COMPLEXITY ADAPTIVE QUANTIZATION FOR INTRA-FRAMES IN VERY LOW BIT RATE VIDEO CODING

F. Pan, Z. G. Li, K. P. Lim, X. Lin, S. Rahardja, D. J. Wu & Wu Si
Institute for Infocomm Research
21 Heng Mui Keng Terrace, Singapore 119613

ABSTRACT

Conventional rate control scheme focuses on the problem of finding an optimal quantization value for P- and B- frames. No rate control is available for the encoding of I-frames. This could pose big problems due to the large number of bits an I-frame could generate, and due to the fact that the number of bits varies drastically from sequence to sequence depending on their image complexity. This problem becomes severe especially at very low bit rate. Therefore, a mechanism to allocate data bits to an I-frame according to its complexity is indispensable in order to have constant coding quality. This paper presents a mechanism to establish for I-frames the generic relationship among the quantization value, data bits, and their image complexity. Experimental results show that this generic relationship provides a fairly accurate estimation of quantization value for an I-frame at given data bits and image complexity, and is very useful in controlling the data bits generated by an I-frame.

1. INTRODUCTION

Standardized hybrid video coders (MPEG-1/2/4, H.261//3/4) result in highly variable output bit rate. In practive, a buffer us placed at the output of the encoder, which monitors the amount of data bits generated and feedbacks to the quantization unit to adjust the quantization value so that the bit rate can be controlled. The size of the buffer is limited by the transmission delay, and the bit rate should be regulated to avoid buffer overflow or underflow. Therefore the rate control algorithm is one of the critical problems in designing a video compression system and has been widely studied in last two decades [1-4].

Conventional rate control algorithms assume a constant ratio for the number of bits generated by an I-, P- and B- frames. For example, the ratio of coded bits between an I-frame and a P-frame, W_P, is set to 3, i.e., an encoded I-frame would consume 3 times as much data bits as that of an encoded P-frame. A constant quantization value Q_I is thus suggested for all the I-frames (typically $Q_I = 15$). This could pose problems due to higher number of bits an I-frame could generate. Experiments show that the actual ratio, W_P, varies drastically from one sequence to another at constant quantization value Q_I. A bigger ratio of W_P means that an I-frame would consume larger number of bits, and leave insufficient data bits for the coding of the following P- and B- frames. It might also cause the buffer level surge at the beginning of a GOP, resulting in continuous frame skippings. As the bandwidth becomes narrower, the bit budget for I-frames is a growing burden for the whole coded bit stream. This paper proposes an adaptive rate control algorithm for I-frames by providing a mechanism to estimate their quantization value according to the spatial complexity, local motion information, and target bit rate. Experimental results show that this algorithm is very effective in controlling the data bits generated by an I-frame, and can improve the objective and subjective quality significantly.

The rest of the paper is organized as follows; Section 2 presents a new metrics to measure the spatial complexity of an I-frame. Section 3 derives a generic relationship, for I-frames, among the quantization value, data bits, and their spatial

0-7803-8603-5/04/$20.00 ©2004 IEEE

complexity. Section 4 presents the further adjustment of Q_I by using local motion information. Experimental results will be presented in Section 5 and Section 6 gives the conclusions.

2. SPATIAL COMPLEXITY MEASUREMENT OF AN I-FRAME

As is mentioned, the bits generated by an I-frame are dependent on the quantization value and its spatial complexity. There are many possible image complexity measures, such as edge intensity, gray level variance, and entropy variance [5] etc. In this paper, a new spatial complexity measure, the mean absolute value of DCT coefficients, MAV_{DCT}, is proposed based on DCT (BDCT) coefficients of the image,

$$MAV_{DCT} = \frac{1}{N \times M} \sum_{u=0,v=0}^{N-1,M-1} ABS(F(u,v)) \quad (1)$$

where $F(u,v)$ is the 8×8 BDCT coefficients of the I-frame, and N×M is the resolution of the images. MAV_{DCT} has the following unique features: 1) In most well known video coding standards such as MPEG-1/2/4 and H.26x, intra image uses BDCT transform; 2) The number of bits generated by the intra image are closely associated with their DCT coefficients; 3) Using the mean absolute values, instead of sum of absolute values, of the BDCT coefficients makes the measure independent of the video formats.

