ADAPTIVE MAD PREDICTION AND REFINED R-Q MODEL FOR H.264/AVC RATE CONTROL

Y. Liu [1], Z. G. Li [2], and Y. C. Soh [1]

[1] Centre for Modeling and Control of Complex Systems
School of Electrical and Electronic Engineering
Nanyang Technological University
50 Nanyang Avenue, Singapore 639798
eycsoh@ntu.edu.sg

[2] Media Division
Institute for Infocomm Research
21 Heng Mui Keng Terrace, Singapore 119613
ezgli@i2r.a-star.org.sg

ABSTRACT

This paper presents an improved rate control scheme for the H.264/AVC video coding scheme. By analyzing the relationship between direct mean absolute difference (MAD) and actual MAD, a new MAD prediction scheme is introduced to enhance traditional linear MAD prediction model, which is unable to predict abrupt MAD fluctuations. Our proposed adaptive model could reduce MAD prediction error by up to 34%. One simple sum bit quadratic R-Q model is also presented to solve the problem caused by inaccurate texture bits estimation of H.264/AVC. With the new MAD prediction model and R-Q model, our proposed scheme could reduce the mismatch between actual frame bits and target frame bits by up to 32%, and the buffer occupancy is much closer to the ideal status. Meanwhile, reconstructed video quality is also improved by up to 0.21dB at low bitrate.

1. INTRODUCTION

The demands for multimedia services are rapidly increasing and the expectation of quality from these services is also becoming higher. In many video applications, the compressed video signal is to be delivered through constant bit rate channels, but the information within each frame is inherently compressed into variable bits. Adjusting only the encoding parameters to allocate extremely accurate fixed bits to every encoded frame is apparently not applicable due to the complex video encoding process. Therefore, one buffer should be added between the video encoder and the transmission channel to regulate the bitstream before transmission. Rate control has played an important role in video services. With proper rate control strategy, the encoding parameters could be adjusted to prevent the buffer from overflowing and underflowing.

The rate control has been widely studied in digital video coding standards and applications, such as TM5 [**?**] for MPEG-2, VM8 [3] for MPEG-4, JVT-G012 [1] for H.264/AVC and so on. In the rate control scheme for previous standards (MPEG-1, MPEG-2, H.261, H.263), the mean absolute difference (MAD) of residual frame which is often used as a measure of the coding complexity could be calculated based on the motion information between reference frame and predicted frame. As QP value is not involved in the motion estimation step, the actual MAD could be derived independently on QP value. As an indication of the coding complexity associated with texture information, MAD is an important variable for adjusting the QP value for rate control. Therefore the QP value could be naturally adjusted between the motion compensation step and the spatial quantization step to meet the required coding bits, according to the following R-Q model:

$$R_{texture}[i] \quad = \quad MAD_{actual}[i] \times (\frac{X1[i]}{Q[i]} + \frac{X2[i]}{Q[i]^2}) \qquad (1)$$

where $R_{texture}[i]$ denotes the amount of bits to encode texture information of frame i, $MAD_{actual}[i]$ denotes the actual MAD of the residual frame after motion compensation, $Q[i]$ is the quantization step for frame i (see [1] for the relationship between quantization step and quantization parameter), $X1[i]$ and $X2[i]$ are the first-order and second-order parameters of this quadratic R-Q model, which would be updated after encoding every frame.

One of the distinguishing features for the coding efficiency improvement of H.264/AVC is the optional implementation of rate distortion optimization (RDO) to determine motion information. Therefore, the quantization parameter (QP) affects both the motion estimation and spatial residual quantization. In this way, the actual MAD changes with the QP value adjustment as the adjusted QP value also influences the motion estimation result. So QP value adjustment based on actual MAD requires the processing of motion estimation multiple times, it is not desirable in real time applications as motion estimation is the most time consuming part of the whole video coding procedure. To solve this chicken-egg dilemma, one simple linear model is proposed in [1] to predict the MAD before motion compensation:

$$MAD_{pred,linear}[i] = Y1[i] \times MAD_{actual}[i-1] + Y2[i] \qquad (2)$$

where $MAD_{pred,linear}[i]$ denotes the predicted MAD of frame i, $MAD_{actual}[i-1]$ denotes the actual MAD of the previous residual frame, $Y1[i]$ and $Y2[i]$ are the first-order and zero-order parameters of this linear prediction model, which would be updated after encoding every frame.

