Video Rate Control for Streaming and Local Recording Optimized for Mobile Devices

Mehdi Rezaei, Stephan Wenger, Moncef Gabbouj

Abstract—**In this paper we propose a real-time, low-complexity video rate control algorithm designed to obey buffer constraints. The algorithm is optimized for streaming and local recording applications in mobile devices. Today, most mobile phones include a digital camera that can be used to capture video. The on-phone processor technology has become powerful enough to encode video in real-time. The resulting file can, for example, be archived in the phone's memory, or (progressively or as one block) downloaded, through the 3G mobile network, Bluetooth, or WLAN, to Internet-connected computer systems. From here, all forms of multimedia transmission, such as streaming, file sharing, or multimedia mail become possible.**

In local recording and streaming applications on a mobile phone, we assume that no memory for storage of uncompressed video is available. Therefore, look-ahead and multi-path rate control are not possible. Furthermore, considering the processing power and, more importantly, battery life constraints in mobile devices, the proposed algorithm needs to be as simple as possible.

The described algorithm implements a variable bitrate (VBR) by controlling the quantization scale (QS) on a per picture basis. The QS is calculated based on two other QSs, which correspond to constant rate and constant quality rate controls. The algorithm utilizes the variable bitrate benefits as much as possible so as to minimize the variation of the QS scale, and to provide encoded video with high visual quality. Although it strictly obeys buffering constraints as discussed later, the experimental results show that it allows encoded video at average quality levels significantly higher than reported in earlier works.

***Index Terms*—Control, Rate, Recording, Streaming, Video.**

I. Introduction

Mobile devices with a built-in digital video camera and networks access are becoming popular. In addition to conversational services, where low delay, constant bitrate rate control algorithms are required, local recording and streaming applications (with the mobile device acting as capture and encoder) are also of importance. Consider a mobile phone with a digital video camera, which can capture, encode and record the video in real-time. The modern networks such as 3G mobile phone networks, Bluetooth, or WLAN, allow for the close-to real-time (or real-time) transmission of the encoded video file to a connected PC. Consequently, all forms of multimedia communication such as streaming, file sharing, progressive file download and others become possible.

The rate control constraints for local recording and streaming differ significantly from those for low delay video communication over bandwidth-limited links. In real-time video communication applications, a constant, short-term average bitrate is required to ensure low delay. In contrast, for streaming and local recording applications, a constant long-term average bitrate is sufficient and a significant short-term variation in bitrate is acceptable.

In comparison with constant bitrate video, a variable bitrate video rate control can provide better visual quality and coding efficiency for most video content. A video rate control algorithm can work in different regions in the rate-distortion space between the constant rate region and the constant quality region. Normally, a VBR rate control algorithm operates closer to the constant quality optimum, which results in a better average quality [1]. Minimizing the overall distortion is, in many cases, is roughly equivalent to minimizing the variation in quality [2].

Many real-time variable rate control algorithms have been proposed, specific to local recording [1], [3]-[5] and streaming [6], [7]. The key concept in [1] is to suppress the fluctuation in quantization scale as much as possible. The buffer constraint, which is essential in streaming applications so as to allow for well-defined buffering requirements at the player, is not considered in this algorithm. The methods proposed in [3] and [5] attempt to satisfy a target bit-budget constraint. In other words, they utilize the total storage size as a constraint for encoding a number of frames. Due to the (potentially) open-ended recording session, in our application a target file size is unknown and the algorithm is hence not appropriate. The algorithm used in [4] performs bit allocation according to the coding complexity and does not obey any bit rate constraints. Therefore, depending on the content activity, it produces extreme bit rate variations, which do not obey buffering constraints. Therefore, the algorithm is not suitable for streaming. The algorithm presented in [6] is a low complexity frame-layer bit rate control for streaming video applications. Although this algorithm utilizes a virtual buffer, two other parameters predominantly control its operation: a large time interval and a large bit budget. The virtual buffer, which is essential in streaming, does not play an active role in this

Manuscript received February 28, 2005. Mehdi Rezaei is researcher with the Institute of Signal Processing, Tampere University of Technology (email: mehdi.rezaei@tut.fi). Stephan Wenger is a Principal Scientist with Nokia Research Center, Tampere, Finland (email: Stephan.Wenger@nokia.com). Moncef Gabbouj is Professor and head of Institute of Signal processing, Tampere University of Technology, Finland (email: Moncef.Gabbouj@tut.fi).

algorithm. They assume in case of overflow, the decoder can notify the transmission protocol to stop sending bits. The SPEM (Smooth Pursuit Eye Movement) rate control scheme introduced in [7] is designed for real time encoding and streaming over a constant bit rate channel. This method works near the constant bit rate region, and cannot utilize the variable bit rate benefits. The long window rate control proposed in [8] is a two pass variable rate control for streaming applications. In the first pass they determine the coding complexity of each frame. In this pass also the video sequence is segmented to non-overlapping temporal windows. In optimal case each window is a semantic unit or scene. In the second pass a bit budget is allocated to each window according to target rate and width of the window. In each window a bit budget is allocated to each frame according to the frame complexity, the buffer fullness and allocated bit to the window. This algorithm aims to provide uniform quality for each temporal window, and the quality changes occur only at windows boundaries.

