A Frame-Layer Rate Control Algorithm for H.264 Using Rate-Dependent Mode Selection

Jun-Yup Kim, Seung-Hwan Kim, and Yo-Sung Ho

Gwangju Institute of Science and Technology (GIST),
1 Oryong-dong Buk-gu, Gwangju, 500-712, Korea
{jykim77, kshkim, hoyo}@gist.ac.kr

Abstract. Since the H.264 video coding standard becomes quite popular for transmission of diverse multimedia data due to its high coding efficiency, bit allocation and buffer management schemes are required for various applications. In this paper, we propose a new bit rate control algorithm for H.264 based on an accurate bit allocation strategy, where we introduce a reference quantization parameter (QP) table for the intra frame. For the inter frame, we propose a new bit allocation scheme by adjusting QP considering overhead bits and rate-dependent mode selection. Experimental results demonstrate that the number of the actual coding bits and the number of estimated bits match well such that we can avoid severe degradation of picture quality at the end of the group of pictures (GOP). Moreover, we have improved coding efficiency and average PSNR by up to 0.25dB.

Keywords: H.264 video coding, Rate control, Rate-dependent mode selection.

1 Introduction

Although rate control schemes are generally not included in the normative parts of video coding standards, they play important roles in video coding algorithms. Rate control is thus the necessary part of the encoder, and has been widely studied in several video coding standards, such as MPEG-2, MPEG-4, and H.263 [1], [2], [3], [4], [5]. However, since the quantization parameter (QP) should be determined before the rate distortion optimization (RDO) operation is applied in H.264, the study of rate control for H.264 is more difficult than those of previous standards.

Ma [6] proposed a rate control scheme based on the TM5 model of MPEG-2, assuming a linear relationship between rate and distortion. A rate control algorithm based on VM8 for MPEG-4 has been proposed and adopted by the H.264 test model [7]. The test model for H.264 employs the linear model for predicting the distortion of the current basic unit. In this scheme, the target bit is estimated based only on the buffer fullness, regardless of the current frame complexity. This may lead to drastic drops in PSNR values, especially in the case of high motion scenes or scene changes. In order to improve the video quality at scene changes, Jiang [8] introduced an MAD ratio as a measure of motion complexity. In his approach, a bit budget was allocated to each frame according to the MAD ratio of the current frame.

Generally, the rate control algorithm consists of three parts. The first part is to allocate an appropriate number of coding bits to each picture. The bit allocation process is

Y.-S. Ho and H.J. Kim (Eds.): PCM 2005, Part II, LNCS 3768, pp. 477 – 488, 2005.

constrained by the test model defined in the video coding standard. The second part is to adjust QP for each coding unit (e.g. the macroblock or the frame) in order to achieve the target bit rate constraint. In other words, the key point is to find the relation between the rate and QP. The final part is to update the coding parameters.

However, the rate control scheme for H.264 has some problems. The first problem is the selection of the initial QP. H.264 includes a different framework for motion estimation and motion compensation (ME/MC) using multiple reference frames and various coding modes. Hence, errors caused by an inappropriate initial QP are largely propagated in the whole frames of the group of pictures (GOP). The second problem is the abrupt change in PSNR values due to unsuitable bit allocation. The improper bit allocation results in a large fluctuation of PSNR values that degrades the subjective video quality significantly.

In this paper, we propose solutions to overcome those problems. In order to solve the initial QP problem, we have done extensive experiments, and have obtained the best QP range that depends on both the target bit rate and the image size. In order to avoid the abrupt fluctuation of PSNR values, we take an adaptive buffer control strategy by which we adjust the target buffer level using the distance from the intra frame. We select the rate-dependent optimum mode to reduce the bit consumption for the inter frame when the number of actual coding bits is greater than that of threshold bits.

This paper is organized as follows. After we describe the rate control scheme for estimating the target bits in Section 2, we propose a new bit allocation scheme for intra frame and inter frame in Section 3. Then, experimental results are presented in Section 4, and we conclude this paper in Section 5.

