PRACTICAL RATE CONTROL ALGORITHM FOR TEMPORAL SCALABILITY IN SCALABLE VIDEO CODING

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

A rate control algorithm for hierarchical B-pictures in Scalable Video Coding (SVC) is proposed in this work. The complex inter-frame dependency issue is effectively addressed by the Q-distance policy decision rule while the statistical smoothing effect enables the GOP-based precise bit rate control. The simplicity of the decision processes greatly reduces the encoder complexity providing an efficient and effective rate control algorithm with hierarchical B-pictures. Experimental results verify the significant performance gain by the proposed algorithm.

***Index Terms*—** Rate control, temporal scalability, scalable video coding

1. INTRODUCTION

A rate control algorithm provides an efficient encoder controlling tool with video codecs enabling the production of the compressed bit stream at a target bit rate precisely. With its unique GOP structure and intractable three dimensional scalability, H.264/SVC depicts an interesting but extremely involved rate control problem. That is, it is highly required to develop an efficient and effective rate control algorithm for the practical encoder control of H.264/SVC.

Our first target is the rate control of hierarchical-B pictures which lies in the spatial and the quality base layer. A GOP of H.264/SVC is composed of a key frame of I- or P-pictures and hierarchically aligned B-pictures to implement the temporal scalability and to achieve high coding performance. The generic coding efficiency issue of the scalable video is effectively addressed by the temporal layer QP-cascading, by which the quantization parameter (QP) for each frame is determined based on the temporal layer it belongs to [1]. With the complex inter-dependency of the hierarchical B-pictures as the major obstacle, a rate control algorithm has to determine the QP of each frame to produce a rate-distortion (R-D) optimized bit stream at a target bit rate.

*Correspongding author

This work was supported by National Natural Science Foundation of China under contract No. 61071082, National Basic Research 973 Program of China under Contract 2009CB320907 and Beijing Municipal Natural Science Foundation under contract No. 4102025.

The conventional trellis-based solution framework is performed by Pranantha *et al.* in [2] and Xu *et al.* proposed an rate control algorithm for hierarchical B-pictures, which determines the frame QPs based on scaling factors [3]. Whereas the first approach suffers from the complexity requirement for the R-D data generation, the coding gain by the latter is relatively small when compared to that by QP-cascading scheme employed in the current reference software Joint Scalable Video Model (JSVM). The proposal JVT-W043 [4] introduced the quadratic rate control scheme into the base layer of JSVM, which has been adopted.

With the development of a practical algorithm as the main objective, we propose a two-pass rate control algorithm for hierarchical B-pictures. The experimental results show that the proposed algorithm precisely controls the output bit rate with a significant coding gain at a highly reduced complexity. The proposed algorithm is composed of a sequence of two decision stages which separately tackles the R-D performance and the rate control issues. The pre-analysis of an original input GOP addresses the R-D performance issue while avoiding the excessive complexity of the exhaustive R-D analysis as in [2]. Moreover, the simplified GOP rate model in [5] enables the precise bit rate control in a single pass of pre-encoding.

The rest of this paper is organized as follows. The problem formulation and the whole solution framework are formulated and described in Sec.2. Sec.3 has demonstrated the practical rate control algorithm for temporal scalability. Experimental results are given in Sec.4. Finally, concluding remarks are given in Sec.5.

2. SOLUTION FRAMEWORK

2.1. Problem Description

The production of an R-D optimized compressed bit stream at a target bit rate is the final objective of rate control algorithms. Conventionally, the Lagrangian optimization method has provided a clear problem formulation and a neat solution framework with the R-D optimized rate control problem. However, despite the advantage of its ability to provide a clear analytical expression, it can hardly serve as a solution framework to the rate control problems considering highly complicated

978-1-4577-1302-6/11/$26.00

inter-frame dependency.

