On-the-Fly Tone Mapping for Backward-Compatible High Dynamic Range Image/Video Compression

Zicong Mai*, Hassan Mansour*, Rafal Mantiuk†, Panos Nasiopoulos*, Rabab Ward*, and Wolfgang Heidrich†
*Department of Electrical and Computer Engineering
†Department of Computer Science
University of British Columbia, Vancouver, BC, Canada

***Abstract*— In this paper, we propose a real-time tone-mapping scheme for backward compatible high dynamic range (HDR) video compression. The appropriate choice of a tone-mapping operator (TMO) can significantly improve the HDR quality reconstructed from a low dynamic range (LDR) version. We develop a statistical model that approximates the mean square error (MSE) distortion resulting from the combined processes of tone-mapping and compression. Using this model, we formulate a numerical optimization problem to find the tone-curve that minimizes the expected MSE in the reconstructed HDR sequence. We then simplify the developed model in order to reduce the computational complexity of the optimization problem to a closed-form solution. Performance evaluations show that the proposed methods provide superior performance in terms of HDR MSE and SSIM compared to existing tone-mapping schemes. It is also shown that the LDR image quality resulting from the proposed methods matches that produced by perceptually-based TMOs.**

I. Introduction

High dynamic range (HDR) scenes are characterized by an enhanced luminance range with an intra-frame contrast that exceeds 5-6 orders of magnitude ($10^6 : 1$). Such content can be well represented by advanced display technologies that utilize a bit-depth precision of at least 10-bit-per-color-channel and employ narrow-wavelength LED light sources. However, the majority of existing digital display devices can only support low dynamic range (LDR) 8-bit video content with a limited contrast that ranges between 2 or 3 orders of magnitude.

The broad deployment of HDR video formats relies on their backward compatibility with 8-bit devices. Such compatibility can be achieved if the HDR video stream contains 1) a backward-compatible 8-bit video layer which could be directly displayed on existing devices, and 2) additional information which along with this 8-bit layer can yield a good quality reconstructed version of the original HDR content. Such a stream can also contain a residual layer, i.e., enhancement layer, to further improve the quality of the HDR reconstruction.

Backward compatible HDR video encoding has received significant interest recently. Mantiuk et al. [1] derived a color space of encoding HDR content based on the luminance threshold sensitivity of the human visual system. For still image compression, backward compatibility can be achieved by encoding a tone-mapped copy of the HDR image together with a residual [2] or a ratio image [3] that allows the reconstruction of the original HDR image. Further implementations for backward-compatible video coding can be found in [4]–[7].

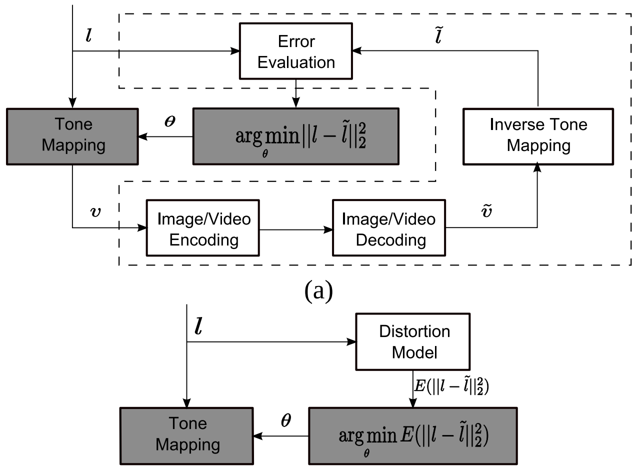


Fig. 1. System overview of the proposed tone-mapping method. (a) demonstrates the ideal scenario where the actual H.264/AVC encoding is employed. (b) shows the practical scenario which is addressed by this paper.

In this paper, we address the problem of finding an optimal tone-curve for such a backward-compatible encoding scheme. The remainder of this paper is organized as follows: the problem statement is presented in Section II. In Section III, the proposed tone-mapping approach is discussed in detail. Section IV shows the performance of the proposed methods. Conclusions are drawn in Section V.

II. Backward-compatible HDR Video Challenges

The performance of a backward-compatible HDR video and image encoding system depends on the coding efficiency of the LDR base layer and the HDR enhancement layer. Performance gains can be achieved by finding a TMO that preserves the necessary information in the LDR base layer so that after it passes through the inverse TMO process, the resultant HDR reconstructed signal is of high quality.