2 Target Bit Estimation for H.264

Before we discuss the rate control scheme, we need to determine the initial QP. In order to select the initial QP for H.264, we calculate

$$bpp = \frac{1.0 \times bit\_rate}{frame\_rate \times image\_width \times image\_height} \tag{1}$$

where *bpp* represents bit per pixel. The initial QP is then selected from 10, 20, 25 and 35 using the *bpp* value of Eq. (1). Figure 1 illustrates the decision of the initial QP according to *bpp*. However, the above specific initial QP is not adaptive and the range between the initial QPs is much large. In Section 3, we propose a new method for the initial QP selection.

Generally, the frame-layer rate control scheme for H.264 consists of two stages: pre-encoding and post-encoding. The objective of the pre-encoding stage is to determine QP for all frames and it has two sub-steps: (a) determine a target bit for each P-frame and (b) compute QP using a quadratic rate-distortion (R-D) model and perform RDO. In the post-encoding stages, we update the model parameters continually and control frame-skipping.

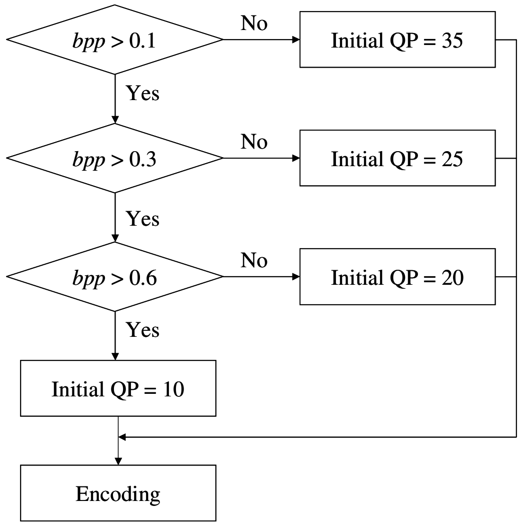


Fig. 1. Flow Chart for Selection of the Initial QP

In order to estimate target bits for the current frame, we employ a fluid traffic model based on the linear tracking theory [9]. For explaining simply, we assume one GOP consisting of first I-frame and subsequent P-frames. Let N denote the total number of frames in a GOP, n_j $(j=1,2,\ldots,N)$ denote the j^{th} frame, and $B_c(n_j)$ denote the occupancy of the virtual buffer after coding the j^{th} frame. The fluid traffic model is described by

$$B_c(n_{j+1}) = \min\{\max\{0, B_c(n_j) + A(n_j) - \frac{u(n_j)}{F_r}\}, B_s\}$$
$$B_c(n_1) = \frac{B_s}{8} \tag{2}$$

where $A(n_j)$ is the number of bits generated by the j^{th} frame, $u(n_j)$ is the available channel bandwidth, F_r is the predefined frame rate, and B_s is the buffer size. In order to determine the target bits for the current P-frame, we take the following two steps.

STEP 1: Allocate the bit budget among pictures. Since the QP of the first P frame is given at the GOP layer in this scheme, the initial target buffer level $Tbl(n_2)$ is set by

$$Tbl(n_2) = B_c(n_2) \tag{3}$$

Then the target buffer levels of other P-frames in the GOP are predefined by

$$Tbl(n_{j+1}) = Tbl(n_j) - \frac{Tbl(n_2) - B_s/8}{N_p - 1} \tag{4}$$

where N_p is the total number of P-frames in the GOP.

STEP 2: Calculate the target bit rate. Using the linear tracking theory, we can determine the number of target bits allocated for the j^{th} frame based on the target buffer level, the frame rate, the available channel bandwidth, and the actual buffer occupancy

$$T_{buf}(n_j) = \frac{u(n_j)}{F_r} + \gamma(Tbl(n_j) - B_c(n_j)) \tag{5}$$

where γ is a constant and its typical value is 0.75. Meanwhile, the average value of remaining bits is also computed by

$$T_r = \frac{R_r}{N_r} \tag{6}$$

where R_r is the number of bits remaining for encoding, and N_r is the number of P frames remaining for encoding. Therefore, the estimated target bit $T_{estimated}$ is a weighted combination of T_{buf} and T_r

$$T_{estimated}(n_j) = \beta \times T_r(n_j) + (1-\beta) \times T_{buf}(n_j) \tag{7}$$

