RATE-DISTORTION OPTIMAL BIT ALLOCATION FOR HIGH DYNAMIC RANGE VIDEO COMPRESSION

Chul Lee and Chang-Su Kim

School of Electrical Engineering, Korea University, Seoul, Korea
E-mails: kayne@korea.ac.kr, changsukim@korea.ac.kr

ABSTRACT

An efficient algorithm to compress high dynamic range (HDR) videos is proposed in this work. We separate an HDR video sequence into a tone-mapped low dynamic range (LDR) sequence and a ratio sequence. Then, we encode those two sequences using the standard H.264/AVC codec. During the encoding, we allocate a limited amount of bit budget to the LDR sequence and the ratio sequence adaptively to maximize the qualities of both the LDR and HDR sequences. While a conventional LDR decoder uses only the LDR stream, an HDR decoder can reconstruct the HDR video using the LDR stream and the ratio stream. Simulation results demonstrate that the proposed algorithm provides higher performance than the conventional methods.

Keywords: High dynamic range (HDR) video, tone mapping, H.264/AVC, rate-distortion optimization

1. INTRODUCTION

The dynamic range of a digital image is referred to as the ratio of the intensities between the brightest pixel and the darkest pixel. While the dynamic ranges of conventional display devices are less than two orders of magnitude, real world scenes have much higher dynamic ranges. Also, human eyes can perceive more than six orders of magnitude. Therefore conventional devices cannot display realistic scenes which correspond with the human vision. In order to overcome this limitation, a lot of efforts have been made in the research from acquisition to display images which can handle the full dynamic range. Such images that have higher dynamic ranges than conventional devices are called high dynamic range (HDR) images [1].

HDR imaging technologies are already adopted in some industries such as games, movies and surveillance systems [2]. It is expected that this trend would continue and HDR imaging would be used in general applications in near future. However, in the intermediate stage, HDR images should be displayed on conventional display devices. In this case, the conventional low dynamic range (LDR) devices cannot display HDR images due to their limited dynamic range and conversion is required. This conversion from HDR to LDR is called tone mapping, and various algorithms have been proposed [1].

Due to their higher dynamic range, HDR images and videos require a huge amount of storage space and transmission bandwidth compared with conventional LDR data. However, only a few algorithms have been proposed to compress HDR data. In [3], Ward *et al.* proposed a backward-compatible compression algorithm for HDR images. Their algorithm encodes the tone-mapped LDR image and the ratio image, which represents ratio between the HDR and LDR images, using the standard JPEG codec. In [4], Okuda *et al.* improved the coding performance by adopting an inverse tone mapping function to predict the HDR images from tone-mapped LDR images, and used a wavelet encoder to compress ratio images.

While recent video coding standards support the encoding of videos with an extended bit depth up to 12 bits, they cannot reconstruct full dynamic ranges of HDR videos faithfully. So, to address this limitation, Mantiuk *et al.* [5] proposed a luminance quantization method, which is optimized for the human perception. Their algorithm compresses the quantized videos, using the MPEG-4 codec, with an additional scheme to transform high contrast blocks. In [6], Mantiuk *et al.* extended the work in [5] to offer the backward compatibility with LDR devices. Some recent approaches tried to add the HDR ability to the existing video coding standard as a bit-depth scalable video coding (SVC) [7].

In this work, we propose a novel video compression algorithm for HDR video sequences based on rate-distortion (R-D) optimization. The proposed algorithm separates an HDR video into an LDR sequence and a ratio sequence, and encodes them using the state-of-the-art H.264/AVC codec [8]. Whereas conventional methods focus on maximizing quality of the reconstructed HDR sequence, we develop a rate-distortion optimized scheme to allocate a limited bit budget to the LDR and ratio sequences efficiently to maximize the qualities of the reconstructed LDR and HDR sequences.

2. TONE MAPPING

The proposed algorithm uses the gradient domain tone mapping for high dynamic range videos proposed in [9]. It obtains temporally coherent high quality LDR sequence so as

---

This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MEST) (No. R01-2008-000-20292-0)

to improve the coding performance.

2.1. Gradient Domain Tone Mapping

The gradient domain tone mapping, which was proposed in [10], is based on an assumption that the dynamic range of an image can be compressed by reducing gradients.

