INTER-DEPENDENT RATE-DISTORTION MODELING FOR VIDEO CODING AND ITS APPLICATION TO RATE CONTROL

Yuan Li, Huizhu Jia[*], Pan Ma, Chuang Zhu, Xiaodong Xie and Wen Gao, Fellow, IEEE

National Engineering Laboratory for Video Technology
Peking University, Beijing, China
{yuanli, hzjia}@pku.edu.cn, {mapan, czhu}@jdl.ac.cn, {donxie, wgao}@pku.edu.cn

ABSTRACT

Rate control scheme using independent rate-distortion (R-D) model at the minimum coding unit (macroblock) level has been widely discussed in the literature where R-D optimization is performed without consideration of the inter-dependencies of different coding units. In this paper, we extend these techniques to the more general situations – rate control for inter-dependent video coding units. The inter-dependent distortion-quantization (D-Q) model and rate-quantization (R-Q) model are formulated separately based on the analysis of the relationship between the spatial-domain residual and the transform-domain residual. Then a window-based rate control scheme with frame bit allocation and video quality optimization is proposed, which uses the approximated R-D model to reduce the computational complexity. Simulation results demonstrate that the proposed algorithm shows excellent peak signal-to-noise ratio (PSNR) performance under the bit rate constraint. This one-pass rate control scheme is highly practical for the real-time video coding application.

Index Terms—Rate control, rate-distortion (R-D) model, video coding, video quality, H.264/AVC

1. INTRODUCTION

Rate control is essential to apply the modern video coding standard such as MPEG-2 [1], H.264/AVC [2] and AVS [3] to real-life application. Various video coding applications introduce strict bit rate constraint to the bit stream due to the limited transmission bandwidth or the storage size. Rate control scheme is responsible to achieve the bit budget by adjusting the quantization parameters (QPs) or trading the compressed video quality in a video encoder. Meanwhile, other optimizations are also needed to be considered such as system latency, buffer occupation and smoothness in objective or subjective video quality.

R-D models are widely used in rate control as the theoretical foundation in the literature which can be classified into two categories. The first category of R-D models assumes that the video coding units are independent with each other. Under this assumption, Chiang *et al*. [4] proposed a quadratic R-D model to calculate the target bit rate for each frame, which was adopted in both MPEG-4 and H.264/AVC. In [5], Ribas-Corbera *et al*. used a MB-level R-D model to choose the QP, which was adopted by H.263. He *et al*. [6] proposed a linear ρ-domain R-D model, which used the percentage of zero coefficients after quantization to approximate the bit rate. To tackle the inherent dilemma between rate control and R-D optimization (RDO) in H.264/AVC, Ma *et al*. [7] used the true quantization step size to establish the R-D model and proposed a rate control scheme with partial two-pass process at macroblock (MB) level.

The second category of R-D models takes the dependency introduced by the prediction process into consideration. Ramchandran *et al*. [8] provided a trellis-based solution for an arbitrary set of QPs for each coding unit. The computational complexity grew exponentially with the increase of dependent frame numbers. In [9], Lin *et al*. used interpolation to establish the approximated R-D curves. Liu *et al*. [10] analyzed the dependent temporal-spatial bit allocation problem and proposed two iteration algorithms to reduce the computational complexity. In scalable video coding, [11] proposed a GOP-based distortion model for different temporal layers according to the dependency between base layer and enhancement layer. The algorithms of [9][10][11] need to encode the source video several times, which are not suitable for real-time applications.

The above-mentioned methods are established by heuristic analyses and statistical examinations. However, the theoretical inter-dependent R-D model among different coding units needs to be further developed. In this paper, we set up the inter-dependent D-Q model and inter-dependent R-Q model based on the analysis of the relationship between the spatial-domain residual and the transform-domain residual. Then a window-based rate control scheme with

This work is partially supported by grants from the Chinese National Natural Science Foundation under contract No. 61171139 and No. 61035001, and National High Technology Research and Development Program of China (863 Program) under contract No. 2012AA011703.

complexity-based frame bit allocation and video quality optimization is proposed, which uses the approximated R-D model to reduce the computational complexity. Simulation results demonstrate that the proposed algorithm shows excellent PSNR performance under the bit rate constraint.

The rest of this paper is organized as follows. The theoretical analysis of the inter-dependent R-D model is given in Section II. Then the proposed window-based rate control scheme is presented in Section III. Section IV shows the experimental results of the rate control scheme. At last, we give a conclusion in Section V.

