RATE CONTROL FOR JVT VIDEO CODING SCHEME WITH HRD CONSIDERATIONS

Siwei Ma[1], Wen Gao[1], Feng Wu[2], and Yan Lu[3]
[1]Institute of Computing Technology, Chinese Academy of Science, Beijing, 100080, China
[2]Microsoft Research Asia, Beijing, 100080, China
[3]Department of Computer Science, Harbin Institute of Technology, Harbin, 150001, China
{swma, wgao}@ict.ac.cn, ylu@ieee.org

ABSTRACT

The recent JVT video coding scheme (MPEG-4 AVC/H.264) is a promising technique due to its high coding efficiency. Hypothetical Reference Decoder (HRD) is a very important part in JVT video coding, which represents a set of normative requirements on bitstream for the purpose of avoiding buffer overflow and underflow. The problem of HRD requirements can be solved by rate control. This paper proposes an effective rate control scheme for JVT video coding with HRD considerations. First, bit allocation with HRD constraints is presented, and second, based on a simple rate distortion model, a single pass rate control is implemented on both frame level and macroblock level. Experimental results show that the proposed rate control algorithm can achieve the target bit rate with very little bit rate or image quality fluctuation, and meanwhile it can well meet the HRD requirements. Furthermore, the proposed algorithm is so simple that it only introduces little computation complexity. Therefore, it can be used in real time video coding.

1. INTRODUCTION

The JVT video coding, also known as MPEG-4 Advanced Video Coding (AVC) in ISO/IEC and H.264 Recommendation in ITU-T, is jointly developed by ISO/IEC and ITU-T. The JVT video coding can achieve much higher coding efficiency than any other existing video coding standards. In JVT video coding, the Hypothetical Reference Decoder (HRD) is a normative part, which represents a set of normative requirements on bitstreams [1]. These constraints can be enforced by an encoder through rate control.

HRD is conceptually connected to the output of an encoder and consists of a decoder buffer, a decoder and a display unit. A mathematical model, also known as leaky bucket, is usually employed to characterize the hypothetical decoder and its input buffer called coded picture buffers (CPB). A constrained arrival time leaky bucket (CAT-LB) model has been defined in JVT FCD document [1]. Bits flow in the decoder buffer at a constant rate, and are removed from the decoder buffer in chunks. A HRD compliant bitstream must be decoded in the CPB without overflow and underflow.

The problem of HRD requirements can be solved by rate control implemented in the encoder. A rate control algorithm may target at a variable bit rate (VBR) or a constant bit rate (CBR). In our previous researches, a macroblock-based rate control scheme by considering both bit allocation and Rate-Distortion Optimization (RDO) has been proposed for JVT video coding. However, the constraints requested by HRD have not been considered. Therefore, this paper proposes an improved rate control scheme for JVT video coding with HRD or CAT_LB constraints. First, in the bit allocation process of rate control, the target bit for each picture is clipped with an upper and lower bound required by HRD to guarantee the CPB neithor overflow nor underflow. Since the rate distortion model is usually not accurate enough, further limit to the target bits is necessary. And second, after bit allocation, rate control must maintain the target bit by adjusting the quantization parameter at frame or macroblock level adaptively. Sometimes frame may be skipped if the buffer occupancy is too high.

The rest of the paper is organized as follows. Section 2 describes the CAT_LB model requirements for rate control. In Section 3, the strategy of the proposed rate control algorithm is described in detail. The experimental results are discussed in Section 4. And finally, Section 5 concludes this paper.

2. HYPOTHETICAL REFERENCE DECODER

The section outlines the operation of CAT_LB model in JVT video coding. More detailed description can be found in [1]. In Figure 1, R is the bit rate at which the CPB is filled. $b(n)$ is the size in bits of picture n. $t_{ai}(n)$ is the time when the first bit of picture n enters the CPB, called the initial arrival time of picture n. $t_{af}(n)$ is the time when the last bit of picture n enters the CPB, called the final arrival time of picture n. $t_r(n)$ is the time when the picture n is removed from the CPB, called the removal time. $t_{ai}(n)$ and $t_{af}(n)$ can be computed by:

$$t_{ai}(n) = \begin{cases} 0 & n = 0, the\ first\ picture \\ t_{af}(n-1) & t_{af}(n-1) \geq t_{ai,earliest}(n) \\ t_{ai,earliest}(n) & t_{af}(n-1) < t_{ai,earliest}(n) \end{cases}, \quad (1)$$