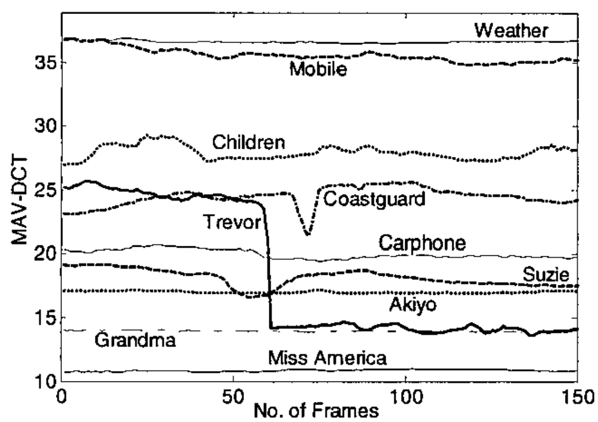

Figure 1. MAV_{DCT} versus frame number

Figure 1 shows the MAV_{DCT} values of a number of sequences. It can be seen that the spatial complexity of the sequences varies significantly, ranging from 10 to 38. MAV_{DCT} also changes within the same sequence at the junction of scene changes.

3. RELATIONSHIP OF Q_I, B_I AND MAV_{DCT}

There are typically two ways in deducing relationship of Q_I, B_I, and MAV_{DCT}: 1) Based on the stationarity and ergodicity assumptions of the statistical behavior of the image/video, the analytic approach constructs a mathematical model by analyzing the characteristics and behaviors of the image under the quantization process. However for real image/video, these assumptions are not always true [5]; 2) the empirical approach that derives the input/output relationship base on the observed data. This approach is rather useful in practice, and has been widely used nowadays [1,2]. The approach in this paper belongs to empirical approach category, and we will only deal with QCIF video format. It is simple to extend this method to any other video formats.

In order to cover the possible nonstationarity and data dependency, we have conducted extensive experiments of intra-coding on various test sequences by using different preset Q_I. It has been observed that images from 25 test sequences have shown very consistent statistical behaviors in terms of the relationships between Q_I, B_I, and their corresponding MAV_{DCT} values. These relationships are fitted by using the following equation,

$$Q_I = f(B_I) \times MAV_{DCT}^{\,g(B_I)} \quad (2)$$

Note that $f(B_I)$ and $g(B_I)$ are the two coefficients that vary according to the generated bits B_I. Equation (2) shows that, at the same number of bits, the more complicated (bigger MAV_{DCT}) an intra image has, the bigger its quantization value Q_I should be.

In order to establish the relationship among all the three parameters, i.e., Q_I, MAV_{DCT} and B_I, we need to find the relationships, for Equation (2), of $f(B_I)$ and $g(B_I)$ versus B_I. These relationships can also be closely fitted using the following two equations,

$$f(B_I) = 16.34 \times B_I^{-2.05} \quad (3)$$

$$g(B_I) = 0.29 \times \ln(B_I) + 1.0 \quad (4)$$

Therefore, by substituting Equations (3) and (4) into Equation (2), we have successfully established the analytical formula between Q_I, B_I, and MAV_{DCT} of an I-frame, as shown in Equation (5).

$$Q_I = \frac{16.34}{B_I^{2.05}} \times MAV_{DCT}^{\ 1.0+0.29\times\ln(B_I)} \qquad (5)$$

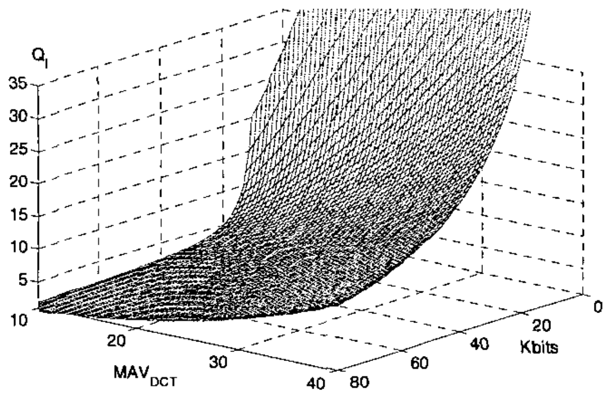

Figure 2. Q_I versus B_I and MAV_{DCT}

Note that B_I in the above equation is in Kbits. Figure 2 shows the 3-D plotting of the above relationship. It can be seen from Figure 6 that, at the same quantization value, Q_I, the bigger the MAV_{DCT} an intra image has, the more data bits it generates; and similarly, to generate the same number of data bits, the bigger the MAV_{DCT} an intra image has, the bigger its quantization value Q_I must be. Thus Equation (5) provides an effective way to control the number of coded bits before actually encoding it.

