A PERCEPTUAL BASED RATE CONTROL SCHEME FOR MPEG-2

K.T.Choi, S.C.Chan and T.S.Ng
Department of Electrical and Electronic Engineering
The University of Hong Kong, Pokfulam Road, Hong Kong
Email: scchan@hkueee.hku.hk

ABSTRACT

In this paper, a new perceptual based rate control algorithm for MPEG-2 is presented. The algorithm first determines the target bit count for each frame using simple statistical models. Precise bit allocation is used to adjust the macroblock quantization scale factors to meet the given bit count, with the uniform visual fidelity as the primary objective. Since the buffer is very stable, it is less sensitive to transmission delay. Experimental results showed that it gave better visual quality and less buffer variations as compared to the TM5 rate control scheme.

1. INTRODUCTION

The MPEG-2 [1] is an important video coding standard for coding high quality video such as high-definition television, and high-quality digital video recording. In constant bit rate application, rate control has played an important role in improving and stabilising the video quality. The well-known TM5 rate control scheme determines the current target bit count using previous encoded frame statistics, and adjusts the quantization step size based on the buffer occupancy. Therefore, the quantization error of adjacent blocks can vary significantly and it generates visible blocking artifacts. Recently, there has been a growing interest in rate-distortion optimal techniques for both bit allocation and rate control. The rate control problem can be formulated as a constrained optimization problem and solved by the Lagrangian or minimax technique [5-7]. Ortega *et al.* [5], proposed a rate control scheme which used dynamic programming to search for the true global optimal solution. Lin *et al.* [8] proposed a faster algorithm by encoding the source with only a few quantization steps and the rate-distortion value for other quantization steps is found using interpolation. However, the source video has to be encoded several times and the complexity is still very high. In a previous work [3], we had proposed a new buffer control algorithm for motion-compensated hybrid DPCM/DCT coding. The quantization scale factors for each macroblock are determined to meet a given target bit budget. The salient features of that scheme are that (1) the quantization scale factors are determined using information of the whole picture; (2) it has precise control of the buffer; and (3) it tries to allocate the given number of bits as efficient as possible in a rate-distortion sense. However, this scheme is not directly applicable to MPEG-2 coder, since the encoding of the I, P, and B pictures are not completely independent. In fact, P and B pictures are predicted from the I pictures and so on.

In this paper, a new perceptual based rate control algorithm is presented to solve this problem. It first determines the target bit count for each frame using certain statistical model. The bit allocation algorithm proposed in [3] can then be used to determine the macroblock quantization scale factors to meet the given bit count, with the uniform visual fidelity as the primary objective. More precisely, the *frame bit allocation* determines the bit counts to be spent for each frame using an accurate statistical rate-distortion model. Then, we perform the *macroblock bit allocation*, similar to [3], to meet exactly the target bit count found in the *frame bit allocation.* Experimental results showed that the proposed scheme gave better visual quality as compared to the TM5 model at comparable PSNR values. The buffer is more stable and is less sensitive to transmission delay. The proposed algorithm is discussed in Sections 2 and 3. In Section 4, the simulation results and discussion are presented, followed by the conclusion in Section 5.

2. FRAME BIT ALLOCATION

There are several statistical approaches that can be used to deteremine the approximate number of bits required for each frame. Here, we shall make use of the method in [2]. The rate distortion function is modeled as a quadratic function between the bit rate and the average quantization scale of a picture [2]. In particular, the target bit rates T_i, T_p, T_b for the I, P, and B pictures can be found from the following equations:

$$K_p Q_i = K_i Q_p$$
$$K_b Q_i = K_i Q_b$$
$$T_i + N_p T_p + T_b T_b = R_{GOP}$$
$$T_i = a_I Q_i^{-1} + b_I Q_i^{-2}$$

$$T_p = a_2 Q_p^{-1} + b_2 Q_p^{-2}$$
$$T_b = a_3 Q_b^{-1} + b_3 Q_b^{-2} ;$$

where K_i, K_p, K_b are constants determining the relative quality for I, P, and B pictures; Q_i, Q_p, Q_b are average quantization steps for I, P, and B picture; N_p *and* N_b are the number of frames to be encoded for P and B pictures and R_{GOP} is the remaining number of bits in the current GOP.

