SINGLE PASS DEPENDENT BIT ALLOCATION FOR SPATIAL SCALABILITY CODING OF H.264/SVC

Randa Atta, Rehab F. Abdel-Kader, and Amera Abd-AlRahem

Electrical Engineering Department, Faculty of Engineering, Port Said University,
Port Fouad 42523, Port Said, Egypt
E-mails: {r.atta, rehabfarouk, amera.gamea}@eng.psu.edu.eg

ABSTRACT

This paper investigates the problem of bit allocation for spatial scalability coding of H.264/SVC. Little prior work deal with the H.264/SVC bit allocation problem considering the correlation between the enhancement and base layers. Nevertheless, most of the bit allocation algorithms suffer from high computational complexity which grows significantly with the number of layers. In this paper, a single-pass spatial layer bit allocation algorithm, based on dependent Rate-Distortion modeling is proposed. In this algorithm, the R-D model parameters are adaptively updated during the coding process. Experimental results demonstrate that the proposed algorithm achieves a significant improvement in the coding gain as compared to the multi-pass model-based algorithm and the Joint Scalable Video Model reference software algorithm.

Index Terms— Dependent R-D models, bit allocation, H.264/scalable video coding (SVC), spatial scalability.

1. INTRODUCTION

Scalable video coding (SVC) is an extension of the H.264/AVC standard that supports the video content adaptation to meet the heterogeneity of user devices and networks. It encodes an input video sequence once but allows the decoding of partial bit streams according to the target application requirements. Thus, SVC can be efficiently used to support the various capabilities of modern video applications such as conferencing, monitoring, surveillance, streaming and broadcasting, including variations in display resolution and available bandwidth.

H.264/SVC supports temporal, spatial and quality scalability modes. For spatial scalability, the lowest spatial resolution layer, the base layer (BL), is compatible with the H.264/AVC bit stream. The higher spatial resolution layers, the enhancement layers (ELs), are encoded using the inter-layer prediction technique. It allows using information from lower layers in the prediction of the enhancement layers. Therefore, the redundancies between spatial layers are discarded and hence the total coding performance is improved.

Rate control (RC) is an important stage in the video coding process. It adjusts the output bitrate of the coded bit stream to get the best tradeoff between the reconstructed video quality and the actual channel bandwidth. Two types of RC approaches have been distinguished with respect to the end application: constant bit rate (CBR) and variable bit rate (VBR) control algorithms. For the CBR coding, numerous RC schemes have been recommended in the video coding standards such as the Verification Model Version 8 for MPEG-4, the Test Model Version 8 for H.263, and Joint Model for H.264/AVC. Rate control has been extensively studied for single layer video coding [1]. Although the H.264/AVC rate control schemes can be directly applied to the H.264/SVC base layer, their efficiency drastically decreases when applied at the enhancement layers. This is because they are not designed to deal with inter-layer dependency.

Recently, several RC algorithms have been developed for SVC [2-8], including the temporal layer RC algorithms and spatial layer RC algorithms. For temporal-layer RC, Xu *et al.* [2] proposed a frame level rate control for hierarchical B-picture. Based on Scaling-factors, the target bits are allocated for each frame and the quantization parameters (QPs) are estimated. Cho *et al.* [3] proposed a group of pictures (GOP)-based dependent distortion model that takes into account the inter-dependency between the temporal layers. Then, a multi-pass RC algorithm is applied to yield a good rate-distortion (R-D) performance tradeoff.

For spatial layer RC, it is a very challenging problem because of the inter-layer prediction which makes the R-D characteristics of one enhancement layer dependent on its preceding layers. Even the current SVC reference software supports RC for the base layer only. It adopts a bottom-up scheme to generate a scalable bit stream [2]. The encoding process is performed in an ordered path, starting with the lower BL to the succeeding ELs.

