FRAME-BASED BIT ALLOCATION FOR SPATIAL SCALABILITY IN H.264/SVC

Jiaying Liu[1*], *Yongjin Cho*[2] *and Zongming Guo*[1]

[1]Institute of Computer Science and Technology, Peking University, Beijing, P.R. China 100871
[2]Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089

ABSTRACT

The spatial scalability of H.264/SVC is achieved by a multi-layer approach, where an enhancement layer is dependent on its preceding layers. To address this dependent issue, we propose a model-based spatial layer bit allocation algorithm for H.264/SVC in this work. The inter-layer dependency is decoupled by analyzing the signal flow in the H.264/SVC encoder. We show that the rate and the distortion (R-D) characteristics of a dependent layer with a frame as a basic coding unit. Finally, a low complexity spatial layer bit allocation scheme is developed using the proposed frame-based R-D models. It is shown by experimental results that our proposed bit allocation algorithm can achieve the coding performance close to the optimal R-D performance of full search and is highly improved from current reference codec JSVM.

Index Terms— Frame-based bit allocation, dependent rate and distortion model, H.264/SVC

1. INTRODUCTION

Scalable video coding (SVC) has recently been standardized to extend the capabilities of the H.264/AVC standard [1]. It addresses the application need of a more flexible format of coded video in heterogeneous and time-varying environments. H.264/SVC supports spatial, temporal and quality scalabilities while keeping a good balance in decoder complexity and coding efficiency.

In this work, we will focus on the bit allocation problem for the spatial scalability, which is a challenging task due to inter-layer dependency. The H.264/SVC reference codec JSVM specifies a bottom-up approach to produce a scalable bit stream. That is, the encoding process starts from the bottom-most base layer (BL) and subsequent enhancement layers (EL's) are encoded in an ordered manner. However, the current version of JSVM does not support any encoding tool for bit allocation among spatial layers.

Bit allocation algorithms for inter-frame dependency have been examined since MPEG. For example, Ramachandran *et al.* studied the dependent bit allocation problem with a trellis-based solution framework in [2]. Lin and Ortega [3] speeded up the scheme by encoding the source with only a few quantization steps and using interpolation to find the rate distortion value for other quantization steps. However, the complexities of these algorithms are extremely high. So they can not be practically extended to solve a dependent bit allocation problem involved with multiple layers in H.264/SVC. In the latest bit allocation algorithms proposed for H.264/SVC, the property of inter-layer dependency is not properly addressed in the problem formulation. Liu *et al.* [4] proposed a rate control algorithm for the spatial and coarse-grain SNR (CGS) scalability of H.264/SVC. The proposed algorithm operates on a fixed rate of each layer and implements an MB-layer bit allocation scheme. The spatial-layer bit allocation problem is not addressed at all. To solve this inter-layer dependent issue, a GOP-based bit allocation problem for the spatial scalability of H.264/SVC was studied in the authors' previous work [5] with two spatial layers. The GOP-based bit allocation algorithm demands a longer delay in the encoding process and, therefore, it is not suitable for real-time conversational applications.

In this paper, we focus on allocating bits among spatial dependent layers with a frame as the basic coding unit. Please note the frame here is a set of the BL frame and corresponding EL frames, which correspond to the same position in each spatial layer. After analyzing the input visual signals to the EL quantizer, we use the Cauchy density to estimate the distribution of ac coefficients of EL input sequence, and find that the parameter in Cauchy probability density function (pdf) is related with BL quantization. Thus, the impact of the EL quantization is isolated in a frame-based distortion model. Based on the observation of the rate dependency, an approximated frame-based rate model is presented. To provide a simple yet effective model of the R-D characteristics of dependent layers, the details are given in Sec.3. The proposed algorithm allows to allocate bits simultaneously while providing a trade-off between BL and ELs so as to improve the efficiency of spatial scalable coding.

The rest of the paper is organized as follows. The frame-based spatial layer bit allocation problem is formulated in Sec.2. The dependent frame-based R-D models are analyzed and simplified in Sec.3. The bit allocation algorithm is proposed in Sec.4. Experimental results are given in Sec.5. Finally, concluding remarks are given in Sec.6.

