16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

RATE-DISTORTION OPTIMIZED COMPRESSION OF HIGH DYNAMIC RANGE VIDEOS

Chul Lee and Chang-Su Kim

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

ABSTRACT

An efficient algorithm to compress high dynamic range (HDR) videos is proposed in this work. We employ the motion information to separate an HDR video sequence into a tone-mapped low dynamic range (LDR) sequence and a ratio sequence. Then, those two sequences are encoded using the standard H.264/AVC codec. During the encoding, we allocate the limited bit budget to the LDR sequence and the ratio sequence optimally to maximize the qualities of the LDR and HDR sequences. Conventional LDR devices use only the LDR stream, whereas HDR devices reconstruct the HDR video using the LDR stream and the ratio stream. Simulation results show that the proposed algorithm provides better performance than the conventional methods.

1. INTRODUCTION

The dynamic range of a digital image is defined 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 via adaptation. Images that have higher dynamic ranges than conventional display devices are called HDR images. In order to display HDR images on conventional devices, we need tone mapping techniques [1].

A lot of efforts have been made in the research and development of HDR imaging technologies from acquisition to display. Entertainment industries already started to employ HDR technologies. For example, video games 'Need for Speed: Most Wanted' and 'Half-Life 2: Lost Coast' and movies 'Harry Potter and the Sorcerer's Stone' and 'Men in Black II' used HDR technologies to render computer graphics scenes [2, 3]. In addition, graphics card manufacturers, such as NVIDIA® and ATI®, supply products supporting HDR rendering. It is expected that this trend would continue and HDR imaging would be used in general applications in near future [1].

HDR images and videos require a big amount of storage space and transmission bandwidth. However, only a few algorithms have been proposed to compress HDR images or videos. In [4], Ward and Simmons proposed a backward-compatible compression algorithm for HDR images. Their algorithm first obtains a lower dynamic range (LDR) image from an input HDR image using a tone-mapping scheme. Then, it encodes the tone-mapped LDR image and the ratio image between the HDR and LDR images using the standard JPEG codec. At the decoder side, a naive JPEG decoder decodes only the LDR image, while an HDR-enabled decoder can reconstruct the HDR image. In [5], to improve the coding performance, Okuda *et al.* adopted an inverse tone mapping function and a new luminance compensation scheme, and used a wavelet encoder to compress ratio images.

Recent video coding standards support the encoding of videos with an extended bit depth up to 12 bits, but they cannot reconstruct full dynamic ranges of HDR videos faithfully. Mantiuk *et al*. [6] proposed a luminance quantization method, which is optimized for the contrast threshold perception in the human visual system (HVS). Their algorithm compresses the quantized videos, using the MPEG-4 codec, with an additional scheme to transform high contrast blocks. In [7], Mantiuk *et al*. extended the work in [6] to offer the backward compatibility with LDR devices. Also, they proposed several tools to achieve higher flexibility and coding gain.

In this work, we propose a novel algorithm to compress HDR videos. The proposed algorithm uses a tone mapping algorithm to separate an HDR video into an LDR sequence and a ratio sequence. By exploiting the motion information in the tone mapping, the proposed algorithm reduces flickering artifacts in the LDR sequence. The proposed algorithm uses the state-of-the-art H.264/AVC codec [8] to encode LDR and ratio sequences. We develop a rate-distortion optimized scheme to allocate a limited bit budget to the LDR and ratio sequences efficiently to maximize the quality of the reconstructed sequences. Simulation results demonstrate that the proposed algorithm provides better performance than the conventional algorithm [7].

The remainder of this paper is organized as follows. Section 2 reviews the gradient domain tone mapping of HDR videos, and Section 3 describes the proposed HDR video encoding algorithm. Section 4 provides experimental results. Finally, Section 5 concludes the paper.

2. TONE MAPPING

We use the gradient domain tone mapping algorithm of HDR videos, proposed in [9], since it reduces flickering artifacts in LDR sequences and improve the compression performance. For the sake of completeness, we briefly review the algorithm in this section.

2.1 Gradient Domain Tone Mapping

The gradient domain tone mapping, which was first proposed in [10], is based on the assumption that we can compress the dynamic range of an input image, while preserving the details, by attenuating large gradients.

This research was supported by the Ministry of Knowledge Economy, Korea, under the Information Technology Research Center support program supervised by the Institute of Information Technology Advancement (grant number IITA-2008-C1090-0801-0017).

