RATE-DISTORTION OPTIMIZATION FOR HIGH DYNAMIC RANGE VIDEO CODING

Maryam Azimi, Mahsa T. Pourazad, Panos Nasiopoulos

Student Member, IEEE, Member, IEEE, Senior Member, IEEE

ABSTRACT

The genuine viewing quality of High Dynamic Range (HDR) data brought an evolution to content production and display manufacture markets. However, the efficiency of compression and transmission of HDR content still needs to be improved. In this paper we modify the H.264/AVC standard to better characterize HDR content. We propose a new Lagrangian multiplier that strikes a balance between the bit-rate and distortion of the HDR video. The updated Lagrange multiplier was implemented on the H.264/AVC reference software. Our experiment results show that the HDR-VDP2 quality scores of the videos encoded by the HDR-accustomed encoder are higher than the ones encoded with the reference encoder. Moreover, subjective tests confirmed that the visual quality of the compressed HDR videos using our proposed method is higher than the one encoded by the reference codec[1].

Index Terms— High Dynamic Range, HDR, Low Dynamic Range, video encoder, subjective evaluation, Lagrangian multiplier

1. INTRODUCTION

High Dynamic Range (HDR) video content depicts the captured scene with color and brightness range closer to the natural scene characteristics than those found in Low Dynamic Range (LDR) content. While content producers and display manufacturers are recently employing HDR technology in an effort to provide true-life viewing experience, compression and transmission of HDR content still remain a challenge. HDR content can be represented in high-bit-depth data (10 to 16 bits per color component). Although the existing LDR video coding standards, such as H.264/AV [1] and High Efficiency Video Coding Standard (HEVC) [2, 3], support high bit depth data, they are not optimized for coding HDR data.

In general, the coding process introduces compression distortions to the video as a result of the block-based prediction and motion estimation procedures. To minimize the ultimate distortion of the coded stream so that its bitrate does not exceed a maximum bitrate setting, the encoder employs rate-distortion optimization (RDO). The RDO schemes employed by the existing standards use LDR quality metrics, such as PSNR and Structural Similarity (SSIM) [4], to measure the distortion during the coding process. Both SSIM and PSNR are LDR-driven quality metrics in the sense that they are designed to assess the quality of the LDR data and as such their performance on HDR content is not guaranteed (see [5] for more details). Accordingly using the RDO schemes, which are designed based on LDR quality metrics for coding HDR content, does not necessarily create a balance between distortion and bitrate.

One approach to address this problem is to adapt the HDR content to LDR format so that the application of LDR metrics on HDR content becomes possible. In this regard, the researchers in [6] propose to employ a perceptually uniform (PU) encoding method. Another efficient technique in [7] uses a multi-exposure method to preprocess the HDR content. Based on this method, HDR data is tone mapped at different exposures and is uniformly distributed over the dynamic range of the data. Then the LDR Quality metric is employed on each exposure of the data, which is essentially an LDR version of the content. The average of the quality of all the exposures is the final HDR data quality by the LDR metric. Unfortunately the correlation between the results of LDR metrics and the subjective evaluations of the HDR preprocessed by [6] and [7] is very low [5]. Ideally, the best compression performance will be achieved by optimizing rate distortion using an HDR quality metric.

In this paper we propose a new Lagrangian multiplier for the RDO process that strikes a balance between the bit-rate and distortion of the HDR content measured by an HDR quality metric. The proposed Lagrange multiplier is implemented in the H.264/AVC reference encoder software. To evaluate the efficiency of the proposed scheme, the performance of the HDR-accustomed encoder is compared with that of reference H.264/AVC encoder over a HDR data set. Moreover, we perform subjective tests to evaluate the quality of encoded HDR streams using a prototype HDR TV display.

The rest of the paper is organized as follows: Section 2 elaborates on the methodology, Section 3 describes the subjective test procedure, the experiment results are reported

[1] This work was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC), the Institute for Computing, Information and Cognitive Systems (ICICS) at UBC, and TELUS.

and discussed in Section 4, and conclusions are drawn in Section 5.

2. METHODOLOGY

The first step in our method is choosing an appropriate HDR quality metric as the quality predictor of the final encoded video. Among the existing HDR metrics we chose the HDR-VDP-2 quality metric, which yields the best performance when measuring compression distortion (highest correlation with subjective quality scores) [5, 8]. HDR-VDP-2 is a full-reference quality metric, which works with higher level of luminance in images. Its output is a quality score for the distorted HDR image between 0 and 100, where 0 represents the lowest quality and 100 represents equal quality with respect to the reference HDR image.

