Frame Layer Rate Control for Dual Frame Motion Compensation

1Da Liu, 1Debin Zhao, 2Siwe Ma and 2Wen Gao, Fellow, IEEE
1Harbin Institute of Technology, Harbin, China
2Peking University

Abstract — Rate control plays an important role in video coding. When jump update dual frame motion compensation (JU-DFMC) is utilized in video coding, two kinds of frames are existed according to bit allocation, one is low quality frame (LQF), the other is high quality frame (HQF). For each current frame (LQF or HQF), one short term reference frame (STR) and one long term reference frame (LTR) are utilized for motion compensation. Owning to this kind of coding structure, rate control for JU-DFMC is different from traditional methods. In this paper, a rate control scheme for JU-DFMC is proposed. Firstly, a linear bit allocation method is proposed for LOFs to obtain smooth quality. Secondly, different source rate and quantization stepsize (Qstep) models are proposed to different frames (LQFs or HQFs). Thirdly, overhead prediction and MAD prediction are presented. Experimental results show that the proposed method can bring accurate rate control for JU-DFMC, at the same time the PSNR is better than the other methods.

Index Terms — Rate control, video coding, dual frame motion compensation, bit allocation, R-Q model.

1. INTRODUCTION

![Fig. 1. Dual frame motion compensation.](image)

With the development of wireless access and video coding techniques, the utilization of wireless video transmission, such as multimedia in 3G, video conference, video supervision, increases rapidly. These wireless applications require significant bandwidth. However, much existing wireless spectrum is under-used, and the available radio spectrum is limited [1].

Cognitive radio [2][3] is a kind of method to solve the problem. Base on a fixed low bandwidth, cognitive radio can rent extra bandwidth by shifting to a different spectrum in short periods of time, such as spectrum owning to fire department, which is not under-used [1] all the time. The other kind of method is in wireless multi-channel video communication, such as a CDMA High Data Rate (HDR) system [4] [5]. For the portion of the bandwidth which the user does not utilize at the present time, the channel scheduler divides it among the other users in turn.

Both methods (cognitive radio or multi-channel system) can be seen as a constant low bandwidth channel that is occasionally supplemented by a substantially larger bandwidth for a short interval of time. In video communication, the larger bandwidth can be utilized to code a frame with higher quality, which is utilized as reference frames for following several frames to improve the overall quality. Besides this reference frame, each previously decoded frame is also utilized as reference frame for each current frame. This is a special case of multi-reference frame, called jump update dual frame motion compensation (JU-DFMC).

In JU-DFMC, two kinds of frames are existed according bit allocation. As shown in Fig. 1, the first kind of frames has relatively lower quality and is called low quality frame (LQF), such as the $i+k$th frame $f_{i+k}^{Q}$ ($k=-N,-N+1,-N+2,-N+3,\ldots,-2,0,1$). And the second one has relatively higher quality and is called high quality frame (HQF), such as the $i$-th frame $f_{i}^{Q}$ and the $i-N$-th frame $f_{i-N}^{Q}$. For each current frame (LQF or HQF), two reference frame are utilized for motion prediction. The first reference frame is the most recently decoded frame, called short-term reference frame (STR). And the second reference frame is a HQF from the past, which is named as long-term reference frame (LTR). The LTR (HQF) is periodically updated. It remains static for $N$ frames, and jumps forward to be a frame which has two frame distance back from the current frame. For example, as shown in Fig. 1, for frame from time instant $i-N+1$ to $i$, LTR is HQF $f_{i-N}^{Q}$. When the encoder moves on to coding frame $i+1$, the STR will slide forward by one to frame $i$, and the LTR will jump forward by $N$, namely from frame $i-N-1$ to frame $i-1$. After that, the LTR remains fixed for $N$ frames, and then jumps forward again. $N$ is the jump update parameter.

In recent years, several dual frame motion compensation
(DFMC) related research have been done. In [6] and [7], it was shown that overall PSNR was influenced by the different extra bandwidth and the period giving to the LTR. In [8], the update period of the LTR was set to ten frames. The PSNR of nine frames that follow the LTR frame was utilized to determine how many bits can be allocated to the LTR. In [9], simulated annealing was utilized to select the LTR; however, its computational complexity is very high. In the works [10]-[13], DFMC based methods were proposed for video transmission over noisy channels. In [14], the end-to-end transmission delay of DFMC was studied.

