FRAME LEVEL RATE CONTROL FOR H.264/AVC WITH NOVEL RATE-QUANTIZATION MODEL

Sudeng Hu[1], *Hanli Wang*[2,*], *Sam Kwong*[1,*], *Tiesong Zhao*[1]

[1]Department of Computer Science, City University of Hong Kong, Hong Kong, China
[2]Department of Computer Science, Tongji University, Shanghai 201804, China
Email: sudenghu@gmail.com, hanli.wang@ieee.org,
cssamk@cityu.edu.hk, ztiesong2@student.cityu.edu.hk

ABSTRACT

In this paper, a frame level rate control algorithm is proposed with a novel Rate-Quantization (R-Q) model for H.264/AVC. Firstly, a two-stage rate control scheme is adopted to decouple the inter-dependency between Rate Distortion Optimization (RDO) and rate control. Secondly, in order to predict the frame complexity accurately, instead of the Mean Absolute Difference (MAD) of the residual signal, bits information in the RDO-based mode decision process is employed to predict the frame complexity. Thirdly, a self-adaptive exponential R-Q model is proposed for rate control. Experimental results reveal that the proposed R-Q model can estimate the actual output bits very well, and the novel rate control scheme has excellent performance both in bit rate accuracy and coding efficiency as compared to JVT-W043 and the FixedQp tool in the Joint Scalable Video Model reference software.

Keywords— Rate control, H.264/AVC, R-Q model.

1. INTRODUCTION

Rate Control (RC) is employed to regulate output bit stream according to the bandwidth limitation, meanwhile aiming to optimize the visual quality. Usually, this purpose is achieved through two steps. At the first step, efficient bit budget is distributed for each coding unit such as a frame or a macroblock (MB). At the second step, a proper quantization parameter is chosen to achieve this bit budget.

In order to reach the target bit accurately at the second step, several RC algorithms have been developed to model the R-Q relationship. In [1], based on the assumption that DCT coefficients are Laplacian distributed, a second-order quadratic model is derived. However it does not take frame complexity into consideration. In [2], the classic quadratic model is applied and the frame complexity is considered by predicting the MAD of encoding frame, which is implemented as the non-normative RC algorithm in H.264/AVC Joint Model (JM) reference softwares. In [3, 4, 5], simplified linear R-Q models are proposed which have comparable coding efficiency to the quadratic model. In [6], based on the assumption that DCT coefficients are Cauchy distributed, an exponential model is developed. In [7], the relationship between rate and quantization is built through ρ, which is the percentage of zeros among the quantized transform coefficients.

As stated above, the model parameters including the frame complexity in terms of MAD and zero coefficients' percentage ρ are crucial to construct accurate R-Q models. However, it is not trivial to obtain these parameters. Regarding MAD, the traditional way to characterize it is not very effective, since MAD is indirectly related to the number of output bits and thus it generally does not hold strong relationship with the actual output bit rate. Moreover, due to the inter-dependency between RC and RDO, the MAD of a coding frame is usually predicted from its previous frame, which further decreases the accuracy of the complexity estimation. On the other hand, as far as the ρ-domain method is concerned, although it has a relatively close relation with the consumed bit rate, more Discrete Cosine Transform (DCT) and quantization computations are required for model parameter generation.

In this paper, in order to improve the R-Q model performance, instead of utilizing MAD or ρ, the information of bit consumption in the RDO process is employed to predict the complexity of a coding frame. With this novel complexity measure, an improved exponential R-Q model is developed for RC realization. The rest of this paper is organized as follows. The frame complexity prediction with two-stage RC scheme is presented in Section 2. Then, the self-adaptive exponential R-Q model is introduced in Section 3, and the proposed overall RC algorithm is described therein. In Section 4, experimental results are presented to illustrate the efficiency of the proposed RC algorithm. Finally, conclusions are given in Section 5.

