ADAPTIVE FRAME LAYER RATE CONTROL FOR H.264

Z. G. Li, F. Pan, K. P. Lim, G. N. Feng, X. Lin, S. Rahardja and D. J. Wu

Signal Processing Program, Institute for Infocomm Research
21 Heng Mui Keng Terrace, Singapore 119613
Email: {ezgli, efpan, kplim, gnfeng, linxiao, rsusanto, djwu} @i2r.a-star.org.sg

ABSTRACT

This paper proposes an adaptive frame layer rate control scheme for H.264 by introducing a linear model to predict the mean absolute difference (MAD) of current frame by that of previous one. The target bit rate for each frame is computed by adopting a fluid flow traffic model and linear tracking theory. The corresponding quantization parameter is computed by using a quadratic rate-distortion model. The rate distortion optimization (RDO) is then performed for all macroblocks (MBs) in the current frame by the quantization parameter. Both constant bit rate (CBR) and variable bit rate (VBR) cases are studied. The average PSNR is improved up to 0.75 dB compared to an encoder with fixed quantization parameter.

1. INTRODUCTION

An encoder employs rate control as a means to constrain the varying bit rate characteristics of the coded bitstream in order to produce high quality coded frame at the target bit rates. Rate control has been widely studied for other standards, like MPEG 2, MPEG 4, H.263, and so on [1, 3, 2, 7, 4]. There are two types of rate control methods. The first type assumes that the bandwidth desired by the coding process is always guaranteed by the network [1, 3] and does not adjust its target bit rates according to the network status, while the second one utilizes the available bandwidth provided by the network [4], which can be either a constant bit rate (CBR) or a variable bit rate (VBR). Compared to the first type, the second one is adaptive to unpredictable and varying network conditions and is thus more attractive for the video over IP.

Due to high coding efficiency, H.264 will be in the final draft international standard (FDIS) stage soon. However, rate control in H.264 has not been well studied. The topic is very complex because the quantization parameters are involved in both rate control and rate distortion optimization (RDO). There exists a problem when rate control is implemented: to perform RDO for MBs in the current frame, a quantization parameter should be firstly determined for each MB by using the MAD of current frame and/or the MB [1, 2, 7, 8]. However, the MAD of current frame and/or the MB is only available after performing the RDO. The following linear model (1) is proposed to solve this problem.

\[
MAD_{vbr} = a_1 MAD_{vbr}^{prev} + a_2
\]  

where \(MAD_{vbr}^{prev}\) is the exact MAD of previous frame, \(MAD_{vbr}\) is the predicted MAD of current frame, \(a_1\) and \(a_2\) are two coefficients of prediction model, their initial values are 1 and 0, respectively, and they are updated after coding each frame.

We shall now present a fluid flow traffic model. Let \(N_{pop}^i\) denote the total number of frames in a GOP, \(n_i, j (i = 1, 2, \cdots, N_{pop})\) denote the jth frame in the ith GOP, and \(B_i(n_j)\) denote the occupancy of virtual buffer after coding the jth frame. It can be shown from the fluid flow traffic model that [4]

\[
\begin{align*}
B_i(n_{j+1}) &= \min \{ \max \{0, B_i(n_{j+1}), A(n_{j+1}) \} + \frac{v_{j+1}}{F_i}, B_i \} \\
B_i(n_{j+1}) &= \frac{B_i(n_j)}{F_i} + \frac{B_i(n_{j+1})}{F_i}
\end{align*}
\]

where \(A(n_{j+1})\) is the actual bits generated by the jth frame in the ith GOP, \(B_i(n_{j+1})\) is the available channel bandwidth which can be either constant or time varying, \(F_i\) is the predefined frame rate, and \(B_i\) is the buffer size and its maximum value is determined based on different level and different profile [8]. Note that the initial level of the virtual buffer is set to \(B_i/8\). The function of model (2) is to compute the target bit for the current coding picture.