Substituting $MAD_{pred,linear}[i]$ for $MAD_{actual}[i]$ into Eq. (1), the QP value for H.264/AVC rate control could be derived as:

$$R_{texture}[i] = MAD_{pred,linear}[i] \times (\frac{X1[i]}{Q[i]} + \frac{X2[i]}{Q[i]^2}) \quad (3)$$

However, there are two problems in the above quadratic model for the H.264/AVC rate control scheme:

1) If MAD fluctuats due to high motion or scene change in the test sequence, the linear model preforms poorly for such sudden changes. And the updated model parameters after the abnormal training data lead to disastrous effect in subsequent MAD linear predictions.

2) In the rate control scheme for previous standards (MPEG-1, MPEG-2, H.261, H.263), the amount of bits used for non-texture encoding in the current frame is considered the same as previous frame. So the amount of texture bits is simply derived by

$$R_{texture}[i] = R_{sum}[i] - R_{non-texture}[i-1] \quad (4)$$

In H.264/AVC, because more motion modes, more reference frames and quarter-pel motion estimation could be activated, the bits for encoding motion information takes a higher percentage than the previous standards (MPEG-1, MPEG-2, H.261, H.263). And the amount of non-texture bits also fluctuates with unpredictable property. Therefore Eq. (4) is not suitable to estimate $R_{texture}[i]$ for H.264/AVC standard.

If the above mentioned two problems could be solved properly, one refined R-Q quadratic model could help to further improve the rate control scheme for H.264/AVC.

The remainder of this paper is organized as follows. In Section 2, we observe the relationship between MAD_{direct} and MAD_{actual} and propose the direct MAD prediction model, one prediction model switching strategy is also introduced to adaptively choose linear model or direct model. In Section 3, we propose the refined R-Q model to solve the inaccurate texture bit prediction problem. With our adaptive MAD prediction method and refined R-Q model, rate control experimental results and discussions are presented in Section 4. Finally, conclusions are drawn in Section 5.

2. IMPROVED MAD PREDICTION

The MAD is used as an index for frame complexity in [3]. In H.264/AVC rate control scheme [1], the predicted MAD by linear model is not accurate when abrupt MAD changes happen due to high motion or scene change. Moreover, this linear prediction model is only updated by available MAD data, which means that this linear model only propagate the inherited property of the previous MAD data. Thus, the updated model is unable to predict future change, and is less sensitive to input data fluctuation. This is not desirable for a model which is to be used to capture abrupt input changes.

If some other information, that is helpful for predicting MAD could be collected before motion determination, then the chicken-egg dilemma could be solved. After analyzing the original input frames, we find that MAD_{direct} is somewhat related to MAD_{actual}. MAD_{direct} is a measure to evaluate the difference between the current original frame and the previous reconstructed frame, it could be calculated without motion information (as all motion vectors are set to zero). The curves of MAD_{direct} and MAD_{actual} for different test sequences are show in Fig. 1.

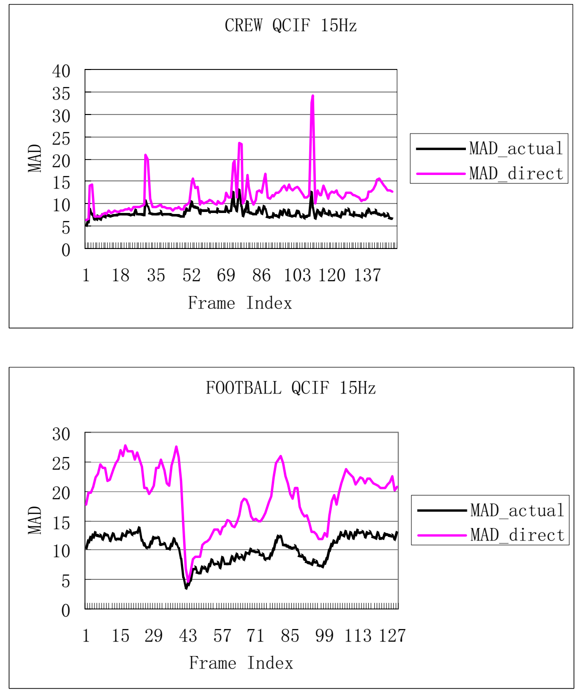

Fig. 1. The comparison of actual MAD and direct MAD. Analyze the P frames from CREW@QCIF-15Hz and FOOTBALL@QCIF-15Hz sequences, IPPP structure with QP=40.

