A BIT-DEPTH SCALABLE VIDEO CODING APPROACH CONSIDERING SPATIAL GRADATION RESTORATION

Masaru TAKEUCHI[1], Yasutaka MATSUO[1,2], Yuta YAMAMURA[1], Jiro KATTO[1], Kazuhisa IGUCHI[2]

[1] Department of Computer Science, Waseda University
[2] NHK Science & Technology Research Laboratories

ABSTRACT

Bit-depth scalable coding method is an approach that generates multiplexed bit-streams that can be decoded to two video sequences, for Standard Dynamic Range (SDR) environment and for High Dynamic Range (HDR) environment. This paper presents a bit-depth scalable coding method that is considered as gradation restoration in inter-layer prediction process. Our proposed inter-layer prediction uses a histogram interpolation method that enables to generate more mid-gray brightness levels than traditional inverse tone mapping methods with one-to-one correspondence.

Index Terms— High Bit-depth, Bit-Depth Scalable Video Coding, High Dynamic Range, Inverse Tone Mapping.

1. INTRODUCTION

Nowadays, imaging devices like displays and cameras have made remarkable progresses and have enabled us to represent higher definition and higher dynamic range (HDR) images. Imaging devices that can handle higher bit-depth images, for example ten to twelve bit-per-sample, have begun to be more familiar. Moreover, interface systems which support transmitting high bit-depth images such as Displayport and HDMI, have already been standardized.

Digital video contents represented by eight-bit-per-sample still have been popular and widely used in numerous applications such as digital television, DVD, camcorder, video conference, etc. However, the eight-bit, 256 levels may not be sufficient to represent the brief details of luminance and color information of HDR images that are captured from the real world. Hence, it has been expected that image and video contents that are represented by high bit-depth become more popular.

In addition, the technique for properly displaying an HDR video on a SDR (i.e. eight bits-per-pixel) display device has been studied, which is referred as tone mapping (TM) [1]. Due to the different capabilities of end display devices, an HDR video source is expected to be displayed in various bit-depths. Hence, a bit-depth scalable video coding solution is required, and some approaches have been proposed during the international standardization work known as the Joint Video Team (JVT) and Video Coding Expert Group (VCEG) [2,3,4,5,6].

In the bit-depth scalable video coding solution, eight-bit SDR video sequences that are generated from HDR sequences for a backward compatibility, are coded as the base layer bit-stream. HDR, high bit-depth video sequences are predicted from the decoded images of the base layer stream in the local decoder using inverse tone mapping (ITM) [7], that is the backward process of the TM. The residue signals and side information that are necessary for the ITM, are coded in the enhancement layer.

However, conventional studies about the ITM used in inter-layer prediction are not considering restoration of gradation information. In this study, we propose a bit-depth scalable coding method using well-optimized inter-layer prediction for gradation restoration.

2. OVERVIEW OF BIT-DEPTH SCALABLE VIDEO CODING

A simple solution for coding HDR video sequence into one bit-stream from which the sequence for HDR environment and the one for SDR environment can be decoded, is simulcast like Figure 1.

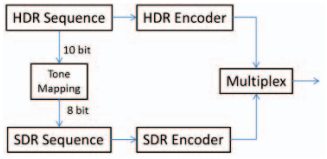

Fig. 1. Simulcast.

The sequences for SDR environment are generated from HDR sequences using TM methods. In this solution, two bit-streams are simply multiplexed into one bit-stream. However, two sequences decoded from each stream might be clearly similar to each other, despite they differ from each other in bit-depth and target environment. Thus, this simulcast solution has much redundancy.

To avoid this redundancy, the methods called bit-depth scalable video coding have been proposed in recent years [2,3,4,5,6]. Its general encoder structure is shown in Fig. 2.

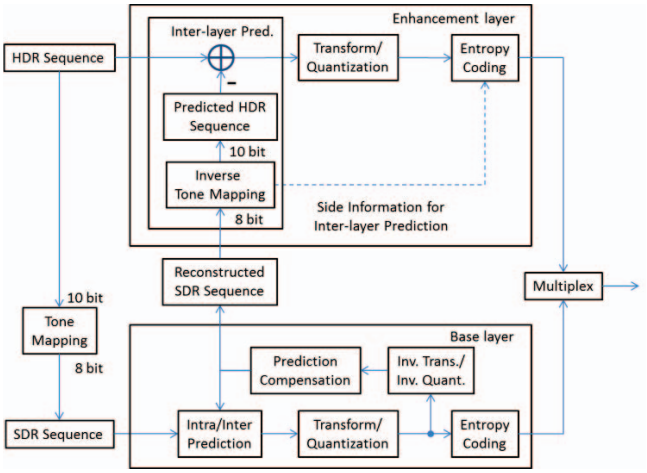

Fig. 2. Bit-depth scalable encoder with one motion compensation loop.

