Improved Rate Controlling with Efficient Bit-Rate Abundance Measurement and Prediction in H.264/AVC Video Coding

Din-Yuan Chan, Lin-Yu Liao, Yu-Hsien Chen

Dept. of Computer Science and Information Engineering and Dept. of
Electrical Engineering, National Chia-Yi University
Chia-Yi City 60004, R.O.C.
Tel:05-2717726 Fax:05-2717740
`dychan@mail.ncyu.edu.tw`
`realjerry@gmail.com`
`jiojima@gmail.com`

***Abstract*—The paper proposes a DCT-spectral bit-amount abundance (BAE) instead of the mean of absolute difference (MAD) used in JM12.4 to improve the bandwidth utilization of rate controlling for H.264/AVC video coding. The BAE costs the H.264 encoder the low extra computation complexity as a new rate-estimation metric. It can not only have a good mapping of coding bit count and, particularly, but also offer well prediction for itself. In this study, the rate controller uses the predicted value of BAE from the stored frame rather than the actual one under the constraint of one-pass processing in QP-decision [1]. And, it adopts an efficient match-testing method using a directional-contrast vector to find the highest-correlation predictive MB in the neighbourhood of co-located MB to support the rate-parameter estimation. Experiment results show that the proposed rate controlling effectively improves the H.264/AVC coding quality in comparison with the rate controlling suggested by JM12.4 in terms of PSNR and buffer fullness stabilization with negligible increase in computational complexity.**

***Keywords*—Rate controlling, H.264/AVC, Bit-rate estimation, Bit-amount abundance, Mean of absolute difference**

1. INTRODUCTION

The H.264/AVC video standard [1] is a video coding milestone in visual communication development. Unlike other existing video encoders, a quite complete rate control mechanism has been almost concurrently opened by JM [2] with announce of compliant H.264 specification. Based on the Lagrangian optimization, the mechanism is implemented by the rate-distortion optimization technique [3], which what adopts some mature opinions such as the buffer-management based bit-allocation in [4] and the QP decision formula in [5]. Since the actually adopted quantization parameters (QPs) of MBs are determined after RDO [6] or a later stage [4], the constrained optimization problem offers an optimal approach with the assumption that to estimate the bitrate complexity of a coding source can merely depend on the quantization-irrelevant parameter such as MAD [6]. Although, two rate estimation parameters based on ρ-domain [7] have well included the quantization effects, those rate-estimation parameters has an application obstacle to the QP decision in H.264 without prior pseudo-coding. In the H.264/AVC video coding, the constraint of one-pass rate controlling need be complied [1] in general for limiting the complexity of already quite sophisticated H.264/AVC. A basic solution strategy is to employ the estimate of content-complexity of coding source rather than the actual value obtained by pre-quantization (or pseudo-coding). Hence, in this study, based on JM12.4, we propose a simple, yet effective improved rate-controller associated with a new parameter, say DBA, and a new definition of referable MB for the DBA predication. The DBA of current MB is predicted by the DBA of a new-defined referable MB, which is searched from the stored reference frame. By our experiments, DBA can offer the higher accuracies of bitrate mapping and its own prediction, both, than MAD and ρ. Hence, the effectiveness of QP decision can be substantially raised by the improved rate controlling relative to the implementation specified in [6]. In addition, the calculations related to the DBA are easier than that to the ρ-based parameters in rate estimation such that the complexity increase due to the improvement of rate controlling is regarded negligible.

2. PROPOSED BIT-RATE ABUNDANCE ESTIMATION

At present, one direct issue to improve the H.264 rate controlling is the creation of a low-complexity reliability parameter with much more accuracy over MAD in the QP decision. In this study, one well-designed parameter (i.e., DBA) indeed provides the property described above to entirely replace MAD. When the higher accuracy is induced from the higher rate estimation supported by such a parameter, the parameter is impossibly completely irrelevant to the quantizer. Basically, this parameter is obtained in the pseudo-coding or pre-quantization of need due to the chicken-egg dilemma [6]. Thus, it shall become problematic when forbidding the multi-pass processing in complex codecs just like the H.264/AVC codec. Therefore, we follow the well-suggested procedure in JM12.4 [2] to basically solve the problem, and further make ingenious but efficient modification in the utilization of data prediction and coefficient adaptation, etc.

