H.264/SVC TEMPORAL BIT ALLOCATION WITH DEPENDENT DISTORTION MODEL

Yongjin Cho[†], *Jiaying Liu*[*], *Do-Kyoung Kwon*[††] *and C.-C. Jay Kuo*[†]

[†]Ming Hsieh Department of EE, University of Southern California, Los Angeles, CA 90089
[*]Institute of Computer Science and Technology, Peking University, Beijing, China 100871
[††]Texas Instruments, 12500 TI Blvd., TX 72543

ABSTRACT

The bit allocation problem for hierarchical B-pictures in H.264/SVC is studied with a GOP-based dependent distortion model in this work. Inter-dependency between temporal layers of H.264/SVC is often neglected because of the complexity involved, which often leads to poorer rate control performance. To address this shortcoming, we propose a distortion model that takes inter-dependency into consideration while preserving the low complexity of the encoding process. It is demonstrated by experimental results that the new distortion model results in a highly efficient bit allocation scheme, which outperforms the rate control algorithm in the JSVM 9.12 reference codec by a significant margin.

***Index Terms*—** Dependent R-D modeling, optimal bit allocation, H.264/SVC

1. INTRODUCTION

H.264/SVC is a scalable extension of H.264/AVC which was standardized recently. As with other video compression standards, a good rate control algorithm is critical for an H.264/SVC video encoder to achieve a better rate-distortion (R-D) tradeoff, which is the main objective of this research. Rate control for H.264/SVC is challenging due to its unique GOP structure and spatial-temporal-quality scalability. H.264/SVC video has highly complicated dependency relation among scalability layers, which makes rate control of H.264/SVC more difficult than that of all previous video coding standards. We focus on H.264/SVC with temporal scalability in this research. A group of pictures (GOP) of H.264/SVC video consists of a key frame (of I- or P-pictures) and hierarchically aligned B-pictures, which are adopted to achieve temporal scalability with good coding performance. Under this setting, H.264/SVC rate control can be formulated as a bit allocation problem among temporal layers, where temporal dependency greatly influences the coding efficiency. This temporal layer bit allocation problem has no precedence in any rate control problem encountered in the past and deserves in-depth investigation.

Rate control algorithms for H.264/SVC have been studied before, *e.g.*, [1, 2]. However, the dependency between temporal layers was not addressed explicitly. Thus, the improvement in coding efficiency is somehow limited. A trellis-based optimal solution to the dependent bit allocation problem was proposed by Ramchandran *et al.* [3] to analyze the dependency of I-, P- and B-frames in MPEG-2. Lin *et al.* [4] addressed the same problem using a piecewise linear approximation of the R-D characteristics of dependent frames. Both approaches result in successful rate control algorithms with good coding efficiency for MPEG-2. However, the complexity of these two rate control algorithms is too high, which grows exponentially as the number of layers increases. Besides, the dependency of I-, P- and B-frames in MPEG-2 is different from that of temporal layers in H.264/SVC. The major contribution of this work is the proposal of a novel GOP-based dependent distortion model targeting at H.264/SVC temporal scalability. The dependent distortion model is relatively simple yet accurate enough to provide good R-D performance tradeoff. Based on this distortion model and a straightforward rate model, the optimal bit allocation problem can be formulated using the Lagrange approach and solved numerically. It is shown by experimental results that the new distortion model results in a highly efficient bit allocation scheme, which outperforms the rate control algorithm in the JSVM 9.12 reference codec by a significant margin.

The rest of this paper is organized as follows. The dependent rate and distortion characteristics of H.264/SVC bit streams are reported in Sec. 2. Then, we propose a model to describe the GOP-based dependent distortion characteristics in Sec. 3, which is called the dependent distortion model. As an application, the temporal layer bit allocation problem for hierarchical B-pictures of H.264/SVC is examined in Sec. 4. Finally, concluding remarks and future research directions are given in Sec. 6.

