Available online at www.sciencedirect.com

SCIENCE DIRECT®

J. Vis. Commun. Image R. 17 (2006) 376–406

JOURNAL OF VISUAL Communication & IMAGE Representation

www.elsevier.com/locate/jvci

Adaptive rate control for H.264

Z.G. Li [a,*], W. Gao [b], F. Pan [a], S.W. Ma [b], K.P. Lim [a], G.N. Feng [a], X. Lin [a], S. Rahardja [a], H.Q. Lu [b], Y. Lu [b]

[a] *Media Division, Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613, Singapore*

[b] *Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China*

Received 29 April 2004; accepted 30 April 2005
Available online 11 August 2005

Abstract

This paper presents a rate control scheme for H.264 by introducing the concept of basic unit and a linear prediction model. The basic unit can be a macroblock (MB), a slice, or a frame. The linear model is used to predict the mean absolute differences (MADs) of the remaining basic units in the current stored picture by those of the co-located basic units in the previous stored picture. The target bits for the current stored picture are computed by adopting a fluid flow traffic model and linear tracking theory, and are further bounded by two values that are derived by taking the hypothetical reference decoder (HRD) into consideration. The remaining bits are allocated to the remaining basic units in the current stored picture according to their predicted MADs. The corresponding quantization parameter is computed by using a quadratic rate-distortion model. The rate distortion optimization (RDO) is then performed for all MBs in the current basic unit by the quantization parameter. Both constant bit rate and variable bit rate cases are studied. The average PSNR is improved by up to 0.8 dB for an encoder using our scheme compared to an encoder using a fixed quantization parameter. With our scheme, an H.264 encoder can be adaptive to time varying channel bandwidth that is available for the coding process.

Keywords: H.264; Rate control; Adaptive; Basic unit

* Corresponding author.
E-mail address: ezgli@i2r.a-star.edu.sg (Z.G. Li).

1047-3203/$ - see front matter
doi:10.1016/j.jvcir.2005.04.004

1. Introduction

Video compression has made steady progress over the last 20 years, with 2X bit-rate reduction every 5 years. With more and more applications, video coding moves beyond mere compression. A joint source-networking coding will provide the best solutions for these applications [1]. One of the key issues for the joint source-networking coding is the bit rate adaptation of the coding system. An encoder employs rate control as a way to regulate varying bit rate characteristics of the coded bit-stream to produce high quality decoded pictures at bit rates that are provided by the network. Specifically, the rate control is used to compute quantization parameters for the current frame and the number of skipped frames according to the specified bit rate and the statistics of the current frame, like mean absolution difference (MAD), mean squared error (MSE), and head bits of each predefined unit (which can be a macroblock (MB), a frame, or an object) [2–4,6,14]. Rate control is thus a necessary part of an encoder. There are five key parts on rate control: (1) a rate-distortion (R-D) model at the predefined unit level. The R-D model sets up a relationship among quantization parameter, the statistics of the current unit, and the available number of bits; (2) the frame level bits allocation by using the statistics of the current frame and the coded frames and the available channel bandwidth; (3) the unit level bits allocation by using the statistics of all units in the current frame; (4) quantization parameter of each instantaneous decoding refresh (IDR) picture, which is determined by the length of group of pictures (GOP) and the available channel bandwidth; and (5) the number of skipped frames.

The rate control has been widely studied in standards, like MPEG 2, MPEG 4, H.263, and so on [2–4,6,14]. These scheme focused on the case that the available channel bandwidth for the coding process is constant. However, this cannot be guaranteed by the current Internet with the best effort service or the Internet with the relative differentiated quality of service (QoS) [5]. It is thus necessary to design a rate control scheme such that a video encoder can be adaptive to time varying channel bandwidth that is available for the coding process. Meanwhile, the rate control in H.264 is more complex than TM5, Q2, and TMN 8 in the sense the statistics of the current frame is available in TM 5, Q2, and TMN 8 while it is not available for the rate control in H.264 [2,3,14]. This is because that the quantization parameters are involved in both rate control and rate distortion optimization (RDO) in H.264 while it is only involved in rate control in MPEG 2, MPEG 4, and H.263. There exists a typical chicken and egg dilemma when the rate control is implemented for H.264: to perform RDO for an MB, a quantization parameter should be first determined for the MB by using the MAD of the current MB. However, the MAD of the current MB is only available after performing the RDO. An intuition method is to predict the statistics of each MB in the current frame by that of co-located MB in the previous frame. However, there exist problems associated with the prediction at the MB level. For example, imagine the case that an MB in the current frame is an Intra-MB while the co-located MB is an Inter-one. Thus, Li et al. [9] provided a frame layer rate control scheme for H.264 by introducing a linear model to predict the MAD of the current stored picture by the actual MAD of the previous stored picture, which solves the chicken and egg dilemma that is mentioned above. Jiang et al. [11] proposed an interesting scheme in to improve video distortion, due to scene change, by more accurately predict frame complexity using the statistics of previous encoded frames. However, both schemes focused on the frame level and the bit fluctuation may be too large.

To solve this problem, we introduce the concept of basic unit, which can be either a slice, or a frame. All MBs in the same basic unit share a common quantization parameter [12]. The linear model is adopted to predict the MADs of the remaining basic units in the current stored picture by the actual MADs of the co-located basic units in the previous stored picture. With the linear model and the concept of basic unit, our scheme is described in detail as follows: the target bits for the current frame are computed by utilizing a fluid flow traffic model and linear tracking theory, and are determined by the frame rate, the current buffer occupancy, the target buffer level and the available channel bandwidth. The target bits are further bounded by two values that are derived by taking the hypothetical reference decoder (HRD) into consideration. The remaining bits for the current frame are allocated to the remaining basic units according to their predicted MADs. A quadratic R-D model is used to compute the corresponding quantization parameter, which is then used to perform the RDO for each MB in the current basic unit. We focus on the case that the available channel bandwidth for the coding process is time varying while our scheme is also applicable to the case that the bandwidth is constant. The available channel bandwidth can be estimated by feedback information provided by the receiver [5]. Compared to an H.264 encoder using fixed quantization parameter, our scheme can improve the average PSNR up to 0.8 dB. The improved average PSNR is 0.47 dB for all H.264 test sequences under normal test condition. The major contributions of this paper are listed as follows:

1. Introduce the concept of basic unit to obtain a good trade-off between bit fluctuation and coding efficiency;
2. Propose a linear model to predict the MADs of the remaining basic units in the current stored picture by those of the co-located basic units in the previous stored picture and solve the chicken and egg dilemma;
3. Provide an effective buffer regulation scheme which can be utilized in both CBR case and VBR case while the existing schemes focus on the CBR case;
4. Propose an adaptive scheme to compute the quantization parameter of the IDR picture in each GOP. With our scheme, the quantization parameter is adaptive to both the GOP length and the available channel bandwidth.

The rest of this paper is organized as follows: the concept of basic unit and two models are given in the following section. GOP layer rate control scheme, frame layer rate control scheme, and basic unit layer rate control scheme are presented in Sections 3–5, respectively. Section 6 contains extensive experimental results to illustrate the efficiency of our scheme. Finally, concluding remarks are given in Section 7.

2. Preliminary knowledge

In this section, we shall first present the problem associated with the rate control for H.264.

2.1. The chicken and egg dilemma

The coding process of a MB that is related to the rate control is illustrated in Fig. 1 and is given by [15,7]

$$\text{Statistics of current frame} \to \text{Rate control} \to \text{Quantization parameter}$$
$$\to \text{RDO} \to \text{Statistics of current frame} \to \text{Coding}.$$

Since quantization parameters are involved in both rate control and RDO, there exists a dilemma when the rate control is implemented: to perform RDO for an MB, a quantization parameter should be first determined for the MB by using the MAD of the MB and the number of header bits. However, the MAD of the current MB and the number of header bits are only available after performing the RDO. This is a typical chicken and egg dilemma. Because of this, the rate control for H.264 is more difficult than those for MPEG 2, MPEG 4, and H.263.

To study the rate control for H.264, we should find a way to estimate the MAD of the current MB. Besides this, we also need to compute target bits for the current frame and to determine the number of contiguous MBs that share a quantization parameter. To solve the problems, we also need the following preliminary knowledge.

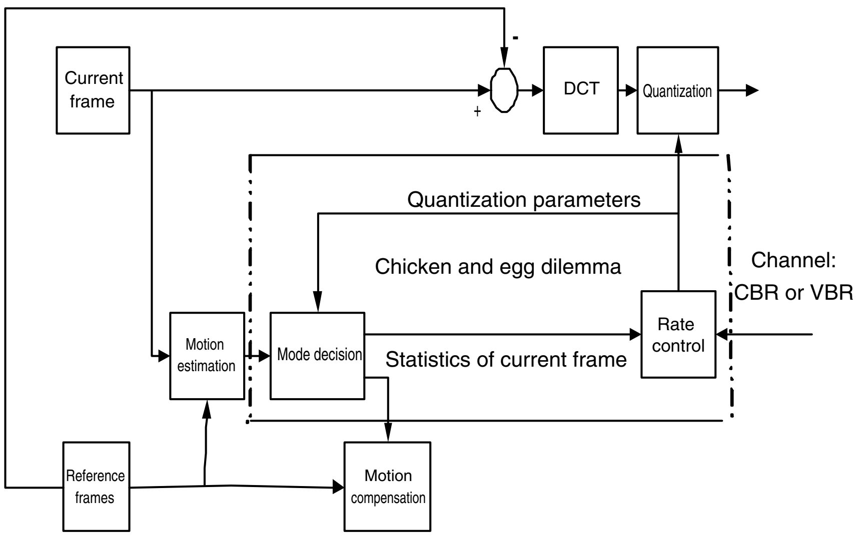


Fig. 1. The coding process of a MB related to the rate control.

2.2. The definition of basic unit

The concept of basic unit is defined by [7].

Definition 1. Suppose that a frame is composed of N_{mbpic} MBs. A basic unit is defined to be a group of contiguous MBs which is composed of N_{mbunit} MBs where N_{mbunit} is a fraction of N_{mbpic}.

Denote the total number of basic units in a frame by N_{unit}, which is computed by

$$N_{\text{unit}} = \frac{N_{\text{mbpic}}}{N_{\text{mbunit}}}. \tag{1}$$

All MBs in the same basic unit share a common quantization parameter. It should be mentioned that the authors of [3] tried to achieve a similar objective at lower bit rate by setting weight to each MBs. The concept of the basic unit can also be extended to study the rate control for MPEG 2, MPEG 4, and H.263.

2.3. A fluid flow traffic model

We shall now present a fluid flow traffic model to compute the target bits for the current coding picture. Let N_i denote the total number of frames in the ith group of picture (GOP), and $V_i(j)(i = 1, 2, \ldots, j = 1, 2, \ldots, N_i)$ denote the occupancy of virtual buffer before coding the jth frame. It can be shown from the fluid flow traffic model that [6]

$$\begin{cases} V_i(j+1) = \max\{V_i(j) + b_i(j) - \frac{R_i(j)}{f}, 0\} \\ V_1(1) = V_s/8 \\ V_{i+1}(1) = V_i(N_i) \end{cases}, \tag{2}$$

where $b_i(j)$ is the coded bits of the jth frame in the ith GOP, $R_i(j)$ is the available channel bandwidth which can be either constant or time varying, f is the predefined frame rate, and V_s is the buffer size whose maximum value is determined based on different level and different profile [15].

