RAW IMAGE ENCODING BASED ON POLYNOMIAL APPROXIMATION

Masahiro Okuda[*]

The University of Kitakyushu
Dpt. of Information and Media Sciences

Nicola Adami[†]

University of Brescia
Dpt. of Electronic for the Automation

ABSTRACT

In this paper, we propose a coding algorithm for raw images with high dynamic ranges. Our encoder has two layers. In the first layer, 24 bit low dynamic range image is encoded by a conventional codec, and then the ratio image that represents the difference between the decoded 24 bit image and the raw image is encoded in the second layer. Experiments shows compression efficiency is significantly improved by taking an inverse tone mapping into account.

***Index Terms*—** Raw Image, Compression, Polynomial Approximation

1. INTRODUCTION

Digital cameras generally use CCD or CMOS image sensors. The charge captured by these sensors during exposure time is read out and converted to pixel values with some processing such as white balance adjustment, filtering, color transform, and compression [1]. Then the digital images are stored mostly in JPEG format. The image sensors in most of the digital cameras inherently have a capability of capturing higher dynamic range and many of recent models can provide higher dynamic range images as well as the JPEG, the so called 'raw images'. The raw images can be seen as 'scene-referred' because no image processing is done and the stored pixel values are linearly proportional to actual scene radiance [1], [2]. Besides, films used by conventional analog cameras, both the negatives and the positives, can represent higher dynamic range. One can obtain the scene-referred images with high-end film scanners.

On the other hand, the conventional 24 bit image format is called 'output-referred'. Before storing it in JPEG, the color is transformed to the output-referred color formats such as sRGB, which have nonlinearity to compensate CRT monitors' gamma curves. Then the dynamic range of the images are reduced to 8 bits per color by quantization. These operations discards a significant amount of information.

The raw image is often preferable since it allows users control settings such as white balance, contrast, sharpness to obtain desired effect easily. With the growth of performance of CPU, display, and printers, the demand for the high quality images is increasing. Since most of the raw images have 12 to 16 bits for each color component that is equivalent to 16 to 256 times higher dynamic range, image compression is essential. The image coding standard, JPEG 2000, provides seamless compression from 1 to 16 bits per color channel [3]. Ruifeng et al.'s scheme [4] verifies the validity of JPEG2000 for the high dynamic range image compression. Spaulding [5] proposes two layer encoding for gamut extended images. In the first layer, an image with clipped gamut (output-referred image) is encoded. And then in the second layer, the residual information that represents the difference between an gamut extended image and the decoded image in the first layer is encoded. The main advantage of this approach is that the format is compatible to existing file formats, and no extra efforts are needed to extract the output-referred image. In the field of Computer Graphics, similar concepts are adopted for the high dynamic image compression [6], [8].

In this paper, we extend and elaborate on the conventional methods [5], [6], [8]. We approximate by polynomials a tone mapping function that transforms a scene-referred raw image to its corresponding output-referred 24 bit image. We encode the difference between the raw image and the tone mapped image. The polynomial approximation significantly improves coding efficiency.

2. RELATED WORK

The technologies of in-camera processing and developments are usually camera-specific and not open to the public. Moreover the raw file formats used by camera manufacturers often have poor compression performance. The two layer coding method previously proposed by [5] addresses these problems. In this framework, a pair of a raw image and its tone mapped 24 bit image is given. They first compress the 24 bit image encoded by using JPEG, and then a residual image that repre-

[*]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.

sents the difference between the two images is encoded. Although large error occurs in very bright and dark regions, it removes correlations in large portions of the image. Thus the residual image is highly compressible. Ward et al. [6] adopts a similar method to the high dynamic range image encoding. They compute a ratio image instead by dividing the high dynamic range image by its 24 bit version and then it is encoded. Mantiuk et al. applies the approach to high dynamic range video coding[8].

Although these methods achieve high rate-distortion performance, they pay little attention to tone mapping operation that transforms the raw image to the 24 bit version. When the tone mapping that includes the in-camera processing and developments is estimated well, higher compression may be possible. To our knowledge, only [8] tackles this problem. They compute the look up table (LUT) that represents the tone mapping and then compensate the difference of the two images. Although the LUT well approximates the curve of the tone mapping, the resulted curve is not smooth. The lack of smoothness may yield extra energy in high frequency, which is not desirable in a sense of compression efficiency. The remaining of this paper describes our algorithm based on polynomial approximation and shows our method outperforms the conventional methods.

