A PREDICTIVE H.263 BIT-RATE CONTROL SCHEME BASED ON SCENE INFORMATION

Pohsiang Hsu and K. J. Ray Liu

Department of Electrical and Computer Engineering
University of Maryland at College Park
College Park, Maryland, USA

ABSTRACT

For many real-time visual communication applications such as video telephony, the transmission channel in used is typically constant bit rate (e.g. PSTN). Under this circumstance, the encoder's output bit-rate must be regulated to meet the transmission bandwidth in order to guarantee a constant frame rate and delay. In this work, we propose a new predictive rate control scheme based on the activity of the scene to be coded for H.263 encoder

1. INTRODUCTION

In the past few years, emerging demands for digital video communication applications such as video conferencing, video e-mailing and video telephony over the public switched telephone networks (PSTN) has prompted the creation of the ITU-T H.263 video compression standard [1] to replace the H.261 video compression standard.

Similar to most video compression standards, H.263 relies on block based motion estimation/compensation to remove the temporal redundancy and discrete cosine transform (DCT) to remove the spatial redundancy in the digitized video signals. In addition, run length, fixed length, and variable length coding are applied to achieve further compression. Due to the type of compression used, the H.263 encoder intrinsically produces variable bit-rates, where the number of bits generated varies from frame to frame depending on the contents of the input video signal.

Typically, a buffer is placed between the encoder and the channel to smooth out the variation of the encoder's output bit-rate. The encoder will generate bits and store them in the buffer while the transmitter will remove the bits from the buffer. When the source rate exceeds the transmission rate, the buffer acts as a temporary storage for the encoded bits so that they may be transmitted later allowing the encoding operation to continue. However, when the buffer is full, the encoder must cease generating bits by dropping frames thereby causing an interruption to the smoothness of the video. The selection of the buffer size is dependent on the maximum delay allowed. A large buffer size tends to ensure smoother video but causes longer delay while a small buffer size guarantees low delay but may be more susceptible to dropping frames due to buffer overflow.

For many real-time visual communication applications such as video telephony, the transmission channel in used is typically constant bit rate (e.g. PSTN). Under this circumstance, the encoder's output bit-rate must be regulated to meet the transmission bandwidth in order to guarantee a constant frame rate and delay. Given the channel rate and a target frame rate, we derive the target bits per frame by distribute the channel rate equally to each frame. Rate control is typically based on varying the quantization parameter to meet the target bit budget for each frame. Many rate control scheme for the block-based DCT/motion compensation video encoders have been proposed in the past. These rate control algorithms are typically focused on MPEG video coding standards. Rate-Distortion based rate control for MPEG using Lagrangian optimization [4] have been proposed in the literature. These methods require extensive rate-distortion measurements and are unsuitable for real time applications. Model based approaches based on rate-quantization modeling have also been proposed in the literature. In [7], each frame is subjected to several pass of actual encoding with different quantization parameter to generate a rate-quantization model used in rate control. This method is robust to scene changes but requires actual coding to estimate the rate-quantization curve. In [8], a normalized rate-quantization model has been proposed and it's parameter are computed from training data. The complexity of this approach is low but the performance of the prediction might be dependent on the effectiveness of the training data. For applications requiring low end-to-end delay and low complexity such as video telephony, predictive rate control approaches has been popular solutions. Indeed, the H.263 Test Model 5 Rate control belongs to this category. Typically, the predictive schemes [3], [6] (for MPEG) make rate allocation decisions based on currently available information such as the state of buffer fullness, block activities, etc.

In this paper, we propose a new predictive rate control scheme based on the activity of the scene to be coded for H.263 encoder. The scheme will be applied to the forward predicted frame and is based on the following three steps. First, we estimate the number of bits available for encoding the motion residuals. Second, select a frame quantization parameter based on motion compensated residuals of the current frame and the previous frame. Third, modify the frame quantization parameter for each macroblock so that the bits available for encoding the motion residuals are distributed according to the amount of block activities as indicated by the motion compensated residuals.

