Abstract

In some video applications, a variable bit rate (VBR) video bitstream will be transmitted over a constant bit rate (CBR) transmission channel, in which a channel buffer is employed. In this study, a rate control scheme for H.264 video transmission is proposed.

The target number of bits for each video frame is first obtained. Three rate-quantization (R-Q) models are proposed to approximate the relationship between the number of coding bits for a video frame, \( B \), and the quantization step size, \( Q \), in the frame layer. The best model is selected to determine the candidate QP for the current video frame. After the coding complexity of each macroblock (MB) is estimated by using the candidate QP, the Lagrange multiplier is used to optimally select the quantization step sizes \( \{Q_{mb,1}, Q_{mb,2}, \ldots, Q_{mb,N}\} \) for the \( N \) MBs (macroblocks) in a video frame to minimize the distortion, subject to the given bit budget constraint. Based on the simulation results obtained in this study, the proposed approach can meet the target bit rate more accurately, keep a larger average frame rate, and higher PSNR values than the two quadratic approaches for comparison.

1. Introduction

In some video applications, a variable bit rate (VBR) video bitstream will be transmitted over a constant bit rate (CBR) transmission channel, in which a channel buffer is usually employed [1]. In this study, a rate control scheme for H.264 video transmission is proposed.

Rate control can be generally separated into two steps: (1) to allocate a target number of bits for each allocation unit, (2) to apply a rate control scheme to meet the target number of bits. For target bit allocation, it is important to distribute an adequate bit budget for each allocation unit with an identical perceptual quality. A feasible bit allocation strategy will allocate the available bits according to the content activities of each allocation unit [2]. Several models are proposed to estimate the target number of bits allocated to a video frame or an MB [2]-[5]. The quantization scale can be employed to control the bit rate according to the content activities of a video sequence and the planned buffer fullness. For example, the quantization scale for an MB in MPEG TM5 [2] is decided by the normalized MB activity property calculated from the minimum variance of the luminance frame- or field-based subblock of 8×8 pixels in an MB as well as the planned buffer fullness. Jordi and Lei [6] presented a rate control method for operating typical DCT video codecs for video communications. One type of rate control approaches uses an explicit bit model to predict the number of compressed bits when a certain quantization scale is used, as the control schemes developed in [7]-[8]. The paper is organized as follows. The proposed rate control scheme is addressed in Section 2. Simulation results are given in Section 3, followed by concluding remarks.

2. Proposed rate control scheme

2.1. Frame-layer rate control

In the frame layer, the sum of the absolute difference after the Hadamard transform (SATD) [1], \( \delta \), of a video frame is employed as the coding complexity of the video frame. The target number of bits for a video frame, \( T \), is simply given by

\[
T = \min\{\max\{\tilde{\delta}_c + R_y - B_f, 0.5 \times B_c + R_y - B_f\}\},
\]

where \( \tilde{\delta}_c \) is the SATD of the current frame, \( \tilde{\delta}_{\text{avg}} \) is the average SATD of all previous coded frames, \( R_y \) is the number of remaining bits for the video sequence, \( N_r \) is the number of remaining frames for the video sequence,
$R_p$ is the average number of bits per video frame, $B_s$ is the buffer size, and $B_i$ is the current buffer fullness (bits).

As the relationships between $\delta$ and $B$ for different QP’s and that between $Q$ and $B$ for different $\delta$’s shown in Figs. 1 and 2, respectively, three models, namely, the quadratic formula, the logarithmic formula, and the exponential formula are employed to approximate the relationship between $Q$ and $B$ as

$$B_i = X_1 \times Q_{i}^{-2} + X_2 \times \delta_i \times Q_{i}^{-1} + X_3,$$  \hspace{0.5cm} (2)

$$B_i = Y_1 \times \delta_i \times (\ln Q_{i})^{-1} + Y_2,$$ \hspace{0.5cm} (3)

$$B_i = Z_1 \times \delta_i \times 2^{-Q_{i}} + Z_2,$$ \hspace{0.5cm} (4)

where $X_1$, $X_2$, $X_3$, $Y_1$, $Y_2$, $Z_1$, and $Z_2$ are the model parameters for the three models and $\delta_i$ is the SATD of the $i$-th video frame. The best model among the three models will be finally used to estimate the best candidate QP for each video frame. Note that the quantization array (table) $A(i)$ is employed here [1].

### 2.2. Selection of candidate QP for each frame

The proposed procedure for selecting the candidate QP, $Q_{i,frame}$, for each frame is described as follows.

