BIT ALLOCATION FOR JOINT SPATIAL-QUALITY SCALABILITY IN H.264/SVC

Jiaying Liu[†], *Zongming Guo*[†] *and Yongjin Cho*[††]

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

ABSTRACT

In this work, we propose a model-based layer bit allocation algorithm for joint spatial-quality (S-Q) scalability in H.264/SVC. The complicated inter-layer dependency is decoupled by the proposed spatial and quality rate and distortion (R-D) models. We show that the R-D characteristics of a dependent layer can be represented by a number of independent functions with GOP as a basic coding unit. Then, the joint bit allocation problem is formulated as a two-step optimization problem by the Lagrangian multiplier method, which can be numerically solved using the proposed R-D models. Finally, we develop a low-complexity bit allocation algorithm for the combined spatial and quality scalability in H.264/SVC. It is shown by experimental results that our proposed bit allocation algorithm achieves the coding performance significantly improved from current reference software JSVM.

Index Terms— H.264/SVC, dependent R-D models, joint spatial-quality scalability

1. INTRODUCTION

H.264/SVC is recently standardized as a scalable extension of H.264/AVC [1]. The SVC video stream is equipped with great flexibility and adaptability in terms of frame rates, display resolutions and quality levels. With three dimensions of the scalability, each coding unit in an H.264/SVC video is subject to highly complicated inter-dependency, which provides one of the major challenges for H.264/SVC bit allocation algorithms. Thus, our works focus on the bit allocation of the combined spatial and quality scalability in H.264/SVC, where the layer dependency greatly influences the overall coding efficiency.

Bit allocation algorithms for inter-frame dependency have been examined since MPEG-2. 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.

Recently, a model-based optimal spatial layer bit allocation problem is investigated by the authors' previous work [5], where dependent R-D models for the spatial scalability are proposed. Near optimal R-D performance was reported, which was mainly contributed to the successful modeling of the spatial layer R-D characteristics. In this work, we further simplify the spatial layer R-D models, and propose a novel quality layer dependent R-D models. The proposed dependent R-D models are relatively simple, but it enables an accurate and robust bit allocation scheme for the combined spatial and quality scalability of H.264/SVC. The optimal joint S-Q bit allocation problem is formulated using the Lagrangian optimization framework and solved numerically by the gradient method. The experimental results show that the new R-D models achieve a highly efficient bit allocation scheme, where significant coding gain could be observed from FixedQPEncoder tool of JSVM 9.12 reference codec [6].

The rest of this paper is organized as follows. The dependent rate and distortion characteristics of spatial and quality layer in H.264/SVC video are analyzed and modeled in Sec.2. As an application, the joint spatial-quality layer bit allocation problem is examined in Sec.3. Experimental results are given in Sec. 4. Finally, concluding remarks are given in Sec.5.

2. R-D CHARACTERISTICS OF DEPENDENT LAYER

An S-Q plane, which represents the combined spatial and quality scalability, is illustrated in Fig.1.

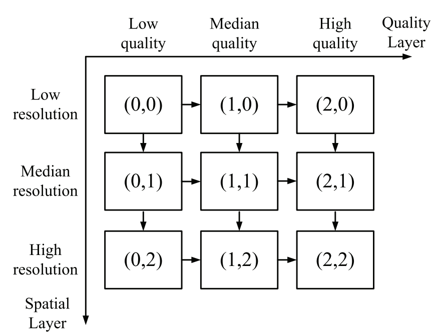


Fig. 1. Illustration of combined spatial and quality scalability, the S-Q plane of 3 spatial layer and 3 quality layers.

The interaction among coding units in the S-Q plane greatly influences the overall coding efficiency of a scalable bit stream. To represent such interaction, the R-D characteristics of a coding unit (i, j) is generally expressed by multi-variable functions of

$$R_{i,j}(q_{0,0}, \cdots, q_{i,j}) \quad \text{and} \quad D_{i,j}(q_{0,0}, \cdots, q_{i,j}),$$

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.