The encoder codes an SDR sequence as the base layer. On the other hand, it doesn't simply code an HDR sequence. In the enhancement layer, it codes residual signals between an original HDR image and a predicted HDR image obtained from locally decoded base layer images.

Bit-depth scalable approaches use the ITM in the inter-layer prediction process to predict the HDR signal from the reconstructed SDR signals decoded from the base layer bit-stream. In the ITM process, HDR signals are recovered from SDR signals according to the tone mapping operator (TMO) like scaling factors and/or look-up-tables. Information necessary for the inter-layer prediction like TMO must be also coded as side information.

However, it could be pointed out that ITM methods using scaling factors and/or look-up-tables have problems that they cannot generate HDR levels enough to represent spatial gradation in HDR images because a SDR level is corresponding to only one HDR level. Figure 3 shows an example of a predicted signal and its residue.

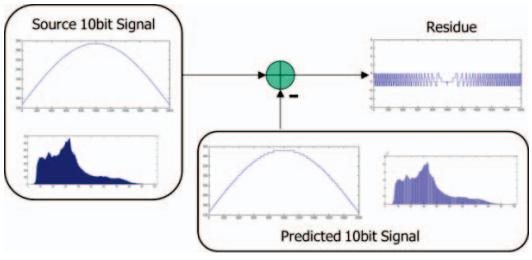

Fig. 3. An example of histograms of source and predicted signals and resultant residue signal.

An example of source signal that has smooth waveform is shown at left side, and its histogram that has smooth distribution is also shown under its waveform. Pixel values of the predicted signal are biased according to a scaling factor and/or a look-up-table, so that the histogram looks like a comb and a waveform has no smooth gradation. The residue that must be coded as the enhancement layer has a complex form that is not easy to be coded. To reduce the power of residue, the inter-layer prediction method should generate more mid-levels and the histogram of a predicted signal must be much similar to the one of an original signal. In next section, we propose a method that considers the restoration of spatial gradation by attempting to bit-depth transformation with one-to-many correspondence.

3. PROPOSAL METHOD

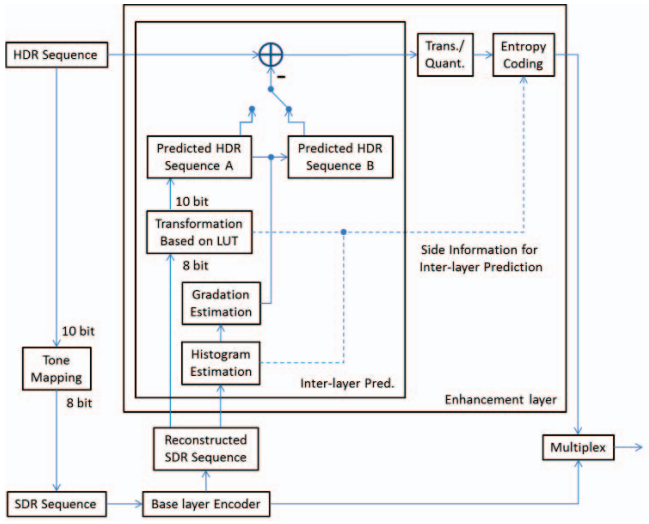

Fig. 4. The framework of the proposed method.

Figure 4 shows the framework of the proposed approach. Our proposal estimates HDR signals by two means. The one is simply generated by using a look-up-table. The other is by the method called histogram estimation and gradation estimation for generating well-gradation-restored HDR signals. Then, a predicted signal, that the enhancement layer can code as a residue more efficiently, is adaptively chosen and coded with flag information overhead. Later, details of our proposal are explained.

3.1. Histogram estimation(HE)

The goal through two processes, histogram and gradation estimations, is to generate a predicted signal that has a histogram similar to the one of original. The original could be estimated from the comb-like histogram to some extent. The simplest solution is a linear interpolation method as Figure 5 shows.

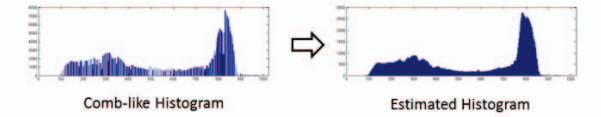

Fig. 5. Estimated histogram.