2.4. A linear MAD prediction model

Suppose that the $(jL + 1)$th picture is a stored picture and the number of successive non-stored pictures between two stored pictures are L. The chicken and egg dilemma can be solved by a linear model

$$\tilde{\delta}_{l,i}(jL+1) = a_1\delta_{l,i}(jL - L + 1) + a_2, \tag{3}$$

where $\tilde{\delta}_{l,i}(jL + 1)$ is the predicted MAD of the lth basic unit in the current stored picture in the ith GOP, $\delta_{l,i}(jL - L + 1)$ is the actual MAD of the lth basic unit in the previous stored picture, a_1 and a_2 are two coefficients of the prediction model. The initial values of a_1 and a_2 are set to 1 and 0, respectively. They are updated after coding each basic unit [7].

It should be mentioned that $\delta_{l,i}(jL - L + 1)$ is a reference value for the prediction. There are many other choices for the reference value. For example, 16×16 based motion estimation and compensation can be proceeded for all MBs in the current picture to obtain the rough knowledge of the MADs and the numbers of header bits for them. We can also use the most possible mode used in the previous frame to perform motion estimation and compensation to get the rough information. The resulting MAD can be used as the reference value. For simplicity, we choose $\delta_{l,i}(jL - L + 1)$ as the reference value. Similarly, there also exist many models for the prediction and for simplicity, we choose the linear model.

2.5. A quadratic rate-distortion (R-D) model

Our quadratic mode is derived from that in [2,4] by taking the quantization scheme of H.264 into consideration, and is provided by

$$\hat{b}_{l,i}(jL+1) = c_1 \frac{\tilde{\delta}_{l,i}(jL+1)}{Q_{\text{step},l,i}(jL+1)} + c_2 \frac{\tilde{\delta}_{l,i}(jL+1)}{Q^2_{\text{step},l,i}(jL+1)}, \tag{4}$$

where $\hat{b}_{l,i}(jL+1)$ is the number of texture bits for the lth basic unit, c_1 and c_2 are two coefficients of the quadratic model, and $Q_{\text{step}}, l, i(jL+1)$ is the quantization step for the lth basic unit. The relationship between the quantization parameter, $QP_{l,i}(jL+1)$, and the quantization step, $Q_{\text{step}}, l, ijL + 1$ is given as follows [15]:

$$Q_{\text{step},l,i}(jL+1) = 2^{QP_{l,i}(jL+1)/6} * d(QP_{l,i}(jL+1)\%6), \tag{5}$$

$$d(0) = 0.675; d(1) = 0.6875; d(2) = 0.8125; d(3) = 0.875; d(4) = 1;$$
$$d(5) = 1.125. \tag{6}$$

It should be mentioned that the quantization parameter instead of the quantization step is used in the quadratic R-D model in [2,4].

2.6. HRD conformance

A lower bound and an upper bound for the target bits of each frame are determined by taking the hypothetical reference decoder (HRD) into consideration. The lower bound and the upper bound for the jth frame in the ith GOP are denoted by $Z_i(j)$ and $U_i(j)$, respectively.

Let $t_{r,i}(j)$ denote the remove time of the jth frame in the ith GOP from the coded picture buffer (CPB). Also let $be(t)$ be the bit equivalent of a time duration t with the conversion factor being the buffer arrival rate [15]. Obviously, $be(t)$ has the following property:

$$be(t_2) - be(t_1) \leqslant (t_2 - t_1)R; t_2 \geqslant t_1 \geqslant 0, \tag{7}$$

where R is the maximum channel bandwidth within the interval $[t_1, t_2]$.

The upper bound and the lower bound are initialized as follows [8]:

$$Z_i(1) = B_{i-1}(N_{i-1}) + \frac{R_i(1)}{f}, \tag{8}$$

$$U_i(1) = B_{i-1}(N_{i-1}) + be(t_{r,1}(1))\omega, \tag{9}$$

where $B_{i-1}(N_{i-1})$ is the remaining bits of the $(i-1)$th GOP and $B_0(N_0) = 0$. The value of ω is 0.9.

The successive bounds $U_i(j)$ and $Z_i(j)$ are computed iteratively as [10]

$$Z_i(j) = Z_i(j-1) + \frac{R_i(j)}{f} - b_i(j-1), \tag{10}$$

$$U_i(j) = U_i(j-1) + \left(\frac{R_i(j)}{f} - b_i(j-1)\right)\omega. \tag{11}$$

It can be easily shown that a rate control scheme conforms to the HRD if two conditions

$$t_{r,1}(1)R \leqslant V_s \tag{12}$$

$$Z_i(j) + be(t_{r,1}(1)) - V_s \leqslant b_i(j) \leqslant \frac{R_i(j)}{f} \tag{13}$$

are satisfied.

With the concept of the basic unit, models (2) and (3), our scheme works step by step as follows:

1. Compute the target bits for the current frame by using the fluid traffic model (2), linear tracking theory [13] and the bounds $U_i(j)$ and $Z_i(j)$.
2. Predict the MADs of the remaining basic units in the current stored picture by using linear model (3) and the actual MADs of the co-located basic units in the previous stored picture.
3. Allocate the remaining bits to the remaining basic units in the current stored picture according to the predicted MADs.
4. Compute the corresponding quantization parameter by using the quadratic R-D model (4) and (6).
5. Perform RDO for each MB in the current basic unit by the quantization parameter derived from step 4 [15,16].
6. Update the parameters a_1 and a_2 of model (3) and c_1 and c_2 of model (4).

Clearly, the chicken and egg dilemma is solved, even though it may not be the best solution. Our rate control consists of three levels: the GOP level rate control, the frame level rate control and an optional basic unit layer rate control. The details are provided in the following three sections.

3. GOP-level rate control

In this level, the total number of bits allocated to each GOP is computed and the initial quantization parameter of each GOP is set.

3.1. Total number of bits

The initial value of bits allocated for the ith GOP is computed as

$$B_i(1) = \frac{R_i(1)}{f} N_i + B_{i-1}(N_{i-1}). \tag{14}$$

It can be shown from (14) that the coding results of the latter GOPs depend on those of the former GOPs. To ensure that all GOPs have a uniform quality, each GOP should use its own budget.

Since the channel bandwidth may vary at any time, B_i is updated frame by frame as below.

$$B_i(j) = B_i(j-1) - b_i(j-1) + \frac{R_i(j) - R_i(j-1)}{f}(N_i - j + 1); \quad j = 2, 3, \ldots, N_i. \tag{15}$$

In the case of the CBR, i.e., $R_i(j) = R_i(j-1)$, Eq. (15) is simplified as

$$B_i(j) = B_i(j-1) - b_i(j-1)\ ;\ j = 2, 3, \ldots, N_i. \tag{16}$$

In other words, Eq. (15) is also applicable to the CBR case.

3.2. Initial quantization parameter

In our scheme, the initial quantization parameter of the ith GOP, $QP_i(1)$, is predefined based on the available channel bandwidth and the GOP length. Normally, a small $QP_1(1)$ should be chosen if the available channel bandwidth is wide and a big $QP_1(1)$ should be used if it is narrow. The IDR picture and the first stored picture of the GOP are coded by $QP_i(1)$. The other $QP_i(1)$ is computed by

$$QP_i(1) = \max\left\{\min\left\{\frac{\sum_{j=1}^{N_p(i-1)} \bar{Q}P_{i-1}(jL+1)}{N_p(i-1)} - \min\left\{2, \frac{N_{i-1}}{15}\right\}, QP_{i-1}(1) + 2\right\}, QP_{i-1}(1) - 2\right\}, \tag{17}$$

where $N_p(i-1)$ is the total number of stored pictures in the $(i-1)$th GOP and $\bar{Q}P_{i-1}(jL+1)$ is the average value of all quantization parameters of the jth stored picture in the $(i-1)$th GOP. $QP_i(1)$ is further adjusted by

$$QP_i(1) = QP_i(1) - 1; \quad \text{if} QP_i(1) > QP_{i-1}(N_p(i-1)L+1) - 2, \tag{18}$$

where $QP_{i-1}(N_p(i-1)L+1)$ is the quantization parameter of the last stored picture in the $(i-1)$th GOP. Clearly, $QP_i(1)$ is adaptive to both the GOP length and the available channel bandwidth.

The IDR picture and the first stored picture in each GOP are coded by $QP_i(1)$ if the basic unit is chosen as a frame. Otherwise, only the IDR picture in each GOP is coded by $QP_i(1)$.

4. Frame level rate control

The frame level rate control scheme consists of two stages: pre-encoding and post-encoding.

4.1. Pre-encoding stage

The objective of this stage is to compute quantization parameters of all frames. For simplicity, different methods are provided to compute the quantization parameters of the stored picture and the non-stored picture.

4.1.1. Quantization parameters of non-stored pictures

To simplify our algorithm, the quantization parameters of non-stored pictures are obtained through a linear interpolation method as follows:

Suppose that the $(jL+1)$th and the $(jL+L+1)$th frames are stored pictures and the quantization parameters of two adjacent stored pictures are $QP_i(jL)$ and $QP_i(jL+L+1)$, respectively. The quantization parameter of the kth $(1 \leqslant k \leqslant L)$ non-stored picture is given according to the following two cases:

Case 1. When $L=1$, there is only one non-stored picture between two stored pictures. The quantization parameter is computed by

$$QP_i(j+1) = \begin{cases} \frac{QP_i(j)+QP_i(j+2)+2}{2} & \text{if} \quad QP_i(j) \neq QP_i(j+2) \\ QP_i(j)+2 & \text{Otherwise} \end{cases}. \tag{19}$$

Case 2. When $L>1$, there are more than one non-stored picture between two stored pictures. The quantization parameters are computed by

$$\begin{aligned} QP_i(jL+k) &= QP_i(jL+1)+\alpha+\max\left\{\min\left\{\frac{(QP_i(jL+L+1)-QP_i(jL+1))(k-1)}{L-1},\right.\right. \\ &\left.\left. 2(k-1)\right\}, -2(k-1)\right\}, \quad k=2,3,\ldots,L, \end{aligned} \tag{20}$$

where α is a function of the difference between the quantization parameter of the first non-stored picture and $QP_i(j)$, and is given as

$$\alpha = \begin{cases} -3 & QP_i(jL+L+1) - QP_i(jL+1) \leqslant -2L-3 \\ -2 & QP_i(jL+L+1) - QP_i(jL+1) = -2L-2 \\ -1 & QP_i(jL+L+1) - QP_i(jL+1) = -2L-1 \\ 0 & QP_i(jL+L+1) - QP_i(jL+1) = -2L \\ 1 & QP_i(jL+L+1) - QP_i(jL+1) = -2L+1 \\ 2 & \text{Otherwise} \end{cases}. \tag{21}$$

The cases that $QP_i(jL + L + 1) - QP_i(jL + 1) < -2L + 1$ can only occur at time instant that the video sequence switches from one GOP to another GOP. It can be verified that $|QP_i(jL + k + 1) - QP_i(jL + k)| < 2$.

The final quantization parameter $QP_i(jL + 1 + k)$ for the non-stored picture is further bounded by 0 and 51.

Remark 1. Since the non-stored pictures are not used to predict other pictures, relatively larger quantization parameters are chosen for them to save more bits for the stored pictures. The overall prediction gain is thus improved. Meanwhile, the smoothness of visual quality is also kept by setting the maximum value of the difference between quantization parameters of two adjacent pictures as 2. □

4.1.2. Quantization parameters of stored pictures

The quantization parameters of stored pictures are computed by the following two steps:

Step 1. Determine target bits for each stored picture.