where $t_{af}(n) = t_{ai}(n) + b(n)/R$. The removal time $t_r(n)$ of picture n can be computed by:

$$t_r(n) = \begin{cases} initial_cpb_removal_delay/9000 & n = 0 \\ t_r(0) + t_c \times cpb_removal_delay(n) & n \neq 0 \end{cases}, \quad (2)$$

where

$$t_{ai,earliest}(n) = t_c \times cpb_removal_delay(n) - t_0 = t_r(n) - t_r(0) - t_0.$$

t_0 is *initial_cpb_removal_delay_offset*/9000;

initial_cpb_removal_delay_offset, *cpb_removal_delay*(n), *initial_cpb_removal_delay* and t_c have been defined in [1].

0-7803-7750-8/03/$17.00 ©2003 IEEE.
ICIP 2003
TCL-1035
TTE Technology Inc. d/b/a TCL North America v. InterDigital, Inc. IPR2026-00417
AMAZON-1035
Amazon v. InterDigital, IPR2026-00195

To ensure the CPB does not overflow and underflow, the bits allocated to picture n by rate control must not surpass an upper bound U_n. For CBR a lower bound L_n must be set to ensure the bit stream enters into the CPB continuously. That is to say the following equation must be true:

$$t_{af}(n) \geq t_{ai,earliest}(n+1). \quad (3)$$

From (1) and (3), we have

$$b(n) \geq (t_{ai,earliest}(n+1) - t_{ai}(n))R. \quad (4)$$

So the lower bound is

$$L_n = \max((t_{ai,ealiest}(n+1) - t_{ai}(n))R, 0). \quad (5)$$

According to the rate control considerations in [1], it is stated that the CPB should neithor overflow nor underflow if the following equation remains true:

$$te[b(n)] \leq t_r(n) - t_{ai}(n), \quad (6)$$

where $be[t]$ and $te[b]$ denote the bit equivalent of a time t and the time equivalent of a number of bits b respectively, and

$$te[b] = b/R, be[t] = tR. \quad (7)$$

From (6) and (7), we have

$$b(n) = be[te[b(n)]] \leq be[t_r(n) - t_{ai}(n)]. \quad (8)$$

So the upper bound U_n for the picture n is

$$U_n = (t_r(n) - t_{ai}(n))R. \quad (9)$$

Since sometimes the rate control is not very accurate, further limits need to be done on U_n or L_n through the bit allocation.

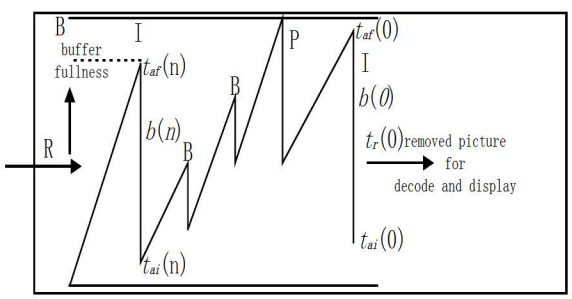

Figure 1: A Hypothetical Reference Decoder (Coded picture buffer occupancy in HRD model)

3. RATE CONTROL ALGORITHM

One of the major tasks of rate control is to allocate a target bit for each picture with minimum distortion under bit constraint. Rate control must maintain the target bit rate through tuning the quantization parameter of picture or macroblock. In addition, rate control must guarantee the CPB neithor overflow nor underflow. In the past, the relationship between the rate and distortion of video coding has been fully studied. In [4], a source model for transform video coder is derived from rate-distortion theory and is used to variable frame rate coding. In [5], an efficient quadratic rate distortion model is proposed and has been widely used in the rate control scheme of MPEG-4 verification model. However, both [4] and [5] need to compute the complexity parameters.

In this paper, a simplified rate quantization model from [6][7] is employed, i.e,

$$R^t(Q^t) = X^t / Q^t \qquad t = I, P, B. \quad (10)$$

where t is the picture type, and X^t is a complexity measure for each picture type. According to (10), the quantization parameter Q^t for the picture can be calculated from an estimated complexity measure X^t and the bits R^t of a picture. In the practical applications, the complexity measure of the current picture is usually replaced with that of the previously coded picture. In detail, the proposed rate control includes three steps.

Step 1. Bit allocation.