The modified gradient map $\nabla g(x, y)$ can be obtained by multiplying the input HDR image gradient $\nabla h(x, y)$ by the attenuation function $\Phi(x, y)$, given by

$$g(x, y) = \nabla h(x, y)\Phi(x, y) \tag{1}$$

The LDR image l, which has closest gradients to $g(x, y)$, can be obtained by minimizing a cost function, given by

$$\begin{aligned} C_1 &= \sum_{x,y} \|\nabla l(x, y) - g(x, y)\|^2 \\ &= \sum_{x,y} \big\{[l(x+1, y) - l(x, y) - g_x(x, y)]^2 \\ &\quad + [l(x, y+1) - l(x, y) - g_y(x, y)]^2\big\} \end{aligned} \tag{2}$$

where g_x and g_y stand for the x and y components of the gradient, respectively. We differentiate C_1 respect to $l(x, y)$ and set it to 0, which yields the equation

$$\begin{aligned} & l(x+1, y) + l(x-1, y) + l(x, y+1) + l(x, y-1) - 4l(x, y) \\ &= \; g_x(x, y) - g_x(x-1, y) + g_y(x, y) - g_y(x, y-1). \end{aligned}$$

The equation is a discrete approximation of the Poisson equation and can be solved by a numerical algorithm.

2.2. Tone Mapping of HDR Videos

If the tone mapping scheme is applied to each frame of a video sequence, it produces flickering artifacts since it doesn't take temporal coherency into account. To alleviate these artifacts, Lee and Kim proposed a video tone mapping algorithm, which obtains high quality temporally coherent LDR sequence by exploiting motion information [9].

The Lee and Kim's algorithm first finds pixelwise motion vector field between successive frames $h(x, y, t-1)$ and $h(x, y, t)$. Then, they assume that the HDR and LDR sequences have the same motion vectors since they represent the same objects, and let $p(x, y)$ denote the pixel of the previous LDR frame, specified by the motion vector. Therefore, in addition to the original cost function in (2), a new cost function is introduced, given by

$$C_2 = \sum_{x,y} \big\{l(x, y, t) - p(x, y)\big\}^2. \tag{3}$$

Then, the overall cost function can be defined as a weighted sum of the two costs,

$$C = C_1 + \lambda C_2, \tag{4}$$

where, λ is a weighting coefficient which controls relative importance between the two costs.

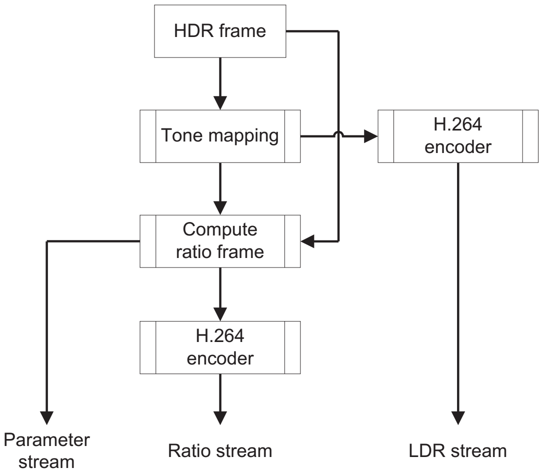


Fig. 1. The complete block diagram of the proposed HDR video encoder.

By differentiating C respect to $l(x, y, t)$ and setting it 0, we have the solution

$$\begin{aligned} & l(x+1, y) + l(x-1, y) + l(x, y+1) \\ &\quad + l(x, y-1) - (4+\lambda)l(x, y) \\ &= \; g_x(x, y) - g_x(x-1, y) \\ &\quad + g_y(x, y) - g_y(x, y-1) - \lambda p(x, y), \end{aligned} \tag{5}$$

where the time index t in $l(x, y, t)$ is omitted for simpler notations. The set of equations can be solved using a numerical method [9].

3. HDR VIDEO COMPRESSION

Fig. 1 shows the complete block diagram of the proposed HDR video encoder. The encoder takes HDR frames as input, and convert them into LDR frames using the tone mapping operator in Section 2. We encode the LDR frames with the standard H.264/AVC encoder, which offers backward compatibility. Then, the LDR frames are compared with the original HDR frames to yield ratio frames, which are also encoded using the H.264/AVC. A set of parameters are used in this process and transmitted to the decoder as side information.