The bits that are allocated to the current stored picture should be adjusted according to the current buffer occupancy and the picture complexity. Step 1 is composed of the following two sub-steps:

Step 1.1. Predefine a target buffer level for each stored picture in the current GOP.

After coding the first stored picture in the ith GOP, the value of the target buffer level is initialized as

$$S_i(2) = V_i(2), \tag{22}$$

where $V_i(2)$ is the actual buffer occupancy after coding the first stored picture in the ith GOP.

Suppose that the $(jL + 1)$th picture is a stored picture. The target buffer level for the picture is determined by

$$S_i(jL+1) = S_i(jL-L+1) - \frac{S_i(2) - V_s/8}{N_p(i) - 1} + \frac{\bar{W}_{p,i}(jL-L+1)(L+1)R_i(jL+1)}{f(\bar{W}_{p,i}(jL-L+1) + \bar{W}_{b,i}(jL)L)} - \frac{R_i(jL+1)}{f}. \tag{23}$$

In the above equation, $\bar{W}_{p,i}(jL - L + 1)$ is the average complexity weight of stored pictures that have been coded, $\bar{W}_{b,i}(jL)$ is the complexity weight of non-stored pictures that have been coded. They are computed as follows:

$$\bar{W}_{p,i}(jL - L + 1) = \frac{W_{p,i}(jL - L + 1)}{8} + \frac{7\bar{W}_{p,i}(jL - 2L + 1)}{8}, \tag{24}$$

$$\bar{W}_{b,i}(jL) = \frac{W_{b,i}(jL)}{8} + \frac{7\bar{W}_{b,i}(jL - 1)}{8}, \tag{25}$$

$$W_{p,i}(jL - L + 1) = b_i(jL - L + 1)\bar{QP}_i(jL - L + 1), \tag{26}$$

$$W_{b,i}(jL) = \frac{b_i(jL)\bar{QP}_i(jL)}{1.36}, \tag{27}$$

where $\bar{QP}_i(j)$ is the average value of quantization parameters in the jth frame, W_p and W_b are the weight of the current/previous stored picture and non-stored picture.

In the case that there is no non-stored picture between two stored pictures, i.e., $L = 1$, Eq. (23) can be simplified as

$$S_i(j + 1) = S_i(j) - \frac{S_i(2) - V_s/8}{N_p(i) - 1}. \tag{28}$$

It can be easily shown that $S_i(N_i)$ is about $V_s/8$. Thus, if the actual buffer fullness is exactly the same as the predefined target buffer level, it means that each GOP uses its own budget. However, since the rate-distortion (R-D) model and the linear MAD prediction model are not accurate [2,4], there usually exists difference between the actual buffer occupancy and the target buffer level. We therefore need to compute target bits for each stored picture to reduce this difference. This is achieved by the following sub-step.

Step 1.2. Compute target bits for the current stored picture.

Using model (2) and linear tracking theory [13], the target bits allocated for the jth stored picture in the ith GOP are determined based on the target buffer level, the frame rate, the available channel bandwidth and the actual buffer occupancy as follows:

$$\tilde{T}_i(jL + 1) = \frac{R_i(jL + 1)}{f} + \gamma(S_i(jL + 1) - V_i(jL + 1)),$$

where γ is a constant, its typical value is 0.125.

If the number of generated bits is around $\tilde{T}_i(jL + 1)$, it can be easily shown that

$$V_i(jL + 2) - S_i(jL + 2) \approx (1 - \gamma)(V_i(jL + 1) - S_i(jL + 1)). \tag{29}$$

Therefore, a tight buffer regulation can be achieved by choosing a large γ. However, a small γ is selected in our scheme to improve video distortion that is due to scene change.

Meanwhile, the number of remaining bits should also be considered when the target bit is computed, i.e.,

$$\hat{T}_i(jL+1) = \frac{W_{p,i}(jL-L+1)B_i(jL)}{W_{p,i}(jL-L+1)N_{p,r} + W_{b,i}(jL)N_{b,r}},$$

where $N_{p,r}$ and $N_{b,r}$ are the numbers of the remaining stored pictures and the remaining non-stored pictures, respectively.

The target bits are a weighted combination of the items $\tilde{T}_i(jL+1)$ and $\hat{T}_i(jL+1)$, i.e.,

$$T_i(jL+1) = \beta\hat{T}_i(jL+1) + (1-\beta)\tilde{T}_i(jL+1), \tag{30}$$

where β is a constant and its typical value is 0.875.

It can be known from (30) that a tight buffer regulation can be achieved by choosing a small β. Similarly, a large β is selected in our scheme to improve video distortion that is due to scene change.

To maintain the quality of the coded frames, the target bit $T_i(jL+1)$ is bounded by

$$T_i(jL+1) = \max\left\{T_i(jL+1), m_{hdr,i}(jL-L+1) + \frac{R_i(jL+1)}{4f}\right\}. \tag{31}$$

where $m_{hdr,i}(jL-L+1)$ is the number of bits used for the header and motion vectors by the previous stored picture.

To conform with the HRD, the target bits are further bounded by

$$T_i(jL+1) = \min\{\max\{Z_i(jL+1), T_i(jL+1)\}, U_i(jL+1)\}. \tag{32}$$

Step 2. Compute the quantization parameter of the current stored picture and perform RDO for each MB in the current picture.

The MAD of the current store picture is predicted by model (3) using the actual MAD of the previous stored picture. The quantization parameter $QP_i(jL+1)$ corresponding to the target bits is then computed by using the quadratic model (4) and (6). The details on this can be found in [2,4,6], it is thus not elaborated in this section. To maintain the smoothness of visual quality among successive frames, the quantization parameter $QP_i(jL+1)$ is further adjusted by

$$\begin{aligned} QP_i(jL+1) = {} & \min\{QP_i(jL-L+1)+2, \\ & \max\{QP_i(jL-L+1)-2, QP_i(jL+1)\}\}, \end{aligned} \tag{33}$$

where $QP_i(jL-L+1)$ is the quantization parameter of the previous stored pictures.

Meanwhile, it should be bounded by the global bounds provided by H.264, i.e.,

$$QP_i(jL+1) = \min\{\max\{QP_i(jL+1), 0\}, 51\}. \tag{34}$$

The quantization parameter is then applied to perform RDO for each MB in the current stored picture by using the method provided in [15,16].

4.2. Post-encoding stage

There are three tasks in this stage: update the parameters a_1 and a_2 of model (3), update the parameters c_1 and c_2 of quadratic R-D models (4) and (6), and determine the number of frames that need to be skipped.

After encoding a frame, the parameters a_1 and a_2 of model (3), as well as c_1 and c_2 of quadratic R-D models (4) and (6) are updated. A linear regression method similar to that in [2,4] is used where the sliding window size is computed by using the method provided in [6] instead of that in [2,4].

The actual number of bits generated, $b_i(j)$ is added to the current buffer. To prevent the updated buffer from overflow, the frame skipping parameter N_{post} is set to zero and increased until the buffer condition

$$V_i(j + N_{\text{post}}) < 0.8V_s \tag{35}$$

is satisfied [2], where the buffer fullness is updated as

$$V_i(j + l + 1) = V_i(j + l) - \frac{R_i(j + l)}{f}; \ 1 \leqslant l \leqslant N_{\text{post}} - 1.$$

5. Basic unit level rate control

If the basic unit is not selected as a frame, an additional basic unit level rate control should be added to our scheme.

Same as the frame layer rate control, the quantization parameters for the IDR pictures and the non-stored pictures are the same for all basic units in the same frame. They are computed almost the same as that at the frame level provided that $QP_i(jL + 1)$ and $QP_i(jL + L + 1)$ are replaced by the average values of quantization parameters of all basic units in the corresponding stored pictures.

In the remaining part of this section, we shall provide the basic unit layer rate control for each stored picture.

Same as the frame level, we shall first determine target bits for each stored picture. The process is the same as that at the frame layer. The bits are then allocated to the remaining basic units according to their predicted MADs.

The basic unit level rate control selects the values of quantization parameters of all basic units in a stored picture, so that the sum of generated bits is close to the frame target $T_i(jL + 1)$. The following is a step-by-step description of this method.

Step 1. Predict the MADs of the remaining basic units in the current stored picture by model (3) using the actual MADs of the co-located basic units in the previous stored picture.

Step 2. Compute the number of texture bits $\hat{b}_{l,i}(jL + 1)$ for the current basic unit. This step is composed of the following three substeps:

Step 2.1. Compute the target bits for the current basic unit.

Let $T_{r,i}(jL+1)$ denote the number of remaining bits for the the remaining basic units in the current stored picture, and the initial value of $T_{r,i}(jL+1)$ is $T_i(jL+1)$. The target bits for the lth basic unit are given by

$$\tilde{b}_{l,i}(jL+1) = T_{r,i}(jL+1)\frac{\tilde{\delta}^2_{l,i}(jL+1)}{\sum_{k=l}^{N_{unit}}\tilde{\delta}^2_{k,i}(jL+1)}. \tag{36}$$

Remark 2. Since only $\delta_{k,i}(jL+1)$ is available by using our simple method, our basic unit layer bit allocation scheme is thus designed as (36). The RDO cost should be used if 16×16 or 8×8 based motion estimation and compensation is performed for the current frame. It is desirable to use the available information as much as possible in the bit allocation scheme. □

Step 2.2. Compute the average number of header bits that are generated by all coded basic units:

$$\tilde{m}_{hdr,l} = \tilde{m}_{hdr,l-1}\left(1-\frac{1}{l}\right)+\frac{\hat{m}_{hdr,l}}{l} \tag{37}$$

$$m_{hdr,l} = \tilde{m}_{hdr,l}\frac{l}{N_{\text{unit}}}+m_{hdr,1}\left(1-\frac{l}{N_{\text{unit}}}\right); \quad 1\leqslant l\leqslant N_{\text{unit}}, \tag{38}$$

where $\hat{m}_{hdr,l}$ is the actual number of header bits generated by the lth basic unit in the current stored picture, $m_{hdr,1}$ is the estimate from all basic units in the previous frame.

Remark 3. If the estimation of motion cost is available, it can also be used to achieve a more accurate estimation of the head bits for the current basic unit. □

Step 2.3. Compute the number of texture bits $\hat{b}_{l,i}(jL+1)$ as below.

$$\hat{b}_{l,i}(jL+1) = \tilde{b}_{l,i}(jL+1) - m_{hdr,l}. \tag{39}$$

To maintain the quality of each basic unit, $\hat{b}_{l,i}(jL+1)$ is further bounded by

$$\hat{b}_{l,i}(jL+1) = \max\left\{\hat{b}_{l,i}(jL+1), \frac{R_i(jL+1)}{4fN_{\text{unit}}}\right\}. \tag{40}$$

Step 3. Compute the quantization parameter of the current basic unit by using the quadratic R-D model (4) and (6). We need to consider the following three cases:

Case 1. When the basic unit is the first one in the current stored picture, the quantization parameter is given by

$$QP_{1,i}(jL+1) = \bar{QP}_i(jL-L+1), \tag{41}$$

where $\bar{QP}_i(jL-L+1)$ is the average value of quantization parameters for all basic units in the previous stored picture.