Our proposed approach attempts to find the best global (spatially invariant) tone-mapping curve that minimizes the mean square error (MSE) between the original HDR content and the reconstructed version obtained after tone-mapping, compression, decompression, and inverse tone-mapping . This

This work was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC).

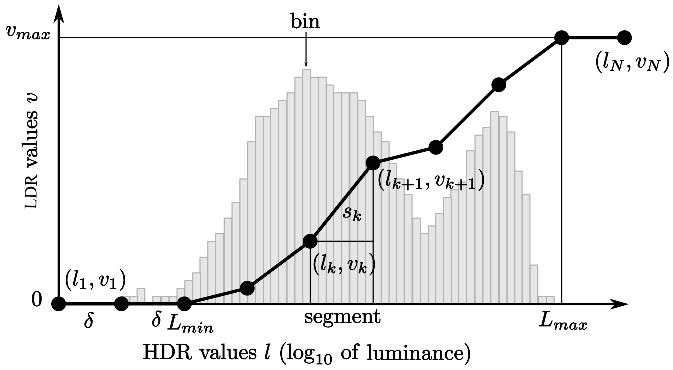


Fig. 2. Parameterization of a tone-mapping curve. The bar-plot in the background represents an image histogram used to compute $p(l)$.

process is illustrated in Fig. 1. Let l denote the input HDR image/frame, $\tilde{l}$ be the reconstructed HDR frame produced after inverse-tone-mapping, and θ be the set of parameters that control the tone-mapping operator, then our goal is to find the tone-mapping parameters θ that minimize $||l - \tilde{l}||_2^2$.

The above optimization problem can be solved by exhaustive search, repeatedly tone-mapping, encoding, decoding and then inverse tone-mapping, until the best set of TMO parameters θ^* is found (see Fig. 1a). Even though this approach guarantees an optimal solution, this framework requires an unacceptable computational cost. To overcome this problem, we estimate the distortion due to tone-mapping, encoding, decoding, and inverse tone-mapping with a statistical distortion model, as illustrated in Fig. 1b. Then, we show that under certain assumptions that are valid for natural images, an immediate closed-form solution for this problem can be found.

In this paper we consider only luminance/luma channels. To tone-map color images, we use the same tone-curve for the three color channels. Such approach was shown to well preserve color appearance for moderate contrast compression [8]. Encoding of the enhancement layer (for the residual data) is not considered in this paper. The rationale comes from our effort to achieve the best possible HDR reconstruction and thus the smallest possible residual. As a result, the cost of encoding any additional refinement layer would be minimized.

III. On-the-Fly Tone Mapping Solution

In this section we describe in detail how we parameterize the tone-mapping function, approximate encoding distortions with a statistical model and then find a closed-form solution for an optimal tone-curve.

A. Tone-Mapping Curve

The global tone-mapping curve is a function that maps HDR luminance values to LDR pixel values, and it is usually continuous and non-decreasing.

To keep the problem analytically tractable, we parameterize the tone-mapping curve as a piece-wise linear function with the nodes (l_k, v_k), as shown in Fig. 2. We represent HDR values as the base-10 logarithm of relative luminance values ($l = \log_{10}(L)$ where L is the luminance of the HDR image). Each segment has a constant width in HDR values equal to δ (0.1 in our implementation). The tone-mapping curve can then be uniquely specified by a set of slopes s_k, which forms a vector of tone-mapping parameters θ. Using this parameterization, the forward tone-mapping function is defined as:

$$\tilde{v} = (l - l_k) \cdot s_k + v_k, \tag{1}$$

where v is the pixel value, and k is the segment corresponding to HDR value l, that is $l_k \le l < l_{k+1}$. The inverse mapping function is then:

$$\tilde{l}(v; s_k) = \begin{cases} \dfrac{v - v_k}{s_k} + l_k & \text{for} \quad s_k > 0 \\ \sum\limits_{l \in S0} l \cdot p_L(l) & \text{for} \quad s_k = 0 \end{cases}, \tag{2}$$

where $s_k \in \{s_1..s_N\}$. When the slope is zero ($s_k = 0$), $\tilde{l}(v; s_k)$ is assigned an expected HDR pixel value for the entire range S_0 in which the slope is equal zero. $p_L(l)$ is the probability of HDR pixel value l.

