ADAPTIVE RATE CONTROL FOR H.264

Z. G. Li, F. Pan, K. P. Lim, X. Lin, and S. Rahardja

Media Division, Institute for Infocomm Research
21 Heng Mui Keng Terrace, Singapore 119613
{ezgli,efpan,kplim,linxiao,rsusanto}@i2r.a-star.org.sg

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. It can be used to obtain a trade-off between the overall coding efficiency and the bits fluctuation. The linear model is used to solve the chicken and egg dilemma existing in the rate control of H.264. Both constant bit rate (CBR) and variable bit rate (VBR) cases are studied. Our scheme has been adopted by H.264.

1. INTRODUCTION

Rate control is used to compute quantization parameters for the current frame according to the specified bit rate and the statistics of the current frame, like the MAD and the head bits of each macroblock (MB). The rate control in H.264 is more complex than Q2 and TMN 8 in the sense that the statistics of the current frame is available in Q2 and TMN 8 while it is not available for the rate control of H.264 [1, 2]. This is because that the quantization parameters are involved in both rate control and rate distortion optimization (RDO) of H.264 while it is only involved in rate control of 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 statistics of the current frame. However, the statistics of the current frame 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 the co-located MB in the previous frame. However, there exists problem associated with the prediction at the MB level. For example, image the case that an MB in the current frame is an Intra-MB while the co-located MB is an Inter-one. To solve this problem, we introduce the concept of basic unit, which can be either an MB, a slice, or a frame. All MBs in the same basic unit share a common quantization parameter. The choice of basic unit depends on the size of coded picture buffer (CPB). An adaptive basic unit layer rate control scheme is then presented for H.264 by further introducing a linear model 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. This solves the chicken and egg dilemma mentioned above. 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 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 rate-distortion (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. Compared to an H.264 encoder using fixed quantization parameter, our scheme can improve the average PSNR up to 0.78dB. The improved average PSNR is 0.43 dB for all H.264 test sequences under normal test condition. Our scheme has been adopted by H.264.

2. PRELIMINARY KNOWLEDGE

2.1. The Chicken and Egg Dilemma

The coding process of a MB related to the rate control of H.264 is given by [3]: Statistics of current frame → Rate control → Quantization parameter → RDO → Statistics of current frame → Entropy coding.

Clearly, there exists a chicken and egg dilemma when the rate control is implemented. Because of this, the rate control of H.264 is more difficult than TM 5, Q2 and TMN 8. To study the rate control for H.264, we need to find a way to estimate the statistics of the current frame. Besides this, we also need to determine the number of contiguous MBs that share a quantization parameter.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

2.2. The Definition of Basic Unit

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} [4]. Denote the total number of basic units in a frame by N_{unit}, which is computed by $N_{unit} = N_{mbpic}/N_{mbunit}$.

2.3. A Linear MAD Prediction Model

Suppose that the jth picture is a stored picture and the number of successive non-stored pictures between two stored pictures are L. The following linear model (1) is proposed to solve the chicken and egg dilemma [4].

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

where $\tilde{\delta}_{l,i}(j)$ is the predicted MAD of the lth basic unit in the current stored picture in the ith GOP, $\delta_{l,i}(j-L-1)$ 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.

Remark 1 It should be mentioned that $\delta_{l,i}(j - L - 1)$ is a reference value for the prediction. There are many other choices for the reference value. For example, 16×16 based motion estimation and motion compensation can be proceeded for all MBs in the current picture before performing RDO. 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 resulted MAD can also be used as the reference value. For the simplicity, we choose $\delta_{l,i}(j - L - 1)$ as the reference value. The number of header bits can be also obtained in this way. Similarly, there also exist many models for the prediction and for the simplicity, we choose the linear model.

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 follows:

$$B_i(1) = \frac{R_i(1)}{f} N_i + B_{i-1}(N_{i-1}) \tag{2}$$

where $R_i(j)$ is the available channel bandwidth which can be either constant or time varying, N_{i-1} denote the total number of frames in the $(i-1)$th GOP, and f is the predefined frame rate.

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

$$\begin{aligned} & B_i(j) = B_i(j-1) - b_i(j-1) \\ & + \frac{R_i(j) - R_i(j-1)}{f}(N_i - j + 1)\ ;\ j = 2, 3, \cdots, N_i \end{aligned}$$

where $b_i(j)$ is the coded bits of the jth frame in the ith GOP.