3.1. Ratio Frames

We separate an HDR frame into an LDR frame and residual data that represents the differences between the HDR and LDR frames using tone mapping operator. In previous works, the residual data represents the ratios, obtained by dividing the HDR value by the LDR value at each pixel position [3], or the prediction error between the original HDR value and predicted value from the LDR value [4, 6]. However, since the prediction errors are less temporally coherent and contain more high frequency components than the ratios, their compression efficiency is not as good as that of ratios.

Therefore, an in [3], we employ the ratio frame, given by

$$r(x,y) = \log \frac{h(x,y)}{l(x,y)+\epsilon}, \tag{6}$$

where $h(x,y)$ and $l(x,y)$ are the pixel values of the HDR and LDR frames, and a small constant ϵ keeps $r(x,y)$ stable when both $h(x,y)$ and $l(x,y)$ are small.

3.2. Quality Control of the LDR and Ratio Frames

We have two sequences to encode: an LDR and a ratio sequence. Their qualities should be controlled to use the limited bit budget efficiently. While previous algorithms aimed to get the best performance for the HDR sequence only, we consider the qualities of both the HDR and the LDR sequences. Specifically, we control the quantization parameters for the LDR sequence ($\mathrm{QP}_{\mathrm{LDR}}$) and the ratio sequence ($\mathrm{QP}_{\mathrm{ratio}}$) so that the distortions of the reconstructed LDR sequence (D_{LDR}) and the HDR sequence (D_{HDR}) are minimized, subject to the overall bit budget. We solve this constrained optimization problem by minimizing the Lagrangian cost function, given by

$$J = D_{\mathrm{LDR}} + \mu D_{\mathrm{HDR}} + \lambda(R_{\mathrm{LDR}} + R_{\mathrm{ratio}}), \tag{7}$$

where R_{LDR} and R_{ratio} are bit rates for the LDR and ratio sequences, respectively. The distortions are measured by the sum of squared differences between the original and the reconstructed pixels. λ is a Lagrangian multiplier, which controls the tradeoff between the rates and the distortions, and μ is another multiplier, which determines the relative importance of the HDR sequence as compared with the LDR sequence. In this work, we determine μ as

$$\mu = \frac{\sum_{x,y} l(x,y)^2}{\sum_{x,y} h(x,y)^2}.$$

We model relation between the LDR, HDR and ratio frames as

$$r = a \cdot \frac{h}{l} + b.$$

Then, the squared error of the HDR frame is given by

$$\begin{aligned}(h-\hat{h})^2 &= \frac{1}{a^2}\left\{l(r-b) - \hat{l}(\hat{r}-b)\right\}^2 \\ &= \frac{1}{a^2}\left\{l(r-b) - \hat{l}(r-b) + \hat{l}(r-b) + \hat{l}(\hat{r}-b)\right\}^2,\end{aligned}$$

where r and l denote pixel values of the original ratio and LDR frames, and $\hat{r}$ and $\hat{l}$ denote the reconstructed values, respectively. To simplify the model, we assume that r and l are uncorrelated random variables and the averages of l and $\hat{l}$ are the same. Then, the model can be approximated by

$$\begin{aligned}D_{\mathrm{HDR}} &= E[(h-\hat{h})^2] \\ &\simeq \frac{1}{a^2}E\left[\{l(r-b) - \hat{l}(r-b)\}^2 + \{\hat{l}(r-b) - \hat{l}(\hat{r}-b)\}^2\right] \\ &\simeq \frac{1}{a^2}E[(r-b)^2(l-\hat{l})^2] + \frac{1}{a^2}E[\hat{l}^2(r-\hat{r})^2] \\ &\simeq \frac{1}{a^2}E[(r-b)^2]E[(l-\hat{l})^2] + \frac{1}{a^2}E[l^2]E[(r-\hat{r})^2] \\ &= C_{\mathrm{ratio}}D_{\mathrm{LDR}} + C_{\mathrm{LDR}}D_{\mathrm{ratio}},\end{aligned} \tag{8}$$

where $C_{\mathrm{ratio}} = \frac{1}{a^2}E[(r-b)^2]$ and $C_{\mathrm{LDR}} = \frac{1}{a^2}E[l^2]$. Also, the rates and distortions of the LDR and HDR pixel values are approximated by

$$D_{\mathrm{LDR}} = \alpha_{\mathrm{LDR}} 2^{-2R_{\mathrm{LDR}}} \tag{9}$$

$$D_{\mathrm{ratio}} = \alpha_{\mathrm{ratio}} 2^{-2R_{\mathrm{ratio}}}. \tag{10}$$