0-7803-6536-4/00/$10.00 (c) 2000 IEEE

Authorized licensed use limited to: IEEE Xplore. Downloaded on December 22, 2008 at 01:44 from IEEE Xplore. Restrictions apply.

2. H.263 TMN5 RATE CONTROL

The H.263 Test Model, Near Term, Version 5 (TMN5) describes an implementation of H.263 encoder that serves as a basis for experimentation and development. In this section, we describe the real time rate control method specified in TMN5. The mechanisms used for regulating the output bit-rate are changing the quantization step size on macroblock level and dropping frames. We note that the TMN5 rate control strategy ensures that buffer overflow will not occur by dropping frames but this also implies that no minimum frame rate can be guaranteed.

Basically, the TMN5 rate control scheme consists of two main steps. The first step is to determine how many frames to skip after encoding a frame and the second step is to determine the quantization parameter for each macroblock in a picture so as to meet the specified target number of bits. The decisions are based on the state of the buffer fullness (i.e. the number of bits in the output buffer.) Let us first introduce some necessary notations that will be used to describe the scheme.

Let us denote the desired target frame rate by f_{target} and the channel bit-rate by R. Let FR denote the source input frame rate which typically equals to 25 or 30 Hz (frames per second). The number of bit transmitted per source frame is $R_t = R/FR$ which can be seen as the number of bits transmitted during the time spanned by two adjacent frames. Assuming that we distribute the bandwidth evenly to each encoded frame, the target number of bits per encoded frame B_{target} is

$$B_{target} = \frac{R}{f_{target}}. \qquad (1)$$

Let B_n denote the number of bits used to encode the nth frame, MB denote the total number of macroblocks in a picture, and $B_{n,k}$, $k = 0, 1, \cdots, MB$ be the number of bits used to encode from the first macroblock up to but not including the kth macroblock for the nth frame. Clearly, we have the total number of bits used to encode the nth frame B_n is equal to $B_{n,MB}$. In addition, let us denote the buffer fullness as b_{enc} and we initialize it as

$$b_{enc} = B_{target} + TBF, \qquad (2)$$

where TBF is the target buffer fullness which is a specified threshold set as

$$TBF = 3 \times \frac{R}{FR}. \qquad (3)$$

Basically, the rate controller will determine the number of frames to be dropped after each encoded frame so that the buffer fullness b_{enc} remain below the target buffer fullness TBF.

In TMN5, the first frame of the GOP (frame number $n = 0$) is intra-coded using a default quantizer QP (typically 16) for all of the macroblocks in the frame. Then, the following steps are applied repetitively to regulate the bit-rate of the P-Frames. Let n denote the frame that was just encoded. First, the buffer fullness is updated as

$$b_{enc} = b_{enc} + B_n. \qquad (4)$$

Then, we determine the number of frames to be skipped by finding an integer f_{skip} such that

$$b_{enc} - f_{skip} \cdot R_t \leq TBF \leq b_{enc} - (f_{skip} - 1) \cdot R_t. \qquad (5)$$

Once we determine f_{skip}, the next frame to be encoded is $n = n + f_{skip} + 1$.

For each frame to be encoded, we have to select quantizer step sizes for each macroblock in the frame. In TMN5, stationarity between temporally close frames are assumed which means that the results of the rate control on the previous frame will be used in performing rate control of the current frame. Therefore, the selection of the quantizer step sizes is based on the average quantizer step size QP_{prev} of the previously encoded frame, the previous target bit per frame deviation, and the current target bit deviation. First, a global adjustment factor G_{adj} is computed in the beginning as,

$$G_{adj} = \frac{B_{n-f_{skip}-1} - b_{target}}{2 \cdot B_{target}} \qquad (6)$$

Then, for each macroblock at the beginning of the macroblock line, we compute a local adjustment factor L_{adj} as,

$$L_{adj} = \frac{12 \cdot (B_{n,k} - \frac{k}{MB} B_{target})}{R}, \qquad (7)$$

where k denote the macroblock number. Then, each macroblock in a macroblock line will be quantized using the same step size computed as,

$$QP_{n,k} = QP_{prev}(1 + G_{adj} + L_{adj}). \qquad (8)$$

However in H.263 syntax, each adjustment of quantizer step size within a picture is limited to 2. Therefore, we have

$$QP_{n,k} = QP_{n,k-1} \bullet max(2, |QP_{n,k} - QP_{n,k-1}|), \qquad (9)$$

where $QP_{n,k-1}$ was the last quantizer step size used before the adjustment.

