LOW-DELAY WINDOW-BASED RATE CONTROL SCHEME FOR VIDEO QUALITY OPTIMIZATION IN VIDEO ENCODER

Yuan Li, Huizhu Jia[*], *Chuang Zhu, Meng Li, Xiaodong Xie and Wen Gao, Fellow, IEEE*

National Engineering Laboratory for Video Technology
Peking University, Beijing, China

ABSTRACT

The consistent video quality and encoding latency due to buffering are two important aspects in designing rate control scheme for the application of real-time video coding system. To well balance these two contrary objectives, we firstly analyze the constraint of buffer latency and the definition of a "consistent" video quality. Then a window-based rate control scheme is proposed with one window for controlling the rate and latency, while the other window for optimizing video quality. By applying low complexity frame level rate-distortion model in the testing sequences, our proposed method shows excellent performance in balancing the encoder buffer latency and optimized video quality. Besides, this one-pass rate control scheme is highly practical for the real-time video coding application.

Index Terms — Rate control, buffer latency, video quality, window-based, video encoder

1. INTRODUCTION

Modern video coding standards such as MPEG-2 [1], H.264/AVC [2] and AVS [3] are widely used for their excellent performance in video compression. However, in certain real-time encoding applications, such as video stream transmission and video communication etc., there exits strict constraints of bit rate and system latency. Meanwhile, the consistency of video quality is also required to be considered. Under these situations, rate control aims at achieving both shortened system latency and consistent video quality for the required bit rate budget.

Typical rate control can be classified as two categories: constant bit rate (CBR) and variable bit rate (VBR). In CBR applications, a short-term average bit rate must be reached regardless of the frame's characteristic. Since the video contents vary from frame to frame, CBR will cause the unavoidable fluctuation of video quality. On the other hand, VBR usually uses the global information of video content to allocate bits among different frames, which can guarantee the uniform video quality. However, this mechanism needs a two-pass or multi-pass encoding process, which is not suitable for real-time encoding application.

Many CBR rate control schemes are proposed and developed in literature. The TM5 in MPEG-2 [4] uses group of pictures (GOP)-based bit allocation, which allocates equal bit budget to all GOPs. Within a GOP, the frame bit allocation uses a fixed weighting factor according to the frame type. The assumption of this mechanism is that the video content is stationary, which is not always true and will cause fluctuation of video quality. Xie *et al*. [5] proposed a sequence-based bit allocation scheme by tracking the non-stationary characteristics in a video sequence. The consistent video quality can be achieved. However, the system latency is not discussed. In [6], Xu *et al*. proposed a window model about the picture quality and the buffer occupancy. To reduce the buffer latency, Ribas-Corbera *et al*. [7] applied a near-constant bit budget for each frame and introduced frame skipping mechanism. In [8], He *et al*. used the similar near-constant frame bit allocation and smoothed the rate shape by varying the distortion in a small range.

In summary, the balancing of encoder buffer latency and consistent video quality is not discussed sufficiently in these above-mentioned methods. In this paper, we firstly analyze the constraint of buffer latency and the definition of consistent video quality. Then a window-based rate control scheme is proposed with one window for controlling the rate and latency, while the other window for optimizing video quality. By applying low complexity frame level rate-distortion model in the testing sequences, our proposed method shows excellent performance in balancing the encoder buffer latency and consistent video quality.

The rest of this paper is organized as follows. In Section 2, the constraint of encoder buffer latency and consistent video quality are analyzed. Then the proposed window-based rate control scheme is presented in Section 3. Section 4 provides the experimental results and discussions. At last, we give a conclusion in Section 5.

2. PROBLEM ANALYSIS

2.1. Bit rate and buffer delay constraint

In real-time CBR applications, the coded bit stream is transmitted to the end users (decoder) through a bandwidth-fixed channel, with a bit rate constraint R_C. Let F and R_T be the frame rate and frame bit budget, respectively. Adopting a constant frame bit allocation by (1), similar to [7] [8],

$$R_T = \frac{R_C}{F} \qquad (1)$$

can reach the target bit rate constraint. In this mechanism, the latency is neglected since the bits of each encoded frame are assumed to be sent to the channel immediately. However, this scheme will cause the fluctuation of the video quality, especially when high motion or scene change occurs.