2. PROBLEM FORMULATION

Let N be the number of spatial layers in a frame. $R_k(Q_1,\ldots,Q_k)$ and $D_k(Q_1,\ldots,Q_k)$ are the rate and distortion model of the k-th layer with respect to a quantization vector $(Q_1,\ldots,Q_k)$. Given the bit budget R_T of the current encoding frame, the bit allocation problem can be formulated as

$$\mathbf{Q}^* = (Q_1^*,\ldots,Q_N^*) = \arg\min_{Q_k\in\mathcal{Q}} \sum_{k=1}^{N} \omega_k \cdot D_k(Q_1,\cdots,Q_k) \tag{1}$$
$$s.t. \sum_{k=1}^{N} R_k(Q_1,\ldots,Q_k) \le R_T,$$

where $\mathbf{Q}^* = (Q_1^*,\ldots,Q_N^*)$ is the selected Q vector for all spatial layers, $\mathcal{Q}$ is the set of all quantization candidates, and ω_k is the weighting factor representing the corresponding importance of the k-th layer.

The Lagrangian multiplier method converts the constrained optimization problem in Eq.(1) to an equivalent unconstrained optimization problem by introducing the Lagrange cost function as

[*] This work was supported by National Science Foundation of China and Microsoft Research Asia under 60833013 and National Basic Research Program (973 Program) of China under contract No. 2009CB320907. The authors also acknowledge the generous support of Prof. C.-C. Jay Kuo at the University of Southern California.

978-1-4244-4291-1/09/$25.00 ©2009 IEEE

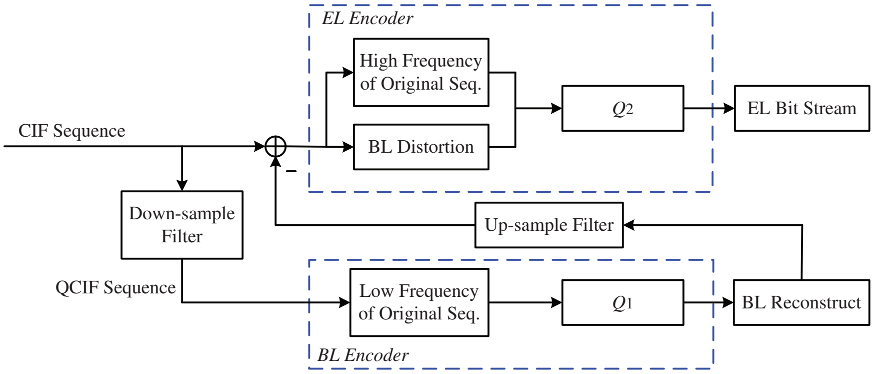


Fig. 1. Decomposition of the H.264/SVC encoder diagram.

$$\mathbf{Q}^* = \arg\min_{Q_k \in \mathcal{Q}} J(\mathbf{Q}, \lambda),$$
$$J(\mathbf{Q}, \lambda) = \sum_{k=1}^{N} \omega_k \cdot D_k(\cdot) + \lambda \cdot \left(\sum_{k=1}^{N} R_k(\cdot) - R_T \right), \tag{2}$$

where λ is the Lagrange multiplier.

Without loss of generality, we consider bit allocation in a simple two-layer scenario first, and assume the equal importance of these two layers; namely, $\omega_1 = \omega_2 = 1$. The solution can be easily generalized to a multi-layer scenario. Mathematically, the Lagrange cost function is expressed as

$$\begin{aligned} J(\mathbf{Q}, \lambda) = D_1(Q_1) + D_2(Q_1, Q_2) \\ +\lambda \cdot \left[(R_1(Q_1) + R_2(Q_1, Q_2) - R_T \right]. \end{aligned} \tag{3}$$

3. DISTORTION AND RATE MODELING OF DEPENDENT LAYERS

Generally speaking, the R-D characteristics of a dependent layer are represented by a function having the quantization step sizes of the reference layer and the dependent layer as the variables. The impact of an individual variable on the R-D characteristics of a dependent layer has to be known to solve the bit allocation problem.