Case 2. When $T_r(jL+1) < 0$, the quantization parameter should be greater than that of previous basic unit such that the sum of generated bits is close to $T_i(jL+1)$, i.e.,

$$QP_{l,i}(jL+1) = QP_{l-1,i}(jL+1) + 1. \tag{42}$$

To maintain the smoothness of visual quality, the quantization parameter is further bounded by

$$\begin{aligned} QP_{l,i}(jL+1) = {} & \max\{\bar{QP}_i(jL-L+1) - 2, \\ & \min\{\bar{QP}_i(jL-L+1) + 2, QP_{l,i}(jL+1)\}\}. \end{aligned} \tag{43}$$

Two major functions of bounds are to maintain the smoothness of visual quality and to compensate the inaccuracy of model (3).

Case 3. Otherwise, we shall first compute a quantization parameter $QP_{l,i}(jL+1)$ by using the quadratic model (4) and (6). Similar to case 2, it is bounded by

$$\begin{aligned} QP_{l,i}(jL+1) = {} & \max\{QP_{l-1,i}(jL+1) - 1, \\ & \min\{QP_{l,i}(jL+1), QP_{l-1,i}(jL+1) + 1\}\}. \end{aligned} \tag{44}$$

The objective is to reduce the blocking artifacts. Meanwhile, to maintain the smoothness of visual quality, it is further bounded by (43).

Step 4. Perform RDO for all MBs in the current basic unit.

Step 5. Update the number of remaining bits T_r and update the parameters of the linear MAD prediction model and the quadratic R-D model.

Step 6. After coding the current stored picture, $\bar{QP}_i(jL+1)$ is updated.

Remark 4. To obtain a good trade-off between average PSNR and bit fluctuation, N_{mbunit} is recommended to be the number of MBs in a row for field coding, adaptive field/frame coding and MB adaptive field/frame coding, and N_{unit} is recommended to be 9 for other applications. □

Remark 5. To reduce the number of bits used for the difference among the quantization parameters of MBs, the syntax of H.264 could be modified by inserting a number in the beginning of the bit stream to indicate the exact number of MBs in the basic unit. We then only need to code the difference among the quantization parameters of basic unit instead of those of MBs. □

6. Experimental results

Extensive experiments have been carried out in this section by using [18]. The test conditions are given in Table 1 [17]:

An encoder employs rate control as a way to regulate varying bit rate characteristics of the coded bitstream to produce high quality coded frames at a given target bit rate. The target bit rate can be given in many ways. For example, it can be a predefined CBR, or a VBR determined by the status of the Internet.

We shall first test the effect of different basic units on the bit fluctuation and the average PSNR. We consider two video sequences with QCIF size where the target bit rate is 64 kb/s, the frame rate is 15 frames/s, the GOP length is 150, QP_0 is 25, and RDO is switched off. The experimental results are illustrated in Fig. 2, and Tables 2 and 3. It is shown that a frame layer rate control could achieve a high PSNR but a big bit fluctuation, while MB layer rate control could have small bit fluctuation with slight loss in PSNR. A basic unit layer rate control can be used to obtain a good trade-off.

We now consider the case that the target bit rate is time varying. Test sequences are Foreman, News, Container, and Silent with QCIF size(4:2:0), Mobile, Paris, and Tempete with CIF size (4:2:0); Source frame rate is 30 frames/s, encoded frame rate is 15 frames/s; 150 frames are used for each sequence and the GOP length is 150; QP for the first I frame is 21 for sequences with QCIF size and 26 for sequences with CIF size, the time varying bandwidth is 128 kb/s (the first 59 frames $1 \leqslant j \leqslant 59$) and 192 kb/s (the remaining 91 frames $60 \leqslant j \leqslant 150$) for video sequences with QCIF size, is 512kb/s (the first 59 frames $1 \leqslant j \leqslant 59$) and 786 kb/s (the remaining 91 frames $60 \leqslant j \leqslant 150$) for video sequences with CIF size. The experimental results are listed in the Table 4 and Figs. 3–9. It is shown that the actual bit rate is kept close to the

Table 1
Testing condition

MV resolution	1/4 pel
Hadamard	On
RDO	On (IBBP), Off (IPP)
Search range	±32
Restrict search range	2
Reference frame	1
Symbol	CABAC
GOP structure	IPP (VBR, CBR), IBBP (CBR)
Platform	JM6.1c

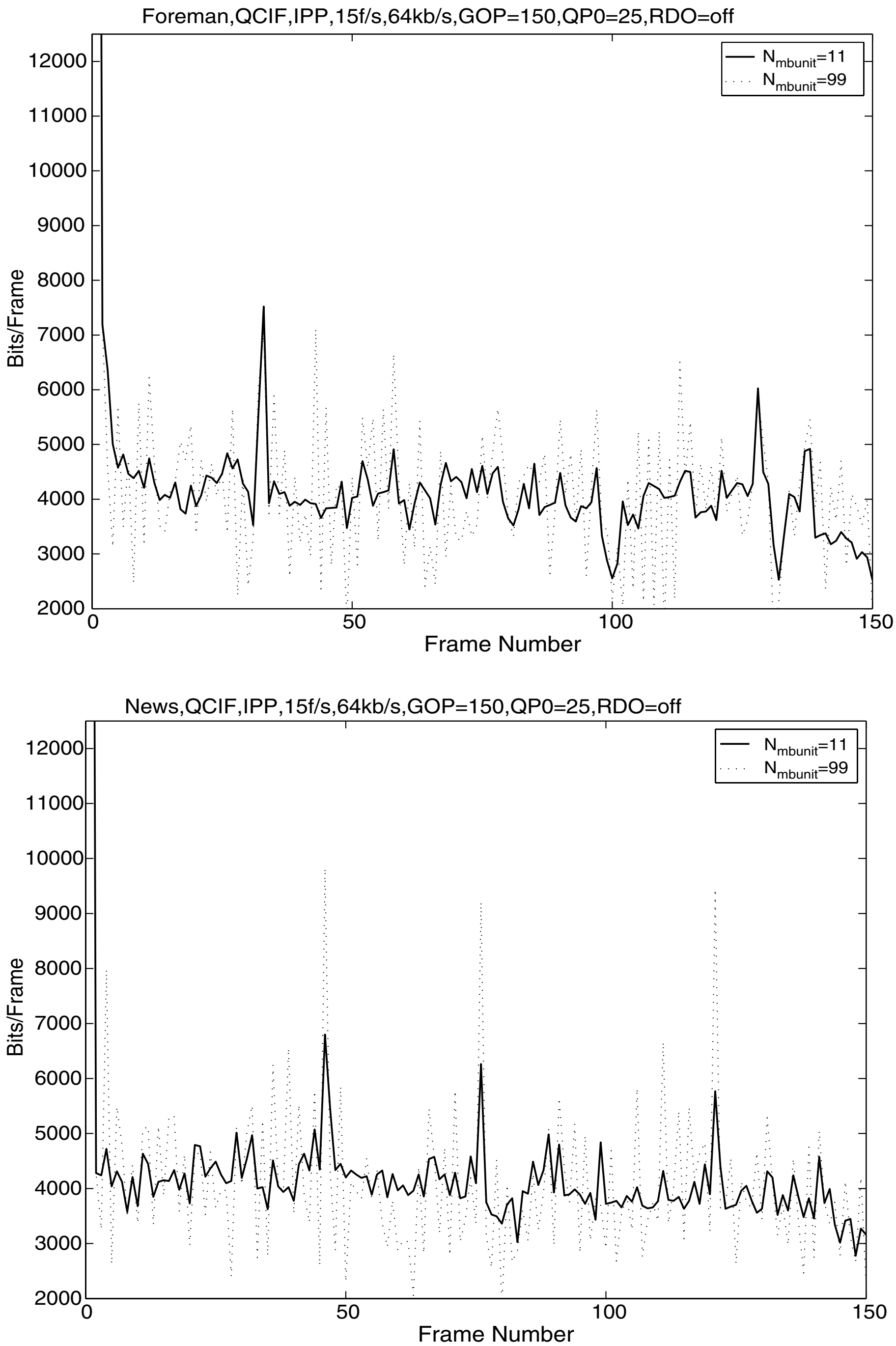


Fig. 2. Comparison of bit fluctuation among different choices of basic unit for video sequences Foreman and News.

Table 2
Comparison of average PSNR among different choices of basic unit

Sequences	N_{mbunit}	Y (dB)	U (dB)	V (dB)
News	99 (frame)	37.96	41.22	41.58
News	11 (slice)	37.89	41.14	41.64
News	1 (MB)	37.66	41.23	41.61
Foreman	99 (frame)	33.94	38.50	39.76
Foreman	11 (slice)	33.85	38.47	39.73
Foreman	1 (MB)	33.74	38.38	39.62

Table 3
Comparison of bit fluctuation among different choices of basic unit

Sequences	N_{mbunit}	Mean (bits)	Standard deviation
News	99 (frame)	4060	1303
News	11 (slice)	4063	969
News	1 (MB)	4072	533
Foreman	99 (frame)	4052	1153
Foreman	11 (slice)	4056	636
Foreman	1 (MB)	4078	876

Table 4
Experimental results for our adaptive rate control under time varying channel

Video sequence	Average PSNR (dB)	Skipped frames	Resulted bandwidth (kb/s)
Foreman	38.36	0	166.64
News	43.64	0	166.81
Container	43.42	0	166.78
Silent	42.66	0	166.87
Paris	39.82	0	668.78
Tempete	34.12	0	642.35
Mobile	31.94	0	668.31

predefined bandwidth. With our scheme, the H.264 encoder can be adaptive to the time varying channel bandwidth that is available to the coding process. Meanwhile, the buffer neither underflows nor overflows. The available channel bandwidth is fully utilized. However, the actual buffer occupancy does not track the target buffer level well. This is due to the following two reasons: (1) the estimation of the header bits is not accurate and the header bits occupy a large part of the total bits for a basic unit; and (2) the estimation of the MAD is also not accurate enough.

We finally compare the coding efficiency of an H.264 encoder with our rate control to that of an H.264 encoder with a fixed quantization parameter. In other words, the target bit rate is generated by first coding a video sequence with a fixed quantization parameter, and then serves as the input to the H.264 encoder with our rate control. This type of comparison is recommended by the ad hoc group of H.264 according to [17].

