SINGLE PASS DEPENDENT BIT ALLOCATION FOR H.264 TEMPORAL SCALABILITY

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

[††]Institute of Computer Science and Technology, Peking University, Beijing, China 100871
[†]Ming Hsieh Department of EE, University of Southern California, Los Angeles, CA 90089

ABSTRACT

In this paper, we propose a single-pass dependent bit allocation algorithm for H.264/SVC hierarchical B-pictures. To develop a practical bit allocation algorithm, we use the number of skipped blocks and the ratio of the mean absolute difference (MAD) as features to measure the inter-layer signal dependence of input video signals. The proposed algorithm performs bit allocation at the target bit rate with two steps: the group-of-picture (GOP) based rate control and adaptive temporal layer quantization parameter (QP) decision. The superior performance of the proposed algorithm is demonstrated by experimental results, which is compared with two other one-pass bit allocation algorithms in the literature.

***Index Terms*—** Dependent R-D, bit allocation, SVC, single pass

1. INTRODUCTION

By hierarchical B-pictures, we refer to a Group-of-Pictures (GOP) structure that is composed of hierarchically aligned B-pictures. It employs generalized B-pictures that can be used as a reference to following inter-coded frames. Although it introduces a structural encoding delay of one GOP size, it usually provides much higher coding efficiency than the conventional GOP structures [1]. Moreover, due to its natural capability of providing the temporal scalability, it is employed as a GOP structure of H.264/SVC, the scalable extension of H.264/AVC [2].

Due to the complex inter-layer dependence of hierarchical B-pictures, the development of an efficient and effective bit allocation algorithm for H.264/SVC is a challenging task. There are several bit allocation algorithms that considered the inter-layer dependence in the literature before. Schwarz *et al.* [1] proposed the *QP cascading* scheme that applies a fixed quantization parameter (QP) difference between adjacent temporal layers. Liu *et al.* [3] introduced constant weights to temporal layers in their H.264/SVC rate control algorithm. Although these algorithms achieve good coding efficiency, they are limited in two aspects. First, the inter-layer dependence is heuristically addressed. Second, the input video characteristics are not taken into account. For these reasons, the optimality of these bit allocation algorithms cannot be guaranteed.

A framework of an optimal dependent bit allocation for MPEG video was investigated by Ramachandran *et al.* [4]. However, due to the high computational requirement, it cannot be used for practical applications. More recently, a model based dependent bit allocation for hierarchical B-pictures was presented in [5], where a dependent distortion model was proposed to simplify the complexity of dependent bit allocation. Although the algorithm given in [5] achieves huge complexity reduction as compared with that in [4], it still demands a number of encoding passes in order to decide model parameters.

Here, we propose a greatly simplified dependent bit allocation algorithm to reduce the complexity requirement furthermore. Given the fact that the bit allocation problem is essentially a QP decision problem among coding units, we propose an adaptive QP decision mechanism that incorporates the input video characteristics efficiently. The proposed algorithm consists of two QP decision methods. First, the QP of TL-0 key frames is determined by the GOP-based rate model. Then, the QPs of the remaining temporal layers are adaptively determined by considering inter-layer dependency within a GOP. It will be shown by experimental results that the proposed algorithm outperforms the one in the JSVM 9.12 as well as that proposed by Liu *et al.* [3].

The rest of this article is organized as follows. We first formulate the temporal layer bit allocation problem and discuss the solution framework in Sec.2. Then, the proposed bit allocation algorithm is described in Sec.3 and experimental results are shown in Sec.4. Finally, concluding remarks are given in Sec.5.

2. PROBLEM DESCRIPTION

2.1. Simplified Problem Formulation

In this section, we analyze the optimal solution framework in [5], the optimization equation is numerically solved by par-

*Correspongding author
This work was supported by National Natural Science Foundation of China under contract No.61101078, National Basic Research Program of China under contract No.2009CB320907 and Doctoral Fund of Ministry of Education of China under contract No.20110001120117.

tial differentiation. That is, the solution can be acquired by differentiating the cost function with respect to variables $q_i's$ and λ and setting them to zero. Then, we have $N_T + 1$ equations:

$$\frac{\partial J}{\partial q_i} = \omega_i \cdot b_0 \cdot \beta_0 \cdot q_i^{\beta_0 - 1} - \lambda \cdot a_i \cdot \alpha_i \cdot q_i^{-\alpha_i - 1} = 0, \quad (1)$$

$$\frac{\partial J}{\partial \lambda} = a_0 \cdot q^{-\alpha_0} + \cdots + a_{N_T-1} \cdot q^{-\alpha_{N_T-1}} - R_T = 0. \quad (2)$$

where $i = 0, \cdots, N_T - 1$.The objective is to solve the above system of equations for QP_i's (or q_i's) such that the cost function is minimized.