In this step, a target bit is allocated to each picture. In CBR, typically a group of picture (GOP) is given a constant bits constraint. Bit allocation for a picture in a GOP can be resolved through a Lagrangian multiplier method [7][8]. In TM5, the target bit for a picture is computed by:

$$T^t = X^t G / \Sigma_t X^t N^t \quad (11)$$

G is the available bits to the GOP decided by the target rate, frame rate and GOP size. t is the picture type. N^t is the number of pictures with the picture type of t.

Let T_n^t denotes the target bits for the picture n. Before coding picture n, the X_{n-1}^t is used to replace X_n^t. T_n^t can be adjusted adaptively in coding process as follows:

$$T_n^t = (1-\alpha) X_{n-1}^t G / \Sigma_t X_{n-1}^t N^t + \alpha T_{prev} \quad (12)$$

where T_{prev} is the coded bits of previously coded picture. To guarantee the lowest image quality, T_n^t is usually bounded with:

$$T_n^t = \max(R/8F, (1-\lambda) X_{n-1}^t G / \Sigma_t X_{n-1}^t N^t + \lambda T_{prev}) \quad (13)$$

where F is the frame rate. Therefore, the target bit T for picture n must be clipped with:

$$T = \begin{cases} \beta U_n & T_n^t > \beta U_n, \beta \leq 1 \\ L_n & T_n^t < L_n \\ T_n^t & other \end{cases} \quad (14)$$

In our experiments, β =0.9.

Step 2. Computing the quantization parameter for the current frame.

According to (14), the quantization parameter Q_n^t for picture n:

$$Q_n^t = X_n^t / R_n = \lambda^t X_{n-1}^t / T \quad (15)$$

λ^t can be approximated from previous coded picture, but for simplicity, $\lambda^I = 0.9$, and $\lambda^P = \lambda^B = 1$ are used in our simulating system. To avoid quality fluctuation, Q_n^t is usually limited by:

$$Q_n^t = \min(\max(Q_n^t, Q_{n-1}^t - 2, Q_{prev} - 4), Q_{n-1}^t + 2, Q_{prev} + 4) \quad (16)$$

where Q_{prev} is the quantization parameter of previous coded picture.

For the first picture, there are many empirical parameters in (11). In TM5, an initial X^t is used for the bit allocation for the first t type picture. In [5], a fixed quantization parameter is used for the first I picture. These empirical would affect the quality of first few frames or GOPs. In [5], it is reported that the average performance variation may have 1dB. In AVC standard, the quantization parameter may range from 0 to 51. It is not reasonable to use a fix quantization parameter for the first frame for all video sequences. Here, for a given bit rate an estimated quantization parameter is decided for the first picture by bits per pixel (*bpp*).

$$bpp = R / FN$$

$$Q_0^I = \begin{cases} 40 & bpp \le l1 \\ 30 & l1 < bpp \le l2 \\ 20 & l2 < bpp \le l3 \\ 10 & bpp > l3 \end{cases} \tag{17}$$

$$Q_0^P = Q_0^B = Q_0^I + 2$$

N is the number of pixels in a picture. In our experiments $L1$=0.1, $L2$=0.6, $L3$=1.5.

After coding a picture, the coded picture buffer and picture complexity measure are updated. If the buffer occupancy is over a threshold the next frame may be skipped.

Step 3. Computing the macroblock quantization parameter.

To enforce the CBP constraints, the quantization parameter of a macroblock can be adjusted adaptively. In TM5, the quantization parameter of macroblock m is decided by the virtual buffer occupancy. The virtual buffer model is used here, but the initial buffer occupancy is decided by the quantization parameter computed by frame level rate control. The quantization parameter $Q^t_{n,m}$ for macroblock n is decided by:

$$\begin{aligned} Q^t_{n,m} &= \min(\max(d^t_{n,m} / r, Q^t_n), Q^t_n + 6, 51); \\ d^t_{n,m} &= d^t_{n,m-1} + b^t_m - T / MB_CNT \\ d^t_{n,0} &= d^t_{n-1,MB_CNT} \end{aligned} \tag{18}$$

$d^t_{n,m}$ is the virtual buffer occupancy before coding macroblock m. r is the virtual buffer size, and the virtual buffer is initialized with:

$$d^t_{0,0} = rQ^t_0 \tag{19}$$

4. EXPERIMENTS AND RESULTS

In order to evaluate the performance of the proposed algorithm, this section presents the experimental results on typical test sequences. AVC with the proposed rate control and without rate control is tested, respectively. JM5.0 developed by JVT serves as the platform. Table 1 illustrates the coding results of the proposed rate control scheme. The sequence format and testing conditions are also shown in Table 1. From the table, we can see that the proposed algorithm can efficiently control the bit-rate at different resolution, frame rate. The error between target bit rate and real bit bate are very small, which usually does not exceed 1% and the quantization parameter for the first frame usually is very close to the average quantization parameter, which would improve the coding performance of whole sequence.