Xu *et al.* [4] proposed a RC scheme for spatial and coarse-grain-SNR (CGS) scalable coding based on an improved TMN8 model and two-pass refinement of QP. However, the inter-dependency between spatial layers is not addressed at all because of the complexity concern which leads to poor RC performance. Some recent researches addressed this shortcoming and investigated the spatial

H.264/SVC bit allocation problem regarding these inter-layer dependencies. Liu *et al*. [5] proposed a rate control scheme that can be utilized in temporal, spatial, CGS, and combined scalable ELs. Inter-layer dependency was taken into account and an adaptive mean absolute difference (MAD) prediction method was used for the linear rate-quantization (R-Q) model. Nevertheless, the accurate control is at the expense of the computational complexity. Recently, Hu *et al*. [6] proposed a spatial-layer RC algorithm. First, an adaptive Qp initialization method for both BL and ELs is introduced. Then, a two-stage Qp estimation schema is designed to improve the RC performance. This was achieved by implementing a frame complexity prediction method and an adaptive model-parameter technique. Jing *et al.* [7] proposed an adaptive R-Q model for spatial enhancement layer. An efficient coding complexity estimation method is introduced by taking into account the inter-layer dependency between the spatial layers without any pre-encoding process. Liu *et al.* [8] developed a multi-pass model-based spatial layer RC algorithm. The relationship between the BL and ELs was taken into consideration. Most of the mentioned algorithms are inconvenient for real time applications because of their high computational complexity.

Therefore, in this paper, a single pass frame-based bit allocation algorithm for spatial SVC is presented. The proposed rate control algorithm initializes and updates the R-D model parameters presented in [8] during a single pass. Thus the computational complexity problem associated with the estimation of the R-D model parameters of [8] in three passes is avoided with the proposed algorithm. Furthermore, the estimated parameters in [8] are not updated during the encoding of each frame in the video sequence that may lead to the decrease of the coding efficiency. Therefore in the proposed algorithm the initialized values of the R-D model parameters are updated after the encoding of each frame.

The rest of this paper is organized as follows. The spatial layer bit allocation problem is formulated in Section 2. The proposed bit allocation algorithm with adaptive R-D model is described in Section 3. The simulation results are presented in Section 4. Finally, the paper is concluded in Section 5.

2. PROBLEM FORMULATION

Most RC schemes suitable for real-time applications employ R-D models for QP estimation. In this section, we consider the bit allocation problem for the two-layer case (i.e., N = 2). Since the H.264/SVC base layer is compatible with H.264/AVC and the bit allocation via Cauchy-density-based R-D models is one of the most precise solutions at frame level. Therefore, the RD characteristics of the H.264/SVC base layer denoted by $D_1(Q_1)$ and $R_1(Q_1)$ can be described using the models proposed by Kamaci *et al.* [1]:

$$D_1(Q_1) = b \cdot (Q_1)^{\beta_1} \text{ and } R_1(Q_1) = a \cdot (Q_1)^{-\alpha} \quad (1)$$

where Q_1 is the BL quantization step-size, and a, b, α and β_1 are R-D model parameters. It is obvious from (1) that the rate and distortion of the BL is primarily determined by its own quantization step-size Q_1.

For the enhancement layer, the R-D characteristics denoted by $D_2(Q_1,Q_2)$ and $R_2(Q_1,Q_2)$ were derived in [7] as a function of the quantization step-size of the base layer Q_1 and the enhancement layer Q_2 as follows:

$$D_2(\mathrm{Q}_1, \mathrm{Q}_2) \approx (\zeta\,\mathrm{Q}_1 + \upsilon) \, . \, (Q_2)^{\beta_2} \quad (2)$$

$$R_2(Q_1, Q_2) = \begin{cases} r.\,R_1(Q_1) + (\mathrm{s}-\mathrm{r}).\,R_1(\mathrm{Q}_2/2), & \mathrm{Q}_2 \leq 2\mathrm{Q}_1 \\ s.\,R_1(\mathrm{Q}_2/2) & ,\mathrm{Q}_2 \geq 2\mathrm{Q}_1 \end{cases} \quad (3)$$

where ζ, υ, and β_2 are the distortion model parameters. s and r are the slopes of the rate model line when $QP2 = QP1 + 6$ and $QP2 \leq QP1 + 6$, respectively.