3 New Frame-Layer Rate Control

3.1 Bit Allocation for Intra Frame

Like the rate control of MPEG video coding standards, the frame-layer rate control for H.264 can be used to encode the video sequence by the unit of GOP. Each GOP consists of at least one intra frame and other frames which are motion compensated from the intra frame directly or indirectly. Hence, the quality of the intra frame determines coding efficiency of all other frames in the same GOP through motion compensation. Especially, since H.264 includes good ME/MC processes, the quality of the intra frame is more important than that of the previous video coding standards. In other words, if we encode the intra frame with high quality, we have improved the picture quality of succeeding frames in the same GOP by allocating lower bits.

Figure 2 shows the importance of selecting the initial QP value for the intra frame. As shown in Fig. 2, we have improved coding performance by up to 0.91dB with the optimal initial QP.

In order to choose the optimal initial QP, *bpp* is calculated with the I-frame and the initial P-frame because the I-frame and the first P-frame deal with still images.

$$bpp_{\text{mod}} = \frac{a \times bit\_rate}{b \times image\_width \times image\_height} \tag{8}$$

where a is a constant and the typical value is 0.4, and b is $1+\varepsilon$ ($0<\varepsilon\leq 1$) which represents the sum of weighting factors for the I-frame and the initial P-frame.

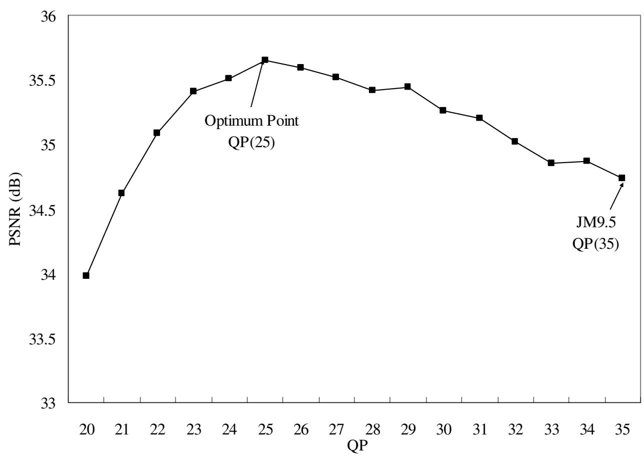


Fig. 2. Initial QP and Average PSNR Value for NEWS Sequence in 64kbps

Table 1. Initial QP and bpp_{mod} Value

Initial QP	bpp_{mod}	Initial QP	bpp_{mod}
10	2.7548927	23	0.7803819
11	2.5195049	24	0.7021517
12	2.2879709	25	0.6440709
13	2.1126631	26	0.5706939
14	1.9089462	27	0.5178741
15	1.7430161	28	0.4701441
16	1.6042719	29	0.4167456
17	1.4442603	30	0.3821154
18	1.3093566	31	0.3501026
19	1.2066104	32	0.3106850
20	1.0757050	33	0.2785275
21	0.9686448	34	0.2536564
22	0.8777515	35	0.2247080

In order to decide the initial QP based on bpp_{mod}, we introduce a reference table that is derived from extensive experiments with various test sequences. For simplicity, we set the b to two which represents that the weighting factor ε for the initial P-frame is one. Table 1 lists the relationship between the QP and bpp_{mod}.

From Table 1, we can choose the initial QP adaptively and improve coding efficiency by allocating a proper number of bits for the I-frame and the initial P-frame. In Table 1, we limit the range of the initial QP from 10 to 35 because the other range is not appropriate for practical applications.

3.2 Bit Allocation for Inter Frame

3.2.1 Improved Buffer Control

In Eq. (4), the target buffer level is controlled by the uniform allocation. However, each frame has a different weighting value according to the temporal distance from the reference I-frame. In other words, since early P-frames are used as references by many subsequent P-frames in the same GOP, P-frames that are closer to the reference I-frame should be allocated with more target bits than other P-frames.