2. INTER-DEPENDENT RATE-DISTORTION MODELING

In this section, we establish the so-called inter-dependent R-D model, which is based on the theoretical analysis about the relationship between the spatial-domain residual and the transform-domain residual.

2.1. Inter-dependency problem analysis

Inter-frame prediction used in video coding increases the compression performance dramatically, meanwhile causing the dependency problem in R-D based rate control. Both distortion and bit rate of an inter frame (either P or B frame) will be affected by the QP variation of its reference frame (either I or P frame). For example, considering two frames with one referencing the other, the rate control problem can be formulated as minimizing the total distortion under the bit rate constraint. With the traditional coding-unit-independent assumption, the formulation is

$$\min_{Q_1,Q_2}(D_1(Q_1)+D_2(Q_2))$$
$$\text{such that } (R_1(Q_1)+R_2(Q_2)) \le R_{budget} \quad (1)$$

where Q_1, $D_1(Q_1)$, $R_1(Q_1)$ are the QP, distortion and bit rate of the first frame, Q_2, $D_2(Q_2)$, $R_2(Q_2)$ are the corresponding parameters of the second frame which gets the predicted value from the first frame. Actually, the distortion and bit rate of the second frame have strong dependency with its reference frame. Considering this inter-frame dependency, formulation (1) can thus be rewritten as

$$\min_{Q_1,Q_2}(D_1(Q_1)+D_2(Q_1,Q_2))$$
$$\text{such that } (R_1(Q_1)+R_2(Q_1,Q_2)) \le R_{budget}\,. \quad (2)$$

To address this dependency problem, a trellis-based solution is used in [8]. However, the real bit rate and distortion need to be obtained first. R-D model based solutions are proposed in [9][11], which need multi-pass encoding to derive the R-D models. These solutions are not suitable for real-time video coding applications since a "*forward*" bit allocation is usually needed in rate control scheme. To further reduce the computational complexity, we aim to establish the R-D model using the spatial-domain information which can be easily obtained in the pre-analysis process before the actual encoding is performed [6][12].

In the inter-frame coding process, the residual pixels of the second frame, which directly contribute to the bit rate and the distortion, are generated by the subtraction of the original pixels of the second frame, represented as $org_{2,(i,j)}$, and its reference pixels from the first frame. The reference pixels in the first frame are the reconstructed pixels which contain the distortion due to the quantization, represented as $rec_{1,(u,v)}$. We use the mean of absolute difference (MAD) at frame level to represent the residual information as

$$MAD_2 = \frac{1}{M \times N}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left|org_{2,(i,j)} - rec_{1,(u,v)}\right| \quad (3)$$

where MAD_2 is the real MAD of the second frame, M, N are the frame width and height respectively. The $rec_{1,(u,v)}$ can be calculated by the subtraction of the original pixels and the error which is the distortion and represented by $err_{1,(u.v)}$.

$$rec_{1,(u,v)} = org_{1,(u.v)} - err_{1,(u.v)} \quad (4)$$

From (3) and (4), we can get that

$$MAD_2 = \frac{1}{M \times N}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left|org_{2,(i,j)} - (org_{1,(u,v)} - err_{1,(u,v)})\right|$$
$$\approx \frac{1}{M \times N}\left(\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left|org_{2,(i,j)} - org_{1,(u,v)}\right| + \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} err_{1,(u,v)} - \beta\right)$$
$$= MAD\_O_2 + \frac{1}{M \times N}\left(\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} err_{1,(u,v)} - \beta\right) \quad (5)$$

where MAD_O_2 represents the MAD between the original pixels of the second frame and the first frame, which can be easily obtained by the pre-analysis. β has a direct relation with ($org_{2,(i,j)}$-$org_{1,(i,j)}$) which is less than zero. The second item of (5) is related to the distortion which is usually represented by the mean squared error (MSE) as

$$MSE_1 = \frac{1}{M \times N}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}(org_{1,(i,j)} - rec_{1,(i,j)})^2$$
$$= \frac{1}{M \times N}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} err_{1,(i,j)}^2 \quad (6)$$

Since the second item of (5), which only contains the distortion of the referenced pixels but not the whole frame, is partial to (6), we can use (6) as the approximation of it,