A modified gradient map $g(x,y)$ is obtained by multiplying the gradient map $\nabla h(x,y)$ of the original HDR image with an attenuation function $\Phi(x,y)$.

$$g(x,y) = \nabla h(x,y)\Phi(x,y).$$

The output LDR image l, whose gradient is the closest to g, can be obtained numerically by minimizing a cost function. After the discrete approximation of gradient operators, the cost function to be minimized is given by

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

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

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

This is the discrete approximation of the Poisson equation

$$\nabla^2 l = \operatorname{div} g,$$

which can be solved efficiently due to its sparsity.

2.2 Tone Mapping of HDR Videos

If the tone mapping algorithm in [10] is applied to each frame of a video independently, the resultant LDR sequence contains severe flickering artifacts. In [9], Lee and Kim proposed a video tone mapping algorithm, which obtains a high-quality, temporally coherent LDR sequence by exploiting the motion information.

The Lee and Kim's algorithm first estimates a pixelwise motion vector field between two successive HDR frames $h(x,y,t)$ and $h(x,y,t-1)$. As shown in Figure 1, let (v_x, v_y) denote the motion vector for a pixel $h(x,y,t)$. Then, the same motion vector is used to estimate the LDR pixel value $l(x,y,t)$. Let $p(x,y)$ denote the pixel in the previous frame, specified by the motion vector (v_x, v_y), *i.e.*,

$$p(x,y) = l(x - v_x, y - v_y, t-1).$$

After the tone mapping, $l(x,y,t)$ and $p(x,y)$ should have similar values, since they represent the same object point at successive time instances. Therefore, in addition to the original cost function C_1 in (1), a new term is introduced for the temporal coherence, given by

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

Then, the overall cost is defined as a weighted sum of the two costs

$$C = C_1 + \lambda C_2, \tag{3}$$

where λ is a weighting coefficient.

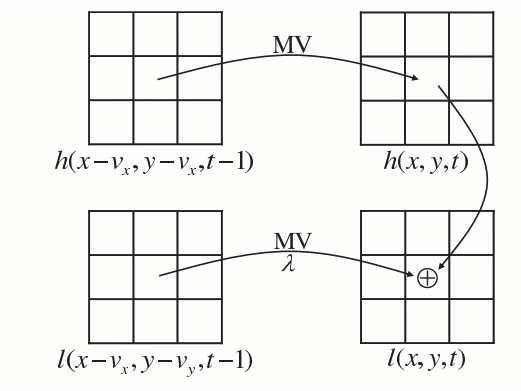


Figure 1: A pixelwise motion vector field is obtained between the HDR frames $h(x,y,t-1)$ and $h(x,y,t)$. It is then used as a constraint for the temporal coherence between the LDR frames $l(x,y,t-1)$ and $l(x,y,t)$.

By differentiating C respect to $l(x,y,t)$, setting it to 0 and rearranging terms, we have the modified Poisson equation

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

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

Figure 2 shows the block diagram of the proposed HDR video encoder. The encoder takes HDR frames as input, and separates them into an LDR sequence and a ratio sequence using the tone mapping operator in Section 2. We encode the LDR frames with the standard H.264/AVC encoder, offering backward compatibility. Then, the LDR frames are reconstructed and compared with the original HDR frames to yield ratio frames. The ratio frames are then filtered to enhance the compression efficiency. Finally, the ratio frames are also encoded using the H.264/AVC encoder. A set of parameters are used in the process of obtaining the ratio frames, and transmitted to the decoder as side information.

3.1 Ratio Frames

An HDR frame is separated into a tone-mapped LDR frame and residual data that represent the differences between the HDR and LDR frames. In [4], the residual data consist of the ratios, obtained by dividing the HDR value by the LDR value at each pixel position. In general, the HDR value and the LDR value are correlated. Using the correlation, in [5, 7], the HDR value is predicted from the LDR value, and the prediction error is encoded, instead of the ratio, as the residual data. However, the prediction errors are less temporally coherent and contain more high frequency components than the ratios. Thus, they cannot be compressed efficiently.