2. RATE AND DISTORTION CHARACTERISTICS OF H.264/SVC VIDEO

Hierarchical Temporal Layers of H.264/SVC

The hierarchical prediction structure of H.264/SVC for temporal scalability provision is illustrated in Fig. 1. The rate and distortion characteristics of B-frames heavily depend on the fidelity of reference frames. Thus, it is important to understand the dependent rate and distortion characteristics. Generally speaking, we can express the rate and distortion characteristics of a dependent layer as multi-variable functions of $R_i(q_0, \cdots, q_i)$ and $D_i(q_0, \cdots, q_i)$, where q_i is the quantization step size of the i^{th} coding unit. We describe observations of the rate and distortion characteristics in this section.

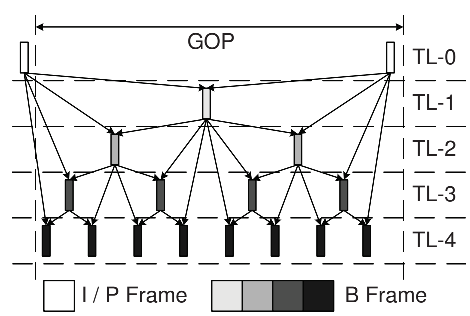
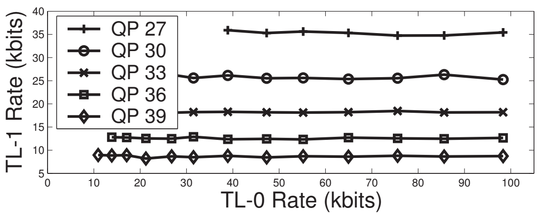


Fig. 1. Illustration of a GOP of H.264/SVC, which consists of five temporal layers (denoted by TL-0, $\cdots$, TL-4) formed by hierarchical B-pictures.

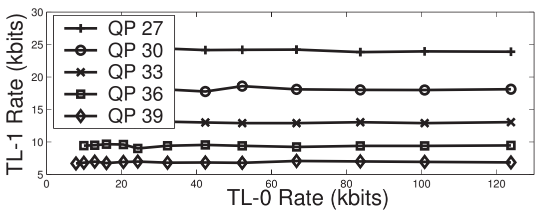
Rate characteristics

We show the bit rate of a dependent layer (*i.e.,* layer TL-1) as a function of the bit rate of its reference layer (*i.e.,* layer TL-0) in Fig. 2. It is clear that the bit rate of layer TL-1 is mainly determined by its own QP and it is independent of the bit rate of its reference layer, TL-0. This observation holds generally. As a result, we can express the rate of a dependent layer as a function of its own quantization step size as

$$R_i(q_0, q_1, \cdots, q_i) \approx R_i(q_i). \tag{1}$$



(a) Football, QCIF



(b) Foreman, CIF

Fig. 2. Illustration of rate dependency where the x-axis is the bit rate of layer TL-0 (reference layer) and the y-axis is the bit rate of layer TL-1 (dependent layer).

Distortion characteristics

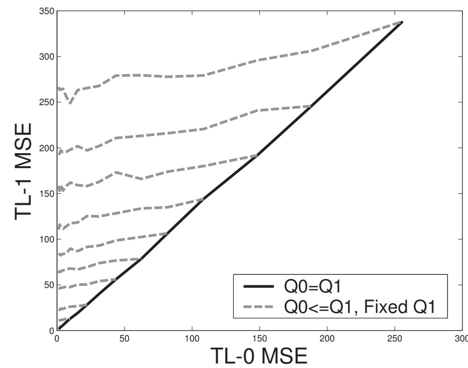
The characterization of the distortion behavior of temporal layers is more involved than that of the rate. The distortion of a dependent layer is plotted with respect to the distortion of their reference layers in Fig. 3, where distortion pairs $(D_0(q), D_1(q_0, q_1))$ and $(D_0(q), D_2(q_0, q_1, q_2))$ are shown in Figs. 3 (a) and (b), respectively. There are two types of curves in Fig. 3(a), reflecting two different settings of a two-variable TL-1 distortion function denoted by $D_1(q_0, q_1)$. The solid curve on the diagonal plots the TL-1 distortion when q_0 and q_1 have the same values. For dashed branches, the value of q_0 varies whereas the value of q_1 is fixed at each branch. Similarly, there are three types of curves in Fig. 3(b), represented by solid, dashed and dotted curves. To understand the distortion of a layer TL-2 denoted by $D_2(q_0, q_1, q_2)$, we consider the following three settings among quantization parameters q_0, q_1 and q_2:

1. $q_0 = q_1 = q_2$;
2. q_2 is fixed and q_0 and q_1 vary with $q_0 = q_1 \leq q_2$;
3. q_1 and q_2 are fixed and q_0 varies with $q_0 \leq q_1 \leq q_2$.

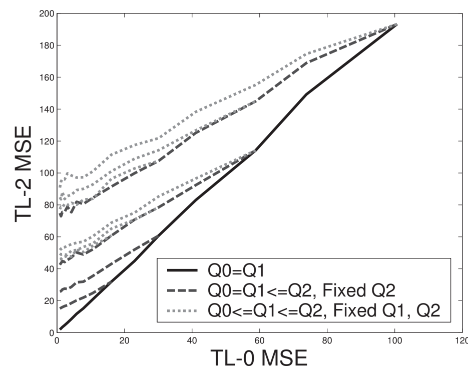
The solid diagonal and dashed branches from the diagonal in Fig. 3(b) show the first and the second settings, respectively. The dotted branches from dashed branches correspond to the third setting.

Careful observations with these controlled settings suggest the following two properties about distortion characteristics of dependent temporal layers.

1. The distortion of a dependent temporal layer is linearly proportional to the distortion of its reference layers.



(a) Football, QCIF, TL-1



(b) Foreman, CIF, TL-2

Fig. 3. Illustration of distortion characteristics of temporal layers (a) TL-1 and (b) TL-2.

2. The slopes remains approximately constant for curves under the same setting of q's. For example, distortion branches of layer TL-2 in the second and the third settings are basically in parallel. In other words, branches of the same setting have the same slope.

We will develop a dependent distortion model for temporal layers based on the above two properties in the next section.

3. PROPOSED GOP DISTORTION MODEL

The proposed GOP distortion model is constructed using a two-step modeling procedure.

1. The distortion of a dependent layer is modeled as the linear sum of the distortion function of layer TL-0 evaluated at parameters $q_0, q_1, \cdots, q_i$. Mathematically, this is written as

$$D_i(q_0, \cdots, q_i) = \zeta_{i,0} \cdot D_0(q_0) + \cdots + \zeta_{i,i} \cdot D_0(q_i), \tag{2}$$

where $\zeta_{i,j}$'s are model parameters that reflect the contribution of layer j to the distortion of the current layer, TL-i, where $j \leq i$.

2. The distortion of all layers are simply added to construct the GOP distortion model. That is,

$$\begin{aligned} &D_{GOP}(q_0, \cdots, q_{N_T-1}) \\ &\quad = D_0(q_0) + \cdots + D_{N_T-1}(q_0, \cdots, q_{N_T-1}) \\ &\quad = D_0(q_0) + \cdots + \sum_{j=0}^{N_T-1} \zeta_{N_T-1,j} \cdot D_0(q_j) \qquad (3) \\ &\quad = \omega_0 \cdot D_0(q_0) + \cdots + \omega_{N_T-1} \cdot D_0(q_{N_T} - 1), \end{aligned}$$

where ω_i is the model parameter and N_T is the number of temporal layers.