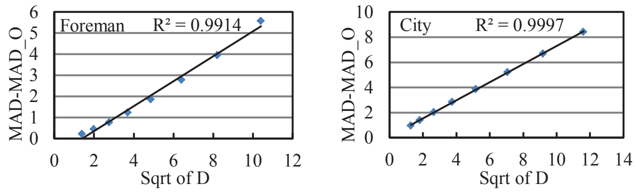
$$\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} err_{1,(u,v)} \approx \alpha \times \sqrt{M \times N \times MSE_1} \quad (7)$$

where α is the parameter that has a direct relationship with the MAD_ref_1/MAD_1, MAD_ref_1 represents the MAD of the referenced pixels in the first frame. Then, from (5) and (7), we can get that

$$MAD_2 = MAD\_O_2 + k \times \sqrt{D_1} + t \quad (8)$$

where k and t are model parameters. We test some video sequences to verify this relationship. The experimental results about the CIF and 720P sequences are shown in Fig. 1. The correlation between the square root of distortion of reference frame and the subtraction of MAD and MAD_O is larger than 0.99, which shows a good linear relation between them.

From (8), the relationship between the residual of the



(a) (b)

Fig. 1. Relationship about the distortion of the reference frame and the dependent residual. (a) *Foreman* (CIF), second frame, QP from 18 to 46; (b) *City* (720P), third frame, QP from 18 to 46.

second frame and the distortion of the first frame is established. We can get the real complexity information from the pre-analysis plus the distortion of the reference frame without the actual encoding. Cooperating with the D-Q model about the complexity and distortion as well as the R-Q model about the complexity and bit rate which will be described in the following subsection, the "*forward*" bit allocation and video quality optimization can be achieved only according to the spatial-domain residual information.

2.2. Inter-dependent D-Q model

Setting up the inter-dependent D-Q model contains two steps: 1) set up the relationship between the distortion, quantization step size and the true MAD. 2) combining it into (8) to get the inter-dependent D-Q model.

Modern hybrid video coding standards adopt discrete cosine transform (DCT) to convert the predicted residual block, represented by S, to a transformed matrix, represented by X, then use the quantization and entropy coding to achieve compression. The DCT process can be expressed as follows [13]

$$X = ASA^T \tag{9}$$

where T denotes transposition and

$$A(k,n) = \begin{cases} \dfrac{1}{\sqrt{N}}, k = 0, 0 \le n \le N-1 \\ \sqrt{\dfrac{2}{N}} \cos\dfrac{\pi(2n+1)k}{2N}, 1 \le k \le N-1, 0 \le n \le N-1 \end{cases} \tag{10}$$

where N is 4 in H.264/AVC.

Since the pixel values of S can be approximated by a Laplacian distribution with a zero mean and a separable covariance $r(m,n) = \sigma_S^2 \rho^{|m|} \rho^{|n|}$ [14], the variance of the (u,v)th DCT coefficient $\sigma_X^2(u,v)$ can be represented as [15]

$$\sigma_X^2(u,v) = \sigma_S^2 [ARA^T]_{u,u} [ARA^T]_{v,v} \tag{11}$$

where
$$R = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix} \tag{12}$$

ρ is the correlation coefficient and $[\bullet]_{u,u}$ denotes the (u,u)th component of the matrix. With ρ=0.6 as a typical value [14], we can get that

$$\sigma_X^2(u,v) = \sigma_S^2 C(u,v) \tag{13}$$

where $C(u,v)$ is a matrix with constants. With the Laplacian distribution and zero mean, the variance of S can be approximated by $\sigma_S=\sqrt{2}MAD$ [14]. This will lead to $\sigma_X(u,v)=\sqrt{2C(u,v)}MAD$. Assume the (u,v)th transformed coefficient $X(u,v)$ is Laplacian distributed as [16]

$$f_{X(u,v)}(x) = \frac{\lambda(u,v)}{2} e^{-\lambda(u,v)|x|} \tag{14}$$

where f denotes the probability density function (PDF) and

$$\lambda(u,v) = \frac{\sqrt{2}}{\sigma_{X(u,v)}} = \frac{1}{\sqrt{C(u,v)}MAD} \tag{15}$$

With PDF (14), we can calculate the distortion by summing up the distortion in each quantization interval as [17]

$$D(u,v) = \int_{-(Q-\gamma Q)}^{Q-\gamma Q} x^2 f_{X(u,v)}(x)dx + 2\sum_{n=1}^{\infty} \int_{nQ-\gamma Q}^{(n+1)Q-\gamma Q} (x-nQ)^2 f_{X(u,v)}(x)dx \tag{16}$$

where Q is the quantization step size, γQ denotes the rounding offset and γ is between (0,1), such as 1/6 in H.264/AVC [2]. Substituting (14) into (16), we can get [17]

$$D(u,v) = \frac{\lambda Q e^{\gamma\lambda Q}(2+\lambda Q - 2\gamma\lambda Q) + 2 - 2e^{\lambda Q}}{\lambda^2(1-e^{\lambda Q})} \tag{17}$$

where λ represents the $\lambda(u,v)$ in (15).