The model parameters a_1, a_2, a_3 and b_1, b_2, b_3 can be found by statistical linear regression analysis using previous pictures. Based on the above model, we can calculate the target bit rate before encoding using Table 1. Due to the improved accuracy of the rate distortion function, the fluctuation of the bit count after encoding is reduced. The resulting visual quality becomes more uniform across the whole video sequence. In the next section, further reduction in the bit count variation can be achieved when more precise bit allocation buffer control algorithm is applied to each frame.

3. MACROBLOCK BIT ALLOCATION

After the target bit rate for each frame is determined, we can perform bit allocation [3] for the I, P, and B pictures. At this level, the bit allocation problem involves minimizing the overall distortion of all N macroblocks with respect to a set of quantization scale factor Q_n (where $n = 1,\ldots,N$):

$$D = \sum_{n=1}^{N} D_n(Q_n)$$

subject to the bit rate constraint T_{pict}:

$$R = R_{overhead} + \sum_{n=1}^{N} R_n(Q_n) \le T_{pict}$$

$D_n(Q_n)$ and $R_n(Q_n)$ denote respectively the distortion and the bit rate generated by the n-th macroblock quantizer Q_n. $R_{overhead}$ is the overhead of a picture including the number of bits generated for the picture, the slice and the macroblock headers including motion vectors. Since $R_{overhead}$ depends on the motion vectors, it is assumed that the motion vectors have been determined by full search or some other means prior to the bit allocation. The T_{pict} is one of the target bit rates T_i, T_p, *and* T_b. In order to adapt to the local image content and to exploit spatial masking effect, we weight the distortion measure by the local spatial activity factor [4]. The activity measure act_j of the macroblock j is chosen as the minimum among the four (8x8) block luminance variance σ_{sb}

$$act_j = 1 + min(\sigma_{sb})$$

The activity measure for the macroblock j is then normalized against the mean activity value *avg_act* of the previous encoded picture of the same type. The normalized macroblock local activity N_act_j is defined as

$$N\_act_j = \frac{2 * act_j + avg\_act}{act_j + 2 * avg\_act}$$

The perceptual distortion $D^*_n(Q_n)$ for the j-th macroblock is obtained by multiplying the original distortion $D_n(Q_n)$ by the spatial activity measure N_act_j as follow,

$$D^*_n(Q_n) = D_n(Q_n) \bullet N\_act_j$$

The proposed macroblock bit allocation scheme aims to vary the quantizer scale of each MB until the frame bit budget determined by the frame bit allocation scheme is reached. The algorithm is summarized as follows [3]:

1. Initialize all quantization scale factors Q_n to maximum value Q_{max}, and the bit count B.

2. Calculate functions $D^*_n(Q_n)$ and $R_n(Q_n)$ for $n = 1,\ldots,N$ and $q = 1,\ldots,Q_{max}$.

3. Calculate the efficiency of the k-th block as

$$\lambda_k = \max_q \frac{-\Delta D|_{Q_k \to q}}{\Delta R|_{Q_k \to q}}$$

where $k = 1,\ldots,N$, and $\Delta D|_{Q_k \to q}$ and $\Delta R|_{Q_k \to q}$ are the change in distortion and the change in overall bit-rate used for all macroblocks when the quantization scale factor of the k-th macroblock Q_k is replaced by q. As each quantizer is independent, these increments can be calculated as

$$\Delta D|_{Q_k \to q} = D^*_k(q) - D^*_k(Q_k)$$
$$\Delta R|_{Q_k \to q} = R_k(q) - R_k(Q_k)$$

4. Find the macroblock m with maximum λ_k :

$$\lambda_m = \max_k \lambda_k$$

This means that the ratio of decrease in distortion to increase in bit-rate is maximized over all possible reduced values for each quantization scale factor.

5. Update the quantization scale factor $Q_m = q$ and the current bit count B. Calculate new value of λ_m and insert it into the list.

6. Repeat Steps **4** and **5** while $T_{pict} > B + R_{overhead}$.

Under the *macroblock bit allocation*, the bit count assigned to each individual picture in Section 2 is exactly met. Better perceptual quality is achieved since the bit allocation is adapted to human visual perception using local spatial activity. The proposed scheme not only meets the CBR transmission constraint but also minimizes the perceptual coding distortion. Both the buffer requirement and the delay can be minimized, with better rate distortion trade off.

4. RESULTS AND DISCUSSION

Simulation is performed on composite video sequence at 30 frame/sec with the picture size of 352 x 240 pels. The composite sequence with a total of 490 frames consists of the video sequence "Table tennis", "Mobile and Calendar", "Football", and "Flower Garden". The target average bit rate is about 1.5 Mbit/sec with a buffer size of 96 msec. There are 15 pictures in each GOP with two B-frames between two anchor frames. To demonstrate the performance of the proposed scheme, we compare it with the rate control method in MPEG TM5 model.

