

ACQUISITION AND ENCODING OF HIGH DYNAMIC RANGE IMAGES USING INVERSE TONE MAPPING

Takao Jinno, Masahiro Okuda*

The University of Kitakyushu
Dpt. of Information and Media Sciences

Nicola Adami[†]

University of Brescia
Dpt. of Electronic for the Automation

ABSTRACT

A two layer coding algorithm for high dynamic range images is discussed. In the first layer, a low dynamic range image is encoded by a conventional codec, and then the residual information that represents the difference between an original and the decoded images in the first layer is encoded in the second layer, which realizes compatibility with conventional image file formats. Our method utilizes the approximation of an inverse tone mapping function that reduces the high dynamic range to a displayable range. Our algorithm significantly improves a compression performance, compared to conventional methods.

Keywords: High Dynamic Range Images, Two-layer Encoding, Inverse Tone Mapping Function

1. INTRODUCTION

Since the sizes of the HDR images are often huge, development for functional compression is one of the research topics. The image coding standard, JPEG 2000, provides seamless compression from 1 to 16 bits per color channel [?]. Xu et al.'s scheme [?] verifies the validity of JPEG2000 for the high dynamic range images. An HDR video compression scheme which uses MPEG2 has also been proposed [?]. Spaulding [?] proposes a two layer encoding for gamut extended images. In the first layer, an image with clipped gamut is encoded. And then in the second layer, the residual information that represents the difference between the original gamut extended image and the decoded image in the first layer is encoded. The main advantage of this approach is that the format is backward compatible to existing file formats, and no extra efforts are needed to extract the 24 bit image. In the field of Computer Graphics, similar concepts are adopted for the high dynamic range image compression [?], [?].

In this paper, we extend and elaborate on the conventional two-layer coding methods [?] - [?]. We approximate a tone mapping function that reduces dynamic ranges of the HDR images to displayable ranges by a single function. We encode the difference between the HDR image and its tone mapped version, indicated as Low Dynamic Range (LDR) Image. This approximation significantly improves coding efficiency.

*This work was partly supported by a Grant-in-Aid for Young Sciences (#14750305) of Japan Society for the Promotion of Science, fund from MEXT via Kitakyushu innovative cluster project.

[†]Thanks to the support of Invitation Fellowship Programs of Japan Society for the Promotion of Science.

2. ACQUISITION OF HDR IMAGES

The HDR images are synthesized using multiple images with different exposures. Compensating the motion between the images is a challenging problem. Kang et.al [?] proposes the motion estimation for HDR video synthesis. Our acquisition algorithm is also based on a simple motion compensation algorithm in which the motion is estimated by block matching. In our block search, the error of exposure is calculated instead of pixel values, that is a key of our method. Fig.?? depicts the results compared with the state of the art software "Photomatix"¹. One can see from the figure that the motion is well compensated in our method.



Fig. 1. (left) Our HDRI, (right) Photomatix

3. HDR IMAGE CODING

3.1. Summary

The proposed method realizes the two-layer compression similar to [?]. That is, starting from the original HDR, a tone mapped image (LDR) is generated or given as input, and then encoded by using JPEG. A residue information is also encoded in order to enable the reconstruction of the HDR from the knowledge of its LDR representation. The most significant difference from the conventional methods is the use of an Inverse Tone Mapping Function (ITMF). Fig.?? depicts the encoding steps. Our method assumes that an original HDR and its tone mapped LDR image are given as inputs. The LDR image is encoded by JPEG, followed by JPEG decoding. The intensity Y is transformed by the ITMF to expand the reduced dynamic range. Then we compare it with the intensity of the HDR image and calculate a residue. The residue is encoded by the wavelet-based image compression. The ITMF used here is image-dependent function. Thus the parameters to describe the ITMF are calculated using the

¹<http://www.hdrsoft.com/>

given two images on line and then they are sent to the decoder as well as the encoded LDR and the residue. At low bit rates, only the intensity of the HDR image is encoded and the chroma is estimated by the LDR image. At higher bit rate, color difference between the HDR and the LDR with color compensation is compressed by the wavelet. Each step is explained more in detail in the following sections.