With (15) and (17), the relationship between distortion, quantization step size and MAD is established. However, this model is too complicated to be implemented. A simpler approximate model needs to be developed. Our approximation bases on two useful observations. First, in (17) there is an item of $1/\lambda^2$, which directly relates to MAD^2 according to (15). Second, it has been stated that the distortion has an approximate exponential relation with QP [7]. With $Q=2^{(QP-4)/6}$ in H.264/AVC, the distortion has a linear relation with Q. Based on these observations, we propose a simple yet accurate distortion model as

$$D = \alpha(MAD^2 + Q) + \beta \tag{18}$$

where α and β are model parameters. Substituting (8) into (18), we can further get the inter-dependent D-Q model. Here, an approximate formulation which is easy for applying is given as

$$D_2 = a(Q_2 + MAD\_O_2^2 + k^2 D_1) + b \tag{19}$$

where a and b are model parameters, k is the same as in (8).

To verify the accuracy of the dependent D-Q model (19), we tested some video sequences with variable resolutions and QPs. The model accuracy is represented as

$$\text{Accuracy}=\left(1-\frac{|\text{Estimated value-Actual value}|}{\text{Actual value}}\right)\times 100\% \tag{20}$$

Table 1 shows the detailed results. The average accuracy of the inter-dependent D-Q model is more than 90%.

Table 1. Verification of the inter-dependent D-Q model

Resolution	Sequence	QP	Accuracy
CIF	Foreman	28	94.2%
		36	92.6%
	News	28	90.5%
		36	91.3%
720P	City	32	93.4%
		40	95.2%
	Crew	32	91.6%
		40	92.1%

2.3. Inter-dependent R-Q model

The relationship between bit rate and quantization step size under the independent coding assumption has been studied in the literature. By assuming that the transformed coefficients are Laplacian [16] or Cauchy [18] distributed, the bit rate can be derived from calculating the entropy of the quantized DCT coefficients. However, these R-Q models are complicated and not suitable for rate control applications. To reduce the computational complexity, many R-Q models based on the relationship between bit rate and coding complexities are proposed for rate control scheme [4][5][7], where the coding complexity is usually represented by MAD or sum of absolute difference (SAD). For the low computational complexity purpose, our investigation of the inter-dependent R-Q model is based on the linear relationship between bit rate, SAD and quantization step size as

$$R = a_1 \frac{SAD}{Q} + b_1 \tag{21}$$

where SAD equals to $M\times N\times MAD$, a_1 and b_1 are the model parameters. The SAD in (21) is obtained in the actual frame coding process. By substituting (8) into (21), we can get the inter-dependent R-Q model directly as

$$R_2 = a_1 \frac{SAD\_O_2 + k\sqrt{D_1} + t}{Q_2} + b_1 \tag{22}$$

where SAD_O is the SAD between original frames similar to MAD_O. However, the item of $\sqrt{D}$ will also be introduced. To further simplify the R-Q model, we tested some sequences to statistically analyze (21) and the following R-Q model.

$$R = a_2 \frac{SAD\_O}{Q} + b_2 \tag{23}$$

where a_2 and b_2 are model parameters. Fig. 2. shows the R-Q curves of (21) and (23) for different test sequences. From the figure, an approximate linear relationship is observed between R and SAD_O/Q in (23), similar to that in (21). With the R-Q model (23), the bit rate can be estimated only by the SAD_O and quantization step size, while the distortion effect of the reference frame can be neglected. In other words, the complex inter-dependent R-Q issue is converted into a simple independent issue.

