REAL-TIME H.263+ FRAME RATE CONTROL FOR LOW BIT RATE VBR VIDEO

Hwangjun Song, Jongwon Kim and C.-C. Jay Kuo

Integrated Media Systems Center and Department of Electrical Engineering-Systems
University of Southern California, Los Angeles, California 90089-2564
E-mail: {hwangjun,jongwon,ckkuo}@sipi.usc.edu

ABSTRACT

H.263+ is an emerging video compression standard for the low bit rate communication. Most H.263+ rate control algorithms, e.g. the one used in the test model of the near-term (TMN), focus on macroblock level bit allocation and low latency by assuming a constant frame rate video encoding sent through a constant bit rate (CBR) channel. These algorithms cannot handle the fluctuation of the channel bandwidth well so that the transmitted video quality is degraded seriously. In this work, we propose a new H.263+ rate control scheme which supports the variable bit rate (VBR) channel through the frame rate adjustment. In particular, a fast realization of encoding frame rate control based on motion information within a sliding window is developed to efficiently determine the tradeoff between spatial and temporal qualities. Experimental results are provided to demonstrate the superior performance of the proposed scheme.

1. INTRODUCTION

Visual communications have gained more and more attention due to the fast development of digital communication techniques and various network systems recently. Since raw video contains a large amount of data, an efficient compression technique is essential. Among various video coding standards, ITU-T (International Telecommunication Union) has presented a sequence of H.26x standards for low bit rate visual communications, among which H.263+ is a new member. Enhanced versions of H.263 and H.263+ such as H.264/++ and H.26L are now under development.

It is well known that VBR (variable bit rate) video supports better quality than CBR (constant bit rate) video [1]. Recently, wired/wireless packet switching networks and their associated protocols with various features are widely implemented. These systems can support VBR channels to increase the network efficiency and enhance VBR data service quality. Real-time H.263+ rate control for low bit rate VBR channel is studied in this work.

Most H.263+ rate control algorithms, e.g. the one used in the test model of the near-term (TMN), focus on macroblock level bit allocation and low latency by assuming a constant frame rate video encoding sent through a CBR channel. These algorithms cannot handle the fluctuation of the channel bandwidth well, and the transmitted video quality is degraded seriously. Here, we propose a new H.263+ rate control scheme which supports the VBR channel through the frame rate adjustment. In particular, a fast realization of encoding frame rate control based on motion information within a sliding window is developed to efficiently determine the tradeoff between spatial and temporal qualities.

2. RATE AND DISTORTION MODELS FOR FRAME LAYER

We consider a frame layer R-D model to construct both the rate and distortion models with respect to the averaged quantization parameter (QP) of all macroblocks in each frame. To be more specific, the quadratic rate model and the affine distortion model are employed, and rate control results of the macroblock layer are used to determine the coefficients of the frame layer R-D model. Thus, the additional computational complexity required for the frame layer R-D modeling is very small and negligible.

In terms of mathematics, the rate and distortion models can be written, respectively, as:

\[
\hat{R}(\tilde{q}) = (a\tilde{q}^{-1} + b\tilde{q}^{-2})MAD(f_{\text{ref}}, f_{\text{cur}}),
\]

\[
\hat{D}(\tilde{q}) = a'\tilde{q} + b',
\]

where \(a, b, a', b'\) are model coefficients, \(f_{\text{ref}}\) is the reconstructed reference frame at the previous time instance, \(f_{\text{cur}}\) is the uncompressed image at the current time instance, \(\tilde{q}\) is average QP of all macroblocks in a frame, \(\hat{R}(\tilde{q})\) and \(\hat{D}(\tilde{q})\) are the rate and distortion models of a frame, respectively, and \(MAD(f_{\text{ref}}, f_{\text{cur}})\) is the mean of absolute difference between \(f_{\text{ref}}\) and \(f_{\text{cur}}\). Note that \(MAD(f_{\text{ref}}, f_{\text{cur}})\) takes into account the dependency among frames. Coefficients \(a, b, a', b'\) are determined by using the linear regression method. Conventionally, the R-D curve is computed based on integer QP. In our case, \(\tilde{q}\) can be a floating point number since \(\tilde{q}\) is the average QP of all macroblocks in a frame. We use an outlier removal process to improve the model accuracy. The R-D modeling method provides a very good approximation for all test sequences in our experiment.