**Step 1.** Target bit allocation using Eq. (1).

**Step 2.** The candidate quantization step sizes, $Q_{a,i}$, $Q_{b,i}$, $Q_{c,i}$, using the quadratic formula, the logarithmic formula, and the exponential formula of the current $i$th video frame are given, respectively, by

$$Q_{a,i} = \sqrt{(X_2 \times 2^{-\delta_i} - 4 \times X_1 \times (X_3 - T_i)) - X_2 \times \delta_i},$$

$$Q_{b,i} = \exp(-T_i),$$

$$Q_{c,i} = \log_{10}\left(\frac{Z_1 \times \delta_i}{T_i - Z_2}\right).$$  \hspace{0.5cm} (5)

The candidate quantization step size $Q_i$ of the current $i$th video frame is given by

$$Q_i = \begin{cases} Q_{a,i}, & \text{if } 1.1 \times err_\alpha < err_a \text{ and } 1.1 \times err_\beta < err_c, \\ Q_{b,i}, & \text{if } 1.1 \times err_\alpha < err_a \text{ and } 1.1 \times err_\beta < err_c, \\ Q_{c,i}, & \text{if } 1.1 \times err_\beta < err_c \text{ and } 1.1 \times err_\alpha < err_a, \\ \text{avg}(Q_{a,i}, Q_{b,i}, Q_{c,i}), & \text{otherwise}, \end{cases}$$ \hspace{0.5cm} (6)

where $err_\alpha$, $err_\beta$, and $err_c$ are estimated, respectively, by

$$err_\alpha = \frac{\sum_{k=1}^{w} X_1 \times Q_{a,k}^{-2} + X_2 \times \delta_{a,k} \times Q_{a,k}^{-1} + X_3 - B_{i+k-1}}{w},$$

$$err_\beta = \frac{\sum_{k=1}^{w} Y_1 \times \delta_{a,k} \times (\ln Q_{a,k})^{-1} + Y_2 - B_{i+k-1}}{w},$$

$$err_c = \frac{\sum_{k=1}^{w} Z_1 \times \delta_{a,k} \times 2^{-Q_{a,k}} + Z_2 - B_{i+k-1}}{w}.$$ \hspace{0.5cm} (7)

Finally, the candidate QP for the current $i$th video frame is determined as the nearest integer in the quantization array (table) $A(i)$.

**Step 3.** Perform macroblock layer rate control using $Q_{i,frame}$.

**Step 4.** Update $X_i$, $X_2$, $X_3$, $Y_i$, $Y_2$, $Z_i$ and $Z_2$ using the encoding relationships between $B_i$ and $Q_i$ in the previous $w$ video frames.

**Step 5.** $N_c = N_c - 1$. If $N_c = 0$, stop; otherwise, go to Step 1 to process the next frame.

### 2.3. Macroblock-layer rate control

...
The bit estimation model relating the number of coding bits for the header and motion vectors for the video frame. It can be derived that

\[ Q_{mb,i} = \frac{k}{\tilde{B}_i - H_i} \times \frac{\alpha_i}{\sigma_i} \times \sum_{i=1}^{N} \sigma_i, \quad i = 1,2,\ldots,N. \]  

Moreover, using the previous i-1 encoded MBs, the optimized QP for the ith MB becomes

\[ Q_{mb,i} = \sqrt{k} \times \frac{\alpha_i}{\sigma_i} \times \sum_{i=1}^{N} \sigma_i. \]  

3. Simulation results

Here three video sequences, “Container,” “Hall Monitor,” and “Mother Daughter,” are employed. Each sequence containing 100 frames is encoded at three channel bit rates (R), 24 kbps, 33.6 kbps, and 48 kbps, with the frame rate (F) being 10 fps. The first frame of each video sequence is intra coded with QP=35, and the other frames are all P-frames.

A well-known quadratic rate control scheme [3] is implemented on H.264 in both the frame layer and the macroblock layer, denoted as Quadratic_1 and Quadratic_2, for comparison. Four performance measures, namely, (1) the average PSNR, PSNRseq, (2) the average accuracy of achieving the target number of bits, ACC,seq, (3) the average accuracy of achieving the average target number of bits per frame, ACC_MFs, and (4) the number of skipped video frames, for a video sequence are employed.

In terms of PSNRseq (dB), the performance comparison between the two quadratic approaches for comparison and the proposed approach for the three video sequences is listed in Table 1. In terms of ACCseq, ACC_MFs, the average bit rate, and the number of skipped video frames, the performance comparison between the two quadratic approaches for comparison and the proposed approach for the three video sequences is listed in Table 2. In terms of the number of coding bits per video frame, the performance comparison between the two quadratic approaches for comparison and the proposed approach for the video sequence “Container” at channel bit rate 48 kbps is shown in Fig. 3.

