

INTER-LAYER BIT ALLOCATION FOR SCALABLE HIGH-EFFICIENCY VIDEO CODING

Vo Phuong Binh^{a*}

^aThe Faculty of Information Technology, Dalat University, Lamdong, Vietnam

Article history

Received: January 04th, 2016

Received in revised form: March 07th, 2016

Accepted: March 16th, 2016

Abstract

Bit allocation is essential for a video encoder to accurately control the generated bits, and thus greatly influences the visual quality. In this paper, an improved bit allocation algorithm is proposed at the frame level for the emerging Scalable High-efficiency Video Coding (SHVC) standard. At the spatial base and enhancement layers, the bit budget is derived jointly from the hierarchical level and the visual complexity of the current frame, where the latter is measured by the inter-layer predicted MAD (Mean Absolute Difference) to allocate the bit budget of each frame. Experimental results show that the proposed method achieves more accurate bitrates with higher visual quality in the average PSNR up to 1.40dB, and controls buffer occupancy more satisfactorily, as compared with the-state-of-the-art approaches in the literature.

Keywords: Bit Allocation; Mean Absolute Difference (MAD); Rate Control; Scalable High-efficiency Video Coding (SHVC); Scalable Video Coding (SVC).

1. INTRODUCTION

Videos find wide applications. With a variety of end devices and network environments, a single-layer coded video content will not adapt all its needs to various constraints, such as display resolution, network bandwidth, and computational capability. Scalable Video Coding (SVC), also termed layered coding technically, has been proposed as an efficient solution to address this issue. Each SVC layer includes a video bit-stream corresponding to a specified frame rate, resolution, or fidelity. The basic High Efficiency Video Coding (HEVC) or H.265 [1] specifies a single-layer video

* Corresponding author: binhvp@dlu.edu.vn

coding structure while it also supports a temporal multi-layer video coding by using the hierarchical B-picture structure, which was adopted in H.264/SVC [2]. Spatial and quality (SNR) scalability is developed in HEVC as an important extension [3], commonly known as Scalable High Efficiency Video Coding (SHVC). Consequently, SHVC provides fully scalabilities in the temporal (frame rate), spatial (resolution), and SNR (fidelity) domains.

Rate control (RC) for a video encoder is a mechanism that modifies the encoding parameters to maintain a target bit rate. A good RC algorithm also attempts to optimize the video quality, minimize the fluctuation of PSNR in the coded sequence, and prevent the buffer overflow and underflow for a hypothetical reference decoder (HRD). RC is generally fulfilled by adjusting the quantization parameter (QP) to regulate the bit rate [4]. A larger QP that corresponds to a larger quantization step size reduces the number of generated bits, while the reconstructed image block will have a larger distortion.

Two main steps are involved in an RC algorithm to determine QP, namely *bit allocation* and *QP estimation*. The bit allocation step aims to assign a bit budget for each of the coding segments, such as a group of picture (GOP), a picture (frame), or a coding unit (CU). Then, the QP estimation step manages to compute a QP value based on the allocated bit budget for each coding segment. Therefore, bit allocation is a very important part of an RC algorithm to achieve a proper QP.

Some bit allocation methods have been proposed for the RC algorithm of HEVC. The pixel-wise (PW) based on bit allocation algorithm in [5] considered the buffer occupancy to prevent the buffer overflow or underflow. Lee et al. [6] presented a frame-level bit allocation algorithm for HEVC that utilized the average remaining bits in the GOP, additional to the buffer-occupancy constraint. In [7], a proposed bit allocation algorithm utilized the hierarchical structure and the relationship between a coding frame and its reference frame. Note that these algorithms are not applied to the SHVC.

The RC algorithm of the SHVC reference software (SHM), SHM9.0 [8], was mainly based on the two RC algorithms of HEVC for spatial layers [9, 10]. The hierarchical bit allocation (HBA) algorithm in [9] considered the hierarchical level and buffer occupancy of the current GOP. The adaptive bit allocation (ABA) algorithm in [10] further improved the algorithm in [9] by incorporating a $R-\lambda$ model estimated from the video content of the previous GOP. However, both of [9, 10] do not consider the visual content of the current frame, which is important for allocating a proper bit budget to the current frame.