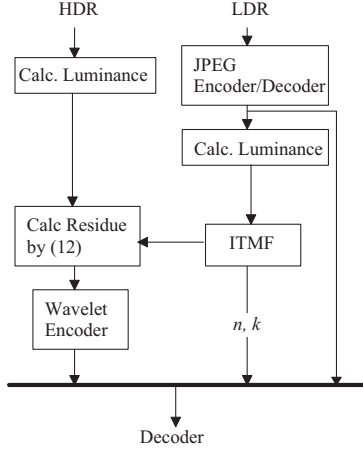


Fig. 2. Outline of Our Method

3.2. Image Dependant Inverse Tone Mapping Function

We call the procedure for converting an HDR to LDR image, the Tone Mapping (TM), and the inverse procedure of the TM, the Inverse Tone Mapping (ITM). In this section, we treat the problem of finding the Inverse Tone Mapping Functions (ITMF), given an image pair of HDR and tone mapped LDR. Our attempt here is to closely approximate the relationship between HDR and tone mapped LDR by ITMF, which will improve a coding efficiency.

Many tone mapping algorithms have already been proposed. In some methods, global operations are applied to the HDR image. Some other algorithms locally map each pixel of HDR to LDR. (A comprehensive review of the existing tone mapping operations can be found in [?].) Although the local tone mapping operations can not inherently described by a function, many of them can be approximated by the s-shape. Thus regardless of which type of mapping is actually used to create the LDR image, in this paper, we approximate it only by a single function.

The response of the retina is well approximated by the s-shape function. To fairly approximate the s-shape, we have adopted the Hill function after testing other functions such as polynomials, sigmoid functions. The formula of the Hill function f is given by

$$y = f(x) = \frac{x^n}{x^n + k^n}. \quad (1)$$

where the function is normalized by 1. This function has some desirable properties:

1. By using the function one can easily express the s-shape with long saturated regions
2. For $x > 0$ and $k > 0$, the function monotonically increases (see Appendix).
3. Response is controlled by only two parameters, n and k .

4. In the log domain, the inverse of the function is linear to the two parameters $\log(k)$ and $1/n$, which will be further discussed later.

The first two properties are preferable for the approximation of the response of the retina, while the others make it easy to design and handle the inverse tone mapping functions. Further discussion on this function in a viewpoint of human perception can be found in [?] and [?].

3.3. Closed Form Solution

Here we derive the ITMF to convert the LDR to HDR, which is optimal in least squares sense. The inverse of f in (1), $g = f^{-1}$ is given by

$$x = g(y) = k \left(\frac{y}{1-y} \right)^{\frac{1}{n}}, \quad (2)$$

Then the conversion from LDR to HDR image by the ITMF is expressed as

$$\mathcal{H}'_y(i, j) = g(\mathcal{L}_y(i, j)) = k \left(\frac{\mathcal{L}_y(i, j)}{1 - \mathcal{L}_y(i, j)} \right)^{\frac{1}{n}}, \quad (3)$$

where $\mathcal{H}'_y(i, j)$ and $\mathcal{L}_y(i, j)$ is the luminance Y of the reconstructed HDR image and the LDR image, respectively. Then we define the error between $\mathcal{H}'_y(i, j)$ and the original HDR image $\mathcal{H}_y(i, j)$ by

$$e(i, j) = \log(\mathcal{H}_y(i, j)) - \log(\mathcal{H}'_y(i, j)) \quad (4)$$

The error in the log domain put more weights on low luminance. This metric matches human perceptions more and suppress the overestimate of high luminance. Moreover it linearizes the function w.r.t. the parameters $\log(k)$ and $1/n$ as follows.

$$\log(\mathcal{H}'_y(i, j)) = \log(k) + \frac{1}{n} \log \left(\frac{\mathcal{L}_y(i, j)}{1 - \mathcal{L}_y(i, j)} \right) \quad (5)$$

Then the design problem of ITMF is stated as the minimization of the mean squared error of (5),

$$E = \frac{1}{M} \sum_{i,j} e(i, j)^2 \quad (6)$$

where M is the number of pixels. Since (5) is a quadratic form with respect to $\log(k)$ and $1/n$, the optimum solution that minimizes (6) is uniquely determined. In the end, solving the two equations

$$\frac{\partial E}{\partial(\log(k))} = 0, \quad \frac{\partial E}{\partial(1/n)} = 0 \quad (7)$$

yields a closed form solution for n and k .