The major obstacle against the employment of the Lagrangian framework is the intractability of the dependency relationship of frames, which makes it infeasible to develop analytical models of the rate and the distortion characteristics of dependent frames. For this reason, we divide the R-D optimized rate control problem into consecutive sub-problems of a frame-layer *Q-distance policy* and *Q-root* decision problems to draw a tractable solution framework.

2.2. Problem Formulation

The GOP rate control via frame layer bit allocation is formulated as a frame layer *Q-selection policy* decision problem of which the result is a vector of frame QPs. Mathematically,

$$\mathbf{Q}^* = \operatorname*{arg\,min}_{\mathbf{Q}\in\mathcal{Q}^{N_{GOP}}} \{D_{GOP}\} \quad \text{Subject to } R_{GOP} \leq R_{T,GOP}, \tag{1}$$

where $D_{GOP} = \sum_{i=1}^{N_{GOP}} D_i$, $R_{GOP} = \sum_{i=1}^{N_{GOP}} R_i$, $\mathbf{Q}^* = [Q_1^*, Q_2^*, \ldots, Q_{N_{GOP}}^*] \in \mathcal{Q}^{N_{GOP}}$, N_{GOP} is the number of frames in a GOP, $\mathcal{Q} = \{0, \ldots, 51\}$ is the set of all possible Q values and $R_{T,GOP}$ is the target GOP rate.

The *Q-selection policy* decision problem is divided into two tasks of 1) the *Q-distance policy* and 2) the *Q-root* decision problems. We define a vector $\mathbf{\Delta} = [\delta_1, \delta_2, \ldots, \delta_{N_{GOP}-1}]$ to represent the QP differences of B-frames from its reference frames. Then, the solution to the *Q-distance policy* decision problem seeks for an optimal QP-difference vector resulting in the minimum GOP distortion at any given GOP rate. Mathematically,

$$\mathbf{\Delta}^* = \operatorname*{arg\,min}_{\mathbf{\Delta}\in\mathbf{\Delta}^{N_{GOP}-1}} \{D_{GOP}|R_{GOP}\} \tag{2}$$

Given a QP-difference vector, the *Q-root* decision stage follows seeking for the key frame QP (Q_{key}^*) which meets the target bit rate constraints. That is,

$$Q_{key}^*|\mathbf{\Delta} = \operatorname*{arg\,max}_{Q\in\mathcal{Q}}\{R_{GOP}(Q|\mathbf{\Delta})\} \quad \text{Subject to } R_{GOP} \leq R_{T,GOP}. \tag{3}$$

2.3. Q-distance Policy Decision

The frame layer inter-GOP *Q-distance policy* decision is considered in this stage by the QP-difference vector ($\mathbf{\Delta}$) decision process. For inter-predictively coded frames, it is desirable that frames with high temporal variations are assigned large number of bits for the overall coding performance enhancement of a GOP. Conversely, frames consuming more bits at the same QP are necessarily of higher temporal variations than frames requiring less bits.

The QP-cascading in H.264/SVC provides a very effective tool for the improvement of coding efficiency of the hierarchical B-pictures. However, in the current JSVM implementation, its usage is mislead in that only the temporal layer is considered for the internal QP assignment. That is, the characteristics of individual frames are not properly exploited resulting in a possible coding gain loss within a GOP. Given the base QP (Q_{base}) of the lowest temporal layer, the temporal layer QP allocation is determined by

$$Q_i = Q_{base} - 6 \times \log_2(SF_i),$$

where i refers to the temporal layer and the scaling factor SF_i is determined by the temporal layer and the filtering and the prediction options [6].

In the proposed algorithm, the B-frame complexity introduced in [5] is employed for the *Q-distance policy* decision process for its simplicity. The complexity can be computed from original pictures by

$$S_B = \sum_{y=0}^{H}\sum_{x=0}^{W} \min\begin{pmatrix} |f_i(x,y) - g_i(x,y)|, \\ |f_i(x,y) - h_i(x,y)| \end{pmatrix}, \tag{4}$$

where f_i is the current frame and g_i and h_i refer to the closest forward and backward reference frame of f_i respectively.