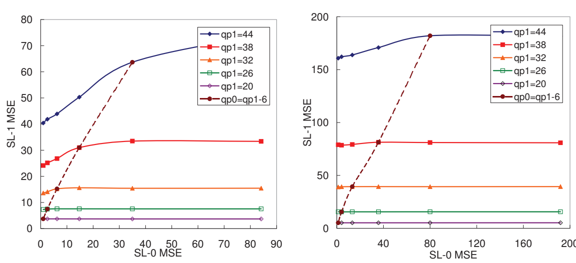
where $q_{i,j}$ is the quantization step size of a coding unit at i-th quality layer (QL-i) and j-th spatial layer (SL-j). We describe observations of the R-D characteristics, by which we could develop a linear decomposition of the complex R-D expressions.

A. Spatial layer distortion characteristic

The distortion characteristic of spatial layer (SL) is studied in [5]. In this work, we simplify the model of dependent spatial layers, which can be expressed as a linear sum of distortion function D_0 evaluated at the quantization step sizes of all participating spatial layers. Fig. 2 shows the distortion of spatial enhancement layer (*i.e.,* SL-1), denoted by $D_1(q_0, q_1)$, as a function of spatial base layer (SL-0), $D_0(q_0)$. We have the following main observations:

1. For a fixed q_1 value, we observe a linear region between $D_1(q_0, q_1)$ and $D_0(q_0)$ when $D_0(q_0)$ is small.
2. The slope in the linear region reflects the dependency between layers. The slopes remain approximately constant for lines, which are basically in parallel. Here we assume that the slopes of all branches in the linear region are the same.
3. Distortion $D_1(q_0, q_1)$ becomes flat after the inflection point which is located at $qp_0 \approx qp_1 - 6$, where qp_0 and qp_1 are quantization parameters corresponding to quantization step sizes q_0 and q_1, respectively.

When $qp_1 - qp_0 = 6$, we have that the corresponding quantization step size is halved, *i.e.,* $q_0 = q_1/2$. Intuitively, the flat distortion phenomenon can be explained as follows. When q_1 is relatively small as compared with q_0, the SL-1 distortion is primarily determined by q_1 alone.



(a) Foreman (b) Mobile

Fig. 2. Illustration of distortion dependency where the x-axis is the distortion of layer SL-0 (reference layer) and the y-axis is the distortion of layer SL-1.

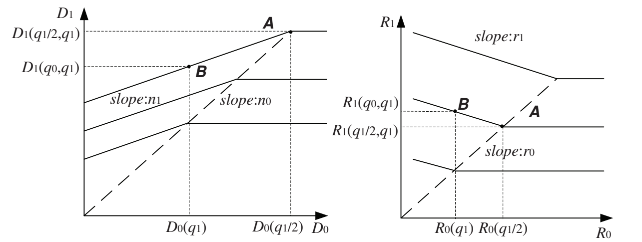
Taking account of the above analysis, SL-1 distortion model can be derived as,

$$D_1(q_0, q_1) = \begin{cases} n_1 \cdot D_0(q_0) + (n_0 - n_1) \cdot D_0(q_1/2), & q_0 \leq q_1/2, \\ n_0 \cdot D_0(q_1/2), & q_0 > q_1/2. \end{cases} \tag{1}$$

where n_0 is the slope of the dash line, which presents the distortion relation between SL-0 and SL-1 when $q_0 = q_1/2$, n_1 is the slope of the branch in the linear region, which presents the distortion relation when $q_0 < q_1/2$ as indicated in Fig.3 (a).

B. Spatial layer rate characteristic

To derive the rate model of the SL layer, denoted by $R_1(q_0, q_1)$, we plot $R_1(q_0, q_1)$ as a function of $R_0(q_0)$ for some video sequences in Fig. 4. 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,

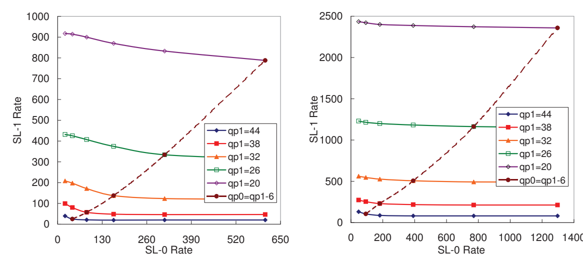


(a) Distortion Model (b) Rate Model

Fig. 3. The proposed R-D models of spatial dependent layers.