The proposed DBA is a unified parameter composed of two ingredients that one is universally known as the energy of quantized DCT spectrum, which denoted as $\chi(QP)$ is the function of QP, another is a frequency characteristic called frequency-saturation parameter (FSP) denoted as ϕ to efficiently express the concentration and the range of quantized source in frequency-domain. The FSP is concerned with the occupancy ratio of non-zero DCT coefficients and the quantized DCT bandwidth. They are two key factors to be intimately relevant to the coding bit rate regardless of the magnitudes of DCT coefficients. After introducing the effect of DCT coefficients' magnitudes, to compute the DBA of coded source symbolized by $\psi(q)$ is modelled by

$$\psi(q) = \chi(q)\cdot\phi \tag{1}$$

while QP=q. The resultant $\psi(q)$ will be used to obtain the characteristic rate curve for estimating the coding bitrate denoted by R(q) whose estimation result is expressed by $\tilde{R}(q)$. In addition to the accuracy of rate estimation using (1), another merit of formularizing DBA by (1) is the analytical facilitation of cause and consequence about bitrates due to the separable individual observation of influential factors.

2.1. Linearity-like DBA-domain Bitrate Estimation

The advantage of the bitrate solely depending on DBA is making both the bit allocation and the adaptation of DBA estimation equation easy for the H.264/AVC rate control. By the observations of $\chi(QP)$ shown by Fig. 1, we found that there exists a slightly-bending parabola curve relation between $\chi(QP)$ and the bitrate of coded MB texture that the curve passes through the origin after the regression procedure. Assume that $y(x_i)$ can be approximated by $\hat{y}(x_i)$ via the regression procedure and the mean of $y(x_i)$ is $\bar{y}(x_i)$ at x_i, which is the ith sampled value of x. Then, the coefficient of determination to evaluate the regression performance can be formularized as follows:

$$R^2 = \frac{\underset{all\ i}{mean}\left((\hat{y}(x_i)-\bar{y}(x_i))^2\right)}{\underset{all\ i}{mean}\left((y(x_i)-\bar{y}(x_i))^2\right)} \tag{2}$$

where $\underset{alli}{mean}()$ expresses the function of computing the average in parentheses throughout all the sampled points, i.e. x_i for all i's. When R^2 is closer to 1, a better approximation with lower regression error is indicated.

One shall have an opportunity to employ a factor, i.e. ϕ, to multiply $\chi(QP)$ to obtain $\psi(q)$ with a linear or linear-like parabola relation to the bitrate of coded MB texture. From the mathematical opinion, the MB-texture bitrate could also be approximated by a slightly-bending parabola curve of ϕ. Observe the order of H.264/AVC entropy coding procedure and the key factors described above, ϕ is to concisely reflect accurate prediction of coding bitrate and can be well-designed as

$$\frac{1}{\Gamma}\left(1+\kappa_{\text{mod}e}\frac{N_{nonzero}}{N_{DCT\_BW}}\right)\cdot\left(N_{DCT\_BW}\right) \tag{3}$$

where Γ is for normalization; we can tune κ_{mode} to carry out an accurate relation in estimating the bitrate for different coding modes that κ_{mode} is a variable of different fixed values according to various types of block partition. By (3), as Fig. 2 shown, the linear-like parabola-curve relation can be obtained via the regression approximation with lower regression error than ρ in [7].

However, both ϕ and ρ can not be directly applied in practical QP decision for the lack of information of varying frame energies during encoding. Hence, the realistic meaning of straight line relation of ρ and coding bit rate in regression is not beyond that of linear-like parabola-curve relation of ϕ and coding bit rate.

