RELATION BETWEEN VIDEO BITRATE AND FRAME SIZE
IN ARBITRARY DOWNSIZING TRANSCODING

HaiWei Sun, Yap-Peng Tan and YongQing Liang

School of Electrical & Electronic Engineering
Nanyang Technological University, Singapore

ABSTRACT

In arbitrary downsizing video transcoding [1], although the transcoder can adjust the size of a pre-coded video to meet the requirement of different transmission channels or receiving devices, there is however no model to illustrate the relation between bitrate and frame size. In this paper, we propose a new method to model the relation between bitrate and frame size under low bitrate condition. With this new method, we can re-estimate the bitrate according to required frame size or select a suitable frame size for a given target bitrate, while maintaining good video quality. Experimental results are presented to demonstrate the effectiveness of the proposed video transcoding methods.

1. INTRODUCTION

By using arbitrary downsizing video transcoding technique [1], we have more choices in adjusting the spatial resolution of a pre-compressed video according to the available channel bandwidth. Given a target bitrate, however, it is not clear which reduced size can provide optimal video quality. Moreover, if users want to select the display size, the bitrate which is required for the size is also not clear. Thus, we focus on modelling the relation between the frame size of the transcoded video and its required bitrate. With the relation between bitrate and frame size, we can apply it into two application as follows:

1) If the available channel bandwidth is insufficient, we can select a reduced frame size for the transcoded video to meet the available bandwidth.

2) If users want to change frame size during video transmission, we can estimate a suitable bitrate which is required to transcode the video.

In order to estimate the relation between bitrate and frame size, we want to build a model to estimate the required bitrate given a pre-coded video with its frame size and bitrate, while maintaining video quality or distortion from pre-coded video. To simplify the estimation of the relation, same mean quantization parameter Q for all macroblocks is used in both pre-coded video and downsizing transcoded video.

Since we care more for downsizing transcoding under low bitrate, we set Q as 32 and frame rate as 10 to stimulate low bitrate case in our experiments. Six standard test videos, “Tempe”, “Foreman”, “Stefan”, “Mobile & Calendar (MC)”, “Mother & Daughter (MD)”, and “News” are used in our experiments.

2. RELATION BETWEEN VIDEO BITRATE AND FRAME SIZE

The total number of bits allocated for a frame can be estimated by the total number of bits used to code motion vectors ($R_{MV}$), frame residue ($R_r$), overhead ($R_H$) as follows [3].

$$ R = R_{MV} + R_r + R_H $$

where parameter $R_M$ in our experiments is treated as bits allocated to express layer information, including bits allocated for picture layer, bits allocated for macroblock layer, bits allocated for block layer, etc [6]. To estimate the relation between bitrate and frame size, we consider $R_{MV}$, $R_r$, $R_H$ separately.

2.1. RELATION BETWEEN BITRATE FOR MOTION VECTOR AND FRAME SIZE

Considering the relation between bitrate for motion vector and frame size, we notice that papers [3, 4, 5] present methods to estimate motion vector rate, or bits/pixel for motion vector as follows:

$$ R_M = \frac{1}{B^2} \log_2 \left( \frac{4e^2B}{\Delta^2} \right) + \frac{1}{B^2} \log_2 \left( \frac{\sigma_V^2 \ln \left( \frac{1}{a} \right)}{\tilde{a}} \right) \quad (2) $$

where parameters $\sigma_V^2$ and $\tilde{a}$ are the average variance and average correlation coefficient of motion vectors for one frame. Parameter $\Delta$ is the accuracy of motion vectors. Parameter $B$ is the size of motion compensated block for frame residue.

So the total number of bits for motion vectors ($R_{MV}$) within a frame can be expressed as the total number of pixels times the motion vector rate. Thus, we obtain an expression for $R_{MV}$ as:

$$ R_{MV} = S \cdot A \times R_M $$

where parameter $S$ is the total number of macroblocks in a frame. Parameter $A$ is the total number of pixels in a macroblock (i.e., $A=16^2$). Thus, $S \cdot A$ stands for the total number of pixels in one frame.