3.1. Frame-based Distortion Modeling

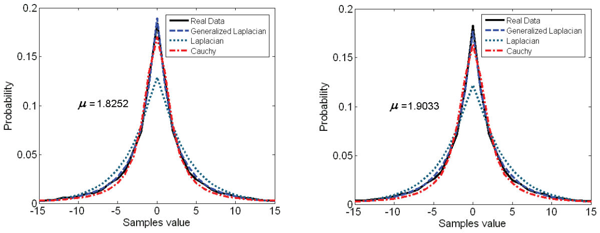
It is a challenge in distortion modeling of dependent layers; namely, to determine the impact of each individual variable on the distortion of the target layer. To achieve this goal, we attempt to analyze the processing of the input video signal in the H.264/SVC encoder. In Fig. 1, we depict the H.264/SVC encoder whose input is a CIF sequence and output is a bit stream consisting of two spatial layers, *i.e.*, a BL and a dependent EL. We are mainly interested in the input and the output of the EL quantizer (*i.e.,* Q_2) in Fig. 1.

We first obtain a low frequency component of the input CIF video by the down-sampling process. The lowpass filtered signal is fed into the BL encoder to produce the BL reconstruction signal, which corresponds to a quantized version of the low-frequency video using quantization step size Q_1. The reconstructed BL is used as a basis to predict the low frequency component of the input to reduce inter-layer redundancy. Then, we use the differential signal between the original and the interpolated BL signals as the input to the EL encoder. This differential signal actually consists of two parts: 1) the high frequency component and 2) the distortion in the low frequency component due to the quantization effect in the BL.

Since the frame is used as the basic unit, we can consider using the distributions of the DCT coefficients of each frame to describe the feature of EL differential signal. By testing and considering both the accuracy and the complexity issues, we approximate the DCT coefficients distribution successfully by zero-mean Cauchy distribution with parameter μ, having the pdf

$$p(x) = \frac{1}{\pi} \frac{\mu}{\mu^2 + x^2}, \quad x \in \mathbf{R}, \tag{4}$$

where the parameter μ depends on the EL input differential signal.

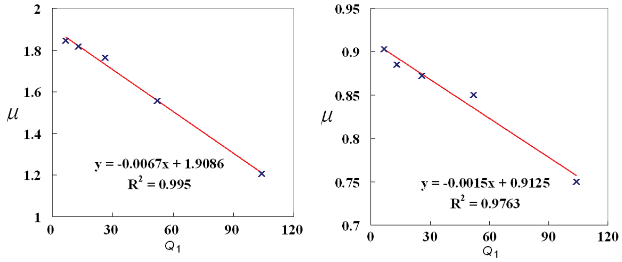


(a) City, $Q_1 = 14$

(b) City, $Q_1 = 32$

Fig. 2. DCT coefficients fitting result of different EL input signal based on different Q_1.

Fig. 2 illustrates the curve fitting results with Fig. 2(a) and Fig. 2(b) depicting the influence of different Q_1. The main difference between the coefficients distribution figures of different EL input signal is the height of pdf, which corresponds to different parameter values of μ. Thus, we further analyze the relationship between the only variable μ in Cauchy pdf and BL quantization step size to investigate the relationship between Q_1 and the EL input differential signal.



(a) City

(b) Foreman

Fig. 3. Illustration of linear relationship between μ and Q_1.

We observe the relationship between the parameter μ in Cauchy pdf from the differential input and the BL encoder step size Q_1 for various test sequences shown in Fig. 3. The data in these figures indicate a linear model between μ and Q_1, which can be written as

$$\mu = \eta \cdot Q_1 + \varphi, \tag{5}$$

where η and φ are model parameters.

