A RATE CONTROL ALGORITHM FOR LOW-DELAY H.264 VIDEO CODING WITH STORED-B PICTURES

Sergio Sanz-Rodríguez[1], Manuel de-Frutos-López, Iván González-Díaz, Jesús Cid-Sueiro

Department of Signal Theory and Communications
Universidad Carlos III, Leganés (Madrid), Spain

ABSTRACT

A rate control (RC) algorithm for H.264 video coding with stored-B (SB) pictures is proposed for low-delay applications. Different models for P and SB pictures are defined for a better QP and MAD estimation. Furthermore, a novel saw-tooth shaped model of target buffer level has also been introduced for a proper bit allocation in GOP structures with SB pictures. Our experimental results show that this proposal outperforms the reference software RC in terms of buffer occupancy and target bit rate adjustment at the expense of slight quality reduction.

***Index Terms*—** H.264, rate control, stored-B pictures, bit allocation, low-delay.

1. INTRODUCTION

The new video applications demand both high quality and low bit rates. H.264/AVC achieves much higher coding efficiency than that of the previous standards. Its main improvements are concentrated in motion compensation, with variable block size, quarter-pixel motion vectors and up to sixteen reference frames. New entropy coding techniques are also defined including CABAC and CAVLC modes.

Beside these new tools, a RC scheme (not standardised) must be defined to comply with network constraints (available rate or delay), in transmission or storage applications. An adaptive RC scheme has been adopted by the Joint Video Team (JVT) for H.264 [1]. It has been widely studied to improve the target bit assignment [2, 3], its behaviour in non-stationary situations [4] and its efficiency for specific applications [3, 5]. Other strategies have also been proposed, such as [6], to enhance the quality. Most of these proposals have been tested with IP...P coding patterns.

Bipredictive (B) pictures allow further rate reduction than predictive (P) by combining blocks from two previously coded frames, reducing residual data. Reference proximity is also a key factor for an accurate prediction and therefore a better quality can be achieved by means of SB pictures. Unfortunately, the use of this kind of pictures has been ignored in previous works concerning RC in H.264.

B pictures have the effect of a noticeable delay ($\tau = \tau_R + \tau_B$), where τ_R is an inherent delay involved in the reordering of pictures, and τ_B is a delay introduced due to the buffer size needed to smooth rate variations, especially with the RC described in [1]. Therefore, working with B pictures in low-delay applications is not usual. However, the proposed RC focuses on solving this drawback, by including SB picture management. A better buffer control and target rate adjustment are achieved, and τ_B is considerably reduced.

The rest of the paper is organised as follows. In section 2, the main improvements over [1] are described. Section 3 makes a description of experiments. Finally, a set of conclusions and further lines are established in section 4.

2. PROPOSED RATE CONTROL FOR GOP STRUCTURES WITH STORED-B PICTURES

The proposed RC is based on [1] and works in three levels: GOP layer, picture layer and basic unit layer. In the following subsections we describe the contributions of this paper for SB picture management at each level.

2.1. GOP layer

The GOP layer sets the initial value of the quantization parameter, $QP_i(1)$, for each GOP *i*, with *i*>1, by averaging the QP values of all inter stored pictures (P and SB types). Thus, more data than in the original RC are used to compute the average QP value:

$$QP_i(1) = \frac{SumQP(i-1)}{N_s(i-1)} - min\left\{2, \frac{N_{i-1}}{15}\right\}$$
$$QP_i(1) = max\{QP_{i-1}(1) - 2, min\{QP_{i-1}(1) + 2, QP_i(1)\}\}$$
$$QP_i(1) = QP_i(1) - 1 \text{ if } QP_i(1) > QP_{i-1}(N_{i-1}) - 2$$

[1]This work has been supported by PRODYS, S. L. and the regional grant UC3M-TEC-05-059. Thanks to J. C. Pujol and R. Santos.

where $N_s(i-1)$ is the total number of P and SB pictures in the $(i-1)^{th}$ GOP, $SumQP(i-1)$ is the sum of their average picture QP values, and N_{i-1} is the total number of pictures in the last GOP.