3. PROPOSED ALGORITHM

The outline of our encoder is depicted in Fig.1. Just like the conventional methods mentioned above, our algorithm starts with the two images, a raw image and its 24 bit tone mapped version. The output is a bitstream that contains the compressed 24 bit image and residual information that represents the differences between the two inputs. We do not impose any restrictions on the tone mapping. We only assume that the 24 bit image is transformed from the raw image by in-camera processing, user specified image processing, etc.

The encoder first compresses the 24 bit image by JPEG, and then decompress it We denote the luminance channels of the raw image and its 24 bit version as I_R and I_L, respectively. The decoded one of I_L is denoted by $\hat{I}_L$. The decompressed luminance $\hat{I}_L$ is then transformed to $f(\hat{I}_L)$ by a polynomial model described later. In the final step, the residual image is encoded by a wavelet based image encoder.

3.1. Polynomial Approximation

The mapping function from I_R to I_L in general depends on imaging systems such as cameras and film scanners, and development procedures. The paper [9] examines various types of cameras to find camera response functions that map radiance to output referred color space. According to [9], there is considerable variation between response curves. Moreover in a development process, the user may intentionally change

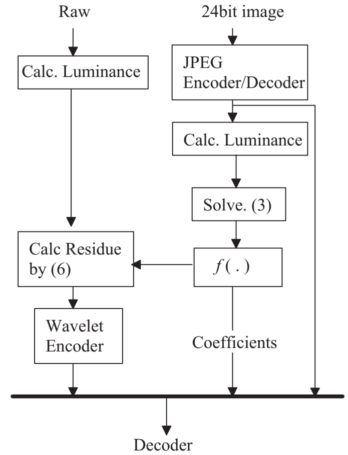


Fig. 1. Encoder

tonal curves of the images. For these reasons, in order to flexibly approximate the functions, we adopt a polynomial model,

$$f(x) = \sum_{n=0}^{N} a_n x^n \tag{1}$$

In addition to minimizing error between an actual mapping function and the polynomial model, we wish to make the approximation monotonic. In most of actual cases, the mapping function monotonically increases, which assures one-to-one mapping. Our approximation, however, does not impose the monotonic property to f, since the conditional approximation problem requires numerical optimization methods with more computational complexity. Instead we incorporate a smoothness term to the formulation of a squared error function to be minimized. It is essential to take the smoothness into consideration, since an obtained curve with some jaggy would bring extra high frequency components to smooth regions of the image, which makes the compression of the residual image inefficient. Then, the approximation problem is stated as the minimization of an error function:

$$E = \sum_{l=0}^{L-1} \sum_{n=0}^{N} \left\{ I_R(l) - f(\hat{I}_L(l)) \right\}^2 + \alpha \sum_{k=0}^{K} f'(k/K)^2, \tag{2}$$

where l and L are a pixel index and the number of pixels, respectively, and K is an integer grid density. Note that in our setting the 24 bit image I_L is normalized by 1, that means the domain of f is in the range $[0,\ 1]$. The weight α balances the trade-off of the two terms, the preciseness and the smoothness. In our experiment $\alpha = 0.01$ is used. To make the function f smoother one can add second or higher degree of derivatives, but in our experiments the minimization of higher degree of derivatives does not improve final compression performance very much.

Then eq.(2) is a quadratic form with respect to the coefficients a_n, the minimization is achieved by solving a set of linear equations,

$$\frac{\partial E}{\partial a_n}, \quad n = 0, 1, 2, \cdots, N. \tag{3}$$

We have confirmed that the order N ranging from 3 to 5 give satisfactory results. For most examples we have tested, in the result the monotonic property is satisfied with the help of the smoothness term. A poorer performance would be obtained with lower degree, while a higher degree polynomial tends to oscillate too much and be non-monotonic.

