S(x_{max}, \lambda_i)}\right] \qquad (9)$$

where n is the bitdepth of the LDR image and the brackets represent the rounding operation. Note that the factor $2^n - 1$ does not affect the shape of the tone curve.

The tone mapping operation only consists in applying this LUT to every pixel of the original HDR image. The image obtained is then compressed with an LDR encoder. The parameters used for the construction of the tone mapping curve (i.e. GMM parameters, xmin, xmax) must also be encoded without loss.

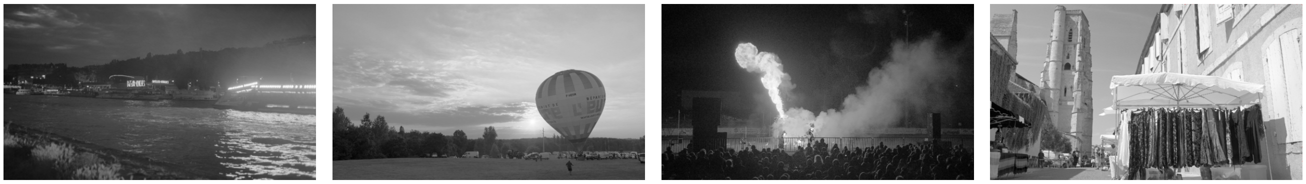
On the decoder side, the first step is to decode the LDR image and the model parameters. Knowing the parameters, the operations described in equations (7) and (8) can be performed. From the tabulated function $\frac{\partial S}{\partial x}$, the inverse of the integral is computed numerically to obtain an inverse tone mapping LUT. Finally, the inverse LUT is applied to the decoded LDR image to reconstruct the HDR image.

4.3. Determination of the Lagrangian Multiplier

In the previous section, all the computations were based on the parameter λ_i. This value must be optimized with respect to the rate distortion performance of the complete HDR compression scheme. For a given encoder (e.g. HEVC [1], MPEG-4 H264/AVC [12], JPEG2000 [13]) and LDR bitdepth, a law giving the optimal value λ_i^* as a function of the encoding quality parameter (e.g. QP parameter in HEVC) must be determined. In our implementation, input images with a bitdepth of 16 bits were tone mapped to n bits and encoded using HEVC at different QP. In that case, the following law was found to give nearly optimal rate distortion results for a large set of images:

$$\lambda_i^* = 100 \cdot 2^{0.37 \cdot (QP + 6 \cdot (n-8))} \qquad (10)$$

This model was found by encoding several images over a large range of QP and λ_i values. For a given image, at each QP, the encoding was performed several times by varying the value of λ_i. Given a QP value, the Rate Distortion (RD) point obtained with the optimal λ_i is on the convex hull of the set



(a) Péniches (b) Mongolfière (c) Cracheur (d) Marché

Fig. 2. Images from the NEVEx project used for our experiment

of all the RD points achievable. Thus, the optimal Lagrangian multiplier λ_i^* can be determined for each QP. The operation was performed for several images and an exponential law was fitted to the experimental data to derive the formula in equation (10).

5. EXPERIMENTAL RESULTS

For the experiments, only a luma channel was encoded. The method can be generalized to color images by computing an optimized tone curve for each channel separately. The HEVC standard was used to encode the LDR image.

Since the input images are originally in the OpenEXR half float RGB format [14], a logarithmic encoding was performed to convert the RGB floating point data to 16 bit integers. We consider this encoding as perceptually uniform. The obtained log RGB values were then converted to YUV using the BT.709 primaries. Only the luma channel Y was compressed with our method by tone mapping to 10 bits and encoding with HEVC using the YUV 4:0:0 chroma format and 10 bit input bitdepth. In all the tests, the number of gaussian functions in the GMM was fixed to 6. We did not observe significant changes in the probability density function by increasing this value.

Our distortion measure is the PSNR computed on the reconstructed 16 bit integer data obtained after applying the inverse curve to the 10 bit decoded image. Thus, the peak signal value used in the PSNR formula is 65535. Note that this 16 bit data is proportionnal to the logarithm of luminance. Therefore, we obtain a better visual indicator than a PSNR applied directly to luminance values.

The results are shown for the input images *Péniches*, *Mongolfière*, *Cracheur* and *Marché* in figure 2. These images are taken from sequences produced by Technicolor within the framework of the french collaborative project NEVEx.

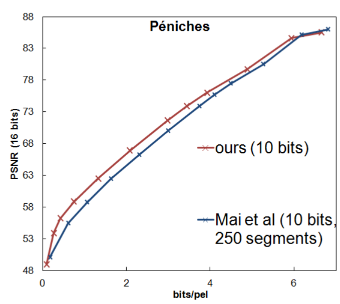
Our method was compared to the distortion optimized tone mapping developed by *Mai et al.* [10]. In our implementation their curve is applied directly on the 16 bit integer data. Therefore, their logarithm function is not applied again and the histogram bin size was chosen to have 250 segments.

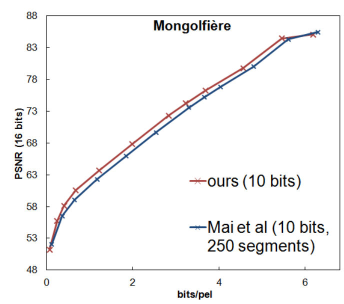
The resulting rate distortion curves are shown in figure 3. The curves were generated by varying the QP value from -12 to 32 in the HEVC encoder configuration.