The video transmission of JU-DFMC needs rate control to achieve consistent good quality. A number of model-based rate control algorithms have been reported. In [15], target bits were distributed proportional to the size of objects. In [16], linear correlation between rate and the percentage of zeros among the quantized transform coefficients was adopted for rate control. In [17], a constraint for the least-mean-square estimation of the model parameters was presented for rate control. In [18]-[20], single-pass and two-pass Variable Bit-Rate (VBR) rate control algorithms were proposed for digital storage application. In [21], a rate control algorithm based on Evolution Strategy (ES) was proposed for off-line H.264/AVC video encoding. To get better subjective visual quality, a practical two-pass VBR rate control algorithm was proposed in [22], and the relationship between quantization step and the given target distortion was considered to reduce quality fluctuation in [23]. For rate control in H.264, a linear total rate and \( \frac{1}{Q_{step}} \) model was developed in [24], and a texture rate and \( Q_{step} \) quadratic model was employed in [25]. In the further study [26], a sum bits and quantization parameter (QP) quadratic model was proposed. To obtain better quality, lagrange multiplier was computed for mode decision in [27], and a model which relates source bits to complexity of coded blocks was proposed for rate control in [28].

However, rate control for JU-DFMC has not been fully studied yet. In the paper, a frame layer rate control scheme for JU-DFMC is proposed. In the proposed scheme, extra bandwidth allocation method, i.e., bits allocation, Rate-Qstep models, MAD and overhead prediction are separately presented.

The rest of the paper is organized as follows. The proposed bit allocation, Rate-Qstep model, MAD and overhead prediction are separately given in Section 2, Section 3 and Section 4. The experimental results and discussions are provided in Section 5. Finally, Section 6 concludes this paper.

2. BIT ALLOCATION

2.1 Bit allocation relationship in LQFs

With the coding of each current LQF following HQF, the temporal distance from HQF (which is utilized as LTR) increases, the prediction performance from HQF decreases. For obtaining the same quality in LQFs, the bit allocation in LQFs increases with the enhancement of temporal distance between LQFs and HQF. An example of the bit allocation and temporal distance is shown in Fig. 2. line \( a \) and line \( b \) are the bit allocation in LQFs when LTR and STR are all utilized as reference frame. The difference is that for line \( a \), PSNR in HQF and LQF are separately about 35.8dB and 32.2dB; for line \( b \), PSNR in HQF and LQF are separately about 34.3dB and 32.32dB. As can be seen, in any case, the bit allocation in LQFs is nearly linear with the temporal distance between the LQF and the previous HQF,

\[
R_i = X \times D_i + Y ,
\]

where \( R_i \) is the bit allocation in LQF \( f_i^{LQ} \), \( D_i \) is temporal distance between LQF \( f_i^{LQ} \) and previous LTR, \( X \) and \( Y \) are linear parameters.

2.2 GOP level bit allocation

With the coding of each current LQF following previous HQF (LTR), the prediction performance from HQF decreases, the prediction error variance \( \sigma^2 \) from HQF increases. When the prediction error variance \( \sigma^2 \) in the current LQF from previous HQF is [29]

\[
\sigma^2_j \geq \frac{1}{\gamma} \sigma^2_0 + \sigma^2_j ,
\]

the LQFs coding is terminated, the next frame is encoded as HQF (will be utilized as LTR), where \( \sigma^2_0 \) is the prediction error variance in the first LQF after HQF when the reference frame is HQF, \( \sigma^2_j \) is the average prediction error variance in encoded LQFs when the reference frame is LQF.

In JU-DFMC of the work, one HQF and following LQFs comprise a GOP (group of pictures). Before encoding the
current GOP, the GOP length is initialized the same as that in previous GOP, \( N \). And suppose the overall remaining frames waiting to be encoded is \( M \), while the remaining bits are \( R_s \), then the target bit allocation \( T(j) \) in the GOP (suppose the \( j \)th GOP) is calculated as
\[
T(j) = \frac{N}{M} \times R_s .
\] (3)

In calculating bit allocation in the first GOP, GOP length is initialized as 10. However, the actual GOP length is also determined by (2).