B. Statistical Distortion Model

As mentioned earlier in section II, accurately computing the distorted HDR values $\tilde{l}$ would be too computationally demanding. Instead, we estimate the error $(\tilde{l} - l)^2$ assuming that H.264/AVC distortions follow a known probability distribution p_{H264}, and we find that p_{H264} can be modelled using a generalized Gaussian distribution function. Under this assumption, the expected value of the error $(\tilde{l} - l)^2$ is:

$$\sum_{l=l_{min}}^{l_{max}} \sum_{v=0}^{v_{max}} \sum_{\tilde{v}=0}^{v_{max}} (\tilde{l}(\tilde{v}; s_k) - l)^2 \cdot p_{H264}(v - \tilde{v} | v) \cdot p(v|l) \cdot p_L(l), \tag{3}$$

where $p(v|l)$ is given by the following expression

$$p(v|l) = \begin{cases} 1, & \text{for} \quad v = \lfloor (l - l_k) \cdot s_k + v_k \rfloor \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

for a given tone-curve since the TMO is a many-to-one mapping. $\lfloor \cdot \rfloor$ denotes the floor function.

Assuming that the compression error probability is independent of the LDR pixel value v, we can simplify the expression above by removing the dependency of p_{H264} on v, i.e. dropping the summation over v. Consequently, the continuously relaxed objective function is written as:

$$\varepsilon(s_k) = \sum_{l=l_{min}}^{l_{max}} \sum_{\tilde{v}=0}^{v_{max}} (\tilde{l}(\tilde{v}; s_k) - l)^2 \cdot p_{H264}(v - \tilde{v}) \cdot p_L(l) \tag{5}$$

C. Optimization Problem

The optimum tone curve can be found by minimizing the function $\varepsilon(s_k)$ with respect to the segment slopes s_k:

$$\arg\min_{s_1..s_N} \varepsilon(s_k) \tag{6}$$

subject to:

$$\begin{gathered} s_{min} \le s_k \le s_{max} \quad \text{for} \quad k = 1..N \\ \sum_{k=1}^{N} s_k \cdot \delta = v_{max}. \end{gathered} \tag{7}$$

The first constraint restricts slopes to the allowable range, while the second ensures that the tone curve spans exactly the range of pixel values from 0 to v_{max}. Considering the sensitivity of the human vision, s_{max} is set to be $(\log_{10}(1.01))^{-1}$.

D. Closed-form Solution

The distortion model in (5) gives a good estimate of compression errors but has relatively high computational complexity for real-time applications. With an additional assumption, we can find a closed-form solution with almost no noticeable impact on the compression performance.

If we assume local linearity of the tone-curve, so that the slope at the non-distorted pixel value v and that at the distorted pixel value $\tilde{v}$ is the same, we can then substitute $\tilde{l}(\tilde{v}; s_k)$ and l in the distortion model (5) using the inverse mapping function in (2), this gives:

$$\varepsilon(s_k) \approx \sum_{l=l_{min}}^{l_{max}} \sum_{\tilde{v}=0}^{v_{max}} p_{H264}(v-\tilde{v}) \cdot p_L(l) \cdot \left(\frac{v-\tilde{v}}{s_k}\right)^2. \quad (8)$$

After reorganizing we get:

$$\begin{aligned} \varepsilon(s_k) &\approx \sum_{l=l_{min}}^{l_{max}} \frac{p_L(l)}{s_k^2} \cdot \sum_{\tilde{v}=0}^{v_{max}} p_{H264}(v-\tilde{v}) \cdot (v-\tilde{v})^2 \\ &= \sum_{l=l_{min}}^{l_{max}} \frac{p_L(l)}{s_k^2} \cdot Var(v-\tilde{v}). \end{aligned} \quad (9)$$

Since the variance of $(v-\tilde{v})$ does not depend on the slopes s_k, it does not affect the location of the global minimum of $\varepsilon(s_k)$ and thus can be omitted when searching for the minimum.

Our local linearity assumption holds in most cases for two reasons: 1) the distortion distribution p_{H264} has high kurtosis so that most of the distorted pixels are likely to lie in the same segment as the non-distorted pixel v. 2) Even if a distorted pixel $\tilde{v}$ moves to another segment, the slopes of two neighboring segments are usually very close to each other.