$$R_1(q_0, q_1) = \begin{cases} r_1 \cdot R_0(q_0) + (r_0 - r_1) R_0(q_1/2), & q_0 > q_1/2, \\ r_0 \cdot R_0(q_1/2), & q_0 \leq q_1/2. \end{cases} \tag{2}$$

where r_0 and r_1 are the slope of the line when $q_0 = q_1/2$ and q_1 fixed, respectively. The proposed rate model is plotted in Fig.3 (b).



(a) Foreman (b) Mobile

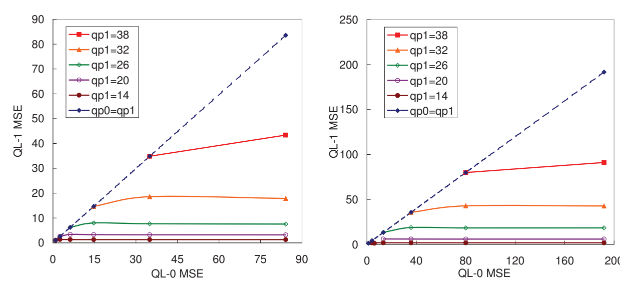
Fig. 4. Illustration of rate dependency where the x-axis is the rate of layer SL-0 (reference layer) and the y-axis is the rate of layer SL-1.

C. Quality layer distortion characteristics

The quality layer (QL) distortion characteristics are depicted in Fig. 5, where the distortion of quality enhancement layers (*i.e.,* QL-1) is plotted as a function of the quality base layer (QL-0). A linear relation among the quality base and enhancement layers could be observe and as a result, we can express a QL distortion function by

$$D_1(q_0, q_1) = s_0 \cdot D_0(q_1), \tag{3}$$

where s_0 is the model parameter, which is the slope of the distortion relation when $qp_1 = qp_0$.



(a) Foreman (b) Mobile

Fig. 5. Illustration of distortion dependency where the x-axis is the distortion of layer QL-0 (reference layer) and the y-axis is the distortion of layer QL-1.

This simple distortion relation can also be explained by the signal decomposition principle. Since the quantization step size of the enhancement layer has to be finer than that of the base layer, the enhancement layer distortion with respect to the original frame becomes relatively independent of the base layer quantization.

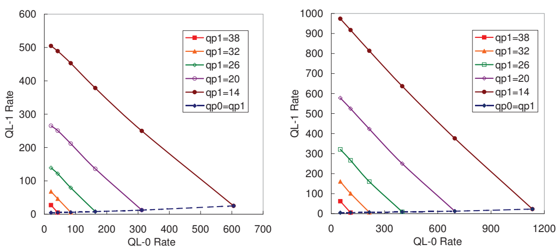
D. Quality layer rate characteristics

The rate characteristics of a quality layer is more involved than the distortion characteristics and thus it demands a careful investigation. We show the rate of QL-1 as a function of QL-0 rate in Fig. 6, where the q_i's are controlled to generate the rate points.

We have the following two main observations about the rate characteristics of quality layers.

1. The rate of a quality enhancement layer is approximately linearly proportional to the rate of the quality base layer.
2. The lines of the same q_1-setting have a common slope. That is, each type of curves are approximately parallel to each other.

As shown in the abstract model Fig.7, the solid lines present the relation that the QL-1 rate decreases as the QL-0 rate increases. This is because higher bit rates of the base layer result in the reduction of the information to be coded in the enhancement layer and, as a result, the enhancement layer rate is reduced when qp_1 is fixed. Based on



(a) Foreman (b) Bus

Fig. 6. Illustration of rate dependency where the x-axis is the rate of layer QL-0 (reference layer) and the y-axis is the rate of layer QL-1.