2. MEASUREMENT OF FRAME COMPLEXITY

2.1. Two-Stage RC Scheme

The RDO-based mode decision implemented in H.264/AVC improves the coding efficiency significantly as compared with pre-

This work is supported by the Hong Kong Research Grants Council General Research Fund, under Projects 9041236 (CityU 114707) and 9041353 (CityU 115408), and the German Alexander von Humboldt Research Fellowship. *Corresponding author.

vious coding standards. In mode decision, the best coding mode is selected with the minimal Rate-Distortion (R-D) cost according to [8]:

$$J = D_{mode} + \lambda R_{mode}, \qquad (1)$$

where D_{mode} and R_{mode} are the distortion and the number of output bits of a MB encoded in a specific INTER or INTRA mode; λ is the Lagrangian multiplier depending on quantization parameter Q_p. For each MB, all possible encoding modes are tried and the encoding mode with the minimal RD cost by minimizing J in Eq. (1) is chosen as the best.

The RDO-based mode decision process causes the so-called "Chicken-and-Egg" dilemma [9] in rate control for H.264/AVC, because Q_p is required before RDO process but residual signal and its related information is unavailable before that stage. In [10], a two-Q_p scheme is proposed to decouple the interdependency problem between RDO and rate control: one is used for mode decision and the other is employed in quantization. Theoretically, to reach the R-D optimization, it requires these two Q_p values to be equal. However, it is observed that a small mismatch between these two Q_p values will not decrease the coding efficiency significantly.

Inspired by [10], a two-stage frame level rate control scheme is developed in this work. At the first stage, a predefine Q_p denoted as Q_{p1} is adopted in RDO process for all MBs of a frame. At the second stage, the other Q_p denoted as Q_{p2} is calculated for quantization to generate the number of target bits. Before encoding the ith frame in a sequence, the $Q_{p1}(i)$ is set according to the previous Q_{p2} value. More specifically, it is set to $\hat{Q}_p(i)$ which is updated as

$$\hat{Q}_p(i) = w_q \cdot Q_{p2}(i-1) + (1 - w_q) \cdot \hat{Q}_p(i-1), \qquad (2)$$

where w_q is the weighting parameter, which is set to 0.7 based on experiments. For Q_{p2} at the second stage, it is calculated according to the R-Q model that will be discussed in Section 3.

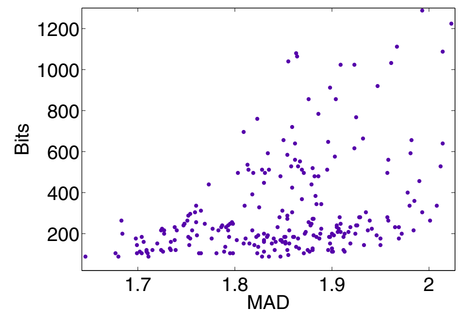
2.2. Proposed Frame Complexity Measurement

Due to the variant content of different frames, the frames even encoded with the same Q_p often produce quite different numbers of output bits. In order to build an accurate R-Q model, the MAD of residual signal is usually employed to describe the frame complexity in non-normative RC algorithms of video coding standards. The following model is considered as the classic R-Q model:

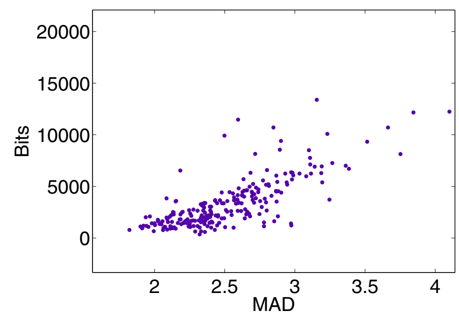
$$R = \alpha_1 \cdot \frac{MAD}{Q_{step}} + \alpha_2 \cdot \frac{MAD}{Q_{step}^2}, \qquad (3)$$

where α_1 and α_2 are model parameters and Q_{step} is the quantization step size. In this model, MAD is utilized to model the linear relation between complexity and the number of output bits. However, as illustrated in Fig. 1, this kind of linear relationship is not very strong and consequently it will decrease the accuracy of the R-Q model.

In order to estimate the frame complexity accurately, instead of using MAD, bits information available in the RDO-based



(a)



(b)

Fig. 1. The relationship between MAD and the number of output bits for frames. (a) "Container" sequence, QCIF format, 241 frames, $Q_p = 28$. (b) "Foreman" sequence, QCIF format, 241 frames, $Q_p = 28$.