3. FRAME RATE CONTROL FOR UNCONSTRAINED VBR VIDEO

Based on the rate and distortion models described in the previous section, we describe how to choose the encoding frame rate to minimize degradation in motion smoothness, and how to control the bit rates for each frame and adjust the Lagrange multiplier \(\lambda\). It is worthwhile to emphasize that the proposed rate control algorithm includes existing H.263+ macroblock layer rate control algorithms as low level component. Thus, our work is compatible with existing rate control schemes [2].
3.1. Motion-based Frame Rate Control with A Sliding Window

One objective of our rate control scheme is to keep the quality of P-frames nearly constant (or varying very slowly). Since each P frame is used as a reference frame for the following P frames, quality degradation propagates to later frames when a P frame is degraded severely. It is observed that human eyes are sensitive to the abrupt encoding frame rate (or interval) change. Our scheme aims at the reduction of temporal degradation in terms of motion jerkiness perceived by human beings. At the same time, no encoding time delay is imposed for real-time processing. The next encoding frame position is estimated by motion information within a sliding window of fixed length to avoid an abrupt frame rate change. By adjusting the frame rate, we can avoid or reduce the sudden frame skipping in existing rate control algorithms, which degrades motion smoothness disastrously. Two problems have to be addressed for frame rate control. They are: when the frame rate should be changed, and how to change the encoding frame rate to preserve motion smoothness. They are considered as follows. The overview of the framework is shown in Fig. 1, where sliding windows of 12 frames width are used as a control unit for frame rate selection.

We need some measure to detect motion change in video. Here, the histogram of difference (HOD) is adopted since HOD is very sensitive to local motion in video. After the HOD values of consecutive frames in a sliding window are calculated, the estimated HOD value $\hat{D}_h$ for the next frame can be calculated by

$$\hat{D}_h = D_h + \omega_h s_h,$$

(3)

where $\omega_h$ is the HOD value between the two last encoded frames in the sliding window, $\omega_h$ is a weighting factor and $s_h$ is the slope of approximating line which minimizes the mean square error of HOD in the sliding window. It is interesting to point out that $s_h$ is related to motion change in video. The positive value means that the motion becomes faster while the negative value means that the motion becomes slower. Also, a larger value of $|s_h|$ implies a larger motion change.

Based on the motion change information in the sliding window, we can determine the rule for the change of the encoding frame interval. Let $\Delta D_h$ denote the mean of all HOD values of frames in the sliding window. We can adjust the encoding frame rate based on the difference $\delta D_h = \hat{D}_h - D_h$ as follows.

- If $\delta D_h \leq -TH_h$, the encoding frame interval is decreased by $\Delta F^{\text{int}}(F^{\text{cur}}_h)$.
- If $|\delta D_h| < TH_h$, the encoding frame interval remains the same.

In above, the threshold value $TH_h$ is chosen based on the averaged $HOD$ over the the sliding window, and $F^{\text{int}}_h$ is the current encoding frame interval. In addition, it is observed that for slow and steady motion change.

- If $B_i^{\text{res}} > TH_B1$, the encoding frame interval is decreased by $\Delta F^{\text{int}}(F^{\text{cur}}_h)$.
- If $B_i^{\text{res}} < TH_B0$ and the current encoding frame interval is greater than the frame capturing interval, the encoding frame interval is decreased by $\Delta F^{\text{int}}(F^{\text{cur}}_h)$.

where $TH_B0$ and $TH_B1$ are threshold values.