The optimization problem defined in (6) can now be rewritten accordingly and solved analytically by calculating the first order Karush-Kuhn-Tucker (KKT) optimality conditions of the corresponding Lagrangian, and the resulting slopes s_k are given by:

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

IV. Experimental Results and Discussion

This section evaluates the performance of our models by the comparison with existing tone-mapping methods. The chosen TMOs are the photographic TMO [9], the adaptive logarithmic TMO [10] and the display adaptive TMO [11]. In [12], a study shows that the photographic TMO and the adaptive logarithmic TMO were found to outperform other popular tone-mapping methods in backward-compatible HDR image compression. Display adaptive TMO is a recent tone-mapping algorithm at the time of writing and it employs a similar optimization loop as our technique. In the experiments below, all tone-mapped images are compressed/decompressed using the H.264/AVC reference software [13]. To reconstruct an HDR image from a decoded LDR image, an inverse tone-mapping function is stored as a lookup table with each encoded image.

Fig. 3 compares the distortion of the reconstructed HDR image versus the compressed LDR bit rate for different TMOs, averaged over 40 images. This test demonstrates how successful each TMO is at delivering a good quality HDR by inverse tone mapping the corresponding LDR signal. The results show that our proposed methods clearly outperform the other methods in terms of MSE. The difference is not very large for low bit rates (heavy compression), but the performance of our model dramatically improves when the compression reaches the point of the medium and light compression. For HDR MSE = -3 (in log10 scale), which corresponds to QP = 25, we save about 50% of the bit-rate compared to the best performing competitive TMO for the same quality.

Although our models are designed for minimizing MSE, the results also indicate that the proposed TMOs show superior performance for the advanced quality metric, SSIM.

Fig. 4 display the tone-mapped LDR images for three images, which demonstrate that the images tone-mapped using our method also provide good quality.

V. Conclusion

In this paper, we showed that the appropriate choice of a tone-mapping operator can significantly improve the reconstructed HDR quality. We developed a statistical model that approximates the distortion resulting from the combined processes of tone-mapping and compression. Using this model, we formulated an optimization problem that finds the tone-curve which minimizes the expected HDR MSE. With an additional assumption, the optimization problem can be solved analytically with a closed-form solution. The closed-form solution is computationally efficient and has a performance compatible to our developed statistical model. Although our models are designed to minimize MSE, the extensive performance evaluations show that the proposed methods provide excellent performance in terms of SSIM and the LDR image quality, in addition to an outstanding performance in MSE.

References

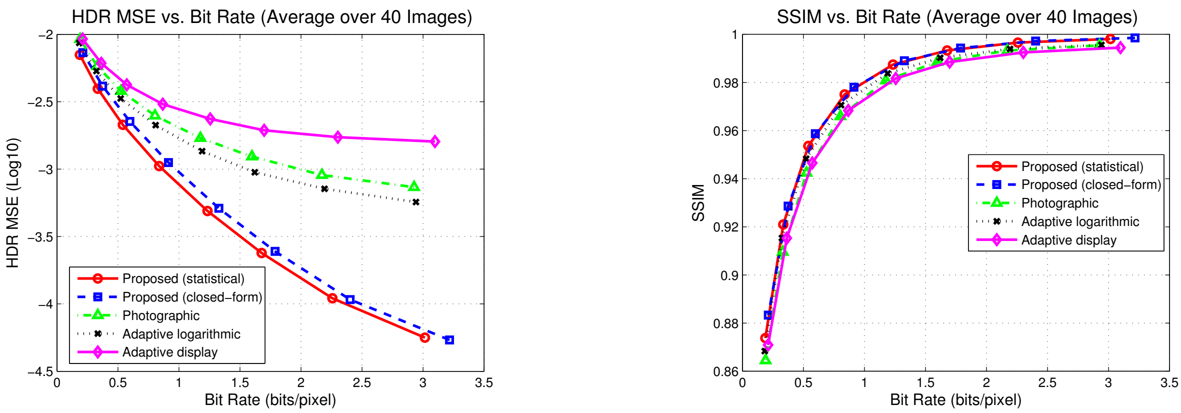
[1] R. Mantiuk, G. Krawczyk, K. Myszkowski, and H.-P. Seidel, “Perception-motivated high dynamic range video encoding,” *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 730–738, 2004.