By (1) and (3), we have the following equation to express the DBA of MB denoted as $\psi_{MB}(q)$:

$$\begin{aligned}\psi_{MB}(q) &= sum\left(DCT(q)\right)\cdot\frac{1}{\Gamma}\left(1+\kappa_{\mathrm{mod}\,e}\frac{N_{nonzero}}{N_{DCT\_BW}}\right)\cdot\left(N_{DCT\_BW}\right)\\ &= sum\left(DCT(q)\right)\cdot\frac{\left(N_{DCT\_BW}+\kappa_{\mathrm{mod}\,e}\times N_{nonzero}\right)}{\Gamma}\equiv\chi(q)\cdot\phi\end{aligned}\tag{4}$$

when QP=q. As tabulated in Table 1, the coefficient of determination for the mapping the coding bitrate by ϕ could be larger than that by ρ (the percentage of zero DCT coefficients quantized) [7] for a fixed coding frame given. This is a significant advantage to construct a rate-complexity (R-C) model for accurate bitrate estimation. Although the regressed function of ϕ mapping the bitrate is a parabola curve, i.e. two-order polynomial, rather than a simple linear, employing DBA in estimation of bitrate is not more complicated at all than the rate estimation parameter involving ρ [7]. In effect, the complexity of utilizing the curve of $\alpha(QP)$ versus R(QP) is even simpler than that of utilizing the two characteristic rate curves proposed in [7]. This is because neither ϕ or ρ can be straightforward applied in R-C-Q model to obtain a rate estimate without the information of spectral energy. Rather, just like the ρ -domain rate estimation parameter, DBA can offer a bitrate estimate for a coding frame or MB. Particularly, the relation of $\alpha(q)$ and R(q) can be considerably approximated by a linear function, while QP=q, such that a simplest mapping relation will be utilized to govern the bitrate estimation and then the QP decision.

3. DBA UTILIZATION IN RATE CONTROL IMPLEMENTATION

Observing experimental results in Fig.1 and Fig.2, we know that the DBA in (1) theoretically can provide a good mapping of coding bit rate with a slight-bending parabolic curve through the origin. The bitrate of coding MB's texture corresponding to $\psi_{MB}(q)$ in (4) can be estimated by a non-bias 2-order polynomial:

$$\tilde{b}_{MB} = a_2\psi_{MB}(q)^2 + a_1\psi_{MB}(q) \tag{5}$$

where a_1 and a_2 are adaptive model coefficients adapted after coding a frame. The "slight-bending" property of parabolic curve in bit rate approximation gives $a_1 >> a_2$.

3.1. MB-Layer DBA Prediction for MB and Frame DBA Estimation

Our proposed QP-decision improvement mechanism relative to that reported in JVT[2] is substituting DAB for MAD to express MB's bitrate complexity, and improving the prediction process to set the estimate of DAB. In JVT[2], the MB-layer MAD is estimated by a linear prediction equation using co-located MB's DBA in the reference frame for solving a chicken-egg problem in rate controlling, so does the MB-layer DBA value of MB in the improvement. To fix the co-located MB of coding MB as its only predictive reference appears rough because of the possible existence of different coding modes, and might be not plausible for high-activity video fragments. Therefore, the proposed scheme adopts the neighboring highest-correlation MB searched rather than fixedly the co-located MB in the reference frame for predicting the complexity of each coding MB in terms of DBA. For suppressing the possible non-consistence influence caused by different quantizers, we define a match-testing vector using a directional-contrast vector covering five directional contrasts for the similarity comparison in relatively searching a highest-correlation MB of current coding MB. The highest-correlation MB is searched from the square range of 25-neighboring MBs centered by the co-located MB on the reference frame. Such a MB could have near characteristics such as the coding mode to the current coding MB and, thus, be a suitable predictive MB for the latter. The directional-contrast vector is defined as $V_C=\left[\bar{Y}_{LU}-\bar{Y}_{RU},\bar{Y}_{RU}-\bar{Y}_{RD},\bar{Y}_{RD}-\bar{Y}_{LD},\bar{Y}_{LD}-\bar{Y}_{RU},\bar{Y}_S-\bar{Y}_C\right]$,where $\bar{Y}_{LU}$, $\bar{Y}_{RU}$, $\bar{Y}_{RD}$ and $\bar{Y}_{LD}$ are the luminance means of left-upper, right-upper, right-down and left-down 8×8 sub-blocks, respectively, and $\bar{Y}_S$ expresses the luma mean of 192 pixels, which forms a 8-pixel-width strip encompassing the 8×8 center block, whose luma mean is denoted as $\bar{Y}_C$, in a MB. The former four elements and the last element of V_C represent directional circular-ordered differences of four 8×8 sub-blocks and an outer-to-inner difference, respectively, for a MB. Such an outer-to-inner difference is associated with the possibility of 4×4-block partitioning (finer than