2.2. Picture layer

At this level, the QP value is computed for each picture. A certain amount of target bits for the j^{th} picture in the i^{th} GOP is allocated, then the QP value is determined by a quadratic model and, after coding the picture, the model coefficients are updated (see [1] for details).

2.2.1. Bit allocation

In the original RC, B pictures are not used as reference and thus the quality assigned to them is worse than that of P pictures. Therefore, the original RC does not need to estimate any amount of target bits and simply uses a linear interpolation method to obtain the QP value for B pictures. Besides of this, when SB pictures are going to be used as references, they should be encoded with better quality to improve the inter prediction. Furthermore, in the original RC the buffer fullness is only monitored before codifying P pictures. This supposes a serious problem in low-delay applications for GOP structures with B pictures, because of the potential overflow/underflow risk in video sequences with scene cuts or high complexity.

In order to circumvent the aforementioned drawbacks, this paper focuses on bit allocation enhancement by including both P and SB pictures in formulas. The linear model of target buffer fullness in [1] is a suitable solution for low-delay constrains, but it is optimized for IP...P coding patterns. An alternative is proposed which also models the buffer time variation when SB pictures are included in the GOP. This kind of pictures needs fewer bits than P pictures to achieve similar peak signal-noise ratio (PSNR) values. As consequence, the target bits for P and SB pictures should not be the same and then, the original linear model becomes unsuitable for the bit allocation task. A new saw-tooth shaped model is assumed (see Fig. 1). The target buffer level after encoding the j^{th} picture in the i^{th} GOP, $S_i(j)$, is determined by:

$$S_i(j) = S_i(j-1) - \frac{A - GOP\_TBL}{B} - C \cdot \frac{R_i(j)}{f} \cdot \left(1 - \frac{D \cdot (L+1)}{\overline{W}_{p,i}(1) + \overline{W}_{sb,i}(1) \cdot L}\right)$$

$$A = \begin{cases} S_i(2) & , j \le L+1 \\ S_i(L+1) & , j > L+1 \end{cases}, \quad B = \begin{cases} N_s(i) - 1 & , j \le L+1 \\ N_s(i) - L & , j > L+1 \end{cases},$$

$$C = \begin{cases} 0 & , i = 1 \\ 1 & , i > 1 \end{cases} \text{ and } D = \begin{cases} \overline{W}_{p,i}(1) & , P \\ \overline{W}_{sb,i}(1) & , stored - B \end{cases}$$

where $\overline{W}_{p,i}(1)$ and $\overline{W}_{sb,i}(1)$ are the average complexity weights of P and SB pictures in the intra picture of the i^{th} GOP computed as in [1], respectively, L is the number of successive B pictures between two P pictures, $R_i(j)$ is the actual target bit rate, and f is the frame rate. The GOP target buffer level (*GOP_TBL*) is set to zero in the experiments.

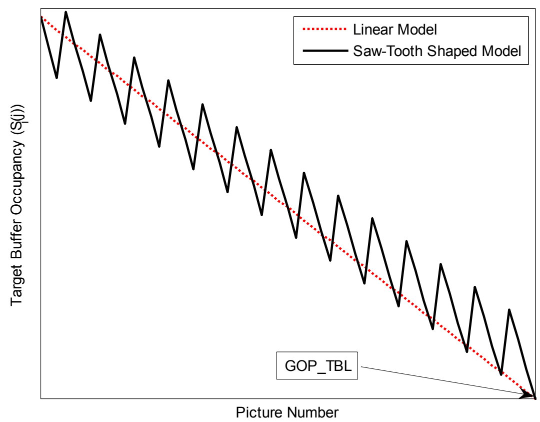

Fig. 1. Target buffer occupancy for a GOP with only P pictures (linear model) and including SB pictures (saw-tooth shaped model).