3. A PREDICTIVE H.263 BIT-RATE CONTROL BASED ON SCENE INFORMATION

The new scheme is similar to the TMN5 rate control in the overall structure. It consists of two main steps, frames skipping and quantization parameter selection for each macroblock. The frame skipping mechanism is the same as TMN5 rate control scheme. As in TMN5, the first frame of the GOP is intra-coded using a default quantizer QP (typically 16) for all the macroblocks in the frame. Therefore, we will be focusing our attention on the process of quantization parameter selection for the predicted frames that follows.

One main problem of the TMN5 rate control is the fact that the scheme tries to distribute the bits evenly to each macroblock as indicated by equation 7. This implies that the high activity blocks and the low activity blocks are given the same amount of bits to describe them. Since the high activity blocks usually occur around edges of moving objects, they convey important visual information. On the other hand, low activity blocks typically belong to the background region. Moreover, low activity block typically

0-7803-6536-4/00/$10.00 (c) 2000 IEEE

Authorized licensed use limited to: IEEE Xplore. Downloaded on December 22, 2008 at 01:44 from IEEE Xplore. Restrictions apply.

requires smaller number of bits to encode than a high activity block given the same quality constraint. Therefore, the equal distribution of bits implies the low activity block receives too many bits compared to the high activity blocks, which will result in the loss of perceptual quality.

To remedy this problem, we proposed to distribute the bits unevenly to the macroblocks according to their activity level. Thus, we need to define an efficient measure of activity level for each macroblock. Some popular measures of activity levels of blocks used in image and video coding includes variance, sum of squared values, and sum of absolute values. In our approach, we define the activity level for each macroblock by summing the squared values of the motion compensated residual for that macroblock. It is important to point out that the motion compensated residual does not need to be computed since it is readily available after the motion estimation process and right before the bit allocation process. Therefore, the computational cost of obtaining the activity level for each macroblock is not too high. Intuitively, the choice of this activity measure can be justified by the fact that typically moving objects result in large motion compensated residual while stationary background results in small motion compensated residual.

Therefore, we proposed to modify the local adjustment of the quantization parameter for the kth macroblock as follows:

$$L_{adj} = c * (B_{n,k} - \frac{\sum_{n=0}^{k-1} S_n}{S} B_{target}), \qquad (10)$$

where S_n, $n = 0, 1, \cdots, MB-1$ denotes the sum of squared value of motion residual (i.e. activity) of the nth macroblock, c is a penalty factor for not meeting the bit allocation for the previous blocks, and S is defined as the total activity of the frame,

$$S = \sum_{n=0}^{MB-1} S_n, \qquad (11)$$

In TMN5 rate control, the initial quantization parameter chosen for the frame to be encoded is adjusted based on the average quantization parameter used in the previous frame, and the previous target bit per frame deviation. This can be seen from the global adjustment parameter in equation 6. Therefore, the selection of the initial quantization parameter for the current frame in TMN rate control is based solely on the information obtained in the previous frame. While this approach performs well when the complexity of the current frame is similar to that of the previous frame, it can be problematic when the complexities of the two frames are not so similar. Suppose we know the complexity of the current frame and the previous frame before rate control takes place. Then, we can use the information to change the global adjustment parameter to reflect the difference in complexity. Specifically, we can increase/decrease the global adjustment when the complexity of the current frame is higher/lower than the previous frame.