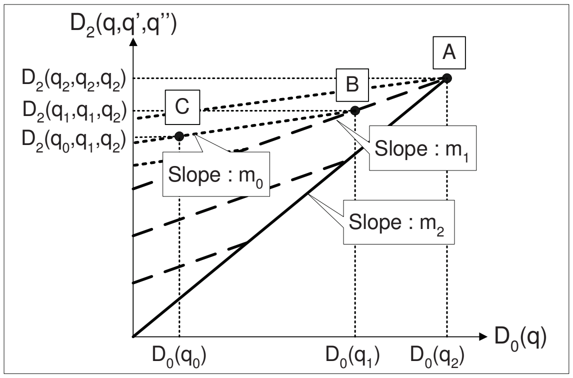


Fig. 4. Illustration of the TL-2 distortion modeling procedure

Eq. (2) can be derived mathematically as follows. Consider three representative points A, B and C on these three types of curves as shown in Fig. 4, where A, B and C correspond to cases 1, 2 and 3 of the q setting, respectively. Without loss of generality, the distortion of the temporal layer, TL-2, at an arbitrary point can be determined by analyzing the distortion at points A, B and C as depicted in the figure. That is, we have

1. A: $D_2(q_2, q_2, q_2) = m_2 \cdot D_0(q_2)$
2. B: $D_2(q_1, q_1, q_2) = m_1 \cdot D_0(q_1) + (m_2 - m_1) \cdot D_0(q_2)$
3. C: $D_2(q_0, q_1, q_2) = m_0 \cdot D_0(q_0) + (m_1 - m_0) \cdot D_0(q_1) + (m_2 - m_1) \cdot D_0(q_2)$.

Finally, by substituting coefficients with $\zeta_{2,i}$ we can obtain the distortion model of temporal layer TL-2 as

$$D_2(q_0, q_1, q_2) = \zeta_{2,0} \cdot D_0(q_0) + \zeta_{2,1} \cdot D_0(q_1) + \zeta_{2,2} \cdot D_0(q_2),$$

where

$$\zeta_{2,0} = m_0, \quad \zeta_{2,1} = m_1 - m_0 \quad \text{and} \quad \zeta_{2,2} = m_2 - m_1$$

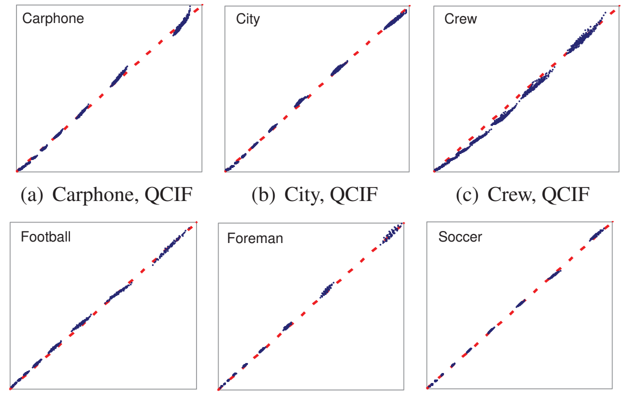
are model parameters. The distortion model of temporal layer TL-2 can be generalized to that of temporal layer TL-i given i slopes of $m_0, \cdots, m_{i-1}$. That is, parameters of the distortion model of temporal layer TL-i as given in Eq. 2 can be expressed as

$$\zeta_{i,0} = m_0 \text{ and } \zeta_{i,j} = m_j - m_{j-1}, \quad \forall j \in 1, \cdots, i-1$$

Finally, the model parameters of the GOP distortion model (ω_i) are determined simply by adding $\zeta_{i,j}$'s; namely,

$$\omega_i = \sum_{j=i}^{N_T-1} \zeta_{i,j}.$$

We generated 2,730 and 810 GOP distortion samples in QCIF and CIF, respectively, to examine the validity of the proposed GOP distortion model. The number of frames in a GOP is set to 16, which corresponds to 5 temporal layers. Accuracy and robustness of the proposed distortion model are verified in Fig. 5, which presents the modeling result with respect to the real GOP distortion values on the horizontal axis.



(a) Carphone, QCIF (b) City, QCIF (c) Crew, QCIF

(d) Football, CIF (e) Foreman, CIF (f) Soccer, CIF

Fig. 5. Illustration of the 16 frame (5-TL) GOP distortion modeling result with respect to red dashed $y = x$ line.