In this paper, we propose a bit allocation algorithm to calculate the bit budget of each frame for each of the SHVC spatial layers. The bit budget is allocated based on both the hierarchical level and the visual complexity of the current frame. The visual complexity is estimated by the inter-layer MAD prediction. The bit allocation algorithm extends our previous work for H.264/SVC [11] that incorporates the visual complexity and the corresponding temporal frame level. Experimental results substantiate the superiority of the proposed method.

The rest of this paper is organized as follows. Section 2 provides a brief description of the bit allocation methods in SHM9.0. The proposed bit allocation algorithm for SHVC is presented in Section 3. Section 4 shows the experimental results to demonstrate the efficiency of the proposed algorithm as compared with the-state-of-the-art approaches in the literature. Finally, conclusions are presented in Section 5.

2. BIT ALLOCATION METHODS FOR SHVC IN SHM9.0

Bit allocation is implemented at the first step of each two-step RC algorithm of spatial layers in the SHM. In SHM9.0 [8], the target bits for the current frame T_{CurrPic} in a GOP (Group of Pictures) is determined as follows:

$$T_{\text{CurrPic}} = \frac{T_{\text{GOP}} - \text{Coded}_{\text{GOP}}}{\sum_{\text{NotCoded}} \omega_i} \times \omega_{\text{CurrPic}} \quad (1)$$

$$T_{\text{GOP}} = \left[R_{\text{PicAvg}} + \frac{R_{\text{PicAvg}} \times N_{\text{coded}} - R_{\text{coded}}}{SW} \right] \times N_{\text{GOP}} \quad (2)$$

where T_{GOP} is the bit budget of the current GOP; R_{PicAvg} is the average target bits per picture determined by the target bit rate R and frame rate f : $R_{\text{PicAvg}} = R / f$; N_{coded} is the number of coded frames; R_{coded} is the generated bits of coded frames; SW is the size of the smooth window set to 40 in SHM9.0; N_{GOP} is the number of frames in each GOP; $Coded_{\text{GOP}}$ is the coded bits of the current GOP before encoding the current frame; ω_{CurrPic} and ω_i are the weight of the current frame and i th frame in the current GOP, respectively.

In SHM9.0, there are two methods to determine the weight ω_i of the i th frame. The HBA method [9] determines ω_i based on the hierarchical level and bpp (bits per pixel), where the larger the hierarchical level is, the smaller the weight value is assigned. SHM9.0 also supports the ABA method [10] based on the following R- λ model [9]:

$$\lambda = \alpha \times bpp^\beta \quad (3)$$

$$bpp = \frac{T}{w \times h} \quad (4)$$

where λ is the slope of rate-distortion (R-D) curve; α and β are parameters of the R- λ model updated after encoding each frame; bpp is the number of bits per pixel; T is the target bits of the current frame; w and h are the width and height of the frame respectively. Then, the weight ω_i is determined by utilizing indirectly the video content of the previous GOP based on the parameters of the R- λ model.

3. PROPOSED METHOD

The visual complexity of a frame is one of the most important characteristics for allocating a proper bit budget to achieve good R-D performance. As presented in Section 2, the bit allocation methods at the frame level in SHM9.0 do not utilize the complexity of the current frame and the visual quality may thus be unsatisfactory due to inadequate bit allocation. In this section, the bit allocation algorithm is proposed based on both the hierarchical level and the visual complexity measured by MAD, as will be explained in the following subsections.

3.1. Relationship between the number of output bits and MAD

The QP corresponds to the quantization level for residual transform coefficients after inter/intra-predictions. Therefore, encoding with a fixed QP produces coded video sequences of relatively stable quality in terms of PSNR. However, encoding with a fixed QP does not ensure a constant bitrate. In addition to the QP, the generated bitrate is closely associated with visual complexity. The MAD of a frame of height H and width W is defined as follows:

$$\text{MAD} = \frac{1}{H \times W} \times \sum_{x=1}^H \sum_{y=1}^W |\text{Pic}_{\text{Org}}(x, y) - \text{Pic}_{\text{Pred}}(x, y)| \quad (5)$$

where $\text{Pic}_{\text{Org}}(x, y)$ and $\text{Pic}_{\text{Pred}}(x, y)$ are the pixel values at position (x, y) of the original and predicted frames, respectively. $\text{Pic}_{\text{Pred}}(x, y)$ is obtained using motion estimation and motion compensation, usually performed in blocks, such as the prediction units (PUs) in HEVC. The relationship between the number of output bits and MAD for encoding test sequences using HEVC with a fixed QP, plotted in Figure 1, exhibits a near-linear relationship. This relationship is considered in designing the proposed bit allocation algorithm to minimize the PSNR fluctuation with the bitrate and buffer constraint.

