IMPACT OF TEMPORAL COHERENCE-BASED TONE MAPPING ON VIDEO COMPRESSION

Ronan Boitard[1,2], *Dominique Thoreau*[1]

[1]Technicolor
975 Av. des Champs Blancs,
35576 Cesson-Sevigne, France

Rémi Cozot[2], *Kadi Bouatouch*[2]

[2]IRISA
263 Av. du Général Leclerc,
35000 Rennes, France

ABSTRACT

Tone Mapping Operators (TMOs) aim at converting real world high dynamic range (HDR) images captured with HDR cameras, into low dynamic range (LDR) images that can be displayed on LDR displays. Even though most of the designed solutions provide good results for still HDR images, they are not efficient for tone mapping video sequences. The main issue is their inability to preserve the temporal correlation inherent in a video sequence. This has a consequence on the video compression efficiency. In this work, we show that higher compression ratios can be obtained by preserving the temporal coherency of a sequence. We evaluate temporal coherency and video compression in regard to two aspects. The first one evaluates the quality of the decoded LDR sequences after applying different TMOs. The second aspect assesses the quality of a reconstructed HDR sequence when a TMO and a codec are applied.

Index Terms— Video Tone Mapping, HDR Video, Video Compression, Temporal Coherency

1. INTRODUCTION

Tone Mapping Operators (TMOs) compress High Dynamic Range (HDR) images or video sequences to a lower dynamic range. Thanks to recent developments in HDR acquisition materials [1], HDR contents are now available. The Academy Color Encoding Specification (ACES) even recommends to use HDR representation (OpenEXR half-float format) in their workflow.

Many studies have evaluated the subjective quality of TMOs [2, 3]. However, to our knowledge, none of the existing evaluations addresses the efficiency of the TMOs regarding compression schemes known as codec (coder-decoder). Such an evaluation is needed as tone mapped contents are to be used in many applications (broadcasting, media storage, etc.) that require compression.

A codec removes redundant information using the spatial and temporal correlation of a video signal. This allows to only encode the information that cannot be deduced on the decoder side. The more correlated a signal, the higher the compression ratio. Tone mapping separately each frame of a video sequence results in abrupt changes of the tone map curve, which reduces the temporal correlation.

In this paper, we show, for a targeted bit-rate, that using temporally coherent TMOs results in decoded tone mapped video sequences of higher quality. In addition to display purposes, TMOs are also used in every scalable backward compatible HDR video codec. That is why, we also evaluate the temporal coherency in regard to the preservation of the original HDR video sequence after applying a TMO and a codec.

The next section introduces three TMOs, one designed for still images and two that preserve temporal coherency. It also gives an overview of the standard ITU-T H.265 / MPEG-H Part 2 'High Efficiency Video Codec' (HEVC) [4] as well as a scalable backward compatible HDR video codec. Section 3 presents three methods to evaluate the compression efficiency of different TMOs as well as some results. Finally, conclusion and future work are presented.

2. RELATED WORK

Recall that most of the existing TMOs are designed for still images. When applied to video sequences, the temporal coherency is not preserved. Indeed, TMOs optimize the preservation of the HDR dynamic range unto the LDR one. As a consequence, successive frames may be tone mapped quite differently which is source of flickering and temporal brightness incoherency in a video sequence.

In this section, we first present a TMO for still images, say the Photographic Tone Reproduction (PTR) [5]. Then, we describe an extension of the PTR algorithm [6] that reduces flickering artifacts when tone mapping video sequences. Finally, the last technique [7], that adapts to any TMO, aims at preserving the temporal brightness coherency of the original HDR video sequence.

To evaluate the compression efficiency of these TMOs, we use version 10 of the HEVC test Model (HM 10.0) [8]. We briefly describe the HM in section 2.3. We also give an overview of a scalable backward compatible video codec.

2.1. Photographic Tone Reproduction (PTR)

The PTR algorithm uses a technique, designed by Adams [9], to rescale HDR frames which is analogous to setting exposure in a camera:

$$L_s = \frac{a}{\kappa} \cdot L_w \, , \tag{1}$$

$$\kappa = exp\left(\frac{1}{N} \cdot \sum_{n=1}^{N} log(\delta + L_w(n))\right) \, , \tag{2}$$

where a is the chosen exposure, L_w the HDR luminance and L_s the scaled luminance. The key value κ is a subjective indication of the image's overall brightness level. It is computed using Equation (2), where δ is a small value to avoid singularity and N the number of pixels in the image. The tone map curve is a sigmoid function given by Equation (3):

$$L_d = \frac{L_s}{1 + L_s} \cdot \left(1 + \frac{L_s}{L_{white}^2}\right) \, , \tag{3}$$

where L_{white} is used to burn out areas with high luminance value and L_d is the tone map LDR luma. Two parameters (a and L_{white}) are then necessary to compute the TMO results. In [5], a is set to 18% and if L_{white} is set to infinity, then the term $1 + \frac{L_s}{L_{white}^2}$ has no effect, it can hence be removed.

