REAL-TIME HDR VIDEO TONE MAPPING USING HIGH EFFICIENCY VIDEO CODING

Mingyun Kang *Joo Ho Lee* *Inchang Choi* *Min H. Kim*

Korea Advanced Institute of Science and Technology (KAIST)

ABSTRACT

High-dynamic-range (HDR) video streams have been delivered through high efficiency video coding (HEVC). HDR video tone mapping is additionally required but is operated separately to adjust the content's dynamic range for each display device. HDR video tone mapping and HEVC encoding share common computational processes for spatial and temporal coherence in a video stream; however, they have been developed and implemented independently with their own computational budgets. In this work, we propose a practical HDR video tone-mapping method that combines two overlapping computational blocks in HDR video tone-mapping and HEVC compression frameworks with the objective of achieving real-time HDR video tone mapping. We utilize precomputed coding blocks and motion vectors so as to achieve spatial and temporal coherence of HDR video tone-mapping in the decoding stage, even without introducing substantial computational cost. Results demonstrate that our method can achieve real-time performance without compromising the video quality, which is highly comparable to that of state-of-the-art video tone-mapping methods.

1. INTRODUCTION

High-dynamic-range (HDR) video contents have been delivered through traditional video compression standards with adjustment engineering by SMPTE [1]. Different from traditional low dynamic range (LDR) video processing, an additional process, so-called HDR video tone mapping, is required to adjust the dynamic range of contents into the range of each display device [2, 3].

An HDR tone-mapping operator (TMO) consists of two different types of processes: A global tone-curve function that adjusts the dynamic range of contents for each pixel [4, 5] and a local adaptation operator that enhances spatially-varying luminance adaptation levels, often implemented as a low-pass filter [6] or a bilateral filter [7]. In addition, HDR video tone mapping requires another temporal operator that mimics temporal adaptation in the human visual system [3]. Despite the recent advances in HDR imaging technologies [8, 9], the applicability of HDR video has been restricted by the additional cost of video tone mapping. In this work, we propose a novel HDR video tone-mapping method that can reduce the computational cost by combining common processes in the high efficiency video coding (HEVC) compression [1] and HDR video tone-mapping frameworks.

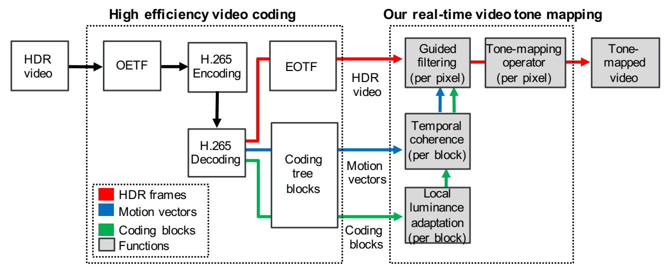

Fig. 1: Overview of our practical HDR tone-mapping framework combined with the HEVC standard.

Recently, several video compression standards, such as MPEG-2 Video, MPEG-4 AVC, and HEVC, have been proposed and broadly used for efficient distribution of high-resolution video streams. The video compression standard consists of two different stages: coding and decoding [10]. In the coding stage, according to local similarity, each video frame is segmented to variously sized coding blocks in HEVC. Also, inter-frame, bidirectional motion vectors of coding blocks are estimated using an optical flow operator so that the differences between intermediate frames can be coded sparsely with temporal coherency to save the stream bandwidth of coding blocks.

Modern HDR video tone-mapping operators and the HEVC standard for video compression share similar features, but with different objectives of tone-mapping and compression, respectively. First, a local adaption operator in tone mapping is applied to adjust the parameters of the tone-curve function by accounting for the regional similarity of luminance within a scene [4, 6]. Coding blocks in the compression standard are estimated by accounting for the local similarity of image structures to reduce the number of coefficients for frequency bases, resulting in different sizes of coding blocks [11]. These two spatial operators consider the local similarity within image structures. Second, while a temporal coherence operator is commonly employed to avoid the typical flickering artifacts in video tone mapping via temporal filtering [3], inter-frame motion vectors of coding blocks in video compression are also calculated via temporal filtering to reduce the redundancy of coding blocks by accounting for temporal coherency [11]. The HDR video tone mapping operators and HEVC standards share these similar computation processes that we combine in this work.