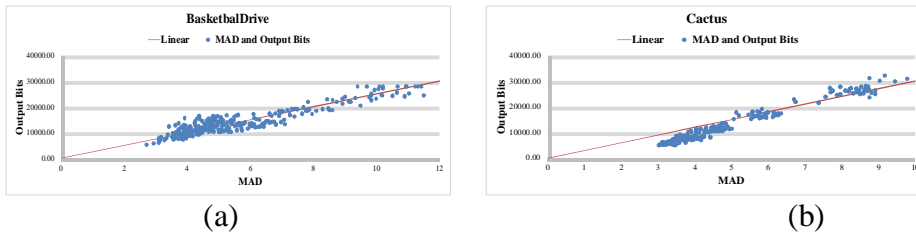


Figure 1. Relationship between number of output bits and MAD with fixed QP encoding for (a) BasketballDrive and (b) Cactus sequences.

3.2. Estimating the visual complexity at the base layer

The major challenge in using MAD in bit allocation is that the actual MAD of the current frame is available *after* motion compensation and is thus unavailable during bit allocation. Although pre-encoding the current frame with a specific QP can produce an accurately estimated MAD, this approach involves large computation and is impractical. Instead, the MAD of the current frame is typically predicted from the actual

MAD of the previously coded frame, which is available during encoding. At the base layer, the conventional linear MAD prediction is utilized according to the autoregressive model described in [12]:

$$MAD(i) = a \times MAD_{\text{actual}}(i-1) + b \quad (6)$$

where $MAD(i)$ is the predicted MAD of the current frame, and $MAD_{\text{actual}}(i-1)$ is the actual MAD of the previously coded frame. In (6), the parameters a and b are initially set as 1 and 0, respectively, and updated after each frame is encoded through linear regression and by using the outlier removal strategy described in [13].

3.3. Estimating the visual complexity at the enhancement layer

Experimental results for the relationship between MADs of the base layer (layer 0) and enhancement layer (layer 1) are illustrated in Figure 2. These results reveal that the MAD values of the enhancement and base layers product a near-directly proportional relationship.

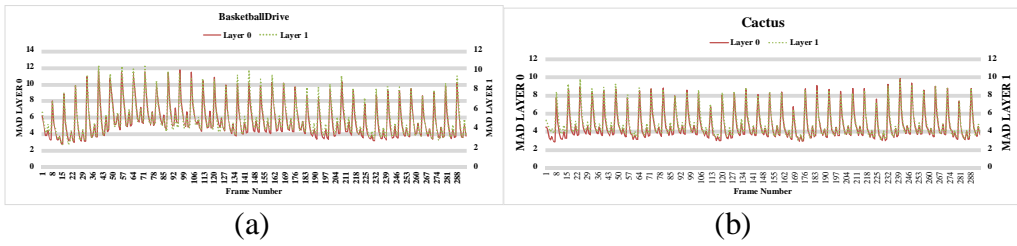


Figure 2. Relationship between MADs of the base and enhancement layers for (a) BasketballDrive and (b) Cactus

According to the above experimental results, a new MAD prediction model for the enhancement layer using the encoding results from both the base layer and previous temporal frames is proposed. The new prediction model is defined as:

$$MAD(i) = \omega \times MAD_{\text{el,inter}}(i) + (1 - \omega) \times MAD_{\text{el,temp}}(i) \quad (7)$$

Where ω is a weighting factor, calculated as

$$\omega = \text{Min} \left(\frac{|MAD_{\text{bl,pred}}(i) - MAD_{\text{bl,act}}(i)|}{MAD_{\text{bl,act}}(i)}, 1 \right) \quad (8)$$

and subscripts ‘el’ and ‘bl’ indicate the enhancement layer and the base layer; $MAD_{bl,act}(i)$ and $MAD_{bl,pred}(i)$ refer to the actual and predicted MAD of the co-located frame of the i th frame in the enhancement layer; the $\text{Min}(x, y)$ function returns the smallest value between x and y ; $MAD_{el,temp}(i)$ and $MAD_{el,inter}(i)$ indicate the temporally predicted MAD and the inter-layer predicted MAD of the i th frame in the enhancement layer.