In order to control the buffer adaptively for H.264, we employ the linear weighting function [10] for the target buffer level. Let n_j be the distance between P-frame and its reference I-frame, N_p be the GOP length, and σ be an adjustable parameter. Then, improved target buffer control is defined by.

$$Tbl(n_{j+1}) = Tbl(n_j) - \frac{Tbl(n_2) - B_s/8}{N_p - 1} + \Delta Tbl(n_j) \tag{9}$$

$$\Delta Tbl(n_j) = \sigma \frac{N_p - 2n_j}{N_p - 2}, \qquad n = 1, \cdots, N_p - 1 \tag{10}$$

where σ is equal to one third of the average target buffer level.

3.2.2 Optimal Bit Allocation for Inter Frame

The rate control scheme for H.264 has a problem that the difference value between the estimated coding bits and the actual coding bits is not reflected efficiently when QP is determined for the inter frame.

Figure 3 shows that the estimated coding bits and the actual coding bits do not match well in the AKIYO sequence when the GOP size is 30. This problem causes that all the target bits are abruptly consumed and picture quality of remaining frames is seriously degraded.

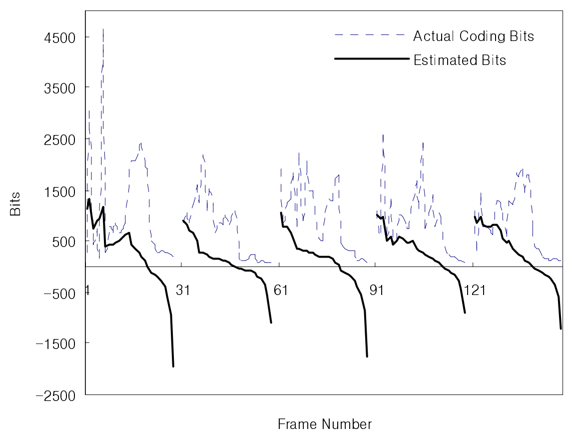


Fig. 3. Estimated Bits and Actual Coding Bits for Akiyo (48kbps)

Our rate control algorithm consists of two steps for this mismatch problem.

STEP 1: Consider the overhead bits when determining QP. The size of the overhead bits for mode and motion information is abruptly changed in H.264 depending on various coding modes. Therefore, the estimated target bits need to reflect the change of both the texture bits and the overhead bits efficiently.

Let $T_{overhead}(j\text{-}1)$ denote the bits used for motion and header information of previous frame, and $T_{texutre}(j\text{-}1)$ denote the number of bits for texture(i.e., residual) coding of previous frame. Then, we distribute estimated target bits $T_{estimated}(j)$ to $T_{overhead}(j)$ and $T_{texutre}(j)$, effectively

$$T_{estimated}(j) = T_{texture}(j) + T_{overhead}(j) \tag{11}$$

If we assume $T_{overhead}(j) = T_{overhead}(j\text{-}1)$, the quantization step $Q_{step}(j)$ of the current frame is computed based on a quadratic R-D model

$$\begin{aligned} T_{texture}(j) &= T_{estimated}(j) - T_{overhead}(j-1) \\ &= X_1 \frac{MAD}{Q_{step}(j)} + X_2 \frac{MAD}{Q_{step}^2(j)} \end{aligned} \tag{12}$$

where X_1 and X_2 are the first-order and second-order model parameters respectively, and MAD is calculated by a linear model.

In the proposed scheme, we adjust $QP(j)$ simply by adding 2 when $T_{estimated}(j)$ is smaller than $T_{overhead}(j\text{-}1)$. In other case, $QP(j)$ is calculated to the corresponding quantization step $Q_{step}(j)$ with the method provided by H.264.

$$QP(j) = \begin{cases} QP(j-1)+2, & T_{estimated}(j) < T_{overhead}(j-1) \\ round(6\log_2 Q_{step}(j)) + 4, & otherwise \end{cases} \tag{13}$$

In order to maintain the smoothness of visual quality among successive frames, we restrict the $QP(j)$ as follows

$$\begin{aligned} QP(j) &= Max\{QP(j-1) - \Delta QP,\ QP(j)\} \\ QP(j) &= Max\{QP(j-1) + \Delta QP,\ QP(j)\} \end{aligned} \tag{14}$$

where ΔQP is the varying range of QP and its typical value is 2.