First, we can compute the relationship between q_0 and q_i with (1). That is,

$$q_i = \left[\frac{\omega_0}{\omega_i} \cdot \frac{a_i \cdot \alpha_i}{a_0 \cdot \alpha_0} \right]^{\frac{1}{\beta_0 + \alpha_i}} \cdot q_0^{\frac{\beta_0 + \alpha_0}{\beta_0 + \alpha_i}}. \quad (3)$$

Based on the approximated q-QP relation

$$q = 0.6267 \cdot e^{0.1155 \cdot QP},$$

we can rewrite Eq. (3) in terms of QP as

$$QP_i = C_{i,0} \cdot QP_0 + C_{i,1}, \quad (4)$$

where

$$C_{i,0} = \frac{\beta_0 + \alpha_0}{\beta_0 + \alpha_i}, \quad (5)$$

$$C_{i,1} = \frac{231}{2000} \cdot \left[\frac{1}{\beta_0 + \alpha_i} \cdot \ln \left(\frac{\omega_0}{\omega_i} \cdot \frac{a_i \cdot \alpha_i}{a_0 \cdot \alpha_0} \right) + \frac{\alpha_0 - \alpha_i}{\beta_0 + \alpha_i} \cdot \ln 0.6267 \right]$$

Then, the original bit allocation problem with a number of unknown model parameters is successfully simplified to Eq. (4) with only two parameters $C_{i,0}$ and $C_{i,1}$. Thus, instead of determining model paratmers a_i, α_i, b_i, β_i and ζ_i (or ω_i), we may focus on the selection of parameters $C_{i,0}$ and $C_{i,1}$ directly. Our approach to this problem will be detailed in the next section.

3. SINGLE PASS BIT ALLOCATION ALGORITHM

In this section, we first study the characteristics of parameters $C_{i,0}$ and $C_{i,1}$ in Eq. (4). Then, based on this investigation, we propose a simplified dependent bit allocation algorithm for hierarchical B-pictures. Generally speaking, a good rate control algorithm should satisfy two requirements: 1) accurate rate control to meet the target bit rate and 2) the R-D efficient bit allocation among coding units. We address these two issues by GOP-based rate modeling and temporal layer QP decision, respectively.

3.1. Characteristics of Model Parameters

Although we have a simple expression relating QP_i to QP_0 in Eq. (4), coefficients $C_{i,0}$ and $C_{i,1}$ are still functions of model parameters a_i, α_i, b_i, β_i and ζ_i (or ω_i). It is not convenient to re-compute model parameters for every GOP, which demands the coding of the same GOP several times using different quantization step sizes. For this reason, we adopt an empirical approach to determine the characteristics of $C_{i,0}$ and $C_{i,1}$ directly.

Table 1. Coefficients for the QP decision

Sequence	Coefficient	GOP=8		
		TL-1	TL-2	TL-3
Foreman	C0	0.99	0.98	0.95
	C1	0.80	3.61	6.28
Hall	C0	0.97	0.94	0.92
	C1	2.49	5.76	9.47

The average values of parameters $C_{i,0}$ and $C_{i,1}$ in Eq. (4) for several test sequences are shown in Table 1. We see from Table 1 that $C_{i,0}$ does not vary a lot with a value close to one while $C_{i,1}$ varies more significantly according to the characteristics of the underlying video. Generally speaking, $C_{i,1}$ is larger for low motion sequences such as Hall, which have stronger inter-layer dependency. Actually, by approximating $C_{i,0}$ by one in Eq. (4), we have

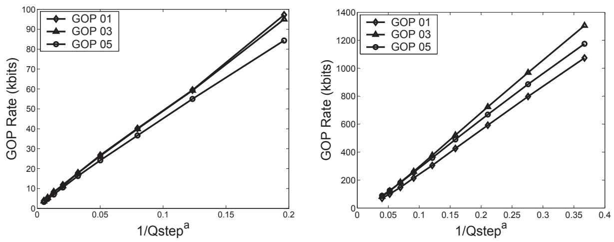
$$C_{i,1} \approx QP_i - QP_0 = \Delta QP_i. \quad (6)$$

The result in Eq. (6) is simple yet interesting. That is, parameter $C_{i,1}$ is used to predict the QP difference between QP_i and QP_0. When there exists inter-layer dependence, we can adopt a sequence of QP values with a larger gap between them.

3.2. GOP Rate Modeling

The rate of a video coding unit is often expressed as a function of the quantization step size. One common approach in rate modeling is to examine the statistical characteristics of transform coefficients. That is, we study the histogram of transform coefficients of macroblocks (MBs) and fit it with a certain probability distribution. For example, the quadratic rate model is a direct consequence of the Laplacian distribution assumption of source statistics. Recently, Kamaci *et al.* [6] proposed another frame-based rate model based on the assumption of the the Cauchy distribution of transform coefficients.

