IMPROVED FRAME-LAYER RATE CONTROL FOR H.264 USING MAD RATIO

Minqiang Jiang, Xiaoquan Yi, and Nam Ling
Department of Computer Engineering
Santa Clara University
Santa Clara, California 95053, USA
Email: {mjiang, xyi, nling}@scu.edu

ABSTRACT
In recent years, rate control plays an increasing important role in real-time video communication applications using MPEG-4 AVC/H.264. An important step in many existing rate control algorithms, which employs the quadratic rate-distortion (R-D) model, is to determine the target bits for each P frame. This paper aims in improving video distortion, due to high motions or scene changes, by more accurately predicting frame complexity using the statistics of previously encoded frames. We use mean absolute difference (MAD) ratio as a measure for global frame encoding complexity. Bit budget is allocated to frames according to their MAD ratio, combined with the bits computed based on their buffer status. Simulation results show that the H.264 coder, using our proposed algorithm with virtually little computational complexity added, effectively alleviates PSNR surges and sharp drops for frames caused by high motions or scene changes.

1. INTRODUCTION
Many existing video encoding rate control schemes (e.g. [2]-[4]) are based on a quadratic rate-quantizer model [1]. In [2], Lee et al. proposed a scalable rate control scheme. It was shown that their algorithm based on a more accurate second-order R-D model can be simultaneously applied to different coding contexts, such as frame level, object level, and macroblock (MB) level, and that an improved sliding window method was used to select data points for updating model parameters. Pan et al. [3] proposed several improvements over the MPEG-4 Q2 rate control scheme [11], using new frame bit allocation method and target buffer function to further improve the perceptual quality of the reconstructed video. A new least-mean-square estimation method of the R-D model parameters was developed in [4]. However, the target bit estimation for a frame, as an important factor in determining the quantization parameter (QP), is not allocated based on frame complexity in the aforementioned rate control schemes.

As we know, the key role of quadratic rate-quantizer model is to determine the QP prior to encoding a frame. After initialization, the subsequent important step is to estimate the target bits for the new P frame [2], [3], [5]. First, the initial target is determined by the number of remaining bits (95%) and the actual bits used for the previous P frame (5%). Then, this initial estimate is scaled based on buffer fullness and buffer size. The purpose of the adjustment is to get a better target bit estimation and maintain the target buffer level at 50% after encoding each frame. In [6], the number of target bits selected throughout the whole video sequence is nearly constant, and is usually corrected by a small value from the feedback of the buffer fullness. In [7] and [8], the target bit number is a weighted combination of remaining bits (50%) when there is no B frames and bits calculated from buffer regulation.

To estimate the target bits for each frame, a common straightforward way is used in [1]-[6], namely, an equal number of bits is allocated to each frame regardless of its complexity. In other words, the initial target bit count for a P frame at time \( t + 1 \) is estimated as the quotient of the remaining bits being divided by the remaining number of P frames. There are basically three problems associated with this method. First, it reflects little global frame complexities over a group of pictures (GOP), resulting in poor target bit estimation for different frames. Second, the target bits count is determined solely based on the buffer fullness. That is, a buffer nearly full will allocate less target bits to a new frame while a nearly empty buffer will allocate more bits, which will lead to a much smaller quantization parameter regardless of the complexity of frame content. This may result in non-uniform distortion over a video sequence. Third, all the algorithms mentioned above do not handle scene changes efficiently. Many macroblocks (MBs) in the subsequent frame after scene change may need to be encoded in intra-mode and need more bits or else it may cause a serious degradation in picture quality. The scheme in [7] and [8] slightly improves bit allocation by considering the distance of each P frame from the initial I frame in a GOP.

In this paper, we focus on solving the above issues by using the mean absolute difference (MAD) ratio [10] and an adaptive initial target bits estimation strategy among different frames according to the coding complexities of each frame, to improve the quality of frames with high motions or scene changes. The organization of the rest paper is as follows. The next section reviews the rate control algorithm for estimating the target bits presented in [7], which forms the base for our new algorithm. Section 3 presents our detailed proposed algorithm. Our extensive experiment results are provided in Section 4. This paper concludes with Section 5.

2. TARGET BIT ESTIMATION
According to [7], frame layer rate control scheme consists of two stages: pre-encoding and post-encoding. The objective of the first stage is to compute quantization parameter for all frames and it comprises of two sub steps: (a) determine a target bit for each P frame and (b) compute the quantization parameter and perform rate distortion optimization (RDO). In the pre-encoding stage, a quadratic rate-distortion (R-D) model is used to calculate the corresponding quantization parameters. In the post-encoding stage, the model parameters are continually updated and frame-skipping control is performed. In this section, we will only review and summarize the method used for estimating the target bits. For complete algorithm description, please refer to [7].