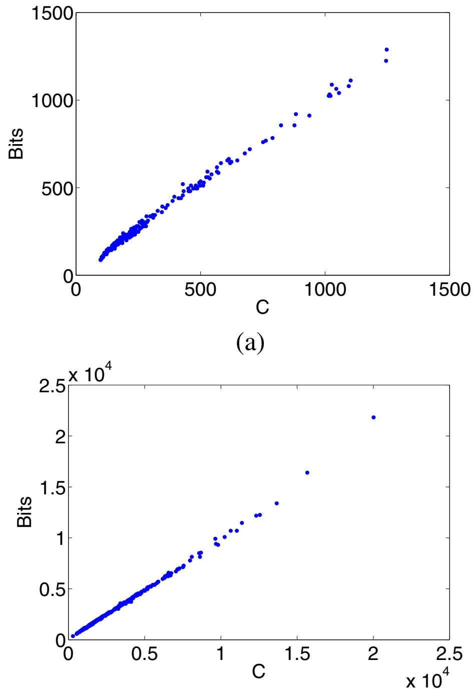
mode decision process is investigated to predict the frame complexity. This is because at the mode decision stage, the approximate number of output bits of each MB will be estimated and thus available. Let $R^h_{best}(i)$ and $R^t_{best}(i)$ denote the approximate number of header bits and texture bits for encoding the ith MB in the mode decision process, respectively, the complexity of the corresponding frame is defined as

$$C = C_h + C_t, \qquad (4)$$

where the header complexity C_h and texture complexity C_t are

$$C_h = \sum_{i=1}^{N} R^h_{best}(i), \; C_t = \sum_{i=1}^{N} R^t_{best}(i), \qquad (5)$$

in which N is the number of MBs inside a frame. Then, the proposed frame complexity C is used to derive the quantization parameter Q_{p2} for the subsequent quantization process to generate the actual output bits. In order to show the efficiency of the proposed frame complexity C, the relationship between C and the actual number of output bits is illustrated in Fig. 2, where it can be observed that there is a quasi-linear relationship between C and the actual number of output bits, and thus C is better than MAD for complexity predication purposes.



(a)

(b)

Fig. 2. The relationship between C and actual output bits for frames. (a) "Container" sequence, QCIF format, 241 frames, $Q_{p1} = Q_{p2} = 28$. (b) "Foreman" sequence, QCIF format, 241 frames, $Q_{p1} = Q_{p2} = 28$.

3. PROPOSED RC ALGORITHM WITH NOVEL R-Q MODEL

3.1. Exponential R-Q Model

Before quantization and entropy coding, the two-dimensional DCT is applied to reduce the spatial redundancy. The probability distribution of the DCT coefficients become important to build a reasonable R-Q model. Several distribution models are proposed to model the actual distribution of DCT coefficients, and among them the Cauchy distribution is reported to have better accuracy than other models in [6]. The Cauchy distribution-based R-Q model is expressed by

$$R_t = a \cdot Q_{step}^{-\beta}, \tag{6}$$

where a is the complexity related parameter and β is the model parameter associated with DCT coefficients distribution characteristics. In this model, a could be updated in the encoding process, while β is limited to a set of constant values according to different frame types, e.g., $\{0.75, 0.8, 0.85\}$ for I frame, $\{1.2, 1.4, 1.6\}$ for P frame, $\{1.6, 1.8, 2.0\}$ for B frame.

In the proposed two-stage RC algorithm, after the first stage, the model parameter a can be estimated according to texture complexity as

$$a = C_t \cdot Q_{step1}^{\beta}, \tag{7}$$

where Q_{step1} is the quantization step size corresponding to Q_{p1} used at the first stage. Therefore, the model of texture bits with