Although the rate model in [6] was derived from the statistics of a single frame, it can be easily extended to a GOP. That is, the GOP rate model should be consistent with the frame rate model since a GOP consists of multiple frames whose MB transform coefficients follow the same statistical characteristics. Actually, based on the same idea, the GOP rate was modeled as an inverse relation with the GOP average quantization step size in [7].



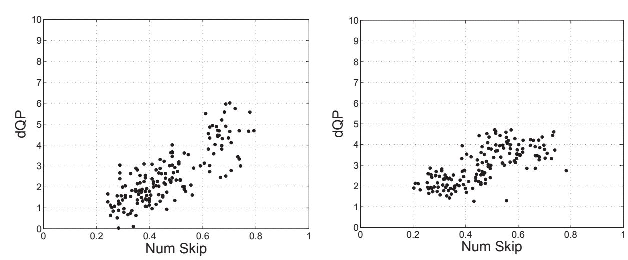
(a) City (b) Football

Fig. 1. The GOP rate characteristics with respect to the quantization step size of the key frame for test sequences: (a) City, QCIF ($\alpha = 1.3$), (b) Football, CIF ($\alpha = 0.8$).

We show the GOP rate characteristics with respect to the quantization step size of the TL-0 key frame, where GOP N denotes the N-th GOP, for QCIF and CIF sequences in Fig.1. As shown in these figures, we see that the GOP rate characteristics can be modeled by the same statistical source as the frame. Furthermore, it is observed that the rate characteristics of adjacent GOPs from the same sequence are close to each other. As a result, the QP of the TL-0 key frame in a target GOP can be predicted from the GOP rate model constructed by the rate statistics of its previous GOP.

3.3. Temporal Layer QP Decision

The analysis in Sec.3.1 provides a guideline for temporal layer QP decision to achieve high coding efficiency. Based on the discussion in Sec.3.1, we set $C_{i,0}$ to one and treat $C_{i,1}$ as the QP difference (ΔQP_i) between TL-i and TL-0. In this section, we present a simple method that determines ΔQP_i adaptively according to the characteristics of the input video.



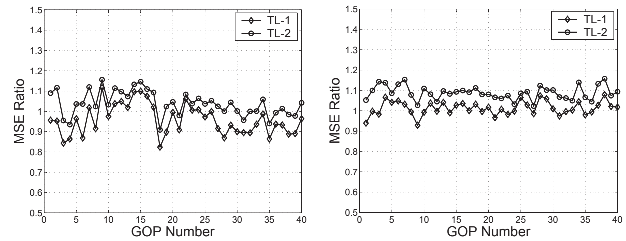
(a) City, QCIF (b) Tempete, CIF

Fig. 2. Illustration of the relationship between ΔQP and the number of skipped MBs: (a) City-QCIF, (b) Tempete-CIF.

First, we use the number of skipped MBs as a measure of inter-layer dependency. Fig. 2 shows ΔQP_i as a function of the number of skipped MBs, where ΔQP_i is obtained during the optimal bit allocation process as given in [5] and the number of skipped MBs is normalized by the total number of MBs in a frame. We see that ΔQP_i is roughly proportional to the number of skipped MB's, which can be written as

$$\Delta QP = C \cdot \frac{N_{skip}}{N_{MB}}, \tag{7}$$

where N_{skip} and N_{MB} are the number of skipped MBs and the number of MBs in a frame, respectively.



(a) Hall, QCIF (b) Mobile, CIF

Fig. 3. Illustration of the MSE ratios of test sequences: (a) Hall-QCIF, (b) Mobile-CIF.

Since the linear approximation in Fig. 2 is a little bit rough, we consider the ratio of mean squared errors (MSEs) between TL-0 and TL-i, *i.e.*, $\frac{MSE_i}{MSE_0}$, as another tool for bit allocation. We observe from the experimental data of the temporal layer bit allocation in [5] that MSE ratios between TL-0 and TL-i are stable and they do not vary much for different GOPs regardless of the QP differences between temporal layers. In Fig. 3, we plot the MSE ratios at different temporal layers and GOPs. Based on this observation, we can impose the following condition:

$$R_L \leq \frac{MSE_i}{MSE_0} \leq R_H, \tag{8}$$

where R_L and R_H are the lower and the upper threshold ratio values. Since the MSE ratio can be computed only after the quantization of the target coding unit, we employ the ratio of the MAD after the motion estimation instead of the MSE ratio in the proposed algorithm.

However, we still need to address the *chicken-and-egg* dilemma of the H.264 video encoder. That is, the rate-distortion optimization (RDO) process requires a QP value (QP_{RDO}) while the QP has to be determined by the number of skipped MBs and the MAD ratio that are only available after the RDO process. To address the dilemma, we consider the QPs for the RDO and the quantization process separately. They are determined by the following procedure.