4. Concluding remarks

Based on the simulation results obtained in this study, several observations can be found. (1) On the average, the larger the channel bit rate is, the better the rate control results will be. (2) The less content activities of a video sequence are, the more accuracy of achieving the target bit rate will be. (3) Based on Table 2, the proposed approach achieves a bit rate closer to the target channel bit rate than the two quadratic approaches for comparison do, especially for the test video sequences containing quick motions or scene...
changes. (4) Based on Fig. 3, the proposed approach achieves a nearly constant number of coding bits per video frame (with smaller oscillations), whereas the two quadratic approaches for comparison oscillate with the larger amplitudes. (5) Based on Table 1, the $PSNR_{seq}$ values of the proposed approach are slightly better than that of the two quadratic approaches for comparison. (6) Based on Table 2, in some situations, Quadratic_1 skips the largest number of video frames, Quadratic_2 skips a medium number of video frames, and the proposed approach skip very few video frames.

As a summary, the proposed approach can meet the target bit rate more accurately, keep a larger average frame rate, and higher $PSNR$ values than that of the two quadratic approaches for comparison. This shows the feasibility of the proposed approach.

Fig. 3. The number of coding bits per video frame for the “Container” sequence by the two quadratic approaches for comparison and the proposed approach at channel bit rate 48 kbps. The horizontal straight line indicates the average target bits per frame ($M = R/F$).

5. References


Table 1. The simulation results, in terms of $PSNR_{seq}$ (dB), for the three video sequences by the two quadratic approaches for comparison and the proposed approach.

<table>
<thead>
<tr>
<th>Video sequence</th>
<th>Target channel bit rate (kbps)</th>
<th>Quadratic 1</th>
<th>Quadratic 2</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$PSNR_{seq}$</td>
<td>$PSNR_{seq}$</td>
<td>$PSNR_{seq}$</td>
<td>$PSNR_{seq}$</td>
</tr>
<tr>
<td>Container</td>
<td>24.00</td>
<td>36.73</td>
<td>35.65</td>
<td>37.18</td>
</tr>
<tr>
<td></td>
<td>33.60</td>
<td>37.99</td>
<td>37.37</td>
<td>38.41</td>
</tr>
<tr>
<td></td>
<td>48.00</td>
<td>39.33</td>
<td>39.02</td>
<td>39.78</td>
</tr>
<tr>
<td>Hall Monitor</td>
<td>24.00</td>
<td>37.11</td>
<td>36.48</td>
<td>37.49</td>
</tr>
<tr>
<td></td>
<td>33.60</td>
<td>38.66</td>
<td>38.22</td>
<td>39.13</td>
</tr>
<tr>
<td></td>
<td>48.00</td>
<td>40.15</td>
<td>39.87</td>
<td>40.43</td>
</tr>
<tr>
<td>Mother and Daughter</td>
<td>24.00</td>
<td>38.17</td>
<td>37.68</td>
<td>38.50</td>
</tr>
<tr>
<td></td>
<td>33.60</td>
<td>39.53</td>
<td>39.30</td>
<td>39.87</td>
</tr>
<tr>
<td></td>
<td>48.00</td>
<td>40.92</td>
<td>40.86</td>
<td>41.26</td>
</tr>
</tbody>
</table>

Table 2. The coding results, in terms of the average bit rate, $AF_{seq}$, $AC_{seq}$, and the number of the skipped video frames, for the three video sequences by the two quadratic approaches for comparison and the proposed approach.

<table>
<thead>
<tr>
<th>image sequence</th>
<th>Target channel bit rate (kbps)</th>
<th>Quadratic 1</th>
<th>Quadratic 2</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obtained average bit rate (kbps)</td>
<td>$AF_{seq}$ (%)</td>
<td>$AC_{seq}$ (%)</td>
<td>Frames skipped</td>
</tr>
<tr>
<td>Container</td>
<td>24.00</td>
<td>23.35</td>
<td>31.05</td>
<td>24.88</td>
</tr>
<tr>
<td></td>
<td>33.60</td>
<td>32.81</td>
<td>26.86</td>
<td>23.23</td>
</tr>
<tr>
<td></td>
<td>48.00</td>
<td>47.15</td>
<td>25.21</td>
<td>21.92</td>
</tr>
<tr>
<td>Hall Monitor</td>
<td>24.00</td>
<td>23.77</td>
<td>20.95</td>
<td>15.62</td>
</tr>
<tr>
<td></td>
<td>33.60</td>
<td>33.53</td>
<td>14.66</td>
<td>12.34</td>
</tr>
<tr>
<td></td>
<td>48.00</td>
<td>47.70</td>
<td>14.05</td>
<td>12.89</td>
</tr>
<tr>
<td>Mother and Daughter</td>
<td>24.00</td>
<td>23.58</td>
<td>28.92</td>
<td>27.21</td>
</tr>
<tr>
<td></td>
<td>33.60</td>
<td>33.11</td>
<td>29.48</td>
<td>26.88</td>
</tr>
<tr>
<td></td>
<td>48.00</td>
<td>47.26</td>
<td>23.62</td>
<td>24.87</td>
</tr>
</tbody>
</table>