BIT-DEPTH SCALABLE VIDEO CODING USING INTER-LAYER PREDICTION FROM HIGH BIT-DEPTH LAYER

Jui-Chiu Chiang, Wen-Ting Kuo

Department of Electrical Engineering
National Chung Cheng University, Chia-Yi, 621, Taiwan, ROC.
rachel@ee.ccu.edu.tw, wtkuo@samlab.ee.ccu.edu.tw

ABSTRACT

Scalable video coding (SVC) is currently developed as an extension of H.264/AVC video coding standard. In this paper, we propose three H.264/AVC compliant bit-depth scalable video coding schemes, named LH mode (Low Bit-depth to High Bit-depth), HL mode (High Bit-depth to Low Bit-depth) and combined LH-HL mode for different applications. All these schemes efficiently exploit the high correlation between the high bit-depth layer and the low bit-depth layer on Macroblock level. Experimental results indicate that the HL mode outperforms the other two schemes and it achieves up to 7 *dB* improvement over the simulcast where the high bit-depth video and low bit-depth representations are 12-bit and 8-bit, respectively.

Index Terms—Scalable video coding, Bit-depth, High dynamic range, Inter-layer prediction

1. INTRODUCTION

The need to transmit digital video/audio over wired/wireless channels increases with the mature development of multimedia processing and also the wild deployment of internet service. However, due to the unstable channel condition, the successful delivery over network can not be ensured always. To make the transmission over heterogonous networks more flexible and efficient, scalable video coding [1-2] was proposed to overcome such situations and aim at improving transmission efficiency. The idea behind SVC is to provide a fully scalable video bitstream in terms of spatial, temporal, and quality scalability. Currently, the acquisition of high dynamic range (HDR) image has become easier than before. In addition, HDR images receive considerable attention in many practical applications [3]. For example, in HDMI (High Definition Multimedia Interface) 1.3, the supported bit-depth has been extended from 8 bit up to 16 bit per channel, which can definitely offer human being an impression with more reality. Thus, JVT has issued a "Call for Proposals" to standardize bit-depth scalable video coding and make it compliant with current H.264 [4] and SVC specifications. However, the bandwidth required for the encoded high bit-depth image/video is huge undoubtedly. Besides, not all the displayer can present the HDR video on the receiver side. Therefore, it is desirable to design proper algorithms to overcome these problems.

Many efforts [5-10] were made to develop efficient bit-depth scalable video coding in the past few years. In [7], Winken *et al* proposed a coding method where the high bit-depth video sequence is first converted into a low bit-depth format, which can be regarded as the base layer and it will be encoded by H.264 standard. Then the reconstructed base layer image can be inversely processed as a prediction for the high bit-depth image. The difference between the original high bit-depth image and the predicted one is treated as the enhancement layer. Segall [8] proposed another scalable bit-depth video coding algorithm on Macroblock (MB) level. Similarly, the base layer in this scheme is generated by tone mapping of the high bit-depth input and will be encoded by H.264. For the high bit-depth input, not only inter/intra prediction is used to remove the redundancy, but also the inter-layer prediction is exploited. Besides, both high bit-depth and low bit-depth layers use the same motion information estimated in the base layer. In [9-10], an implementation focusing on combined spatial and bit-depth scalability has been addressed. To achieve a better coding efficiency, Wu *et al* [9] recommend that the inverse tone mapping should be realized before the spatial upsampling. Moreover, the residual of low bit-depth layer will be upsampled and regarded as a prediction for the residual of high bit-depth layer [10]. In this way, the redundancy removal performs better than the works in [7-8]. Up to now, those bit-depth scalable coding schemes use the low bit-depth information to predict the high bit-depth layer. In this paper, we will present an approach where the source for inter-layer prediction is the high bit-depth layer, not the low bit-depth one. We believe that the information contained in the high bit-depth layer is more accurate than that in the low bit-depth layer. Thus, a better coding efficiency is expected to predict the low bit-depth layer from the high bit-depth layer.

This paper is organized as follows: Section 2 introduces the LH mode, which is similar to most current methods. Section 3 presents the HL mode and the combined LH-HL mode in detail. Section 4 demonstrates experimental results and Section VI contains the conclusion.

2. LH MODE SCHME

In attempt to make the generated bitstream embedded and compatible to H.264 standard, the current bit-depth scalable coding schemes almost employed inter-layer prediction information from the low bit-depth layer. The proposed LH mode adopts this idea, and some modifications compared to other methods will be described later.