3.2. Gradation estimation(GE)

This process actually generates a predicted HDR signal that has a histogram estimated in the previous process. For this purpose, our process replaces predicted HDR pixel values with new values, which are varying and follow the estimated histogram.

First, to make distinction among the pixels that have the same pixel values, we calculate the evaluation values for each pixel on an output image with considering spatial information. Evaluation value R is given by

$$R(i,j) = X(i,j) + \frac{\mathrm{Step}(X(i,j))}{2^{HDR\_bitdepth}-1} \cdot O(X(i,j)) \quad (1)$$

where (i,j) is a pixel coordinate, X is predicted HDR signal generated by simple LUT transformation, $\mathrm{Step}(k)$ is the value range of new pixel values which k is replaced with, $HDR\_bitdepth$ is the bit-depth in the HDR signal, $O(\cdot)$ is the kernel to add the information of neighboring pixels, where we use a Gausian kernel, of which size is 7 and variance is 0.45.

Owing to the second term of Eq.(1), pixels that have the same values in X have different values in R. Then, we assign new values based on the estimated histogram. The higher R the pixel has, the higher values in the estimated histogram is assigned.

3.3. Blockwise mode selection

By comparing two prediction methods, with GE/HE and without GE/HE, we figured that the power of residue with GE/HE tended to decrease in spatial gradation area as Figure 6 shows.

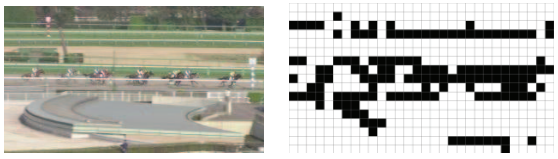

Fig. 6. An example of evaluation between two prediction methods for each 64x64 block. (Left: the target frame, Right: the result)
White: the prediction errors will decrease with HE/GE. Black: the prediction errors will increase with HE/GE.

For the purpose of decreasing the prediction errors, we adopt block-wise prediction mode selection. However, selected modes for each block must be also coded as side information. The more blocks a frame is divided into, the less MSE we could achieve and the more bits the selection information requires.

To cope with this problem, we let our method enable to support variable blockwise selection. Each variable block has a quad-tree structure. Each structure of the variable block is decided as the result of optimization (i.e. rate-distortion optimization) that is based on the cost function J, which J is given by

$$J = \mathrm{SSD} + \lambda \cdot \mathrm{Genbit} \quad (2)$$

where λ is a Lagrange multiplier which is temporary defined as $0.85 \times 2^{\frac{QP}{3}}$, QP is the QP value of I slice in base layer encoder, SSD is the sum of squared differences, Genbit is generated bits for describing structure of current block.

4. RESULTS

4.1. Experiment conditions

Our experiments use three video sequences shown in Figure 7. All sequence are 1920x1080 size in YCbCr 4:2:0 format, their frame rates are 30 frames/sec, and target bitrates in base layer are varied from 30 to 60 Mbps. target bitrates in enhancement layer is set to 10 Mbps. We use JM Encoder [8] as the base layer encoder and also as the residue encoder in the enhancement layer. The GOP structure in base layer is IBPBPBPBPBPBPBP. We use non-linear TMO that is generated by Lloyd-max algorithm [9] to reduce errors caused by TM process for luminance plane, and linear TMO for chroma planes. The SDR base layer is eight bit-per-sample and HDR enhancement layer is in ten bit-per-sample. We evaluate two different proposed inter-layer prediction methods. The first is the proposed method without blockwise mode selection. This method applies HE/GE to an entire frame. The other is the variable blockwise method. We set the maximum size of process unit to 256x256, and maximum search depth to 3. Each proposal is compared with the method without HE/GE as the anchor.

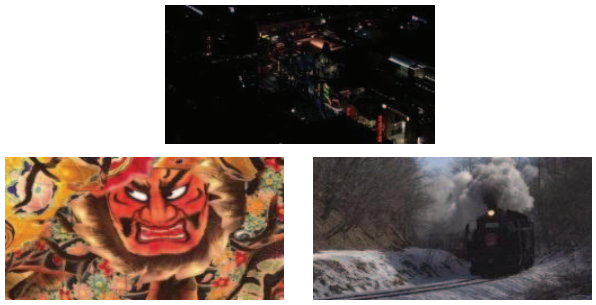

Fig. 7. Video sequences used in experiments (Night, Nebuta, Steam Locomotive Train)

4.2. Results and discussions

The BDRATE [10] evaluations against the anchor are shown in Table 2.