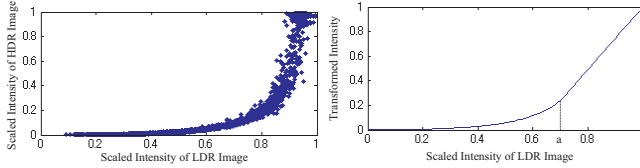
In the tone mapping procedures, to reduce the dynamic range of the HDR images, special attention should be paid for bright area (luminance more than a few thousands in cd/m^2), where human sensitivity is saturated. The simplest operation is to clip the values higher than a threshold and they are mapped to 255. To improve the quality, more sophisticated operations have been proposed[?], in which the local operators determine the mapping depending on the neighbors of each pixel. Due to these, the approximation described in this section often fails in the bright regions, especially in the case that the LDR images are coarsely quantized by the JPEG. Fig.??(a)

shows the plot of the LDR pixel values versus the HDR pixel values in corresponding positions. One can see from this figure that the variance of the HDR pixel values increases as the luminance gets high. Due to this, in the bright area the optimal approximation in least squares sense does not necessarily give an optimal compression performance. Thus, in our algorithm we simply approximate the bright area by a first order polynomial. Fig.??(b) shows the obtained ITMF. In this example, the region $\mathcal{L}_y(i, j) \in [0, a]$ is approximated by the hill function and the linear mapping is done in $\mathcal{L}_y(i, j) \in (a, 1]$ (in this paper, we suppose that LDR images are normalized to 1). The boundary a is determined as follows. First we divide the region of the LDR intensity $[0, 1]$ to Q bins. For each bin, the variance of HDR intensity in corresponding positions, $\sigma_q^2 = \text{var}(H_q)$, $q = 1, 2, \dots, Q$, where

$$H_q = \{\mathcal{H}_y(i, j) \mid i, j, (q-1)/Q \leq \mathcal{L}_y(i, j) < q/Q\}$$

$$q = 1, 2, \dots, Q.$$

Next we find the smallest value q^* that satisfies $\sigma_{q^*}^2 > c \cdot \sigma_{hdr}^2$, and then we set $a = (q^* - 1)/Q$, where σ_{hdr}^2 is a variance of the HDR image and c is a constant. For most examples, it works well with $c = 0.01$.



(a) LDR-HDR plots, (b) Inverse Tone Mapping Function

Fig. 3. Approximation by ITMF

3.4. Residue Compression

As explained in the previous section, the luminance of the LDR image is transformed by the ITMF in (??). Then the residual image \mathcal{R} is created by using the original HDR image, $\mathcal{H}_y(i, j)$ and the transformed image, $\mathcal{H}'_y(i, j)$

$$\mathcal{R}(i, j) = \left(\frac{\mathcal{H}_y(i, j)}{\mathcal{H}'_y(i, j) + \epsilon} \right)^\gamma, \quad (8)$$

where γ is a constant less than 1. When both of the two images, \mathcal{H}_y and \mathcal{H}'_y have very small values, slight difference between them will be overvalued by (??) more than the case that they have large values. Thus this division process implicitly weights the small pixel values more. γ in (??) plays a role in relieving this weight effect. ϵ is introduced to prevent the values from being too large when $\mathcal{H}'_y(i, j)$ approaches to zero.

Then the image is reconstructed in the decoding steps by

$$\bar{\mathcal{H}}_y(i, j) = \mathcal{R}(i, j)^{\frac{1}{\gamma}} (\mathcal{H}'_y(i, j) + \epsilon), \quad (9)$$

where $\bar{\mathcal{H}}_y$ is the decoded image.