Note that when we take $\lambda_i = 0$, our method is very close to that of *Mai et al.* The main difference relies on the estimation of the probability density function. This is in accordance with the fact that only distortion is taken into account in [10]. This can be observed in the experimental results. For low QP values (and thus low λ_i), the results of both methods are close. The bit rate in this case is very high (from 6 to 8 bits per pixel for the tested images encoded at 10 bits). For lower bit rates, our method gives better results thanks to the rate distortion optimization technique.

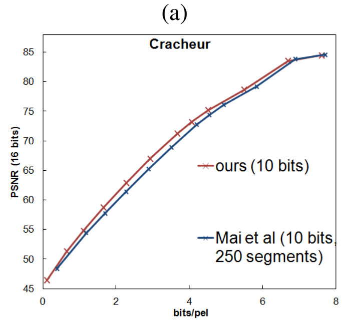
Table 1 shows the bit rate savings obtained in comparison to *Mai et al.*'s method using Bjontegaard metric [15] for two different QP ranges. Low bitrates are computed for QP from 12 to 32 and high bit rates are computed for QP from -4 to 4.

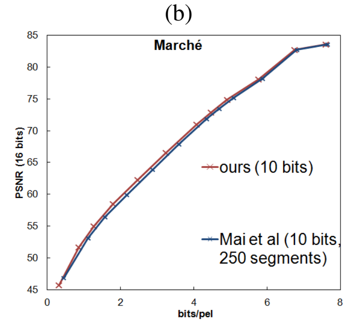
The image *Marché* gives the lowest gains. This can be explained by the fact that the histogram of this image is more "uniform" than those of the other images. In this case, the tone curves obtained by both methods are similar.





(a) (b)





(c) (d)

Fig. 3. Rate distortion results for four HDR images

Image	Low bitrates	High bitrates
Péniches	36.7%	5.4%
Mongolfière	21.2%	5.2%
Cracheur	13.9%	4.0%
Marché	6.9%	1.9%

Table 1. Bitrate savings compared to *Mai et al.* [10]. QP from 12 to 32 are used for low bitrates and QP -4 to 4 are used for high bitrates.

In comparison to [10], however, our method has an increased complexity. On average, for the tested images, the execution time of our tone mapping represents 56% of the total encoding time (i.e. tone mapping followed by HEVC encoding), but this figure drops to 1.3% in [10]. This is due to the Expectation Maximization algorithm used in our implementation for estimating the probability density function. Since, this step is only performed on the encoder side, the decoding time remains unchanged. Note that if a real time encoding is needed, the estimation of the p.d.f. can be alternatively performed in a similar way as in [10], by computing a histogram with a small number of bins. In this case, the probability for each bin must be transmitted to the decoder in place of the Gaussians parameters.

6. CONCLUSION

In this article, we developed a global tone mapping operator designed for compressing a High Dynamic Range image using a low bit depth encoder. Based on a statistical model of the compression scheme, we found a closed form solution for the optimal tone curve with respect to rate of the encoded tone mapped image and the distortion of the reconstructed HDR image. Our implementation based on the HEVC encoder showed improved rate distortion results compared to the existing global tone mapping operator that only minimizes the distortion without considering the rate.

7. REFERENCES

[1] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1649–1668, Dec. 2012.

[2] G. Ward and M. Simmons, "JPEG-HDR: A backwards-compatible, high dynamic range extension to jpeg," *Proceedings of the Thirteenth Color Imaging Conference*, 2005.

[3] G. Ward and M. Simmons, "Subband encoding of high dynamic range imagery," *Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, pp. 83–90, 2004.

[4] R. Mantiuk, A. Efremov, K. Myszkowski, and H.-P. Seidel, "Backward compatible high dynamic range MPEG video compression," *ACM SIGGRAPH*, Jul. 2006.

[5] F. Banterle, K. Debattista, P. Ledda, and A. Chalmers, "A gpu-friendly method for high dynamic range texture compression using inverse tone mapping," *Proceedings of graphics interface*, pp. 41–48, 2008.

[6] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM SIGGRAPH*, pp. 267–276, Jul. 2002.

[7] L. Yang and D. Wu, "Adaptive quantization using piecewise companding and scaling for gaussian mixture," *Visual Communication and Image Representation*, pp. 959–971, Oct. 2012.

[8] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, pp. 129–137, 1982.

[9] Y.Zhang, E. Reinhard, and D. Bull, "Perception-based high dynamic range video compression with optimal bit-depth transformation," *18th IEEE International Conference on Image Processing (ICIP)*, pp. 1321–1324, Sept. 2011.

[10] Z. Mai, H. Mansour, R. Mantiuk, P. Nasiopoulos, R. Ward, and W. Heidrich, "On-the-fly tone mapping for backward-compatible high dynamic range image/video compression," *ISCAS*, pp. 1831–1834, Jun. 2010.

[11] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, pp. 1–38, 1977.

[12] T. Wiegand, G. J. Sullivan, B. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13-7, pp. 560–576, Jul. 2003.

[13] "The JPEG-2000 still image compression standard," *ISO/IEC JTC Standard 1/SC29/WG1*, 2005.

[14] R. Bogart, F. Kainz, and D. Hess, "OpenEXR image file format," *ACM Siggraph, Sketches & Applications*, pp. 836–844, Jul. 2003.

[15] G. Bjontegaard, "Calculation of average PSNR differences between RD curves," *document VCEG-M33, ITU-T VCEG Meeting*, 2001.