Parameters $TH_h$, $TH_B0$ and $TH_B1$ work as thresholding values for controlling the tradeoff between temporal and spatial quality. When the frame rate is changed, the frame rate is kept constant for a period of time as long as that of the sliding window to avoid the frequent occurrence of frame rate change. The first rule is for short-term motion change while the second rule is for the long-term motion. Furthermore, the following empirical rule to choose $\Delta F^{\text{int}}(F^{\text{cur}}_h)$ is adopted in our experiment:

$$\Delta F^{\text{int}}(F^{\text{cur}}_h) = \lceil 0.3 \cdot F^{\text{cur}}_h \rceil,$$

where $\lceil z \rceil$ means the smallest integer greater than $z$.

3.2. Frame Layer Rate Control with Adaptive Lagrange Multiplier

For every chosen frame, we have to allocate the bit rate (i.e. decide the spatial quality of each frame) so as to optimize the distribution of bit budgets over encoded frames. In our approach, a pseudo-GOP unit (size of up to 120 captured frames and moving along with the above sliding window) is introduced for H.263+ to provide a long-term guideline for bandwidth consumption based on the CBR channel. Then, we consider a new low complexity formulation of the rate allocation problem based on R-D models as stated below.

Determine $Q_i$, $i = 1, 2, \ldots, N_{\text{pop}}$, to minimize

$$\sum_{i=1}^{N_{\text{pop}}} \{\tilde{D}_i(\tilde{q}) + \omega \|\tilde{D}_i(\tilde{q}) - D_{i-1}\|\},$$

(4)

subject to

$$\sum_{i=1}^{N_{\text{pop}}} R_i \leq B_{\text{pop}},$$

(5)

where $\tilde{D}_i$ is the estimated distortion of the current frame, $D_{i-1}$ is the actual distortion of the previous frame, and $\omega$ is the weighting factor for quality change between adjacent frames. The second term of (4) is included to reduce the flickering artifact caused by the abrupt quality change between adjacent frames.

By using the Lagrangian method, we can define a penalty function for the ith frame by combining the cost function and the constraint through a Lagrange multiplier. As mentioned, to satisfy
the bit budget constraint smoothly, we use CBR video transmission as a reference.

\[ P_i(\tilde{q}_i) = \tilde{D}_i(\tilde{q}_i) + \omega ||\tilde{D}_i(\tilde{q}_i) - D_{i-1}|| + \lambda_i \cdot \max \{ \tilde{B}_i^{res}, 0 \}, \]

\[ \tilde{B}_i^{res}(\tilde{q}_i) = \sum_{j=1}^{i-1} R_j + \tilde{R}_i(\tilde{q}_i) - t_i \cdot C_{cbr}, \]  

where \( P_i(\tilde{q}_i) \) is the cost function for the \( i \)th frame, \( \lambda_i \) is the Lagrange multiplier for the \( i \)th frame, \( R_j \) is the used bit rate for the \( j \)th frame, \( t_i \) is the time instance of the \( i \)th encoded frame after the start of GOP and the average channel bandwidth

\[ C_{cbr} = \frac{B_{gop}}{T_{gop}}, \]

where \( T_{gop} \) is the time duration of a GOP.

Based on the rate and distortion models in Section 2, we can determine the optimal QP to minimize the above penalty function. It was shown in [3] that \( P_i(\tilde{q}_i) \) is in general a convex function. Thus, we can get its optimal solution by using the gradient method, i.e.

\[ \tilde{q}_i^* = \arg \min_{\tilde{q}_i} P_i(\tilde{q}_i). \]  

Notice that what we finally need is not \( \tilde{q}_i^* \), but \( \tilde{R}_i(\tilde{q}_i^*) \) which is the target bit budget for the \( i \)th frame. Also, if we consider one special case of the above optimization problem such as \( \tilde{B}_i^{res} \leq 0 \) and \( \tilde{R}_i(\tilde{q}_i^*) \leq (t_{i+1} - t_i)C_{cbr} \) case, then, it is not difficult to see that \( \tilde{R}_i(\tilde{q}_i^*) = (t_{i+1} - t_i)C_{cbr} \). In this case, we go back to the conventional macroblock layer rate control algorithms of a lower level, which try to allocate the bit budget to each macroblock with the solution \( \tilde{R}_i(q_i^*) \). However, if \( \tilde{B}_i^{res} > 0 \), we are able to provide a better solution than that given by existing macroblock layer rate control schemes.