The target bits for the j^{th} picture in the i^{th} GOP, $\tilde{T}_i(j)$, depends on the combination of the target bit rate, the frame rate and the actual buffer occupancy, $V_i(j)$:

$$\tilde{T}_i(j) = \frac{R_i(j)}{f} + \gamma \cdot (S_i(j) - V_i(j))$$

where γ is a constant. The number of remaining bits, $B_i(j)$, should also be considered for a better adjustment:

$$\hat{T}_i(j) = \frac{K \cdot B_i(j)}{W_{p,i}(j-1) \cdot N_{p,r} + W_{sb,i}(j-1) \cdot N_{sb,r}}$$

$$K = \begin{cases} W_{p,i}(j-1) & , P \\ W_{sb,i}(j-1) & , stored - B \end{cases}$$

where $N_{p,r}$ and $N_{sb,r}$ are the number of remaining P and SB pictures, respectively. Usually, SB pictures need fewer bits than P pictures to achieve similar qualities, and therefore $W_{sb,i}(j-1) < W_{p,i}(j-1)$. Then the target bits are a weighted combination of $\tilde{T}_i(j)$ and $\hat{T}_i(j)$:

$$T_i(j) = \beta \cdot \hat{T}_i(j) + (1-\beta) \cdot \tilde{T}_i(j)$$

where β is a constant. Finally, $T_i(j)$ is bounded to conform to hypothetical reference decoder (HRD) constraints [1].

2.2.2. QP estimation

The QP for j^{th} picture is obtained through a quadratic model which depends on the target bits and the mean absolute difference (MAD). However, MAD is only available after picture encoding (egg-chicken dilemma), so a lineal regression method is used to predict it. Similarly to [7] the proposed algorithm works with different prediction models (for MAD and QP) for each type of stored picture.

2.3. Basic unit layer

A basic unit (BU) is a group of MBs in raster scan order which share the same QP. The number of MBs in a BU must be a fraction of the total number of MBs in a picture. Since the aim of BU layer is a fine adjustment of target bits, the buffer occupancy levels are decreased, possibly at the expense of slight PSNR reductions.

An additional feature of proposed algorithm is that BU layer is now included for SB pictures. Therefore, separate prediction models are also used for both P and SB pictures here. Another minor change is that, in order to obtain more reliable predictions, co-located BUs must belong to the last stored picture of the same type when computing MAD estimation for current BU.

3. EXPERIMENTS AND RESULTS

The proposed RC for low-delay applications and GOP structures with SB pictures is compared with the original RC [1] adopted by the JVT software version JM 9.2 [8]. Two sets of sequences have been defined. The first one, including scene cuts and high complexity sequences, are linkings of "Paris" and "Football" (named "Parifoot" in Table 1) and linkings of "Garden", "Stefan" and "Football". Sequences from the second set are typically used in video coding experiments ("Mobile", "Football" and "Highway"). A total of 300 pictures at 30 frames per second have been encoded for each sequence. The GOP structure is defined as two SB pictures between I or P pictures and an intra picture each second. R-D optimization has been disabled (for real-time coding applications) and symbol mode is CAVLC. The γ and β values are set to 0.75 in the experiments with the proposed RC. In JM 9.2 RC, both γ and β have been chosen based on experiments with a minimal peak buffer occupancy (PBO) criterion. Their final values have been set to 0.25 and 0.90 respectively (default values suggested in [1]). Several BU sizes have been tested, but a value of 22 MBs (a row of MBs in CIF format) have been chosen as an appropriate size for low-delay applications.

Luminance PSNR (average and standard deviation), target bit rate adjustment and PBO have been used to compare the performance in both RC algorithms. With the proposed RC, the buffer occupancy is better controlled (see Fig. 3 and Fig. 5). In general, PBO is higher for any bit rate and sequence (see Fig. 2 and Fig. 4) with JM 9.2 RC. Therefore, the adjustment to the target rate is finer. On the other hand, a certain PSNR reduction must be accepted (up to 0.38 dB in "PariFoot" at 256 kbits/s). PSNR fluctuations are also lower with our RC (see Table 1).

