Bit Allocation for Spatial Scalability in H.264/SVC

Jiaying Liu †1, Yongjin Cho ††2, Zongming Guo †3 , C.-C. Jay Kuo ††4

† *Institute of Computer Science and Technology, Peking University, Beijing, P.R. China 100871*
{1 liujiaying, 3 guozongming}@icst.pku.edu.cn

†† *Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089*
{2 yongjinc, 4 cckuo}@sipi.usc.edu

***Abstract*—We propose a model-based spatial layer bit allocation algorithm for H.264/SVC in this work. The spatial scalability of H.264/SVC is achieved by a multi-layer approach, where an enhancement layer is bound by the dependency on its preceding layers. The inter-layer dependency is decoupled in our analysis by a careful examination of the signal flow in the H.264/SVC encoder. We show that the rate and the distortion (R-D) characteristics of a dependent layer can be represented by a number of independent functions with a group of pictures (GOP) as a basic coding unit. Finally, a low complexity spatial layer bit allocation scheme is developed using the proposed GOP-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 significantly improved from current reference software JSVM.**

I. INTRODUCTION

H.264/SVC is recently standardized as a scalable extension of H.264/AVC standard [1]. A SVC video stream is equipped with great flexibility and adaptability in terms of frame rates, display resolutions and quality levels. These characteristics of SVC video offer attractive solutions for service providers who are in face of a heterogeneous service environment. However, the provision of scalability is traded-off by lower coding efficiency, which has been the major reason that prevents scalable video from its practical employment.

In this work, we focus on the problem of spatial bit allocation for H.264/SVC, which is a challenging task due to inter-layer dependency. The H.264/SVC reference encoder (Joint Scalable Video Model, JSVM) specifies a bottom-up approach to produce a scalable bit stream [2]. 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 rate control 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 [3]. Although it can yield the optimal solution, the complexity of the solution grows exponentially as the number of dependent frames increases. For this reason, it can be used only as a performance benchmark rather than a practical solution. Lin and Ortega [4] 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. This scheme used the spline interpolation for I frames and the piecewise linear interpolation for P frames. However, the complexity of this algorithms is still high since the source video has to be encoded several times. So it 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.* [5] 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.

In this work, we consider the inter-layer bit allocation problem for H.264/SVC with spatial scalability using a model-based approach. To provide a simple yet effective model of the R-D characteristics of dependent layers in a GOP, we focus on the input and the output visual signals to the EL quantizer and propose a method to decouple the influence of each layer's quantization choice on the R-D characteristics of a dependent enhancement layer. Simply speaking, the complexity of the input signal to the EL quantizer is modeled as a function of the BL quantization and, as a result, the impact of the EL quantization is successfully isolated in the proposed R-D model. The details are given in Sec. III. The proposed algorithm allows to allocate bits simultaneously while providing a trade-off between base and enhancement layers so as to improve the efficiency of spatial scalable coding.

The rest of the paper is organized as follows. The spatial layer bit allocation problem is formulated in Sec. II. The dependent GOP-based R-D model is analyzed and simplified in Sec. III. The bit allocation algorithm is proposed in Sec. IV. Experimental results are given in Sec. V. Finally, concluding remarks are given in Sec. VI.

II. PROBLEM FORMULATION

The rate control of our proposed H.264/SVC encoder is achieved by the spatial layer bit allocation followed by the frame layer bit allocation. We first determine the bit budget allocation over spatial layers, and then assign each frame in the same spatial layer a certain amount of bit budget for its encoding. Since the frame-layer bit allocation problem has a lot of precedents in the literature, we mainly focus on the spatial layer bit allocation in this work.

In H.264/SVC, the spatial scalability is achieved by a multi-layer coding approach. That is, a video signal with a high spatial resolution is encoded in such a way that the output bit stream provides multiple layers of various spatial resolutions. When a bit budget constraint is imposed, it is essential for an encoder to efficiently distribute the bit budget to each spatial layer for the optimal coding efficiency. The rate and the distortion of a coded video stream are determined by the choice of quantization step-size Q. In the following, the problem is formulated as the spatial layer Q-decision problem.