Once the BL and EL R-D models are selected, the objective is to obtain the optimal quantization step-size for each spatial layer that minimizes the total distortion under a given bitrate constraint. The optimal solution can be obtained by differentiating the Lagrangian cost function with respect to Q_1, Q_2, and λ to yield the three equations as in [8]:

$$b\,\beta_1 \, . \, Q_1^{(\beta_1-1)} + \zeta \, . \, Q_2^{\beta_2} - a\alpha\,(1+r)\,Q_1^{(-\alpha-1)} \, . \, \lambda = 0$$

$$\upsilon\,\beta_2 \, . \, Q_2^{(\beta_2-1)} - 1/2\,a\alpha \, . \, (s-r)(Q_2/2)^{(-\alpha-1)} \, . \, \lambda = 0$$

$$a \, . \, (1+r)\,Q_1^{-\alpha} + a \, . \, (s-r)\,(Q_2/2)^{-\alpha} - R_T = 0 \quad (4)$$

where R_T is the bit budget of the current frame, and λ is the lagrange multiplier. Newton method can be used to numerically solve these non-linear equations to obtain the values of Q_1 and Q_2. Then, the value of QP_1 and QP_2 can be directly determined and assigned to the corresponding frame layer to be actually encoded.

According to [8], the bit allocation algorithm requires four iterations for encoding each video sequence. Three encoding passes are needed to build all required R and D models. Computing the model parameters has to be done in the pre-encoding stage. In addition, it is needed to encode a video sequence once to meet the target bit budget. Consequently, the pre-processing encoding iterations yield a high computational complexity. Moreover, the model parameters obtained at the pre-encoding stage are not updated during the encoding of the video sequence which significantly degrades the performance of the bit allocation algorithm [8].

In the following section, the presented rate control algorithm for two layer spatial scalability video coding is described. Rather than estimating the model parameters by encoding the video sequence three times as in [8], the proposed algorithm initializes and updates the R-D model parameters. This causes our scheme to be single-pass rate control algorithm and hence more convenient for real-time applications.

3. THE PROPOSED BIT ALLOCATION WITH ADAPTIVE R-D MODELS

3.1. Initial model parameters estimation

In order to solve the three equations in (4), the initial values of the R-D model parameters for the two spatial layers of the SVC encoder must be defined. The frame coding structure employed in this paper is IPPP structure. After encoding the first I- and P-frames with initial quantization parameters, the BL bits (R_1^P) and distortion (D_1^P) obtained from encoding the lower resolution P-frame are calculated. The bits and distorition (R_2^P, D_2^P) of the corresponding EL frame are also calculated using the BL and EL quantization step sizes (Q_1^P, Q_2^P).

Now we consider the initialization of model parameters using the first set of pictures. In the Cauchy-based BL R-D models (1), parameter α is restricted to a set of predefined fixed values which is identified for the P frames as in [1]:

$$\alpha_P = \begin{cases} 1.2, & if\ R_1^P/N_P > 0.1 \\ 1.6, & if\ R_1^P/N_P < 0.05 \\ 1.4, & else \end{cases} \tag{5}$$

where N_P is the number of pixels per frame (e.g., for a 4:2:0 format QCIF sequence, $N_P = 176 \times 144 \times 1.5$).

Several experiments were performed to estimate the values of the distortion parameters, it was found that parameters β_1 and β_2 are restricted to the two sets which are given as:

$$\beta_1^P = \begin{cases} 1.4, & if\ \ D_1^P/N_P < 30 \\ 1.6, & if\ \ 30 < D_1^P/N_P < 40 \\ 1.8, & if\ \ D_1^P/N_P > 40 \end{cases} \tag{6}$$

$$\beta_2^P = \begin{cases} 1.0, & if\ \ D_2^P/N_P < 20 \\ 1.2, & if\ \ 20 < D_2^P/N_P < 35 \\ 1.4, & if\ \ D_2^P/N_P > 35 \end{cases} \tag{7}$$

Since there is a high correlation between the statistics of two contiguous frames, the information from one frame can be used as a basis to encode the next frame. Therefore, the information from the previously encoded P-frame can be employed to initialize the BL R-D model parameters a_p and b_p as follows:

$$a_P = \frac{R_1^P}{(Q_1^P)^{-\alpha_P}} \tag{8}$$

$$b_P = \frac{D_1^P}{(Q_1^P)^{\beta_1}} \tag{9}$$

The EL R-D model parameters ζ and υ for can be predicted from previously encoded I- and P-frames using

$$\zeta_P = \left(\frac{D_2^P}{(Q_2^P)^{\beta_2}} - \frac{D_2^I}{(Q_2^I)^{\beta_2}}\right) / (Q_1^P - Q_1^I) \tag{10}$$

$$\upsilon_P = \frac{D_2^P}{(Q_2^P)^{\beta_2}} - \zeta\, Q_1^P \tag{11}$$

where Q_1^I and Q_2^I are the quantization step sizes for the first I-frame at the base and enhancement layers respectively, and D_2^I is the output distortion for the first I-frame at the EL.