The main issue with this algorithm is that flickering artifacts appear for abrupt changes of the key value. To avoid these artifacts, Kang et al. [10] proposed to filter the current frame's key value using a set of the n previous frames key values. Ramsey et al. [6] further refined this idea by making n adaptive. This adaptation discards outliers using a min/max threshold. We refer to this technique as $PTR{+}R$ hereinafter. Some other solutions [11] also reduce flickering artifacts using temporal close frames. However, none of these solutions preserve the temporal brightness coherency.

2.2. Brightness Coherency for Video Tone Mapping

To better preserve the temporal brightness coherency, Boitard et al. [7] developed a method which adapts to any TMO through a post-processing operation. We refer to this technique as BC hereinafter. This technique uses the frame key value κ (computed with equation 2) to preserve the HDR brightness ratio in the tone map LDR sequence.

The HDR brightness ratio is equal to the LDR brightness ratio if:

$$\frac{\kappa_F^{HDR}}{\kappa_V^{HDR}} = \frac{\kappa_F^{LDR}}{\kappa_V^{LDR}} \, , \tag{4}$$

where κ_F^{HDR} is the current HDR frame key value and κ_V^{HDR} the key value of the brightest frame. To satisfy Equation (4), the tone map luma L_d is scaled according to Equation (5) to get the new tone map luma L'_d:

$$L'_d = L_d \cdot \left(Of + (1 - Of) \cdot \frac{\kappa_F^{HDR} \cdot \kappa_V^{LDR}}{\kappa_V^{HDR} \cdot \kappa_F^{LDR}}\right) = R \cdot L_d \tag{5}$$

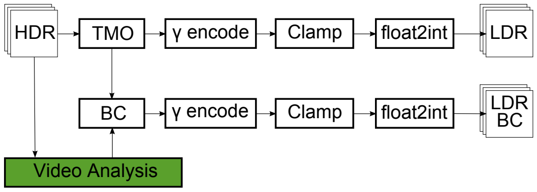

Fig. 1: General workflow for tone mapping a video sequence with and without the BC algorithm. The video analysis (green box) is performed as a preprocessing step. "γ encode" is required to compensate for the use of the bits relative to how humans perceive light (usually $\gamma = 1/2.2$). "Clamp" removes the values outside the range [0, 1]. "float2int" converts floating point values into integers.

where κ_F is computed using Equation (2), L_d and L'_d are respectively the tone map LDR luma and the post-processed LDR luma. Of is a user-defined parameter to prevent low scale ratio. In order to determine the anchor, i.e. the frame with the maximum HDR frame key value, a video analysis is performed prior to the tone mapping operation. Figure 1 depicts the workflow used to tone map a video with and without preserving the brightness coherency. This solution preserves the relative HDR brightness levels in the LDR tone map results.

2.3. High Efficiency Video Codec

HEVC is the successor of the ITU-T H.264 / MPEG-4 Part 10 'Advanced Video Coding' (AVC) codec. Developed by the Joint Collaborative Team on Video Coding (JCT-VC) group, it was released in January 2013 and is reported to double AVC compression ratio. The HEVC test Model (HM) is currently in its version 10.

HEVC is a block-based codec that exploits both spatial and temporal correlations between the code values of the pixels to achieve a high compression ratio. To exploit the spatial and temporal correlation, blocks are predicted using two processes. Intra-prediction relies on spatial correlation to predict the current block using blocks already decoded in the current frame. Inter-prediction exploits the temporal correlation by predicting the current block using blocks from a set of previous/subsequent decoded frames. The predicted block is then subtracted from the original block leaving only the residuals to be encoded. In video compression, it is generally considered that the closest the prediction, the more efficient the compression.

As this paper focuses on temporal coherency, a detailed explanation of the inter-prediction is given hereafter. To predict the current block, a block-based motion estimation is performed to find the best temporal prediction. It consists in finding a block that minimizes the distortion with the current block to be encoded. An example of distortion metrics is the Sum of Absolute Differences (SAD) or the Mean Square Er-

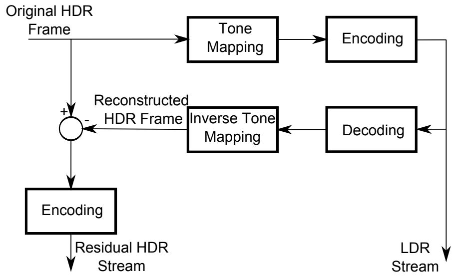

Fig. 2: Example of the workflow of a backward compatible scalable HDR video codec.

ror (MSE).

