WINDOW-LEVEL RATE CONTROL FOR SMOOTH PICTURE QUALITY AND SMOOTH BUFFER OCCUPANCY

Long Xu^{*1},*Zhihang Wang*^{1},*Lei Deng*^{2},*Xiangyang Ji*^{1},*Debin Zhao*^{1}, *and Wen Gao*^{1,2}

^{1}Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China
^{2}Institute of Digital Media, Peking University, Beijing 1000871, China
{lxu,zhihangwang,ldeng,xyji,dbzhao,wgao}@jdl.ac.cn

ABSTRACT

Traditionally, rate control consists of bit allocation and QP decision on R-QP model. In bit allocation, target bits is further confined if buffer overflows. Meanwhile, rate control should also give as smooth as possible picture quality. However, there is no explicit relationship between picture quality and encoding parameters, so the coding result on picture quality is usually unpredictable and uncontrollable. On the condition of smooth picture quality, the smooth buffer occupancy is preferable certainly. In our work, we first proposed a "window model"formulating the size of window and variations of picture quality and buffer occupancy. Thus, given the constraint on picture quality and buffer occupancy, the compliant coding result about them can be expected employing window model. Second, a window-level rate distortion (R-D) model inspired by the traditional ρ-domain model is introduced. Lastly, the evaluation of our proposal is presented with elaborate experiments.

Index Terms— Rate Control, Smooth Picture Quality, ρ-domain, Window, Buffer Occupancy

1. INTRODUCTION

In traditional rate control schemes, such as MPEG-2 [1] and H.264/AVC [2], bit allocation is performed according to the complexity and coding type of each picture. Since we can NOT get the actual picture complexity in advance, the inefficient bit allocation is performed as scene changes or high motion object occurs. As a result, the picture quality and bits output vary much more usually. Meanwhile, it is hard to match the encoding bit rate to the target bit rate employing the traditional R-D models. Recently, the ρ-domain model is proposed in the macroblock level which can match the target bit rate more accurately than other methods. However, all these employ the traditional bit allocation technique, so the good coding performance with respect to the smooth picture quality and smooth buffer occupancy can be hardly achieved. In addition, there is no theoretical model on picture quality and buffer variations, so the coding efficiency on them is unexpectable and uncontrollable. In this paper, we introduced a "window model", which formulates the relation between the size of window and the variations of picture quality and buffer occupancy. Moreover, a new R-D model on window-level, which is derived from that of traditional ρ-domain [3] has also been proposed. Our proposed algorithm has three advantages. First, the best compromise between picture quality and buffer occupancy can be derived from the theoretical model; second, there is no need of the conventional bit allocation on frame-level, and we only need to allocate the bits quota among windows in the average bit rate; third, as smooth as possible picture quality can be obtained if only the buffer constraint allows.

The rest of this paper is organized as follows. Section 2 gives the definition of the window model. Section 3 describes the establishment of M-QP tables and window-level ρ-QP tables. The corresponding rate control algorithm on window-level is proposed in Section 4. Section 5 gives the experiments of our proposed algorithm. Section 6 concludes this paper.

2. WINDOW MODEL

As we know, the efficiency of a rate control algorithm can be depicted by the match between the target bit rate and the actual bit rate, the average PSNR improvement, the picture quality consistency, and the compliant buffer constraint. Generally, we use the match between the target bit rate and the actual bit rate to evaluate the efficiency of the R-D model. The average PSNR improvement is used to evaluate the efficiency of a rate control algorithm. But in fact, the consistent visual quality is essential as it faces the end-users. Meanwhile, the smooth buffer occupancy makes the fluent transmission and decoding process of video stream under the limited capacity of channel and decoding devices. In real applications, violating buffer constraint may cause the annoying jitter event. We integrated the picture quality and buffer occupancy into our proposed window model, where the size of a window (L) is determined by the variations of the picture quality and

---

[*] This work was partially supported by National Science Foundation (No.60803013).

the buffer occupancy which are qualified by their variances ($\sigma_{QP}^2, \sigma_R^2$). Suppose the QP and bits output (corresponding to buffer occupancy) variations provided by encoding requirements are ΔQ and ΔR respectively. In this paper, we try to find the relationship between L, ΔQ and ΔR from the theory of *error analysis method.*