As discussed, the cited papers assume different types of constrains for the two applications. We believe, however, that a unified set of constraints, expressed in a buffer model, is equally appropriate.

In this paper we propose a VBR rate control algorithm with buffer constraint, which operates near the constant quality region in the rate-distortion space. We tested this algorithm on a large number of different video sequences. The experimental results indicate that, while strictly obeying buffering constraints, average quality levels can be obtained that surpass a two-pass variable rate control algorithm [8].

This paper is organized as follow: Section II and III present overview and detailed description, respectively, of the new rate control algorithm. Simulation results are provided in Section IV. A short summary is presented in Section V.

II. Algorithm Overview

The proposed rate control algorithm can be used with all video encoders complying with standards such as H.263, MPEG-4 part 2, and H.264, as well as to non-standard algorithms utilizing similar operation principles. It can support all frame types such as IDR, I, P, and B-frames, but in this paper we consider only Intra frames (I-frame) and Inter frames (P-frame). The control operates on two levels: the frame level and the SOF (Set Of Frame) level.

The SOF is roughly comparable to a Group of Pictures (GOP) as known from the MPEG family of standards, in that it comprises a fixed number of frames and starts with an I-frame. However, within the SOF, the frame types are not fixed and selected based on their rate-distortion properties. When only I-frames and P-frames are allowed, in practice this results in the insertion of an I-frame after a detected scene cut. However, if other frame types were enabled, much more complex SOF structures would be possible. The key reason for introducing the SOF concept in this paper is that it keeps mid-term variations of quality under control, and hence results in a higher average quality [1], [2].

The algorithm utilizes a simple first order Rate-Distortion model and a virtual buffer of relatively large size. The final quantization scales are calculated based on the local coding complexity and the global coding complexity. The local complexity denotes here the coding complexity of the current frame. It is estimated from the results of the encoding of the previous frame. The global complexity is an estimate for the coding complexity of all P-frames in the current SOF, It is estimated from the complexity of all the encoded frames in the video sequence. A global quantization scale (QS) is calculated for all P-frames in each SOF based on the global complexity and buffer fullness. Also a local QS is calculated for each frame based on the local complexity. The global QS and the local QS can be considered corresponding to a QS obtained by constant quality RC and constant rate RC, respectively.

The final QS for variable rate control is calculated based on these two QSs, using a nonlinear modulation and smoothing mechanism in order to minimize the total variation of the quantization scale. It has been shown experimentally that this modulation process improves the visual quality and compression performance significantly. By adjusting the weighting parameters of the modulation process, we can tune the rate controller to a wide range of applications from constant quality to constant rate applications.

III. Details of Algorithm

We summarize the algorithm in three steps: Definitions and Initialization, Bit Allocation, and Quantization Scale Calculation. Although the following descriptions seem to indicate complexity, the algorithm is very lightweight from a computational complexity point-of-view as all operations discussed are exercised only once per SOF or picture.

A. Definitions and Initialization

We use a simple first order R-D model as below:

$$R_j = \frac{S_j}{Q_j} + H_j \ , \qquad (1)$$

where $j=I,P$ denotes Intra or Inter frame. Q_j is quantization scale. R_j, H_j and S_j denote total bits (coefficient bits and header bits), header bits and coding complexity respectively.

Several parameters must be initialized before the start of the coding process. Target bit rate (TR), Target frame rate (FR), Number of frames in one SOF (SOF) and Receiver virtual buffer size (B_S) are defined by the user. Other parameters such as SOF target bit (R_{SOFT}), Current SOF rate (R_{SOF}) and Buffer fullness (B) are initialized by the algorithm.