Sequence @Rate (kbits/s)	Algorithm	Average PSNR (dB)	PSNR Deviation (dB)	Obtained Rate (kbits/s)
PariFoot @256	JM 9.2	27.53	3.26	256.99
	Proposed	27.15	2.73	256.40
PariFoot @512	JM 9.2	30.98	3.43	513.75
	Proposed	30.71	3.09	512.20
PariFoot @1024	JM 9.2	35.17	4.16	1029.25
	Proposed	34.99	3.80	1024.04
Mobile @256	JM 9.2	25.71	0.88	257.41
	Proposed	25.36	0.91	256.51
Mobile @512	JM 9.2	29.06	0.90	513.07
	Proposed	28.74	0.81	512.72
Mobile @1024	JM 9.2	32.39	1.37	1025.35
	Proposed	32.21	1.20	1024.70
Highway @256	JM 9.2	38.41	0.69	256.81
	Proposed	38.30	0.59	256.69
Highway @512	JM 9.2	39.54	0.83	512.98
	Proposed	39.51	0.65	512.63
Highway @1024	JM 9.2	40.74	0.65	1025.97
	Proposed	40.72	0.68	1024.55

Table 1. Average PSNR, PSNR standard deviation and obtained bit rate for different target bit rates and sequences.

4. CONCLUSIONS AND FURTHER WORK

A new rate control algorithm for low delay applications with stored-B pictures has been defined. It shows a better performance in terms of buffer occupancy and target rate adjustment at the expense of slight PSNR reduction than the original RC. The proposed algorithm achieves lower PSNR variation with a perceptual quality improvement. The best results have been obtained in sequences with high complexity or scene cuts.

A large BU size achieves a better quality, since a smaller one supposes a finer buffer control. As consequence, a further work on dynamic BU size selection for quality enhancement could be the next step. A variable GOP size selection with early scene-cut detector could also be another area of interest for an additional buffer size reduction and quality improvement.

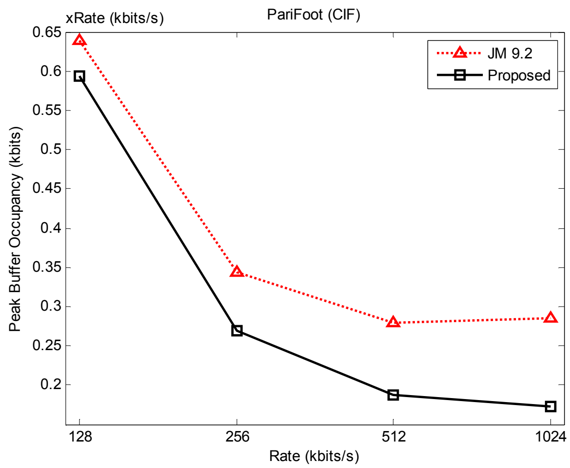

Fig. 2. Peak buffer occupancy vs. rate. "PariFoot".

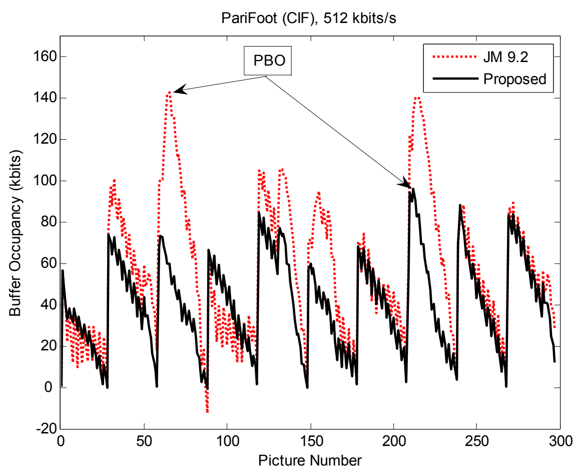

Fig. 3. Buffer occupancy variation. "PariFoot". 512 kbits/s.