Therefore, as in [4], we employ the ratio frame, given by

$$r(x,y) = \log\left(\frac{h(x,y)}{l(x,y) + \varepsilon} \right), \tag{4}$$

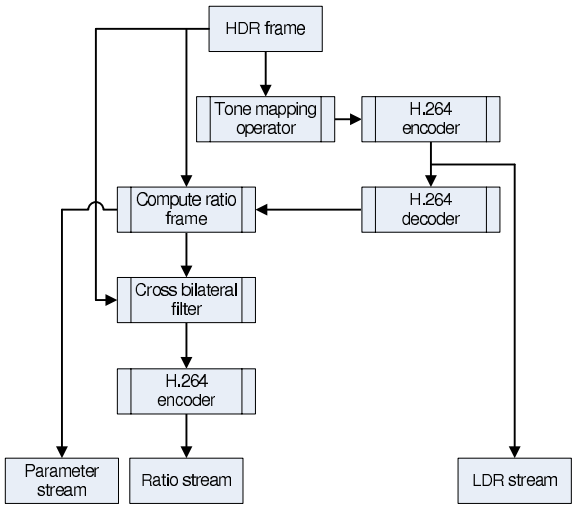


Figure 2: The block diagram of the proposed HDR video encoder.

where $h(x,y)$ and $l(x,y)$ are the luminance values of the HDR and LDR frames. A small constant ε in (4) prevents the division by zero. The logarithm representation suppresses unstable values when both $h(x,y)$ and $l(x,y)$ are small. Also, note that HVS is well modeled by the logarithm. The pixel values of the ratio frame are then normalized to the range $[0,255]$.

Notice that the ratio frame in (4) is obtained by dividing the original HDR frame by the reconstructed LDR frame. Thus, it contains lots of high frequency components, which are from the tone mapping operator, the quantization noises in the LDR frame encoding, and the division. Those high frequency components reduce the coding gain of the ratio frame. To improve the coding gain, we should remove noise components, while preserving sharp edge information. In [4], the spatial resolution of the ratio frame is reduced to erase high frequency components and reduce the bit rate. However, this approach degrades the quality of the reconstructed frame, although the pre-correction or post-correction techniques are employed. In [7], a filter, based on the properties of HVS, is used to remove invisible high frequency components. But, it requires complex computations.

In this work, we employ the cross bilateral filter [12], which is a variant of the bilateral filter, to remove noise components. The filter is given by

$$r'(\mathbf{p}) = \frac{\sum_{\mathbf{q}\in S} G_{\sigma_s}\big(\|\mathbf{p}-\mathbf{q}\|\big) G_{\sigma_r}\big(h(\mathbf{p})-h(\mathbf{q})\big) r(\mathbf{q})}{\sum_{\mathbf{q}\in S} G_{\sigma_s}\big(\|\mathbf{p}-\mathbf{q}\|\big) G_{\sigma_r}\big(h(\mathbf{p})-h(\mathbf{q})\big)}, \quad (5)$$

where S denotes the set of neighboring pixels of $\mathbf{p}$, $h(\mathbf{p})$ and $r(\mathbf{q})$ are the values of the HDR and the ratio frames at pixel position $\mathbf{p}$. Also, $G_\sigma(x)$ is the Gaussian kernel with variance σ^2. Thus, the cross bilateral filter is an adaptive moving average filter, which assigns higher weights to neighboring pixels that have similar values in the HDR frame. Consequently, it smooths the ratio frame, while retaining high frequency components corresponding to the edges in the HDR frame. Figure 3 shows an example of the cross bilateral filtering. After the filtering, noises are removed within each objects, while the edge information is sharpened.

(a) (b)

Figure 3: An example of the cross bilateral filtering: (a) before filtering and (b) after filtering.

3.2 Quality Control of LDR and Ratio Frames

We have two sequences to encode: an LDR sequence and a ratio sequence. Their qualities should be controlled to use the limited bit budget effectively. Ward *et al*. [4] used the highest quality setting for the ratio frame, and Mantiuk *et al*. [7] used the same quality setting for both sequences. These strategies aim to get the best performance for the HDR sequence only.

On the other hand, we consider the qualities of both the HDR sequence and the LDR sequence. 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 reconstructed HDR sequence (D_{HDR}) are minimized, subject to the constraint on the overall bit budget. This constrained optimization problem can be solved by minimizing the Lagrangian cost function, given by

$$J = D_{\mathrm{LDR}} + \mu D_{\mathrm{HDR}} + \lambda\big(R_{\mathrm{LDR}} + R_{\mathrm{ratio}}\big), \quad (6)$$

where R_{LDR} and R_{ratio} are the 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, μ is set to 1.