From the result of Table 2(a), we can recognize that coding efficiency has gained by using the proposed inter-layer prediction method. In the proposal A, the average BDRATE reduction of the inter-layer predicted HDR signal is 1.186[%] against the anchor. That of finally decoded HDR signal is 1.243[%]. From this result, it can be concluded that the proposed inter-layer prediction method

Table 2: BDRATE gains against the anchor [%].

(a) Proposal A (without blockwise selection)

Sequence No.	Night	Nebuta	SLT
Predicted	-0.610	-0.958	-1.990
Decoded	-1.269	-0.549	-1.912

(b) Proposal B (with variable blockwise selection)

Sequence No.	Night	Nebuta	SLT
Predicted	-0.795	-0.958	-2.011
Decoded	-0.898	-0.546	-1.899

Predicted: Inter-layer predicted HDR signal from reconstructed SDR signal. Decoded: Finally decoded HDR signal.

doesn't only reduce the power of the residue signal but also contributes to gaining coding efficiency of the residue signal.

To evaluate subjective quality of the decoded images using HE/GE, we compare those of the anchor and our proposal. Figure 8 shows detailed parts of these images, of which contrasts are enhanced. The anchor image seems to have some artifacts like block noise. On the other hand, our proposal, with HE/GE, has less noise and represents spatial gradation well.

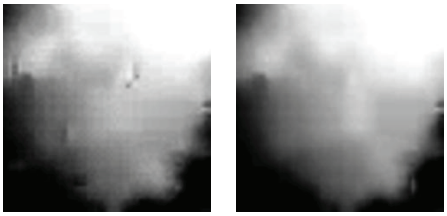

Fig. 8. Parts of decoded HDR images stretched contrast Left: the anchor(without HE/GE), Right: Proposal A

In Table 2(b), the result of the proposal B with variable blockwise selection is shown. Compared with the evaluation of predicted HDR signal by the proposal A, proposal B wins at all sequences. However, these gains aren't so much. To describe this situation, we show the actual structure obtained as the result of optimization in Figure 9. We can recognize that selected prediction modes are biased to white, with HE/GE. This tendency is not only at this frame, seen in almost all sequences under the experiment condition. We thought that these bias disadvantages the blockwise method that requires additional useless flag information in this situation.

Fig. 9. The example of selected prediction method and optimized variable block structure.

On the other hand, the evaluation of finally decoded video loses against the proposal A. This could be due to the approach of variable block structure optimization. The cost function used at this time for deciding the suitable quad-tree structures is optimized to reduce BDRATE of the inter-layer predicted HDR signal and not the one of finally decoded pictures.

5. CONCLUSIONS

This paper proposed an inter-layer prediction approach and a bit-depth scalable coding method considering spatial gradation restoration using the inter-layer prediction. By applying the prediction method that is specialized for generating mid-levels, coding efficiency is gained. In addition, we also verify that visual quality of the decoded images is gained subjectively. Furthermore, we let our method enable to support variable blockwise adaptive mode selection, and the experimental results show BDRATE reductions HDR signal prediction. As future work, we will need more study about the mode optimization, and also the process after the inter-layer prediction.

6. REFERENCES

[1] E. Reinhard et al., "Photographic Tone Reproduction for Digital Images," *ACM Trans. Graphics*, vol. 21, pp. 267–276, July 2002

[2] M. Winken et al., "Bit-Depth Scalable Video Coding," *Proc. ICIP*, pp. 5-8, Sept.-Oct. 2007

[3] Shan Liu et al., "Bit-depth Scalable Coding for High Dynamic Range Video" *VCIP*, vol. 6822, Jan. 2008

[4] A. Segall, "Scalable Coding of High Dynamic Range Video" *Proc. ICIP*, pp.1-4, Sept.-Oct. 2007.

[5] Y. Gao, et al., "H.264/Advanced Video Coding (AVC) Backward-Compatible Bit-Depth Scalable Coding," *IEEE Trans. Circuits and Systems for Video Technology*, vol.19, no.4, pp.500-510, April 2009

[6] Y. Wu, et al., "Bit Depth Scalable Coding," *IMCE2007*, pp.1139-1142, 2-5 July 2007

[7] Francesco Banterle et al., "A framework for inverse tone mapping" *The Visual Computer: International Journal of Computer Graphics*, vol. 23, Issue 7, May.2007

[8] JM 18.0: http://iphome.hhi.de/suehring/tml/

[9] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Information Theory*, vol. IT-28, pp. 129-136, Mar. 1982

[10] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves", Doc. VCEG-M33, Apr. 2001