Let N be the number of spatial layers. $R_k(Q_1, \ldots, Q_k)$ and $D_k(Q_1, \ldots, Q_k)$ are GOP-based distortion and rate model of the k-th layer with respect to a quantization vector $(Q_1, \ldots, Q_k)$. Given the bit budget R_{total} of a GOP, the bit allocation problem can be formulated as

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

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. Note that $Q_1, \cdots, Q_{k-1}$ in the R-D function of the k-th layer indicate that the coding performance of the k-th layer is dependent upon previous coded $(k-1)$ layers.

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

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

where λ is the Lagrangian multiplier.

Without loss of generality, we consider bit allocation in a simple two-layer scenario first. The solution can be easily generalized to a multi-layer scenario. Mathematically, the Lagrangian cost function can be expressed as

$$J(\mathbf{Q}, \lambda) = \omega_1 \cdot D_1(Q_1) + \omega_2 \cdot D_2(Q_1, Q_2)$$
$$+\lambda \cdot (R_1(Q_1) + R_2(Q_1, Q_2) - R_{total}). \tag{3}$$

In the following discussion, we assume the equal importance of these two layers; namely, $\omega_1 = \omega_2 = 1$.

To solve the dependent quantization decision problem in Eq. (3), one solution is to conduct a full search over all possible combinations of admissible quantization choices. However, since the search space grows exponentially as the number of layers increases, the complexity of full search is prohibitively large. We address the complexity issue by modeling the R-D characteristics of dependent layers in the next section.

III. Distortion and Rate Modeling of Dependent Layers

Generally speaking, the R-D characteristics of a dependent layer are represented by a function consisting of quantization step sizes for the reference layer and the dependent layer. The impact of an individual quantization parameter on the R-D characteristics of the dependent layer has to be known to solve the bit allocation problem. For dependent R-D modeling, we convert the multi-variable function into a number of single-variable functions which simplifies the solution framework to the bit allocation problem greatly.

A. GOP-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 base layer (BL)

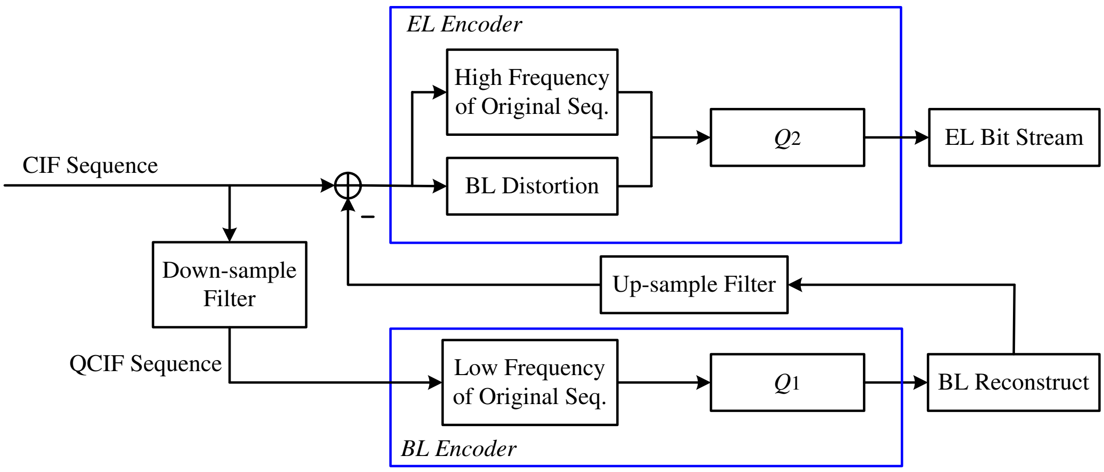


Fig. 1. Decomposition of the H.264/SVC encoder diagram.

and a dependent enhancement layer (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 extrapolated BL signals as the input to the EL encoder. Because of the influence of the BL encoding on the input to the EL encoder, the differential signal depends on the BL encoding with quantization step size Q_1.

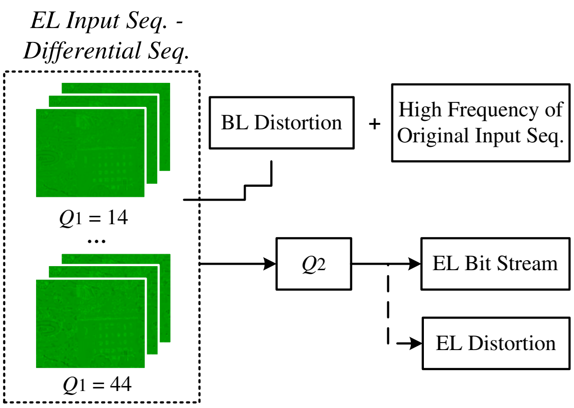


Fig. 2. Coding of the enhancement layer.