To estimate target bits for the current frame, a fluid traffic model is employed, which is based on the linear tracking theory [3]. For simplicity, let us assume one group of pictures (GOP) is used and the video sequence is encoded first as an I frame, and subsequently P frames. To illustrate rate control modeling, let \( n_j \) denotes the total number of frames in a GOP, \( n_j \) denotes the \( j \)th frame, and \( B_v(n, j) \) denotes the occupancy of virtual buffer after coding the \( j \)th frame. The fluid traffic model is stated as.

To estimate the target bits for each frame, a common straightforward way is used in [1]-[6], namely, an equal number of bits is allocated to
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. The determination of target bits for current P frame is composed of two steps.

**Step 1** Target buffer occupancy determination. Since the quantization parameter of the first P frame is given at the GOP layer in this algorithm, the initial value of target buffer level is set as

\[
T_{bl}(n_1) = B_s(n_1)
\]

Then the target buffer levels of other P frames in the GOP are predefined by using the following function

\[
T_{bl}(n_{j+1}) = T_{bl}(n_j) - \frac{T_{bl}(n_j) - B_s(n_j)}{N_p - 1}
\]

where \(N_p\) is the total number of P frames in the GOP.

**Step 2** Target bit rate computation. By using linear tracking theory, the target bits allocated for the \(j\)th frame is determined based on the target buffer level, the frame rate, the available channel bandwidth, and the actual buffer occupancy as follows:

\[
T_{buf} = \frac{u(n_j)}{F_r} + \gamma(T_{bl}(n_j) - B_s(n_j))
\]

where \(\gamma\) is a constant and its typical value is 0.75. Meanwhile, the remaining bits are also computed as

\[
T_r = \frac{R}{N_r}
\]

where \(R\) is the number of bits remaining for encoding this sequence, and \(N_r\) is the number of P frames remaining for encoding. The final target bit is a weighted combination of \(T_r\) and \(T_{buf}\)

\[
T = \beta T_r + (1 - \beta) T_{buf}
\]

where \(\beta\) is a weighting factor and its typical value is 0.5.

Our proposed improvement: In (5), the remaining bits \(R\) is allocated to all non-coded frames equally. In our new scheme, we focus on \(T_r\), that is, the remaining bits should be un-equally distributed to all non-coded frames according to frame complexities in the initial target bit estimation step. In other words, \(T_r\) will no longer be the same for different frames with different encoding complexities, therefore we have a better target bit \(T\) for the current frame. Details of the improvements will be discussed in Section 3.

### 3. OUR IMPROVED SCHEME USING MAD RATIO

It is well known that MAD of the residual component can be a good indication of encoding complexity [2, 4]. In the quadratic rate-quantizer (R-Q) model, the encoding complexity is usually substituted by MAD. To solve the distribution of the bit budget to different frames, we propose a new frame complexity parameter, namely, MAD RATIO.

#### 3.1. Measure of Frame Encoding Complexity

Similar as in [10], MAD ratio is used to measure the complexity of a frame, which is the ratio of the predicted MAD of current frame to the average MAD of all previously encoded P frames in the GOP. It can be easily calculated from the following equation:

\[
MAD_{RATIO}(i) = \frac{MAD_P}{\sum_{j=1}^{i-1} MAD_{A_j}}/(i-1)
\]

where

\(MAD_P\) Predicted MAD of current frame \(i\);

\(MAD_{A_j}\) Actual MAD of previously encoded P frame \(j\);

\(i - 1\) The number of previously encoded P frames in the GOP.

Note that \(MAD_P\) is calculated using a linear prediction model [7]

\[
MAD_P = a_1 \cdot MAD_{A_{j-1}} + a_2
\]

where \(a_1\) and \(a_2\) are the 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 frame. We use the average MAD of all previous P frames as an indication of the average complexity for the whole video sequence. It is updated after encoding each frame. Figure 1 (a) shows the MAD RATIO values of a number of test sequences. It can be seen that MAD RATIO varies significantly from one sequence to another. MAD RATIO also varies within the same sequence at the junction of scene changes, such as in the cases of "Suzie_Trevor". Figure 1 (b) depicts the inverse of MAD RATIO for three sequences. We observe that the inverses of MAD RATIO are consistent with the PSNR plots (dotted line) in Figures 2 (a) and (b). The MAD RATIO forms our measure for global encoding complexity.