8×8-block partitioning). Then, the match-testing vector is composed of these five elements via normalization. It is able to tersely describe the existing difference about the strength and the number of classed pixels within a MB such that the vector is directly concerned with the tendency of block partition in a MB, i.e. the possible coding modes in H.264/AVC standard.

The highest-correlation MB is to offer the DBA estimate of coding MB via a simple one-order prediction shown by

$$\tilde{\psi}_{MB}^{f}(i,j)=\alpha_1(f)\left(\psi_{matched}(f,i,j)\right)+\alpha_0(f) \qquad (6)$$

where f and (i,j) are frame index and 2-D MB index, respectively, and predictor coefficients $\alpha_1(f)$ and $\alpha_0(f)$ are frame-adaptive; $\tilde{\psi}_{MB}^{f}(i,j)$ expresses the DBA estimate of (i,j)th MB, whose highest-correlation MB owns the DBA value equal to $\psi_{matched}(f,i,j)$, in the fth coding frame. Let the true DBA of (i,j)th MB in the fth coding frame is denoted as $\psi_{MB}^{f}(i,j)$ such that $\psi_{matched}(f,i,j)$ will be one element in $\left\{\psi_{MB}^{f-1}(i+k,j+k)\,|\,k=0,\pm 1,\pm 2\right\}$. In (6), the adaptation of $\alpha_1(f)$ and $\alpha_0(f)$ will directly follow the related method suggested in JM12.4. In our scheme, to compute the DBA estimates of all MBs within a coding frame (6) is like a pre-processing. It is performed prior to the QP decisions of those MBs. Thus, the estimate of coding frame's DBA can equal the summation of DBA estimates of its MBs so that the DBA of fth coding frame is

$$\tilde{\psi}_{Frame}(f)=\sum_{i=0}^{N_H-1}\sum_{j=0}^{N_V-1}\tilde{\psi}_{MB}^{f}(i,j) \qquad (7)$$

where N_H and N_V express the count of horizontal MBs and that of vertical MBs, respectively, in a coding frame.

3.2. Confirming the DBA Estimates of MBs and Their Shrinkages

It is basic that the bit-rate of coding frame shall be restricted under the available buffer space for avoiding buffer overflow. Therefore, when the predicted bit count of a coding frame causes buffer fullness to exceed a well-designed upper-bound margin, the DBAs of MBs in this coding frame shall be shrunken to reduce their bit consumption. For making fast shrinkage for all MBs' DBA estimates if necessary, the proposed rate control will merely shrink the estimate of coding frame DBA and then compute a fixed shrinking ratio for all MBs in this frame. The detail is described as follows.

By using (5), the coding bitrate estimate of fth frame $\tilde{b}_{Frame}(f)$ can be obtained by equation: $\tilde{b}_{Frame}(f)=a_2\tilde{\psi}_{Frame}(f)^2+a_1\tilde{\psi}_{Frame}(f)$. Thus, theoretically, the bit target of fth coding frame $T_{Frame}(f)$ could be allowed as $\tilde{b}_{Frame}(f)$ but limited under an upper bound $T_{up\_bound}(f)$ so that

$$T_{Frame}(f)=\min\left\{\tilde{b}_{Frame}(f),T_{up\_bound}(f)\right\} \qquad (8)$$