Recently, several works that exploit the applicability of an HDR tone-mapping operator to the encoding stage of the video compression standard for reducing the number of bits with fewer quantization artifacts have been introduced [12, 13, 14, 15]. They propose to apply an HDR tone-mapping operator for converting floating to integer numbers to reduce quantization errors. In contrast, we propose a novel tone-mapping algorithm that adopts the coding-block information precomputed in the HEVC coding stage toward HDR video tone mapping to save costs for the real-time performance of HDR video tone mapping. To the best of our knowledge, the proposed tone-mapping method is the first work that combines HDR video tone mapping with encoding information in HEVC for real-time video tone mapping, and can be easily adopted into existing video decoders through simple modification.

2. HDR VIDEO TONE MAPPING FOR HEVC

The key challenge for HDR video tone mapping is the computational burden for estimating (1) local adaptation of luminance and (2) temporal coherence in adaptation, in addition to computing a global tone-curve function for every pixel. Although computing spatiotemporal adaptation information for every frame is significantly expensive, it has been an inevitable cost to achieve high-quality HDR tone-mapping. We are motivated to substitute local luminance adaptation and coherent temporal adaptation by utilizing the spatiotemporal information of coding blocks and motion vectors computed in the encoding stage of video compression. See Figure 1 for an overview. Different from other video tone-mapping methods, we first estimate luminance parameters for a tone-curve function per coding block for local adaptation of luminance. We then propagate the parameters across frames through motion vectors to achieve temporal coherence. Our novel approach can save costs in HDR video tone mapping significantly without compromising the image quality or suffering from flickering artifacts.

2.1. Tone Mapping Operator

A global tone mapping operator can be used to compress the dynamic range of HDR contents in our video tone-mapping framework. Here, we start from an efficient logarithmic tone-curve function proposed by Drago et al. [4]. We apply the original function to the luminance channel in the CIE Yxy space as:

$$L_d = \frac{0.01 \cdot L_{dmax}}{\log(L_{wadapt}+1)} \cdot \frac{\log(L_w+1)}{\log\left(2+8\cdot\left(\left(\frac{L_w}{L_{wadapt}}\right)^{\frac{\log(b)}{\log(0.5)}}\right)\right)},$$

where L_{dmax} is the maximum luminance of a display, L_w is the scene luminance normalized to L_{dmax}, $\log()$ indicates a 10-based logarithmic function, b is a bias parameter that controls the contrast ratio in tone mapping, and the target L_{dmax} is set to 100 in our experiment.

Different from the original formulation, we revise the calculation of the luminance adaptation parameter L_{wadapt} by computing the local adaptation level per block:

$$L_{wadapt} = e \cdot L_{wmax} / \log\left(\frac{\bar{L}_w+1}{\bar{L}_w}\right), \quad (1)$$

where e is a user parameter of the exposure factor that controls the overall brightness level of tone mapping, L_{wmax} is the maximum luminance per coding block, and $\bar{L}_w$ is the geometric mean of luminance of the scene so that we can combine both local and global adaptation as a single operator.

Transfer Functions. Pixel values in an HDR frame are recorded as floating-point numbers with single or half precision [16]. Since the bit depth of recent HEVC is limited up to 10 or 12-bit integer representation of a pixel value, an additional signal transfer function is required to distribute HDR signals through the integer-based codec workflow. In the coding stage in HEVC, an opto-electronic transfer function (OETF) is applied to adjust the contrast of linear HDR signals for the purpose of converting floating-point numbers of signals to integers [1]; for instance, the perceptual quantization (PQ) based on Barten [17] or the hybrid log-gamma (HLG) [18] has been used. In the decoding stage, an electro-optical transfer function (EOTF) or the inverse OETF is used to recover linear HDR video signals. Note that the transfer function distributes HDR signals through the HEVC codec and that an HDR tone-mapping operator is an independent, additional step to adjust the dynamic range of contents for rendering on display.

2.2. Local Adaptation via Coding Blocks

For simulating local luminance adaption in the human visual system, different edge-aware filtering methods have been devised to separate HDR frames into base and detail layers in HDR video tone-mapping studies [19, 3, 20]. To avoid the computational burden by spatial filtering, we do not separate a frame into base and detail layers but instead utilize coding blocks to account for locally adapted luminance levels as tone-mapping parameters on each block (L_{wadapt} in Eq. (1)).

We found that each block in HEVC does not present significant structural variation except intra-prediction modes (refer to the below paragraph). The various-sized block segmentation for video compression resembles edge-aware filtering for local luminance adaptation in HDR tone-mapping. We, therefore, estimate the parameters of a global tone-mapping operator (Section 2.1) for each coding block to account for local luminance adaptation, rather than applying an edge-aware filter individually on each frame.

