<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>A perceptual based rate control scheme for MPEG-2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Choi, KT; Chan, SC; Ng, TS</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>1998</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/54077">http://hdl.handle.net/10722/54077</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>©1998 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.: This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.</td>
</tr>
</tbody>
</table>
A PERCEPTUAL BASED RATE CONTROL SCHEME FOR MPEG-2

K.T. Choi, S.C. Chan and T.S. Ng
Department of Electrical and Electronic Engineering
The University of Hong Kong, Pokfulam Road, Hong Kong
Email: scchan@hkuee.hku.hk

ABSTRACT

In this paper, a new perceptual based rate control algorithm for MPEG-2 is presented. The algorithm first determines the target bit count for each frame using simple statistical models. Precise bit allocation is used to adjust the macroblock quantization scale factors to meet the given bit count, with the uniform visual fidelity as the primary objective. Since the buffer is very stable, it is less sensitive to transmission delay. Experimental results showed that it gave better visual quality and less buffer variations as compared to the TM5 rate control scheme.

1. INTRODUCTION

The MPEG-2 [1] is an important video coding standard for coding high quality video such as high-definition television, and high-quality digital video recording. In constant bit rate application, rate control has played an important role in improving and stabilising the video quality. The well-known TM5 rate control scheme determines the current target bit count using previous encoded frame statistics, and adjusts the quantization step size based on the buffer occupancy. Therefore, the quantization error of adjacent blocks can vary significantly and it generates visible blocking artifacts. Recently, there has been a growing interest in rate-distortion optimal techniques for both bit allocation and rate control. The rate control problem can be formulated as a constrained optimization problem and solved by the Lagrangian or minimax technique [5-7]. Ortega et al. [5], proposed a rate control scheme which used dynamic programming to search for the true global optimal solution. Lin et al. [8] proposed a faster algorithm by encoding the source with only a few quantization steps and the rate-distortion value for other quantization steps is found using interpolation. However, the source video has to be encoded several times and the complexity is still very high. In a previous work [3], we had proposed a new buffer control algorithm for motion-compensated hybrid DPCM/DCT coding. The quantization scale factors for each macroblock are determined to meet a given target bit budget. The salient features of that scheme are that (1) the quantization scale factors are determined using information of the whole picture; (2) it has precise control of the buffer; and (3) it tries to allocate the given number of bits as efficient as possible in a rate-distortion sense. However, this scheme is not directly applicable to MPEG-2 coder, since the encoding of the I, P, and B pictures are not completely independent. In fact, P and B pictures are predicted from the I pictures and so on.

In this paper, a new perceptual based rate control algorithm is presented to solve this problem. It first determines the target bit count for each frame using certain statistical model. The bit allocation algorithm proposed in [3] can then be used to determine the macroblock quantization scale factors to meet the given bit count, with the uniform visual fidelity as the primary objective. More precisely, the frame bit allocation determines the bit counts to be spent for each frame using an accurate statistical rate-distortion model. Then, we perform the macroblock bit allocation, similar to [3], to meet exactly the target bit count found in the frame bit allocation. Experimental results showed that the proposed scheme gave better visual quality as compared to the TM5 model at comparable PSNR values. The buffer is more stable and is less sensitive to transmission delay. The proposed algorithm is discussed in Sections 2 and 3. In Section 4, the simulation results and discussion are presented, followed by the conclusion in Section 5.

2. FRAME BIT ALLOCATION

There are several statistical approaches that can be used to determine the approximate number of bits required for each frame. Here, we shall make use of the method in [2]. The rate distortion function is modeled as a quadratic function between the bit rate and the average quantization scale of a picture [2]. In particular, the target bit rates $T_i$, $T_p$, $T_b$ for the I, P, and B pictures can be found from the following equations:

$$k_p Q_i = k_i Q_p$$
$$k_b Q_i = k_i Q_b$$
$$T_i + N_p T_p + T_b T_b = RGOP$$
$$T_i = a_i Q_i^{-1} + b_i Q_i^2$$
The activity measure for the macroblock $j$ is then normalized against the mean activity value $\text{avg}_j$ of the previous encoded picture of the same type. The normalized macroblock local activity $N_{-act}j$ is defined as

\[ N_{-act}j \triangleq \frac{2 * \text{act}_j + \text{avg}_j}{\text{act}_j + 2 * \text{avg}_j} \]

The perceptual distortion $D^*_n(Q_n)$ for the $j$-th macroblock is obtained by multiplying the original distortion $D_n(Q_n)$ by the spatial activity measure $N_{-act}j$ as follow,

\[ D^*_n(Q_n) = D_n(Q_n) \cdot N_{-act}j \]

The proposed macroblock bit allocation scheme aims to vary the quantizer scale of each MB until the frame bit budget determined by the frame bit allocation scheme is reached. The algorithm is summarized as follows [3]:

1. Initialize all quantization scale factors $Q_n$ to maximum value $Q_{max}$, and the bit count $B$.

2. Calculate functions $D^*_n(Q_n)$ and $R_n(Q_n)$ for $n = 1,...,N$ and $q = 1,...,Q_{max}$.

3. Calculate the efficiency of the $k$-th block as

\[ \lambda_k = \frac{\Delta D_k \to q}{\Delta R_k \to q} \]

where $k = 1,...,N$, and $\Delta D_k \to q$ and $\Delta R_k \to q$ are the change in distortion and the change in overall bit-rate used for all macroblocks when the quantization scale factor of the $k$-th macroblock $Q_k$ is replaced by $q$. As each quantizer is independent, these increments can be calculated as

\[ \Delta D_k \to q = D^*_k(q) - D^*_k(Q_k) \]
\[ \Delta R_k \to q = R_k(q) - R_k(Q_k) \]

4. Find the macroblock $m$ with maximum $\lambda_k$:

\[ \lambda_m = \max_k \lambda_k \]

This means that the ratio of decrease in distortion to increase in bit-rate is maximized over all possible reduced values for each quantization scale factor.