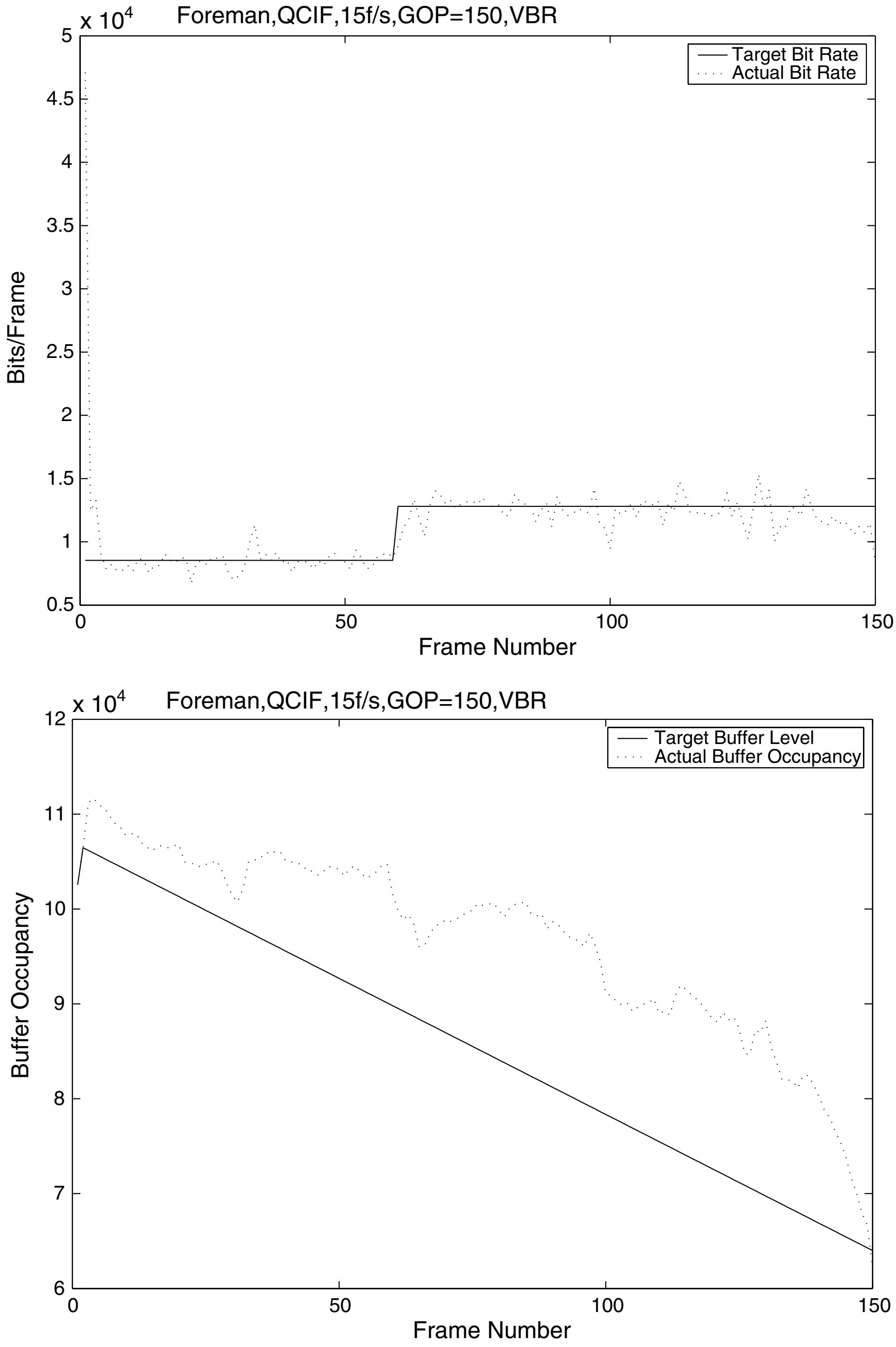


Fig. 3. Experimental results for video sequence Foreman under time varying channel.

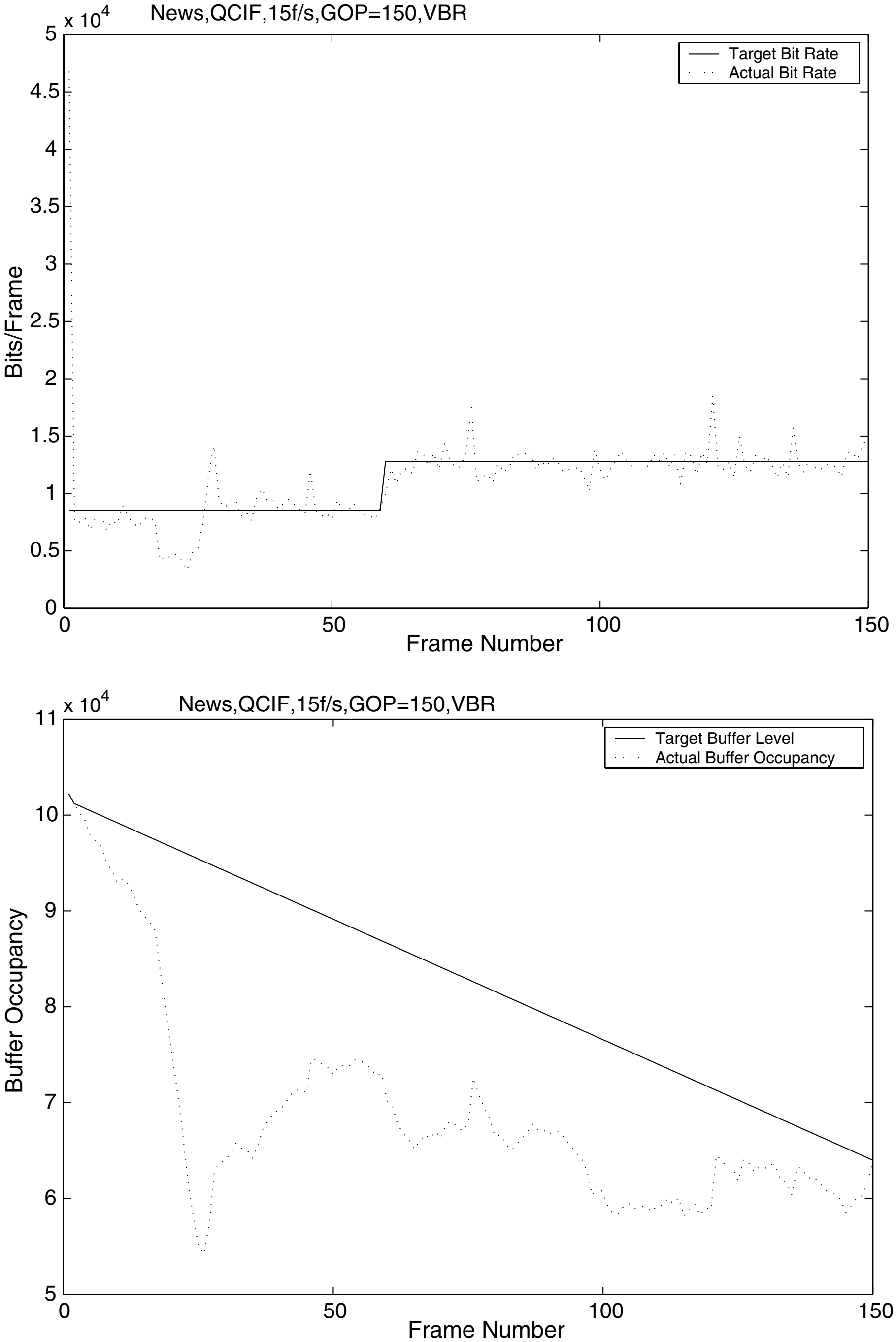
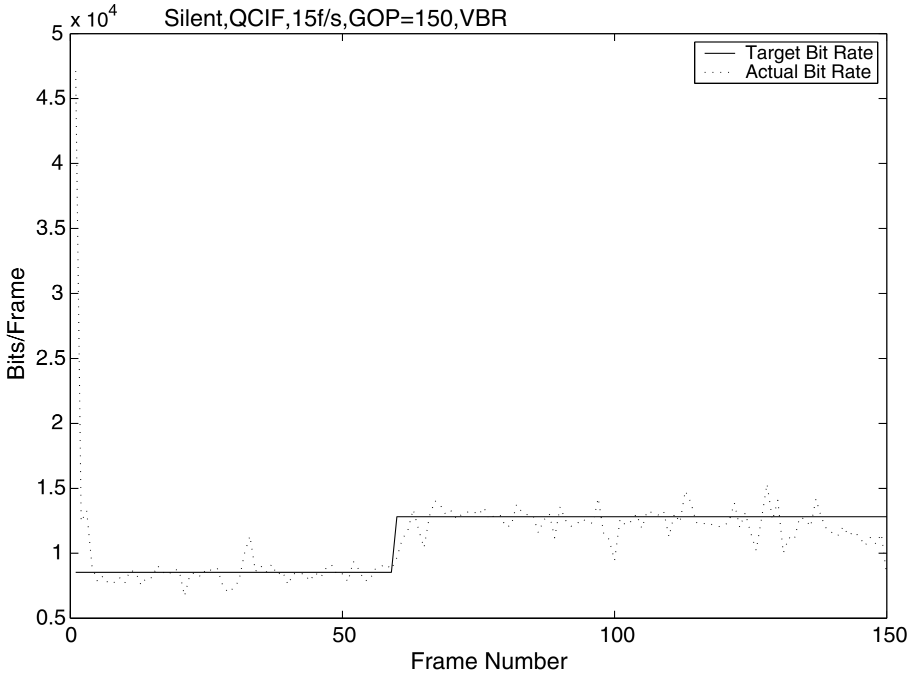


Fig. 4. Experimental results for video sequence News under time varying channel.



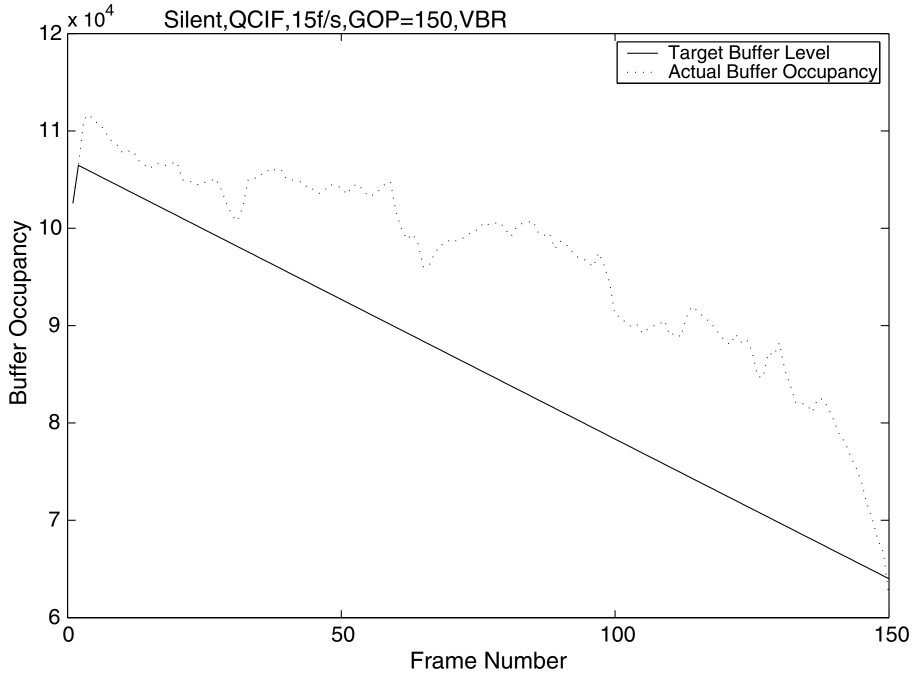


Fig. 5. Experimental results for video sequence Silent under time varying channel.