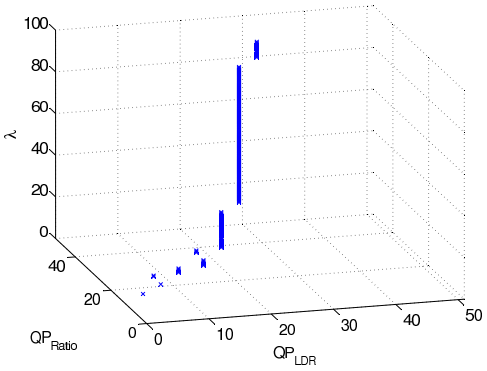
Figure 4 (a) shows, for each λ, the optimal combination $(\mathrm{QP}_{\mathrm{LDR}}, \mathrm{QP}_{\mathrm{ratio}})$ of the quantization parameters that minimizes the cost function J in (6). We project the three-dimensional points onto the $\mathrm{QP}_{\mathrm{LDR}}$-$\mathrm{QP}_{\mathrm{ratio}}$ plane as shown in Figure 4 (b). Then, we model the optimal relationship between $\mathrm{QP}_{\mathrm{LDR}}$ and $\mathrm{QP}_{\mathrm{ratio}}$ with an affine function, given by

$$\mathrm{QP}_{\mathrm{ratio}} = 0.77\mathrm{QP}_{\mathrm{LDR}} + 13.42.$$

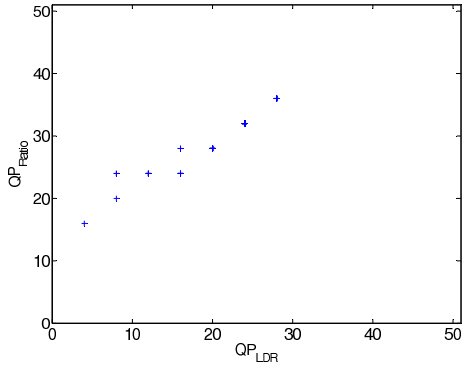
Thus, using this relationship, we control the qualities of both sequences using only one parameter $\mathrm{QP}_{\mathrm{LDR}}$ in this work.

3.3 Color Compensation

While all luminance (Y) and chrominance (U, V) components of the LDR sequence is encoded, only the luminance component of the ratio sequence is defined and encoded. Thus, to reconstruct the color components of the HDR sequence, we use the following system of equations [1], assuming that the amount of saturation is the same for all com-



(a)



(b)

Figure 4: (a) Optimal combinations of the Lagrangian multiplier λ and the quantization parameters QP_{LDR} and QP_{ratio}, and (b) the projection of (a) onto the QP_{LDR}-QP_{ratio} plane.

ponents during the tone mapping.

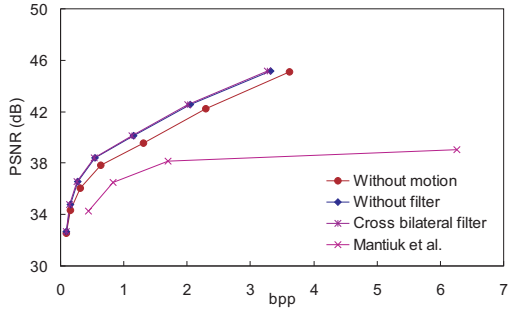
$$\begin{bmatrix} h_r \\ h_g \\ h_b \end{bmatrix} = \begin{bmatrix} h_y \left(\frac{l_r}{l_y}\right)^{1/s} \\ h_y \left(\frac{l_g}{l_y}\right)^{1/s} \\ h_y \left(\frac{l_b}{l_y}\right)^{1/s} \end{bmatrix}, \tag{7}$$

where l_y and h_y are the luminance values of the reconstructed LDR and HDR pixels, respectively. l_r, l_g, and l_b are the R, G, B components of the reconstructed LDR pixel. The exponent s is the saturation parameter less then 1, which controls the amount of coupling between the color components of the LDR and HDR pixels.

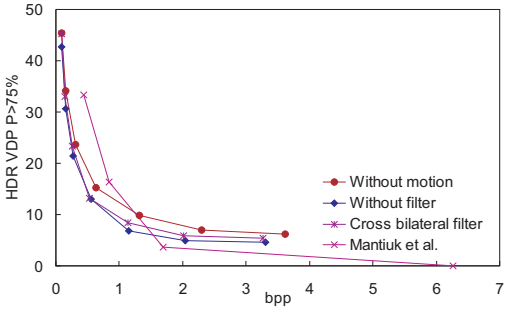
4. EXPERIMENTAL RESULTS

We use the JM10.2 implementation for the baseline profile of the H.264/AVC coder, and the fast approximate implementation of the cross bilateral filter in [13]. The parameter ε in (4) is 0.05, σ_s and σ_r in (5) are 8 and 0.1, respectively, and s in (7) is 0.6.