The Lagrange multiplier method has been widely employed for bit rate allocation in video coding. The Lagrange multiplier \( \lambda \) that satisfies the given bit budget can guarantee the global optimality of the solution for both independent and dependent coding schemes. In [4], Choi and Park attempted to adjust the Lagrange multiplier based on the buffer occupancy, and derived a discrete linear equation for buffer occupancy. Then, they proved the stability of the solution based on Lyapunov theory. Lin and Chen [5] tried to control the Lagrange multiplier to avoid buffer underflow and overflow for the ATM network. Wiegand et al. [6] proposed a frame-to-frame update of \( \lambda \) by using the least mean square adaptation in the selection of an efficient macroblock coding mode.

The optimization of H.263+ coded video constrained by the bit budget formulated in (6) is solved as follows. Since the Lagrange multiplier will be adjusted adaptively on the fly in real-time rate control, only a sub-optimal solution with a low computational complexity will be considered. Thus, we use the following adaptive adjustment rule for \( \lambda \).

\[ \lambda_{i+1} = \lambda_i + \Delta \lambda_i, \]

\[ \Delta \lambda_i = \frac{B_i^{res}}{B_{target,i}} - 1, \]

where \( \lambda_k \) is the Lagrange multiplier for the \( k \)th frame, and

\[ B_i^{res} = \sum_{j=1}^{i} R_j, \]

\[ B_{target,i} = C_{cbr} \cdot t_i. \]

4. FRAME RATE CONTROL FOR TIME VARYING CBR CHANNEL

Time varying CBR channels include the feedback VBR and the renegotiated CBR channel [1], where the available bandwidth is time varying and can be modeled well with a piecewise constant function. This model will be valid if the bandwidth change speed is relatively low in comparison with the duration when the bandwidth is kept at a constant level. This model will be even more reasonable if there are buffers at the encoder and the decoder ends to offset the channel bandwidth fluctuation effect.

Based on the rate and distortion models in Section 2, we can estimate the distortion of current frame. Since \( \tilde{q}_i > 0 \), the estimated distortion \( \tilde{D} \) can be expressed as

\[ \tilde{D} = a'(\frac{\alpha \text{MAD}(f_{ref}, f_{cur})}{2B(F_{cur}^{\text{int}})} + \sqrt{(\alpha \text{MAD}(f_{ref}, f_{cur}))^2 + 4bB(F_{cur}^{\text{int}})\text{MAD}(f_{ref}, f_{cur})})}{2B(F_{cur}^{\text{int}})} + b', \]

\[ B(F_{cur}^{\text{int}}) = C_{cur}C_{cbr}, \]

where \( C_{cbr} \) is the current channel bandwidth of the time varying CBR channel and \( F_{cur}^{\text{int}} \) is the current encoding frame interval under the assumption that the camera captures frames at a rate of 30 fps. Note that \( \tilde{D} \) increases when fast motion change occurs (with an increasing \( \text{MAD} \)) or when the channel bandwidth decreases (with a decreasing \( B(F_{cur}^{\text{int}}) \)) suddenly.

Now, let us consider the rate control scheme. If the spatial quality is below a tolerable level due to fast motion change or sudden channel bandwidth decrease, we should reduce the temporal quality and improve the spatial quality in order to reduce the flickering artifact. At the same time, it is still desirable to control the temporal quality degradation. On the contrary, if the spatial quality is above a certain level, we should increase the temporal quality. Based on the discussion, the encoding frame rate control algorithm can be stated as:

- If \( \tilde{D} > TH_{D1} \), increase the encoding frame interval by \( \Delta F^{\text{int}}(F_{cur}^{\text{int}}) \).
- If \( \tilde{D} < TH_{D2} \) and the current frame interval is greater than the frame capturing interval, the encoding frame interval is decreased by \( \Delta F^{\text{int}}(F_{cur}^{\text{int}}) \).