By substituting the equations into (7) and rearranging terms, we have the cost function, given by

$$\begin{aligned}J \;=\; &\alpha_{\mathrm{LDR}}(1+\mu C_{\mathrm{ratio}})2^{-2R_{\mathrm{LDR}}} + \alpha_{\mathrm{ratio}}\mu C_{\mathrm{LDR}} 2^{-2R_{\mathrm{ratio}}} \\ &+ \lambda(R_{\mathrm{LDR}} + R_{\mathrm{ratio}}).\end{aligned} \tag{11}$$

We differentiate (11) respect to R_{LDR} and R_{ratio} and set them 0, then we have the solution, given by

$$R_{\mathrm{LDR}} = \frac{1}{\ln 4}\ln\left[\alpha_{\mathrm{LDR}}(1+\mu C_{\mathrm{ratio}})\right] + \delta \tag{12}$$

$$R_{\mathrm{ratio}} = \frac{1}{\ln 4}\ln\left[\alpha_{\mathrm{ratio}}\mu C_{\mathrm{LDR}}\right] + \delta, \tag{13}$$

where δ is a constant.

The bit rate and the quantization step size (QSS) have the inverse exponential relation, and an increase of 1 in QP increases QSS by approximately 12% [8]. So, we can model them as

$$\mathrm{QSS} \simeq c\frac{1}{2^R} = (1.12)^{\mathrm{QP}}.$$

Therefore, QP and rate can be modeled as linear function, and using this model the QPs of the LDR and ratio sequences can be derived

$$\mathrm{QP}_{\mathrm{LDR}} - \mathrm{QP}_{\mathrm{ratio}} = \beta \ln\frac{\alpha_{\mathrm{LDR}}}{\alpha_{\mathrm{ratio}}} + \beta \ln\left[\frac{1+\mu \mathrm{C}_{\mathrm{ratio}}}{\mu \mathrm{C}_{\mathrm{LDR}}}\right],$$

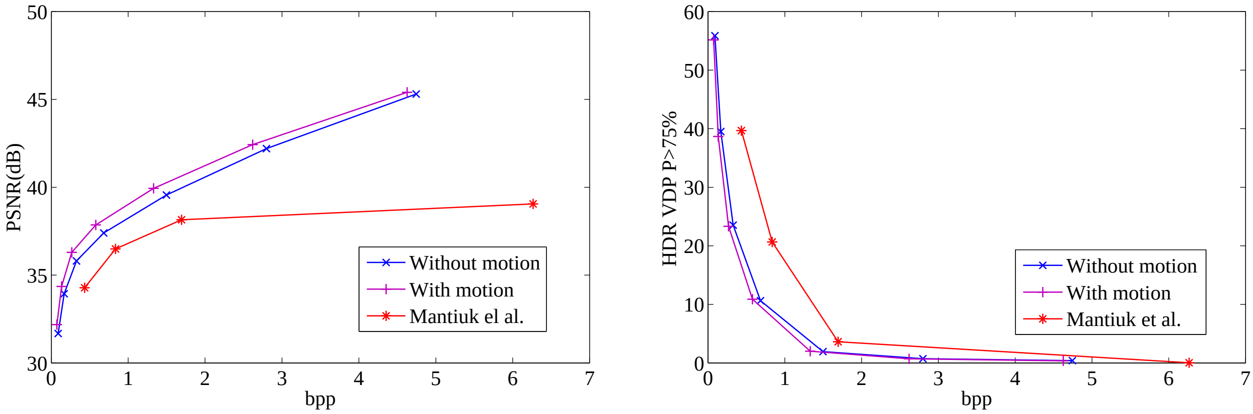
where β is a constants. Therefore, with some test sequences, we model the optimal relationship between $\mathrm{QP}_{\mathrm{LDR}}$ and $\mathrm{QP}_{\mathrm{ratio}}$ to an affine function, given by

$$\mathrm{QP}_{\mathrm{ratio}} = 1.24\mathrm{QP}_{\mathrm{LDR}} + 1.51.$$

4. EXPERIMENTAL RESULTS

We evaluate the performance of the proposed HDR video encoding algorithm using JM10.2 implementation for baseline profile of the H.264/AVC encoder. The parameters λ in (4) and ϵ in (6) are 0.3 and 0.05, respectively.