The temporally predicted MAD is obtained through the linear prediction model defined in equation (6). In a similar way to equation (6), a linear prediction model for the prediction of the MAD of a frame in the enhancement layer, using the actual MAD value of its co-located frame in the base layer is proposed

$$MAD_{el,inter}(i) = t_1 \times MAD_{bl}(i) + t_2 \quad (9)$$

Where $MAD_{bl}(i)$ denotes the actual MAD of the frame in the co-located position in the base layer; t_1 and t_2 are model coefficients updated using a linear regression method after the coding of each frame [13]. It can be seen that the proposed MAD prediction model is completely adaptive, as the weight of the temporal MAD prediction and that of the inter-layer MAD prediction can be adjusted instantly according to the error rate of the linear MAD prediction in the base layer.

3.4. Proposed bit allocation algorithm

For bit allocation at the GOP level and the CU level, we adopt the same methods implemented in SHM9.0. The bit budget for the i th frame at hierarchical level k , denoted by $T(i, k)$, is computed as follows:

$$T(i, k) = (1 - \tau) \times T_1(i, k) + \tau \times T_2(i, k) \quad (10)$$

Where τ is the constant set to 0.1 as in SHM9.0. The first rate term T_1 accounts for the influence of GOP target bit rate to control the buffer occupancy:

$$T_1(i, k) = \frac{T_{\text{GOP}} \times (L - k + 1)}{\sum_{l=1}^L (L - L_l + 1) \times N^l} + (B_r - B(i)) \quad (11)$$

Where T_{GOP} is the allocated bits of the current GOP determined by (2); N^l is the number of frames at the l th hierarchical level in the current GOP; L is the largest hierarchical level, and L_l is the hierarchical level of the l th frame. B_t is the target buffer occupancy, which is set as 40% of the total buffer size in this study, and $B(i)$ is the buffer occupancy before the i^{th} frame is encoded. The second rate term T_2 is calculated based on the visual complexity to achieve better visual quality as follows:

$$T_2(i, k) = \frac{T_r \times (L - k + 1) \times \text{MAD}(i)}{\sum_{l=1}^L (L - L_l + 1) \times N_r^l \times \text{MAD}^l} \quad (12)$$

Where T_r is the remaining bits of the current GOP before encoding the current frame; N_r^l is the number of remaining frames at the l th hierarchical level in the current GOP; $\text{MAD}(i)$ is the visual complexity of the i th current frame determined by (6) and (7) of the base and enhancement layers, respectively; MAD^l is the moving average visual complexity of the l th hierarchical level. Note that MAD^l is updated after encoding the i th frame at the same hierarchical level l as follows:

$$\text{MAD}_{\text{new}}^l = \frac{\text{MAD}(i) + (N_k - 1) \times \text{MAD}_{\text{old}}^l}{N_k} \quad (13)$$

Where N_k is the number of coded frames at the l th hierarchical level.

3.5. Rate control algorithm for SHVC

There are two main steps in the proposed RC algorithm at the frame level for each spatial layer of SHVC multi-layer encoder, including bit allocation and QP estimation as illustrated in Figure 3.

Step 1: Bit allocation is to generate the bit budget of the current frame in the current GOP by (10).

Step 2: QP estimation is to compute the QP value for the current frame of the current GOP based on the R- λ model as in [9]:

$$QP = 4.2005 \times \ln \lambda + 13.7122 \quad (14)$$

Where λ is the slope of R–D curve given in (3). The number of bits per pixel bpp in (3) is determined by (4) based on the bit budget of the current frame in Step 1.