The implication contains two aspects. First, in high bit rate situation (QP is small), the distortion of the reference

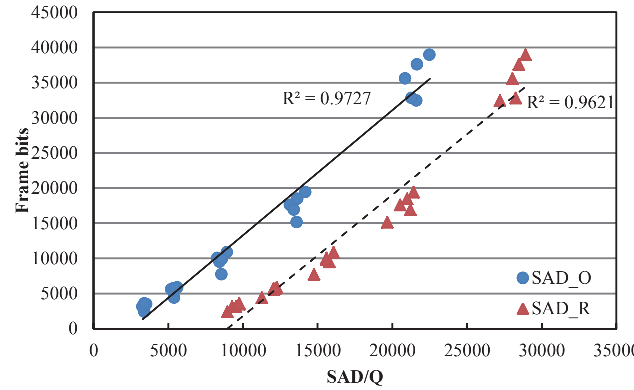


Fig. 2. R-Q curves of (21) and (23) from *Foreman* sequence.

frame is negligible so that $\sqrt{D_1}$ in (22) can be removed. (22) degenerates to (23) which is still a linear relationship. Second, in low bit rate situation (QP is large), $\sqrt{D_1}$ in (22) cannot be neglected. From (18), D_1 has a linear relationship with Q_1 so that $\sqrt{Q_1}$ is introduced to the numerator of (22). Assuming $Q_1 \approx Q_2$ for consistent video quality, we can get that the increase speed of numerator in (22) is slower than that of denominator because of the $\sqrt{Q}$ in the numerator and Q in the denominator. However, in low bit rate, the header bits occupy a significant portion of the total bits, and the percentage of header bits increases as the Q becomes larger. This will compensate for the slow increase speed of $\sqrt{Q}$ and the linear relationship will be held as (23).

3. PROPOSED RATE CONTROL SCHEME WITH INTER-DEPENDENT R-D MODEL

In this section, we propose a window-based rate control scheme with the inter-dependent R-D model described in the last section. The rate control scheme consists of the video quality optimization and complexity-based frame bit allocation, as presented in the following subsections.

3.1. Window-based rate control scheme

Traditional rate control schemes usually allocate the frame bit budget according to the different frame types, which will cause the video quality fluctuation if the frame details vary (high motion or scene change). To address this problem, we propose a bit allocation method among several adjacent frames, which is called a window, according to the coding complexity (represented by SAD). SAD of the current frame or future frames is usually obtained via the prediction of the previous encoded frames in the literature. This is under the assumption that the video content is stationary, which is not always true for different video sequences. To get more accurate SAD for the dependent R-D model, a pre-analysis is used in our scheme. For inter frames, only 16x16 motion search is used, while for intra frames, only few types of intra prediction (such as horizontal, vertical and diagonal) are

Table 2. Experimental results of different rate control algorithms

Format	Sequence	Bit rate (kbps)	RC in JM			ρ-domain RC			Proposed		
			Err (%)	PSNR	Var(D)	Err (%)	PSNR	Var(D)	Err (%)	PSNR	Var(D)
CIF	Foreman	1000	0.19	39.64	14.09	0.23	39.32	17.21	0.13	39.73	7.57
		500	0.20	36.86	75.85	0.25	36.51	87.53	0.15	36.95	42.37
	News	1000	0.08	45.21	0.59	0.09	44.91	0.75	0.04	45.32	0.50
		500	0.06	42.03	28.46	0.10	41.98	30.24	0.04	42.10	13.18
	Akiyo	1000	0.07	48.27	0.43	0.09	47.97	0.52	0.06	48.30	0.40
		500	0.09	45.62	13.11	0.12	45.60	13.57	0.07	45.71	10.47
720P	Night	8000	0.23	38.59	29.91	0.28	38.27	35.46	0.18	39.02	16.84
		5000	0.37	36.93	20.61	0.43	36.51	22.93	0.30	37.41	15.96
	Crew	8000	0.09	41.80	8.89	0.12	41.63	12.64	0.08	42.05	6.23
		5000	0.13	40.76	7.41	0.15	40.57	10.87	0.10	40.83	4.78
	Harbour	8000	0.08	37.07	33.53	0.15	36.74	42.67	0.08	37.35	21.76
		5000	0.12	35.26	21.21	0.17	34.91	26.31	0.12	35.42	16.28
	Average		**0.14**	**40.67**	**21.17**	**0.18**	**40.41**	**25.06**	**0.11**	**40.85**	**13.03**

used. The computational complexity of the pre-analysis is much lower compared to the actual encoding process.

The proposed window-based rate control scheme is performed as follows.