[2] K. E. Spaulding, G. J. Woolfe, and R. L. Joshi, “Using a residual image to extend the color gamut and dynamic range of an sRGB image,” in *Proc. of IS&T PICS Conference*, 2003, pp. 307–314.

[3] G. Ward and M. Simmons, “JPEG-HDR: A backwards-compatible, high dynamic range extension to JPEG,” in *Proceedings of the 13th Color Imaging Conference*, 2005, pp. 283–290.

[4] G. Sullivan, H. Yu, S.-i. Sekiguchi, H. Sun, T. Wedi, S. Wittmann, Y.-L. Lee, A. Segall, and T. Suzuki, “New standardized extensions of mpeg4-avc/h.264 for professional-quality video applications,” *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 1, pp. I–13–16, 2007.

[5] M. Winken, D. Marpe, H. Schwarz, and T. Wiegand, “Bit-depth scalable video coding,” *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 1, pp. I–5–8, 2007.



(a) (b)

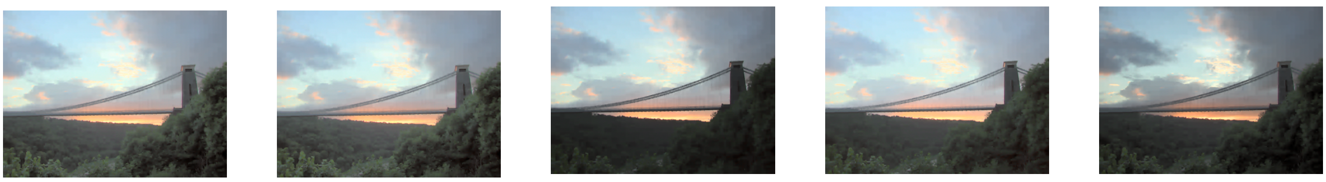
Fig. 3. Comparison with other tone-mapping methods in terms of MSE and SSIM (for the reconstructed HDR image) vs. bit rate, averaged over 40 images.

Statistical | Closed-form solution | Photographic TMO | Adaptive logarithmic TMO | Display Adaptive TMO

(a) Image: AtriumNight; LDR images shown are compressed with QP = 10

Statistical | Closed-form solution | Photographic TMO | Adaptive logarithmic TMO | Display Adaptive TMO

(b) Image: Coby; LDR images shown are compressed with QP = 22



Statistical | Closed-form solution | Photographic TMO | Adaptive logarithmic TMO | Display adaptive TMO

(c) Image: BristolBridge; LDR images shown are compressed with QP = 38

Fig. 4. Tone-mapped images for individual images. Each row shows the tone-mapped image after compression for one image. The compression quantization parameters used for the three images are 10, 22 and 38 respectively.

[6] A. Segall and J. Zhao, "Bit-stream rewriting for SVC-to-AVC conversion," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, 2008, pp. 2776–2779.

[7] Y. Wu, Y. Gao, and Y. Chen, "Bit-depth scalability compatible to H. 264/AVC-scalable extension," *Journal of Visual Communication and Image Representation*, vol. 19, no. 6, pp. 372–381, 2008.

[8] R. Mantiuk, R. Mantiuk, A. Tomaszewska, and W. Heidrich, "Color correction for tone mapping," *Computer Graphics Forum (Proc. of EUROGRAPHICS)*, vol. 28, no. 2, pp. 193–202, 2009.

[9] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 21, no. 3, pp. 267–276, 2002.

[10] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Computer Graphics Forum (Proc. of Eurographics)*, vol. 22, no. 3, pp. 419–426, 2003.

[11] R. Mantiuk, S. Daly, and L. Kerofsky, "Display Adaptive Tone Mapping," *ACM Transactions on Graphics*, vol. 27, no. 3, pp. 68–68, 2008.

[12] R. Mantiuk and H. Seidel, "Modeling a Generic Tone-mapping Operator," *Computer Graphics Forum (Proc. of Eurographics'08)*, vol. 27, no. 2, pp. 699–708, 2008.

[13] "H.264/AVC JM14.2 Reference Software," available: http://iphome.hhi.de/suehring/tml/.