Moreover, after a further analysis, we ignore the second part on the right of the equation (2) and simplify the expression for $R_{MV}$.

The reason is illustrated below.

As for the first part on the right of the equation (2), if we fix $B$ as 8 for the size of block, $\Delta$ as 0.5 for half pixel resolution, we obtain a approximation of the first part:

$$ \frac{1}{B^2} \log_2 \left( \frac{4e^2B}{\Delta^2} \right) \approx \frac{1}{64} \log_2 947 \quad (4) $$

While the second part in (2):

$$ \frac{1}{B^2} \log_2 \left( \frac{\sigma_V^2 \ln \left( \frac{1}{a} \right)}{\tilde{a}} \right) \approx \frac{1}{64} \log_2 (0.002\sigma_V^2) \quad \text{; where } \tilde{a} \text{ is set as } 0.998 \quad (5) $$

We calculate the value of $\sigma_V^2$ by measuring the variance for the $x$ and $y$ components of the block $(8 \times 8$ pixels) motion vectors and average the two variance [3]. If the value of 0.002$\sigma_V^2$ in (5) is comparable with 947 in (4), we should not ignore the second part. However, parameter $\sigma_V^2$ is not large enough. For example, we calculate the the value of $\sigma_V^2$ from “Foreman”:

Table 2 show that the variance is no more than 40 in “Foreman”. So the value of 0.002$\sigma_V^2$ is less than 0.08. So the value of 0.002$\sigma_V^2$ is far less than 947. Moreover, video “Foreman” includes a lot of motions compared to other videos. In other word, the variance of motion vectors from other videos should not be much larger than “Foreman”. Thus, we ignored the second part compared with the
The difference of the estimated results compared to Fullsearch method. Results are shown as follows.

<table>
<thead>
<tr>
<th>Frame size</th>
<th>Foreman</th>
<th>Stegan</th>
<th>News</th>
<th>Md</th>
<th>Mc</th>
</tr>
</thead>
<tbody>
<tr>
<td>364 × 272</td>
<td>364.06</td>
<td>391.87</td>
<td>382.57</td>
<td>312.72</td>
<td>320.47</td>
</tr>
<tr>
<td>256 × 208</td>
<td>279.85</td>
<td>308.27</td>
<td>308.27</td>
<td>238.57</td>
<td>246.32</td>
</tr>
<tr>
<td>256 × 208</td>
<td>279.85</td>
<td>308.27</td>
<td>308.27</td>
<td>238.57</td>
<td>246.32</td>
</tr>
<tr>
<td>176 × 144</td>
<td>176.57</td>
<td>189.13</td>
<td>189.13</td>
<td>119.43</td>
<td>127.19</td>
</tr>
<tr>
<td>128 × 96</td>
<td>128.39</td>
<td>141.95</td>
<td>141.95</td>
<td>72.25</td>
<td>80.01</td>
</tr>
</tbody>
</table>

Table 2. The results from Fullsearch method for frame residue.

The difference of the results for motion vectors is as follows.

<table>
<thead>
<tr>
<th>Frame size</th>
<th>Foreman</th>
<th>Stegan</th>
<th>News</th>
<th>Md</th>
<th>Mc</th>
</tr>
</thead>
<tbody>
<tr>
<td>364 × 272</td>
<td>364.06</td>
<td>391.87</td>
<td>382.57</td>
<td>312.72</td>
<td>320.47</td>
</tr>
<tr>
<td>256 × 208</td>
<td>279.85</td>
<td>308.27</td>
<td>308.27</td>
<td>238.57</td>
<td>246.32</td>
</tr>
<tr>
<td>176 × 144</td>
<td>176.57</td>
<td>189.13</td>
<td>189.13</td>
<td>119.43</td>
<td>127.19</td>
</tr>
<tr>
<td>128 × 96</td>
<td>128.39</td>
<td>141.95</td>
<td>141.95</td>
<td>72.25</td>
<td>80.01</td>
</tr>
</tbody>
</table>

Table 3. The difference of the estimated results compared to Fullsearch method for motion vectors.