Our scheme is composed of two layers: GOP-layer rate control and frame-layer control. They are presented in the following two sections.
3. GOP-LAYER RATE CONTROL

In this layer, we need to compute the total number of bits for each GOP and to determine the starting quantization parameter of each GOP. Same as [8], we assume that the GOP structure is IBBP... PBB or IPP... P, with I being an intra-coded frame, P being a forward predicted frame and B being a bi-directional predicted frame. The length of a GOP is usually 15-30 [7].

3.1. Total Number of Bits

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

\[ T_r(i,n) = \frac{u(n_{i-1})}{F_p} N_{p-p} + \frac{B_e(n_{i-1}, n_{p-p})}{8} \]

(3)

It can be shown from (3) that the coding results of the latter GOPs depend on those of the former GOPs. To ensure that all GOPs have an uniform quality, each GOP should use its own budget. In other words, the buffer occupancy should be kept at \( B_4/8 \) after coding each GOP.

Since the channel bandwidth may vary at any time, \( T_r \) is updated frame by frame as follows:

\[ T_r(n_{i,j}) = T_r(n_{i,j-1}) - A(n_{i,j-1}) + \frac{u(n_{i,j}) - u(n_{i,j-1})}{N_{p-p} - j} \]

(4)

In the case of CBR, i.e. \( u(n_{i,j}) = u(n_{i,j-1}) \), Equation (4) is simplified as

\[ T_r(n_{i,j}) = T_r(n_{i,j-1}) - A(n_{i,j-1}) \]

(5)

In other words, Equation (4) is also applicable to the CBR case.

3.2. Starting Quantization Parameters

In our scheme, the starting quantization parameter of the first GOP is a predefined quantization parameter \( QP_1 \). The I frame and the first P frame of the GOP are coded by \( QP_1 \). \( QP_1 \) is predefined based on the available channel bandwidth and the GOP length. Normally, a small \( QP_1 \) should be chosen if the available channel bandwidth is wide and a big \( QP_1 \) should be used if it is narrow. Under the same bandwidth, \( QP_1 \) reduces by 1 if the GOP length increases by 15.

The starting quantization parameter of other GOPs \( QP_{st} \) is computed by

\[ QP_{st} = \frac{\text{Sum}_{PQPR} - \frac{8T_r(n_{i-1,n_{p-p}})}{T_r(n_{i,n})} \cdot \min\{ N_{p-p}, 15 \}}{2} \]

(6)

where \( N_p \) is the total number of P frame in the previous GOP and \( \text{Sum}_{PQPR} \) is the sum of quantization parameters for all P frames in the previous GOP.

The I frame and the first P frame in other GOPs are coded by \( QP_{st} \). \( T_r(n_{i-1,n_{p-p}}) \) is the number of remaining bits after the previous GOP is coded. The initial quantization parameter \( QP_{st} \) is reduced if there are some bits remained. Conversely, it is increased if the previous GOP over utilizes its budget. Moreover, it is shown from (6) that \( QP_{st} \) is adaptive to both the GOP length and the available channel bandwidth.

4. FRAME LAYER RATE CONTROL

The frame layer 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. We shall first provide a simple method to compute the quantization parameters of B frames.

4.1.1 Quantization parameters of B frames

Since B frames are not used to predict any other frame, the quantization parameters can be greater than those of their adjacent P or I frames such that the bits could be saved for I and P frames. On the other hand, to maintain the smoothness of visual quality, the difference between the quantization parameters of two adjacent frames should not be greater than 2. Based on the observations, the quantization parameters of B frames are obtained through a linear interpolation method as follows:

Suppose that the number of successive B frames between two P frames is \( L \) and the quantization parameters of two adjacent P frames are \( QP_1 \) and \( QP_2 \), respectively. The quantization parameter of the \( i \)th \( (1 \leq i \leq L) \) B frame is given according to the following two cases:

Case 1 \( L = 1 \). In other words, there is only one B frame between two P frames. The quantization parameter is computed by

\[ QB_i = \left\{ \begin{array}{ll} \frac{QP_1 + QP_2 + 2}{2} & \text{id} QP_1 \neq QP_2 \\ \text{Otherwise} & \end{array} \right. \]

(7)