The next step in our scheme is updating the rate-distortion optimization process of H.264/AVC for HDR content. The rate-distortion optimization is a constrained problem, where the ultimate distortion of the coded stream is minimized such that its bitrate does not exceed a maximum bitrate (R_{max}). The RDO problem can be formulated as follows:

$$\min\{D\}, \quad \text{subject to } R \leq R_{max} \tag{1}$$

where D represents the distortion of the coded video and R is its bitrate. As suggested by [9], using the Lagrangian multiplier optimization technique the optimization equation (1) can be converted to:

$$\min\{J\}, \quad \text{where } J = D + \lambda R \tag{2}$$

where λ is the Lagrangian multiplier that strikes a balance between distortion and optimization. By finding the optimal λ, it is assured that the final quality of the encoded video is the best achievable, i.e., it has the least distortion given the rate constraint R_{max}. As shown in [9] and [10] the Lagrangian multiplier can be estimated as (see [9] and [10] for details):

$$\lambda = -\frac{dD}{dR} = c.Q^2 , \tag{3}$$

where Q is the quantization step-size and c is a constant parameter. As it could be observed from (3), λ is dependent on Q. The optimal λ that guarantees minimum distortion for LDR sequences in terms of PSNR has been experimentally calculated as follows [11]:

$$\lambda = 0.85 \times 2^{(\frac{QP}{3} - 4)} \tag{4}$$

This equation is being used in H.264/AVC to balance the relationship between PSNR content distortion and bitrate. To achieve high compression efficiency for HDR content, it is necessary to design of a new Lagrangian multiplier, which utilizes an HDR quality metric.

To find the optimal Lagrangian multiplier (lambda) for HDR video content, the relationship between the quantization parameter (QP) and λ has to be derived for HDR content. To this end, we encode four HDR video sequences using H.264/AVC reference software (JM18.4) with a set of combinations of different lambdas and QPs. The video sequences are chosen from the DML-HDR video dataset [12]. Each sequence is 10 seconds long, 12-bit YUV, and 2K resolution (2048x1080). Three of these videos, namely “Christmas Tree”, “Hallway”, and “Table” are indoor and the other one “Playground” is outdoor. The QP values used in our study range from 24 to 48 (minimum and maximum QP allowed in the JM software for 12-bit data [13]) with a step size of two. The λ values used are 4, 25, 100, 250, 400, 730, and 1000. These λ values have been suggested in [4] for deriving the λ-QP relationship for LDR video content. All four HDR video sequences were encoded with the set of total 91 combinations of λ and QP (13 QP times 7 λ values) . The H.264/AVC encoder settings are as follows: hierarchical B pictures with 3 levels, group of pictures (GOP) size of 8, luma and chroma bit-depth of 12, input video format of YUV 4:2:0 progressive enabled CABAC entropy coding.

Once the compressed videos are obtained, their distortion is calculated based on HDR-VDP-2. At each rate, the λ-QP combination that yields minimum distortion, i.e., the highest HDR-VDP-2 score, is selected as the optimal combination. Fig. 1 depicts the optimal λ and QP relationship for all the four sequences.

It can be observed from Fig.1 that a non-linear relationship holds between QP and λ, where increasing QP results in greater λ. Based on [9] we assume that the approximation function has a similar format as (4). The fitted curve based on (4) results in :

$$\lambda = 0.6203 \times 2^{(0.3492QP - 5.8878)} \tag{5}$$

The Pearson Linear Correlation Coefficient (PCC) and the Spearman Rank Order Correlation Coefficient (SCC) between the approximated function and the actual tested data are 0.92 and 0.89 respectively, which confirms the accuracy and the monotonicity of the approximation function.

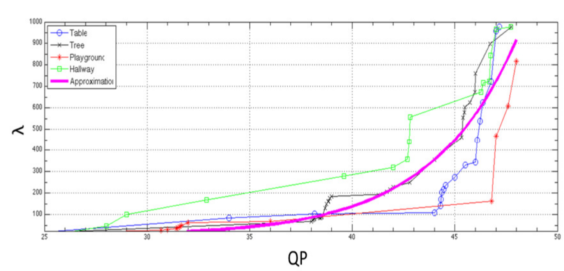

Figure 1. λ-QP relationship

The derived formula was employed in the H.264/AVC reference software (JM18.4) and the related parameters were updated inside the encoder. To evaluate the performance of our scheme, the HDR video sequences were encoded using the modified encoder with the same configuration as the one used for estimating λ. The QP settings used were 28, 32, 36 and 40. Subjective tests were performed to compare the quality of videos coded by the reference encoder with the ones compressed with the modified encoder. The details of our subjective tests are provided in the next section.