Two observations could be extracted from Fig. 1: 1) The fluctuation of MAD_{direct} always reflects a fluctuation of MAD_{actual}, especially for abrupt changes; 2) If the MAD_{direct} curve is nearer the MAD_{actual} curve, similar fluctuations could be observed more obviously. According to the above two observations, we could derive our direct MAD prediction function as follows:

$$\begin{aligned} MAD_{pred,direct}[i] &= MAD_{actual}[i-1] \\ &\times (1 + W \times MAD_{direct,ratio}[i]) \quad (5) \end{aligned}$$

where

$$W = \frac{MAD_{actual}[i-1]}{MAD_{direct}[i-1]}$$

$$MAD_{direct,ratio}[i] = \frac{MAD_{direct}[i] - MAD_{direct}[i-1]}{MAD_{direct}[i-1]}$$

Although the above direct MAD prediction could efficiently predict the actual MAD abrupt change, for some sequences with fine texture details (e.g. *HARBOUR*) or irregular regional motion (e.g. *FOREMAN*), this direct prediction model sometimes does not work such efficiently. Therefore, we introduce a similarity measurement to indicate the efficiency of direct MAD prediction model and linear MAD prediction model:

$$\Gamma_{linear}[i] = \sum_{n=i-S}^{i} |MAD_{pred,linear}[n] - MAD_{actual}[n]|$$

$$\Gamma_{direct}[i] = \sum_{n=i-S}^{i} |MAD_{pred,direct}[n] - MAD_{actual}[n]|$$

where S is the number of MAD samples to measure Γ. The method to predict MAD could be adaptively switched between linear model and direct model, according to Eq. (6) given below. Note that both the way to measure Γ and define the switching policy are not limited, other complicated similarity measurement or switching threshold may achieve more efficient prediction model shift.

$$MAD_{pred,adapt}[i+1] = \begin{cases} MAD_{pred,linear}[i+1] & : \Gamma_{linear}[i] < \Gamma_{direct}[i] \\ MAD_{pred,direct}[i+1] & : \text{else} \end{cases} \quad (6)$$

Fig. 2 shows the comparison of MAD_{actual}, $MAD_{pred,linear}$ and $MAD_{pred,adapt}$. When the MAD abrupt change happens, such as the 31st, 54th frames in *CREW*, the 43rd frame in *FOOTBALL*, our adaptive MAD prediction method could achieve much more accurate result. The prediction errors of $MAD_{pred,linear}$ and $MAD_{pred,adapt}$ in relation to MAD_{actual} are also measured in term of MAD to compare the accuracy of these two prediction schemes. The details are shown in Table. 1.

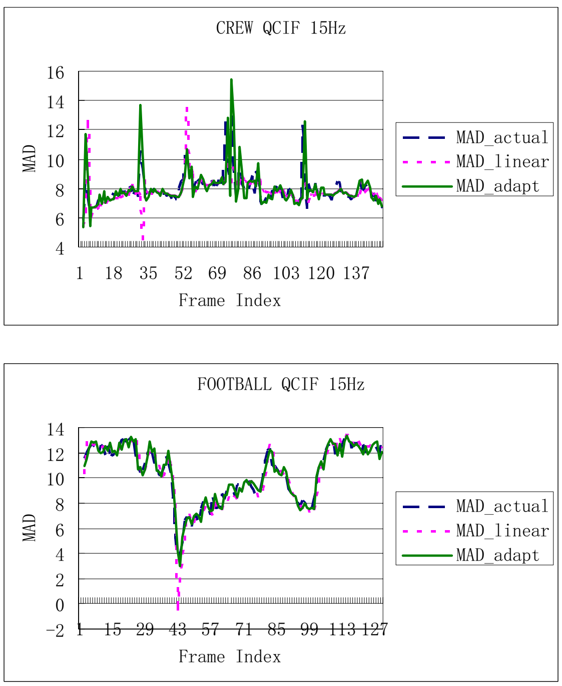

Fig. 2. The comparison of actual MAD, linear predicted MAD and proposed adaptively predicted MAD. Analyze the P frames from CREW@CIF-15Hz and FOOTBALL@CIF-15Hz sequences, respectively. IPPP structure with QP=40.

Table 1. Prediction Error Comparison of $MAD_{pred,linear}$ and $MAD_{pred,adapt}$.

