A Pruning Based Fast Rate Control Algorithm for MPEG Coding

T. G. Venkatesh
Sanyo LSI Technology India Pte. Ltd.,
Bangalore-560066, India
Email: tgvenkatesh@hotmail.com

S. Srinivasan
Department of Electrical Engineering
Indian Institute of Technology,
Chennai - 600 036, India
Email: srini@ee.iitm.ernet.in

Abstract

This paper presents a fast rate control scheme that regulates the coded bit stream generated by an MPEG video encoder such that the resulting variable rate compressed bit stream can be transmitted over a fixed rate channel. A new bit rate regulation scheme is proposed by using a technique called pruning in which only the required number of low frequency coefficients of the Discrete Cosine Transform (DCT) are computed without the need to compute all the coefficients. The use of pruning technique reduces the overall execution time of the encoder implementation.

1. Introduction

With rapid advances in communications and integrated circuit technology, digital transmission of video signals is becoming more and more common place. However, raw digital video signal usually contains a huge amount of information owing to redundant data and requires very high channel speeds for transmission. In order to reduce the requirement of high channel speeds, it is necessary to employ compression techniques that reduce the data rate without sacrificing the quality of images in any large measure. Several video coding standards such as H.261 [1], MPEG [2], etc. have been defined for compressing video signals. It is well known that most of the coding schemes generate variable bit rates (VBR), where the number of transmitted bits change from time to time. However, practical transmission media usually exhibit fixed channel capacities. In order to transmit VBR data through a constant bit rate channel, a buffer is required to regulate the output bit stream generated by the encoder. The buffer delay varies according to the buffer size. A large buffer size causes a long frame delay, while a smaller buffer size leads to the occurrence of buffer overflow which results in a poor video quality. Hence, to reduce the overflow, a feedback mechanism, called the rate control, is put in place that enforces a higher compression when the amount of coded bits in the buffer exceeds a prescribed threshold. Rate control for video coding has been actively studied in recent years. Several rate control schemes [3-5] exist for video transmission. In all these schemes, most of the high frequency coefficients in the DCT transformed block are reduced to zero due to non-uniform quantization. As a result, the effort spent in computing these coefficients is wasted. This drawback is overcome in the proposed work by using a pruning technique, rather than quantization, in order to effect rate control. Pruning [6] is a process whereby the butterfly structure of the DCT computation is modified so as to compute only the required low frequency coefficients, without having to compute all the coefficients.

The paper is organized as follows. The coding algorithm for the MPEG standard is described in Section 2, while Section 3 describes the proposed rate control algorithm. Experimental results are presented in Section 4 and conclusions are drawn in Section 5.
2. Coding Algorithm

Out of several standards, the MPEG is the most popular and commonly used standard for compression of digital video signals. The MPEG video compression relies on two basic techniques, namely, block-based motion compensation for reduction of temporal redundancy and transform (usually DCT) domain based compression for reduction of spatial redundancy. As per the MPEG standards, an input video sequence is divided into groups of pictures (GOP’s) and each picture is coded in one of the three modes, namely, Intra frame (I), Predicted frame (P) or Bidirectionally predicted frame (B). The input frame is partitioned into macroblocks, each of size 16x16 pixels. In the intraframe (I) mode, DCT is employed for transforming each block of 8x8 pixels into frequency domain. In the P and B frame, for each macroblock, the best matching macroblock within a specific tracking range is predicted from the reconstructed previous or future frame by means of motion estimation and compensation techniques. The DCT is then performed for each of the four 8x8 blocks of the remaining residuals. In all three modes (I, P and B), quantization and variable length coding (VLC) are applied to the transformed coefficients.

3. Proposed Rate Control Scheme

The proposed algorithm uses the GOP as the basic coding unit for frame-level bit allocation. Each GOP is allocated an equal number of bits based on the channel rate. For example, if the sequence defines a 10-picture GOP at 30 frames/sec., one third of the channel bit rate is allocated to each group. Within each GOP, the bit allocation is divided among the constituent frames, depending on the frame type. A model is constructed for predicting the number of bits needed to code each frame within the GOP based on the weighting factors k_i, k_p and k_b. If B_i, B_p and B_b are the target number of bits (number of bits allocated per frame) for coding I, P and B frames respectively, then they are computed as

\[ B_i = k_i K, \]  
\[ B_p = k_p K \] and  
\[ B_b = k_b K, \]  

where \( K = \frac{G}{k_i + N_p k_p + (N_{GOP} - N_p - 1) k_b} \). G is the number of bits assigned to the current GOP, and \( N_{GOP} \) and \( N_p \) are the total number of frames and number of P frames in a GOP respectively. The weighting factors \( k_i, k_p \) and \( k_b \) are initially set to 6, 4 and 1 respectively as per the guidelines of the MPEG standard. After encoding all the frames in the GOP, these weighting factors are updated according to the number of bits that were actually produced in the previous GOP. If there is little motion or change in the video, then a greater proportion of the bits is allotted to I frames (\( k_i > k_p \)). On the contrary, if there is a lot of motion or change, then the proportion allotted to I frames is reduced and most of the bits allotted to P frames (\( k_i < k_p \)). In either case, B frames are allotted the minimum number of bits. Before encoding the first frame in a GOP, G is updated as