Therefore, we propose to use the total activity of a frame as an indication of its complexity. Since the total activity of the current frame and the previous frame is available at the time of rate control, we propose the following modification to the global adjustment parameter:

$$G_{adj} = \frac{B_{n-f_{skip}-1} - b_{target}}{2 \cdot B_{target}} * (1 + \alpha * log_{10} \frac{S_{n-f_{skip}-1}}{S}), \qquad (12)$$

where $S_{n-f_{skip}-1}$ indicates the total activity of the previous frame, S indicates the total activity of the current frame, and α is an adjustment parameter. As we can see, the new global adjustment parameter reduces to the global adjustment parameter used in TMN5 rate control when the total activity of the previous frame equals to that of the current frame.

As we pointed out earlier, TMN5 rate control attempts to distribute bits evenly to each macroblock where the bit budget for each macroblock includes the header information, motion vector information, and DCT residual information. In our approach, we attempt to distribute bits according to the activity level, which is mostly related to the DCT residual information. This means that we should only distribute the total number of bits available for the DCT residual.

Since we can obtain the motion vectors from the motion estimation process prior to rate control, we propose to estimate the number of bits required for encoding the motion vectors. The estimate can be obtained by simulating the differential encoding of the motion vectors and count the resulting bits using the H.263 VLC motion vector codeword table. Thus, we attempt to estimate the total number of bits available for the DCT residual by subtracting the estimated number of bits needed for the motion vectors from the original target bit per frame B_{target} prior to rate control. In addition, the bit count for encoding macroblocks $B_{n,k}$ used in the local adjustment parameter is modified by not counting the bits required for encoding the motion vectors.

4. SIMULATION RESULTS

Simulations are conducted using two QCIF color test sequences, "Carphone", and "Foreman" to compare the proposed method with the TMN5 rate control scheme. The frame rates of the two test sequences were 30 frames per second. The target bit rate was set to be 32 kilobits per second with a target frame rate of 10 frame second. The baseline H.263 encoder from Telenor[2] was used for the simulation with the only difference being the rate control scheme used. The first frame was intra-coded using a quantization step size of 16.

Given the target frame rate and target bit rate, the target number of bits allocated to each encoded frame is 3200 bits per frame (i.e. target frame rate * target bit rate). Thus, the goal of the rate controller will be to vary the quantization parameter for the macroblocks to ensure that the frame is encoded using only or as close as possible to 3200 bits.

In Figure 1, we show the bit counts of the encoded frame using the proposed method and the TMN5 rate control method for the two test sequences. The target number of bits per encoded frame is also shown as the dash line in the figure. As we can see, the deviation from the target bits is significant smaller by using the proposed method as

Authorized licensed use limited to: IEEE Xplore. Downloaded on December 22, 2008 at 01:44 from IEEE Xplore. Restrictions apply.

Seq:	Scheme:	PSNR: (dB)	Fr. Rate: (Hz)	Rate: (b/f)	Std: (b/f)
A	TMN	33.65	9.76	3259	441
	Proposed	33.54	10.12	3167	212
B	TMN	31.38	9.51	3344	927.89
	Proposed	31.18	9.63	3310	464.93

Table 1: Performance Evaluation of the Proposed Rate Control Scheme. Sequence A denotes Carphone. Sequence B denotes Foreman. (b/f) denotes bits per frame.

oppose to the TMN5 rate control scheme.

In Table 4, the simulation results for both the TMN5 rate control and the proposed scheme are summarized. The encoded sequence using TMN5 rate control and our proposed method have about the same average PSNR (Peak Signal to Noise Ratio). The average PSNR the encoded sequence using our proposed scheme is slight below the one produced by using TMN5 rate control. However, by considering the fact that the mean frame rate produced by our proposed method is slightly higher, it can be concluded that essential the PSNR performance is about the same. We also listed the average number of bits allocated per frame under the rate column. As we can see, the average number of bits for both methods are close to that of the target bit per frame (3200 bits). The most important statistic is the standard deviation of the bits allocated per frame, which measures the fluctuation of the buffer. As we can see, the proposed method lowers the buffer fluctuation significantly by about 50 percent.

5. CONCLUSION

In this paper, we proposed a new predictive rate control method for the H.263 video encoder. The scheme has a low computational requirement and it is suitable for real time rate control. Efficient bit allocation is achieved by taking into account the activity of the macroblocks in the frame to be encoded. As we can see from the simulation results, the new scheme performs more accurate bit allocation than the TMN rate control scheme without sacrificing the quality of the encoded video.