	Prediction Error	
	$MAD_{pred,linear}$	$MAD_{pred,adapt}$
CREW	0.65	0.43
FOOTBALL	0.60	0.46

From the comparison of Fig. 2 and Table. 1, it could be found that our adaptive model could achieve more accurate predicted MAD than linear MAD prediction model proposed in [1], especially for the sequence with high motion or scene changes. Our model could reduce MAD prediction error by up to 34%.

3. SUM BIT QUADRATIC R-Q MODEL

In previous standards (MPEG-1, MPEG-2, H.261, H.263), the non-texture bits take only a small percentage of sum bits and they vary slightly. So the number of non-texture bits could be considered the same as previous frame. Based on this assumption, the term “R” in the quadratic R-Q model, which is the amount of bits to encode texture information, could be calculated via Eq. (4). On the other hand, H.264/AVC standard newly introduces many complicated motion estimation modes, which leads to some increase of non-texture bits. At the same time, because more efficient motion estimation is employed, the energy in residual frame reduces. Therefore, non-texture bits occupy a higher percentage of the sum bits, especially at low bit rate. And there also exists hard-to-predict fluctuations in non-texture bits. So Eq. (4) could not be used to estimate the amount of texture bits. The prediction of non-texture bits with fluctuation is a very complicated task, and the error of such prediction will finally yield a poor rate control result.

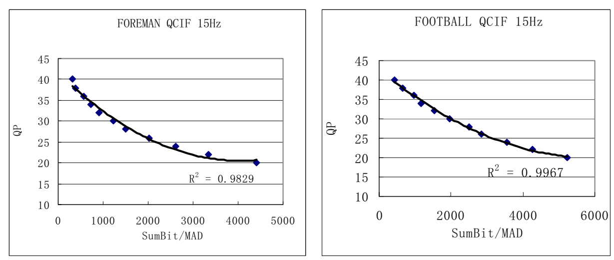

Fig. 3. The relationship sum bit and QP. Randomly choose one P frame from FOREMAN@CIF-15Hz and FOOTBALL@CIF-15Hz sequences, respectively. QP=20:2:40.

To solve this problem, we found that the prediction process of the amount of non-texture bits might be abridged if a simple model could be formulated to relate the amount of sum bits to the QP value. To construct a R-Q model, one approach is to analyze the structure of the video processing and the statistical properties of the video data, which is called the analytical approach. Another approach is to derive the R-Q relationship based on several sampled values, which is called the empirical approach. Although the empirical approach is easy to apply, it does not take into account the video coding process. Our proposed approach is based on the result in [3], which belongs to an analytical approach category, but we make some improvements according to empirical result.

As the QP value affects both the motion determination and texture quantization, only considering the relation of QP and texture bits is insufficient. Suppose there is a relation between sum bit ($SumBit$), MAD (MAD) and QP (Q), i.e. $SumBit = \psi(MAD, Q)$. Based on simulation results, one very accurate quadratic relationship could also be observed between $\frac{SumBit}{MAD}$ and Q, as shown in Fig. 3. R-squared value (R^2), is an indicator from 0 to 1 that reveals how closely the approximated quadratic function correspond to actual data. The approximated quadratic function is most reliable when its R-squared value is at or near 1. From empirical results, the

MAD could also be used as a complexity indicator to encode one frame into sum bits. Therefore, the prediction process of the amount of non-texture bits could be abridged, and the quadratic equation (3) is directly changed to:

$$R_{sum}[i] \quad = \quad MAD_{pred,adapt}[i] \times \left(\frac{X1[i]}{Q[i]} + \frac{X2[i]}{Q[i]^2}\right) \quad (7)$$

where $R_{sum}[i]$ is the target sum bit of the current frame i, which could be calculated according to the buffer status. In this way, $Q[i]$ could be calculated based on $R_{sum}[i]$ and $MAD_{pred,adapt}[i]$, which are both available before motion estimation.

4. EXPERIMENT RESULT

The performance of the H.264/AVC rate control scheme with our proposed adaptive MAD prediction and refined R-Q model is compared with JVT H.264/AVC package JM9.8 [4], in which JVT-G012 [1] is adopted as rate control scheme. All the test sequences used in our experiment are at QCIF (176×144) spatial resolution and 15Hz frame rate with IPPP GOP structure. For simplicity of evaluation, the QP value is set to 40. The objectives of our rate control scheme are to regulate the amount of sum bit to encode each P frame, efficiently utilize the buffer resource and prevent the buffer from overflowing and underflowing.