3.5. Color Compensation

In [?], G. Ward et al. point out that if the TM operation preserves hue of an HDR image, and its saturation change is estimated, the

chroma of the original HDR can be fairly recovered by its LDR by the saturation compensation. Our algorithm adopts a different color compensation scheme. Assume the relation between that RGB values of the HDR and LDR images is approximated by a polynomial,

$$y = p(x) = \sum_{n=0}^N c_n x^n, \quad (10)$$

then we simply apply this to the LDR image, that is,

$$\mathcal{L}_r^c(i, j) = p(\mathcal{L}_r(i, j)), \quad \text{for channel R} \quad (11)$$

The function f can be determined by a simple least squares method, and its coefficients are transformed to the decoder as a side information.

After all the colors are compensated by (??), we decompress the dynamic range as follows

$$\bar{\mathcal{H}}'_y(i, j) = \mathcal{L}_r^c(i, j) \frac{\bar{\mathcal{H}}'_y(i, j)}{\mathcal{L}_y(i, j)}, \quad \text{for channel R}, \quad (12)$$

where $\bar{\mathcal{H}}'_y(i, j)$ is the decoded luminance Y of the HDR image. The same procedure is done for G and B channels as well. This procedure maps the color of the LDR to the HDR image. In a low bit rate, one can spend all bits for the intensity \mathcal{H}_y , and the chroma is compensated by (??). In high bit rates, however, it often gives better results to encode chroma as well. The color encoding is required when the tone mapping operator drastically changes the colors. Thus in the high bit rates, our scheme encodes the chroma difference between the original HDR and color compensated HDR given by (??) as well.

4. EXPERIMENTAL RESULTS

We have tested dozens of HDR images collected from some web sites, many of which are frequently used as sample images. Since the HDR images may have very bright and dark areas in that human do not even perceive luminance change, conventional error metrics (e.g. mean squared error, snr) are not reliable. Thus we instead use two other metrics, the Mean Distance (MD) in CIELAB color space to evaluate color appearance differences, and the Mean Squared Error (MSE) with Daly's nonlinearity. The CIELAB (also called CIE 1976 $L^*a^*b^*$) is a well known uniform color space, which is one of the most reliable evaluation for color images. We use the mean of the absolute difference in this space for color difference evaluation. The second metric evaluates only the luminance. To do this we first transform the intensity with Daly's nonlinearity [?]. Then we calculate the mean squared error.

First we test the performance on several images and compare it with Ward et al's JPEG-HDR [?]. This method encodes the luminance residue, while colors are recovered from saturation information of the LDR image. For fair comparison, we use exactly the same jpeg images as the LDR images for input. We spend all the bits for luminance compression, and we do not encode other color channels. Fig.?? shows comparisons for some images². At very low bit rates, our method is comparable or even worse, but at high bit rates our method gains significant improvements.

²The results are obtained using a software the authors provide

Fig.?? shows a comparison with the most recent HDR coding method [?]. This method is designed mainly for HDR video compression, but it also gives high performances for still images. The plots indicate the results of "Memorial" image³. In these experiments, we use Reinhard et al.'s global tone mapping operator [?] to make LDR images, and we allocate 75 % of bits for the intensity and the rest of them for chroma difference encoding. From these figures, it can be seen that our algorithm performs better than [?].