The *Q-distance policy* decision rule is defined as a binary decision problem represented by

$$\begin{aligned} S_{B_i} \geq \zeta \Rightarrow \delta_i = 1 \\ S_{B_i} < \zeta \Rightarrow \delta_i = 2, \end{aligned} \tag{5}$$

where ζ is a pre-determined threshold value. The value of δ is set to 1 or 2 guided by the monotonicity property while avoiding excessive frame-by-frame quality fluctuations in the hierarchical B-pictures.

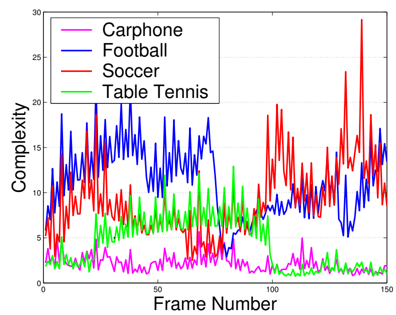
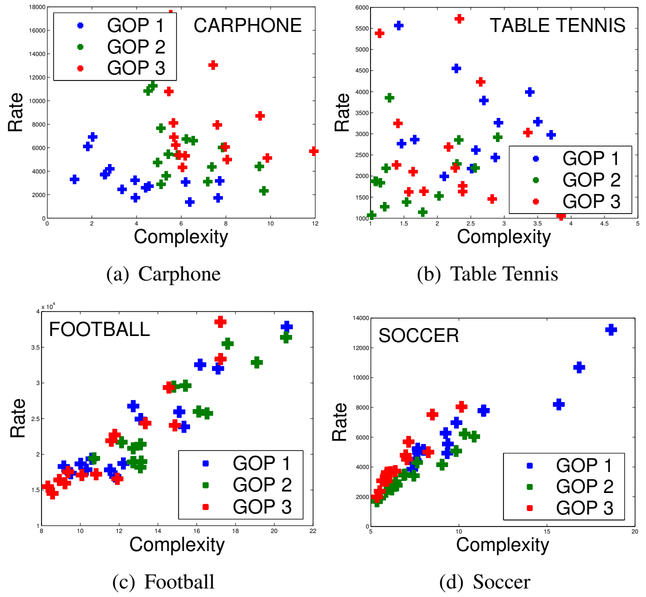


Fig. 1. B-frame complexity calculation

As depicted in Fig.1, sequences with high temporal variations (*e.g.*, Football and Soccer) show dramatic complexity variations over frames while the fluctuation remains relatively stable with low motion sequences such as Carphone and Table Tennis.

Fig.2 empirically justifies the employment of the B-frame complexity for the *Q-distance policy* decision process. Whereas the rate and the complexity are irrelevant in low motion sequences, the relationship between them depicts a strong correlation for the sequences with high temporal variations. This observation plays the key role in determining the threshold value (ζ) in Sec.3.



(a) Carphone (b) Table Tennis

(c) Football (d) Soccer

Fig. 2. Rate vs. Complexity

2.4. Q-root Decision

The simplified linear GOP rate model proposed in [5] is employed to produce the bit stream at the target bit rate, where the rate of a GOP is modeled as a function of the GOP complexity and the quantization step size (Qs). Mathematically,

$$R = \eta \cdot \frac{S}{Q_s}, \tag{6}$$

where η is the model parameter depending on the GOP characteristics.

Since the GOP complexity is computed from the original frames, the GOP complexity (S) is constant given a GOP. For this reason, we can further simplify the expression into

$$R = \alpha \cdot Q_{s_{key}}^{-1}, \tag{7}$$

where $Q_{s_{key}}$ is the quantization step size of the key frame of a GOP.

One of the major advantages of the GOP-based rate control is the statistical smoothing effect. That is, the random behaviors of the smaller encoding units such as frames, slices and macroblocks (MBs) are statistically stabilized resulting in a stationary process. Because the linearity of GOP rate still holds despite the internal variations caused by individual frames' rate fluctuations, the statistical smoothing effect becomes even more desirable for extremely dependent hierarchical B-pictures than for the conventional H.264/AVC GOP structures.