For objective quality evaluation, we use the PSNR measure for the LDR sequence and the VDP measure [11] for the HDR sequence. Figure 5 (a) compares the average PSNR performances on the test sequence in Figure 3. The proposed algorithm is tested in three ways: The red curve shows the performance, when the motion information is not used in the



(a)



(b)

Figure 5: The performance comparison on (a) the LDR sequence and (b) the HDR sequence. The data for the Mantiuk *et al.*'s algorithm were provided by Dr. Mantiuk.

tone mapping and the cross bilateral filter is not employed. The blue curve depicts the case when the motion information is used but the filter is not employed. The purple curve is the case when both the motion information and the filter are employed. We see that the performance is increased significantly by employing the motion information in the tone mapping. Also, the cross bilateral filter slightly improves the PSNR performance. Note that the proposed algorithm provides much higher PSNR performances than the Mantiuk *et al.*'s algorithm [7]. This is partly due to the fact that the compression performance of H.264/AVC is superior to MPEG-4, which is employed in [7]. Also, the proposed algorithm allocates bits to the LDR sequence and the ratio sequence more effectively based on the Lagrangian cost function.

Figure 5 (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. Thus, lower curves represent better performances. Exploiting motion information in the tone mapping also improves the quality of the reconstructed HDR sequence. Since the Mantiuk *et al.*'s algorithm focuses only on the HDR quality, it shows better performance at high bit rates. On the other hand, at low bit rates, 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 at low bit rates.

Figure 6 compares the size of the LDR stream with that of the rest streams. The LDR stream consumes a dominant

(a) (b) (c)

Figure 7: Comparison of the reconstructed HDR frames: (a) the original frame, (b) the Mantiuk *et al.*'s algorithm [7] (bpp=0.43, HDR VDP 75%=33.31%), (c) the proposed algorithm (bpp=0.26, HDR VDP 75%=23.32%). The images were tone-mapped to be printed.

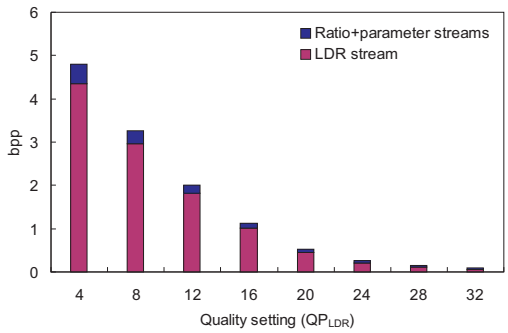


Figure 6: The sizes of the LDR stream and the rest streams.

bit rate, which corresponds to about $70 \sim 90$ % of the total bit rate. This indicates that, in addition to the LDR bit rate, only $10 \sim 30$ % more bit rate is required to support HDR devices. Figure 7 compares the reconstructed HDR frames at similar bit rates. We see that the proposed algorithm provides a higher quality than the Mantiuk *et al.*'s algorithm.

5. CONCLUSIONS

In this work, we proposed a rate-distortion optimized compression algorithm for HDR videos. The proposed algorithm separates an input HDR sequence into a tone-mapped LDR sequence and a ratio sequence. We developed a bit allocation scheme, which maximize the qualities of the reconstructed LDR and HDR sequences subject to the constraint on the 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 at low bit rates.

6. ACKNOWLEDGEMENTS

We thank Dr. Grzegorz Krawczyk for making the HDR sequence in Figure 3 available and Dr. Rafał Mantiuk for providing us the experimental data for comparison and valuable comments.

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*, Jul. 2006.

[3] F. Kains, R. Bogart, D. Hess, P. Schneider, and B. Anderson, "OpenEXR," www.openexr.com, 2003.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] T. Wiegand, G.J. Sullivan, G.Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video conding 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, Oct. 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 predicator for high dynamic range images," in *Proc. IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, pp. 2763–2769, Oct. 2004.

[12] E. Eisemann and F. Durand, "Flash photography enhancement via intrinsic relighting," *ACM Trans. Graphics*, vol. 23, No. 3, pp. 673–678, Aug. 2004.

[13] S. Paris and F. Durand, "A fast approximation of the bilateral filter using a signal processing approach," in *Proc. European Conference on Computer Vision*, 2006.