2.3 Frame level bit allocation

Before encoding the current GOP, the bit allocation ratio between HQF bits and average LQF bits is obtained from the updated value in previous GOP. After coding previous GOP, the bit allocation in HQF and LQFs is fixed. Under the same overall bit-rate of the previous GOP, if the bit allocation between HQF and LQFs changes, the changed mean squared error (MSE) corresponding to the changed bit-rate in each frame can be approximately calculated from the derivation of function
\[
R(D) = \frac{1}{2} \log_2(\frac{\sigma^2}{D}) .
\] (4)

So we have
\[
\Delta D = -2 \ln 2 \times D \times \Delta R ,
\] (5)
where \( \Delta D \) is the change of MSE, \( \Delta R \) is the change of bit-rate, and \( D \) is the MSE in the frame.

After adding the fixed MSE with the changed MSE, and adding the changed bit-rate with the changed bit-rate, the different MSEs in HQF and LQFs corresponding to different bit allocations can be calculated.

The MSE in reference picture is linear with the prediction error variance in next frame. With the change of bit allocation between HQF and LQFs under the same overall bits of the previous GOP, the MSE in HQF and LQFs will change, and then prediction error variance will change as well. When the sum of prediction error variance from HQF and LQF, i.e.,
\[
\sigma_{\text{err}}^2 = \sigma_e^2 + \sigma_R^2 ,
\] (6)
is the smallest, the bit allocation brings the best coding efficiency, then the updated bit allocation ratio \( R_a \) between HQF bits and average LQF bits in previous GOP can be obtained. Since the previous GOP is close to the current GOP, the bit allocation ratio \( R_a \) is utilized for the bit allocation in the current GOP [29].

The target bit allocation \( T_i \) in HQF (suppose at time instant \( i \)) of the GOP (suppose the \( j \)th GOP) is calculated as
\[
T_i = T(j) \times \frac{R_a}{R_a + 1 \times (N - 1)} .
\] (7)

The target bit allocation \( T_i \) for LQF (suppose at time instant \( i \)) of the GOP is calculated as
\[
T_i = \frac{i \times X + Y}{(i + (i + 1) + \ldots + N) \times X + (N - i - 1) \times Y} \times (T(j) - \sum_{i} R_i) ,
\] (8)
where \( R_i \) is the actual bits allocation in frame at time instant \( t = i, i + 1, i + 2, \ldots, i - 1 \), \( X \) and \( Y \) are linear parameter from (1) and are updated with linear regression. For the first GOP, \( R_a \) is initialized as 4.

In video coding, if the frame \( f_i^{LQ} \) is located out of the scope of the initial GOP length, the target bit allocation in LQF \( f_i^{LQ} \) is calculated using the bit allocation relationship in adjacent LQFs
\[
T_i = \frac{1}{n} \sum_{r} [R_{i-r} + \Delta R \times (D_i - D_{i-r})] ,
\] (9)
where \( D_i \) and \( D_{i-r} \) are separately the temporal distance between LQF \( f_i^{LQ} \) and \( f_{i-r}^{LQ} \) and previous LTR, \( \Delta R \) is the average increase of bit allocation between two adjacent LQFs and is updated after encoding each LQF, \( n \) is the number of encoded LQFs in the GOP. If the actual GOP length is less than the initial GOP length, bit allocation in LQFs out of actual GOP length will be abandoned.

To conform with the HRD requirement, the target bits are bounded the same as that in [25]. The detailed proof of (2) and (6) is not given here since it is not the scope of this paper.

3. RATE-QSTEP MODEL

Accurate rate and Qstep model have great impact on the rate control performance. In previous work [27], the source rate is modeled as a function of the quantization stepsize and the complexity of the residual signal. In [28], a simplified source rate model is adopted
\[
R_{\text{sr}} = X \times \frac{MAD}{Q} ,
\] (10)
where \( X \) is model parameters and \( Q \) is the quantization stepsize, MAD is the mean absolute difference between original signal value and prediction signal value, \( R_{\text{sr}} \) is the source bits and is obtained from the difference of total bits and header bits.