The coding scheme of the LH mode is shown in Fig. 1. In our experiment, the low bit-depth input used is 8 bits for each color channel while 12 bits for the high bit-depth input. The low bit-

depth input is obtained after tone mapping of the high bit-depth input and followed by H.264 encoding. Thus, the generated bitstream allows backward compatibility with H.264.

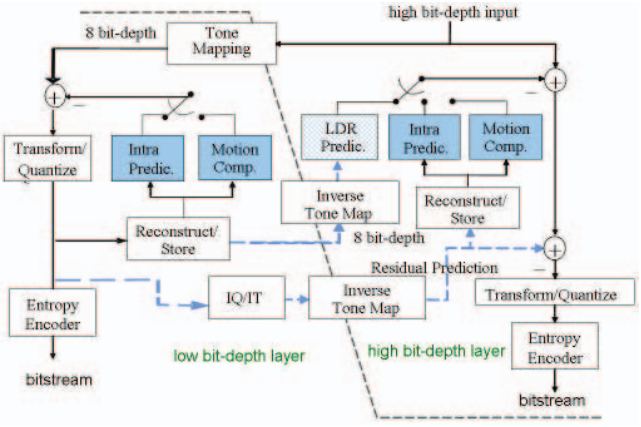

Fig.1 LH mode coding scheme

The right side of Fig. 1 describes the coding procedures for the high bit-depth layer in detail. The same as the low bit-depth layer, the encoding process is carried out on MB level. However, there are three differences compared with the low bit-depth layer. First, in addition to intra/inter prediction, the high bit-depth image has another prediction from the low bit-depth layer by inverse tone mapping of the reconstructed low bit-depth MB. This inter-layer prediction can be seen as the extra intra prediction mode and it is called LDR (low dynamic range) prediction here. Besides, to achieve a better coding efficiency, the residual of the low bit-depth layer will be inverse tone mapped to be further utilized as a prediction for the residual of the high bit-depth layer. It is called residual prediction, which can be applied in two ways. The high bit-depth MB may perform motion estimation after subtracting this residual prediction from its original data or it may subtract this residual after motion compensation. Both methods aim to reduce the redundancy within residuals of the low and the high bit-depth layers. Moreover, different from other approaches, both the low bit-depth layer and the high bit-depth layer have its own motion information in the proposed LH mode.

Fig.2 Bitstream structure of LH mode

2.2 Bitstream structure of LH mode

The bitstream of the LH mode is embedded; it implies that a reasonable truncation of the bitstream can ensure successful reconstruction of low bit-depth images.

Fig. 2 presents one possible setting of LH mode bitstream structures. LDR_I means low bit-depth I frame information. LDR Motion Info and LDR_P stand for the motion information and all associated data for low bit-depth P frames, respectively. It reveals that the generated bitstream in LH mode is backwards compatible to H.264 and it could be extended to include higher bit-depth information as an enhancement layer. For example, the following components are HDR_I_Refine, HDR Motion Info and HDR_P_Refine. They represent the refined information of enhancement layer.

3. HL MODE SCHME

In this section, we propose a new concept for some practical applications. This scheme is called the HL mode, due to the fact that the high bit-depth data is encoded first, and it will provide the low bit-depth layer with some useful information after suitable tone mapping. Consider the situation where channel bandwidth is not a problem and both bit-depth videos can be received by the decoder. In addition, the displayer supports the high bit-depth format. The HL targets at achieving a good coding performance in this situation. If only 8-bit displayer is available in the receiver side, a truncated bistream still guarantees a successful decoding of the low bit-depth video.

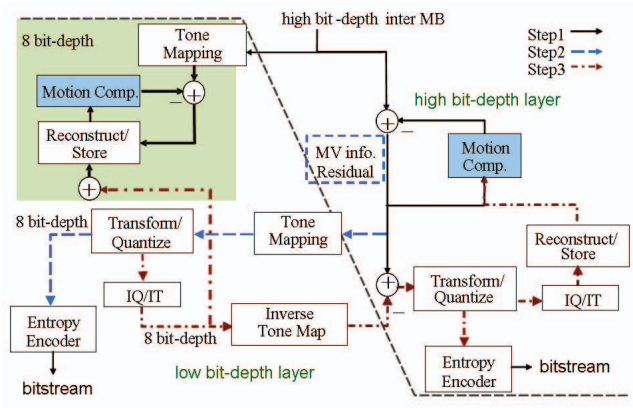

Fig.3 Coding scheme for HL mode inter MB

3.1 Inter-layer prediction in HL mode

First, consider the coding procedure for I frame. The high bit-depth frame will be encoded by H.264 directly and there is no need to encode and transmit the low bit-depth layer, which could be created by tone mapping of the reconstructed high bit-depth I frame. Thus the bitstream contain no data used for reconstructing the low bit-depth I frame only.