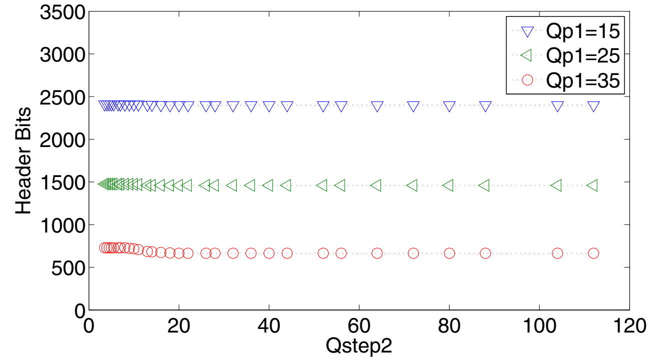


Fig. 3. The relationship between the number of header bits and Q_{step2}. The corresponding Q_{p2} is changed from 15 to 45 while Q_{p1} is set to 15, 25, and 35, respectively.

respect to Q_{step2} (which is the quantization step size corresponding to Q_{p2} used at the second stage) can be written as

$$R_t(Q_{step2}) = C_t Q_{step1}^{\beta} Q_{step2}^{-\beta}. \tag{8}$$

Since the number of header bits mainly comes from Motion Vectors (MVs), mode type and etc., which are decided at the first stage by Q_{p1}. Therefore, it is almost not affected by Q_{step2}, as shown in Fig. 3. Consequently, the total number of frame bits can be modeled as

$$R(Q_{step2}) = C_t Q_{step1}^{\beta} Q_{step2}^{-\beta} + C_h, \tag{9}$$

where R is the number of target bits for the current frame. The classic bit allocation scheme in JVT-W043 [11] is adopted in the current work to calculate the number of target bits R.

In this model, MAD is replaced by C_t as frame complexity for R-Q model. Figure 4 shows the relation between the actual number of output bits and the predicted number of bits by the proposed model with different β values. In Fig. 4, it is obvious that β is crucial to the accuracy of the model. However, when a proper β is chosen, the predicted number of bits matches the actual number of output bits well in a range of Q_{step} values.

3.2. Model Parameter Update

In the R-Q model in Eq. (9), the parameter β is related to distribution of AC DCT coefficients. Usually, it is a predefined constant parameter. However, the distribution of actual AC DCT coefficients of different frames varies significantly in different sequences or even in different frames of the same sequence. Moreover, as shown in Fig. 4, the β value is critical to the model. In the sequel, it is desirable to update β according to the frame characteristics in the encoding process. After coding the ith frame, the actual β for this frame can be calculated according to Eq. (8) as

$$\hat{\beta}(i) = \frac{\ln\left(\hat{R}_t(i)/C_t(i)\right)}{\ln\left(Q_{step1}(i)/Q_{step2}(i)\right)}, \tag{10}$$

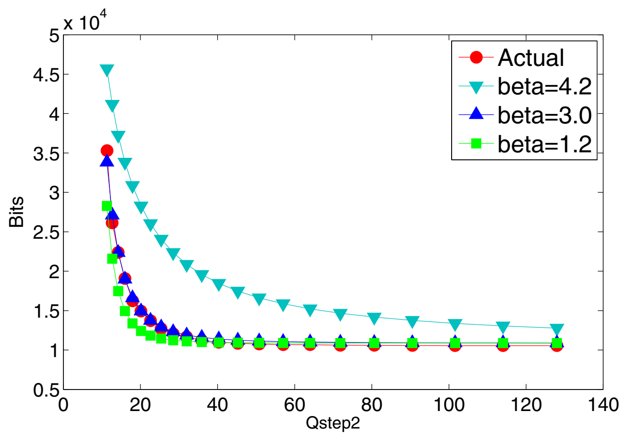


Fig. 4. The actual and predicted R-Q curves. Q_{p2} is changed from 25 to 46, and β is set to 1.2, 3.0, and 4.2 for the proposed R-Q model.

where $\hat{R}_t(i)$ is the number of actual textual output bits for the ith frame.

Since neighboring frames generally exhibit similar characteristics in a video sequence due to the temporal correlation, we assume the similarity in β for nearby frames. As a result, after coding the ith frame, β is updated and predicted for the $(i+1)$th frame as

$$\beta(i+1) = w_\beta \cdot \beta(i) + (1 - w_\beta) \cdot \hat{\beta}(i), \qquad (11)$$

where $\beta(i)$ is the predicted value for the ith frame; w_β is the weighting parameter with the typical value of 0.7 in this work.