For the purpose of smoothing video quality, we need to allocate more bits to the frames which are more complex (in terms of encoding effort), while fewer bits to low-complexity frames. For this, the bit rate will be fluctuant. To smooth the fluctuation of the bit rate, we usually use a buffer, denoted as encoder buffer, between the video encoder and the channel since the transmitting channel can only allow fixed (in each "clock") bit rate. The encoded frame bits are sent to the buffer firstly and then transmitted to the channel with a constant bit rate. It is tolerable that the encoder buffer can maintain certain bits for some frames to smooth the bit rate fluctuation. Because of this, unavoidably a buffer delay is also introduced. When the bits of current frame arrive to this buffer, certain bits for the previous encoded frames are still kept in the buffer, which need to be transmitted to the channel first. Then the encoder buffer delay (T_D) is generated and can be denoted as the time from the bits of current frame arriving to the buffer to when the bits leaving the buffer. T_D can be calculated by (2)

$$T_D = \max\left\{\frac{Q(i)}{R_C} \mid i = 0,1,...,N-1\right\} \qquad (2)$$

where $Q(i)$ denotes the number of bits in the buffer when the bits of ith frame arrive, which can be represented by (3).

$$Q(i) = Q(i-1) + R_T(i) - \frac{R_C}{F} \quad i = 0,1,...N-1 \qquad (3)$$

In the decoder side, there also exists a corresponding relationship similar to (2) and (3). Due to the so-called mirror effect of buffer status [9], we only need to discuss the constraint of the buffer at the encoder side.

From (2), we realize that the buffer delay depends on the maximum number of bits existing in the buffer at any time during the transmission. Reducing the buffer delay implies minimizing the maximum $Q(i)$. For any start time t, considering the consecutive L frames from t to $(t + L - 1)$, we can get (4) from (3)

$$Q(t+L-1) = Q(t) + \sum_{i=t}^{t+L-1} R_T(i) - L \cdot \frac{R_C}{F} \qquad (4)$$

Let

$$\Delta = \sum_{i=t}^{t+L-1} R_T(i) - L \cdot \frac{R_C}{F} \qquad (5)$$

then we can get

$$Q(t+L-1) = Q(t) + \Delta \qquad (6)$$

From (6), it can be seen that to minimize the maximum $Q(i)$ for any time t the Δ should be set to zero. Then (5) becomes

$$\sum_{i=t}^{t+L-1} R_T(i) = L \cdot \frac{R_C}{F} \qquad (7)$$

this means that the total bits for consecutive L frames should be constant for any start time t. This bit rate constraint is stricter than the normally used CBR (GOP as a basic unit) because of the low-delay requirement.

Summary: It is contrary to the common belief that CBR means a constant bit rate per GOP basis, but the real meaning is directly related to the size of the buffer used between the encoder and the transmission channel.

2.2. Consistent video quality constraint

For a given bit budget, not only the high encoding quality of each frame needs to be achieved, but also the fluctuation of the video quality should be minimized for entire sequence in order to maintain a good viewing experience. Since the global video characteristic is not available in one-pass real-time encoding process, the consistent video quality should concern the variance of distortion for all previously encoded frames in respect of the current frame as shown in (8),

$$\text{var}(D_C) = \frac{1}{N}\left(\sum_{i=0}^{N-2}(D_{p,i} - \overline{D_N})^2 + (D_C - \overline{D_N})^2\right) \qquad (8)$$

where D_C denotes the target distortion of current frame, $D_{p,i}$ $i=0,1,\ldots,N\text{-}2$ denotes the distortion of the previously encoded frames, $\overline{D_N}$ denotes the average distortion including all encoded frames and the current frame, which can be represented by (9),

$$\overline{D_N} = \frac{1}{N}\left(\sum_{i=0}^{N-2} D_{p,i} + D_C\right) \qquad (9)$$

Then substituting (9) into (8), we can get

$$\text{var}(D_C) = \frac{N-1}{N^2}D_C^2 - \frac{2D_S}{N^2}D_C + \frac{N^2+N-1}{N^3}D_{SS} + \frac{1-2N}{N^3}D_S^2 \qquad (10)$$

where

$$D_S = \sum_{i=0}^{N-2} D_{p,i} \;,\;\; D_{SS} = \sum_{i=0}^{N-2} D_{p,i}^2 \qquad (11)$$

To minimize var(D_C), let

$$\frac{\partial \, \text{var}(D_C)}{\partial D_C} = 0 \qquad (12)$$

then we get

$$D_C = \frac{D_S}{N-1} = \frac{1}{N-1}\sum_{i=0}^{N-2} D_{p,i} = \overline{D_{N-1}} \qquad (13)$$

From (13), we can conclude that, to minimize the variance of distortion for all previously encoded frames and the currently encoded frame, the distortion of current frame should be equal to the average distortion of the previously encoded frames.