- In the RDO process, the number of skipped MBs ($\hat{N}_{skip}$) is predicted from that of the collocated frame in the previous GOP and QP_{RDO} is determined by $\hat{N}_{skip}$.
- The QP for the quantization (QP_{Quant}) is determined by the updated number of skipped MBs (N_{skip}) after the RDO.

- If the MAD ratio is out of the threshold interval, QP_{Quant} value is changed accordingly.

Kwon *et al.* [7] has justified the separate treatment of QP_{RDO} and QP_{Quant}. That is, the R-D efficiency remains about the same if the absolute difference between them is small.

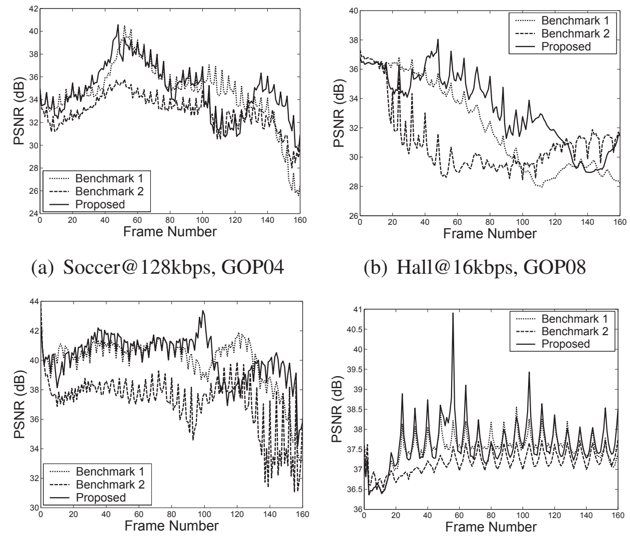
4. EXPERIMENTAL RESULTS

In this section, we compare the performance of the proposed rate control algorithm with two benchmarks. Benchmark 1 and benchmark 2 refer to the rate control algorithm in the JSVM reference software encoder and that proposed by Liu *et al.* in [3], respectively. These two benchmark algorithms are chosen since they also perform rate control in a single pass, and their complexity requirements are comparable. Thus, we concentrate mainly on the comparison of coding efficiency. In the experiment, we have encoded 161 frames at various bit rates depending on the input video characteristics. The GOP size is 8. The proportional constant C in Eq. (7) is set to 0.7, and the lower and higher thresholds, R_L and R_H in Eq. (8), are set to 0.9 and 1.2, respectively. The maximum difference (Δ) between QP_{RDO} and QP_{Quant} is set to 3 to avoid significant degradation of the R-D efficiency.

In Fig. 4, we compare the frame-by-frame PSNR values of the proposed algorithm and Benchmarks 1 and 2. We see that the PSNR level of the proposed algorithm is higher than those of benchmark schemes in most frames. Moreover, the PSNR level of Benchmark 1 has a significant drop at the end of the input sequence, which degrades the overall visual quality. In contrast, the proposed algorithm maintains reasonable quality level, although the quality fluctuation is a little higher than Benchmark 1 for some test sequences.

5. CONCLUSION

A practical single-pass rate control algorithm for H.264/SVC hierarchical B-pictures was developed in this work. The complex dependent bit allocation problem was greatly simplified based on detailed analysis on its optimal solution. It was shown by experimental results that the proposed rate control algorithm outperforms two benchmark rate control algorithms at various target bit rates for different test sequences.



(a) Soccer@128kbps, GOP04 (b) Hall@16kbps, GOP08

(c) Foreman@256kbps, GOP04 (d) Bridge-far@64kbps, GOP08

Fig. 4. Illustration of the frame-by-frame PSNR results of selected test sequences (a) Soccer-QCIF, (b) Hall-QCIF, (c) Foreman-CIF and (d) Bridge-CIF.

6. REFERENCES

[1] Heiko Schwarz, Detlev Marpe, and Thomas Wiegand, “Analysis of hierarchical B-pictures and MCTF,” in *Proc. IEEE International Conf. Multimedia and Expo*, July 2006, pp. 1929–1932.

[2] 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.

[3] Yang Liu, Zhengguo Li, and Yeng Chai 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.

[4] Kannan Ramchandran, Antonio Ortega, and Martin 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.

[5] Yongjin Cho, Jiaying Liu, Do-Kyoung Kwon, and C.-C. Jay Kuo, “H.264/SVC temporal bit allocation with dependent distortion model,” in *Proc. IEEE International Conf. Acoustics, Speech, and Signal Processing*, April 2009.

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

[7] Do-Kyoung Kwon, Yongjin Cho, and C.-C. Jay Kuo, “A simplified rate control scheme for non-conversational H.264 video,” in *International Workshop on Multimedia Signal Processing*, October 2007.