The 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, which is denoted as the "BL Distortion" in Fig. 2. The second part controls the coupling between BL's and EL's encodings. If the BL distortion term is much smaller than the high frequency term, such a coupling effect can be ignored.

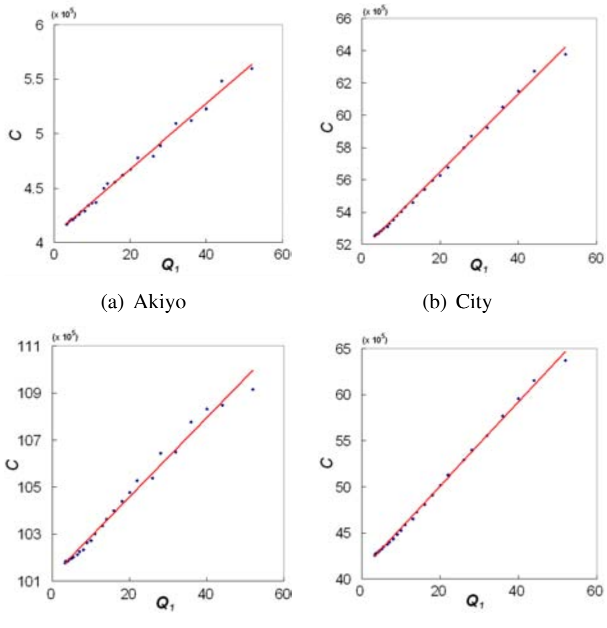
Fig. 2 illustrates the influence of different Q_1 values in the BL encoding on the EL encoding. The GOP complexity proposed in [6] can be used to capture the characteristics of the input signal into the EL quantizer. Thus, we adopt the GOP complexity estimation method given in [6] to investigate the relationship between Q_1 and the EL input differential signal.

Fig. 3 shows the relationship between the estimated GOP complexity (C) of the differential input and the BL encoder step size Q_1 for various test sequences. The data in these figures indicate a linear model between C and Q_1, which can be written as

$$C(Q_1) = c \cdot Q_1 + \theta, \tag{4}$$

where c and θ are model parameters.

With the input to the EL encoder characterized by Eq. (4), the output of the EL encoder is influenced by the EL quantization step size Q_2 only. Fig. 4 shows the distortion of the enhancement layer, denoted by $D_2(Q_1, Q_2)$, as a function



(a) Akiyo (b) City

(c) Mobile (d) Football

Fig. 3. Illustration of the linear relationship between the GOP complexity and Q_1.

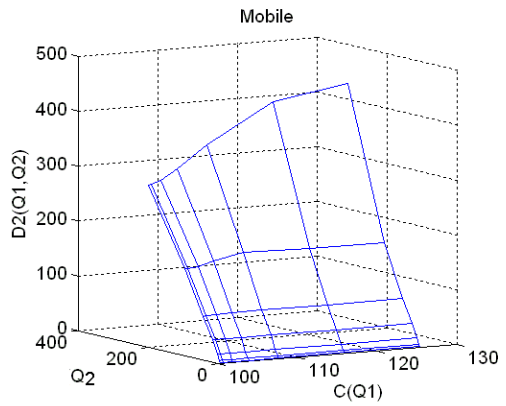
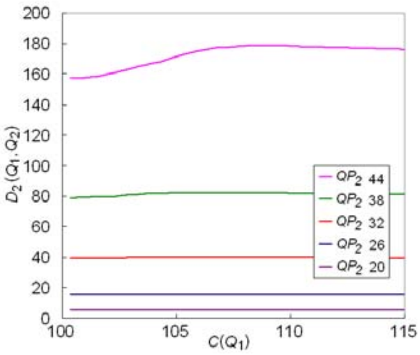


Fig. 4. The EL distortion as a function of the complexity C and quantization step size Q_2 for the Mobile sequence.