3. WINDOW-BASED RATE CONTROL SHCEME

In this section, we will introduce the window-based rate control scheme on the basis of the constraint analysis in last section. The detailed mechanism of frame level bit allocation and quantization parameter (QP) decision are described respectively.

3.1. R-Q model and D-Q model

R-Q model represents the relationship between the QP and the generated bit rate, which is widely used in rate control. Many theoretical R-Q models are proposed in the literature based on the assumption that the quantized DCT coefficients are of a Laplacian distribution [10] [11]. These R-Q models are accurate but with a high computational complexity. For the real-time encoding purpose, we use a simple yet effective linear R-Q model. By the observation that the frames with high complexity cost more bits and vice versa, a linear relationship between the QP, bit rate and frame complexity is introduced as shown in (14)

$$R = \alpha_t \cdot \frac{SAD}{Q_{step}} + \beta_t , \quad t = \text{I, P, B} \qquad (14)$$

where *SAD* is the sum of absolute difference at frame level after prediction, which represents the encoding complexity of the whole frame, while Q_{step} denotes the quantization step size, α_t and β_t are model parameters related to the frame type. A similar R-Q model is also used in [12].

The frame distortion is measured by mean square error (MSE) in our study. The relationship between PSNR and QP has been proved in [13] which is

$$\text{PSNR} = 10\log(\frac{255^2}{\text{MSE}}) = l \times QP + b \qquad (15)$$

where *l* and *b* are constants. By substituting $Q_{step}=2^{(QP-4)/6}$ into (15), we can get the approximate D-Q model as follows

$$D = MSE = k \cdot Q_{step} + t \qquad (16)$$

where *k* and *t* are model parameters. Some sequences are also tested for this linear relationship as shown in Fig. 1.

3.2. Window for bit rate and low delay

In Section 2, we conclude that the total bits for any consecutive *L* frames should be constant to minimize the encoder buffer delay. For this purpose, a sliding window, so called window-R, is introduced to allocate the current frame bit. The window consists of consecutive *L* frames of *L*-1 previously encoded frames plus the current frame. The current frame is the last one in window-R. Let W_R be the total bits of window-R, then W_R can be obtained by (7). The bit budget for the current frame is calculated as follows

$$R_T = W_R - \sum_{i=0}^{L-2} R_{r,i} \qquad (17)$$

where $R_{r,i}$ represents the real bits for previously encoded frames in window-R. This window is sliding frame by frame to allocate the bits for each frame with a fixed size of *L*. By this mechanism, the low buffer latency is guaranteed since the bits for any consecutive *L* frames are restricted as W_R. Besides, different from the window-level (not a sliding window but a "jumping" window) bit allocation in work [6], our method does not introduce extra delay because no pre-analysis is performed on the frames of entire window.
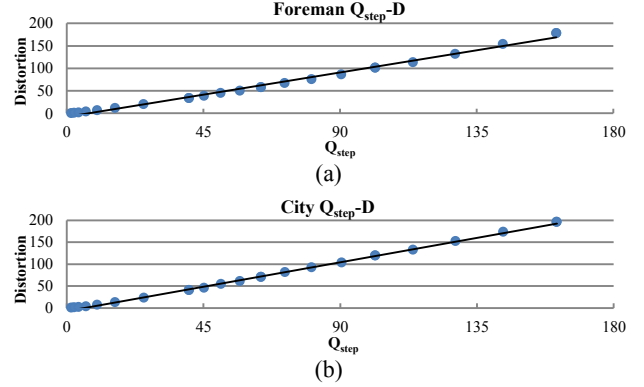

Fig. 1. Relationship between Q_{step} and distortion. (a) "Foreman", CIF format. (b) "City", 720P format.

3.3. Window for consistent video quality

Using the above-mentioned window, the bit rate and buffer latency constraints can be met. However, the consistent video quality is not well considered in this window. If the current frame has high complexity, especially when scene change or high motion occurs, the allocated bits may not be enough for encoding the current frame to obtain the similar distortion as previously encoded frames. To prevent this situation arising, the complexity of future frames should be collected. Hence, another window, so called window-D, consisting of consecutive *M* frames is also introduced. The first frame of window-D is the current frame, while others are future frames to be encoded. Pre-analysis is used to get the complexity (SAD at frame level after prediction) of each frame in window-D. For inter frames, only 16x16 motion search is used, while for intra frames, only few types of prediction (such as horizontal, vertical and diagonal) are used. The computational complexity of the pre-analysis is much lower compared to the entire encoding process.