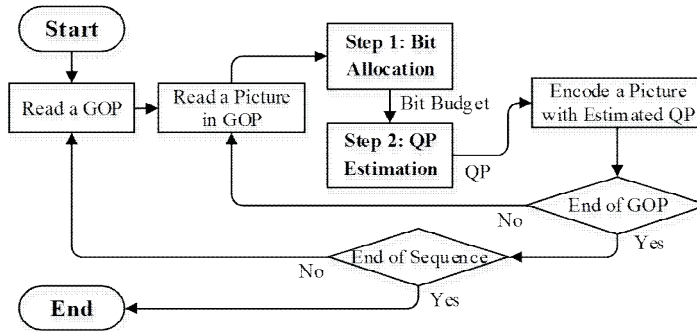


Figure 3. Flow chart of the proposed rate control for each SHVC spatial layer

4. EXPERIMENTAL RESULTS

The proposed method is compared with the bit allocation methods in SHM9.0 [8] including the HBA [14] and ABA [10] algorithms. In addition, the PW method [5], implemented in a few versions before SHM4.0, is used for comparison. The GOP size, which is the length between two consecutive P frames, is set to 8 with the random access main (RA-Main) structure and only the first frame is intra-coded, as the parameter settings of [5, 8] for fair comparisons. The buffer size (in bits) in our experiments is set to 0.25 (in second) multiplied by the target bitrate (in bits/sec). In other words, the decoding delay is limited to 250 ms, which is suitable for low-delay video applications. The buffer fullness is defined as a percentage of the total buffer size and must be between 0% and 100% to prevent buffer underflow and overflow. Four benchmark video sequences, “BasketballDrive” (50Hz), “BQTerrace” (60Hz), “Cactus” (60Hz), and “Vidyo3” (60Hz), all with 300 frames, are tested. Each test sequence was encoded once at the highest bitrate (4096 kbps) at the four target bitrates of the spatial/quality layer listed in Table 1, where a bit-rate referred to a target accumulated bit-rate of a spatial/quality layer. Layer 0 is the base layer with a resolution of 240p (416 × 240 pixels/frame). Layers 1 and 2 are spatial enhancement layers with a resolution of 480p (832 × 480 pixels/frame) and HD (1280 × 720 pixels/frame), respectively. Layer 3 is a CGS quality layer with the same resolution as that of layer 2.

Table 1. Layer settings for the combined scalability experiment

Layer	Resolution (width x height)	Target bitrate (kbps)
0	240p (416 x 240)	512
1	480p (832 x 480)	1024
2	HD (1280 x 720)	2048
3	HD (1280 x 720)	4096

All spatial/quality layers were encoded with a GOP size of 8, and four temporal layers were achieved with temporal sub-streams. All spatial/CGS quality enhancement layers (layers 1, 2, and 3) were predictively encoded with inter-layer and intra-layer predictions. We employ DBR, the differential bit rate, to evaluate the accuracy of the output bit rate R_0 with respect to the desired target bit rate R_t :

$$\text{DBR} = \frac{|R_0 - R_t|}{R_t} \times 100\% \quad (15)$$

The experimental results presented in Table 2 show that the proposed algorithm achieves accurate target bit rates (with average DBR = 0.07%), as compared with the HBA algorithm (with average DBR = 0.11%) and the ABA algorithm (with average DBR = 0.15%). Although the PW method obtains the most accurate target bitrate (with average DBR = 0.02%), its R–D performance is notably the worst (average PSNR = 38.84dB).

The R–D performance of the proposed algorithm (average PSNR = 40.24dB) is superior to those of the ABA algorithm (average PSNR = 39.97dB) and the HBA algorithm (average PSNR = 39.88dB). Recall that the PW and HBA algorithms do not consider the video content.

Table 2. Performance and standard deviation (SD) of PSNR for combined scalability

Sequence	Layer	SHM9.0 - HBA			SHM9.0 - ABA			PW [5]			Proposed		
		DBR (%)	PSNR (dB)	SD	DBR (%)	PSNR (dB)	SD	DBR (%)	PSNR (dB)	SD	DBR (%)	PSNR (dB)	SD
BasketballDrive	0	0.00	36.03	1.35	0.00	36.03	1.14	0.04	35.10	0.94	0.02	36.47	0.47
	1	0.00	36.94	1.63	0.00	36.99	1.25	0.03	35.97	1.01	0.02	37.35	0.81
	2	0.00	38.22	1.94	0.00	38.34	1.32	0.02	37.59	1.38	0.04	38.67	1.04
	3	0.00	40.88	2.24	0.00	41.02	1.37	0.02	40.55	1.49	0.03	41.32	1.28

Table 2. Performance and standard deviation (SD) of PSNR for combined scalability (cont)