Case 2 \( L > 1 \). In other words, there are more than one B frame between two P frames. The quantization parameters are computed by

\[ QB_i = QP_1 + \alpha + \max\{ \min\{ \frac{QP_2 - QP_1}{L - 1}, -2(i - 1) \}, -2(i - 1) \} \]

(8)

where \( \alpha \) is the difference between the quantization parameter of the first B frame and \( QP_1 \), and is given as

\[ \alpha = \left\{ \begin{array}{ll} -3 & QP_2 - QP_1 \leq -2L - 3 \\ -2 & QP_2 - QP_1 = -2L - 2 \\ -1 & QP_2 - QP_1 = -2L - 1 \\ 0 & QP_2 - QP_1 = -2L \\ 1 & QP_2 - QP_1 = -2L + 1 \\ 2 & \text{Otherwise} \end{array} \right. \]

(9)

The cases that \( QP_2 - QP_1 < -2L + 1 \) can only occur at time instant that the video sequence switches from one GOP to another GOP.

The final quantization parameter \( QB_i \) is further adjusted by

\[ QB_i = \min\{ \max\{ QB_i, 1 \}, 51 \} \]

(10)

4.1.2 Quantization parameters of P frames

The quantization parameters of P frames are computed by the following two steps:
Step 1 Determine a target bit for each P frame. Step 1 is composed of the following two sub-steps:

Step 1.1 Macroscopic control (budget allocation among frames).
It is implemented by predefining a target buffer level for each P frame. The function of target buffer level is to compute a target bit for each P frame, which is then used to compute the quantization parameter. Since the quantization parameter of the first P frame is given at the GOP layer, we only need to predefine target buffer levels for other P frames in each GOP.

After coding the first P frame in the i-th GOP, the value of target buffer level is initialized as

$$Tb_l(n_{i,2}) = B_c(n_{i,2})$$ (11)

where $B_c(n_{i,2})$ is the actual buffer occupancy after coding the first P frame in the i-th GOP. The target buffer level for the $(j + 1)$th P frame is determined by

$$Tb_l(n_{i,j+1}) = Tb_l(n_{i,j}) - \frac{Tb_l(n_{i,j}) - B_c(n_{i,j})}{N_p - 1} + \frac{W_p(n_{i,j})(L + 1)u(n_{i,j})}{F_p(W_p(n_{i,j}) + W_b(n_{i,j})L)} - \frac{u(n_{i,j})}{F_p}$$ (12)

In the above equation, $W_p(n_{i,j})$ is the average complexity weight of P frames, $W_b(n_{i,j})$ is the complexity weight of B frames, $Tb_l$ is the target buffer level, $W_p$ and $W_b$ are computed as follows:

$$W_p(n_{i,j}) = \frac{W_p(n_{i,j-1}) + 2W_p(n_{i,j-1})}{8}$$ (13)
$$W_b(n_{i,j}) = \frac{W_b(n_{i,j-1}) + 2W_b(n_{i,j-1})}{8}$$ (14)
$$W_p(n_{i,j}) = S_p(n_{i,j})Q_p(n_{i,j})$$ (15)
$$W_b(n_{i,j}) = \frac{S_b(n_{i,j})Q_b(n_{i,j})}{1.30}$$ (16)

$S_p$ and $S_b$ are the number of bits generated by encoding the corresponding frame, $Q_p$ and $Q_b$ are the quantization parameters, and $W_p$ and $W_b$ are the weight of the current P frame and B frame.

In the case that there is no B frame between two P frames, Equation (12) can be simplified as

$$Tb_l(n_{i,j+1}) = Tb_l(n_{i,j}) - \frac{Tb_l(n_{i,j}) - B_c(n_{i,j})}{N_p - 1}$$ (17)

It can be easily shown that $Tb_l(n_{i,N_p})$ is about $B_c(n_{i,j})/8$. 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 [1, 3], there usually exists difference between the actual buffer level and the target buffer level. We therefore need to compute a target bit for each P frame to reduce this difference. This is achieved by the following microscopic control.

Step 1.2 Microscopic control (target bit rate computation).