In above, \( TH_{D1} \) and \( TH_{D2} \) are two threshold values to be selected. By adopting this rate control scheme, we can avoid the abrupt change of the encoding frame rate and improve the spatial quality. This algorithm can be applied in real-time processing since the computational complexity is very low so that small latency can be guaranteed.

5. EXPERIMENTAL RESULTS

In our experiment, the macroblock layer rate control algorithm of TMN8 [2] is employed, and the implementation is based on the UBC H.263+ source code [7]. Experimental results are presented to demonstrate the performance of our rate control scheme for the unconstrained VBR channel in comparison with TMN8. The three test sequences are the QCIF “Salesman”, “Akiyo” and
Table 1: Performance comparison with TMN8 under unconstrained VBR, where the target average rate is 24kbps and the number in the parenthesis denotes the number of encoded frames.

<table>
<thead>
<tr>
<th>Method</th>
<th>Sequence</th>
<th>Avg</th>
<th>STD</th>
<th>Enc. frms</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMN8 (2-frame skip)</td>
<td>Salesman (28)</td>
<td>31.0904</td>
<td>0.9064</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Akiyo (29)</td>
<td>35.5266</td>
<td>0.9188</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Silent Voice (29)</td>
<td>30.7279</td>
<td>0.3822</td>
<td></td>
</tr>
<tr>
<td>TMN8 (1-frame skip)</td>
<td>Salesman (42)</td>
<td>30.6333</td>
<td>0.8295</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Akiyo (43)</td>
<td>34.8258</td>
<td>0.7010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Silent Voice (42)</td>
<td>30.2198</td>
<td>0.4013</td>
<td></td>
</tr>
<tr>
<td>Proposed method</td>
<td>Salesman (32)</td>
<td>30.9296</td>
<td>0.8976</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Akiyo (37)</td>
<td>35.1607</td>
<td>0.8516</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Silent Voice (31)</td>
<td>30.8802</td>
<td>0.4026</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Performance comparison under time varying CBR for the Silent Voice sequence.

<table>
<thead>
<tr>
<th>Method</th>
<th>Avg</th>
<th>STD</th>
<th>Enc. frms</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMN8 (CBR)</td>
<td>30.9601</td>
<td>0.4784</td>
<td>62</td>
</tr>
<tr>
<td>TMN8 (time varying CBR)</td>
<td>30.8508</td>
<td>0.4918</td>
<td>62</td>
</tr>
<tr>
<td>Proposed method</td>
<td>31.0817</td>
<td>0.9088</td>
<td>56</td>
</tr>
</tbody>
</table>

“Silent Voice”, and the target average bit rate is 24 kbps. In this experiment, we perform the rate control on a sequence consisting of 100 P frames corresponding to about 3.3 seconds (with a frame rate of 30 frames per second).

The performance comparison of three rate control algorithms for the three test sequences is shown in Table 1, where the average PSNR value and the standard deviation of PSNR are computed based on coded frames only. One can see from this table that even though TMN8 with 2-frame skipping has the best average PSNR, but its PSNR values have the largest standard deviation among the three rate control schemes. The other hand, TMN8 with 1-frame skipping has the smallest standard deviation, but the largest average PSNR value. Our algorithm reduces the fluctuation of PSNR of TMN8 with 2-frame skipping and improves the average PSNR of TMN8 with 1-frame skipping. Under time varying CBR, the proposed rate control algorithm can control the encoding frame rate without the obvious motion unsmoothness. We can see the statistical data in Table 2.

6. CONCLUSION

A rate control algorithm for low bit rate unconstrained VBR was proposed in this research. In this rate control scheme, we treat the encoding frame interval (or rate) as a control variable in order to pursue a good trade-off between spatial and temporal qualities. For the low bit rate unconstrained VBR channel, the proposed algorithm can improve human visual perceptual quality by providing a better trade-off between spatial and temporal qualities. Under time varying CBR, the proposed algorithm can adjust the encoding frame rate without obvious motion unsmoothness. This algorithm can be employed for real-time video applications due to its negligible encoding time delay and computational overhead.

7. REFERENCES