3.3. Overall RC Algorithm

Based on the above analyses, the proposed overall RC algorithm includes two stages at the frame level, with the following step-by-step descriptions.

3.3.1. Stage One

Step 1. Q_{p1} is calculated as in Eq. (2).

Step 2. The calculated Q_{p1} is applied to all MBs in the mode decision process for the current frame.

Step 3. C_t and C_h are recorded.

Step 4. The model parameter β is predicted according to Eq. (11).

3.3.2. Stage Two

Step 5. Q_{p2} is calculated in the proposed R-Q model based on the number of target bits.

Step 6. Q_{p2} is clipped as

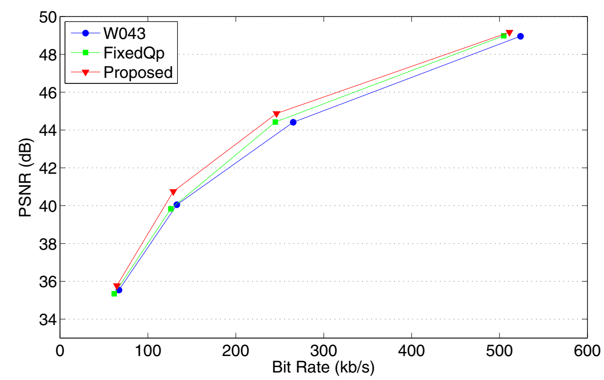
$$Q_{p2} = \max\{Q_{p1} - 3, \min\{Q_{p1} + 3, Q_{p2}\}\}. \qquad (12)$$

Step 7. Q_{p2} is applied in the quantization process for all MBs in the current frame. The number of actual texture output bits $\hat{R}_t$ is recorded, and the actual value of β is updated according to Eq. (10).

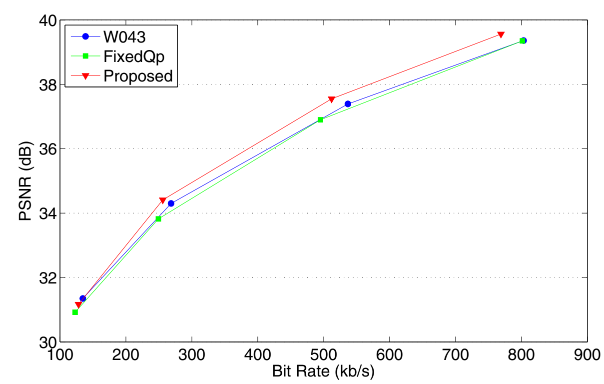
Step 8. Finish encoding the current frame, and process the next frame.

4. EXPERIMENTAL RESULTS

The proposed RC algorithm is implemented in the H.264/SVC-based Joint Scalable Video Model (JSVM) reference software 9.17 [12] where only a single layer is encoded to comply with H.264/AVC encoding conditions. The test frame rate is set to 30 fps. The GOP size is set to 4, where three B-frames are inserted between I/P frames. All video sequences have 241 pictures to be coded.



(a)



(b)

Fig. 5. R-D curves. (a) “Silent” sequence in QCIF format. (b) “Table” sequence in CIF format.

The algorithm JVT-W043 [11] and the FixedQp tool in JSVM are utilized for comparison with the proposed algorithm. In JVT-W043, the classic quadratical model is applied and the FixedQp tool is a multiple-pass RC algorithm where a logarithmic search is applied to find a proper Q_p to meet the target bit rate. Various benchmark video sequences are tested in both QCIF and CIF.

Table 1. Comparison of bit rate accuracy and PSNR.