STEP 2: Adjust the fine-granular QP with the rate-dependent mode selection. In the RDO process, we examine all coding modes for every macroblock and calculate the rate (R) and the distortion (D) for every mode. Hence, the mode with the minimum cost (J) is selected as the optimum mode for every macroblock. However, we control the rate for each macroblock by selecting other modes before actual encoding. We propose a rate-dependent mode selection, where we select the mode by considering the remaining bits

$$J = D + \lambda \times R \tag{15}$$

where the Lagrangian multiplier (λ) is

$$\lambda = 0.85 \times 2^{\frac{QP-12}{3}} \tag{16}$$

In the rate-dependent mode selection, we do not select the optimum mode when the actual coding bits exceed the threshold value composed of estimated bits and the weighting factor. In other words, if the number of actual coding bits for the optimum mode is larger than the threshold value, we select the second optimum mode that has minimum J among the restricted modes whose rates are less than rates of the optimum mode.

Figure 4 shows the flow diagram of the proposed algorithm with two steps. In the proposed scheme, k represents a weighting factor and its typical value is 1.2.

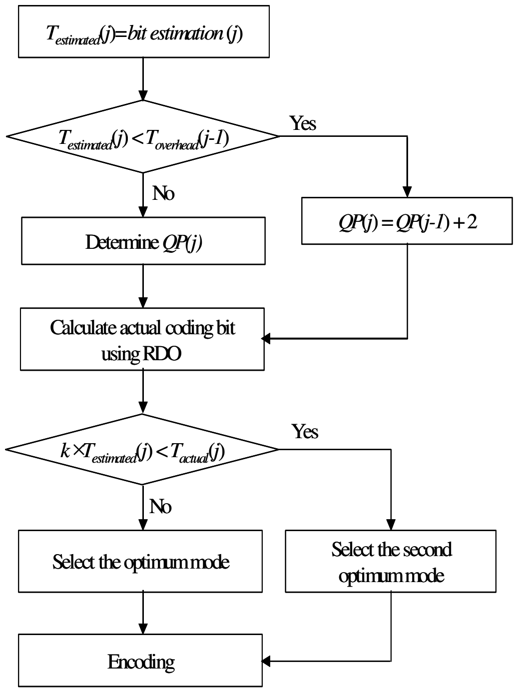


Fig. 4. Flow Diagram of the Proposed Algorithm

4 Experimental Results and Analysis

In order to evaluate performance of the proposed scheme, we have implemented the proposed rate control scheme on the JM9.5 test model software and have compared it with the original JM9.5. The initial QP is set as shown in Table 1, and we experiment 150 frames. Test conditions are given in Table 2.

Table 2. Test Conditions

MV Resolution	1/4
Hadamard	ON
RD Optimization	ON
Search Range	±16
Restrict Search Range	2
Reference Frames	5
Symbol Mode	CAVLC
GOP Structure	IPPP
Intra Period	30
Image Format	QCIF (176×144) CIF (352×288)

Figure 5 depicts the estimated bits and the actual coding bits for the AKIYO sequence. The actual coding bits are well matched to the target bits in the proposed algorithm, and the target bits are efficiently distributed in the whole frames of GOP.