3.2. Residual Image

After the mapping function is designed, we exploit the correlation of the two images, I_R and $f(\hat{I}_L)$. There are two options to remove the correlation from the raw image I_R, calculating their difference:

$$I_E = I_R - f(\hat{I}_L) \tag{4}$$

or their ratio:

$$I_E = \frac{I_R}{f(\hat{I}_L)}. \tag{5}$$

The former is used by [5] and the latter is first introduced in [6], and are also employed by [8]. The ratio (5) is often used in the CG community for the compression of the so-called high dynamic range image whose dynamic range is generally much higher than one of the raw image. The ratio image works well for this application mainly due to the two reasons. First, as is pointed out in [6], the difference (4) does not recover very high dynamic range unless nonlinear quantization is used, while the ratio (5) does. The second reason is that the form of (5) implicitly weight dark regions of the mapped image $f(\hat{I}_L)$. This weighting matches the human perception model in that the sensitivity of visual perception decreases with respect to stimuli as is indicated by the Weber-Fechner's law [10]. For these reasons, we use the ratio (5) rather than (4) to obtain the residual image I_E.

3.3. Smoothing Operator

Although the mapping process reduces the dynamic range of the raw images and discard a large amount of information, large portions of the image suffer only a small information loss and the two images, I_R and I_L, are highly correlated. Accordingly the ratio image is not very compressible since it has a large amount of energy in high frequencies. In order to alleviate this problem, we employ a lowpass filter to the transformed image $f(\hat{I}_L)$. Fig.2 depicts the effects of this operation. Smoothing $f(\hat{I}_L)$ removes some high frequency components of the ratio image, resulting in a significant improvements in compression efficiency. Then finally the ratio image we encode is written by

$$I_E = \frac{I_R}{s(f(\hat{I}_L))}, \tag{6}$$

where $s(\cdot)$ denotes the lowpass operator.

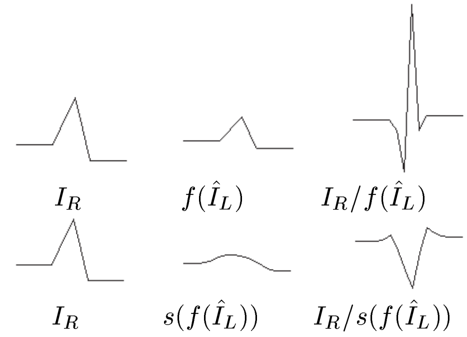


Fig. 2. Smoothing Operator

3.4. Color Reconstruction

In [6], G. Ward et al. point out that if the Tone Mapping (TM) operation preserves hue of an HDR image, and its saturation change is estimated, the color of the original HDR can be fairly recovered by its low dynamic range image by the saturation compensation scheme. Our algorithm adopts a different color compensation scheme. Assume the relation between that RGB values of the raw image and its 24 bit version is roughly approximated by a polynomial, then we simply apply this to each color channel of the 24 bit image. That is, we design three polynomials for RGB channels in the encoder by using the above approximation method, then in the decoder reconstructs each color channel by

$$\hat{I}_R^r(i,j) = \hat{I}_L^r(i,j)\frac{\hat{I}_R(i,j)}{\hat{I}_L(i,j)}, \quad \text{for channel R,} \tag{7}$$

where $\hat{I}$ is a decoded version of I, and the superscript r means the red channel. The same procedure is done for the green and blue channels as well. This procedure maps the color of the 24 bit image to the raw image. Since one does not need to send the color of the ratio image, it is efficient in a sense of coding efficiency. Especially it works well in low bit rates. In high bit rates, however, it often gives better results to encode colors than to spend all bits only for the luminance. The color encoding is required when the tone mapping operator drastically changes the colors.

3.5. Experimental Results

We have tested dozens of raw images captured by eight cameras of four manufacturers. Since the raw 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 squared error in CIELAB color space and VDP-HDR metrics [11]. VDP-HDR is an error estimator designed for the high dynamic range images. This metric mimics the human perception system and estimates how much fraction of pixels are perceived different for two given images.