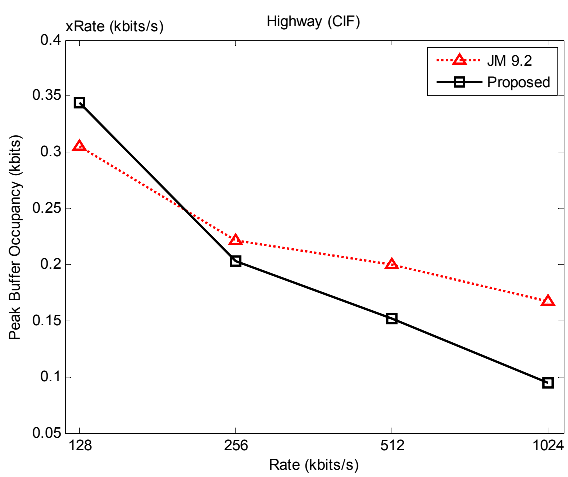

Fig. 4. Peak buffer occupancy vs. rate. "Highway".

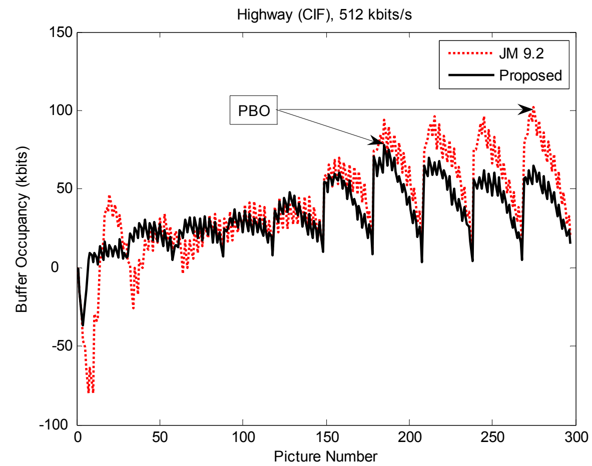

Fig. 5. Buffer occupancy variation. "Highway". 512 kbits/s.

5. REFERENCES

[1] S. Ma, Z. Li, and F. Wu, "Proposed draft of adaptive rate control," JVT-H017, 8th Meeting, Geneva, Switzerland, May 2003.

[2] M. Jiang, and N. Ling, "Low-delay rate control for real-time H.264/AVC video coding.," IEEE Transactions on Multimedia, vol. 8, no. 3, pp. 467–477, 2006.

[3] Y. Wu, L. Shouxun, Z. Yongdong, and L Haiyong, "Optimum bit allocation and rate control for H.264/AVC," JVT-O016, 15th Meeting, Busan, KR, Apr. 2005.

[4] Y. Liu, Z.G. Li, and Y.C. Soh, "Adaptive MAD prediction and refined R-Q model for H.264/AVC rate control," in IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2006, vol. 2, pp. II–905–II–908.

[5] J. Yang, X. Fang, and H. Xiong, "A joint rate control scheme for H.264 encoding of multiple video sequences," IEEE Transactions on Consumer Electronics, vol. 51, no. 2, pp. 617–623, 2005.

[6] A. Matsumura, S. Naito, R. Kawada, and A. Koike, "Effective rate control method for minimizing temporal fluctuations in picture quality applicable for MPEG-4 AVC/H.264 encoding," in IEEE International Conference on Image Processing, ICIP 2005, vol. 1, pp. I–569–72.

[7] S. Ma, W. Gao, and Y. Lu, "Rate control on JVT standard," JVT-D030, 4th Meeting, Klagenfurt, Austria, Jul. 2002.

[8] Karsten Sühring, H.264/AVC software coordination http://iphome.hhi.de/suehring/tml/download/old_jm/