Step 1) Allocating bits for the window as

$$R_W = NR_C / F \qquad (24)$$

where R_W is the bit budget for the window, N is the number of frames in the window, R_C is the bit rate constraint for CBR and F is the frame rate. The window consists of the current encoding frame and the future (N-1) frames;

Step 2) Pre-analyzing the N frames in the window to get SAD_O for each frame;

Step 3) Calculating QP for the current encoding frame (the first frame in the window) according to the video quality optimization and coding complexity as described in the following subsections;

Step 4) Encoding current frame with QP decided in Step 3;

Step 5) Using the actual MAD and distortion of the current frame to update the parameters of the inter-dependent R-D model;

Step 6) Sliding the window by one frame, which means that excluding the encoded frame in Step 4 out of the window and including a new frame which is next to the last frame of the window;

Step 7) Pre-analyzing the last frame of the window, then go to Step 3.

The proposed window-based rate control scheme is totally a one-pass "*forward*" process, which is highly practical for the real-time video coding application. The detailed video quality optimization and complexity-based frame bit allocation are described as follows.

3.2. Video quality optimization

In video coding applications, the quality fluctuation degrades the viewing experience. Reducing the video quality fluctuation is also important in the rate control scheme. Considering the N frames in the window, it can be formulated as to minimize the variance of distortion as

$$\min(\mathrm{var}(D)) \text{ , where } \mathrm{var}(D) = \frac{1}{N}\sum_{i=0}^{N-1}(D_i - \overline{D})^2 \qquad (25)$$

D_i is the distortion of ith frame in the window and $\overline{D}$ denotes the average distortion of the N frames in the window.

Since the N frames are not encoded yet, it is obvious that the solution of (25) is

$$D_0 = D_1 = \cdots = D_{N-1} \qquad (26)$$

Substituting (26) into the dependent D-Q model (19), we can get the relationship between the quantization step size of ith frame Q_i and D_0 as

$$Q_i = (\frac{1}{a} - k^2)D_0 - MAD\_O_i^2 - \frac{b}{a} \qquad (27)$$

It should be noticed that D_0 can always be obtained by (18). If the window is the first window of the video sequence, we can use $MAD_0 \approx MAD\_O_0$ to calculate D_0. Otherwise, MAD_0 can be obtained by (8) since its reference frame has already been encoded. Thus, we can get that

$$D_0 = \alpha(MAD_0^{\ 2} + Q_0) + \beta \qquad (28)$$

Substituting (28) into (27), we can get the relationship between Q_i and Q_0 as

$$Q_i = \theta_i Q_0 + \tau_i \qquad (29)$$

where θ_i and τ_i are decided by the model parameter of (18) (19) and MAD_O_i.

3.3. Complexity-based frame bit allocation

The bit budget for the window has been obtained by (24). Then, we can allocate frame bits according to the coding complexity. By summing up the R-Q model (23) for each frame in the window, we can get that

$$R_W = \sum_{i=0}^{N-1}(a_2 \frac{SAD\_O_i}{Q_i} + b_2) \qquad (30)$$

Using the average quantization step size $\overline{Q}$ to replace Q_i, we can get that

$$\overline{Q} = a_2 \sum_{i=0}^{N-1} SAD\_O_i / (R_W - Nb_2) \qquad (31)$$

From (29) and (31), the quantization step size of the first frame in the window is obtained as

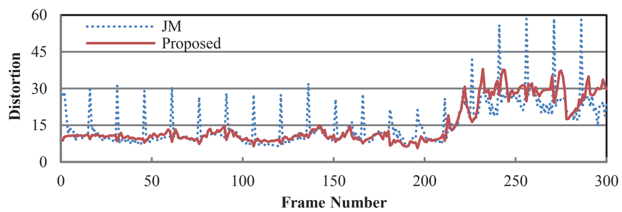


Fig.3. Frame distortion of *Foreman* sequence at 500kbps.

$$Q_0 = (N\overline{Q} - \sum_{i=0}^{N-1} \tau_i) / \sum_{i=0}^{N-1} \theta_i \tag{32}$$

Then QP of the first frame can be calculated by Q_0 and R_1 can be obtained by (23).

4. EXPERIMENTAL RESULTS

The proposed rate control scheme is implemented on the JM18.5 of H.264/AVC. Several video sequences are tested using the configuration as follow: IPPP coding structure with GOP length of 15, 2 reference frames, RDO on and CABAC, 30f/s frame rate. The frame number of the window is set to 5. Tested video sequences include CIF and 720P.