In Figure 2-3, the rate-distortion curves with rate control and without rate control are shown, respectively. From the curves we can see that the rate control can still keep good coding efficiency. Figure 4-5 show the PSNR per frame for the test sequences coded at the same bit rate with proposed rate control and without rate control. In figure 6-7, the buffer occupancy at macroblock rate control is shown. From the plot, after a higher initial buffer occupancy for an initial delay, the proposed rate control can maintain suitable buffer occupancy levels. It proves that the proposed rate control algorithm can prevent buffer overflow and underflow.

Table 1. The generated bit rate with the proposed frame level rate control

Sequence	Average QP	Q_0^I	Target bit-rate (kbps)	Achieved bit-rate (kbps)	PSNR (dB)
Foreman	40	40	35.0	35.11	28.27
News	28	30	44.0	43.84	36.55
Mobile	32	30	776.0	777.17	31.4

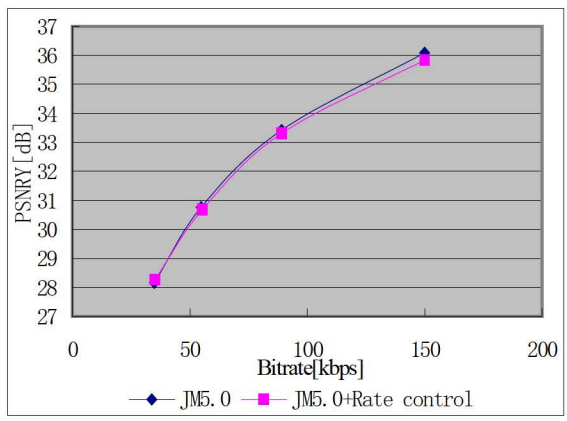

Fig 2. PSNR curve of Foreman sequence with QCIF format at 30fps

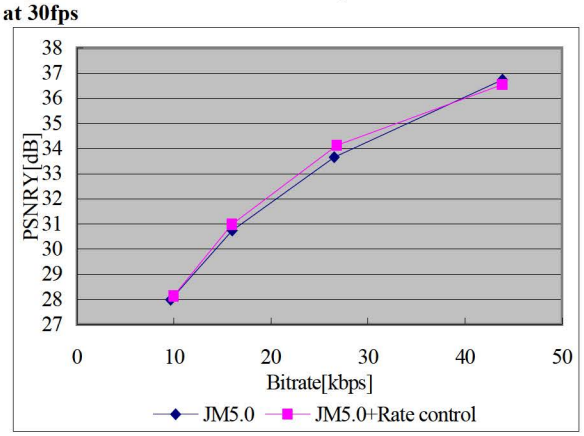

Fig 3. PSNR curve of News sequence with QCIF format at 10fps

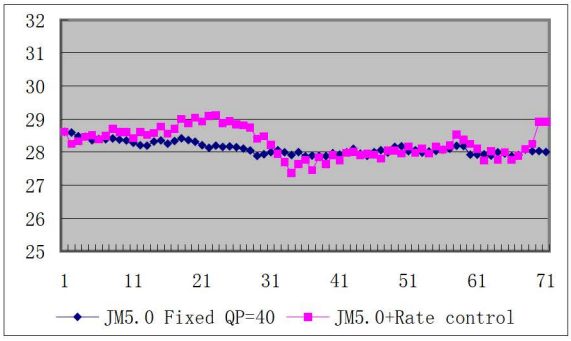

Figure 4. PSNR versus frame curve of News sequence at 10.01kbps with QCIF format, 10fps.