Coding Blocks. In video compression, each frame of the input video is decomposed into tree-structured partitions, so-called coding tree units (CTUs). Each CTU includes luma coding tree blocks (CTBs) and chroma CTBs. While the block size is fixed to 16×16 in MPEG-2 and H.264/AVC, various sizes with the subdivision structure are available from 4×4 to 64×64 in HEVC that we chose [10]. A coding tree block of luma can be partitioned into multiple coding blocks (CBs) in the quadtree structure. Also, a coding block can be divided into either quadrant prediction blocks (PBs) or transform blocks (TBs). A prediction block includes a bidirectional motion vector to copy other CBs at different pixel positions to reduce temporal redundancy. Depending on image structures, a CB can be classified into different shapes of PBs in the prediction mode. We use motion vectors of inter PBs to achieve temporal coherency (Section 2.3). A transform block provides subdivided coefficients to restore non-zero elements more effectively [21].

Intra-Prediction Blocks. Different from inter-coding blocks across frames, intra-prediction blocks in HEVC have different modes depending on image structures in the individual block, such as planar, DC, and angular modes of different orientations [10]. We found that our local smoothness assumption is inapplicable particularly in the case of intra-prediction blocks of a large size. When we have intra-PBs of a large size, we subdivide them into 8-by-8 blocks.

Deblocking via Guided Filtering. We apply the fast guided filter [22] on per-block estimation of the luminance parameters p to mitigate boundary artifacts over blocks, yielding an artifact-free frame q. The input HDR video frame I is used as the filter guidance for each frame.

2.3. Temporal Coherence via Motion Vectors

In this work, we use motion vectors in HEVC to achieve temporal coherence. Once we estimate the tone-mapping parameters L_{wadapt} of luminance per block (Section 2.2), we classify per-block parameters into two groups: intra-coded blocks and inter-coded blocks with motion vectors. Intra-coded blocks are temporally consistent across frames, while inter-coded blocks are copied and blended at different positions [10]. In inter-coded block regions, we update luminance parameters via motion vectors, which associate inter-coded blocks within the range of 32 frames. Finally, once we update per-block estimation of luminance adaptation parameters, we apply additional smooth filtering to the parameters across frames by blending the per-pixel parameters of both the preceding and following frames with a Gaussian weight to enhance temporal coherency.

Motion Vectors. Video frames in HEVC are divided into intra-coded and two different inter-coded frames (or slices for scan lines). An intra-coded frame (I-frame), the so-called keyframe, is independent of other frames but is a complete image frame that includes all coding blocks within the frame. Inter-coded frames are divided into predicted frames (P-frame) and bidirectional predicted frames (B-frame). A P-frame has differences of the current frame against previous ones via motion vectors. A B-frame includes differences of the current one against both the preceding and following frames. There are seminal works that apply motion vectors preliminarily to enhance video quantization and frame interpolation in video compression [14, 23], but we use motion vectors of P-/B-frames differently to achieve temporal coherence for real-time HDR video tone-mapping.

3. RESULTS

As shown in Figure 1, our video tone-mapping algorithm is used together with the state-of-the-art video compression standard, HEVC. We implemented the video compression part based on an open-source version of HEVC [24]. We revise the HEVC framework to extract coding tree block information that includes motion vectors and image data from coding tree units, feeding them with input HDR video signals to our tone-mapping algorithm. Regarding gamma correction, we set the display gamma as 2.44, as described in the original equation [4], but exclude the standard linear part for low signals, as it introduces artifacts in dark regions.

Performance Comparison. To achieve real-time performance, we implemented our tone-mapping algorithm in a parallel manner suitable for GPU computation. The entire computation processes of our tone-mapping are implemented on a GPU platform using CUDA. The performance of our method is compared with that of other real-time tone-mapping algorithms, tested using a machine equipped with an Intel i7-3770 CPU with 32GB of memory and an NVIDIA 1080 Ti GPU with 11GB of memory. Since the original implementations of the previous works are unavailable in public, we implemented three video tone-mapping methods [19, 3, 20].

Table 1 compares the performance of four tone-mapping algorithms to compute a single frame of 1920-by-1080 (1080p). Croci's method [19] is the real-time version of a high-quality video TMO [3], which took 7,545 milliseconds (ms) on the same machine. Eilertsen's method [20] is a real-time video TMO. It decomposes a frame into the base and detail layers based on the partial differential equation (PDE). It increases computational costs significantly. Our method achieves the fastest speed to compute a frame (14.20 ms). However, our method's performance could be degraded with a large filter kernel, resulting in a speed-quality tradeoff.