From Table 3, we find all values of absolute difference is less than 300 bits for motion vector in a frame. So the overall difference can be acceptable. Thus, we apply (6) to estimate bits for coding the motion vectors.

2.2. RELATION BETWEEN BITRATE FOR FRAME RESIDUE AND FRAME SIZE

With the assumption that DCT coefficients of frame residue are approximately uncorrelated and are Laplacian distributed [7], papers [3, 4] present an expression of frame residue rate, or bits/pixel for frame residue:

\[
H(Q) = \begin{cases} 
  H_1(Q) = \frac{1}{2} \log_2 \left( 2e^{\frac{\sigma^2}{Q^2}} \right) & \text{if } \frac{\sigma^2}{Q^2} > \frac{1}{2e} \\
  H_2(Q) = \frac{1}{2} \log_2 \left( 2e^{\frac{\sigma^2}{Q^2}} \right) & \text{if } \frac{\sigma^2}{Q^2} \leq \frac{1}{2e}
\end{cases}
\]  

(9)

So the total number of bits for frame residue \( R_f \) within a frame can be expressed as the total number of pixels times the frame residue rate. We select \( H_2(Q) \) for low bitrate case [5]. Thus, we obtain an expression for \( R_f \):

\[
R_f = S \cdot A \cdot H_2(Q) = S \cdot A \cdot e^{\frac{-\sigma^2}{2Q^2}}
\]  

(10)

where parameter \( \sigma^2 \) is the average variance for a frame in the transcoded video. Parameter \( S \) is the total number of macroblocks within a frame in the transcoded video. Parameter \( A \) is the total number of pixels in a macroblock (i.e., \( A=16^2 \)). Thus, \( S \cdot A \) stands for the total number of pixels in one frame.

However, Equation (9) is obtained before Zig-Zag scan and run length coding. This will make the estimation for frame residue rate higher when (10) is used to estimate the total number of bits for frame residue in a frame. So we revise Equation (10), by keeping \( \frac{\sigma^2}{Q^2} \) in (9) as a relative part for frame residue rate, to:

\[
R_f = S \cdot A \cdot \frac{\sigma^2}{Q^2}
\]  

(11)

From (11), we obtain the total number of bits for frame residue in a frame in the transcoded video:

\[
R_{fr} = S_{fr} \cdot \frac{\sigma^2_{fr}}{Q_{fr}^2}
\]  

(12)

where, parameter \( S_{fr} \) is the total number of macroblocks within a frame in the precoded video. Parameter \( \sigma^2_{fr} \) is the average variance within a frame in the precoded video.

Combining (11) and (12), we obtain an expression for the total number of bits allocated for frame residue in a frame:

\[
R_f = R_{fr} \cdot \frac{S_{fr}}{S_{fr}}
\]  

(13)

Assuming the the size of the precoded video is CIF, we compare the results estimated from Equation (8) with the results obtained from Fullsearch method. Results are shown as follows.

<table>
<thead>
<tr>
<th>Frame size</th>
<th>Foreman</th>
<th>Stegan</th>
<th>News</th>
<th>Md</th>
<th>Mc</th>
</tr>
</thead>
<tbody>
<tr>
<td>364 × 272</td>
<td>364.06</td>
<td>391.87</td>
<td>382.57</td>
<td>312.72</td>
<td>320.47</td>
</tr>
<tr>
<td>256 × 208</td>
<td>279.85</td>
<td>308.27</td>
<td>308.27</td>
<td>238.57</td>
<td>246.32</td>
</tr>
<tr>
<td>176 × 144</td>
<td>176.57</td>
<td>189.13</td>
<td>189.13</td>
<td>119.43</td>
<td>127.19</td>
</tr>
<tr>
<td>128 × 96</td>
<td>128.39</td>
<td>141.95</td>
<td>141.95</td>
<td>72.25</td>
<td>80.01</td>
</tr>
</tbody>
</table>

Table 4. The results from Fullsearch method for frame residue.