Table 2. The Comparison of the Frame Level Target Bits Mismatch.

	Target Bits Mismatch (MAD)	
	scheme in [1]	our scheme
CREW	546	431
FOOTBALL	727	528
HARBOUR	473	323
FOREMAN	484	459

We first compare the difference between frame level actual sum bits and target sum bits in Table. 2 in term of MAD. The target bits mismatch is reduced up to 32% by our scheme.

The buffer fullness status is also compared in Fig. 4. In our simulation, if the buffer occupancy is above 1.0 or below 0, the buffer suffers from overflowing or underflowing. It is obvious that the encoder with our proposed rate control scheme outperforms JVT-G012 [1] in terms of keeping the buffer status close to the target status.

Table 3. The Comparison of Coding Efficiency.

Sequence	Bit Rate (kb/s)		PSNR (dB)		
	[1]	Ours	[1]	Ours	Gain
CREW	30.01	29.98	28.72	28.75	+0.03
FOOTBALL	64.06	64.06	27.51	27.63	+0.12
HARBOUR	20.20	20.12	25.07	25.11	+0.04
FOREMAN	19.05	18.99	27.96	28.19	+0.23

Although our major focus is not to improve the coding efficiency, the encoder with our proposed rate control scheme could also achieve better reconstructed objective quality thanks to the accurate prediction model. In the experiment results presented in Table. 3, the improvement is up to 0.21dB at low bitrate.

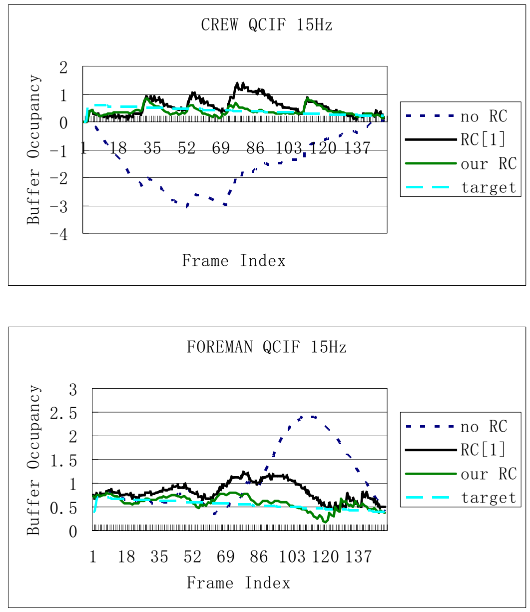

Fig. 4. The comparison of buffer status between different coding scheme. "RC" denotes rate control scheme.

5. CONCLUSION

In this paper, we addressed the problems associated with existing H.264/AVC rate control strategy proposed in [1]: inaccurate MAD prediction and inaccurate estimation for the amount of texture bits. With our proposed methods, the MAD could be predicted adaptively by linear model or direct model to achieve more accurate prediction. Our proposed adaptive MAD model could reduce the prediction error by up to 34%. Moreover one simple quadratic R-Q model between the amount of sum bits and QP value is created by abridging the inaccurate estimation process of texture bits. With the above improvements, one new R-Q model is proposed in this paper for H.264/AVC rate control. From our simulation result, the mismatch between target frame bits and actual frame bits is reduced by up to 32%, and the buffer occupancy is better controlled from overflowing and underflowing. Meanwhile, the reconstructed video quality is improved by up to 0.21dB at low bitrate.

Note that the improvements mentioned in this paper could be further implemented to achieve precise MB level rate control and this will be reported in our future work.

6. REFERENCES

[1] Z. G. Li, F. Pan, K. P. Lim, G. Feng, X. Lin and S. Rahardja, "Adaptive Basic Unit Layer Rate Control for JVT," JVT-G012-rl, 7th Meeting, Pattaya II, Thailand, Mar., 2003.

[2] "MPEG-2 Test Model 5," Doc. ISO/IEC JTC1/SC29 WG11/93-400, Apr., 1993.

[3] H. J. Lee, T. H. Chiang and Y.-Q. Zhang, "Scalable Rate Control for MPEG-4 Video," *IEEE Trans. Circuits, Syst., Video Technol.*, Vol. 10, pp. 878–894, Sep., 2000.

[4] "JM9.8 software package," http://ftp3.itu.ch/av-arch/jvt-site/reference-software/jm98.zip.