With the input signal to the EL encoder characterized by Eq. (5), the output of the EL encoder is influenced by the EL quantization step size Q_2 only. The distortion of EL caused by EL quantization can be estimated accurately for the Cauchy pdf assumption. Assume that we have a uniform quantizer with step size Q_2. The distortion caused by quantization is given by

$$D_2(Q_2) = \sum_{i=-\infty}^{\infty} \int_{(i-\frac{1}{2})Q_2}^{(i+\frac{1}{2})Q_2} |x - iQ_2|^2 p(x)dx. \tag{6}$$

It can be shown that this infinite sum converges and is bounded from above by $Q_2^2/4$. For a Cauchy source, this expression becomes

$$D_2(Q_2) = 2\sum_{i=1}^{M}\left[\frac{\mu Q_2}{\pi} - \frac{i\mu Q_2}{\pi}\ln\left(\frac{\mu^2 + (i+\frac{1}{2})^2 Q_2^2}{\mu^2 + (i-\frac{1}{2})^2 Q_2^2}\right) - \frac{\mu^2 - i^2 Q_2^2}{\pi} \times tan^{-1}\left(\frac{\mu Q_2}{\mu^2 + (i^2 - \frac{1}{4})Q_2^2}\right)\right] + \left[\frac{\mu Q_2}{\pi} - \frac{2\mu^2}{\pi} tan^{-1}\left(\frac{Q_2}{2\mu}\right)\right].$$

In [6], the authors suggest that the distortion depends on μ in addition to Q_2. Although this equation is highly complex, it can be approximated. An approximation leads to

$$D_2(Q_2) \approx b \cdot Q_2^{\beta}, \tag{7}$$

where b is the parameter that depends on μ. Once the sequence is fixed, the value of β is almost constant. They point out that they can solve b as the least square error solution off-line, and Table 1 shows values of b for a possible set of μ values.

Table 1. Parameter b values for corresponding μ

μ	b
1.5634	0.2501
2.8522	0.2087
3.0980	0.1976
3.2928	0.1889

Considering above the analysis on both linear relations, μ is the function of Q_1 and b is function of μ, we can propose the following distortion model of dependent EL layer,

$$D_2(Q_1, Q_2) \approx (\zeta Q_1 + \upsilon) \cdot Q_2^{\beta}, \tag{8}$$

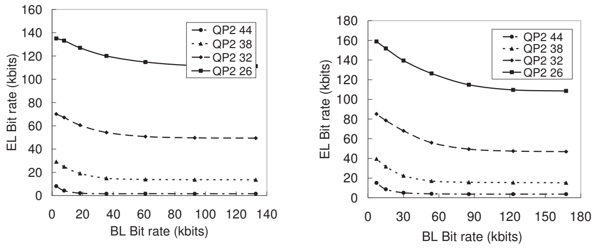
where ζ, υ, and β are empirical parameters.

3.2. Frame-based Rate Model

To derive the frame-based rate model of EL, denoted by $R_2(Q_1, Q_2)$, we plot the bit rate of EL, at four different Q_2 values with respect to the variation rates of the referenced BL. In Fig. 4, it is shown that for each fixed Q_2, increasing $R_1(Q_1)$ (i.e., decreasing Q_1) results in roughly a linear relation decreases in $R_2(Q_1, Q_2)$. However, R_2 does not decrease further beyond the point where $(QP_1 - 6)$ equals to QP_2. That is, the corresponding quantization step size is halved, i.e., $Q_1 = 2Q_2$.

Based on the above observation, we conclude that the rate of a dependent layer can be approximated as

$$R_2(Q_1, Q_2) = \begin{cases} r \cdot R_1(Q_1) + (s - r)R_1(Q_2), & Q_1 \geq 2Q_2, \\ s \cdot R_1(Q_2), & Q_1 < 2Q_2, \end{cases} \tag{9}$$



(a) City, CIF-QCIF (b) Football, CIF-QCIF

Fig. 4. Illustration of rate dependency between two layers.

where s and r are the slope of the line when $Q_1 = 2Q_2$ and $Q_1 \geq 2Q_2$, respectively. The proposed rate model is plotted in Fig. 5.