4. TEMPORAL LAYER BIT ALLOCATION

The temporal layer bit allocation problem is examined as an application of the proposed GOP distortion model. We formulate the problem as the GOP distortion minimization problem with a temporal layer as the basic coding unit. Mathematically, we have

$$\mathbf{q}^* = \arg\min_{\mathbf{q} \in \mathcal{Q}^{N_T}} D_{GOP}(\mathbf{q}) \text{ s.t. } R_{GOP}(\mathbf{q}) \leq R_T, \tag{4}$$

where $\mathbf{q} = [q_0, q_1, \cdots, q_{N_T-1}]$ represents the selected q-vector for each GOP, $\mathcal{Q}$ is the q-space of all admissible quantization step sizes and R_T is the target bit budget constraint. The Lagrangian formulation leads the constrained problem in Eq. 4 to the unconstrained dual of

$$\begin{aligned} J(\mathbf{q}^*, \lambda^*) &= \arg\min_{\mathbf{q} \in \mathcal{Q}^{N_T}} J(\mathbf{q}, \lambda) \\ &= D_{GOP}(\mathbf{q}) + \lambda \cdot (R_{GOP}(\mathbf{q}) - R_T) \\ &= \sum_{i=0}^{N_T-1} \omega_i \cdot D_0(q_i) + \lambda \cdot \left(\sum_{i=0}^{N_T-1} R_i(q_i) - R_T \right). \end{aligned} \tag{5}$$

Note that functions of multiple variables in Eq. (5) are successfully re-written as functions of a single variable by Eq.'s (1) and (3). Moreover, the GOP distortion could be expressed as the linear sum of independent distortion functions ($D_0(q)$'s) by the proposed distortion model.

The R-D functions in Eq. (5) are substituted by frame-based R-D models in [5], where the rate and distortion functions are given by

$$R(q) = a \cdot q^{-\alpha} \text{ and } D(q) = b \cdot q^{\beta}, \tag{6}$$

where a, b, α and β are model parameters. With the R-D models in Eq. (6), the final problem formulation becomes:

$$\begin{aligned} &J(\mathbf{q}^*, \lambda^*) = \arg\min_{\mathbf{q} \in \mathcal{Q}^{N_T}} J(\mathbf{q}, \lambda) = \\ &\arg\min_{\mathbf{q} \in \mathcal{Q}^{N_T}} \left(\sum_{i=0}^{N_T-1} \omega_i \cdot b_0 \cdot q_i^{\beta_0} + \lambda \cdot \sum_{i=0}^{N_T-1} a_i \cdot q_i^{-\alpha_i} \right). \end{aligned} \tag{7}$$

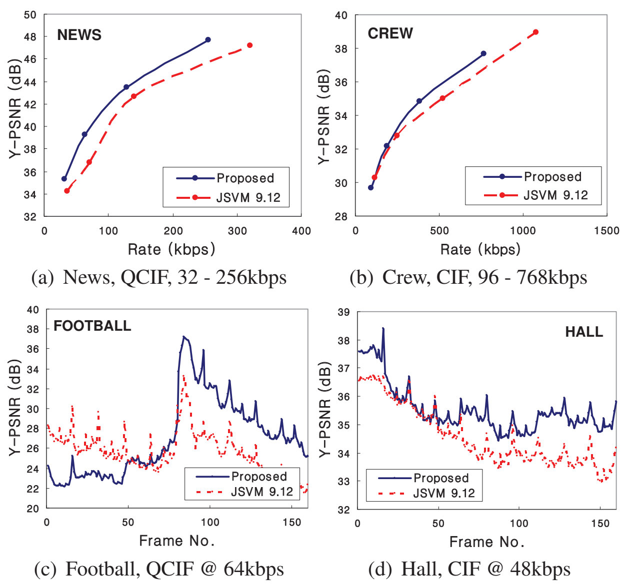
Eq. (7) can be solved by computing partial derivatives with respect to q_i's.