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. 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

$$\begin{aligned} QP_i(1) \ = \ & \max\{\min\{\frac{Sum_{PQP}(i-1)}{N_p(i-1)} - \min\{2, \frac{N_{i-1}}{15}\}, \\ & QP_{i-1}(1) + 2\}, QP_{i-1}(1) - 2\} \end{aligned}$$

where $N_p(i-1)$ is the total number of stored pictures in the $(i-1)$th GOP and $Sum_{PQP}(i-1)$ is the sum of quantization parameters for all stored pictures in the $(i-1)$th GOP. $QP_i(1)$ is further adjusted by

$$QP_i(1) = QP_i(1) - 1\ ;\ \text{if} \quad QP_i(1) > QP_{i-1}(N_{i-1} - L) - 2$$

where $QP_{i-1}(N_{i-1} - L)$ 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.

4. FRAME LEVEL RATE CONTROL

In this level, quantization parameters of non-stored pictures and the target bits for stored pictures are computed.

4.1. Quantization Parameters of Non-stored Pictures

The quantization parameters of non-stored pictures are obtained through a linear interpolation method as follows:

Suppose that the jth and the $(j + L + 1)$th frames are stored pictures and the quantization parameters of two adjacent stored pictures are $QP_i(j)$ and $QP_i(j+L+1)$, respectively. The quantization parameter of the kth$(1 \leq k \leq L)$ 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

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

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

$$QP_i(j+k) = QP_i(j) + \alpha + \max\{-2(k-1),$$
$$\min\{\frac{(QP_i(j+L+1) - QP_i(j))(k-1)}{L-1}, 2(k-1)\}\}$$

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

$$\alpha = \begin{cases} -3 & QP_i(j+L+1) - QP_i(j) \leq -2L-3 \\ -2 & QP_i(j+L+1) - QP_i(j) = -2L-2 \\ -1 & QP_i(j+L+1) - QP_i(j) = -2L-1 \\ 0 & QP_i(j+L+1) - QP_i(j) = -2L \\ 1 & QP_i(j+L+1) - QP_i(j) = -2L+1 \\ 2 & \text{Otherwise} \end{cases}$$

4.2. Target Bits of Stored Pictures

The bits allocated to the current stored picture should be adjusted according to the current buffer occupancy and the picture complexity as follows:

Step 1 Predefine a target buffer level for each stored picture in the current GOP.

Suppose that the jth picture is a stored picture. The target buffer level for the picture is determined by

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

where $S_i(2)$ is reset as $V_i(2)$ after coding the first stored picture in the ith GOP, V_s is the buffer size, $\bar{W}_{p,i}(j-L-1)$ is the average complexity weight of stored pictures that have been coded, $\bar{W}_{b,i}(j-1)$ is the complexity weight of non-stored pictures that have been coded [4].

Step 2 Compute target bits for the current stored picture.

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:

$$\tilde{T}_i(j) \quad = \quad \frac{R_i(j)}{f} + \gamma(S_i(j) - V_i(j))$$

where $V_i(j)$ is the actual buffer occupancy after coding the $(j-1)$th stored picture in the ith GOP, γ is a constant, its typical value is 0.5 when there is no non-store picture and 0.25 otherwise.

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

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

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 $\tilde{T}_i(j)$ and $\hat{T}_i(j)$,

$$T_i(j) \quad = \quad \beta\hat{T}_i(j) + (1-\beta)\tilde{T}_i(j) \tag{3}$$

where β is a constant and its typical value is 0.5 when there is no non-stored picture and 0.9 otherwise. It can be known from (3) that a tight buffer regulation can be achieved by choosing a small β.

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

$$T_i(j) = \max\{T_i(j), m_{hdr,i}(j-L-1) + \frac{R_i(j)}{4f}\} \tag{4}$$

where $m_{hdr,i}(j-L-1)$ 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

$$T_i(j) \quad = \quad \min\{\max\{Z_i(j), T_i(j)\}, U_i(j)\} \tag{5}$$

where $Z_i(j)$ and $U_i(j)$ are derived by taking the HRD into consideration [3].

5. BASIC UNIT LEVEL RATE CONTROL

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 $T_i(j)$. 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 (1) using the actual MADs of the co-located basic units in the previous stored picture.