Window-D is also running frame by frame, together with window-R. Considering the total bit budget for window-D, denoted as W_D, we can see that at time instance *t*, window-R consists of *L*-1 already encoded frames and the current frame, while window-D consists of the current frame and future *M*-1 frames. After *M* frames time (at time instance *t+M*), all *M* frames in window-D at time *t* are already contained in window-R with the window-R sliding *M* times, meanwhile the first *M* frames in window-R at time *t* are excluded from it, which means that the bit budget for window-D at time *t* should be equal to the total bits of the first *M* frames in window-R also at time *t*. Then W_D can be derived as follows,

$$W_D = \sum_{i=0}^{M-1} R_{r,i} \qquad (18)$$

With W_D and SAD of each frame in window-D, using R-Q model (14), the following equation can be derived,

$$W_D = \sum_{i=0}^{M-1} (\alpha_t \cdot \frac{SAD_i}{Q_{step,i}} + \beta_t) , \quad t = \text{I, P, B} \qquad (19)$$

where SAD_i and $Q_{step,i}$ denote frame SAD and quantization step size for ith frame in window-D, respectively. From (13), we can derive that, to minimize the variance of the distortion, $Q_{step,i}$ should be equal to each other, denoted as $Q_{step,D}$. Then $Q_{step,D}$ can be calculated as follows,

$$Q_{step,D} = \frac{\sum_{i=0}^{M-1} \alpha_t \cdot SAD_i}{W_D - \sum_{i=0}^{M-1} \beta_t} \quad , \quad t = \text{I, P, B} \qquad (20)$$

It should be noted that the $Q_{step,D}$ from (20) is only used for encoding the current frame. The $Q_{step,D}$ for the next frame will be recalculated by (20) when the window-D is sliding to the next frame. This window-based bit allocation can guarantee the consistent video quality. The pre-analysis will introduce delays of about M frames. Since M is usually small such as 4 or 5. This delay is tolerable for real-time encoding applications.

3.4. QP decision

From window-R and (13), the distortion of the current frame should be the average distortion of the previously encoded frames in window-R. Then using (16) the corresponding quantization step can be obtained, denoted as $Q_{step,R}$. From window-D, the quantization step of the current frame should be calculated by (20), denoted as $Q_{step,D}$. We can use the average of $Q_{step,R}$ and $Q_{step,D}$, represented by $Q_{step,S}$, to smooth the distortion. Considering the bit rate and buffer latency constraints, the bit allocation of the current frame R_T can be derived from (17). Then using (14) we can get the quantization step $Q_{step,T}$ from R_T. After that, to balance the buffer latency and the consistent video quality, final quantization step of the current frame can be obtained as follows

$$Q_{step} = \lambda \cdot Q_{step,R} + (1 - \lambda) \cdot Q_{step,S} \qquad (21)$$

where λ denotes a weighting factor, which is set to 0.5 in our study. The larger λ makes a lower buffer delay, while the smaller one makes more consistent video quality.

4. EXPERIMENTAL RESULTS

Our proposed rate control scheme is implemented on the JM18.5 of H.264/AVC. Certain video sequences are tested using the configuration as follow: IPPP coding structure, 2 reference frames, RDO on and CABAC, 30f/s frame rate. The size of window-R is 30 and the size of window-D is 5. We test several video sequences including CIF and 720P formats. The bit rate accuracy and the variance of distortion are shown in Table I. From Table I, we can see that our proposed rate control scheme is better in both bit rate accuracy and variance of distortion compared to the earlier works. The better bit rate accuracy gains from the window-based bit allocation, while the variance of distortion gains from the consistent distortion control in window-D.