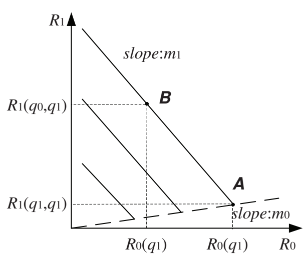


Fig. 7. The proposed rate model for quality dependent layers.

the two properties, we developed a quality layer rate model as the linear sum of the quality base layer (QL-0) rate functions evaluated at parameters q_0, q_1. Mathematically,

$$R_1(q_0, q_1) = m_1 \cdot R_0(q_0) + (m_0 - m_1) \cdot R_0(q_1), \tag{4}$$

where m_0 and m_1 are the slope of the line when $q_0 = q_1$ and q_1 fixed, respectively.

3. JOINT S-Q BIT ALLOCATION

A joint S-Q bit allocation problem is examined as an application of the proposed R-D models. We formulate the joint bit allocation problem into two-step optimization problem: first, when a total 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; second, within each spatial layer, we need to minimize the sum of quality layers' distortion given the certain ratio of the bit budget corresponding to each resolution calculated in the first step. Mathematically,

$$\textbf{Step I}: \mathbf{q_j^*} = \arg\min_{\mathbf{q_j} \in \mathcal{Q}^{N_S}} \sum_{j=o}^{N_S-1} D_j(q_0, \cdots, q_j),$$

$$s.t. \sum_{j=0}^{N_S-1} R_j(q_0, \ldots, q_j) \le R_{total};$$

$$\textbf{Step II}: \mathbf{Q}^* = \arg\min_{\mathbf{Q} \in \mathcal{Q}^{N_Q} \times \mathcal{Q}^{N_S}} \sum_{i=o}^{N_Q-1} D_{i,j}(q_{0,j}, \cdots, q_{i,j}),$$

$$s.t.\ R_{0,j}(q_{0,j}) \le \omega_0 \cdot R_j(\mathbf{q_j}), \ldots, R_{i,j}(q_{i,j}) \le \omega_i \cdot R_j(\mathbf{q_j}), \ldots,$$
$$R_{N_Q-1,j}(q_{N_Q-1,j}) \le \omega_{N_Q-1} \cdot R_j(\mathbf{q_j}), j = 0, \ldots, N_S - 1. \tag{5}$$

where $R_j(\mathbf{q_j})$ is the rate of SL-j, $\mathbf{Q}$ and $\mathbf{q_j}$ are the $N_Q \times N_S$ matrix and the $N_S \times 1$ vector whose elements are the quantization step sizes of the participating coding units, $\mathcal{Q}$ is the space of all admissible quantization step sizes. ω_i is the weighting factor representing the corresponding ratio of the i-th QL.

The Lagrangian multiplier method converts the constrained optimization problem in Eq. (5) to an equivalent unconstrained optimization problem by introducing the Lagrangian cost function as,

$$\textbf{Step I}: J(\mathbf{q_j^*}, \lambda^*) = \arg\min_{\mathbf{q_j} \in \mathcal{Q}^{N_S}} J(\mathbf{q_j}, \lambda)$$

$$= \sum_{j=o}^{N_S-1} D_j(q_0, \cdots, q_j) + \lambda \cdot \left(\sum_{j=0}^{N_S-1} R_j(q_0, \ldots, q_j) - R_{total} \right);$$

$$\textbf{Step II}: J(\mathbf{Q}^*, \Lambda^*) = \arg\min_{\mathbf{Q} \in \mathcal{Q}^{N_Q} \times \mathcal{Q}^{N_S}} J(\mathbf{Q}, \Lambda), j = 0, \ldots, N_S - 1,$$

$$= \sum_{i=o}^{N_Q-1} D_{i,j}(q_{0,j}, \cdots, q_{i,j}) + \lambda_0 \cdot (R_{0,j}(q_{0,j}) - \omega_0 \cdot R_j(\mathbf{q_j})) + \ldots$$
$$+ \lambda_{N_Q-1} \left(R_{N_Q-1,j}(q_{N_Q-1,j}) - \omega_{N_Q-1} \cdot R_j(\mathbf{q_j}) \right). \tag{6}$$

where λ_i is the Lagrangian multiplier.