For the EL rate model parameters, *r* and *s*, it was found empirically by experiments on various video sequences that the values of these parameters are set to -0.25 and 1.1, respectively. By using the equations from (5) to (11) along with the chosen values of *r* and *s* , the BL and EL quantization parameters can be estimated with no need for the pre-encoding stage.

3.2. Parameter update

Since the distribution of the DCT coefficients changes from one frame to the other in the same video sequence, it is required to update the R-D model parameters *a, b,* ζ, and υ for the video sequence frames by exploiting previously coded frames. Practically, for the remaining P-frames in the video sequence, the R-D model parameters are updated as follows:

1. The rate and distortion parameters which are linearly dependent on (i.e., α, β_1, β_2, *r* and *s*) are set as mentioned in the previous subsection.
2. The rate and distortion parameters which are exponentially dependent on (i.e., *a, b,* ζ, υ) are updated as follows:

$$a_P = \delta \times a_P + (1-\delta)\frac{R_1^P}{(Q_1^P)^{-\alpha_P}} \tag{12}$$

$$b_P = \delta \times b_P + (1-\delta)\frac{D_1^P}{(Q_1^P)^{\beta_1}} \tag{13}$$

where δ is a factor, which can be set to 0.5. In (12) and (13), the values of a_P and b_P are acquired from the last encoded frame in the same layer. The values of ζ_P and υ_P are updated using:

$$\zeta_P = \delta \times \zeta_P + (1-\delta)\left[\left(\frac{D_2^P}{(Q_2^P)^{\beta_2}} - \upsilon_P\right) / Q_1^P\right] \tag{14}$$

$$\upsilon_P = \delta \times \upsilon_P + (1-\delta)\left(\frac{D_2^P}{(Q_2^P)^{\beta_2}} - \zeta_P . Q_1^P\right) \tag{15}$$

Finally, after encoding each enhancement P-frame, the model parameters at the frame level have to be updated for the next frame bit allocation.

3.3. Frame based bit allocation algorithm

The pseudo-code of the proposed rate control algorithm with the adaptive R-D models with constant initial parameters is summarized below:

1. Initialization:
2. Compute initial QP values: QP_0^I, QP_0^P
3. QP_0^I = initial QP, and QP_0^I represents $QP_1^I = QP_2^I$.
4. $QP_0^P = QP_0^I + 1$, and QP_0^P represents $QP_1^P = QP_2^P$.
5. Initialize the model parameters: using equations (5)-(11)
6. Initialize the *remaining bits* (R) and *current buffer fullness* (C).
7. $R = N_{GOP} \times Bit\ Rate\ /\ Frame\ Rate$
8. $C = 0$
9. **For** each frame j = 1 to J, do
10. **For** each layer k = 0 to 1, do
11. **If** k = 0 **then** {Base Layer}
12. **If** first I/P picture **then**
13. **Go to** step 17
14. **Else**
15. Target frame bit estimation.
16. QP_1^P, QP_2^P numerical calculation: using (4)
17. Encode picture, and calculate actual BL output bits ($R_1^{I/P}$)
18. **If** k = 1 **then** {Enhancement Layer}
19. Encode picture, and calculate actual EL output bits ($R_2^{I/P}$),
20. Update *remaining bits*, *current buffer fullness*.
21. $R\ -= (R_1^{I/P} + R_2^{I/P} - Bit\ Rate\ /\ Frame\ Rate)$
22. $C\ += (R_1^{I/P} + R_2^{I/P} - Bit\ Rate\ /\ Frame\ Rate)$
23. **If** first I/P picture **then**
24. **Go to** step 9
25. **Else**
26. Update model parameters: using (12)-(15)
27. **End For**
28. **End For**