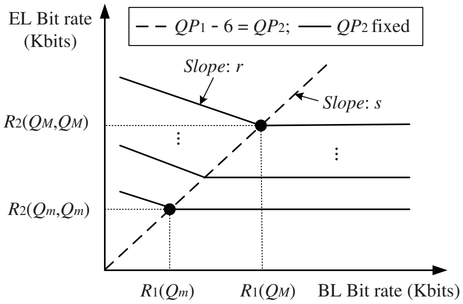


Fig. 5. The proposed rate model for dependent layers

4. PROPOSED SPATIAL-LAYER BIT ALLOCATION ALGORITHM

We address a spatial layer bit allocation problem for H.264/SVC based on the proposed dependent rate and distortion model. This modeling approach enables an analytical solution to the Lagrange equation. For the single variable rate and distortion models, we employ the models proposed by Kamaci *et al.* [6] of the following form:

$$R(Q_i) = a \cdot Q_i^{-\alpha}, \text{ and } D(Q_i) = b \cdot Q_i^{\beta}, \tag{10}$$

where Q_i is the quantization step of BL, a, b, α and β are model parameters.

Then, the Lagrange cost function can be solved with a closed-form solution by applying the proposed frame-based R-D models as given in Eqs. (8) and (9). That is,

$$\begin{aligned} J(\mathbf{Q}, \lambda) = b \cdot Q_1^{\beta_1} + (\zeta Q_1 + \upsilon) \cdot Q_2^{\beta_2} \\ + \lambda \cdot [(1 + r) \cdot aQ_1^{-\alpha} + (s - r) \cdot aQ_2^{-\alpha}]. \end{aligned} \tag{11}$$

To optimize Eq. (11), we first take the partial derivatives with respect to Q_1 and Q_2, which yields the following two equations

$$\begin{aligned} b\beta_1 \cdot Q_1^{(\beta_1 - 1)} + \zeta \cdot Q_2^{\beta_2} - a\alpha(1 + r)Q_1^{-\alpha - 1} \cdot \lambda = 0, \\ \upsilon\beta_2 \cdot Q_2^{(\beta_2 - 1)} - a\alpha(s - r)Q_2^{-\alpha - 1} \cdot \lambda = 0. \end{aligned} \tag{12}$$

Another equation can be derived from the frame bit budget constraint R_T; namely,

$$R_T = (1 + r)a \cdot Q_1^{-\alpha} + (s - r)a \cdot Q_2^{-\alpha}. \tag{13}$$

Based on Eqs. (12) and (13), we can compute Q_1 and Q_2 values that optimize the Lagrange cost function given in Eq. (11). Finally, we can determine the encoder input parameters, QP_1 and QP_2, using the one-to-one correspondence between quantization step size Q and quantization parameter QP.

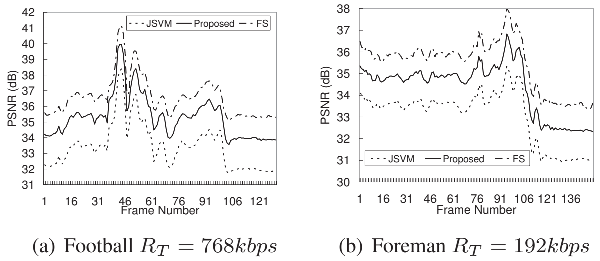
Table 2. Performance of two methods for QCIF-CIF two layers in terms of output rate, PSNR, Δ rate and iteration times.

Sequence	Target rate(kbps)	Method	$PSNR\ (dB)$	Rate (kbps)	Δ Rate	Iteration
Bus	512	Proposed	32.57	495.14	-16.86	4
		JSVM	31.74	527.28	+5.28	40
	768	Proposed	33.98	764.58	-3.42	4
		JSVM	32.18	788.86	+0.32	45
Football	768	Proposed	34.87	779.1	+10.9	4
		JSVM	32.98	776.04	+8.04	57
	1024	Proposed	37.24	1020.88	-3.12	4
		JSVM	34.99	1036.89	+12.89	26
Foreman	256	Proposed	36.85	254.42	-1.58	4
		JSVM	34.90	254.57	-1.43	36
	384	Proposed	38.12	388.23	+4.23	4
		JSVM	36.94	380.66	-3.34	37
Mobile	384	Proposed	29.04	382.65	-1.35	4
		JSVM	27.65	372.88	-11.12	27
	512	Proposed	31.24	520.13	+8.13	4
		JSVM	29.36	514.19	+2.93	39