5. REFERENCES

- [1] P.E. Debevec and J. Malik. "Recovering High Dynamic Range Radiance Maps from Photographs, Proceedings of SIGGRAPH 97, Computer Graphics Proceedings, pp. 369-378 .
- [2] T. Mitsunaga, and S. K. Nayer,"Radiometric Self Calibration," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Vol.1, pp.374-380, Jun, 1999.
- [3] E. Reinhard, S. Pattanaik, G. Ward and P. Debevec, " High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting (Morgan Kaufmann Series in Computer Graphics and Geometric Modeling), Morgan Kaufmann Publisher 2005.
- [4] E. Reinhard, M. Stark, P. Shirley and J. Ferwerda, "Photographic Tone Reproduction for Digital Images", ACM Trans on Graphics. 21, 3, 267-276, 2002
- [5] D. S. Taubman and M. W. Marcellin, "JPEG 2000: Image Compression Fundamentals, Standards and Practice," Kluwer International Series in Engineering and Computer Science,
- [6] R. Xu, S. N. Pattanaik, C. E. Hughes, "High-Dynamic-Range Still-Image Encoding in JPEG 2000. IEEE Computer Graphics and Applications 25(6): 57-64, 2005.
- [7] S. B. Kang, M. Uyttendaele, S. Winder, R. Szaliski, "High Dynamic Range Video", ACM Transactions on Graphics, Volume 22, Issue 3 pp. 319 - 325, 2003.
- [8] R. Mantiuk, G. Krawczyk, K. Myszkowski and H. P. Siedel, "Perception-Motivated High-Dynamic Range Video Encoding, ACM Trans. on Graphics, col23, no.3 2004, pp.773-741.
- [9] R. Xu, S. N. Pattanaik, and C. E. Hughes, "High Dynamic Range Image and Video Data Compression", Technical Report, School of Engineering and Computer Science, University of Central Florida, CS-TR-05-13, 2005.
- [10] K. E. Spaulding, G. J. Woolfe, R. L. Joshi, "Using a Residual Image to Extend the Color Gamut and Dynamic Range of an sRGB Image", The PICS Conference, p. 307-314, 2003
- [11] Ward, Greg, and Maryann Simmons, "Subband Encoding of High Dynamic Range Imagery," First Symposium on Applied Perception in Graphics and Visualization (APGV), August 2004.
- [12] R. Mantiuk, A. Efremov, K. Myszkowski, H.P. Seidel. "Backward Compatible High Dynamic Range MPEG Video Compression". ACM SIGGRAPH 2006
- [13] L. Spillmann, J.S. Wener, "Visual Perception, the Neurophysiological Foundations," Academic Press, San Diego, 1990. .
- [14] S. N. Pattanaik, "High Dynamic Range Imaging", in Tutorial notes, ACM-SIGGRAPH 2004, Los Angeles, Aug 2004.
- [15] S. Daly, "The Visible Difference Predictor: An algorithm for the assessment of image fidelity," in A.B. Watson, editor, Digital Image and Human Vision, pages 179-206, Cambridge, MA, MIT Press, 1993.

³These data are provided by the authors

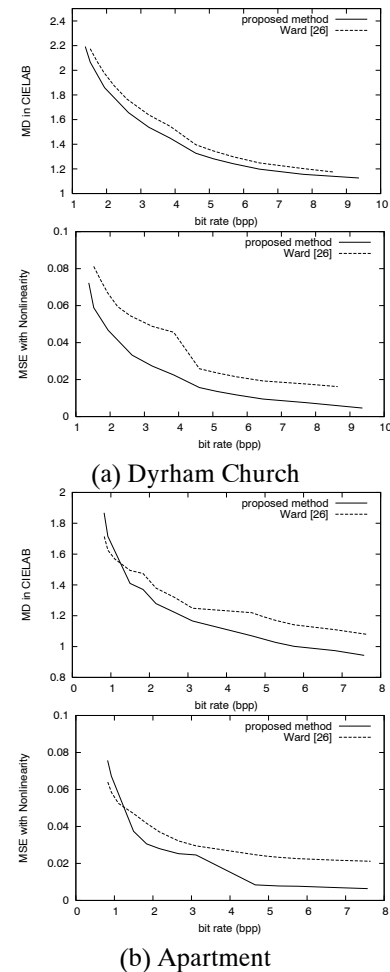


Fig. 4. Comparison with Ward et al.'s method [?]

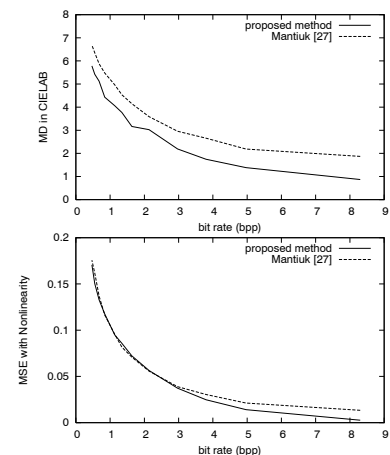


Fig. 5. Comparison with Mantiuk et al.'s method [?]