3. SIMPLIFIED RATE CONTROL ALGORITHM

The proposed rate control algorithm is composed of three stages of pre-analysis for *Q-distance policy* decision, model-based *Q-root* decision and encoding at a target bit rate. Fig.3 explains the GOP-based encoding procedure of the proposed algorithm, where the separation of the R-D performance and the rate control issues is successful.

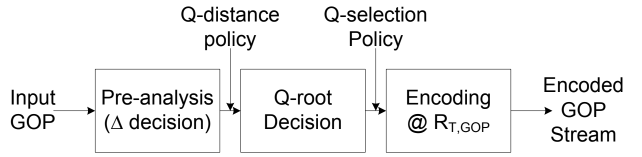


Fig. 3. Proposed two-pass rate control algorithm

The first stage of encoding analyzes the input video for the internal *Q-distance policy* decision which provides the *Q-distance* vector ($\mathbf{\Delta}$) with the *Q-root* decision stage. In the *Q-root* decision stage, the key frame QP (Q_{key}) is determined by combining $\mathbf{\Delta}$ and the simplified GOP rate model in Eq.7.

The encoding procedure of the proposed algorithm is summarized below.

1. *Initial bit allocation to a GOP*
   Based on frame rate F and channel rate C, allocate the target bit budget R_T to a GOP according to
   $$R_T = N_{GOP} \cdot \frac{C}{F} + R_0,$$
   where R_0 compensates for the errors between the target bits and the actual bits of the previous GOP.

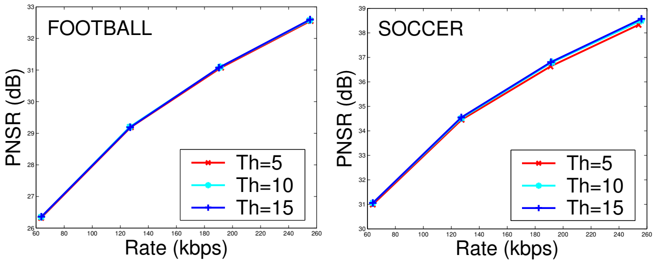
2. *GOP pre-analysis for Q-distance policy decision*
   The *Q-distance* vector $\mathbf{\Delta}$ is determined by the decision rule depicted in Eq. (5).
   $$\Delta_i = S_{B,i} \geq \zeta \ ?\ 1 : 2\,,$$

3. *Q-root decision for Q-selection policy finalization*
   After encoding the current GOP encoded with the key frame QP of the previous GOP, $Q_{key,prev}$, the model parameter η is updated. Based on the updated model, $Q_{key,curr}$ is determined finalizing the decision on *Q-selection policy* for the encoding of the current GOP.

4. *Encoding at the target bit rate*
   The current GOP is encoded with $Q_{key,curr}$ and $\mathbf{\Delta} = [\delta_1, \delta_2, \ldots, \delta_{N_{GOP}-1}]$ to produce the final bit stream of the current GOP.

5. *Iteration*
   Steps 1 through 4 are applied to the following GOPs until the end of the input video.

The threshold, ζ, is determined by a careful investigation of its effect on the performance of the algorithm. As shown in Fig.4, a low threshold value, *e.g.*, $\zeta = 5$, results in the R-D performance degradation. This is because the improper bit assignment to frames with low temporal variations. We set $\zeta = 15$ in the experiment based on the observation of the irrelevance of the rate and the complexity in Sec.2.3 and the dynamic range of the frame complexity depicted in Fig.1.



(a) Football (b) Soccer

Fig. 4. Effect of threshold on performance

4. EXPERIMENTAL RESULT

The proposed rate control algorithm was implemented on JSVM9.12 [6] in our experiments. The experiment is conducted on 241 frames corresponding to 15 GOPs from 4 sequences in the QCIF format for the performance evaluation. The frame rate of each sequence is 30 fps and every key frame in a GOP is encoded as an I frame. The GOP is composed of 16 frames which corresponds to 5 levels of temporal scalability.