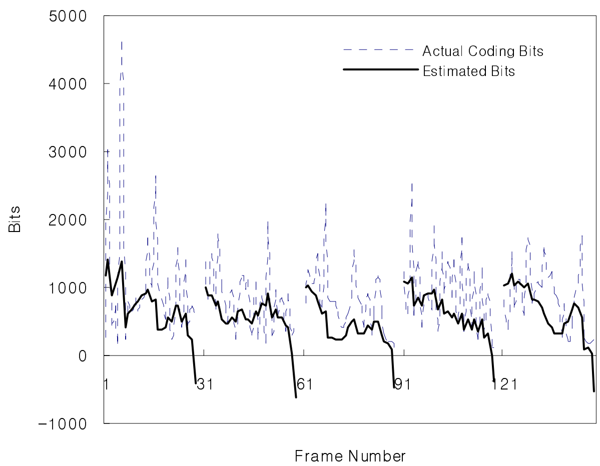
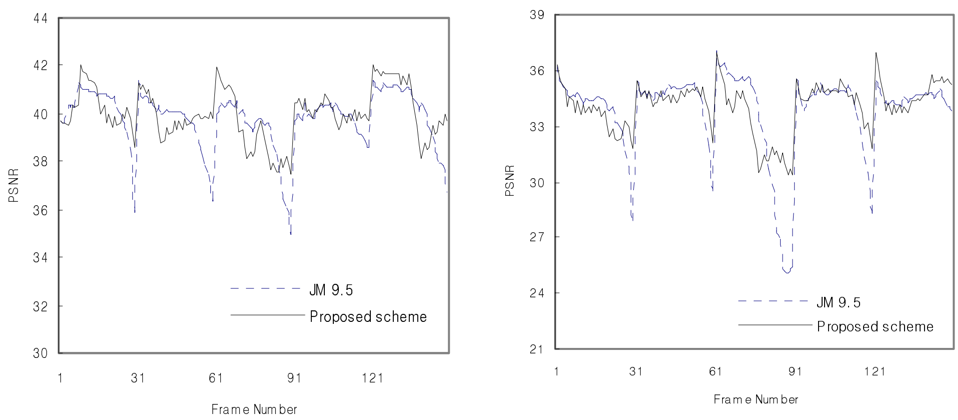


Fig. 5. Estimated Bits and Actual Coding Bits for Akiyo in the Proposed Scheme (48kbps)

Table 3 compares the average PSNR values with JM9.5. As shown in Table 3, we have improved coding efficiency and average PSNR by up to 0.25dB. It also shows the both algorithms can achieve accurate target bit rates.

Table 3. Comparison of the proposed Algorithm with JM9.5

Sequences	Frame Rate	Average PSNR(dB)			Bit Rate(kbps)	
		JM9.5	Proposed	Gain	JM9.5	Proposed
AKIYO	30	36.76	37.21	+0.45	32.78	32.57
SILENT	30	32.88	33.11	+0.23	48.19	48.23
NEWS	30	33.64	33.74	+0.10	48.33	48.41
CARPHONE	30	33.82	34.09	+0.27	48.31	48.39
FOREMAN	30	32.52	32.72	+0.20	64.75	64.64



(a) AKIYO (b) CARPHONE

Fig. 6. PSNR Values for AKIYO and CARPHONE Sequences (48kbps)

Figure 6 represents PNSR values in the AKIYO and CARPHONE sequences in 48kbps. Because of the efficient distribution of the target bits, picture quality of the successive frames does not change abruptly in the proposed algorithm.

In order to evaluate the various environments according to both the variable bit rate and image size, we experiment the CIF image format (352×288). Figure 7 shows rate distortion curve for the FOREMAN sequence. As shown in Fig. 7, the proposed algorithm outperforms the JM9.5 in the all the bit rate. Especially, we have improved better coding efficiency and PSNR value at the low bit rate.

Table 4 compares the standard deviation of PSNR values with JM9.5. In order to evaluate the effect on the human visual system, we employ the standard deviation for PSNR values. As shown in Table 4, the proposed scheme has a smaller variance which is better effect on the human visual system in comparison with JM9.5. Especially, video sequences, such as CARPHONE and FOREMAN, which have low temporal correlation and high complexity between frames, have better visual quality.