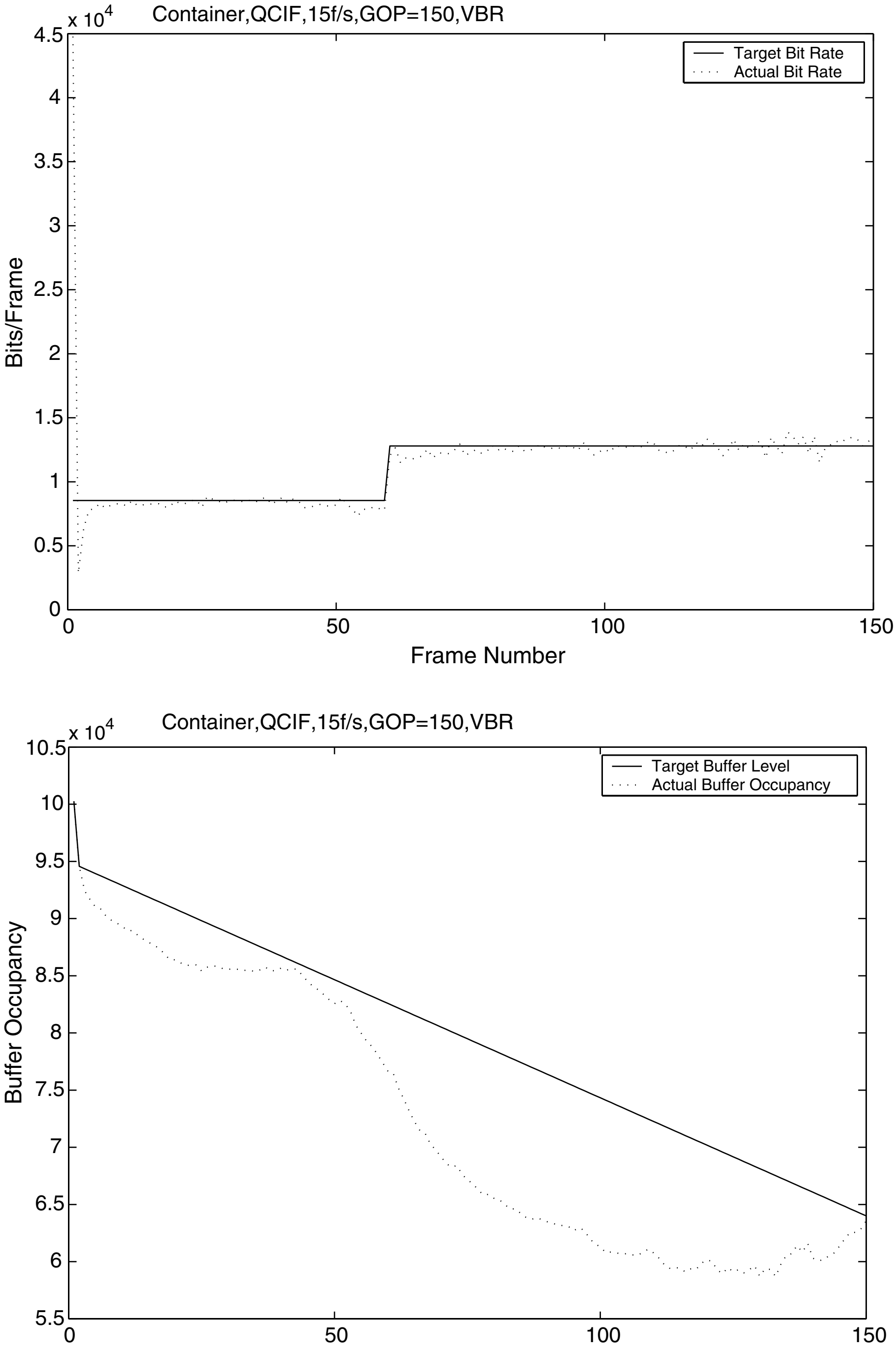


Fig. 6. Experimental results for video sequence Container under time varying channel.

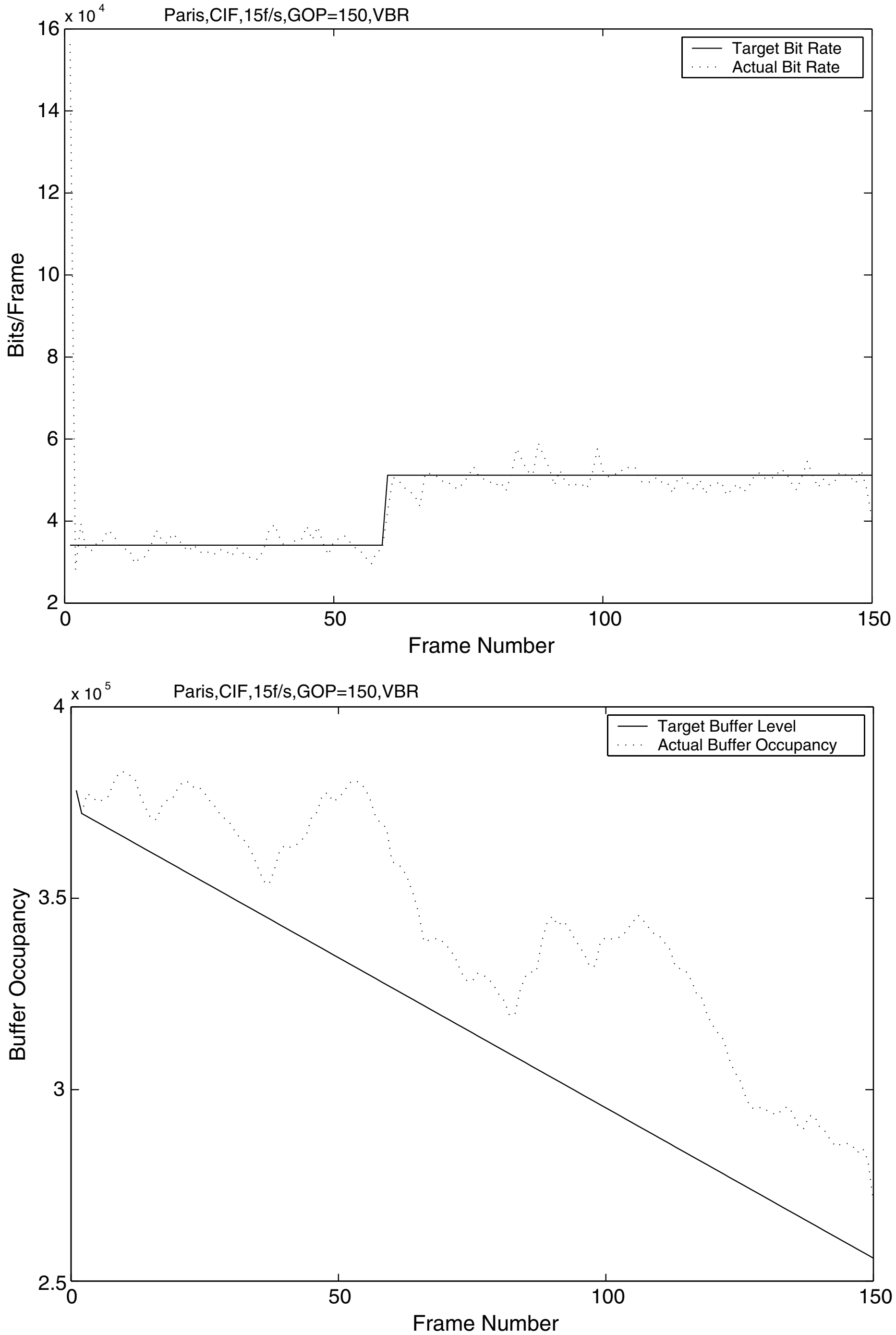


Fig. 7. Experimental results for video sequence Paris under time varying channel.

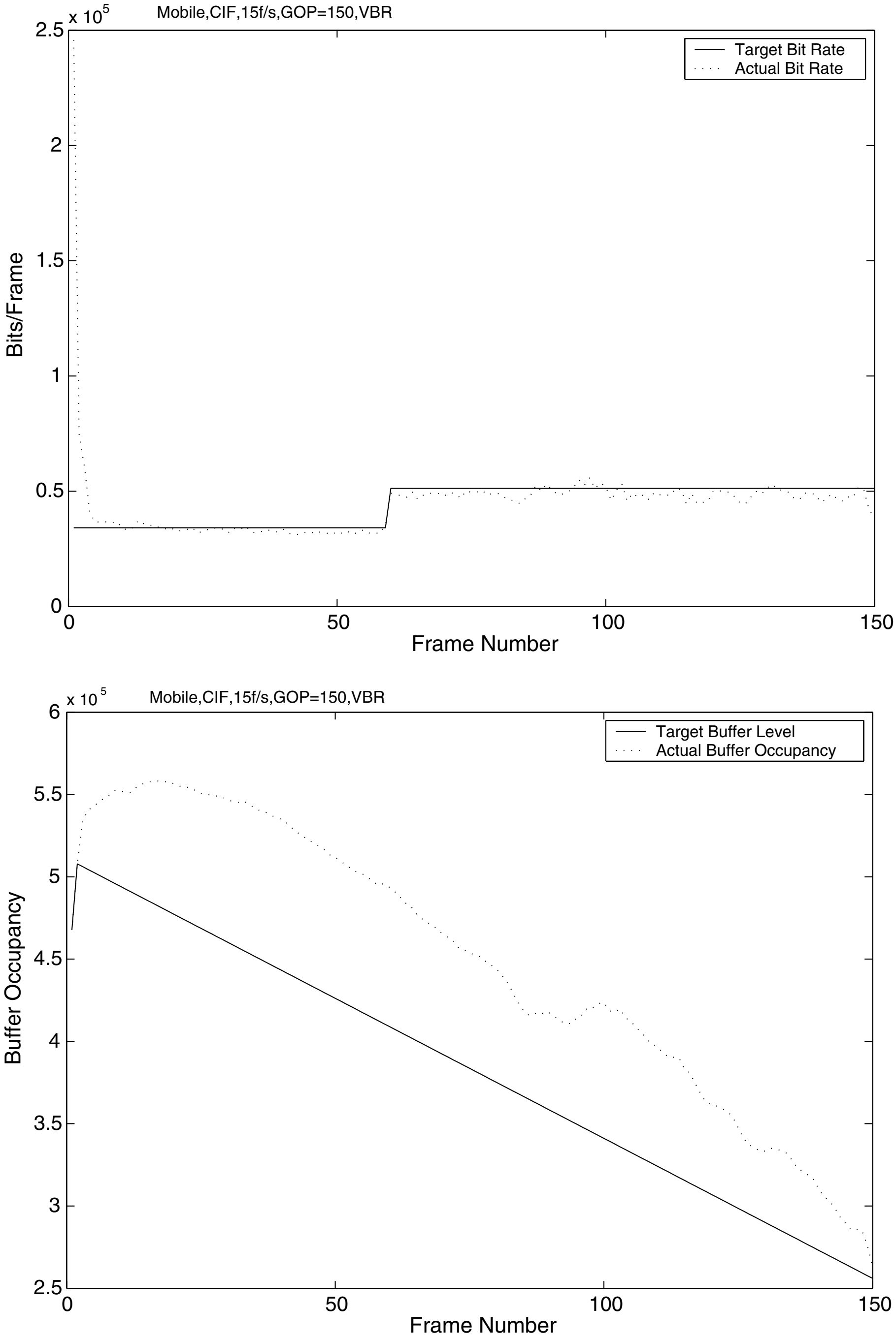


Fig. 8. Experimental results for video sequence Mobile under time varying channel.