The rate control accuracy is represented by the bit rate mismatch between the target bit rate R_{target} and the actual bit rate R_{actual} as follows.

$$Err = \frac{|R_{target} - R_{actual}|}{R_{target}} \times 100\% \tag{33}$$

Table 2 shows the detailed experimental results about different rate control algorithms. The proposed rate control scheme has a better bit rate accuracy than the algorithm in JM and ρ-domain rate control [6].

The average PSNR of our rate control achieves about 0.18dB and 0.44dB gain compared with rate control in JM and ρ-domain rate control, respectively. The PSNR gain can be up to 0.48dB at "Night" sequence. The "*forward*" complexity-base frame bit allocation and the inter-dependent R-Q model contribute to the PSNR gain. Both JM and ρ-domain rate control allocate frame bits just according to the frame type, which cannot guarantee that the complicated frames have enough bit budget to get high PSNR. Our rate control scheme can make a better frame bit allocation to achieve higher PSNR.

Due to the video quality optimization and the inter-dependent D-Q model, the variance of distortion in the proposed rate control is also lower than JM and ρ-domain rate control, as shown in Table 2. The detailed frame distortion is also shown in Fig. 3. Without the distortion consideration, several peak points appear in the figure, which means a video quality fluctuation. On a contrary, the frame distortion of our rate control is much smoother.

5. CONCLUSION

In this paper, we set up the inter-dependent D-Q model and inter-dependent R-Q model based on the analysis of the relationship between the spatial-domain residual and the transform-domain residual. Then a window-based rate control scheme with complexity-based frame bit allocation and video quality optimization is proposed, which uses the approximated R-D model to reduce the computational complexity. Simulation results demonstrate that the proposed algorithm shows excellent PSNR performance under the bit rate constraint.

6. REFERENCES

[1] *ISO/IEC Video Recommendation ITU-T H.262, Generic Coding of Moving Pictures and Associated Audio Information,* MPEG2, 1995.

[2] *ISO/IEC JVT, ITU-T VCEG, ISO/IEC 14496-10 or ITU-T Rec. H.264, Doc. JVT-B118r8*, MPEG4/H.264, Feb. 1, 2002.

[3] *Information Technology Advanced Coding of Audio and Video Part2: Video. Standard Draft, (AVS)*, AVS-P2, Mar. 2005.

[4] T. Chiang and Y. Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 1, pp. 287–311, Feb. 1997.

[5] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 172–185, Feb. 1999.

[6] Z. He, Y. K. Kim, and S. K. Mitra, "Low-delay rate control for DCT video coding via ρ-domain source modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 8, pp. 928–940, Aug. 2001.

[7] S. W. Ma, and W. Gao, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 12, pp. 1533–1544, Dec. 2005.

[8] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 533–545, Sep. 1994.

[9] L. J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 4, pp. 446–459, Aug. 1998.

[10] S. Liu and C. C. J. Kuo, "Joint temporal-spatial bit allocation for video coding with dependency," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 1, pp. 15–26, Jan. 2005.

[11] J. Liu, Y. Cho, Z. Guo and C. C. Jay Kuo, "Bit allocation for spatial scalability coding of H.264/SVC with dependent rate-distortion analysis," *IEEE Trans. Circuits. Syst. Video Technol.*, vol. 20, no. 7, pp. 967–981, July 2010.

[12] L. Xu, D. Zhao, X. Ji, and W. Gao, "Window-level rate control for smooth picture quality and smooth buffer occupancy," *IEEE Trans. Image Processing*, vol. 20, no. 3, pp. 723-734, Mar. 2011.

[13] H. S. Malvar, A Hallapuro, M Karczewicz and L Kerofsky, "Low-complexity transform and quantization in H.264/AVC," *IEEE Trans. Circuits. Syst. Video Technol.*, vol. 13, no. 7, pp. 598–603, July 2003.

[14] I. M. Pao and M. T. Sun, "Modeling DCT coefficients for fast video encoding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 6, pp. 608–616, Jun. 1999.

[15] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[16] E. Lam and J. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1661–1666, Oct. 2000.

[17] X. Li, and A. Kaup, "Laplace distribution based Lagrangian rate distortion optimization for hybrid video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 2, pp. 193–205, Feb. 2009.

[18] N. Kamaci, Y. Altunbasak, and R. M. Mersereau, "Frame bit allocation for the H.264/AVC video coder via cauchy-density-based rate and distortion models," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 8, pp. 994–1006, Aug. 2005.