Inter-prediction is the best way of removing redundant information in a video signal. As shown later on, preserving the temporal coherency, when tone mapping video sequences, improves the motion estimation efficiency and consequently the inter-prediction.

2.4. Backward Compatible HDR Video Codec

If compressing LDR contents is a mandatory operation for storage and transmission, it is even more important for HDR contents. Indeed, current HDR representation formats usually require two to four times the size needed by LDR contents. In addition, due to the small amount of available HDR displays, backward compatibility with legacy display is required. That is why, techniques known as scalable backward compatible have been designed [12]. These techniques follow the workflow depicted in figure 2.

However, these techniques optimize the pair TMO / iTMO (inverse Tone Mapping Operator) to obtain the best reconstruction. The resulting tone map curve is computed so as to preserve the maximum level of information on a per frame basis, usually at the expense of the LDR sequence quality.

3. TEMPORAL COHERENCY AND VIDEO COMPRESSION

Subsequent frames in video sequences are highly correlated. Tone mapping frames individually disrupts the temporal coherency, which reduces this correlation and consequently the compression ratio. That is why, it is essential to devise video TMOs that preserve temporal coherency.

In this section, we study temporal coherency and video compression according to three aspects. First, we analyze the impact that temporal coherency has on motion estimation. Second, we evaluate the quality of a tone mapped video sequence when compressed by a codec. Finally, we assess the efficiency of these TMOs after having reconstructed an HDR video sequence.

TMO	D = 1	D = 2	D = 4	D = 8
PTR	32.85	29.37	26.37	23.35
PTR+BC	38.37	34.91	31.89	28.73
PTR+R	32.84	29.36	26.35	23.38
PTR+R+BC	38.33	34.87	31.87	28.69

Table 1: PSNR of the inter-predicted frame resulting from motion estimation. D represents the distance (in display order) between the reference and the current frame.

3.1. Temporal Coherency and Inter-Prediction

In section 2, we mentioned two issues when tone mapping video sequences. First, flickering artifacts can appear due to abrupt changes of the tone map curve. Second, HDR temporal brightness coherency is not preserved. Both these issues impact the motion estimation performed in a codec. In LDR video compression, the quality of the motion estimation, and hence the inter-prediction, is evaluated by computing the Peak Signal to Noise Ratio (PSNR) between the predicted frame and the original one.

For our study, we tone mapped an HDR sequence of 250 1920x1080 High Definition (HD) frames (named $UnderBridgeHigh$) using 4 different TMOs (PTR, $PTR+BC$, $PTR+R$ and $PTR+R+BC$). We evaluated the performance of each TMO in regard to inter-prediction by analyzing the predicted frame resulting from the motion estimation. The motion estimation was performed at temporal distances corresponding to the hierarchical Group Of Frame (GOF) structure specified in the main profile of HEVC [8]. Table 1 reports the average PSNR of the predicted frame at temporal distances 1, 2, 4 and 8 for a block-size of 8. Results show that preserving the brightness coherency ($PTR+BC$ and $PTR+R+BC$) increases the quality of the predicted frame. Note that reducing flickering artifacts ($PTR+R$) does not improve the inter-prediction for the used video sequences.

It is usually reckoned in video compression that improving the prediction results in a higher compression efficiency as demonstrated in the next section.

3.2. Video Tone Mapping and Compression

Recall that one of the main objective of TMOs is to achieve the best subjective quality. However, when tone mapped contents are compressed using a codec, this quality is degraded. That is why, this section addresses the evaluation of the quality of decoded tone mapped LDR sequences.

For our evaluation, we compare the same TMOs that we used in the previous section. We used two sequences, the $UnderBridgeHigh$ sequence mentioned previously as well as the $Tunnel$ sequence [13]. Figure 3 and 4 plots the rate-distortion curves obtained using the main profile of the HM 10.0 [4]. For additional results and analysis of the sequences, refer to [14]. As expected, the better prediction obtained with

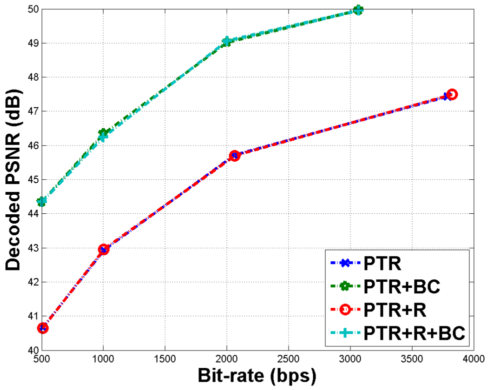

Fig. 3: Rate-distortion curve of the$UnderBridgeHigh$ sequence (HD resolution, 1920x1080) for targeted bit-rates of 500 kilo bits per second (kbps), 1, 2 and 4 Mega bps.