BQTerrace	0	0.00	39.20	1.30	0.00	39.43	1.72	0.00	88.19	0.65	0.02	39.50	0.79
	1	0.00	38.91	0.44	0.00	39.21	0.53	0.02	37.40	0.74	0.02	39.26	0.46
	2	0.01	38.84	0.44	0.00	39.01	0.61	0.03	37.70	0.83	0.04	39.31	0.40
	3	0.00	40.73	0.53	0.00	40.76	0.78	0.00	40.01	0.77	0.05	41.36	0.41
Cactus	0	0.05	36.75	0.37	0.01	36.79	0.71	0.04	35.52	0.69	0.09	37.12	0.19
	1	0.01	36.48	0.49	0.00	36.52	0.60	0.01	34.58	0.40	0.06	36.59	0.20
	2	0.01	37.73	0.68	0.01	37.73	0.59	0.01	35.69	0.41	0.10	37.86	0.28
	3	0.00	40.47	0.80	0.00	40.47	0.60	0.00	38.72	0.60	0.09	40.86	0.39
Vidy3	0	1.67	45.85	0.32	2.41	46.02	0.45	0.03	44.32	0.8	0.18	46.02	0.37
	1	0.01	43.95	0.25	0.02	44.09	0.35	0.00	42.46	0.49	0.14	44.17	0.15
	2	0.00	42.95	0.44	0.01	43.03	0.48	0.00	42.90	0.17	0.07	43.4	0.19
	3	0.00	44.09	0.73	0.00	44.12	0.67	0.01	44.71	0.13	0.07	44.56	0.30
Average		0.11	39.88	0.87	0.15	39.97	0.82	0.02	38.84	0.72	0.07	40.24	0.48

The ABA algorithm infers the complexity of the current frame from the video content of the previous GOP. Consequently, its R–D performance is inferior to the proposed algorithm, especially, for video sequences with non-stationary visual complexity. The proposed method (with average SD = 0.48) generates satisfactorily low PSNR fluctuations in the enhancement layers by more accurately capturing inter-layer correlations. For buffer occupancy comparisons as illustrated in Figure 4 and Figure 5, all algorithms prevent buffer overflow but only the proposed algorithm adequately manages buffer occupancy in all the scalable layers.

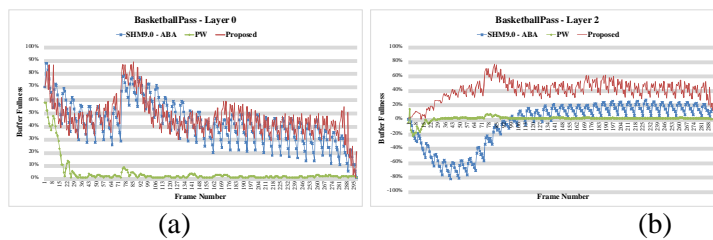


Figure 4. BasketballPass sequence, buffer status in (a) layer 0 and (b) layer 2

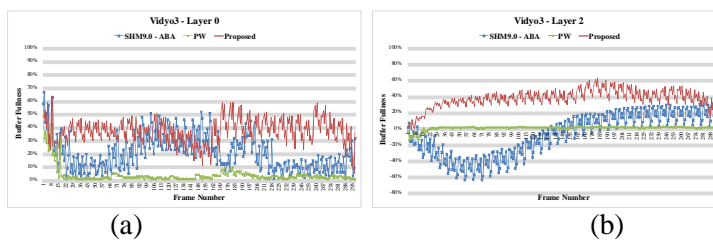


Figure 5. Vidy3 sequence. Buffer status in (a) layer 0 and (b) layer 2

This is because the proposed method yields stable buffer occupancy and allocates bits more adequately compared with the other methods.

The PW method typically incurs buffer underflow in early pictures in enhancement layers. The ABA may incur buffer underflow for video sequences with high target bitrates because only the GOP level and not the frame level buffer occupancy is accounted for in the ABA. For the computational complexity, the average overall encoding time of all the evaluated algorithms are nearly the same, as presented in Table 3, where the HBA method is used as the basis of reference. As described in Section III, the proposed method considered the GOP size and buffer size for allocating the bit budget for each frame.

Table 3. Encoding time comparisons for 4 layers of combined scalability

Sequences	SHM9.0 - ABA	PW [5]	Proposed
	Encoding Time (%)	Encoding Time (%)	Encoding Time (%)
BasketballDrive	101.15%	100.19%	100.09%
BQTerrace	97.80%	100.72%	100.63%
Cactus	100.07%	100.25%	100.15%
Vidyo3	99.07%	100.37%	100.04%
Average	99.52%	100.38%	100.23%

The additional experimental results with GOP size equaling 16 and buffer size set to 0.5 (in second) multiplied by the target bitrate (in bits/sec) presented in Table 4 show that the proposed algorithm also achieve accurate target bit rates (average DBR = 0.6%) with the highest quality and the lowest PSNR fluctuation, as compared with all the remaining algorithms.