Using model (2) and linear tracking theory [6], the target bits allocated for the j-th P frame in the i-th GOP is determined based on the target buffer level, the frame rate, the available channel bandwidth and the actual buffer occupancy as follows:

$$f(n_{i,j}) = \frac{u(n_{i,j})}{F_p} + \gamma(Tb_l(n_{i,j}) - B_c(n_{i,j}))$$

where $\gamma$ is a constant, its typical value is 0.75 when there is no B frame and 0.25 otherwise. If the number of generated bits is around $f(n_{i,j})$, it can be easily shown that

$$B_c(n_{i,j+1}) = Tb_l(n_{i,j+1}) \approx (1 - \gamma)(B_c(n_{i,j}) - Tb_l(n_{i,j}))$$ (18)

Therefore, a tight buffer regulation can be achieved by choosing a large $\gamma$.

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

$$f(n_{i,j}) = \frac{W_p(n_{i,j} - 1)T_c(n_{i,j}) - B_c(n_{i,j})}{W_p(n_{i,j} - 1)N_p(j - 1) + W_b(n_{i,j} - 1)N_b(j - 1)}$$ (19)

The target bit is a weighted combination of the $f(n_{i,j})$ and $f(n_{i,j})$:

$$f(n_{i,j}) = \beta f(n_{i,j}) + (1 - \beta)f(n_{i,j})$$ (19)

where $\beta$ is a constant and its typical value is 0.5 when there is no B frame and 0.9 otherwise. It can be known from (19) that a tight buffer regulation can be achieved by choosing a small $\beta$.

To maintain the quality of the coded frames, the target bit $f(n_{i,j})$ is bounded by

$$f(n_{i,j}) = \max\{f(n_{i,j}), m_{hdr}(n_{i,j} - 1) + u(n_{i,j})\}$$ (20)

where $m_{hdr}(n_{i,j})$ is the number of bits used for the header and motion vectors by the previous P frame.

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

The MAD of the current P frame is predicted by model (1) using the actual MAD of the previous P frame. The quantization parameter $Q_{pe}$ corresponding to the target bit is then computed by using the quadratic model provided in [1, 3]. The details on this can be found in [1, 3, 4], it is thus not elaborated in this section.

To maintain the smoothness of visual quality among successive frames, the quantization parameter $Q_{pe}$ is further adjusted by

$$Q_{pe} = \min\{Q_{pp} + 2, \max\{Q_{pp} - 2, Q_{pe}\}\}$$ (21)

where $Q_{pp}$ is the quantization parameter of the previous P frame. Meanwhile, it should be bounded by the global bounds provided by H.264, i.e.

$$Q_{pe} = \min\{\max\{Q_{pe}, 1\}, 51\}$$ (22)

The quantization parameter is then applied to perform RDO for each MB in the current frame by using the method provided in [8, 9].

4.2. Post-encoding Stage

There are three tasks in this stage: update the parameters $a_1$ and $a_2$ of model (1), the parameters of quadratic R-D model, and determine the number of frames need to be skipped.

After encoding a frame, the parameters $a_1$ and $a_2$ of model (1),
as well as those of quadratic R-D model are updated. A method similar to that in [1, 3] is used where the sliding window size is computed by using the method provided in [4] instead of that in [1, 3].

5. EXPERIMENTAL RESULTS

The testing conditions are given in Table 1 and the experimental conditions are given as follows [10]:

<table>
<thead>
<tr>
<th>MV resolution</th>
<th>1/4 pel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hadamard</td>
<td>on</td>
</tr>
<tr>
<td>RDO</td>
<td>on</td>
</tr>
<tr>
<td>search range</td>
<td>±32</td>
</tr>
<tr>
<td>restrict search range</td>
<td>2</td>
</tr>
<tr>
<td>reference frame</td>
<td>1</td>
</tr>
<tr>
<td>symbol</td>
<td>CABAC</td>
</tr>
<tr>
<td>GOP structure</td>
<td>IBBPBBPBBP</td>
</tr>
</tbody>
</table>