5. EXPERIMENTAL RESULTS

Since there is no spatial layer rate control algorithm in the JSVM, the performance of the proposed algorithm is compared with that of the full search (FS) method. By testing all possible combinations of Q_1 and Q_2 for each spatial layer, an input video is iteratively encoded to find the best R-D performance at each target bit rate. The R-D curve obtained via FS provide the R-D performance bound with two-layer encoding. We also consider comparing with the results of JSVM FixedQPEncoder tool based on the SVC testing conditions JVT-Q205 defined in [7].

The proposed bit allocation algorithm was implemented with JSVM 9.6 in our experiment. The BL and the EL video sequences are QCIF and CIF formats, respectively. The GOP size is set to be 2, which is to avoid the effect of temporal hierarchical-B structure. The frame rate is 15 fps.



(a) Football $R_T = 768kbps$ (b) Foreman $R_T = 192kbps$

Fig. 6. PSNR versus frame for three bit allocation algorithms.

The performances of the proposed algorithm comparing with FS method and JSVM JVT-Q205 are shown in Fig. 6. We see the performance degradation of our method with respect to the optimal solution obtained by the optimal bound at various frame is not large. Significant coding gain could be observed for the proposed algorithm in comparison with JSVM JVT-Q205. Table 2 summarizes the encoding results of the proposed algorithm and JSVM JVT-Q205. The rate control method using the proposed frame-based bit allocation algorithm achieves an average of 1.67-dB PSNR gain over the one that uses current JSVM. It also shows the iteration times of current JSVM FixedQPEncoder. Comparing with fixed 4 times for the proposed algorithm, the complexity of the current JSVM implementation is extremely higher. These experimental results demonstrate the effectiveness and the robustness of the proposed algorithm for video sequences with various spatial characteristics.

6. CONCLUSION

We investigated a model-based spatial layer bit allocation scheme for H.264/SVC in this work. The inter-layer dependence of the distortion and the rate function were successfully modeled. Based on these models, we proposed a bit allocation algorithm that achieved a near optimal R-D performance.

7. REFERENCES

[1] Thomas Wiegand, Gary J. Sullivan, J. Reichel, Heiko Schwarz, and Mathias Wien, “Amendment 3 to ITU-T Rec. H.264 (2005) ISO/IEC 14496-10: 2005,” Scalable Video Coding, July 2007.

[2] Kannan Ramchandran, Antonio Ortega, and Martin Vetterli, “Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders,” *IEEE Trans. on Image Processing*, vol. 3, no. 5, pp. 533–545, Sep. 1994.

[3] Liang-Jin Lin and Antonio Ortega, “Bit-rate control using piecewise approximated rate-distortion characteristics,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 8, no. 4, pp. 446–459, August 1998.

[4] Yang Liu, Zhengguo Li, and Yeng Chai Soh, “Rate control of H.264/AVC scalable extension,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 18, pp. 116–121, Jan. 2008.

[5] JiaYing Liu, YongJin Cho, Zong-Ming Guo, and C.-C. Jay Kuo, “Bit allocation for spatial scalability in H.264/SVC,” in *IEEE International Workshop on Multimedia Signal Processing*, October 2008, pp. 278–283.

[6] Nejat Kamaci, Yucel Altinbasak, and Russel M. Mersereau, “Frame bit allocation for H.264/AVC video coder via Cauchy-density-based rate and distortion models,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 15, no. 8, pp. 994–1006, August 2005.

[7] Thomas Wiegand, Gary J. Sullivan, Heiko Schwarz, and Mathias Wien, “Joint draft 10 of SVC amendment,” Joint Video Team, JVT-Q205, San Jose, CA, USA, April 2007.