First, suppose QP and bits output variations are two Gaussian random variables $\xi(\omega)\,(N(\mu_\xi, \sigma_\xi^2))$ and $\eta(\omega)\,(N(\mu_\eta, \sigma_\eta^2))$. According to the *central limit theory*, i.e.,

$$P(|\frac{1}{n}\sum_{k=1}^{n}\xi_k| < \Delta Q) = P(|\frac{\sum_{k=1}^{n}\xi_k}{\delta\sqrt{n}}| < \Delta Q(\frac{\sqrt{n}}{\delta})) = \Phi(\Delta Q\frac{\sqrt{n}}{\delta})) - \Phi(-\Delta Q\frac{\sqrt{n}}{\delta})) \quad (1)$$

Given certain probability p, we can get $\Delta Q\frac{\sqrt{n}}{\delta} = b(p)$ by looking up normal table $N(0,1)$. Then, the relationship between the size of sample n (n is replaced by L in following content) and sample dispersion ΔQ can be derived as

$$L = \alpha\frac{\sigma_\eta^2}{\Delta Q^2}(\alpha = b(p)). \quad (2)$$

Secondly, according to the classical LOG R-D model

$$R = \log(\frac{2e}{\lambda Q}) + C,$$

we can get the relationship between rate variation and QP variation

$$dR = -\frac{\beta}{Q}dQ. \quad (3)$$

Obviously, the inverse increase between R and Q can be observed from (3). Moreover, the decrease slope of R will slow down gradually as Q increases. Finally, the window model can be deduced from (2) and (3) as

$$L = \alpha\frac{\sigma^2}{\Delta Q^2}\exp(\min\{0, \Delta R - \frac{\beta}{Q}\Delta Q\}), \quad (4)$$

where $\exp(\min\{0, \Delta R - \frac{\beta}{Q}\Delta Q\})$ means L which is deduced from (2) given maximum QP variation ΔQ, will hold when it is satisfied with buffer constraint ΔR. However, it should be scaled by $\exp(\Delta R - \frac{\beta}{Q}\Delta Q)$ if it violates the buffer constraint ΔR. And, α and β are model parameters which are manually initialized and then adaptively updated for each window.

Employing the window model building on the tri-parameters (L, ΔQ, ΔR), L is firstly calculated from the predefined requirements ΔQ and ΔR before encoding a window. Then, the encoding process on the window is implemented. Thus, the encoding results subjected to the predefined requirements can be expected using the proposed window model.

3. WINDOW-LEVEL R-D MODEL

In [3], a novel R-D model on ρ-domain is proposed, where the linear relation between ρ and the bit rate is built as

$$R = \theta(1-\rho) \quad (5)$$

where ρ denotes the number of zero coefficients, R is the target bits allocated for the current frame and θ is the model parameter which is updated adaptively for each MB. After the first MB is coded, θ is updated with the actual bits and the number of zero coefficients of the first MB. Meanwhile, R is updated by subtracting the actual bits of the first MB.

3.1. M-QP table

In ρ-domain rate control, the transformed coefficients should be quantized from minimum QP to maximum QP to establish the ρ-QP table for the QP determination. Such processing is usually of high computing complexity. Fortunately, we can establish the M-QP table beforehand, where each possible magnitude of DCT coefficients is mapped into one QP. Thus, in the stage of building ρ-QP table, which QP quantize a coefficient into zero can be easily looked up M-QP table.

Currently, the popular transform-based coding standards employ integer DCT. Considering the 4×4 DCT of H.264/AVC [6], there are 3 different scale values in the scale matrix, thus we should have 3 M-QP tables.

For 4-pixel DCT, there are 3 parameters in 4×4 DCT matrix, which are denoted as a, b and c respectively. Like the processing of [6], the 4×4 float DCT matrix can be factorized to a diagonal matrix multiplied by an integer DCT matrix as

$$C_f = \begin{bmatrix} a & a & a & a \\ b & c & -c & -b \\ a & -a & -a & a \\ c & -b & b & -c \end{bmatrix} = \begin{bmatrix} x & & & \\ & y & & \\ & & x & \\ & & & y \end{bmatrix} \times C_z \triangleq D \times C_z \quad (6)$$

D is with the ith diagonal element $D(i,i)$ (x, y and z) such that $\|D(i,i)C_z(i)\| = 1$, where $C_z(i)$ is the i-th row of C_z. And C_f and C_z are float DCT and integer DCT matrix respectively. Then, the scaling matrix of 2-D DCT on image can be deduced to

$$Y = C_f X C_f^T = (DC_z)X(DC_z^T) = C_z X C_z^T \otimes E = C_z X C_z^T \otimes \begin{bmatrix} x^2 & xy & x^2 & xy \\ xy & y^2 & xy & y^2 \\ x^2 & xy & x^2 & xy \\ xy & y^2 & xy & y^2 \end{bmatrix} \quad (7)$$

where $\otimes$ indicates that each element of $C_f X C_f^T$ is multiplied by the scaling factor in the same position in matrix E. E is the scaling matrix which contains 3 different values $\{x^2, xy, y^2\}$, so 3 M-QP tables for H.264/AVC DCT should be provided.