In this work, (10) is adopted for rate control. For different kinds of LQFs and HQFs, how to update parameter \( X \) is different.
3.1 R-Q model in the first LQF of a GOP

Fig. 3. R-Q model in the first LQF under different bit allocation ratio $R_a$.

For the first LQF in a GOP, the distance between the LQF and previous HQF is only one frame, but the quality in the LQF is lower than that in HQF. And majority of reference blocks in the LQF come from previous HQF. In the case, parameter $X$ in the first LQF is influenced by bit allocation ratio between the LQF and previous HQF. From the bit allocation method proposed in Section 2, most of bit allocation ratio between the first LQF and previous HQF is in the range of $(0.2, 0.3)$, then a statistic is shown in Fig. 3. As can be seen in the figure, under fixed bit allocation ratio $R_a$ between the LQF and previous HQF, the relationship between $R_a$ and $1/\text{step}_Q$ in the LQF can be accurately approximated by a linear model. And with the increase of bit allocation ratio, linear slope (i.e., parameter $X$) increase progressively. Further more, it can further be seen in Fig. 6, parameter $X$ and temporal distance is nearly linear. Owning to the above influence, from the second LQF to the last LQF in a GOP, parameter $X$ in each current LQF is

\begin{equation}
X_i^{\text{pred}} = X_{i-N_c-1} + k_i \times \left( \frac{T}{R_{r-1}} - \frac{R_{r-N_c-1}}{R_{r-1-N_c-1}} \right),
\end{equation}

where $X_i^{\text{pred}}$ is the predicted parameter $X$ in the current LQF (suppose at time instant $i$), $N_c$ ($j=1,2,\ldots$) is the GOP length of the $j$th GOP before the current GOP, $X_{i-N_c-1}$ is parameter $X$ in the first LQF (at time instant $i-N_c-1$) of previous GOP, $T$ and $R_{r-1}$ are separately the target bit allocation in the current LQF and actual bit allocation in HQF of the current GOP, $R_{r-N_c-1}$ and $R_{r-1-N_c-1}$ are separately the actual bit allocation in the first LQF and HQF in the previous GOP, $k_i$ is updated by a linear regression method after coding the first LQF in each GOP.

3.2 R-Q model from the second LQF to the last LQF in a GOP

From the second LQF to the last LQF in a GOP, a relationship between $R$ and $1/\text{step}_Q$ in each LQF is shown in Fig. 5. In the statistics, the MSE in LTR (HQF) is $1/4$ times that in LQF, LTR and STR are all utilized as reference frame. As can be seen, the relation between $R$ and $1/\text{step}_Q$ in each LQF can be accurately approximated by a linear model, and with the increase of temporal distance between LQFs and LTR, the linear slope in LQFs, i.e., parameter $X$ in LQFs, increases progressively. It can further be seen in Fig. 6, parameter $X$ and temporal distance is nearly linear. Owning to the above influence, from the second LQF to the last LQF in a GOP, parameter $X$ in each current LQF is

Fig. 5. R-Q model in different LQFs.

Fig. 6. Relationship between parameter $X$ and temporal distance.
updated with the parameters in previous encoded LQFs

\[ X_i^{pred} = \frac{1}{i-j-1} \sum_{l=1}^{i-j} \{ X_{l-1} + \Delta X \times (D_l - D_{l-1}) \} , \]  

\( i \) is the temporal location of current LQF, \( l \) is the temporal location of previous LTR, \( D_l \) is the temporal distance between LQF \( i \) and previous LTR, \( X_{l-1} \) is the parameter \( X \) in LQF at time instant \( i-l \), \( \Delta X \) is the average increase of \( X \) between two adjacent LQFs and is updated after coding each LQF.

### 3.3 R-Q model in HQF of a GOP

\[ X_i^{pred} = X_{l-1} - \Delta X \times (D_l - D_{l-1}) + k \times (T_{l-1} - R_{l-1}), \]  

where \( X_{l-1} \) is the predicted parameter \( X \) in the current HQF (suppose at time instant \( l \)), \( X_{l-1} \) is the parameter \( X \) in the previous HQF, \( T_l \) and \( R_{l-1} \) are separately the target bit allocation in the current HQF and actual bit allocation in its STR, \( R_{l-1} \) and \( R_{l-1} \) are separately the actual bit allocation in the previous HQF and its STR, \( k \) is updated by a linear regression method after coding each HQF.