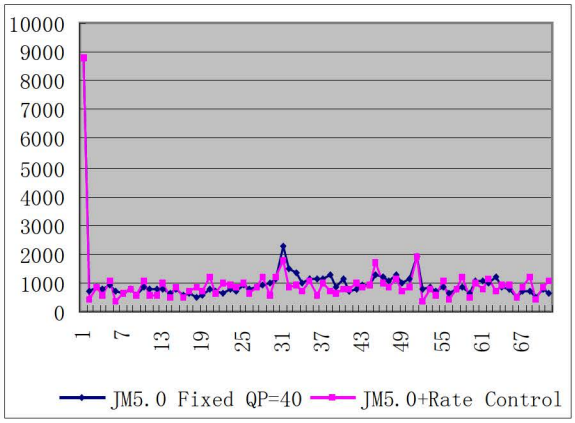

Figure 5. Bit versus frame curve of News sequence at 10.01kbps with QCIF format, 10fps.

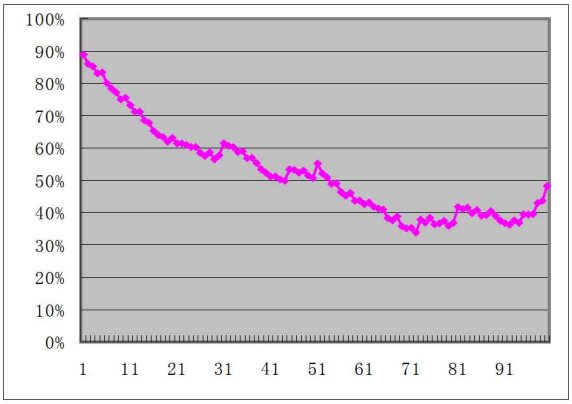

Figure 6. Buffer occupancy of News sequence at 9.96kbps with QCIF IPP format, 10fps.

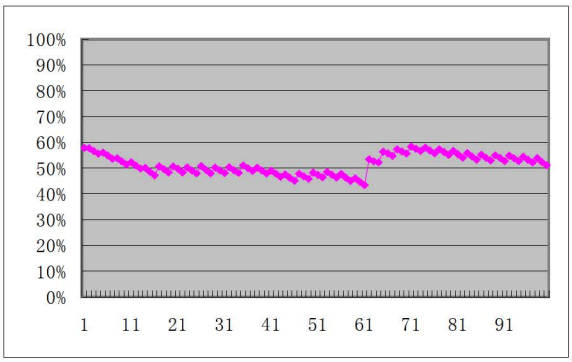

Figure 6. Buffer occupancy of Foreman sequence at 35.13kbps with QCIF IBBP format, 30fps.

5. CONCLUSIONS

This paper has proposed an efficient rate control algorithm with HRD considerations for JVT video coding. The proposed rate control can be realized at both frame level and macroblock level. The experimental results have also been presented. It has been shown that the proposed algorithm can generate the HRD compliant bit stream, and meanwhile the generated bit rate is very close to the target bit rate. The coding efficiency achieved from the proposed rate control is similar to or even better than that of the fixed quantization parameter video coding.

6. ACKNOWLEDGEMENTS

The work in part has been supported by National Hi-Tech Research Program (863) of China (2002AA119010), National Fundamental Research and Development Program (973) of China (2001CCA03300).

7. REFERENCES

[1] T. Wiegand, "Editor's Proposed Draft Text Modifications for Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC), Geneva modifications, draft 7," Document JVT-E146 Annex C, 9 Oct. 2002.

[2] S. Ma, W. Gao and Y. Lu, "Rate control on JVT standard," Document JVT_D030, Klagenfurt, Austria, 22-26 Jul. 2002

[3] S. Ma, W. Gao and Y. Lu, "Improved rate control algorithm on JVT standard," Document JVT_E069, Geneva, Switzerland, 9-17 Oct. 2002.

[4] J. Chen, H. Hang, "Source model for transform video coder and its application-part I, II". IEEE Trans. Circuits Syst. Video Technol. Vol. 7, pp.287-311, Apr. 1997.

[5] T. Chiang, Y. Zhang, "A new rate control scheme using quadratic rate distortion model," IEEE Trans. Circuits Syst. Video Technol. Vol. 7, pp.287-311, Apr. 1997.

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

[7] P. Westerink, R. Rajagopalan and C. Gonzales, "Single-pass constant- and variable-bit-rate MPEG-2 video compression," IBM J.RES. DEVELOP, Vol. 43 No.4 Jul. 1999.

[8] J. Katto and M. Ohta, "Mathematical Analysis of MPEG Compression Capability and Its Application to Rate Control," IEEE ICIP'95, Vol. II, pp. 555-559, 1995.