For inter frame, every MB in the high bit-depth frame will be intra coded or inter coded depending on the rate-distortion cost. If the MB belongs to intra mode, the remaining coding procedure is exactly the same as H.264. The high bit-depth intra MB will be encoded and the associated low bit-depth MB can be obtained at the decoder side after tone mapping of the reconstructed high bit-depth MB. On the other hand, if the MB is decided as inter mode, the following coding procedure is different from H.264. It involves three steps to encode the inter MB in the HL mode. Fig.3 illustrates the coding architecture for the inter MB in HL mode. First, the motion information containing motion vector and MB mode of the high bit-depth MB will be utilized by the low bit-depth MB, accompanying with the associated residual. In the second step, the residual will be processed by tone mapping, followed by Discrete Cosine Transform (DCT), quantization and entropy coding. Then it becomes parts of the embedded bitstream of the low bit-depth MB. It means that the decoder can reconstruct the low bit-depth MB directly using the motion information and residual provided by the high bit-depth MB after suitable processing, as indicated by the green area in Fig. 3. In the final step, the residual coded in the low bit-depth bitstream will be converted back to high bit-depth domain by inverse tone mapping, and the high bit-depth MB may encode simply the difference between its original residual and predicted one come from the encoded residual of low bit-depth MB after inverse tone mapping.

3.2 Bitstream structure of HL mode

The bitstream of HL mode is different from LH mode, as shown in Fig. 4. It starts with the high bit-depth I frame information, represented as HDR_I. Besides, both bit-depth MBs will be reconstructed using the same MV, denoted as HDR Motion Info. The residual of the high bit-depth MB will be processed by tone mapping, DCT, quantization and entropy coding. The resulting data will be called LDR_P while HDR_P_Refine stands for the enhanced residual data for reconstructing the quality-refined high bit-depth MB. Obviously, the size of HL bit-stream is smaller than that of LH mode owing to the absence of the low bit-depth intra MBs and the shared MV of both bit-depth layers for inter MB. Although motion estimation is performed in the high bit-depth MBs, the bitstream of the HL mode is still embedded, which fits in with the essence of SVC.

Fig.4. Bitstream structure of HL mode

3.3 Combined LH-HL mode

As discussed in section 3.1 that the HL mode contains only the high bit-depth I frame data. It will result in bandwidth waste if the decoder side only offers a low bit-depth displayer, especially for small GOP size. Thus, we modify the HL mode to achieve a better performance for such a situation. It is called LH-HL mode where the intra MB will be encoded by LH mode while the inter MB by HL mode. The LH-HL mode targets at creating an embedded bitstream and keeping satisfactory performance at the same time.

4. EXPERIMENTAL RESULTS

We extend H.264 to complete the proposed bit-depth scalable video coding. To evaluate the performance of the proposed algorithms, two 12-bit high bit-depth test sequences [11] Sunrise (960×540) and Library (900×540) are used in this paper. The frame rate for both sequences is 30 Hz. The 8-bit representations are acquired by tone mapping of the original 12-bit sequences. The tone mapping method adopted here is described in [12] while the inverse tone mapping is simply by using table look-up. Both high and low bit-depth layers choose the same quantization parameter (QP) settings; it implies that no extra QP scaling is employed on high bit-depth input. Moreover, and the GOP sizes of 1 and 4 are used, to differentiate the coding efficiency of I frames and P frames in proposed coding schemes.

The rate-distortion performance of the proposed algorithm is shown in Figure 5-8 where a 12-bit single layer coding and the simulcast coding are compared. Fig. 5 and Fig. 6 presents the performance comparison for I frames coding. In this case, the HL mode is equivalent to the single layer coding and the LH-HL mode and the LH mode is the same as the proposed method in [8]. With increased bitrate, a better coding efficiency improvement over the simulcast is observed. Table 1 summarizes the percentage of MB in the inter-layer mode for the LH mode. Here, the inter-layer mode represents the employ of LDR prediction adopted by the high bit-depth layer. It reveals that high bit-depth intra MBs are likely to be predicted from its low bit-depth version once its corresponding low bit-depth MB is reconstructed more precisely and bitrate can be reduced consequently in this way.

Fig 7 and Fig.8 show the performance comparison for the GOP size 4. All the three proposed schemes outperform the simulcast. Moreover, it indicates that the HL mode outperforms the LH, LH-HL mode and the approach proposed in [8]. Table 2 and Table 3 detail the statistical distribution of the inter-layer mode chosen for the inter MB in LH mode and HL mode, respectively. Obviously, the adoption probability of residual prediction in the HL mode is higher than that in the LH mode. Two factors contribute the superior performance of the HL mode to the LH mode. First, the HL mode does not need to transmit the low bit-depth intra MB and only one motion information set is required for both layers. Second, the residual prediction from the high bit-depth layer to the low bit-depth layer is not only efficient but also reliable.