### 4. OVERHEAD AND MAD PREDICTION

#### 4.1 Overhead prediction

The overhead is obtained through temporal prediction. For the first LQF of a GOP, the overhead is predicted from that in the first LQF of the previous GOP. From the second LQF to the next HQF, the overhead bits increase because of inaccurate prediction bring by the enhanced temporal distance between each current frame and LTR. As shown in Fig. 9, the overhead bits in frames (from the first LQF to the next HQF) following a HQF are nearly linear relationship with the temporal distance between each current frame and HQF. So in this case, the predicted overhead \( H_i^{pred} \) (suppose in the \( i \)th frame) can be calculated as

\[ H_i^{pred} = \frac{1}{n} \sum_{j=1}^{n} (H_{j-1} + \Delta H \times (D_j - D_{j-1})), \]  

where \( n \) is the number of encoded LQFs in the GOP, \( D_i \) is the temporal distance between frame \( i \) and previous HQF,
is the actual overhead in the $i$-th frame, $\Delta H$ is the average increase of overhead bits between two frames and is updated with linear regression after coding each frame.

4.2 MAD prediction

MAD prediction method is the same as that in overhead prediction. For the first LQF of a GOP, the MAD is predicted from that in the first LQF of the previous GOP. From the second LQF to the next HQF, MADs increase because of inaccurate prediction bring by the enhanced temporal distance between each current frame and previous LTR. As shown in Fig. 10, the MADs in frames (from the first LQF to the next HQF) following a HQF are nearly linear relationship with the temporal distance between each current frame and HQF. So in this case, the predicted MAD $MAD_i^{\text{pred}}$ (suppose in the $i$th frame) can be calculated as

$$MAD_i^{\text{pred}} = \frac{1}{n} \sum_{i=1}^{n} (MAD_{i-1} + \Delta M \times (D_i - D_{i-1})),$$

where $n$ is the number of encoded LQFs in the GOP, $MAD_{i-1}$ is the actual overhead in the $i$-th frame, $\Delta M$ is the average increase of MADs between two frames and is updated with linear regression after coding each frame.

![Fig. 10. MAD in frames following HQF.](image)

5. EXPERIMENTAL RESULTS

The proposed rate control is adopted by integrating the method into the H.264/AVC reference software JM10.2. The first P frame is set as the first HQF. For bit allocation, Rate-Qstep model, overhead and MAD prediction, the proposed methods in Section 2, 3 and 4 are separately adopted. In all experiments, the buffer size is set to four times of the channel rate and the initial buffer fullness is set to half of the buffer size. To maintain the smoothness of visual quality in LQFs of a GOP, the QP in each current LQF is adjusted by

$$QP_i = \min\{QP_{i-1} + 1, \max\{QP_{i-1} - 2, QP^*_i\}\},$$

where $i$ is the temporal location of the current LQF, $QP^*_i$ is the QP value in the current frame, and $QP_{i-1}$ is the QP

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Target bit rate(kbps)</th>
<th>Rate control scheme</th>
<th>Coded bit rate</th>
<th>PSNR(dB)</th>
<th>Improvement over (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hall (qcif)</td>
<td>71.86</td>
<td>Proposed RC</td>
<td>71.58</td>
<td>39.87</td>
<td>0.25</td>
</tr>
<tr>
<td>Highway (qcif)</td>
<td>40.32</td>
<td>Proposed RC</td>
<td>40.71</td>
<td>36.78</td>
<td>0.16</td>
</tr>
<tr>
<td>Waterfall (cif)</td>
<td>232.82</td>
<td>Proposed RC</td>
<td>232.41</td>
<td>36.11</td>
<td>0.20</td>
</tr>
<tr>
<td>Container (qcif)</td>
<td>17.61</td>
<td>Proposed RC</td>
<td>17.93</td>
<td>34.75</td>
<td>0.24</td>
</tr>
<tr>
<td>News (cif)</td>
<td>443.09</td>
<td>Proposed RC</td>
<td>443.46</td>
<td>39.91</td>
<td>0.20</td>
</tr>
<tr>
<td>Paris (qcif)</td>
<td>253.08</td>
<td>Proposed RC</td>
<td>253.46</td>
<td>37.09</td>
<td>0.18</td>
</tr>
<tr>
<td>Foreman (cif)</td>
<td>658.74</td>
<td>Proposed RC</td>
<td>658.32</td>
<td>39.91</td>
<td>0.20</td>
</tr>
</tbody>
</table>
value in the frame at time instant $i-k$. For the first LQF in a GOP, $k$ is 2; for the remaining LQFs in a GOP, $k$ is 1.