Without loss of generality, we consider bit allocation in a simple scenario with two spatial and two quality layers case first. The solution can be easily generalized to a multi-layer scenario. With the proposed R-D models, the objective function in Eq. (6) can be rewritten as

$$\textbf{Step I}: J(\mathbf{q_j}, \lambda) = (1 + n_1) D_0(q_0) + (n_0 - n_1) D_0(q_1/2)$$
$$+ \lambda \cdot (R_0(q_0) + r_1 \cdot R_0(q_1/2) - R_{total});$$

$$\textbf{Step II}: J(\mathbf{Q}, \Lambda) = D_{0,j}(q_{0,j}) + s_{0,j} \cdot D_{0,j}(q_{1,j})$$
$$+ \lambda_0 \cdot (R_{0,j}(q_{0,j}) - \omega_0 \cdot R_j(\mathbf{q_j})) + \ldots + \lambda_1 \cdot ((1 + m_{1,j}) R_{0,j}(q_{0,j})$$
$$+ (m_{0,j} - m_{1,j}) R_{0,j}(q_{1,j}) - R_j(\mathbf{q_j})), j = 0, \ldots, N_S - 1. \tag{7}$$

For the final closed form expression, the frame-based R-D models in [7], where the rate and distortion functions are given by

$$R(q) = a \cdot q^{-\alpha} \text{ and } D(q) = b \cdot q^{\beta}, \tag{8}$$

where a, b, α and β are model parameters. Finally, the optimization problem becomes:

$$\textbf{Step I}: J(\mathbf{q_j}, \lambda) = (1 + n_1) a_0 \cdot q_0^{-\alpha_0} + (n_0 - n_1) a_0 \cdot (q_1/2)^{-\alpha_0}$$
$$+ \lambda \cdot \left(b_0 \cdot q_0^{\beta_0} + r_1 \cdot b_0 \cdot (q_1/2)^{\beta_0} - R_{total} \right);$$

$$\textbf{Step II}: J(\mathbf{Q}, \Lambda) = a_{0,j} \cdot q_{0,j}^{-\alpha_{0,j}} + s_{0,j} \cdot a_{0,j} \cdot q_{1,j}^{-\alpha_{0,j}}$$
$$+ \lambda_0 \cdot \left(b_{0,j} \cdot q_{0,j}^{\beta_{0,j}} - \omega_0 \cdot R_j(\mathbf{q_j}) \right) + \ldots + \lambda_1 \cdot \left((1 + m_{1,j}) b_{0,j} \cdot q_{0,j}^{\beta_{0,j}} \right.$$
$$\left. + (m_{0,j} - m_{1,j}) \cdot b_{0,j} \cdot q_{1,j}^{\beta_{0,j}} - R_j(\mathbf{q_j}) \right), j = 0, \ldots, N_S - 1. \tag{9}$$

Table 1. Performance of two methods for joint S-Q layers in terms of PSNR,output rate, Δ rate and Δ PSNR.

Sequence	Target rate(kbps)	Method	PSNR (dB)	Rate (kbps)	Δ Rate	ΔPSNR
Bus	512	Proposed	30.47	510.46	-1.54	+0.82
		JSVM	29.65	512.60	+0.60	
	768	Proposed	32.46	770.80	+2.80	+0.42
		JSVM	31.84	756.66	-11.34	
	1024	Proposed	34.03	1023.84	-0.16	+0.61
		JSVM	33.42	1024.58	+0.58	
Football	768	Proposed	33.67	755.40	-12.60	+1.00
		JSVM	31.67	755.86	-12.14	
	1024	Proposed	35.22	1018.73	-5.27	+1.98
		JSVM	33.24	1025.70	+2.89	
	1280	Proposed	36.64	1282.02	+2.02	+1.97
		JSVM	34.67	1280.92	+0.92	

The Lagrange equation in Eq. (9) is solved by computing partial derivatives with respect to $q_{i,j}$'s and gradient method is used to solve the system of non-linear equations from the partial differentiation.