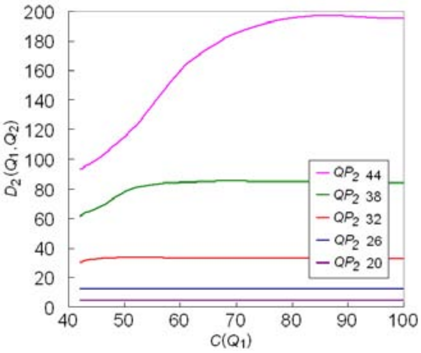
of $C(Q_1)$ and Q_2. In Fig. 5, we plot the projection of the 3D plot onto the $D_2 - C$ plane. We have the following main observations from these experiments:

1) For a fixed Q_2 value, we observe a linear region between between $D_2(Q_1, Q_2)$ and $C(Q_1)$ when $C(Q_1)$ is small.
2) The slope in the linear region reflects the dependency between layers since it is a function of Q_1 and Q_2.
3) Distortion $D_2(Q_1, Q_2)$ becomes flat after the inflection point which is located at $QP_1 \approx QP_2 - 6$, where QP_1 and QP_2 are quantization parameters for quantization step sizes Q_1 and Q_2, respectively.

When $QP_2 - QP_1 = 6$, we have that the corresponding quantization step size is halved, *i.e.,* $Q_1 = \frac{1}{2}Q_2$. Intuitively, the flat distortion phenomenon can be explained as follows. When Q_2 is relatively small as compared with Q_1, the EL distortion is primarily determined by Q_2 alone.



(a) Mobile



(b) Football

Fig. 5. Illustration of distortion dependency between two consecutive layers.

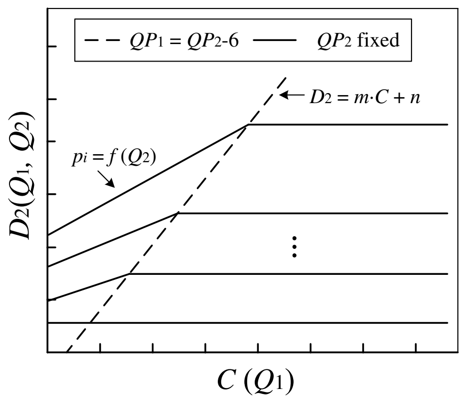


Fig. 6. The proposed distortion model for dependent layers.

Based on the above observation, we propose the following distortion model

$$D_2(Q_1,Q_2) = \begin{cases} p_i \cdot C(Q_1) + (m - p_i) \cdot C(Q_2/2) + n, & QP_1 \le QP_2 - 6, \\ m \cdot C(Q_2/2) + n, & QP_1 > QP_2 - 6. \end{cases} \quad (5)$$

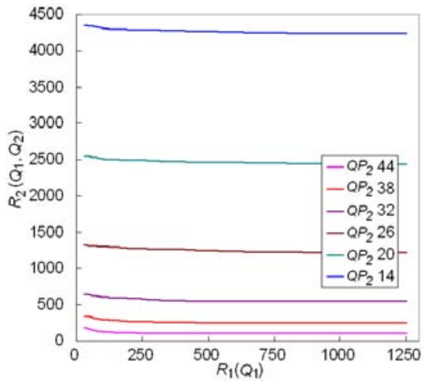
where p_i is the slope, which is a function of Q_2, and m and n are the slope and the intercept of the line with $Q_1 = \frac{1}{2}Q_2$ as indicated in Fig. 6.

B. GOP-Based Rate Model

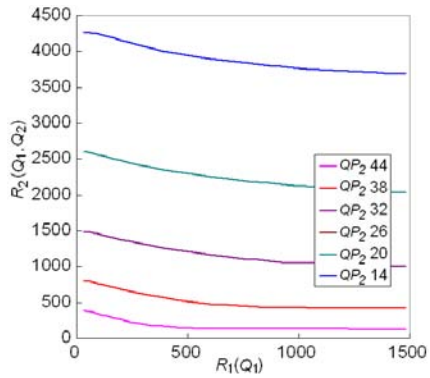
To derive the GOP-based rate model of the enhancement layer, denoted by $R_2(Q_1,Q_2)$, we plot $R_2(Q_1,Q_2)$ as a function of $R_1(Q_1)$ for some video sequences in Fig. 7. We see a set of approximately parallel lines. In other words, the inter-layer rate dependency is relatively low. Thus, we propose the following rate model

$$R_2(Q_1,Q_2) = \begin{cases} r \cdot R_1(Q_1) + (s - r)R_1(Q_2), & QP_1 \ge QP_2, \\ s \cdot R_1(Q_2), & QP_1 < QP_2. \end{cases} \quad (6)$$

where s and r are the slope of the line when $QP_1 = QP_2$ and QP_2 fixed, respectively. The proposed rate model is plotted in Fig. 8.