Besides the proposed rate control scheme, JU-DFMC in [29], JU-DFMC in [9], rate control G012 for H.264 [25] are tested and compared. JU-DFMC [9] did not utilize rate control, however, it has the best performance over any JU-DFMC scheme before it, so its performance is presented for performance comparison. In rate control algorithm G012 [25], two most recently decoded frames are utilized as reference frame.

To evaluate the performance, 7 video sequences including Hall, Highway, Waterfall, Container, News, Paris and Foreman are separately encoded. The test sequences are all encoded at a frame rate 30 fps and in mode of IPPP structure. Each sequence has 120 frames. In the motion estimation, the maximum search range is ±16. The entropy coding is context-adaptive binary arithmetic coding (CABAC).

The coding performance of the proposed rate control for JU-DFMC (Proposed RC), JU-DFMC in [29], JU-DFMC in [9], and rate control algorithm JVT-G012 [25] are shown in Table 1. The encoded bit rate of JU-DFMC in [9] is utilized for target bit rate. For the given target bit rate of the test sequences, the actual bit rate and PSNRs in every scheme are presented respectively. It can be seen that the proposed method can bring accurate rate control for JU-DFMC. At the same time, the proposed method has better performance than the other schemes. Compared with JU-DFMC in [29], the PSNR gains bringing by the proposed rate control method are separately 0.25, 0.16, 0.20, 0.24, 0.20, 0.18 and 0.20 dB in sequences Hall, Highway, Waterfall, Container, News, Paris, and Foreman. Compared with JU-DFMC in [9], the PSNR gains are separately 0.46, 0.23, 1.03, 0.83, 0.59, 0.42 and 0.33dB. It can also be seen that for either static or fast motion sequences, the proposed adaptive JU-DFMC can bring better performance than the other schemes. Furthermore, the rate-distortion curves achieved from the proposed rate control for JU-DFMC (Proposed RC), JU-DFMC in [29], JU-DFMC in [9], and JVT-G012 [25] in some sequences are shown in Fig. 11. These figures prove that at different bit rate, the proposed rate control method can achieve better performance.

Fig. 12 shows the target bits (Target) and actual coded bits (Actual) in every frame of the proposed scheme. The mismatch is in a small range, this shows that the proposed bit allocation achievement method, i.e., QP determination method is effective.

![Rate Distortion curves comparison.](image)

### 6. CONCLUSION

In this paper, a rate control scheme for JU-DFMC has been proposed. In JU-DFMC, two kinds of frames (LQF and HQF) are existed according to bit allocation. For each current frame, one STR and one LTR are utilized for motion prediction. Owning to the coding structure of JU-DFMC, several improvements were introduced to achieve rate control. Firstly, a linear increased bits allocation was proposed for LQFs to obtain smooth quality. Secondly, R-Q models for different LQFs and HQFs were given. Thirdly, overhead prediction and MAD prediction were presented. The proposed method can bring accurate rate control for JU-DFMC. At the same time the PSNR is better than the other methods. The proposed scheme can be widely utilized in multimedia, video conference and video surveillance. Furthermore, in cognitive radio channel or multi-channel system, the proposed bit allocation method can be utilized to instruct the bandwidth allocation. In the future, rate control for the multi-program transmission will be further exploited.

### 7. ACKNOWLEDGEMENT

This work was supported by the National Science Foundation of China under Grant 60736043, and the
National Basic Research Program of China, 973 Program 2009CB320903.

Fig. 12. Target bits and coded bits in every frame of proposed scheme.

8. REFERENCES