TMOs	Eilertsen	Croci	Ours
Temporal filtering (ms)	0.08	1.00	3.00
Spatial filtering (ms)	15.60	22.00	10.20
Tone mapping (ms)	3.43	1.00	1.00
Total time (ms)	19.11	24.00	14.20
Framerate (fps)	52.33	41.67	70.42

Table 1: Performance comparison.

Image-Quality Evaluation. To evaluate the performance of our method regarding the visual quality of tone-mapped

(a) Aydin et al. (time: 7545.00ms) (b) Eilertsen et al. (time: 19.11ms) (c) Croci et al. (time: 24.00ms) (d) Ours (time: 14.20ms)

Fig. 2: Our method is compared with two state-of-the-art methods: Aydin et al. [3] and Eilertsen et al. [20]. Our approach presents comparatively high image quality even with a faster speed (14.20 ms). Refer to the supplemental video.

videos, we compared our method to two state-of-the-art real-time video tone-mapping methods [19, 20] and one offline method [3]. See Figure 2 and the supplemental video for results. Aydin's, Croci's and our method present effective compression of scene contrast, but often losing contrast in dark scenes. Eilersten's method preserves contrast well; however, it often suffers from halo artifacts in bright scenes.

We conducted category judgments to evaluate artifact visibility and image appearance, respectively, where seven participants observed 20 video clips of five different scenes from the Froehlich HDR video datasets [25]. We tone-mapped the video footages by four different algorithms: Aydin et al. [3], Eilertsen et al. [20], Croci et al. [19], and ours, and presented them in random order in a dark room. Refer to the supplemental video for compared video stimuli.

In this experiment, the participants were asked to choose a category of rating for each video footage regarding three image appearance attributes (brightness, contrast and colorfulness) and three visual artifacts (flickering, ghosting and unnaturalness). The experiments were conducted on an LCD display device (Dell U2412M) in controlled environments under dark viewing conditions, where participants were seated at a distance of approximately 60 cm.

Figure 3 presents a statistical analysis of the psychophysical experiments, where we calculate the average responses of the participants and each side of the error bars indicates one standard deviation of the participants. In the top row plots, the middle level indicates the "just right" tone-mapping results; in the bottom row plots, the bottom-level indicates the invisible artifacts. These results validate that although our method achieves the highest frame rate among compared real-time video tone-mapping methods, there is no compromise in video quality, compared with the existing state-of-the-art methods.

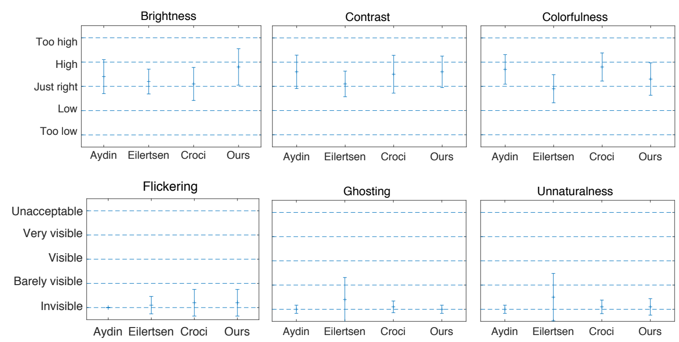

Fig. 3: Statistical analysis of the image quality experiments. Each node indicates the averages of the participants' answers; each wing of the error bars shows one standard deviation.

4. CONCLUSION

We have presented a practical real-time video tone-mapping method that accounts for spatial and temporal adaptations in HDR video frames. We can substitute expensive spatial and temporal operators in HDR video tone-mapping with spatiotemporal information in coding blocks in HEVC to reduce computational cost. This allows us to reduce computational costs in the HDR video tone-mapping framework. Our real-time video results validate the effectiveness of our practical approach without compromising the tone-mapped video quality. The proposed method could be beneficial for rendering HDR video contents on low-cost commodity display devices.

5. ACKNOWLEDGMENT

Min H. Kim acknowledges Korea NRF grants (2016R1A2B-2013031, 2013M3A6A6073718), Cross-Ministry Giga KOREA Project (GK17P0200), ETRI, and an ICT R&D program of MSIT/IITP of Korea (2016-0-00018).

6. REFERENCES

[1] S. Miller, M. Nezamabadi, and S. Daly, "Perceptual signal coding for more efficient usage of bit codes," *SMPTE Motion Imaging Journal*, vol. 122, no. 4, pp. 52–59, May 2013.

[2] Min H. Kim, Tim Weyrich, and Jan Kautz, "Modeling human color perception under extended luminance levels," *ACM TOG*, vol. 28, no. 3, 2009.