Figure 1 showed the zoomed views of the 68th encoded frame using different algorithms are displayed. Both blocky and ringing effects show up in the TM5 case. In particular, the surface of the table and the arm of the player in the encoded image of the proposed algorithm is much smoother than that of the TM5 scheme. Also, the surrounding region of the ball and the paddle appears to be very noise in the TM5 model.

Figure 2 shows the bits generated for the I, P, and B frames, It can be seen that the bit counts is very stable in the proposed approach. The reduction in the standard deviation of the bit counts are 58%, 69% and 75% for I, P, B frames, respectively (Table 2). Also the average peak signal-to-noise ratio (Fig. 2) is improved slightly by about 0.3 dB.

5. CONCLUSIONS

In conclusion, we have proposed a effective rate control strategy for MPEG-2 coder. Simulation result shows that less bit rate fluctuation and better perceptual quality than the TM5 model is achieved.

References

[1] ISO-IEC/JTC1/SC29/WG11/602, "Generic coding of moving pictures and associated audio," MPEG Committee Draft: Recommendation H.262, Nov. 1993.

[2] T. Chiang and Y.Q. Zhang, "A new rate control scheme using quadratic rate distortion model," IEEE Trans. on Circuit and Systems for Video Tech., vol. 7, no.1, pp. 246-250, Feb. 1997.

[3] K.T. Ng, S.C. Chan, and T.S. Ng, "Buffer control algorithm for low bit-rate video compression," in Proc. of ICIP-96, Switzerland, Sept.,96.

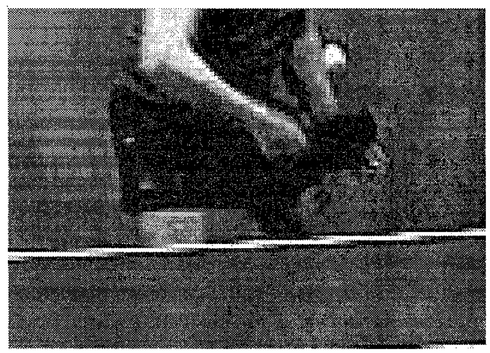
[4] ISO-IEC/JTC1/SC29/WG11, "Test Model 5," April, 1993.

[5] A.Ortega, K.Ramchandran, and M.Vetterli, "Optimal trellis-based buffered compression and fast approximation," IEEE Trans. on Image Proc., 3(1):26-40, Jan. 1994.

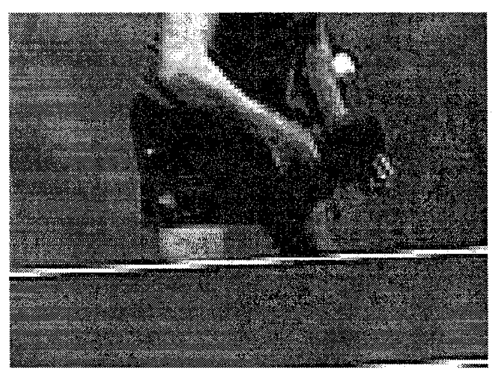
[6] K.Ramchandran, A.Ortega, and M.Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," IEEE Trans. on Image Proc., 3(5):533-545, Sept. 1994.

[7] J.Lee and B.W.Dickinson, "Joint optimization of frame type selection and bit allocation for MPEG video encoders," In Proc. of ICIP 94, volume II, pp. 962-966, Austin, Texas, 1994.

[8] L.J. Lin, A. Ortega, and C.C.J.Kuo, "A gradient-based rate control algorithm with applications to MPEG video," in Proc. ICIP'95, Washington, DC, 1995, vol. III, pp. 392-395.



(a)



(b)

Figure 1: The zoom-in of the decoded 68th frame using (a) TM5 rate control algorithm, and (b) the proposed rate control algorithm.