The performance of the proposed temporal layer rate control algorithm is compared with that of JVT-W043 rate control scheme implemented in the JSVM software reference codec [4]. Table 1 summarizes the experimental results. The proposed algorithm achieves the maximum bit rate mismatch of lower than 1% from various target bit rates ranging from 64 to 256 kbps, and the average coding gain is 0.71dB.

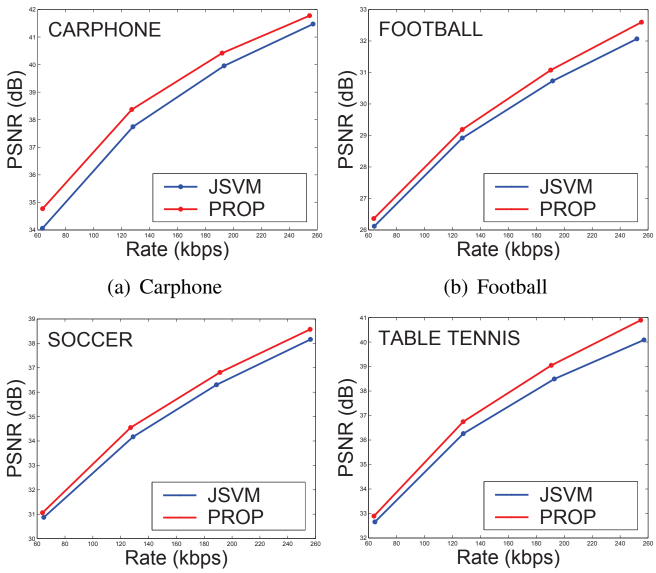
Moreover, while achieving the target bit rate the performance of the proposed algorithm is compared to that of the FixedQPEncoder of JSVM. With the FixedQPEncoder, an input video is iteratively encoded until the target bit rate constraint is satisfied with the same QP assignments for all GOPs in the input video. The average coding gain is 0.43dB with the maximum gain of 0.8dB with the Table Tennis sequence.

5. CONCLUSION

The proposed algorithm achieves a significant coding gain with a precise rate control to the target bit rate for the hierarchical B-pictures. By separately considering the R-D performance and the rate control issues, the encoding process could be significantly simplified.

Table 1. Experimental results comparing with JVT-W043

Sequence	Target Rate	JVT-W043		Proposed	
		Bitrate	PSNR	Bitrate	PSNR
Carphone	64k	63.78	33.75	63.54	34.77
	128k	130.29	37.23	127.33	38.37
	192k	205.84	39.33	191.94	40.42
	256k	274.57	40.84	254.53	41.77
Soccer	64k	66.51	30.91	63.90	31.06
	128k	133.64	34.20	127.15	34.55
	192k	197.58	36.35	191.31	36.81
	256k	264.46	38.03	256.08	38.57



(a) Carphone (b) Football (c) Soccer (d) Table Tennis

Fig. 5. R-D performance comparison

6. REFERENCES

[1] Heiko Schwarz, Detlev Marpe, and Thomas 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, no. 9, pp. 1103–1120, September 2007.

[2] Danu Pranantha, Munchurl Kim, Sangjin Hahm, Byungsun Kim, Keunsik Lee, and Keunsoo Park, "Dependent quantization for scalable video coding," in *Proc. IEEE International Conference Advanced Communication Technology*, 2007.

[3] Long Xu, Wen Gao, Xiangyang Ji, and Debin Zhao, "Rate control for hierarchical B-picture coding with scaling-factors," in *IEEE International Symposium on Circuits and Systems*, 2007.

[4] Athanasios Leontaris, "Rate Control for the Joint Scalable Video Model (JSVM), JVT-W043," Joint Video Team, San Jose, California, USA, April 2007.

[5] 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.

[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.