4. SIMULATION RESULTS

Several experiments have been conducted to evaluate the performance of the proposed rate control algorithm. The two-spatial-layer bit allocation scheme was incorporated into the Joint Scalable Video Model JSVM 9.19. The performance of the proposed algorithm was tested using 3 different video sequences of BL-QCIF@15Hz and EL-CIF@15Hz at various target rates. The frame coding structure was that only the first frame in each spatial layer was intra coded and all subsequent frames were coded as P-frames. The high resolution layer was encoded using the adaptive inter-layer prediction from the base layer and CABAC entropy coding. More simulation parameters are summarized in Table 1, while the other encoding parameters were set to the default values defined in the JSVM software.

The performance of the proposed algorithm was compared to the RC algorithm [8] named as Liu in the rest of this paper as well as the JSVM FixedQPEncoder tool. Table 2 summarizes the encoding results of the three algorithms. It can be seen that the proposed rate control algorithm outperforms both of Liu's algorithm and JSVM FixedQPEncoder tool. It achieves an averaged PSNR gain of 0.53 dB and 1.5 dB over Liu's algorithm and JSVM FixedQPEncoder algorithm, respectively. The table also shows the iteration numbers of the three rate control algorithms. It is clear that the proposed algorithm achieves higher coding gain with a single iteration (pass) (i.e. with lower computational complexity as compared to the other algorithms). Moreover, both the proposed algorithm and the Liu's algorithm control the target bit rate more accurately and keep the actual bit rate closer to its target.

Fig.1 shows the frame-by-frame PSNR values comparison of the three spatial bit allocation algorithms. From this figure, it is shown that the proposed bit allocation algorithm outperforms the JSVM FixedQPEncoder tool significantly while the gap between it and the Liu's algorithm is within 0.6 dB. Fig. 2 illustrates the buffer occupancy at each frame of the proposed algorithm against Liu's algorithm. It is shown that the proposed algorithm is able to keep the buffer status in a stable level away from overflow and underflow. The buffer occupancy is around 50% when each frame has been encoded.

IntraPeriod	-1
NumberReferenceFrames	1
SearchMode	FastSearch
SearchRange	16
InitialQP	32

Table 1. Summary of the simulation parameters.

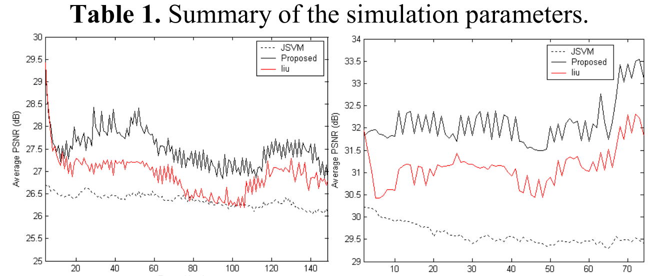

Fig. 1. Frame-by-frame PSNR values for the proposed, Liu, and JSVM. (a) Mobile, R_T = 256 kb/s and (b) Bus, R_T= 384 kb/s.

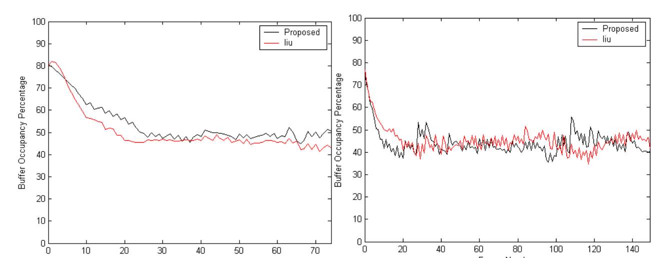

Fig. 2. Buffer occupancy as a function of the frame number number. (a) Bus, R_T = 384 kb/s and (b) Foreman, R_T = 192 kb/s.