		JVT-W043			FixedQp					Proposed				
Sequence	**T** (kb/s)	**R** (kb/s)	**PSNR** (dB)	**E** (%)	**R** (kb/s)	**PSNR** (dB)	**E** (%)	**BP** (dB)	**BB** (kb/s)	**R** (kb/s)	**PSNR** (dB)	**E** (%)	**BP** (dB)	**BB** (kb/s)
Akiyo	**64**	67.13	44.56	4.89	60.38	45.40	5.66	0.97	-15.4	63.74	44.94	0.41	1.20	-18.0
(QCIF)	**128**	135.70	48.15	6.02	127.00	49.01	0.78			127.60	48.66	0.31		
	256	266.90	51.68	4.26	251.60	52.06	1.72			255.40	53.02	0.23		
	512	522.80	56.70	2.11	496.50	56.82	3.03			500.40	57.98	2.27		
Foreman	**64**	66.65	32.84	4.14	63.05	32.84	1.48	0.44	-7.4	63.72	33.05	0.44	0.46	-7.8
(QCIF)	**128**	132.10	36.70	3.20	126.80	36.90	0.94			127.30	36.98	0.55		
	256	262.60	40.62	2.58	255.80	40.97	0.08			254.80	40.95	0.47		
	512	521.50	44.65	1.86	509.30	44.91	0.53			510.80	44.80	0.23		
Paris	**64**	66.93	32.92	4.58	58.38	32.37	8.78	0.26	-3.5	63.94	32.87	0.09	0.34	-4.6
(QCIF)	**128**	133.50	37.76	4.30	118.00	37.12	7.81			128.10	37.81	0.08		
	256	265.50	42.67	3.71	257.30	42.68	0.51			255.90	42.77	0.04		
	512	528.40	47.63	3.20	508.30	47.61	0.72			511.80	47.70	0.04		
Silent	**64**	67.03	35.54	4.73	61.68	35.34	3.63	0.32	-4.8	64.65	35.77	1.02	0.78	-11.4
(QCIF)	**128**	133.10	40.05	3.98	126.10	39.83	1.48			128.90	40.75	0.70		
	256	265.30	44.41	3.63	245.00	44.42	4.30			246.30	44.87	3.79		
	512	523.80	48.96	2.30	504.80	48.98	1.41			511.30	49.17	0.14		
Table	**64**	67.12	34.03	4.88	66.22	34.23	3.47	-0.16	3.2	64.17	34.01	0.27	0.29	-4.9
(QCIF)	**128**	134.20	38.08	4.84	125.10	37.38	2.27			128.10	38.03	0.08		
	256	267.50	41.91	4.49	252.60	41.20	1.33			244.40	41.57	4.53		
	512	565.50	45.71	10.45	508.80	45.56	0.62			512.90	46.01	0.18		
Average				**4.20**			**2.53**	**0.37**	**-5.6**			**0.79**	**0.62**	**-0.9**
Mobile	**128**	137.50	24.97	7.42	123.40	24.38	3.59	0.14	-3.2	128.50	24.93	0.39	0.20	-4.5
(CIF)	**256**	268.40	27.95	4.84	249.60	27.78	2.50			255.60	27.91	0.16		
	512	540.10	30.69	5.49	491.40	30.51	4.02			510.70	30.68	0.25		
	768	802.10	32.44	4.44	769.30	32.43	0.17			765.50	32.42	0.33		
Table	**128**	134.60	31.35	5.16	122.90	30.92	3.98	-0.11	2.6	128.10	31.17	0.08	0.31	-6.5
(CIF)	**256**	268.50	34.30	4.88	249.30	33.82	2.62			256.00	34.41	0.00		
	512	536.70	37.39	4.82	495.10	36.90	3.30			512.00	37.55	0.00		
	768	803.20	39.36	4.58	801.40	39.35	4.35			768.90	39.56	0.12		
Silent	**128**	134.70	33.80	5.23	120.30	33.47	6.02	0.28	-5.3	129.50	34.06	1.17	0.50	-9.3
(CIF)	**256**	268.00	37.18	4.69	256.00	37.20	0.00			258.80	37.43	1.09		
	512	531.00	40.63	3.71	508.10	40.77	0.76			513.50	41.07	0.29		
	768	795.80	42.81	3.62	777.70	42.94	1.26			771.30	43.24	0.43		
News	**128**	136.30	36.93	6.48	126.80	36.78	0.94	0.32	-6.3	128.10	36.83	0.08	0.55	-10.6
(CIF)	**256**	274.00	40.50	7.03	248.20	40.42	3.05			256.10	40.82	0.04		
	512	543.20	43.83	6.09	511.00	43.86	0.20			511.90	44.14	0.02		
	768	809.70	45.69	5.43	771.50	45.49	0.46			765.90	45.82	0.27		
Average				**5.24**			**2.33**	**0.28**	**-3.6**			**0.29**	**0.52**	**-8.7**