Table 4. Additional performance for combined scalability

Sequence	Layer	SHM9.0 - HBA			SHM9.0 - ABA			PW [5]			Proposed		
		DBR (%)	PSNR (dB)	SD	DBR (%)	PSNR (dB)	SD	DBR (%)	PSNR (dB)	SD	DBR (%)	PSNR (dB)	SD
BasketballDrive	0	0.00	36.02	1.32	0.00	36.03	1.15	0.04	35.11	0.95	0.02	36.48	0.46
	1	0.00	36.94	1.61	0.00	36.99	1.27	0.03	35.97	1.03	0.02	37.35	0.79
	2	0.00	38.22	1.95	0.00	38.34	1.31	0.02	37.59	1.37	0.04	38.67	1.05
	3	0.00	40.89	2.21	0.00	41.02	1.38	0.02	40.55	1.48	0.03	41.31	1.26

Table 4. Additional performance for combined scalability (cont)

BQTerrace	0	0.00	39.19	1.31	0.00	39.43	1.71	0.00	38.18	0.64	0.02	39.51	0.79
	1	0.00	38.91	0.44	0.00	39.21	0.53	0.02	37.40	0.75	0.02	39.26	0.46
	2	0.01	38.85	0.45	0.00	39.01	0.61	0.03	37.70	0.81	0.04	39.31	0.42
	3	0.00	40.73	0.53	0.00	40.76	0.79	0.00	40.01	0.79	0.05	41.35	0.41
Cactus	0	0.05	36.76	0.37	0.01	36.81	0.71	0.04	35.52	0.69	0.09	37.13	0.21
	1	0.01	36.48	0.49	0.00	36.52	0.62	0.01	34.58	0.41	0.06	36.59	0.24
	2	0.01	37.73	0.68	0.01	37.73	0.59	0.01	35.69	0.43	0.10	37.86	0.27
	3	0.00	40.47	0.83	0.00	40.47	0.61	0.00	38.72	0.61	0.09	40.86	0.38
Vidyo3	0	1.61	45.84	0.32	2.42	46.02	0.46	0.03	44.31	0.81	0.17	46.01	0.38
	1	0.01	43.95	0.25	0.02	44.09	0.35	0.00	42.46	0.49	0.11	44.16	0.17
	2	0.00	42.95	0.44	0.01	43.03	0.48	0.00	42.90	0.16	0.07	43.40	0.18
	3	0.00	44.09	0.73	0.00	44.12	0.67	0.01	44.71	0.21	0.07	44.56	0.31
Average		0.11	39.88	0.87	0.15	39.97	0.83	0.02	38.84	0.73	0.06	40.24	0.49

5. CONCLUSION

In this paper, an inter-layer bit allocation algorithm for SHVC is proposed. The proposed algorithm determines the bit budget based on both the hierarchical level and the visual complexity of the current frame, where the latter is estimated by the inter-layer predicted MAD. Experimental results show that the proposed method provides accurate bitrates (with average DBR = 0.07%) and more stable visual quality, as compared with the algorithms implemented in SHM9.0. For R–D performance, the proposed method gains 1.40dB, 0.36dB and 0.27dB (average PSNR), as compared with the PW, HBA and ABA methods, respectively. Furthermore, the proposed method achieves enhanced buffer control for all scalable layers, as compared with the-state-of-the-art approaches in the literature.

REFERENCES

- [1] G. J. Sullivan, J. Ohm, H. Woo-Jin, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. on Circuits Syst. Video Technol.*, vol. 22, pp. 1649-1668 (2012).
- [2] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. on Circuits Syst. Video Technol.*, vol. 17, pp. 1103-1120 (2007).
- [3] G. J. Sullivan, J. M. Boyce, C. Ying, J. R. Ohm, C. A. Segall, and A. Vetro, "Standardized extensions of high efficiency video coding (HEVC)," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, pp. 1001-1016 (2013).