4. EXPERIMENTAL RESULTS

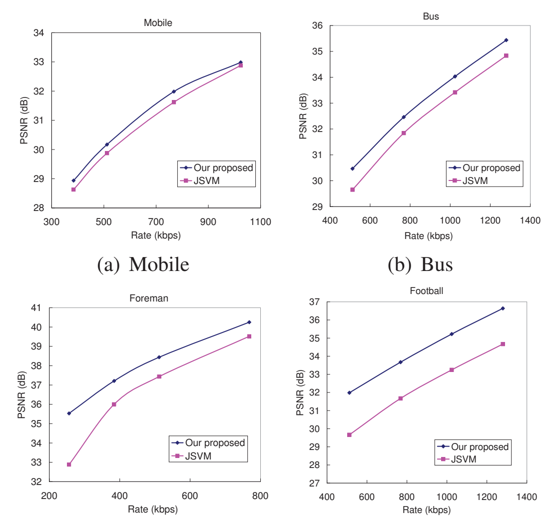
Since there is no spatial/quality layer rate control algorithm in the current version of JSVM, we consider comparing with encoding results of reference JSVM FixedQPEncoder tool based on the SVC testing conditions JVT-Q205 defined in [8]. The logarithmic search algorithm and cascading QPs method are adopted in JSVM FixedQPEncoder tool to find a proper quantization parameter.

The proposed bit allocation algorithm was implemented with JSVM 9.12 [6] in our experiments. It was tested to verify the efficiency of our bit allocation algorithm for joint spatial-quality scalability, with the setting: SL-0 is the base layer encoded without any inter-layer prediction. SL-1 is a spatial enhancement layer by using the adaptive inter-layer prediction from base layer. Within each spatial layer, there are two quality layers, namely QL-0 and QL-1. To compare our scheme with JSVM, the initial setting of QP is the same for both schemes. We have encoded 20 GOPs from four sequences with low to high spatial complexity (Foreman, Football, Bus and Mobile sequences) to verify the performance of the proposed algorithm. The GOP size is set to be 16. In this work, we assume the quality ratio ω_0 is $2/3$.

The PSNR performance of comparison results is shown in Fig. 8(a)-(d) respectively. We see that significant coding gain is achieved by the proposed algorithm in comparison with JSVM9.12. We summarize the coding results using the proposed bit allocation algorithm and the JSVM reference codes in Tables 1. The rate control method using the proposed bit allocation algorithm outperforms the current JSVM by 1.04dB in the PSNR performance on average. Both methods can yield the desired rate with a small deviation (less than 1.5% of the target rate). These experimental results demonstrate the effectiveness and the robustness of the proposed algorithm for video sequences with various spatial characteristics.

5. CONCLUSION

A dependent quality layer R-D models was proposed for the efficient bit allocation of H.264/SVC video. As an application of the proposed R-D models, we examined the joint S-Q bit allocation problem. The major contribution of our work is the decomposition of a multi-variable dependent R-D functions into linear sums of independent single-variable R-D functions. Consequently, the dependent bit allocation problem could be solved at significantly reduced complexity.



(a) Mobile (b) Bus

(c) Foreman (d) Football

Fig. 8. The R-D performance of the proposed bit allocation algorithm compared with JSVM9.12.

6. REFERENCES

[1] T. Wiegand, G. 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] K. Ramchandran, A. Ortega, and M. 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] L. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. on CSVT*, vol. 8, no. 4, pp. 446–459, Aug. 1998.

[4] Y. Liu, Z. Li, and Y. Soh, "Rate control of H.264/AVC scalable extension," *IEEE Trans. on CSVT*, vol. 18, no. 1, pp. 116–121, Jan. 2008.

[5] J. Liu, Y. Cho, Z. Guo, and C.-C. Jay Kuo, "Bit allocation for spatial scalability in H.264/SVC," in *Proc. of International Workshop on MMSP*, 2008.

[6] Joint Video Team of ITU-T VCEG and ISO/IEC MPEG, "Joint Scalable Video Model software 9.12," ftp://garcon.ient.rwth-aachen.de.

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

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