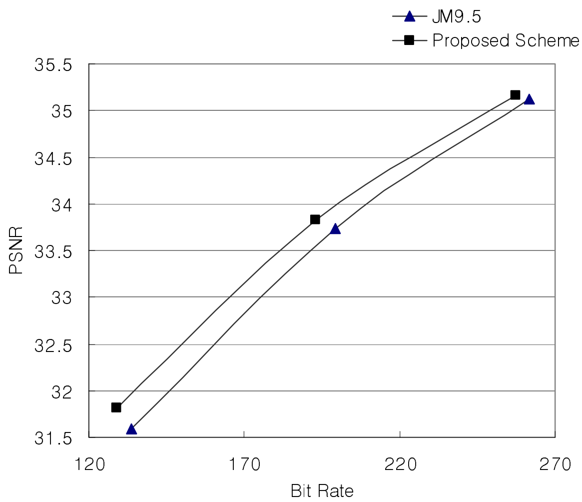


Fig. 7. PSNR Values for AKIYO and CARPHONE Sequences (48kbps)

Table 4. Comparison of the Proposed Algorithm with JM9.5

Sequences	Standard Deviation	
	JM9.5	Proposed
AKIYO	1.439	1.082
SILENT	1.702	1.099
NEWS	1.633	1.289
CARPHONE	2.371	1.293
FOREMAN	2.513	1.844

5 Conclusions

In this paper, we proposed a new rate control algorithm based on accurate bit allocation for H.264 video coding. In the proposed scheme, we introduced the reference table for the intra frame and proposed an optimal bit allocation by adjusting QP such that the number of actual coding bits is as close as possible to the target number of bits, while maintaining the uniform picture quality. In order to determine QP, we consider the overhead bits and control the actual coding bits with the rate-dependent mode selection. Since the number of target bits and the number of actual coding bits for a picture match well in our proposed algorithm, severe degradation of picture quality can be avoided at the end of GOP.

Acknowledgements. This work was supported in part by Gwangju Institute of Science and Technology (GIST), in part by the Ministry of Information and Communication (MIC) through the Realistic Broadcasting Research Center (RBRC), and in part by the Ministry of Education (MOE) through the Brain Korea 21 (BK21) project.

References

1. ISO/IEC JTC1/SC29/WG11 MPEG/N0400: MPEG-2 Test Model5 (1993)
2. Chiang, T., Zhang, Y.Q.: A New Rate Control Scheme Using Quadratic Rate Distortion Model. IEEE Transaction on Circuit and System for Video Technology, Vol. 7, No 1 (1997) 246-250
3. Vetro, A., Sun, H., Wang, Y.: MPEG-4 Rate Control for Multiple Video Objects. IEEE Transaction on Circuit and System for Video Technology, Vol. 9, No 1 (1999) 172-185
4. Lee, H.J., Chiang, T.H., Zhang, Y.Q.: Scalable Rate Control for MPEG-4 Video. IEEE Transaction on Circuit and System for Video Technology, Vol. 10, No 6 (2000) 878-894
5. Corbera, J., Lei, S.: Rate Control for Low-delay Video Communications. ITU Study Group 16. Video Coding Experts Group Documents Q15-A-20 (1997)
6. Ma, S., Gao, W., Wu, F., Lu, Y.: Rate Control for AVC Coding scheme with HRD Considerations. Proceedings of International Conference on Image Processing (2003) 14-17
7. ISO/IEC TC JTC1/SC29 MPEG/N5821: Draft ISO/IEC 14496-5:2002 /PDAM6. (2003)
8. Jiang, M.Q., Yi, X.Q., Ling, N.: Improved Frame-Layer Rate Control for H.264 Using MAD Ratio. Proceedings of IEEE International Symposium on Circuits and Systems, Vol. 3 (2004) 813-816
9. Li, Z.G., Xiao, L., Zhu, C., Pan, F.: A Novel Rate Control Scheme for Video Over the Internet. Proceedings of International Conference on Acoustics, Speech, and Signal Processing, Vol. 2, (2002) 2065-2068
10. Pan, F., Li, Z.G., Lim, K., Feng, G.: A Study of MPEG-4 Rate Control Scheme and Its Implementations. IEEE Transaction on Circuit and System for Video Technology, Vol. 13, No. 5 (2003) 440-446