\[ G = \frac{\text{Channel rate} \times N_{GOP}}{\text{frame rate}} + \Delta, \]  

where \( \Delta \) is the difference between the target number of bits and the output bits of the previous GOP. At the start of the sequence, \( \Delta \) is set to 0.
In the present scheme, the pruning level, which is defined as the number of diagonals of DCT coefficients computed for an 8x8 block of image pixels, is used as the coding parameter for regulating the bit rate. Based on the target number of bits allotted to each frame, the pruning level is predicted for each block so that the encoded output bits are closer to the target bits. The prediction of pruning level for different bit rates is handled separately for intra (I) and non-intra (P and B) frames. For the intra frames, the adaptive pruning technique [7] is used for predicting the pruning level. In adaptive pruning technique, the pruning level for each block is determined based on the energy information packed in a few low frequency coefficients. The adaptive pruning algorithm in [7] is designed for achieving an acceptable picture quality. However, the pruning level obtained from the adaptive pruning expression given in [7] is not enough to achieve the desired bit rate. Therefore a variable parameter called LEV is added to the adaptive pruning expression, resulting in a modified expression given by

$$\text{pruning level} = \left( \frac{c_2 - c_0}{256} \right) + \text{LEV},$$

where $c_0$ is the energy packed in the DC coefficient and $c_2$ is the sum of the energy packed in the first and second AC diagonal coefficients of the DCT matrix for each 8x8 block of the image. LEV varies from 0 to 14, depending upon the desired bit rate. Based on the experiments performed on different video sequences, a lookup table is constructed for obtaining the value of LEV for different bit rates. Thus, before encoding an I frame, LEV for the corresponding B1 is obtained from the lookup table. It is to be noted that LEV varies only from frame to frame and not from block to block.

For the non-intracoded blocks in P and B frames, which contain predominantly high frequency coefficients, the adaptive pruning is not suitable as the pruning levels are predicted from the low frequency coefficients. Hence, pruning levels for the non-intracoded blocks are computed based on the energy of the prediction error blocks, using the information gathered during motion estimation and is given by

$$E = \sum_{i=0}^{2} \sum_{j=0}^{2} \left( e[i][j] \right) ^2,$$

where $e$ is the prediction error block estimated from motion estimation. In the present scheme, these prediction error blocks are classified by their energy values.

![Three-dimensional plots](image)

**Figure 1.** Three-dimensional plots for a random frame of size 256x256 pixels of the salesman sequence (a) bits/block vs energy/block vs pruning level (b) PSNR/block vs energy/block vs pruning level
The bit rate and quality are studied for the blocks that fall in different energy ranges and pruning levels, using several video sequences. The compression and quality behavior for a random frame of the Salesman sequence are shown in Fig. 1. From the plots, it can be seen that, above a pruning level of eight, there are only marginal increases in both the quality and compression. Hence, for each energy category, the minimum pruning level, \( p_{\text{quality}} \), is chosen in such a way that the reconstructed picture will have an acceptable quality. In order to meet the desired bitrate, \( \text{LEV} \), which corresponds to \( B_r \) or \( B_b \), is also used in addition to \( p_{\text{quality}} \). Based on these details, the lookup tables for both P and B frames are developed to generate the required pruning level (sum of \( p_{\text{quality}} \) and \( \text{LEV} \)) for each predicted block.

4. Results

Simulations were carried out using the proposed rate control scheme on the MPEG coded streams of various video sequences as explained above. The format for all the sequences tested was the same, namely, frames of size 256x256 pixels and 25 frames per second. Although any GOP structure can be used, the simulations used a GOP structure IBBPBBPBB. The simulated results for the Salesman and the Car sequences coded at 400kbps and 600 kbps respectively are presented in Fig. 2. Fig. 2(a), 2(b) and 2(c) respectively show the plots for PSNR, encoded output and average pruning level versus the frame number. From Fig. 2(c), it is seen that the pruning level varies depending on the importance of the frame type, i.e., the maximum pruning level of around 7 for I frames and the minimum pruning level (0 to 3) for B frames. Fig. 2(d) shows the buffer occupancy results. Plots for the encoded bit rate (bits/sec) as a function of time for the Salesman sequence coded at 400kbps and the Car sequence coded at 600kbps respectively are shown in Fig. 3(a) and 3(b). The output bit rates are compared with the channel rates (shown by the horizontal lines) to see how the encoded bit rate, coded using pruning level as the rate control parameter, is close to the channel rate in each case.

5. Conclusion

A simple rate control algorithm using pruning technique has been presented for coding the MPEG frame sequences. Pruning level, instead of the conventional quantization, is used as the rate control parameter. Three different lookup tables, one each for I, P and B frame types, are constructed and stored in the memory. These tables output the pruning level required for the desired bit rate. Simulation results demonstrate that the output bit rate is maintained with very little variations. The PSNR of the reconstructed frames remains at a relatively stable level over the frames of a continuous scene. The use of pruning technique reduces the overall execution time of the encoder implementation by up to 50%.

6. References

Figure 2. Results for the salesman sequence and the car sequence coded at channel rates of 400 kpbs and 600 kpbs respectively.

Figure 3. Encoded bit rate vs time (a) the salesman sequence coded at a channel rate of 400kbps and (b) the car sequence coded at a channel rate of 600kbps.