(a) Mobile



(b) Football

Fig. 7. Illustration of rate dependency between two consecutive layers.

IV. Proposed Spatial-Layer Bit Allocation Algorithm

For the single variable rate and distortion models, we employ the models proposed by Kamaci *et al.* [7] of the following form:

$$R(Q_i) = a \cdot Q_i^{-\alpha}, \text{ and } D(Q_i) = b \cdot Q_i^{\beta}, \quad (7)$$

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

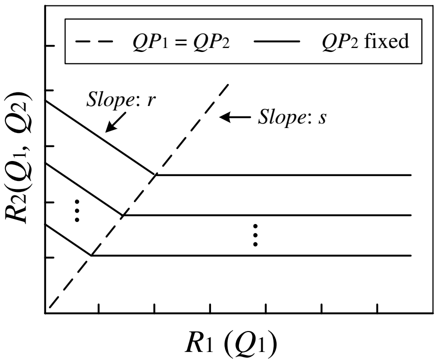


Fig. 8. The proposed rate model for dependent layers.

Then, the Lagrangian cost function in Eq. (3) can be solved with a closed-form solution by applying the proposed GOP-based R-D models as given in Eqs. (5) and (6). That is, we have

$$J(\mathbf{Q}, \lambda) = b \cdot Q_1^{\beta} + cp_i \cdot Q_1 + \frac{c}{2} \cdot (m - p_i)Q_2 + (m\theta + n) + \lambda \cdot [(1+r)a \cdot Q_1^{-\alpha} + (s-r)a \cdot Q_2^{-\alpha} - R_{total}]. \tag{8}$$

To optimize Eq. (8), we first take the partial derivatives with respect to Q_1 and Q_2, which yields the following two equations

$$\begin{aligned} &(b\beta \cdot Q_1^{\beta-1} + cp_i) - [(1+r)a\alpha \cdot Q_1^{-\alpha-1}] \cdot \lambda = 0, \\ &\tfrac{c}{2} \cdot (m - p_i) - [(s-r)a\alpha \cdot Q_2^{-\alpha-1}] \cdot \lambda = 0, \end{aligned} \tag{9}$$

Another equation can be derived from the bit budget constraint $R_{GOP} = R_{total}$; namely,

$$R_{total} = (1+r)a \cdot Q_1^{-\alpha} + (s-r)a \cdot Q_2^{-\alpha}. \tag{10}$$

Based on Eqs. (9) and (10), we can compute Q_1 and Q_2 values that optimize the Lagrangian cost function given in Eq. (8). 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.

V. Experimental Results

Since there is no spatial layer rate control algorithm in the current version of 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 encoding results of reference JSVM FixedQPEncoder tool [8] based on the SVC testing conditions JVT-Q205 defined in [9]. The logarithmic search algorithm and cascading QPs method are adopted in JSVM FixedQPEncoder tool to find a proper quantization parameter. The generalized bit rate value with each trial of quantization parameter for the next trial, until the generated bit rate is within the acceptable range ($\pm 0.5\%$ in our experiments) or the encoding iterations exceed the pre-defined maximal threshold (50 iterations for each layer in our experiments).

TABLE I
The setting of two-layer spatial scalable coding

Layer No.	Spatial Resolution	Temporal Resolution	Initial QP
0	QCIF	15	32
1	CIF	15	32

The proposed bit allocation algorithm was implemented with JSVM 9.6 in our experiment. It was tested to verify the efficiency of our bit allocation algorithm for spatial scalability, with the setting shown in Table I. Layer 0 is the base layer which is encoded without any inter-layer prediction. Layer 1 is a spatial enhancement layer by using the adaptive inter-layer prediction from base layer. To compare our scheme with JSVM9.6, the initial setting of QP is the same for both schemes. We have encoded 10 GOPs from three sequences with low to high spatial complexity (Akiyo, Football and Mobile) to verify the performance of the proposed algorithm. The GOP size is set to be 16.