	Target Rate(kb/s)	Method	PSNR	PSNR Gain Over		Rate (kb/s)	Δ Rate	Iter.
				Liu's	JSVM			
Bus	384	Proposed	32.11	0.95	2.52	384.24	+0.24(0.06%)	1
		Liu's	31.16			380.75	-3.25	4
		JSVM	29.6			388.16	+4.16	27
	512	Proposed	33.59	0.98	2.32	513.32	+1.32 (0.26%)	1
		Liu's	32.61			508.74	-3.26	4
		JSVM	31.27			496.44	-15.56	22
	768	Proposed	35.69	0.84	1.73	768.24	+0. 24(0.03%)	1
		Liu's	34.85			763.22	-4.78	4
		JSVM	33.96			793.71	+25.71	39
Foreman	192	Proposed	36.03	0.06	1.35	190.94	-1.06 (0.55%)	1
		Liu's	35.97			191.09	-0.91	4
		JSVM	34.68			186.35	-5.65	20
	256	Proposed	37.25	0.09	1.16	254.66	-1.34 (0.52%)	1
		Liu's	37.16			254.81	-1.19	4
		JSVM	36.09			258.92	+2.92	8
	384	Proposed	38.73	0.05	0.93	381.96	-2.04 (0.53%)	1
		Liu's	38.68			383.32	-0.68	4
		JSVM	37.80			379.93	-4.07	20
Mobile	256	Proposed	27.60	0.57	1.23	255.01	-0.99(0.39%)	1
		Liu's	27.03			255.78	-0.22	4
		JSVM	26.37			251.14	-4.86	36
	384	Proposed	29.26	0.53	1.12	383.11	-0.89 (0.23%)	1
		Liu's	28.73			384.09	+0.09	4
		JSVM	28.14			376.97	-7.03	15
	512	Proposed	30.41	0.68	1.10	510.60	-1.40 (0.27%)	1
		Liu's	29.73			512.84	+0.84	4
		JSVM	29.31			508.78	-3.22	9

Table 2. Performance of the proposed, Liu, and JSVM bit allocation algorithms.

5. CONCLUSION

In this paper, a single-pass bit allocation algorithm was proposed for spatial scalability of H.264/SVC. The proposed rate control algorithm initializes and updates the R-D model parameters during a single pass. The initialized values of the R-D model parameters are updated after the encoding of each frame. Simulation results indicated that the proposed bit allocation achieves a significant improvement in the PSNR gain compared to the multi-pass model-based algorithm and the Joint Scalable Video Model reference software algorithm while maintaining the stability of the buffer status.

6. REFERENCES

[1] N. Kamaci, Y. Altinbasak, and R. M. Mersereau, "Frame bit allocation for H.264/AVC video coder via Cauchy-density-based rate and distortion models," IEEE Trans. Circuits Syst. Video Technol., vol. 15, no. 8, pp. 994–1006, Aug. 2005.

[2] L. Xu,W. Gao, X. Ji, and D. Zhao, "Rate control for hierarchical B-picture coding with scaling-factors," in Proc. IEEE ISCAS, pp. 49–52, May 2007.

[3] Y. Cho, J. Liu, D.-K. Kwon, and C.-C. J. Kuo, "H.264/SVC temporal bit allocation with dependent distortion model," in Proc. IEEE ICASSP, pp. 641–644, Apr. 2009.

[4] L. Xu, W. Gao, X. Ji, D. Zhao, and S. Ma, "Rate control for spatial scalable coding in SVC," in Proc. Picture Coding Symp., Nov. 2007.

[5] Y. Liu, Z. Li, and Y. C. Soh, "Rate control of H.264/AVC scalable extension," IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 1, pp. 116–121, Jan. 2008.

[6] S. Hu, H. Wang, S. Kwong, and C.-C. J. Kuo, "Novel rate quantization model-based rate control with adaptive initialization for spatial scalable video coding," IEEE Trans. Industrial Electronics, vol. 59, no. 3, pp. 1673-1684, Mar. 2012.

[7] X. Jing, J. Y. Tham, Y. Wang, K. H. Goh, and W. S. Lee "Efficient Rate-Quantization Model for Frame Level Rate Control in Spatially Scalable Video Coding," pp.339-343, Dec. 2012.

[8] J. Liu, Y. Cho, Z. Guo, and C.-C. J. Kuo, "Bit allocation for spatial scalability coding of H.264/SVC with dependent rate distortion analysis," IEEE Trans. Circuits Syst. Video Technol., vol. 20, no. 7, pp. 967-981, Jul. 2010.