Table I. Bit rate accuracy and variance of distortion

Sequence	BR	RC in JM		ρ-domain RC		proposed	
		Error	var D	Error	var D	Error	var D
Foreman	1000	0.19	14.09	0.23	17.21	0.13	7.57
	500	0.20	75.85	0.25	87.53	0.15	42.37
News	1000	0.08	0.59	0.09	0.75	0.04	0.50
	500	0.06	28.46	0.10	30.24	0.04	13.18
Akiyo	1000	0.07	0.43	0.09	0.52	0.06	0.40
	500	0.09	13.11	0.12	13.57	0.07	10.47
Night	8000	0.23	29.91	0.28	35.46	0.18	16.84
	5000	0.37	20.61	0.43	22.93	0.30	15.96
Crew	8000	0.09	8.89	0.12	12.64	0.08	6.23
	5000	0.13	7.41	0.15	10.87	0.10	4.78
Harbour	8000	0.08	33.53	0.15	42.67	0.08	21.76
	5000	0.12	21.21	0.18	26.31	0.11	16.28
Average		**0.14**	**21.17**	**0.18**	**25.06**	**0.11**	**13.03**

BR: Target bit rate (kbps). *Var* D: Variance of distortion.
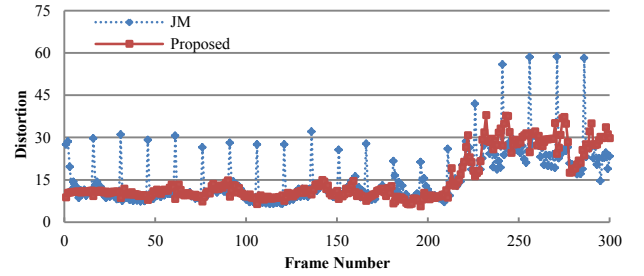Error: Bit rate mismatch between target bit rate and actual bit rate.

Fig. 2 Frame level distortion between JM and proposed RC in Foreman sequence at 500kbps.

The detailed frame level distortion of Foreman sequence at 500kbps is also shown in Fig. 2. JM uses a GOP-based bit allocation and the frame bits are allocated according to the frame type. In JM the first frame of a GOP usually has not enough bits to maintain the quality. The number of peaks in Fig. 2 also demonstrates this undesired phenomenon. In our proposed rate control scheme, the sliding window with pre-analysis can address this fluctuation and maintain the video distortion much smoother.

5. CONCLUSION

In this paper, we first analyze the constraints of buffer latency and consistent video quality. Then a window-based rate control scheme is proposed with one window for controlling the rate and latency, while the other window for consistent video quality. By applying low complexity frame level rate-distortion model, our proposed method shows an excellent performance in balancing the encoder buffer latency and consistent video quality.

6. ACKNOWLEDGEMENT

7. REFERENCES

[1] *ISO/IEC Video Recommendation ITU-T H.262, Generic Coding of Moving Pictures and Associated Audio Information,* MPEG2/H.262, 1995.

[2] *ISO/IEC JVT, ITU-T VCEG, ISO/IEC 14496-10 or ITU-T Rec. H.264, Doc. JVT-B118r8*, MPEG4/H.264, Feb. 1, 2002.

[3] *Information Technology Advanced Coding of Audio and Video Part2: Video. Standard Draft, (AVS)*, AVS-P2, Mar. 2005.

[4] *ISO/IEC, Coding of Moving Pictures and Associated Audio, Test model 5, (MPEG2) MPEG*, JTC1/SC29/WG11, 1994.

[5] B. Xie and W. Zeng, "A sequence-based rate control framework for consistent quality real-time video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 1, pp. 56–71, Jan. 2006.

[6] L. Xu, D. Zhao, X. Ji, L. Deng, S. Kwong and W. Gao, "Window-level rate control for smooth picture quality and smooth buffer occupancy," *IEEE Trans. Image Processing*, vol. 20, no. 3, pp. 723-734, Mar. 2011.

[7] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 172–185, Feb. 1999.

[8] Z. He, Y. K. Kim, and S. K. Mitra, "Low-delay rate control for DCT video coding via ρ-domain source modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 8, pp. 928–940, Aug. 2001.

[9] Z. He, W. Zeng, and C. W. Chen, "Low-pass filtering of rate-distortion functions for quality smoothing in real-time video communication," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 8, pp. 973–981, Aug. 2005.

[10] E. Lam and J. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1661–1666, Oct. 2000.

[11] X. Li, N. Oertel, A. Hutter and A. Kaup, "Laplace distribution based Lagrangian rate distortion optimization for hybrid video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 2, pp. 193–205, Feb. 2009.

[12] S. W. Ma, W. Gao, and Y. Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 12, pp. 1533–1544, Dec. 2005.

[13] K. Takagi, Y. Takishima, and Y. Nakajima, "A study on rate distortion optimization scheme for JVT coder," *in Proc. SPIE*, vol. 5150, 2003, pp. 914–923.