4. ADJUSTMENT OF Q_I BY USING LOCAL MOTION INFORMATION

It is well known that human vision system is less sensitive to spatial errors when there is lots of local motion or temporal changes in video; when the motion is small, spatial errors will become more irritating. This important perceptual phenomenon has not been considered in the conventional rate control algorithms [1-4]. In this paper, we will make use of this phenomenon to further adjust Q_I from Equation (5) as follows,

$$\Delta Q_I = \alpha \times \overline{mv} - \beta \qquad (6)$$

where is the average amplitude of motion vectors of the previous P-frame, which quantifies the amount of motion in the vicinity of the I-frame. The parameters α and β are chosen empirically as $\alpha = \beta = 2$. Equation (6) shows that, if the video has small motion (when $\overline{mv} < 1$ pixel), ΔQ_I is negative so Q_I will be reduced to produce better spatial quality. If the video has high motion, Q_I will be increased accordingly to save bits for P-frames.

5. EXPERIMENTS

The generic relationship described in Equation (5) has been used in our adaptive rate control algorithm. Many experiments have been conducted to evaluate the performance of this scheme. Our experiments uses the publicly available MoMuSys codec, and we set the bit rate to 64 Kbps, frame rate to 15 fps, and GOP size to 50 frames, and. In order to hold a relative constant bit ratio (W_P = 5) of I- and P-frames, the bits allocated to an I-frame should be,

$$B_I = \frac{64\times 50}{15\times(W_P+49)} \times W_P \approx 20\ Kbits \qquad (7)$$

That means that we will allocate about 20 Kbits to an I-frame in a GOP, no matter how complicated that I-frame could be. This is to guarantee that enough bits are reserved for the rest of P- or B- frames in the same GOP. By substituting B = 20 in Equation (5), we have,

$$Q_I = 0.022 \times MAV_{DCT}^{\ 1.933} \qquad (8)$$

The above Q_I will be further adjusted using the local motion information according to Equation (6) during actual coding. Finally Q_I is limited to change from 5 to 25,

$$Q_I = \max[5, \min(Q_I, 25)] \qquad (9)$$

Table 1 shows some of the experimental results by using the above formula. It can be seen that compared to MPEG-4 Q2, for spatially complicated sequences, our scheme reduces the occurrence of frame skipping with comparable average PSNRs (this is achieved by encoding 15 to 23 more frames), thus improves the subjective quality significantly; for spatially simple sequences, we have an average gain in PSNR of between 0.06 to 0.25 dB.

Table 1. Coding result comparisons (QCIF, 64 kbps, GOP 50, W_P=5)

		Q_I	B_I	FS	PSNR
Weather	Q2	15	4186	31/300	30.29
	New	25	2645	8/300	30.14
Children	Q2	15	2687	18/300	27.63
	New	20	2097	3/300	27.87
Coastguard	Q2	15	1499	2/300	30.49
	New	12	1923	0/300	30.54
Carphone	Q2	15	1335	0/300	35.46
	New	11	1986	0/300	35.63
Akiyo	Q2	15	6337	0/300	39.87
	New	7	2046	0/300	40.21
M.America	Q2	15	5973	0/150	41.0
	New	5	1846	0/150	41.06

6. CONCLUSIONS

This paper presents the generic relationship of the quantization value, data bits, and its spatial complexity for I-frames. This relationship provides a mechanism to allocate data bits to an I-frame in a controllable way before actually encoding it. The estimated Q_I is then adjusted by the local motion information to further improve the perceptual quality of the reconstructed video. Experimental results show that this simple scheme can adaptively select optimal quantization values for I-frames. It can effectively reduce the occurrence of frame skipping and improves the overall subjective quality of the reconstructed video.

7. REFERENCES

[1] A. Vetro, H. Sun, and Y. Wang, "MPEG-4 rate control for multiple video objects," IEEE Trans. on CSVT, vol. 9, No. 1, pp. 186-199, February 1999.

[2] Video Group, "MPEG-4 video verification model version 6.0," ISO/IEC JTC 1/SC 29/WG 11, Coding of Moving Pictures and Associated Audio MPEG 1996/N1582, Sevilla, Spain, February 1996.

[3] J. R- Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," IEEE Trans. on CSVT, vol. 9, No. 1, pp. 172-185, February 1999.

[4] Feng Pan, Z. G. Li, K. P. Lim and G. N. Feng, "A study of MPEG-4 rate control scheme and its improvements," IEEE Trans. on CSVT, vol. 13, No. 5, pp. 440-446, May 2003.

[5] H- M. Hang and J- J. Chen, "Source model for transform video coder and its application–Part II: Variable Frame Rate Coding," IEEE Transactions on CSVT, Vol. 7, April 1997.

[6] S. Sakazawa, Y. Yakishima, M. Wada and Y. Hatori, "Coding control scheme for a multi-encoder system,' in 7th International Workshop on Packet Video, pages 83-88, March 1996.