In order to evaluate the accuracy of bit rate achievement, the following measurement is used as

$$E = \frac{|R_t - R_o|}{R_t} \times 100\%, \tag{13}$$

where R_t and R_o are the target bit rate and actual output bit rate, respectively. The mismatch of bit rate at different test target bits by the comparative algorithms are also presented in Tables 1. As shown in Tables 1, the proposed algorithm can achieve much better performance than JVT-W043 and FixedQp in bit rate accuracy.

Since the output bit rate of these three RC algorithms are not matched exactly, the performances of BD-PSNR (denoted by BP) and BD-BR (denoted by BB) [13] are employed in our experiments for a fair comparison. JVT-W043 is set as the benchmark, and both of the FixedQp tool and the proposed algorithm are compared with JVT-W043 in BP and BB. From the experimental results in Tables 1, the FixedQp tool is about 0.37 dB in QCIF and 0.28 dB in CIF better than JVT-W043 on average, while the proposed algorithm is about 0.62 dB in QCIF and 0.52 dB in CIF better than JVT-W043.

The RD curves for two benchmark video sequences are given in Fig. 5, which illustrate the RD performance of the proposed RC algorithm is better than both JVT-W043 and FixedQp tool in a wide range of bit rates.

5. CONCLUSION

In this paper, a novel R-Q model is proposed with a two-stage RC scheme for H.264/AVC. The two-stage RC scheme is able to decouple the inter-dependency between RC and the RDO process. In addition, the proposed R-Q model utilizes the bit information available in the RDO process for effectively predicting the fame complexity, which is more robust and effective. The experimental results demonstrate that the RD performances can be improved significantly with the proposed RC algorithm as compared to JVT-W043 and the FixedQp tool under the test conditions.

6. REFERENCES

[1] 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. 246- 250, Feb. 1997.

[2] Z. Li, F. Pan, K. P. Lim, G. Feng, X. Lin, and S. Rahardja, "Adaptive Basic Unit Layer Rate Control for JVT", Doc. JVT-G012-r1, Pattaya, Thailand, Mar. 2003.

[3] Y. Liu, Z. G. Li, and Y. C. Soh, "A Novel Rate Control Scheme for Low Delay Video Communication of H.264/AVC Standard", *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 17, No. 1, pp. 68-78, Jan. 2007.

[4] S. Ma, W. Gao, and Y. Lu, "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.

[5] H. Wang and S. Kwong, "Rate-Distortion Optimization of Rate Control for H.264 with Adaptive Initial Quantization Parameter Determination", *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 18, No. 1, pp. 140-145, Jan. 2008.

[6] N. Kamaci, Y. Altinbasak, 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.

[7] Z. He and S. K. Mitra, "Optimum Bit Allocation and Accurate Rate Control for Video Coding via ρ-Domain Source Modeling", *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 12, No. 10, pp. 840-849, Oct. 2001.

[8] G. J. Sullivan and T. Wiegand, "Rate-Distortion Optimization for Video Compression", *IEEE Signal Process. Mag.*, Vol. 15, No. 6, pp. 23-50, Nov. 1998.

[9] Z. G. Li, F. Pan, K. P. Lim, and S. Rahardja ,"Adaptive Rate Control for H.264", in *Proc. IEEE Int. Conf. Image Process.*, pp. 449-452, Oct. 2004.

[10] D. Kwon, M. Shen, and C.-C. J. Kuo, "Rate Control for H.264 Video with Enhanced Rate and Distortion models", *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 17, No. 5, pp. 517-529, May 2007.

[11] A. Leontaris and A. M. Tourapis, "Rate Control for the Joint Scalable Video Model (JSVM)", Doc. JVT-W043, San Jose, USA, Apr. 2007.

[12] "Joint Scalable Video Model JSVM 9.17 Software Package", CVS server for JSVM reference softwares, Mar. 2009.

[13] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-Curves", Doc. VCEG-M33, Austin, USA, Apr. 2001.