In (8), $T_{up\_bound}(f)$ is set as ($S_M-B_{f-1}-\frac{B}{F}$), where S_M is the buffer margin defined by the buffer size minus a guard volume, and B_{f-1} expresses the buffer fullness immediately measured after buffering the coding bit count of (f-1)th coding; B and F are bandwidth and frame rate, respectively. The value of ($S_M-B_{f-1}-\frac{B}{F}$) is considered as the maximal supportable buffer space for the fth coding frame. By (8), when $T_{Frame}(f)$ is still kept as $\tilde{b}_{Frame}(f)$, modifying $\tilde{\psi}_{MB}^{f}(i,j)$ is of course unnecessary. On the contrary, if $T_{Frame}(f)$ exceeds $B_{upper\_bound}(f)$ and need be reduced to latter, the required shrinking ratio μ_f is set as

$$\mu_f=\frac{\tilde{\psi}_{Shrinkage}(f)}{\tilde{\psi}_{Frame}(f)}<1 \qquad (9)$$

In (9), the shrunken DBA of this frame can be easily obtained by $\tilde{\psi}_{Shrinkage}(f)=\frac{1}{2a_2}\left(-a_1+\sqrt{a_1^2+4a_2B_{upper\_bound}(f)}\right)$ by solving (5). Then, all $\tilde{\psi}_{MB}^{f}(i,j)$'s in (6) are equally weighted down by multiplying μ_f to linearly shrink the DBA estimates of MBs in the frame. Finally, the QPs of all MBs in the fth coding frame can be concurrently yielded by performing the related formulas proposed in JM12.4 by replacing the MAD estimates by all $\tilde{\psi}_{MB}^{f}(i,j)$'s.

4. EXPERIMENTAL RESULTS

The proposed rate control scheme was implemented in JM12.4. In our experiments, a total number of 100 frames sampled by 30fps are used to each tested QCIF sequence of picture type IPPP…, where the number of available reference frames is set to 5, at 64kbps, 128kbps, 256kbps and 512 kbps. Observe Fig. 3 and Fig. 4, the bitrate and PSNR stabilization can be improved by our proposed scheme compared to the rate controlling recommended in JM12.4 for the sequence "Grandma". Table 2 showed PSNR improvements and Table 3 showed computational time comparisons. In this work, the DBA of coding MB is entirely predicted from the correlation MB in the reference frame without pre-quantization to yield coding MB' own information to strictly limit the increase of computational complexity resulting. In Table 4, we can see that adopting DBA and FSP can acquire lower prediction errors than MAD and ρ, respectively, in rate controlling. Substantially, the less the prediction error of bitrate-associated parameter, the higher the accuracy of decided QP. By our experiments, the increased computation time of utilizing the proposed improvement in rate-controlling is equal to 0.006% on average only relative to the rate controlling of JM12.4 [2]. Hence, the increased PSNR gain tabulated can not appear salient, but could be considered as a solid improvement effect introduced by our proposed algorithm under such a crucial limitation above.

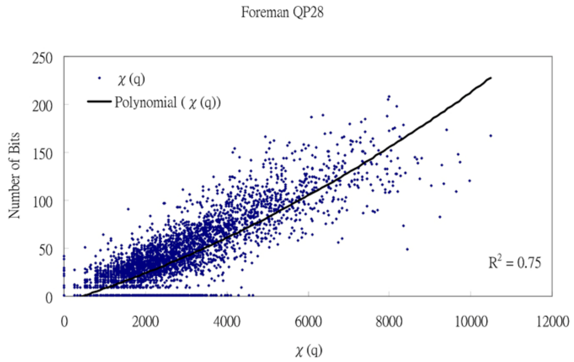
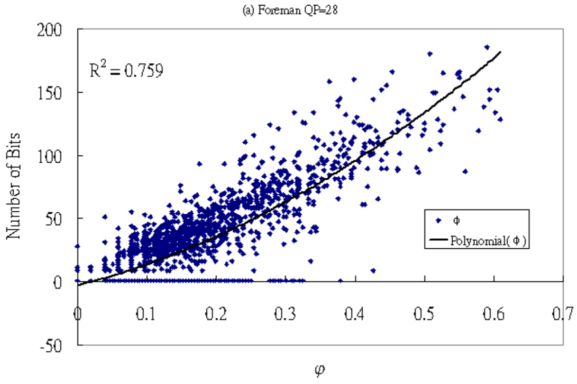