Next, the performance of the low bit-depth representations is assessed under the condition that the entire bitstream is perfectly received. Fig.9 illustrates the performances when the GOP size is 4. The HL mode appears to significantly outperforms LH mode by about 2.1*dB*~5.6*dB*. Thus, it is concluded that if the whole bitstream could be delivered successfully without any truncation and noise corruption, the HL mode can provide a better coding efficiency not only in high bit-depth images but also in low bit-depth ones.

Table 1 Percentage of inter-layer mode MB for I frame in LH mode

	QP 10	QP 15	QP 24	QP 32	QP 40
Sunrise	79.19	75.65	65.49	42.71	20.83
Library	73.41	67.63	53.49	36.37	18.54

Table2. Percentage of inter-layer mode MB for P frames in HL mode

	QP 10	QP 15	QP 24	QP 32	QP 40
LDR pred.	26.78	23.27	20.93	13.41	4.14
Res. pred.	66.21	67.87	60.89	56.67	50.78

Table 3 Percentage of inter-layer mode MB for P frames in HL mode

	QP 10	QP 15	QP 24	QP 32	QP 40
Res.pred.	91.31	91.07	81.92	67.02	50.12

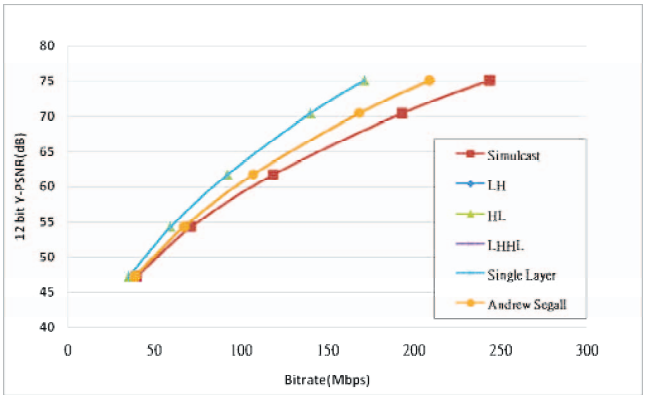

Fig.5 Performance comparison for "Sunrise" (GOP=1)

5. CONCLUSION

In this paper, we proposed three bit-depth scalable video coding architectures suitable for different applications. LH mode provides

an embedded bitstream which is backward compatible to H.264, while HL mode offers better coding efficiency if bandwidth limitation is not an issue. These two modes differ in the direction of the inter-layer prediction. To resolve the limited applicability of the HL mode, a LH-HL mode is proposed to compromise the HL mode with the LH mode where an embedded bistream and an improvement over the LH mode are verified. Experimental results demonstrate the effectiveness of the proposed algorithms. In particular, the major benefit of the proposed algorithms is the ability to adjust and choose a suitable coding approach for specified applications.

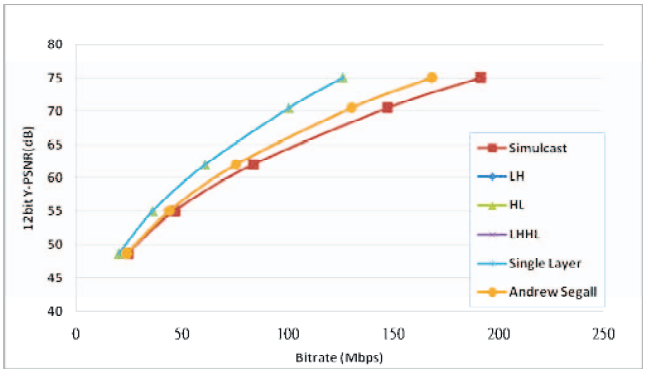

Fig.6 Performance comparison for "Library" (GOP=1)

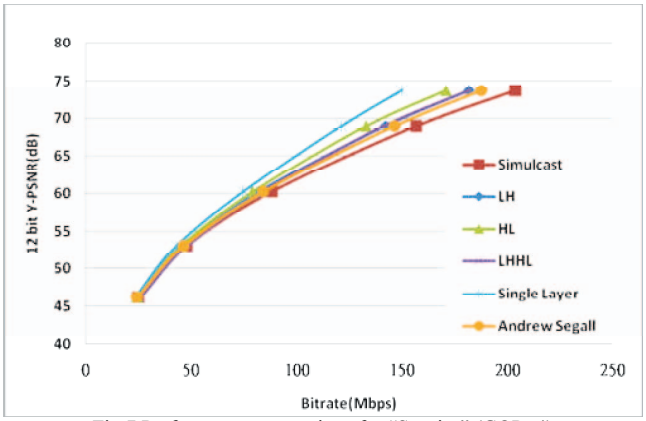

Fig.7 Performance comparison for "Sunrise" (GOP=4)