Table 1 Testing Condition

Testing platform: JM6.1 [5]; Test sequences are container, news, foreman and silent; Format is QCIF size (4:2:0); Input frame rate is 30 frames/s, output frame rate is 15 frames/s; 150 frames are used for each sequence and the GOP length is 30; We compare our scheme with the fixed quantization parameter case where the fixed quantization parameters are 28, 32, 36 and 40, respectively [10]. The initial quantization parameters of our scheme are 26, 30, 34 and 38, respectively. The experimental results are listed in the following tables:

Table 2. QP=28, Comparison results

<table>
<thead>
<tr>
<th>video sequence</th>
<th>PSNR (ours) (fixed QP)</th>
<th>bandwidth (ours) (fixed QP)</th>
<th>PSNR (fixed QP)</th>
<th>bandwidth (fixed QP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>news</td>
<td>37.75dB</td>
<td>60.66kb/s</td>
<td>37.08dB</td>
<td>60.36kb/s</td>
</tr>
<tr>
<td>container</td>
<td>36.86dB</td>
<td>30.49kb/s</td>
<td>36.43dB</td>
<td>30.38kb/s</td>
</tr>
<tr>
<td>foreman</td>
<td>36.27dB</td>
<td>96.21kb/s</td>
<td>35.88dB</td>
<td>93.88kb/s</td>
</tr>
<tr>
<td>silent</td>
<td>36.30dB</td>
<td>63.62kb/s</td>
<td>35.99dB</td>
<td>63.77kb/s</td>
</tr>
</tbody>
</table>

Table 3. QP=32, Comparison results

<table>
<thead>
<tr>
<th>video sequence</th>
<th>PSNR (ours) (fixed QP)</th>
<th>bandwidth (ours) (fixed QP)</th>
<th>PSNR (fixed QP)</th>
<th>bandwidth (fixed QP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>news</td>
<td>34.77dB</td>
<td>37.88kb/s</td>
<td>34.02dB</td>
<td>37.61kb/s</td>
</tr>
<tr>
<td>container</td>
<td>34.21dB</td>
<td>17.30kb/s</td>
<td>33.69dB</td>
<td>17.07kb/s</td>
</tr>
<tr>
<td>foreman</td>
<td>33.74dB</td>
<td>55.75kb/s</td>
<td>33.26dB</td>
<td>53.85kb/s</td>
</tr>
<tr>
<td>silent</td>
<td>33.80dB</td>
<td>37.57kb/s</td>
<td>33.08dB</td>
<td>37.55kb/s</td>
</tr>
</tbody>
</table>

Table 4. QP=36, Comparison results

<table>
<thead>
<tr>
<th>video sequence</th>
<th>PSNR (ours) (fixed QP)</th>
<th>bandwidth (ours) (fixed QP)</th>
<th>PSNR (fixed QP)</th>
<th>bandwidth (fixed QP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>news</td>
<td>31.78dB</td>
<td>23.13kb/s</td>
<td>31.08dB</td>
<td>22.84kb/s</td>
</tr>
<tr>
<td>container</td>
<td>31.55dB</td>
<td>10.57kb/s</td>
<td>31.06dB</td>
<td>10.32kb/s</td>
</tr>
<tr>
<td>foreman</td>
<td>31.25dB</td>
<td>33.08kb/s</td>
<td>30.64dB</td>
<td>31.40kb/s</td>
</tr>
<tr>
<td>silent</td>
<td>30.82dB</td>
<td>21.57kb/s</td>
<td>30.49dB</td>
<td>21.34kb/s</td>
</tr>
</tbody>
</table>

Table 5. QP=40, Comparison results

It can be shown from Tables 2-6 that the average PSNR is improved up to 0.75dB by our scheme. The average value of average improved PSNR is 0.45dB.

6. CONCLUSION

This paper proposed an adaptive frame layer rate control scheme for H.264 by introducing a linear model to predict the mean absolute difference (MAD) of current frame. The target bit rate for each frame is computed by adopting a fluid flow traffic model and linear tracking theory. 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 frame by using the quantization parameter. Both constant bit rate (CBR) and variable bit rate (VBR) cases are studied in this paper.

References