The performance of the proposed algorithm is compared with FS and JSVM method shown in Fig. 9. We see little performance degradation of our method with respect to the optimal solution obtained by FS method at various bit rates. The maximum performance degradation observed is about 0.34 dB for the Mobile sequence. Significant coding gain, around 1dB, is achieved by the proposed algorithm in comparison with JSVM. For the computational complexity issue, the proposed algorithm only need to encode four times to decide the model parameters, but the FS method and current JSVM take 52×52 and average of 40 iteration times respectively, where the complexity are extremely higher. These experimental results demonstrate the effectiveness and the robustness of the proposed algorithm for video sequences with various spatial characteristics.

VI. 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.

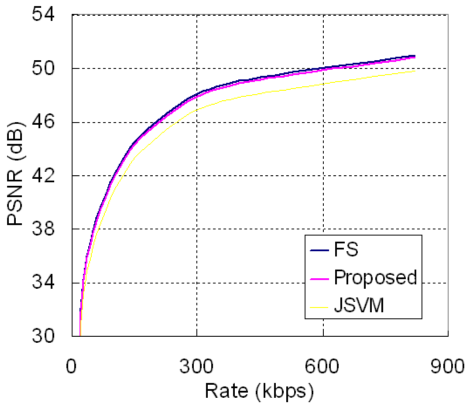
Acknowledgment

This work was supported by National Key Technology R&D Program of China under Grant 2006BAH02A10 and by National High-tech R&D Program (863 Program) of China under Grant 2006AA01Z320, with additional support by the State Scholarship Fund from China Scholarship Council.

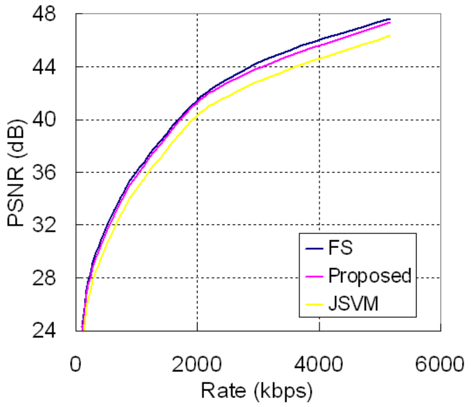
References

[1] T. Wiegand, G. J. Sullivan, J. Reichel, H. Schwarz, and M. Wien, "Amendment 3 to ITU-T Rec. H.264 (2005) ISO/IEC 14496-10: 2005," Scalable Video Coding, July 2007.

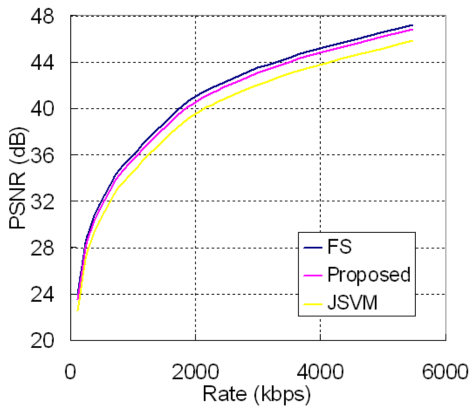
[2] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, pp. 1103–1120, September 2007.



(a) Akiyo



(b) Football



(c) Mobile

Fig. 9. The R-D performance of the proposed bit allocation algorithm compared with the FS method and JSVM.

[3] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 533–545, September 1994.

[4] L.-J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 4, pp. 446–459, August 1998.

[5] Y. Liu, Z. Li, and Y. C. Soh, "Rate control of H.264/AVC scalable extension," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 1, pp. 116–121, January 2008.

[6] D.-K. Kwon, Y. Cho, and C.-C. J. Kuo, "A simplified rate control scheme for non-conversational H.264 video," in *International Workshop on Multimedia Signal Processing*, October 2007.

[7] 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 Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 8, pp. 994–1006, August 2005.

[8] Joint Video Team of ITU-T VCEG and ISO/IEC MPEG. Joint Scalable Video Model software 9.6. [Online]. Available: ftp://garcon.ient.rwth-aachen.de

[9] T. Wiegand, G. J. Sullivan, H. Schwarz, and M. Wien, "Joint draft 10 of SVC amendment," Joint Video Team, JVT-Q205, San Jose, CA, USA, April 2007.