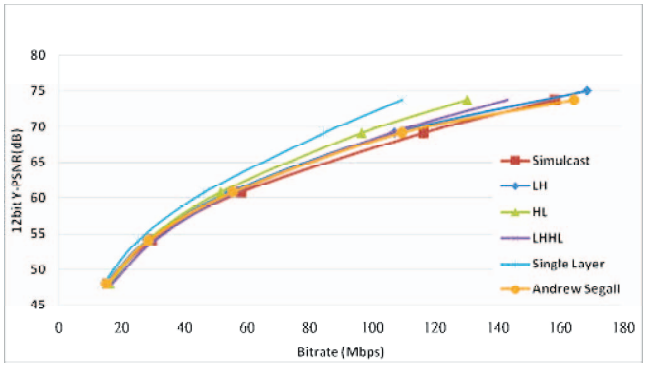

Fig.8 Performance comparison for "Library" (GOP=4)

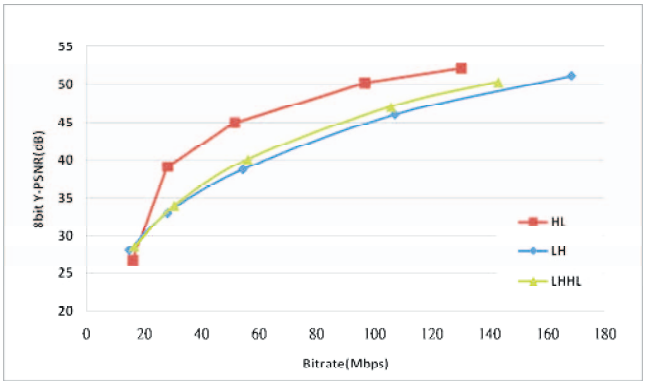

Fig.9 Performance for low bit-depth sequence "library" (GOP=4)

6. REFERENCES

[1] J.Reichel, H.Schwarz, and M. Wien (eds.), "Scalable Video Coding – Joint Draft 9," *Joint Video Team*, Doc. JVT-V201, Marrakech, Jan. 2007.

[2] H.Schwarz, D. Marpe and T. Wiegand, "Overview of the Scalable Video Coding Standard," *IEEE Trans. on Circuits and Systems for Video Technology,*" vol.17, no.9, pp.1103-1120, Sep. 2007.

[3] Y. Gao and Y. Wu, "Applications and Requirement for Color Bit Depth Scalability," *Joint Video Team*, Doc. JVT-U049, Hangzhou, China, Oct. 2006.

[4] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Trans. on Circuits and Systems for Video Technology* vol.13, no.7, pp.560-576, July 2003.

[5] Y. Yu, S. Gordon, and M. Yang, "Improving Compression Performance in Bit Depth SVC with a Prediction Filter," *Joint Video Team*, Doc. JVT-Z049, Antalya, Jan. 2008.

[6] S. Liu, A.Vetro and W. S. Kim, "Inter-layer prediction for SVC Bit-depth Scalability," *Joint Video Team*, Doc. JVT-X075, Geneva, June 2007.

[7] Martin Winken, Detlev Marpe, Heiko Schwarz, and Thomas Wiegand, "Bit-depth Scalable Video Coding," *Proc. IEEE International Conference on Image Processing,* pp.5-8, 2007.

[8] A. Segall, "Scalable Coding of High Dynamic Range Video," *Proc. IEEE International Conference on Image Processing,* pp.1-4, 2007.

[9] Y. Wu, Y. Gao, and Y. Chen, "Bit Depth Scalable Coding," *Proc. IEEE International Conference on Multimedia & Expo*, pp.1139-1142 July 2007.

[10] Y. Wu, Y. Gao, and Y. Chen, "Bit-depth Scalable Coding Based on Macroblock Level Inter-layer Prediction," *Proc. IEEE Symposium Conference on Circuits and Systems*, pp.3442-3445, May 2008.

[11] A. Segall "Donation of Tone Mapped Image Sequences," *Joint Video Team*, Doc. JVT-Y072, Shenzhen, China, October, 2007.

[12] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic Tone Reproduction for Digital Images," *ACM Transactions on Graphics,* vol.23, no.3, pp.267-276, July 2002.