3.2. Frame-level ρ-QP table

Employing ρ-domain model, we first build ρ-QP table representing the relation between the number of zero coefficients and QP before the actual encoding stage. So the pre-analysis is necessary. In [3], only the 16×16 inter mode is used for the P frame in the pre-analysis stage. In fact, the ρ-domain model also can be employed on I and B frames in our analysis. In the traditional ρ-domain rate control scheme, the ρ-QP table for each frame is built on the sum of ρ-QP tables of all macroblocks in the current frame.

3.3. Window-level ρ-QP table

After ρ-QP table of each frame in a window is built, the window-level ρ-QP table is the addition of the frame level ρ-QP tables in the current window.

4. RATE CONTROL ALGORITHM ON WINDOW LEVEL

Considering our proposed window level R-D model and the window model, the algorithm of window level rate control algorithm is detailed as

Step 1. Build M-QP table for each possible magnitude of DCT coefficients;

Step 2. *if (is the first window)* Given window size L; *else* Updating α and β on the bits output and QP of each frame of the previously coded window, and computing L from (4);

Step 3. Pre-analysis of only 16x16 inter prediction for P and B frames and 16x16 intra prediction for I frames in given window;

Step 4. Build frame-level ρ-QP table for each frame and window-level ρ-QP table for current window;

Step 5. Computing QP according to (5) based on window-level ρ-QP table;

Step 6. Updating parameter θ, ρ-QP table and remaining bits on the actual coded bits and QP;

Step 7. *if (is last frame of current window)* go to Step 8; *else* go to Step 5.

Step 8. *if (is last window)* procedure ends, *else* go to Step 2.

5. EXPERIMENTS AND DISCUSSIONS

To evaluate our proposed algorithm, the original [2] and our proposed window-level rate control algorithms on JM11.0 are performed. From window model (4), the larger the window size, the smoother the picture quality can be obtained, which is illustrated in Fig.1 on sequence “Football(352x288)”. Considering large window, there will be large encoder delay due to the pre-analysis processing which is performed on a window and the ρ-QP table is established thereafter. For the real-time coding, we should keep the window size as small as possible. Anyway, the size of window can be qualified in the actual applications with special buffer and picture quality constraint using our proposed algorithm.

To show the efficiency of our proposed window-level R-D model, the experience on standard sequences “Autumn(720x576, 30Hz)” and “Crowds(720x576, 30Hz)” with 300 and 500 frames respectively are performed. Here, the fixed window with size 48 (1.5s) is used. And, the buffer size of original algorithm is infinite for eliminating the influence of buffer constraint in bit allocation. Other encoding parameters are as follows: *Profile=100(High), Level=40, NumberReferenceFrames=2, SearchRange=16*. The QP variation of original and our proposed algorithms are curved in Fig.2-3, where the curve denoted by “Original RC” is for original rate control algorithm, and the curve denoted by “Window” is for our proposed window-level algorithm. The QP variation represented by the standard deviation (σ) is also tabulated in Table.1, where σ_{org} and σ_{window} are for original and our proposed algorithms respectively. From Fig.2-3, the bigger QP variation can be observed employing original algorithm. In Fig.4, the buffer curves represented by the accumulated bits of a window of successive frames are shown for both rate control algorithms, where more bits fluctuation of our proposed algorithm can be observed. The results of two rate controls are also tabulated in Table.2, where mismatch of bit rate ΔR, PSNR improvement ΔPSNR of our proposed algorithm to original algorithm are provided. From Table.2, the average PSNR improvement up to 0.2-0.5dB can be achieved for our proposed algorithm.

6. CONCLUSIONS

In this paper, we first propose a window model, and then the corresponding novel rate control algorithm on it is provided. In the proposed rate control algorithm, the picture quality variation and buffer variation which are two key factors in rate control are controllable both in theory and in practice. Meanwhile, employing window-level rate control, the bit allocation in the conventional rate control algorithms is unnecessary.