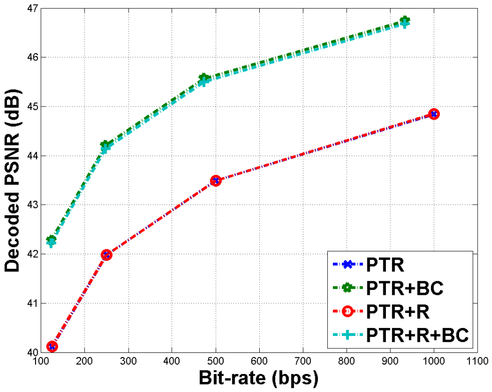

Fig. 4: Rate-distortion curve of the $Tunnel$ sequence (VGA resolution, 640x480) for targeted bit-rates of 125, 250, 500 and 1000 kbps.

the BC technique results in a PSNR gain ranging from 2 to 4 dB.

When the goal is to preserve the quality of an HDR video sequence, the original contents are distorted by the TMO and the used codec. In the next section, we evaluate the quality of a reconstructed HDR video sequence depending on the used TMO with or without applying a codec.

3.3. HDR Video Reconstruction

Recall that scalable backward compatible HDR video codecs optimize the pair TMO / iTMO so as to obtain the best HDR reconstruction. In other words, they focus on the quality of the reconstructed HDR video. Here, we propose a different approach which consists in using TMOs, known to produce good quality LDR video sequence, and assess the quality of a reconstructed HDR video sequence.

We use an inverse Tone Mapping Operator (iTMO) that reconstructs an HDR video sequence from the LDR video sequence or the decoded LDR video sequence (fig. 5). As all the used TMOs are based on the PTR, we apply the same iTMO to reconstruct an HDR video sequence. The different steps of the iTMO are explained as follows:

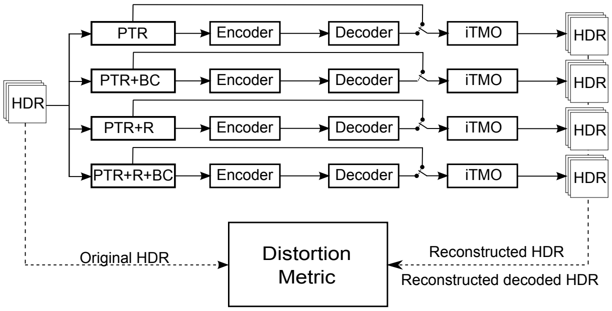

Fig. 5: Proposed workflow to evaluate the quality of the different tone mapped sequences. In this article, the distortion metric is the MSE computed in the log domain.

- convert 8 bits integer to floating point values,
- remove γ encoding $RGB_{Lin} = RGB^{2.2}$,
- compute the luma L'_d,
- invert equation 5 (BC only): $L_d = \frac{L'_d}{R}$,
- invert equation 3 (sigmoid): $Ls = \frac{L_d}{1-L_d}$,
- invert equation 1 (scaling): $L_w = L_s \cdot \frac{\kappa}{\alpha}$.

RGB is the LDR frame, α the same parameter as in equation 1 and κ the key value of either the original HDR frame for the PTR or the temporally filtered key value for the $PTR + R$. To reconstruct the HDR frame, we need the key value computed using equation 2 on the decoder side. We use an HEVC message, namely the tone mapping Supplemental Enhancement Information (SEI) message, which encodes this value [8]. In the case of the BC technique, we also need to encode the scale ratio R used per frame.