3. SUBJECTIVE TESTS

The subjective test procedures were conducted in a room complying with the Recommendation ITU-R BT.500-13 [14]. The tests were designed based on the Double-stimulus continuous quality-scale (DSCQS) method suggested in [14]. In particular, a pair of reference video (uncompressed) and a test video (compressed by the reference or modified encoder) was shown to the subjects consecutively. Subjects were asked to rate the overall perceived quality of both videos. The quality of the reference and the test videos were rated separately on a continuous grading scale ranging from Bad to Excellent and the difference between the tested and the reference sequences' rating was used. Subjects were familiarized with the rating procedure by completing a short training session before the actual test.

The tests lasted approximately 15 minutes and consisted of 36 test videos: 4 original, 16 test sequences encoded by reference H.264/AVC software and 16 encoded by the updated H.264/AVC software. Original videos were inserted as the test videos during the session to examine the subjects' reliability. No outliers were detected in our test (according to the recommendations in [14]). The Mean Opinion Score (MOS) of the subjects for each impaired video was calculated by averaging the scores over the subjects with 95% confidence interval.

The videos were displayed on an HDR TV prototype that consists of a projector with the resolution of 1024×768 and a 40 inch full HD LCD placed in the front. The contrast range of the projector is 2000:1 and is capable of emitting 15000 ANSI lumens. HDR viewing experience is achieved by observing a color stream sent from the LCD screen calibrated on top of a luminance stream from the projector.

18 subjects including 11 female and 7 male participated in the study. They were screened for color blindness and visual acuity by Ishihara and Snellen charts, respectively, prior to the test. Three subjects at a time participated in the test session. The age of the subjects ranged from 21 to 30 years old. None of the subjects were aware of the aim of the test and they were all considered non-experts in the field.

4. RESULTS AND DISCUSSIONS

Figs. 2 (a), (b), (c), and (d) depict the quality of the encoded videos at different bitrates based on the Mean Opinion Score

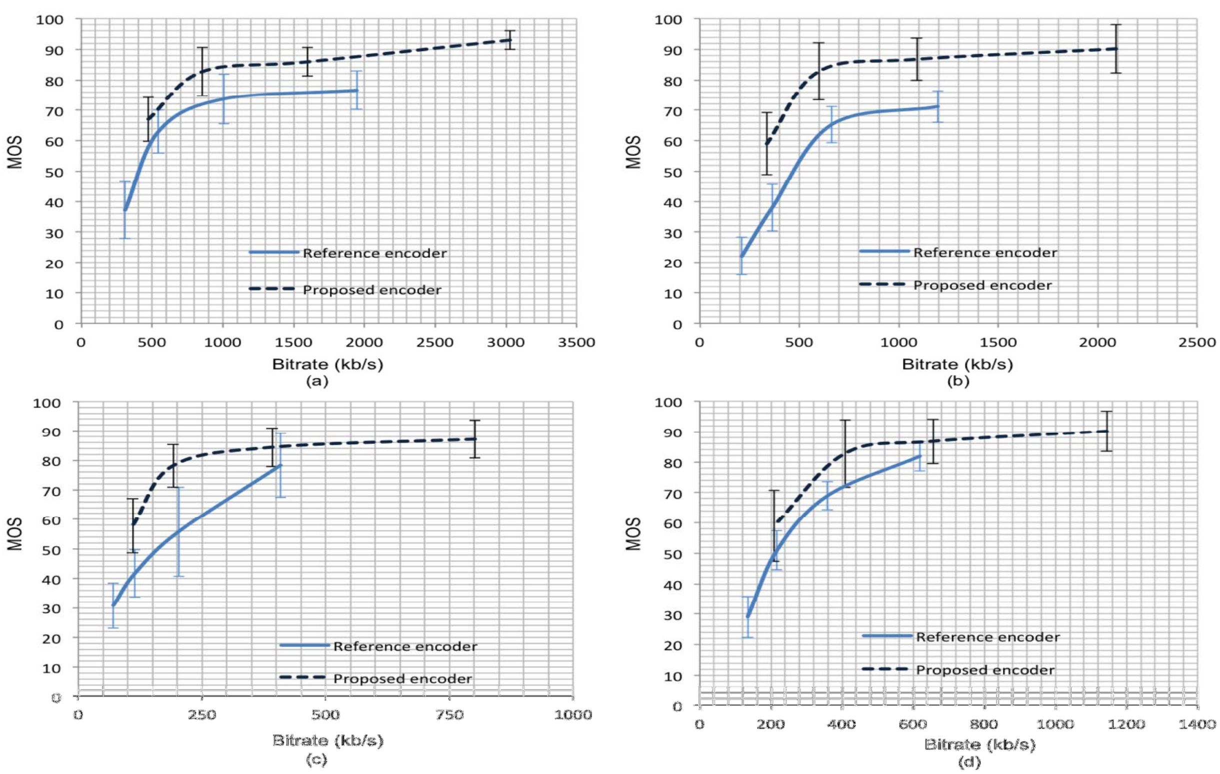

Figure 2. MOS-bitrate comparison of the proposed and reference encoder on (a) Playground, (b) Christmas Tree, (c) Hallway, and (d) Table.