Table 2: Results of rate control for "Crowds"

Target bit rate (kbps)	Original RC			Window RC			Δ PSNR (dB)
	Bit rate (kbps)	PSNR (dB)	Mismatch (%)	Bit rate (kbps)	PSNR (dB)	Mismatch (%)	
17000	17052	42.27	0.31	16977	42.59	-0.14	0.32
6100	6137	35.83	0.62	6092	35.97	-0.12	0.14
1620	1636	30.61	1.02	1619	30.63	-0.04	0.02
420	423	25.98	0.88	419	25.93	0.00	-0.05

Table 1: QP variance of original and our proposed algorithms

Target bit rate(Mbps)	Autumn_colours		Crowds	
	σ_{org}	σ_{window}	σ_{org}	σ_{window}
22	3.91	3.56	1.44	0.78
8.7	3.92	3.59	1.68	0.90
1.8	3.88	3.53	1.41	1.19

7. REFERENCES

[1] ISO/IEC JTC1/SC29/WG11, Coding of Moving Pictures and Associated Audio, "Test model 5", MPEG 1994.

[2] S.W. Ma, Z.G. Li, F. Wu, "Proposed Draft Adaptive Rate Control", Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, Doc. JVT-H017r3, 8th Meeting,Geneva, 20-26 May, 2003.

[3] Z. He and S. K. Mitra, "A linear source model and a unified rate control algorithm for dct video coding," IEEE Trans. on CSVT, vol 12, pp. 970-982, Nov. 2002.

[4] Information technology Advanced coding of audio and video Part2: Video. AVS-P2 Standard draft, Mar 2005.

[5] P.H.Westerink, R. Rajagopalan, etc., "Two-pass MPEG-2 Variable-bit-rate encoding," IBM Journal of Research and Development, vol.43, no.4, July 1999.

[6] http://www.vcodex.com/files/h264_transform.pdf.

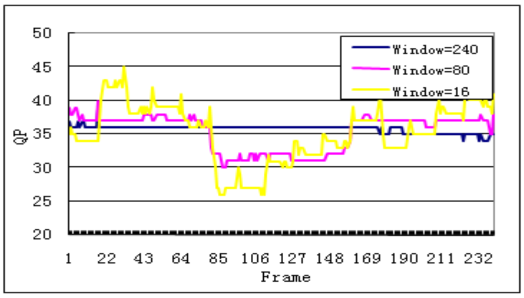

Fig. 1: QP variation of "Football" (window size: 240, 80, 16)

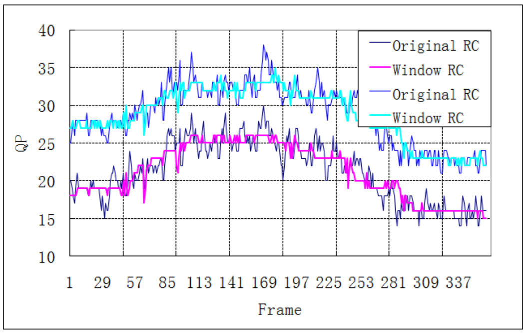

Fig. 2: QP variation of "Autumn"(bit rate: 22Mbps, 8.7Mbps)

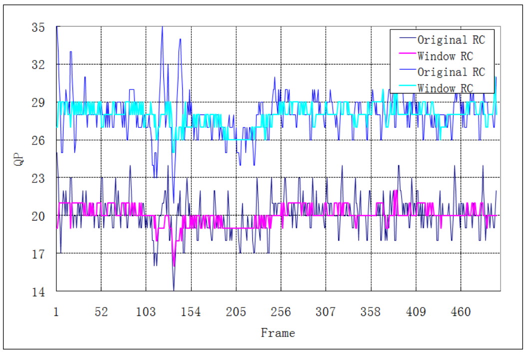

Fig. 3: QP variation of "Crowds" (bit rate: 17Mbps, 6Mbps)

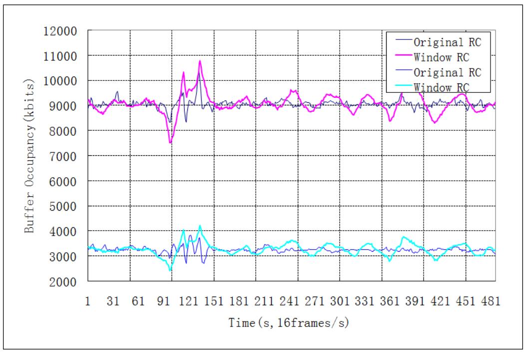

Fig. 4: Buffer status for "Crowds" (bit rate: 17Mbps, 6Mbps)