5. Update the quantization scale factor $Q_m = q$ and the current bit count $B$. Calculate new value of $\lambda_m$ and insert it into the list.

6. Repeat Steps 4 and 5 while $T_{pict} > B + R_{overhead}$.
Under the *macroblock bit allocation*, the bit count assigned to each individual picture in Section 2 is exactly met. Better perceptual quality is achieved since the bit allocation is adapted to human visual perception using local spatial activity. The proposed scheme not only meets the CBR transmission constraint but also minimizes the perceptual coding distortion. Both the buffer requirement and the delay can be minimized, with better rate distortion trade off.

4. RESULTS AND DISCUSSION

Simulation is performed on composite video sequence at 30 frame/sec with the picture size of 352 x 240 pels. The composite sequence with a total of 490 frames consists of the video sequence “Table tennis”, “Mobile and Calendar”, “Football”, and “Flower Garden”. The target average bit rate is about 1.5 Mbit/sec with a buffer size of 96 msec. There are 15 pictures in each GOP with two B-frames between two anchor frames. To demonstrate the performance of the proposed scheme, we compare it with the rate control method in MPEG TM5 model.

Figure 1 showed the zoomed views of the 68th encoded frame using different algorithms are displayed. Both blocky and ringing effects show up in the TM5 case. In particular, the surface of the table and the arm of the player in the encoded image of the proposed algorithm is much smoother than that of the TM5 scheme. Also, the surrounding region of the ball and the paddle appears to be very noise in the TM5 model.

Figure 2 shows the bits generated for the I, P, and B frames. It can be seen that the bit counts is very stable in the proposed approach. The reduction in the standard deviation of the bit counts are 58%, 69% and 75% for I, P, B frames, respectively (Table 2). Also the average peak signal-to-noise ratio (Fig. 2) is improved slightly by about 0.3 dB.

5. CONCLUSIONS

In conclusion, we have proposed a effective rate control strategy for MPEG-2 coder. Simulation result shows that less bit rate fluctuation and better perceptual quality than the TM5 model is achieved.

References


Table 1. Close form solution of the target bit rate for I, P, and B frames, where
\( K_1 = K_i / K_p \), \( K_2 = K_i / K_b \), and \( K_3 = K_p / K_b \), \( \gamma = -RGOP \), \( \delta = \beta^2 - 4\alpha \gamma \).

<table>
<thead>
<tr>
<th>Frame Type</th>
<th>Model Parameters</th>
<th>if ( \delta &gt; 0 )</th>
<th>if ( \delta &lt; 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T_i )</td>
<td>( \alpha = b_1 + N_p b_2 K_1 + N_b b_3 K_2 )</td>
<td>( Q_i^{-1} = \frac{\sqrt{\delta - \beta}}{2\alpha} )</td>
<td>( Q_i^{-1} = \frac{-\gamma}{\beta} )</td>
</tr>
<tr>
<td></td>
<td>( \beta = a_1 + N_p a_2 K_1 + N_b a_3 K_2 )</td>
<td>( T_i = a_1 Q_i^{-1} + b_1 Q_i^{-2} )</td>
<td>( T_i = a_1 Q_i^{-1} )</td>
</tr>
<tr>
<td>( T_p )</td>
<td>( \alpha = N_p b_2 + N_b b_3 K_3^2 )</td>
<td>( Q_p^{-1} = \frac{\sqrt{\delta - \beta}}{2\alpha} )</td>
<td>( Q_p^{-1} = \frac{-\gamma}{\beta} )</td>
</tr>
<tr>
<td></td>
<td>( \beta = N_p a_2 + N_b a_3 K_3 )</td>
<td>( T_p = a_2 Q_p^{-1} + b_2 Q_p^{-2} )</td>
<td>( T_p = a_2 Q_p^{-1} )</td>
</tr>
<tr>
<td>( T_b )</td>
<td>( \alpha = N_p b_2 / K_3^2 + N_b b_3 )</td>
<td>( Q_b^{-1} = \frac{\sqrt{\delta - \beta}}{2\alpha} )</td>
<td>( Q_b^{-1} = \frac{-\gamma}{\beta} )</td>
</tr>
<tr>
<td></td>
<td>( \beta = N_p a_2 / K_3 + N_b a_3 )</td>
<td>( T_b = a_3 Q_b^{-1} + b_3 Q_b^{-2} )</td>
<td>( T_b = a_3 Q_b^{-1} )</td>
</tr>
</tbody>
</table>

Table 2. Standard deviation of the bit counts of the SIF sequence at 1.5 Mbps.

<table>
<thead>
<tr>
<th>Standard Deviation</th>
<th>B</th>
<th>P</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>5.558</td>
<td>4.9374</td>
<td>1.4019</td>
</tr>
<tr>
<td>TM5</td>
<td>13.279</td>
<td>16.02</td>
<td>5.6602</td>
</tr>
<tr>
<td>% in reduction</td>
<td>58.14%</td>
<td>69.18%</td>
<td>75.23%</td>
</tr>
</tbody>
</table>

Figure 2: Upper left, upper right and lower left are the bit rate distribution plot for the I, P, and B frames of the SIF sequence at 1.5 Mbps respectively. Lower right shows the PSNR plot of the luminance for the reconstruction of the first 110 frames. The solid line for the TM5 rate control and the dotted line for the proposed algorithm.