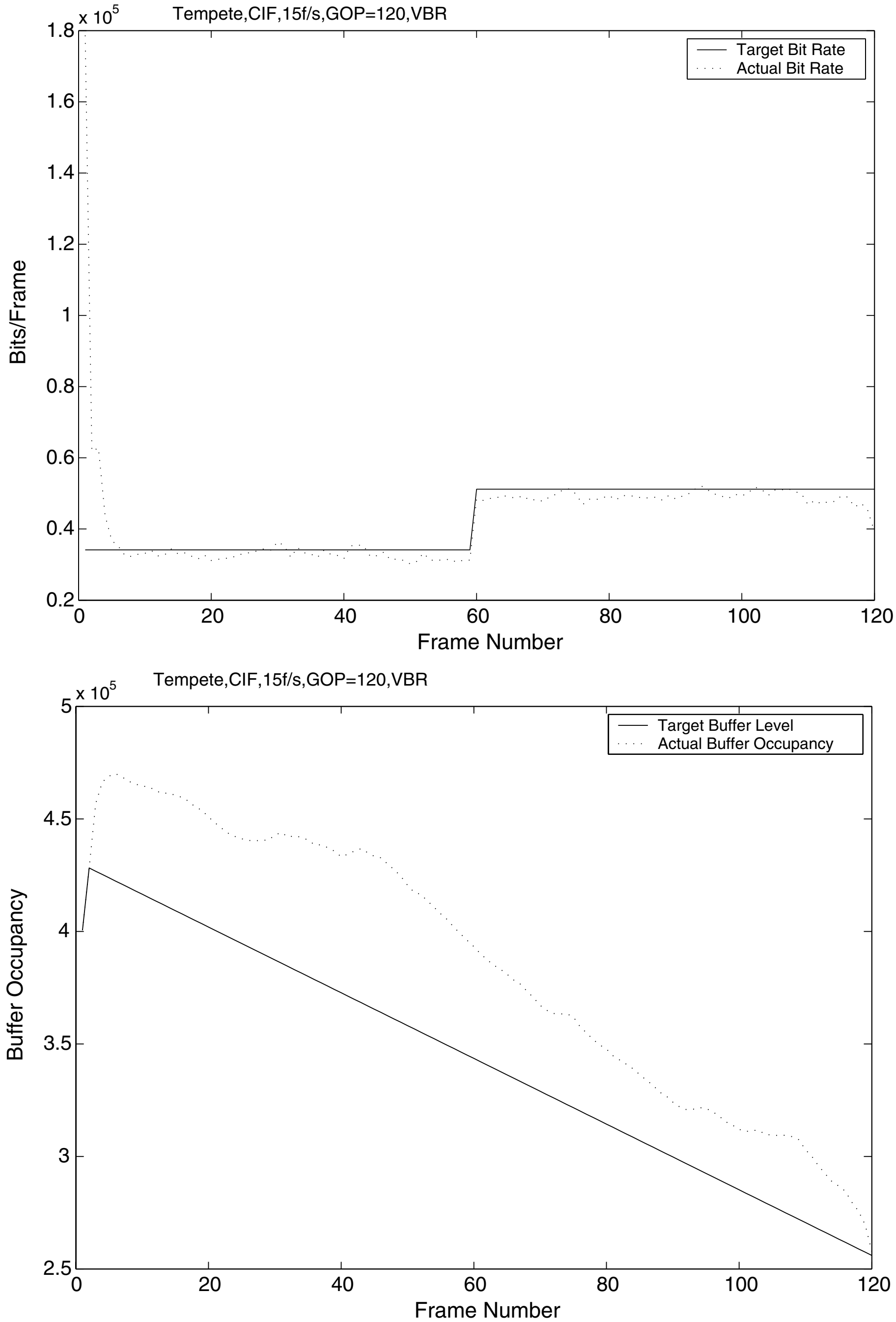
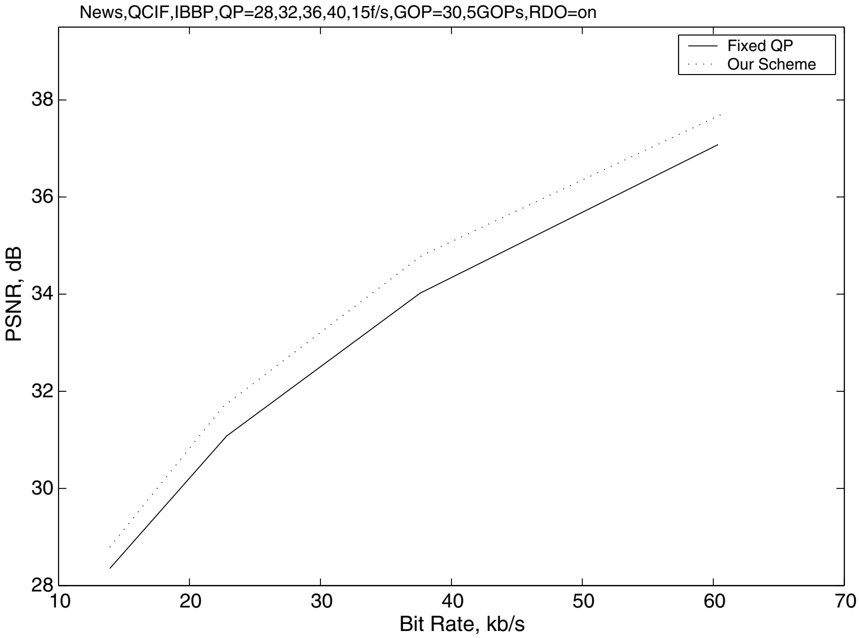


Fig. 9. Experimental results for video sequence Tempete under time varying channel.



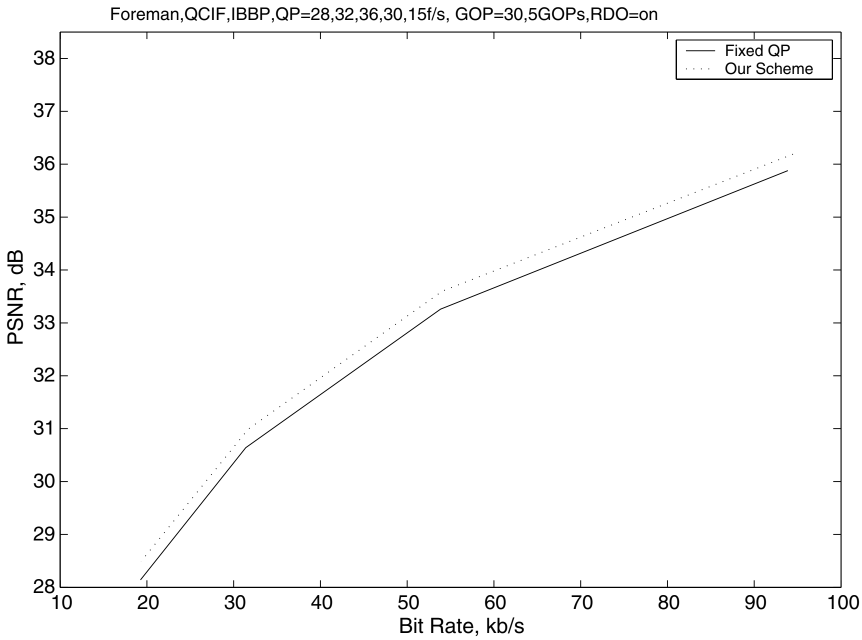
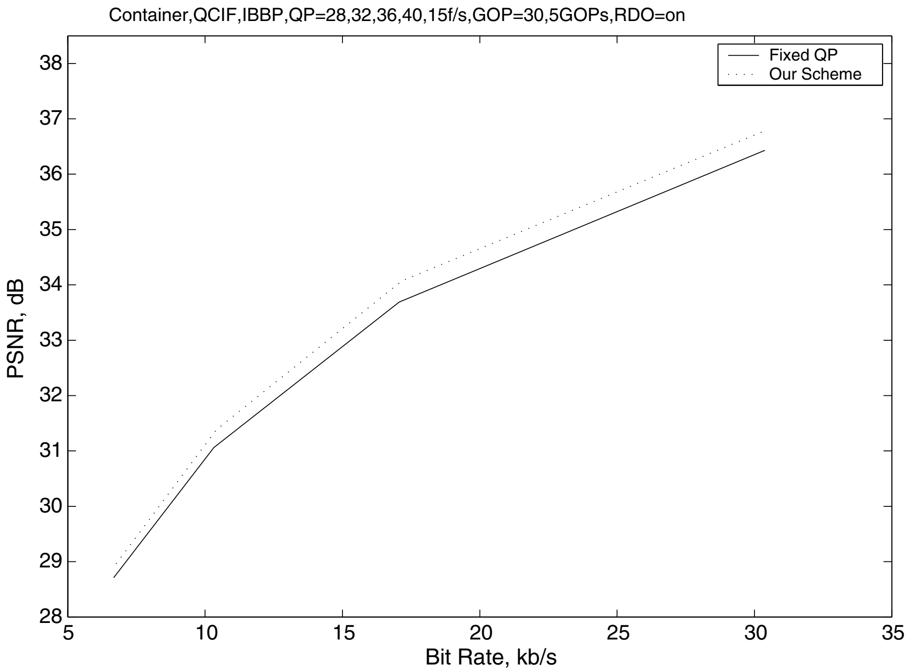


Fig. 10. Comparison results between our scheme and a fixed QP scheme for video sequences News and Foreman under CBR.



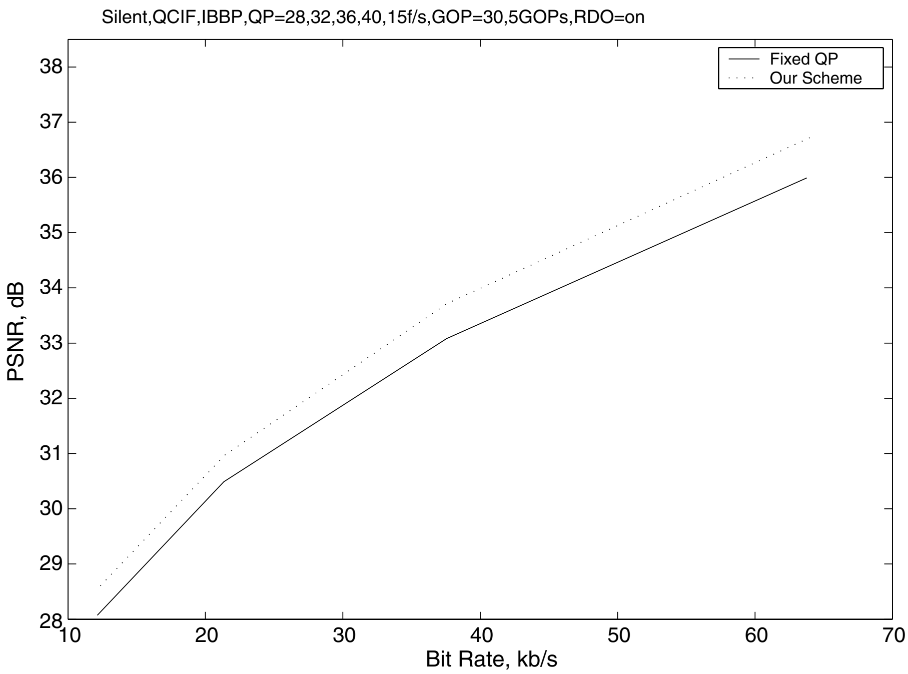


Fig. 11. Comparison results our scheme and a fixed QP scheme for video sequences Container and Silent under CBR.