(MOS) of the viewers. As it can be observed, for the same bitrate, videos encoded by our proposed method achieved significantly higher MOS quality score compared to ones encoded by the reference encoder.

For the same MOS, our proposed encoder yields bitrate savings of 39.17% for the “Hallway”, 35.98% for the “Playground”, 27.02 % for the “Christmas Tree”, and 16.03% of for the “Table” sequence compared to the ones encoded by the reference encoder. On the other hand, for the same bitrate levels, the improvement in MOS was 19.22 for “Hallway”, 11.33 for “Playground”, 14.22 for “Christmas Tree”, and 16.85 for “Table”.

In summary, the above results confirm that changing the lambda to address the HDR content characteristics significantly improves the coding efficiency of the encoder when applied to HDR content.

4. CONCLUSIONS

In this study we proposed an updated Lagrangian multiplier for HDR video content. The updated Lagrange multiplier is shown to have a non-linear dependency on QP values (similar to the one for LDR videos yet with different constants). The proposed Lagrange multiplier was implemented in the H.264/AVC reference software (JM 18.4). The subjective tests confirmed that our proposed HDR-accustomed encoder is significantly more efficient compared to the reference encoder when encoding HDR content.

5. REFERENCES

[1] Wiegand, T.; Sullivan, G.J.; Bjontegaard, G.; Luthra, A., "Overview of the H.264/AVC video coding standard," Circuits and Systems for Video Technology, IEEE Transactions on, vol.13, no.7, pp.560,576, July 2003.

[2] Pourazad, M.T.; Doutre, C.; Azimi, M.; Nasiopoulos, P., "HEVC: The New Gold Standard for Video Compression: How Does HEVC Compare with H.264/AVC," *Consumer Electronics Magazine, IEEE*, vol.1, no.3, pp.36-46, July 2012.

[3] Sullivan, G.J.; Ohm, J.; Woo-Jin Han; Wiegand, T., "Overview of the High Efficiency Video Coding (HEVC) Standard," Circuits and Systems for Video Technology, IEEE Transactions on , vol.22, no.12, pp.1649,1668, Dec. 2012

[4] Zhou Wang; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P., "Image quality assessment: from error visibility to structural similarity," Image Processing, IEEE Transactions on , vol.13, no.4, pp.600,612, April 2004.

[5] Azimi, M.; Banitalebi, A.; Dong. Y.; Pourazad, M.T.; Nasiopoulos, P., “A survey on the performance of the existing full reference HDR video quality metrics: A new HDR video dataset for quality evaluation purposes,” Submitted to Consumer Electronics Magazine, IEEE, February 2014.

[6] T. O. Aydin, R. Mantiuk, and H. P. Seidel, “Extending quality metrics to full luminance range images,” SPIE, Human Vision and Electronic Imaging XIII, San Jose, USA, Jan. 2008.

[7] J. Munkberg, P. Clarberg, J. Hasselgren, and T. Akenine-Möller, “High dynamic range texture compression for graphics hardware,” ACM Transactions on Graphics, vol. 25, no. 3, p. 698, Jul. 2006.

[8] Rafat Mantiuk, Kil Joong Kim, Allan G. Rempel, and Wolfgang Heidrich. 2011. “HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions”. ACM Trans. Graph. 30, 4, Article 40, 14 pages, July 2011.

[9] Sullivan, G.J.; Wiegand, T., "Rate-distortion optimization for video compression," Signal Processing Magazine, IEEE, vol.15, no.6, pp.74-90, Nov 1998.

[10] Wiegand, T.; Girod, B., "Lagrange multiplier selection in hybrid video coder control,"Image Processing, 2001. Proceedings. 2001 International Conference on, vol.3, no., pp.542-545, 2001.

[11] Wiegand, T.; Schwarz, H.; Joch, A.; Kossentini, F.; Sullivan, G.J., "Rate-constrained coder control and comparison of video coding standards," Circuits and Systems for Video Technology, IEEE Transactions on , vol.13, no.7, pp.688,703, July 2003.

[12] Digital Media Lab High Dynamic Range (DML-HDR) video dataset created at the University of British Columbia, available at: http://dml.ece.ubc.ca/data/DML-HDR/

[13] http://iphome.hhi.de/suehring/tml/

[14] Recommendation ITU-R BT.500-13, “Methodology for the subjective assessment of the quality of television pictures,” ITU, 2012.