[3] Tunç Ozan Aydin, Nikolce Stefanoski, Simone Croci, Markus H. Gross, and Aljoscha Smolic, "Temporally coherent local tone mapping of HDR video," *ACM TOG*, vol. 33, no. 6, 2014.

[4] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Wiley CGF*, vol. 22, no. 3, 2003.

[5] Min H. Kim and Jan Kautz, "Consistent tone reproduction," in *Proc. IASTED Int. Conf. Computer Graphics and Imaging (CGIM 2008)*, 2008, pp. 152–159.

[6] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 267–276, July 2002.

[7] Frédo Durand and Julie Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *ACM Trans. Graph*, vol. 21, no. 3, pp. 257–266, 2002.

[8] Inchang Choi, Seung-Hwan Baek, and Min H. Kim, "Reconstructing interlaced high-dynamic-range video using joint learning," *IEEE Transactions on Image Processing (TIP)*, vol. 26, no. 11, pp. 5353 – 5366, 2017.

[9] Min H. Kim and Jan Kautz, "Characterization for high dynamic range imaging," *Computer Graphics Forum (Proc. EUROGRAPHICS 2008)*, vol. 27, no. 2, pp. 691–697, 2008.

[10] Gary J. Sullivan, Jens-Rainer Ohm, Woojin Han, and Thomas Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Techn*, vol. 22, no. 12, pp. 1649–1668, 2012.

[11] D. Grois, D. Marpe, A. Mulayoff, B. Itzhaky, and O. Hadar, "Performance comparison of h.265/mpeg-hevc, vp9, and h.264/mpeg-avc encoders," in *2013 Picture Coding Symposium (PCS)*, 2013, pp. 394–397.

[12] Tsun-Hsien Wang, Wei-Su Wong, Fang-Chu Chen, and Ching-Te Chiu, "Design and implementation of a real-time global tone mapping processor for high dynamic range video," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2007, vol. 6, pp. VI–209.

[13] Yuanyuan Dong, Panos Nasiopoulos, and Mahsa T Pourazad, "HDR video compression using high efficiency video coding (HEVC)," *reproduction*, vol. 6, 2012.

[14] Ronan Boitard, Dominique Thoreau, Rémi Cozot, and Kadi Bouatouch, "Motion-guided quantization for video tone mapping," in *ICME*. 2014, pp. 1–6, IEEE Computer Society.

[15] Cagri Ozcinar, Paul Lauga, Giuseppe Valenzise, and Frédéric Dufaux, "Hdr video coding based on a temporally constrained tone mapping operator," in *Digital Media Industry & Academic Forum (DMIAF)*. IEEE, 2016, pp. 43–47.

[16] R Bogart, F Kainz, and D Hess, "OpenEXR image file format," *ACM SIGGRAPH 2003, Sketches & Applications*, 2003.

[17] Peter GJ Barten, "Formula for the contrast sensitivity of the human eye," in *Image Quality and System Performance*. International Society for Optics and Photonics, 2003, vol. 5294, pp. 231–239.

[18] Tim Borer and Andrew Cotton, "A "display independent" high dynamic range television system," 2015.

[19] S. Croci, T. O. Aydın, N. Stefanoski, M. Gross, and A. Smolic, "Real-time temporally coherent local hdr tone mapping," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sept 2016, pp. 889–893.

[20] Gabriel Eilertsen, Rafal K. Mantiuk, and Jonas Unger, "Real-time noise-aware tone mapping," *ACM Trans. Graph*, vol. 34, no. 6, pp. 198:1–198:15, 2015.

[21] J. Sole, R. Joshi, N. Nguyen, T. Ji, M. Karczewicz, G. Clare, F. Henry, and A. Duenas, "Transform coefficient coding in hevc," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1765–1777, Dec 2012.

[22] Kaiming He, Jian Sun, and Xiaoou Tang, "Guided image filtering," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, 2013.

[23] Lino Coria and Panos Nasiopoulos, "Using temporal correlation for fast and high-detailed video tone mapping," in *Proc. IEEE International Conference on Imaging Systems and Techniques (IST)*, 2010, pp. 329–332.

[24] OpenHEVC, "OpenHEVC," https://github.com/OpenHEVC/openHEVC, 2017.

[25] Jan Froehlich, Stefan Grandinetti, Bernd Eberhardt, Simon Walter, Andreas Schilling, and Harald Brendel, "Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR-displays," in *Digital Photography*, 2014, p. 90230X.