The difference of the results for frame residue is as follows.

<table>
<thead>
<tr>
<th>Frame size</th>
<th>Foreman</th>
<th>Stegan</th>
<th>News</th>
<th>Md</th>
<th>Mc</th>
</tr>
</thead>
<tbody>
<tr>
<td>364 × 272</td>
<td>364.06</td>
<td>391.87</td>
<td>382.57</td>
<td>312.72</td>
<td>320.47</td>
</tr>
<tr>
<td>256 × 208</td>
<td>279.85</td>
<td>308.27</td>
<td>308.27</td>
<td>238.57</td>
<td>246.32</td>
</tr>
<tr>
<td>176 × 144</td>
<td>176.57</td>
<td>189.13</td>
<td>189.13</td>
<td>119.43</td>
<td>127.19</td>
</tr>
<tr>
<td>128 × 96</td>
<td>128.39</td>
<td>141.95</td>
<td>141.95</td>
<td>72.25</td>
<td>80.01</td>
</tr>
</tbody>
</table>

Table 5. The difference of the estimated results compared to Fullsearch method for frame residue.

From 5, the absolute difference for most value (90%) is less than 300 bits for frame residue in a frame. The overall difference can be acceptable. Thus, we apply Equation (13) to estimate bits coded for frame residue.
2.3. RELATION BETWEEN BITRATE FOR OVERHEAD AND FRAME SIZE

Since most bits for overhead is allocated for macroblock and block layer [6], the total number of macroblocks in a frame will affect the overhead mostly. So we use the function of frame size (S) to build the model for the relation for overhead and frame size. To further simplify this model, we assume the average number of bits for overhead per macroblock (R_h) will not change much after downsizing. Thus we set an expression of overhead in a frame:

\[ R_H = S \cdot R_h \]  \hfill (14)

From (14), we obtain the total number of bits for overhead within a frame in the precoded video:

\[ R_{H_{pre}} = S_{pre} \cdot R_h \]  \hfill (15)

where, parameter \( S_{pre} \) is the total number of macroblocks within a frame in the precoded video. Parameter \( R_{H_{pre}} \) is the average number of bits for overhead within a frame in the precoded video.

Combining (14) and (15), we obtain an expression for the total number of bits allocated for overhead in a frame:

\[ R_H = S \cdot \frac{R_{H_{pre}}}{S_{pre}} \]  \hfill (16)

2.4. OVERALL MODEL FOR THE RELATION BETWEEN BITRATE AND FRAME SIZE

Combining (8), (13) and (16), we obtain the overall bitrate (R) model for the relation between video bitrate and frame size:

\[ R = \frac{S}{S_{pre}} (R_{MV} + \frac{R_{H_{pre}}}{S_{pre}} + R_{H}) \]  \hfill (17)

3. CONCLUDING REMARKS

Considering that almost no bitrate will be allocated to the video whose size is close to zero, simple linear model by connecting precoded bitrate and zero bitrate is used to make a comparison with our proposed method (17). Here, the simple linear model is assumed to be the model of relation between bitrate and frame sizes before our investigation in this paper. The bitrate (\( R_l \)) in the linear model can be expressed as follow:

\[ R_l = \frac{S}{S_{pre}} R_{pre} \]  \hfill (18)

Assuming the size of the precoded video is CIF, we compare the results estimated from Equation (17) with the results obtained from the simple linear model. Results are shown in Table 6, 7 and Figure 1. From Table 7, we find the overall absolute difference from simple linear model is smaller than the overall absolute difference from our proposed method (18) for a frame. In other word, using our proposed method will get a more accurate model to build the relation between bitrate and frame sizes.

From all results above, we develop a model which can estimate the relationship between bitrate and video frame sizes. Using this model, we can choose a bitrate according to the frame size required by users or we can select a suitable frame size for a given bitrate, while maintaining a good video quality.

4. REFERENCES