Step 2 Compute the number of texture bits $\hat{b}_{l,i}(j)$ 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 $T_{r,i}(j)$ denote the number of remaining bits for the the remaining basic units in the current stored picture, and the initial value of $T_{r,i}(j)$ is $T_i(j)$. The target bits for the lth basic unit are given by

$$\tilde{b}_{l,i}(j) = T_{r,i}(j)\frac{\tilde{\delta}^2_{l,i}(j)}{\sum_{k=l}^{N_{unit}} \tilde{\delta}^2_{k,i}(j)} \tag{6}$$

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

$$\begin{aligned} \tilde{m}_{hdr,l} &= \tilde{m}_{hdr,l-1}(1 - \frac{1}{l}) + \frac{\hat{m}_{hdr,i}}{l} \\ m_{hdr,l} &= \tilde{m}_{hdr,l}\frac{l}{N_{unit}} + m_{hdr,1}(1 - \frac{l}{N_{unit}}) \end{aligned}$$

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.

Step 2.3 Compute the number of texture bits $\hat{b}_{l,i}(j)$.

$$\hat{b}_{l,i}(j) = \tilde{b}_{l,i}(j) - m_{hdr,l} \tag{7}$$

To maintain the quality of each basic unit, $\hat{b}_{l,i}(j)$ is further bounded by $R_i(j)/(4fN_{unit})$.

Step 3 Compute the quantization parameter of the current basic unit by using the quadratic R-D model [3]. We need to consider the following three cases:

Case 1 The first basic unit in the current stored picture.

$$QP_{1,i}(j) = \bar{QP}_i(j-L-1) \tag{8}$$

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

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

$$QP_{l,i}(j) = QP_{l-1,i}(j) + \Delta_{bu} \tag{9}$$

where Δ_{bu} is a variation range for quantization parameters of two successive basic units [3].

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

$$\begin{aligned} QP_{l,i}(j) &= \max\{0, \bar{QP}_i(j-L-1) - \Delta_{fr}, \\ &\quad \min\{51, \bar{QP}_i(j-L-1) + \Delta_{fr}, QP_{l,i}(j)\}\} \end{aligned}$$

where Δ_{fr} is a further variation range for the quantization parameters [3].

Case 3. Otherwise, we shall first compute a quantization parameter $QP_{l,i}(j)$ by using the quadratic model [3] and the bounds given in case 2.

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

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

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

6. EXPERIMENTAL RESULTS

We 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. This type of comparison is recommended by the adhoc group of H.264 according to [5] and the setting for the experiment is given in [4]. The experimental results of (Y,PSNR) are listed in Table 1. It is shown that compared to an H.264 encoder using fixed quantization parameter, our scheme can improve the average PSNR up to 0.78dB. The improved average PSNR is 0.43 dB for H.264 test sequences under normal test condition.

QP	28	32	36	40
Ours(Paris)	34.23	30.9	27.93	25.35
Fixed QP(Paris)	33.77	30.46	27.6	25.04
Ours(News)	37.73	34.75	31.83	28.86
Fixed QP(News)	37.08	34.02	31.08	28.35
Ours(Silent)	36.71	33.7	30.92	28.47
Fixed QP(Silent)	35.99	33.08	30.49	28.07
Ours(Mobile)	34.23	30.9	27.93	25.35
Fixed QP(Mobile)	33.77	30.46	27.6	25.04

Table 1. Experimental results

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 overall average improvements in PSNR for all the test sequences recommended by H.264 is 0.43dB. An alterative method is given in Remark 1. This will be studied in our future research.

8. REFERENCES

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

[2] J. Ribas-Corbera and S. Lei. "Rate control in DCT video coding for low-delay communications". IEEE Trans. Circuit Syst. Video Technology,9:172–185, 1999.

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

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

[5] Gary Sullivan. " Recommended simulation common conditions for H.26L coding efficiency experiments on low resolution progressive scan source material". VCEG-N81, the 14th meeting: Santa Barbara, CA, USA, Sep. 24-27, 2001.

[6] Z. G. Li, F. Pan, K. P. Lim, G. N. Feng, X. Lin and R. Susanto. "Adaptive frame layer rate control for H.264". In IEEE 2003 International Conference on Multimedia and Expo, Baltimore MD USA, pp. I581-I584, June, 2003.