Figure 1: Plots of R(q) vs. $\chi(q)$ for coding the Foreman QCIF video by QP=28. The coefficient of determination R^2 in constructing the rate estimation curve is 0.75.



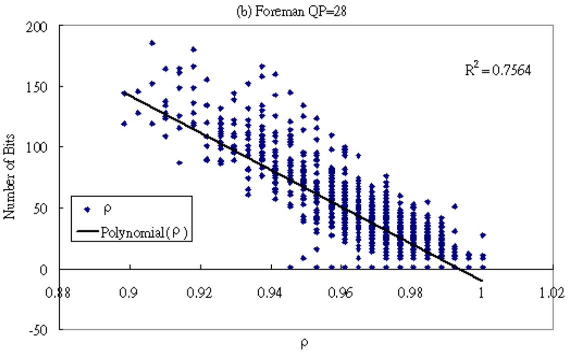


Figure 2: Plots of the relation functions (a) R(QP, ϕ) with R^2 = 0.759 and (b) R(QP, ρ) with R^2 = 0.7564 in regression for coding the QCIF " Foreman" sequence to MBs of P16×16 (i.e., $\kappa_{p16\times16}$).

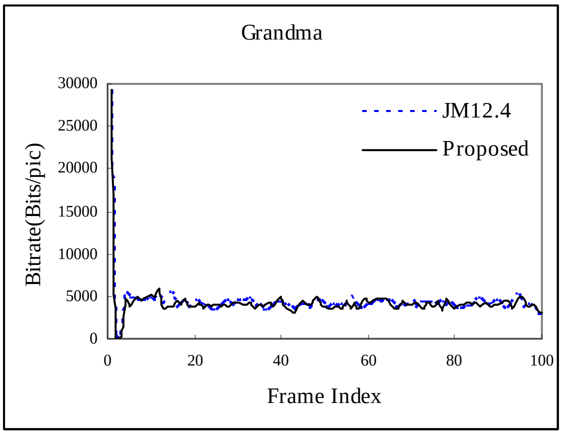


Figure 3: The bitrate comparison of proposed rate control and JM12.4 rate control for "Grandma" sequence at the target bitrate of 128kbps.

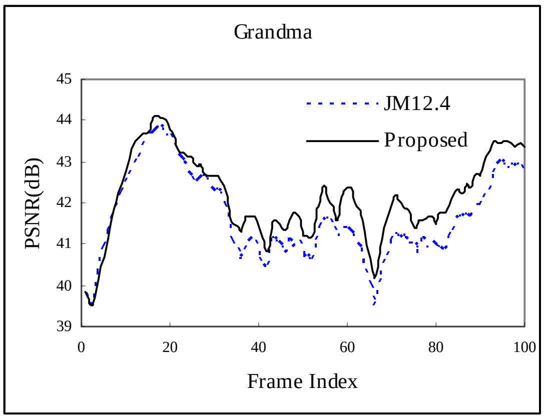


Figure 4: The PSNR comparison for sequence "Grandma" at the target bitrate of 128kbps.

Table 1
The coefficient of determination R^2 of function mapping the coding bitrate by ϕ and that by ρ for a fixed coding frame

P-Frame No.	ϕ	ρ
Foreman 2nd P-Frame	0.9987	0.9984
Foreman 8th P-Frame	0.9973	0.996
Foreman 12th P-Frame	0.9977	0.9967
Foreman 20th P-Frame	0.9981	0.9977

Table 2
Performance comparison of the proposed scheme and JM12.4.