First we compare the mse in CIELAB space with [6][1], shown in Fig. 3. From low to high bit rates, our method gives better results. Next we plot a comparison with the most recent HDR coding method [8] in Fig.4. The method [8] is designed mainly for HDR video compression, but it also gives high performances for still images. The solid and dotted lines indicate the results of our method and the lut-based reconstruction function (RF) in [8]. The results are the average of dozens of tested images. Our gain in compression efficiency is mainly due to the use of the polynomial approximation and wavelet-based coding. The same settings are used except for the polynomial approximation in order to verify the validity of our hill functions. From these figures, it can be seen that our algorithm gain significant improvements. Note that since we use polynomial approximation and wavelet codec, its computational complexity is higher than the conventional methods. In our experiments, the algorithm takes a few seconds for 768x512 images.

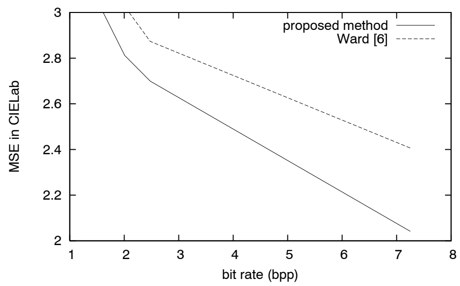


Fig. 3. Comparison with Ward et al's method [6]

4. REFERENCES

[1] J. Nakamura, " Image Sensors And Signal Processing For Digital Still Camera ", Taylor & Francis Group, 2005

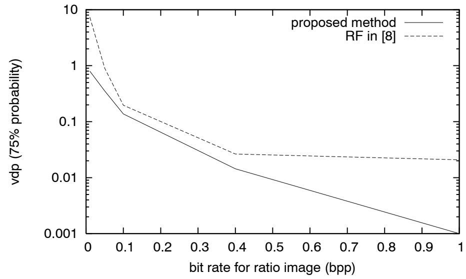
[2] Berthold K. P. Horn, "Robot Vision", MIT Press, MIT-Press, McGraw-Hill, 1986

[3] David S. Taubman and Michael W. Marcellin, JPEG 2000: Image Compression Fundamentals, Standards and Practice, Kluwer International Series in Engineering and Computer Science,

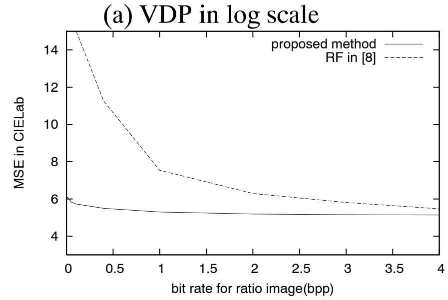
[4] Ruifeng Xu, Sumanta N. Pattanaik, Charles E. Hughes: High-Dynamic-Range Still-Image Encoding in JPEG 2000. IEEE Computer Graphics and Applications 25(6): 57-64 (2005)

[5] Spaulding, Kevin E.; Joshi, Rajan L.; Woolfe, Geoffrey J., "Using a residual image formed from a clipped limited color gamut digital image to represent an extended color gamut digital image", United States Patent 6301393

[1] We use a software provided by the authors. Since the software does not provide results for low bit rates, we cannot see much difference in VDP-HDR metric.



(a) VDP in log scale



(b) Mse in CIELAB

Fig. 4. Comparison with Conventional Method [8]. In these, y-axis shows the average of dozens of test images, and bit rate in x-axis is a bit rate spent only by the ratio image coding.

[6] Ward, Greg, and Maryann Simmons, "JPEG-HDR: A Backwards-Compatible, High Dynamic Range Extension to JPEG," Proceedings of the Thirteenth Color Imaging Conference, November 2005.

[7] 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.

[8] R. Mantiuk, A. Efremov, K. Myszkowski, H.P. Seidel. "Backward Compatible High Dynamic Range MPEG Video Compression". To appear in: Proc. of SIGGRAPH '06

[9] M.D. Grossberg and S.K. Nayar,"What is the Space of Camera Response Functions?,"IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Vol.II, pp.602-609, Jun, 2003.

[10] Spillmann, Lothar; Wener, John S, "Visual Perception, the Neurphysiological Foundations," Academic Press, San Diego, 1990.

[11] Rafal Mantiuk, Karol Myszkowski, Hans-Peter Seidel. Visible Difference Predicator for High Dynamic Range Images. In: Proc. of IEEE International Conference on Systems, Man and Cybernetics, 2004. pp. 2763-2769