We evaluate the qualities using the PSNR measure for the LDR sequence and the VDP measure [11] for the HDR sequence. Figure 2 (a) compares the average PSNR performances on the test sequence. The proposed algorithm is tested in two ways: with and without motion information in the tone mapping. We see that the performance is increased by employing the motion information in the tone mapping. The proposed algorithm provides much higher PSNR performances than the Mantiuk *et al.*'s algorithm [6]. This is partly due to the fact that the compression performance of H.264/AVC is superior to MPEG-4, which is employed in [6]. Also, the proposed algorithm allocates bits to the LDR



(a)

(b)

Fig. 2. The performance comparison of the ”tunnel1” sequence on (a) the LDR sequence and (b) the HDR sequence.

sequence and the ratio sequence more effectively based on the Lagrangian cost function.

Figure 2 (b) shows the VDP performances on the HDR sequences, which compare the average numbers of pixels where HVS perceives the differences between the original and reconstructed values. Exploiting motion information in the tone mapping also improves the quality of the reconstructed HDR sequence. The proposed algorithm provides better performance by allocating the limited bit budget efficiently. To summarize, the proposed algorithm provides significantly better qualities in both LDR and HDR sequences than the Mantiuk *et al.*’s algorithm.

5. CONCLUSION

In this work, we proposed a rate-distortion optimized bit allocation algorithm for HDR video compression. The proposed algorithm separates an input HDR sequence into a tone-mapped LDR sequence and a ratio sequence. We developed a rate-distortion model, which maximizes the qualities of the reconstructed LDR and HDR sequences subject to total bit rate. Simulation results showed that the proposed algorithm provides higher qualities in both LDR and HDR sequences than the Mantiuk *et al.*’s algorithm.

6. REFERENCES

[1] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, *High Dynamic Range Imaging*, Morgan Kaufmann Publishers, 2005.

[2] P. Debevec, E. Reinhard, G. Ward, K. Myszkowski, H. Seetzen, H. Zargarpour, G. McTaggart, and D. Hess, “Course on high dynamic range imaging: theory and applications,” in *ACM SIGGRAPH Course Notes*, vol. 1, Jul. 2006.

[3] G. Ward and M. Simmons, “Subband encoding of high dynamic range imagery,” in *Proc. the 1st Symposium on Applied Perception in Graphics and Visualization*, pp. 83-90, Aug. 2004.

[4] M. Okuda and N. Adami, “Two-layer coding algorithm for high dynamic range images based on luminance compensation,” *J. Visual Commun. Image Rep.*, vol. 18, no. 5, pp. 377-386, Oct. 2007.

[5] R. Mantiuk, G. Krawczyk, K. Myszkowski, and H.-P. Seidel, “Perception-motivated high dynamic range video encoding,” *ACM Trans. Graphics*, vol. 23, no. 3, pp. 733-741, Aug. 2004.

[6] R. Mantiuk, A. Efremov, K. Myszkowski, and H.-P. Seidel, “Backward compatible high dynamic range MPEG video compression,” *ACM Trans. Graphics*, vol. 25, no. 3, pp. 713-723, Jul. 2006.

[7] A. Segall, “Scalable coding of high dynamic range video,” in *Proc. ICIP*, pp. 1-4, vol. 1, Sep. 2007.

[8] T. Wiegand, G.J. Sullivan, G. Bjøntegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560-576, Jul. 2003.

[9] C. Lee and C.-S. Kim, “Gradient domain tone mapping of high dynamic range videos,” in *Proc. ICIP*, vol. 3, pp. III-461-III-464, Sep. 2007.

[10] R. Fattal, D. Lischinski, and M. Werman, “Gradient domain high dynamic range compression,” *ACM Trans. Graphics*, vol. 21, no. 3, pp. 249-256, Jul. 2002.

[11] R. Mantiuk, K. Myszkowski, and H.-P. Seidel, “Visible difference predictor for high dynamic range images,” in *Proc. IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, pp. 2763-2769, Oct. 2004.