Table 1. TL Performance; Rates vs. Y-PSNR, 10 GOP Average

Format	Sequence	R_T	TL	JSVM 9.12		Proposed		ΔR	ΔPSNR	Average ΔPSNR
				Rate (kbps)	PSNR (dB)	Rate	PSNR			
QCIF	Football	64	0	11.84	27.64	11.36	28.02	-0.48	+0.36	+0.94
			1	19.82	27.00	19.03	27.79	-0.79	+0.79	
			2	31.52	26.38	31.04	27.42	-0.48	+1.02	
			3	46.25	25.91	45.84	27.12	-0.41	+1.21	
			4	62.74	25.67	63.51	27.00	+0.77	+1.33	
CIF	News	96	0	37.45	36.88	33.20	38.06	-4.25	+1.38	+1.30
			1	50.07	36.51	47.64	37.86	-2.43	+1.35	
			2	66.34	36.26	64.75	37.58	-1.59	+1.32	
			3	84.70	36.07	82.20	37.33	-2.50	+1.26	
			4	101.25	35.98	94.70	37.15	-6.55	+1.17	

5. EXPERIMENTAL RESULTS

Temporal layer bit allocation is studied with QCIF and CIF sequences using the Lagrange method as shown in Eq. (7). The target bit rates are set to 32 kbps to 768 kbps based on sequence characteristics and formats. Each GOP consists of 16 frames and every key frame in TL-0 is coded as a P-frame except for the 0^{th} I-frame. The performance of the proposed temporal layer bit allocation scheme is verified by the comparison with the rate control algorithm implemented in the JSVM reference software codec.



(a) News, QCIF, 32 - 256kbps (b) Crew, CIF, 96 - 768kbps

(c) Football, QCIF @ 64kbps (d) Hall, CIF @ 48kbps

Fig. 6. Illustration of the coding gain by the proposed temporal layer bit allocation scheme.

Fig. 6 illustrates the bit allocation performance, where Figs. 6(a) and 6(b) show the 10 GOP average over a range of target bit rates while the frame-by-frame Y-PSNR values are plotted in Figs. 6(c) and 6(d). Besides the significant coding gain observed in Fig. 6, test results for sequences of various temporal characteristics suggest the efficiency and robustness of the proposed bit allocation scheme due to successful dependent R-D modeling.

In Table 1, temporal layer coding performance is summarized, where higher performance could be observed at all temporal layers than the JSVM benchmark. At each temporal layer, the coding gain ranges from 0.36 to 1.38 dB with an average gain of 0.94 and 1.30 dB for Football and News sequences, respectively. Inter-layer dependency is successfully captured by the proposed dependent distortion model and, thus, the target bit budget is efficiently allocated to participating temporal layers.

6. CONCLUSION AND FUTURE WORK

A GOP-based dependent distortion model was proposed and applied to the temporal layer bit allocation problem for H.264/SVC in this work. The efficiency and the robustness of the proposed bit allocation scheme could be verified by the experimental result. The major contribution of our work is the decomposition of a multi-variable dependent distortion function into a linear sum of independent single-variable distortion functions. Moreover, the complexity required to solve the dependent bit allocation problem can be reduced significantly. We focused on temporal scalability of H.264/SVC in this work, and would like to investigate rate control for H.264/SVC quality and spatial scalability in the future.

7. REFERENCES

[1] Long Xu, Siwei Ma, Debin Zhao, and Wen Gao, "Rate control for scalable video model," in *Proc. of SPIE on Visual Communications and Image Processing*, July 2005, pp. 525–534.

[2] Yang Liu, Zheng Guo Li, and Yeng Chai Soh, "Rate control of H.264/AVC scalable extension," *IEEE Trans. on Circ. and Sys. for Video Tech.*, vol. 18, no. 1, pp. 116–121, January 2008.

[3] 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, September 1994.

[4] Liang-Jin Lin and Antonio Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. on Circ. and Sys. for Video Tech.*, vol. 8, no. 4, pp. 446–459, August 1998.

[5] 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 Circ. and Sys. for Video Tech.*, vol. 15, no. 8, pp. 994–1006, August 2005.