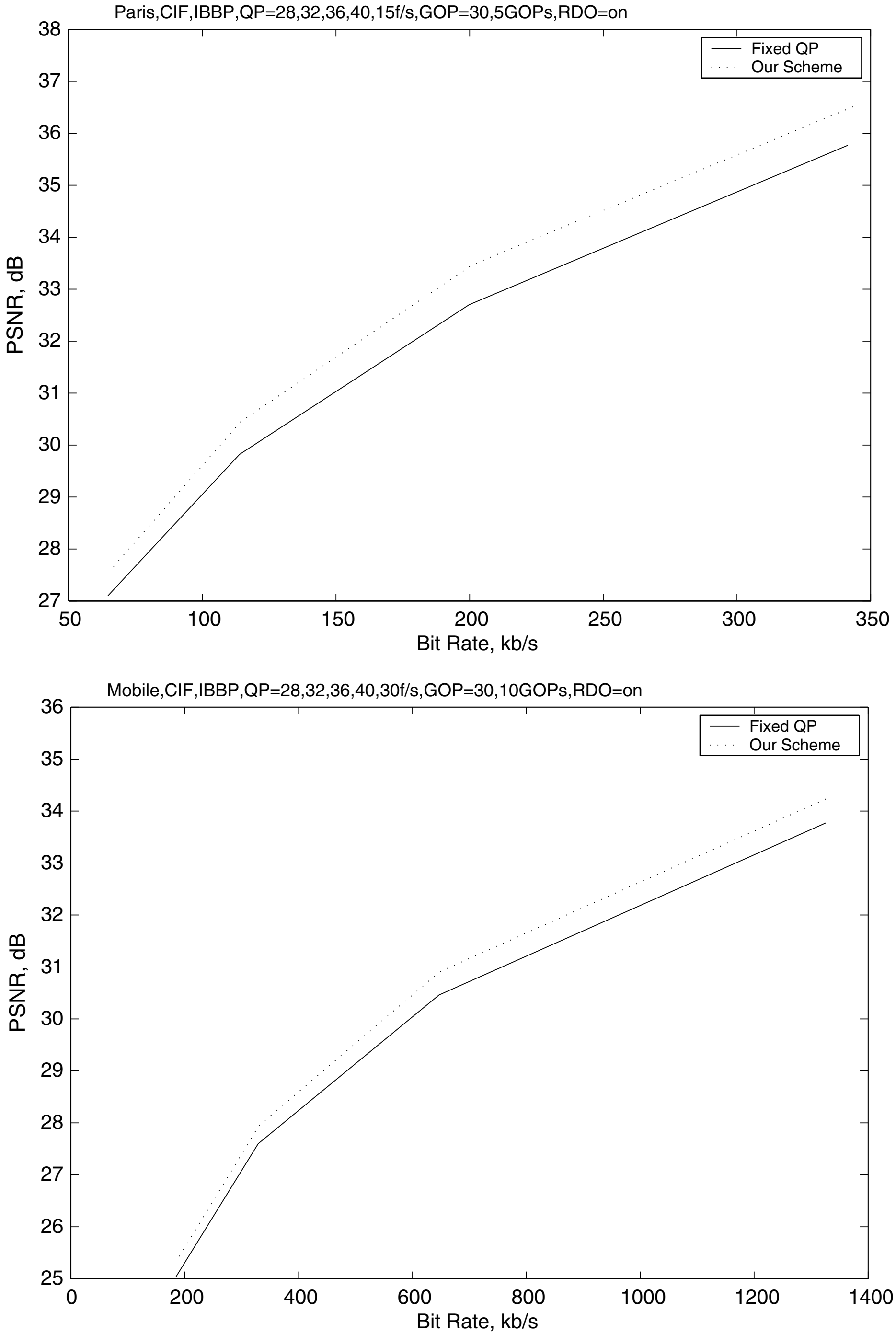


Fig. 12. Comparison results our scheme and a fixed QP scheme for video sequences Paris and Mobile under CBR.

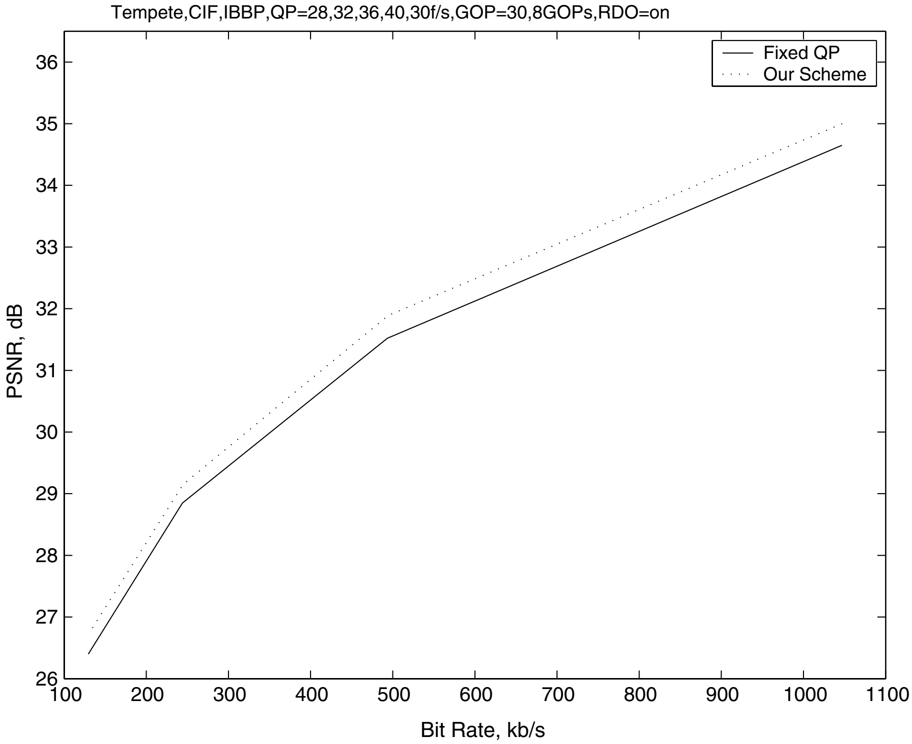


Fig. 13. Comparison results our scheme and a fixed QP scheme for video sequence Tempete under CBR.

Test sequences are Container, Coastguard, News, Silent, and Foreman with QCIF size (4:2:0), Mobile, Tempete, and Paris with CIF size (4:2:0); Source frame rate is 30 frames/s, encoded frame rate is 15 frames/s for Container, News, Silent, Foreman and Paris and 30 for other sequences; 150 frames are used for each sequence with 15 frames/s, 300 frames for sequences with 30 frames/s, and the GOP length is 30; The fixed quantization parameters are set to 28, 32, 36, and 40, respectively [17]. QP_0 in our scheme are set to 26, 30, 34, and 38, respectively.

The experimental results are given in Figs. 10–13. It is shown that compared to an H.264 encoder using fixed quantization parameter, our scheme can improve the average PSNR up to 0.8 dB. The improved average PSNR is 0.47 dB for H.264 test sequences under normal test condition.

7. Conclusions

This paper has proposed an adaptive rate control scheme for H.264 by introducing the concept of basic unit and a linear mean absolute difference (MAD) prediction model. The MADs of the remaining basic units in the current stored picture are predicted by the linear model using the actual MADs of the co-located basic units in the previous stored picture. The target bits for the current stored picture are computed by adopting a fluid flow traffic model and linear tracking theory. The bits are allocated to all remaining basic units in the current picturer according to their predicted MADs. The corresponding quantization parameter is computed by using a quadratic rate-distortion model. The rate distortion optimization is then performed for each macro-block (MB) in the current basic unit by using the quantization parameter. Both constant bit rate (CBR) and variable bit rate (VBR) cases are studied in this paper. The average PSNR is improved by up to 0.8 dB for an encoder using our scheme compared to an encoder using a fixed quantization parameter. The overall average improvements in PSNR for all the test sequences recommended by H.264 is 0.47 dB. With our scheme, an H.264 encoder can be adaptive to time varying channel bandwidth that is available for the coding process.

References

[1] Z.G. Li, C. Zhu, Nam Ling, X.K. Yang, G.N. Feng, S. Wu, F. Pan, A unified architecture for real time video coding systems, IEEE Trans. Circuit. Syst. Video Technol. 13 (2003) 472–487.

[2] H.J. Lee, T.H. Chiang, Y.Q. Zhang, Scalable Rate Control for MPEG–4 Video, IEEE Trans. Circuit Syst. Video Technol. 10 (2000) 878–894.

[3] J. Ribas-Corbera, S. Lei, Rate control in DCT video coding for low-delay communications, IEEE Trans. Circuit Syst. Video Technol. 9 (1999) 172–185.

[4] A. Vetro, H. Sun, Y. Wang, MPEG-4 rate control for multiple video objects, IEEE Trans. Circuit Syst. Video Technol. 9 (1999) 186–199.

[5] Z.G. Li, Y.C. Soh, C.Y. Wen, Switched and Impulsive Systems: Analysis, Design and Application, Springer-Verlag, Berlin, 2005.

[6] Z.G. Li, X. Lin, C. Zhu, F. Pan, A novel rate control scheme for video over the internet, in: Proceedings ICASSP 2002, vol. 2, Florida, USA, 2002. pp. 2065–2068.

[7] Z.G. Li, F. Pan, K.P. Lim, G.N. Feng, X. Lin, R. Susanto, Adaptive basic unit layer rate control for JVT, in: 7th meeting: Pattaya, 2003, PP. 7–15, JVT-G012.

[8] Z.G. Li, W. Gao, Feng Pan, S.W. Ma, K.P. Lim, G.N. Feng, X. Lin, R. Susanto, Y. Lu, H.Q. Lu. Adaptive rate control with HRD consideration, in: 8th Meeting: Geneva, 2003, pp. 23–27, JVT-H014.

[9] Z.G. Li, F. Pan, K.P. Lim, G.N. Feng, X. Lin, R. Susanto, Frame layer rate control for H.264, in: IEEE 2003 International Conference on Multimedia and Expo, Baltimore, MD, USA, 2003, pp. I581–I584.

[10] Z.G. Li, Nam Ling, S. Rahardja, X. Lin, P. Li, An iterative method for hypothetical reference decoder, in: IEEE 2004 International Conference on Multimedia and Expo, Taiwan, 2004, pp. 27–30.

[11] M.Q. Jiang, X.Q. Yi, Nam Ling. Improved frame-layer rate control for H.264 using MAD ratio, in: IEEE 2004 International Symposium on Circuits and Systems (ISCAS 2004), Vancouver, Canada, 2004, pp. 813–816.

[12] Z.G. Li, F. Pan, K.P. Lim, X. Lin, R. Susanto, Adaptive rate control for H.264, in: IEEE 2004 International Conference on Image Processing, Singapore, 2004.

[13] C.T. Chen, Linear System Theory and Design, Rinehart and Winston, New York, 1984.

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

[15] Information Technology-Coding of Audio-Visual Objects-Part 10: Advance Video Coding. ISO/IEC FDIS 14496-10.

[16] Thomas Wiegand and Bernd Girod. Parameter Selection in Lagrangian Hybrid Video Coder Control, in: Proc. IEEE International Conference on Image Processing, Thessaloniki, Greece, 2001.

[17] Gary Sullivan, Recommended simulation common conditions for H.26L coding efficiency experiments on low resolution progressive scan source material, in: VCEG-N81, the 14th Meeting: Santa Barbara, CA, USA, 2001, pp. 24–27.
[18] JM6.1c. ftp://ftp.imtc-files.org/jvt-experts/2003_05_Pattaya/.