Frame Type	*Model Parameters*	*if* $\delta > 0$	*if* $\delta <= 0$
T_i	$\alpha = b_1 + N_p b_2 K_1 + N_b b_3 K_2$ $\beta = a_1 + N_p a_2 K_1 + N_b a_3 K_2$	$Q_i^{-1} = \frac{\sqrt{\delta} - \beta}{2\alpha}$ $T_i = a_1 Q_i^{-1} + b_1 Q_i^{-2}$	$Q_i^{-1} = -\frac{\gamma}{\beta}$ $T_i = a_1 Q_i^{-1}$
T_p	$\alpha = N_p b_2 + N_b b_3 K_3^2$ $\beta = N_p a_2 + N_b a_3 K_3$	$Q_p^{-1} = \frac{\sqrt{\delta} - \beta}{2\alpha}$ $T_p = a_2 Q_p^{-1} + b_2 Q_p^{-2}$	$Q_p^{-1} = -\frac{\gamma}{\beta}$ $T_p = a_2 Q_p^{-1}$
T_b	$\alpha = N_p b_2 / K_3^2 + N_b b_3$ $\beta = N_p a_2 / K_3 + N_b a_3$	$Q_b^{-1} = \frac{\sqrt{\delta} - \beta}{2\alpha}$ $T_b = a_3 Q_b^{-1} + b_3 Q_b^{-2}$	$Q_b^{-1} = -\frac{\gamma}{\beta}$ $T_b = a_3 Q_b^{-1}$

Table 1. Close form solution of the target bit rate for I, P, and B frame, where $K_1 = K_i / K_p$, $K_2 = K_i / K_b$, and $K_3 = K_p / K_b$, $\gamma = -R_{GOP}$, $\delta = \beta^2 - 4\alpha\gamma$.

Standard Deviation (kbits)	I	P	B
Proposed	5.558	4.9374	1.4019
TM5	13.279	16.02	5.6602
% in reduction	58.14%	69.18%	75.23%

Table 2. Standard deviation of the bit counts of the SIF sequence at 1.5 Mbps.

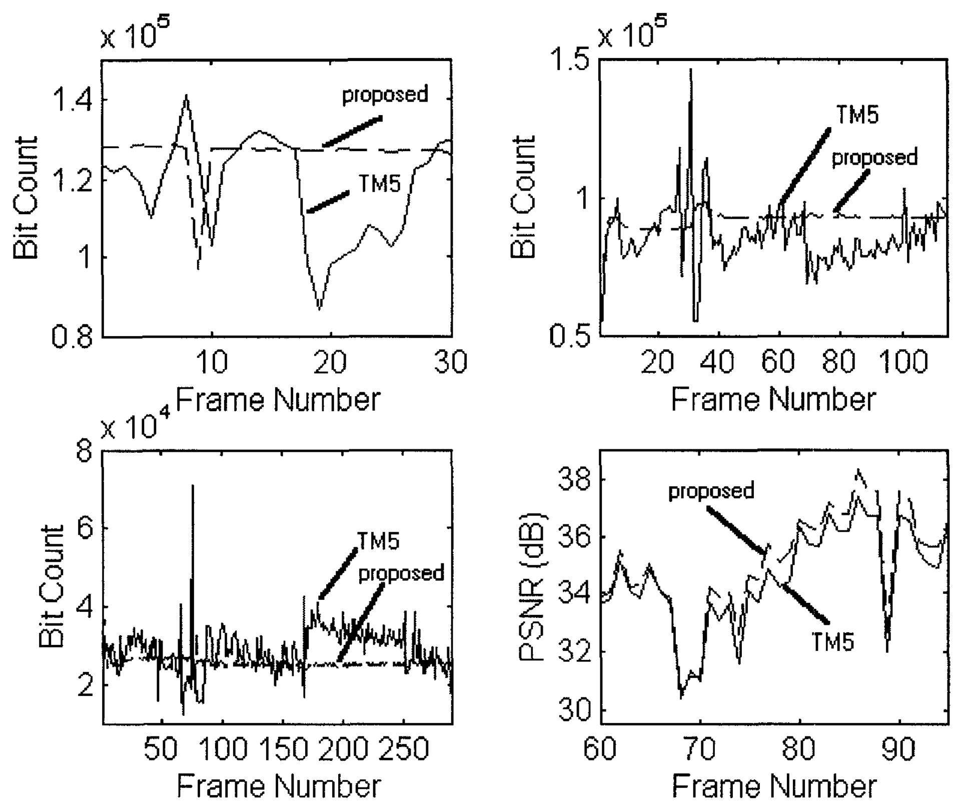


Figure 2: Upper left, upper right and lower left are the bit rate distribution plot for the I, P, and B frames of the SIF sequence at 1.5 Mbps respectively. Lower right shows the PSNR plot of the luminance for the reconstruction of the first 110 frames. The solid line for the TM5 rate control and the dotted line for the proposed algorithm.