To evaluate the impact of both the TMO and the codec, we reconstruct two HDR video sequences, one before applying the codec and another after. Figure 5 represents the workflow of the technique. Table 2 and 3 reports the average distortion (MSE in the log domain) between a reconstructed HDR sequence and the original one. Results show that using the BC technique reduces the quality of the reconstructed HDR sequence even without applying the codec. This can be explained as follows. TMOs assign several HDR values to a same LDR value. This results in a loss of information regarding the HDR sequence. Contrary to other TMOs that use the full available range, the BC technique assigns a variable and reduced range to each frame to preserve the brightness coherency. As a result, a BC-based reconstructed HDR video

TMO	500 kbps	2 Mbps	4 Mbps	Original
PTR	15.34	13.55	11.04	3.28
PTR+BC	27.65	21.79	18.82	4.00
PTR+R	15.39	13.63	11.12	3.38
PTR+R+BC	25.51	21.80	18.81	4.18

Table 2: Average distortion for the $UnderBridgeHigh$ sequence using the MSE metrics in the log domain (for clarity purposes, all values are multiplied by 100).

TMO	125 kbps	500 kbps	1 Mbps	Original
PTR	9.55	7.25	6.66	2.33
PTR+BC	11.30	8.50	7.94	2.71
PTR+R	9.54	7.26	6.68	2.29
PTR+R+BC	11.25	8.48	7.94	2.70

Table 3: Average distortion for the $Tunnel$ sequence using the MSE metrics in the log domain (for clarity purposes, all values are multiplied by 100).

sequence is more distorted due to these reduced ranges. Finally, results show that compression entails a higher distortion than TMOs. See [14] for additional results on the HDR reconstruction.

4. CONCLUSION

In this paper, we studied the relation between the compression of tone mapped video contents and temporal coherency. Two qualities were evaluated in regard to video compression: the LDR video quality and the HDR reconstructed quality. In addition, two types of temporal coherency have been studied: flickering artifacts ($PTR + R$) and temporal brightness coherency (BC).

Results show that the $PTR + R$ technique has no effect either on the quality of the LDR video or on the reconstructed HDR video sequence. On the contrary, the BC-based techniques preserve temporal brightness coherency while achieving a higher compression efficiency. However, they distort the reconstructed HDR video sequence more than the other methods. When the objective is video tone mapping of high quality, the BC technique is preferred. When a reconstructed HDR video sequence is needed, the other TMOs are to be used.

Subjective evaluation must be conducted to validate all the results presented in this paper. In addition, further work should be done on tone mapping using higher bit-depth (10 to 14 bits).

5. REFERENCES

[1] Michael D. Tocci, Chris Kiser, Nora Tocci, and Pradeep Sen, "A versatile HDR video production system," in *ACM SIGGRAPH 2011 papers on - SIGGRAPH '11*. 2011, ACM Press.

[2] Patrick Ledda, Alan Chalmers, Tom Troscianko, and Helge Seetzen, "Evaluation of tone mapping operators using a high dynamic range display," in *ACM SIGGRAPH 2005 Papers*, New York, NY, USA, 2005, SIGGRAPH '05, pp. 640–648, ACM.

[3] Akiko Yoshida, "Perceptual evaluation of tone mapping operators with real-world scenes," in *Proceedings of SPIE*. 2005, vol. 5666, pp. 192–203, SPIE.

[4] Gary J. Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[5] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, July 2002.

[6] SD Ramsey, JT Johnson III, and Charles Hansen, "Adaptive temporal tone mapping," *Computer Graphics and Imaging - 2004*, , no. 3, pp. 3–7, 2004.

[7] Ronan Boitard, Kadi Bouatouch, Remi Cozot, Dominique Thoreau, and Adrien Gruson, "Temporal coherency for video tone mapping," in *Proc. SPIE 8499, Applications of Digital Image Processing XXXV*, 2012.

[8] J.-R. Ohm G. J. Sullivan Y.-K. Wang T. Wiegand B. Bross, W.-J. Han, "High efficiency video coding (hevc) text specification draft 10," ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC) document JCTVC-L1003, Jan 2013.

[9] Ansel Adams, *The Print: The Ansel Adams Photography Series 3*, Little, Brown and Compagny, 1981.

[10] Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski, "High dynamic range video," *ACM Trans. Graph.*, vol. 22, no. 3, July 2003.

[11] Rafal Mantiuk, Scott Daly, and Louis Kerofsky, "Display adaptive tone mapping," *ACM Transactions on Graphics*, vol. 27, no. 3, pp. 1, Aug. 2008.

[12] Zicong Mai, Hassan Mansour, Rafal Mantiuk, Panos Nasiopoulos, Rabab Ward, and Wolfgang Heidrich, "Optimizing a tone curve for backward-compatible high dynamic range image and video compression.," *IEEE transactions on image processing*, June 2011.

[13] G. Krawczyk, "Tunnel video sequence http://www.mpi-inf.mpg.de/resources/hdr/video/," .

[14] More results, "http://people.irisa.fr/Ronan.Boitard/shared/Eusipco2013.pdf," .