Sequence (Target-Bitrate)	**PSNR (dB)**		**Bitrate (kbps)**	
	JM	**Proposed**	**JM**	**Proposed**
Akiyo (64kbps)	40.10	40.52	64.3	64.2
Foreman (64kbps)	33.22	33.08	64.2	64.4
Grandma (64kbps)	37.61	37.89	64.2	64.4
Akiyo (128kbps)	44.75	45.13	128.7	128.3
Foreman (128kbps)	36.45	36.34	128.3	128.4
Grandma (128kbps)	41.64	42.15	128.3	128.3
Akiyo (256kbps)	48.16	48.41	256.8	255.0
Foreman (256kbps)	39.41	39.40	256.2	256.1
Grandma (256kbps)	44.90	45.50	256.9	256.0

Table 3
Sequences coding time comparison of the proposed scheme and JM12.4.

Sequence (Target Bitrate)	**Coding Time (second)**		
	JM12.4	**Proposed**	**Increased (%)**
Akiyo (64kbps)	429.494	429.205	-0.07
Foreman (64kbps)	431.511	434.993	0.806
Grandma (64kbps)	431.904	429.847	-0.476
Akiyo (128kbps)	431.059	429.861	0.232
Foreman (128kbps)	435.096	438.586	0.802
Grandma (128kbps)	431.061	428.123	-0.681
Akiyo (256kbps)	434.369	432.090	-0.524
Foreman (256kbps)	440.202	442.434	0.507
Grandma (256kbps)	431.137	430.720	-0.09

Table 4
The bitrate Prediction error (MAE) comparison of DBA, MAD, FSP and ρ -based bitrate-estimation parameter for sequence Grandma at bitrate 128kpbs.

Parameter Type	**Mean absolute error (MAE)**
DBA (proposed)	0.17955
MAD (JM)	0.25079
FSP (proposed)	0.19657
ρ **-based**	0.21567

5. CONCLUSIONS

The paper presented a key rate parameter called DBA to replace MAD in the MB QP decision equations of JM12.4 [2]. The DBA can provide more accuracy in bit-rate estimation than the rate control parameter obtained in ρ-domain. Since the rate controller merely uses the estimate of DBA rather than its actual value for avoiding pre-quantization, the DBA prediction shall more accuracy yet fast than that applying the co-located MB as the only predictive MB. In order to efficiently suppress the prediction error, in this paper, a low-complex match-testing method to rapidly search the highest-correlation predictive MB is further proposed. Experiment results demonstrate that the proposed rate controlling effectively improves the H.264/AVC coding performance compared to the rate controlling suggested by JM12.4 in PSNRs and bitrate stabilization with slight increment in computational complexity.

ACKNOWLEDGMENT

The authors wish to thanks the support of NSC plan numbered NSC96-222-E-415-006.

REFERENCES

[1] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, “Draft ITU-T recommendation and final draft international standard of joint video specification,” *(ITU-T Rec. H.264|ISO/IEC 14496-10 AVC)*, JVT-G050, 2003.

[2] K. Lim, G. Sullivan, and T. Wiegand, “Text Description of JM Reference Encoding Methods and Decoding Concealment

Methods," *JVT-W057 contribution*, San Jose, Apr. 2007.

[3] G.J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Magazine*, Vol. 15, Nov. 1998, pp. 74–90.

[4] ITU-T Study Group16, Video Coding Experts Group, "Video codec test model," *TMN8*, Jun, 1997.

[5] H. J. Lee, T. Chiang, and Y. Q. Zhang, "Scalable rate control for MPEG-4 video,“ *IEEE Trans. Circuits Systems Video Technology*, Vol. 10 No. 6, Sep. 2000, pp. 878-894.

[6] Z. Li, F. Pan, K.-P. Lim, G. Feng, X. Lin, and S. Rahardja, "Adaptive basic unit layer rate control for JVT,“ *JVT-G012*, March 2003.

[7] Z. He and K. Mitra, "A Unified Rate-Distortion Analysis Framework for Transform Coding," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 11 No. 12, Dec. 2001, pp.1221-1236.