![Fig. 1 – (a) MAD RATIO (b) Inverse of MAD RATIO versus frame number.](image)

**3.2. Adaptive Target Bit Estimation Control**

We found that MAD RATIO is a simple and accurate measure of frame complexity. Therefore, it provides a mechanism to control the target bits estimation. The distribution of the bit count is scaled by a function of
MADRATIO. For computational simplicity, we use the following formula to adjust the initial target bits $T_r$. It can be pseudo-coded as

Calculate the average MAD of all previously coded P frames;
Calculate the MADRATIO using $\text{predicted} \_\text{MAD} \_\text{of} \_\text{the} \_\text{current} \_\text{frame} / \text{average} \_\text{MAD};$

IF (MADRATIO < 0.9) THEN
$T_r = T_r \times 0.05$
ELSE IF (MADRATIO < 1.0) THEN
$T_r = T_r \times \text{MADRATIO} \times 0.6$
ELSE IF (MADRATIO < 1.8) THEN
$T_r = T_r \times \text{MADRATIO} \times 0.7$
ELSE IF (MADRATIO >= 1.8) THEN
$T_r = T_r \times 1.8$

Meanwhile, we increase $\beta$ to 0.7 from 0.5 in (6) so that $T_r$ has more weight than $T_{buf}$. Note that the function parameters used in the above function came from our empirical experiments. The basic idea is to set $T_r$ smaller if the current frame complexity is low and set $T_r$ larger if the current frame complexity is high. For instance, a frame with MADRATIO=0.8 is simple hence it needs less bits and a frame with MADRATIO=2.0 is complex which reflects high motions or scene changes. The objective of this improvement is to save bits from those frames with relatively less complexity and allocate more bits to frames with higher complexity due to high motion or scene changes. The final target bits $T$ for the new P-frame can be calculated using equation (6).

4. EXPERIMENTAL RESULTS

Numerous experiments have been conducted to evaluate the performance of our improved rate-control algorithm under constant bit rate. All test sequences used are in QCIF 4:2:0 format. The test platform is JM6.1 [9]. The bit rates are generated by coding the test sequences with fixed quantization parameters 26, 28, 32, 36, and 40. They are the target bit rates for an encoder with our rate control scheme. The test video “Suzie_Trevor” with scene changes is formed by concatenating two videos “Suzie” and “Trevor”. Results show that our proposed scheme has significant improvements on PSNR surge and sharp drop for various bit rates (e.g., 3 kbits/s to 150 kbits/s) and for various temporal resolutions (e.g., 7.5-30 fps). Some examples are given in Fig. 2. In Fig. 2 (b), due to the scene change, after the 50th frame, the PSNR of “Suzie_Trevor” using rate control scheme in [9] drops to 23.57 dB whereas it drops only to 30.80 dB using our proposed scheme, achieving a gain of 6.23 dB. The average PSNR improvement is 0.53 dB for “Suzie_Trevor”. For most sequences, the average PSNRs are about the same using our scheme and the scheme in [9]. As shown in Table I, our method reduces the standard deviations of PSNR significantly for all pictures. The percentages of reductions are 15, 20, and 46% for the three sequences, respectively.

Table I - PSNRs (mean and standard deviation) and bit rates for sequences (QCIF).

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Carphone</th>
<th>Suzie</th>
<th>Suzie_Trevor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean PSNR (dB)</td>
<td>37.04</td>
<td>37.01</td>
<td>37.1</td>
</tr>
<tr>
<td>PSNR Std. Dev.(dB)</td>
<td>3.23</td>
<td>2.75</td>
<td>1.69</td>
</tr>
<tr>
<td>Bit Rate (kbps)</td>
<td>68.79</td>
<td>68.77</td>
<td>88.84</td>
</tr>
</tbody>
</table>

Fig. 3 shows the buffer-level variations of “Suzie_Trevor”. Results in [9] show underflow, whereas using our proposed scheme, the buffer is virtually never overflowed or underflowed. The actual buffer using our scheme is kept closer to its target buffer level than that using the algorithm in [9]. Figure 4 shows the bit usage for “Suzie_Trevor”. The actual bits provided by our method are closer to the target bit level than that in [9]. PSNR gain is reflected in the improvement in the subjective quality of the reconstructed frames, as shown in Figure 5 for two typical frames of the two test sequences “Carphone” and “Suzie_Trevor”.

5. CONCLUSION

In this paper we presented a few simple yet effective new techniques to improve MPEG-4 AVC/H.264 rate control algorithm so that it can maintain a video stream with a smoother PSNR variation which is highly desirable in real-time video coding and transmission. The proposed approach uses the MAD statistics of previously encoded frames to predict the current frame’s complexity and uses MAD ratio to regulate target bit estimation. Our extensive experimental results show that our proposed approach outperforms the recent rate control algorithm proposed in [9]. Furthermore, the proposed improvements can be extended to MB level rate control.

Fig. 2 - PSNR results for sequence (a). “Carphone” encoded at 68.78 kbps, (b) “Suzie_Trevor” encoded at 55.18 kbps.
Fig. 3 - Buffer fullness results for sequence “Suzie T. Trevor.”

Fig. 4 – Actual bits and target bits for sequence “Suzie T. Trevor.”

6. REFERENCES