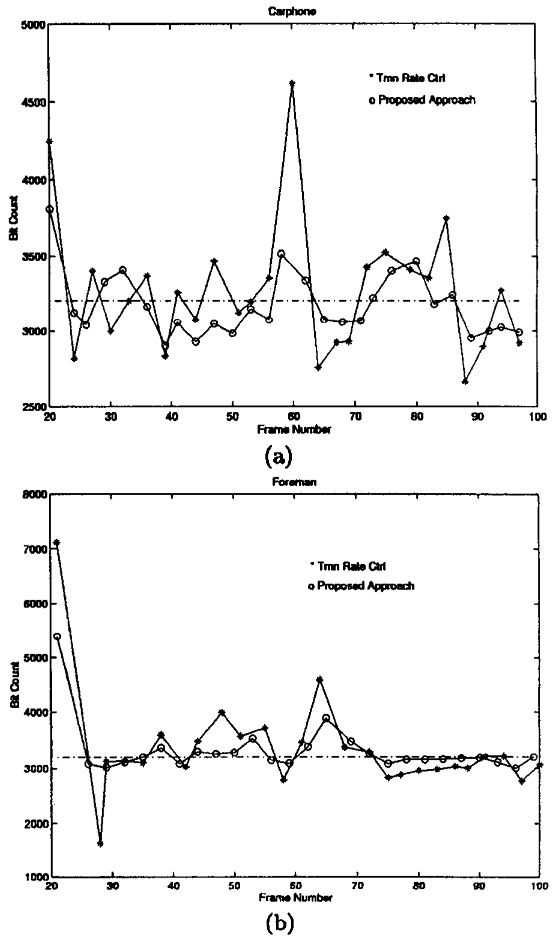

Figure 1: Comparison of Test Model 5 rate control and the proposed scheme for (a) Carphone and (b) Foreman at 32 kilobits per second and 10 frames per second. The dash line is the target number of bit per encoded frame.

6. REFERENCES

[1] ITU Telecom. Standardization Sector of ITU, "Video Coding for Low Bitrate Communication," *ITU-T Recommendation H.263*, March 1996.

[2] Telenor Research. H.263 encoder/decoder tmn1.5 ver. 1.7. *ftp://bonde.nta.no/pub/tmn*,1996.

[3] C. Chen and A. Wong, "A Self-governing Rate-buffer Control Strategy for Pseudo Constant Bit Rate Video Coding," *IEEE Trans. on Image Processing*, Vol. 2, pp. 50-59, Jan. 1993.

[4] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit Allocation for Dependent Quantization with Application to Multiresolution and MPEG Video Coder," *IEEE Trans. on Image Processing*, Vol. 3, No. 5, pp. 533-545, Sep. 1994.

[5] L. Lin and A. Ortega, "Bit-Rate Control Using Piecewise Approximate Rate-Distortion Characteristics," *IEEE Trans. on Circuit and Systems for Video Technology*, Feb. 1998.

[6] K. Fan and K. Kan, "An Active Scene Analysis-Based Approach for Pseudoconstant Bit-Rate Video Coding," *IEEE Trans. on Circuit and Systems for Video Technology*, Vol. 8, No. 2, pp. 159-179, Apr. 1998.

[7] W. Ding and B. Liu, "Rate Control of MPEG Video Coding and Recording by Rate-Quantization Modeling," *IEEE Trans. on Circuit and Systems for Video Technology*, Vol. 6, No. 1, pp. 12-20, Feb. 1996.

[8] K. Yang, A. Jacquin and N. Jayant, "A Normalized Rate-Distortion Model for H.263-compatible Codecs and Its Application to Quantizer Selection," in *Proc. IEEE Int. Conf. on Image Processing*, IEEE, Apr. 1997.

Authorized licensed use limited to: IEEE Xplore. Downloaded on December 22, 2008 at 01:44 from IEEE Xplore. Restrictions apply.