- [4] I. E. G. Richardson, "H.264 and MPEG-4 video compression: video coding for next-generation multimedia," 1st edn. ed: NewYork:Wiley, pp. 256 – 265, (2003).
- [5] H. Choi, J. Yoo, J. Nam, D. Sim, and I. V. Bajic, "Pixel-wise unified rate-quantization model for multi-level rate control," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, pp. 1112-1123 (2013).
- [6] B. Lee, M. Kim, and T. Q. Nguyen, "A frame-level rate control scheme based on texture and nontexture rate models for high efficiency video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, pp. 465-479 (2014).
- [7] S. Wang, S. Ma, S. Wang, D. Zhao, and W. Gao, "Rate-GOP based rate control for high efficiency video coding," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, pp. 1101-1111 (2013).
- [8] SHM9.0 software package [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_SHVCSsoftware/tags/SHM-9.0/.(Sep. 2015)
- [9] B. Li, H. Li, L. Li, and J. Zhang, "Rate control by R-lambda model for HEVC," document JCTVC-K0103, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC, 11th Meeting: Shanghai, China, 10-19 Oct. (2012).
- [10] B. Li, H. Li, and L. Li, "Adaptive bit allocation for R-lambda model rate control in HM," document JCTVC-M0036, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC, 13th Meeting: Incheon, Korea, 18-26 Apr. (2013).
- [11] V. P. Binh and S. H. Yang, "A better bit-allocation algorithm for H.264/SVC," *The Fourth International Symposium on Information and Communication Technology*, pp. 18-26, Dec. (2013).
- [12] Z.-G. Li, F. Pan, K.-P. Lim, G. Feng, X. Lin, and S. Rahardja, "Adaptive basic unit layer rate control for JVT," *document JVT-G012, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 7th Meeting: Pattaya II, Thailand, 7-14 March*, (2003).
- [13] H.-J. Lee, T.-H. Chiang, and Y.-Q. Zhang, "Scalable rate control for MPEG-4 video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp. 878-894 (2000).
- [14] B. Li, H.-Q. Li, L. Li, and J.-L. Zhang, "(lambda) Domain rate control algorithm for high efficiency video coding," *IEEE Trans. on Image Process.*, vol. 23, pp. 3841-3854 (2014).

CẤP PHÁT BÍT SỬ DỤNG THÔNG TIN ĐA LỚP CHO CHUẨN NÉN VIDEO HIỆU QUẢ CAO NHIỀU LỚP SHVC

Võ Phương Bình^{a*}

^aKhoa Công nghệ Thông tin, Trường Đại học Đà Lạt, Lâm Đồng, Việt Nam

*Tác giả liên hệ: Email: binhvp@dlu.edu.vn

Nhận ngày 04 tháng 01 năm 2016

Chỉnh sửa ngày 07 tháng 03 năm 2016 | Chấp nhận đăng ngày 16 tháng 03 năm 2016

Tóm tắt

Cấp phát bit rất cần thiết cho một chuẩn nén video để kiểm soát các bit được tạo ra một cách chính xác, và do đó ảnh hưởng rất lớn đến chất lượng video. Trong bài báo này, thuật toán cấp phát bit được đề xuất ở cấp độ khung ảnh (frame) cho chuẩn nén video hiệu quả cao nhiều lớp SHVC (Scalable High-efficiency Video Coding). Lượng bit được cấp phát dựa trên cấp độ khung ảnh và độ phức tạp của khung ảnh hiện tại, trong đó độ phức tạp khung ảnh được đo bằng MAD (Mean Absolute Difference). MAD của các lớp nâng cao được xác định dựa trên thông tin đa lớp giữa lớp nâng cao và cơ sở. Kết quả thực nghiệm cho thấy rằng phương pháp đề xuất đạt được các tỉ lệ bit (bit-rate) chính xác hơn, chất lượng video tốt hơn với PSNR trung bình cao hơn 1.40dB, và kiểm soát vùng đệm hiệu quả hơn trong việc phòng tránh hiện tượng tràn và lãng phí vùng đệm, so với các phương pháp tiếp cận khác hiện nay cho chuẩn nén video hiệu quả cao nhiều lớp SHVC.

Từ khoá: Cấp phát bit; Hiệu tuyệt đối trung bình (MAD); Kiểm soát tỉ lệ bit; Nén video hiệu quả cao nhiều lớp (SHVC); Nén video nhiều lớp (SVC).