B. Bit Allocation

The bit allocation is performed once for each SOF and not on a per frame basis. The bit budget for the current SOF is calculated by:

$$R_{SOF} = R_{SOFT}\left[2.52\left(\frac{B}{B_S}\right)^3 - 2.68\left(\frac{B}{B_S}\right)^2 + 1.41\left(\frac{B}{B_S}\right) + 0.59\right], \quad (2)$$

where

$$R_{SOFT} = SOF \times TR \div FR, \quad (3)$$

The bit allocation function in (2) has been designed such that it minimizes the variation of the bit budget by gradually driving the buffer fullness toward a reference point. At the reference point, the SOF target rate and the current SOF rate are equal. Moreover this function has been optimized to allow efficient use of the available buffer space, while it also has a fast response time in very high and low buffer fullness conditions. Moreover, to prevent buffer underflow and overflow, we adjust the SOF rate in extreme buffer conditions as follows:

$$if\,(R_{SOFT} + B - R_{SOF} > MFH \times B_S) \Rightarrow$$
$$R_{SOF} = B - MFH \times B_S + R_{SOFT}, \quad (4)$$
$$if\,(R_{SOFT} + B - R_{SOF} < MFL \times B_S) \Rightarrow$$
$$R_{SOF} = R_{SOFT} + B - MFL \times B_S, \quad (5)$$

where MFH and MFL are margin factors for extreme high and low buffer conditions respectively. The margin factors can be selected by the user according to application. Typical values for MFH and MFL are 0.85 and 0.20 respectively. Furthermore, the maximum and minimum values of current SOF rate are clipped. After calculation of SOF target rate the bit budget for I-frames and P-frames is computed according to:

$$R_P = \frac{R_{SOF} \times R_P^{Avg}}{R_{SOFT}}, \quad (6)$$

$$R_I = \frac{R_{SOF} \times R_P^{Avg} \times X_{IP}}{R_{SOFT}}, \quad (7)$$

where R_P and R_I are allocated bit to P-frames and I-frames in the current SOF respectively. The R_P^{Avg} denotes the average number of bits over all encoded P-frames in the video sequence. X_{IP} is the relative average complexity of I-frames to P-frames. In TMN5 of MPEG-2 rate control algorithms, three similar parameters (for the three frame types) are used for the relative complexities, and they are updated after each coded picture based on the newly gathered statistics. We utilize a similar concept, but with two important differences: First, the optimized value of relative complexities in variable bitrate can differ from the constant bitrate. Second, while there is one I-frame in Intra frequencies or scene cuts followed with a large number of P-frames, related parameters to I-frames are updated slowly and averaging on complexity of I-frames as it is done in [5] is not efficient. So if there is a relatively small number of I-frames in the video sequence, a constant optimized value for relative complexity is enough, otherwise it can be updated during encoding.

C. *Quantization Scales Calculation*

We calculate the QS of P-frame based on the global and local quantization scales. Using the R-D model, the global quantization scale (Q_P^{SOF}) is calculated for a given SOF by:

$$Q_P^{SOF} = \frac{S_P^{Avg}}{R_P - H_P^{Avg}}, \quad (8)$$

where S_P^{Avg} and H_P^{Avg} are the averages of complexity and header bits for all encoded P-frames on the video sequence respectively. Using the R-D model, the local QS is calculated for each frame by:

$$Q_P^L = \frac{S_P}{R_P - H_P}, \quad (9)$$

where S_P and H_P are complexity and header bits of the current P-frame. We use the complexity and header bits of the previous encoded P-frame as estimates for the current frame. To remove unnecessary fluctuations, a simple low pass filtering is performed on the local QS by:

$$Q_P^{LSA}(z) = H(z) \times Q_L^P(z), \quad (10)$$

$$H(z) = \frac{m}{m + 1 - z^{-1}}, \quad (11)$$

where Q_P^{LSA} denotes the filtered version of local QS and $H(z)$ implements the impulse response of low pass filter. m is a constant value larger than one and good results are obtained with m=2.7. A modulation step on the difference between the local and global QSs, after scaling and another low pass filtering, provides an additive term to the global QS so as to build the final QS for a P-frame. The modulation is computed according to

$$Q_P^M = \pm\alpha_S \times Q^{Avg} \times \left[1 - \sec h\left[\sigma_S\left(Q_P^{LSA} - Q_P^{SOF}\right)\right]\right], \quad (12)$$

where Q_P^M is modulated QS and Q^{Avg} is the average value of all QSs and it is used as a scale factor here. α_S and σ_S are two group constants for: $S = 1,2,3,4$. These parameters are extracted experimentally and they are used for four different cases as follows:

Case 1: $(B > 0.57B_S)\,\&\,(Q_P^{LSA} \geq Q_P^{SOF})$,

Case 2: $(B > 0.57B_S)\,\&\,(Q_P^{LSA} < Q_P^{SOF})$,

Case 3: $(B \leq 0.57B_S)\,\&\,(Q_P^{LSA} > Q_P^{SOF})$,

Case 4: $(B \leq 0.57B_S)\,\&\,(Q_P^{LSA} \leq Q_P^{SOF})$,

Considering (2), the Buffer fullness is in the reference point, where $B = 0.57B_S$ or $R_{SOF} = R_{SOFT}$. Figure 1 demonstrates a typical normalized modulation function. The modulation gain is small when difference between the local and global quantization scales is small. It prevents unnecessary variations

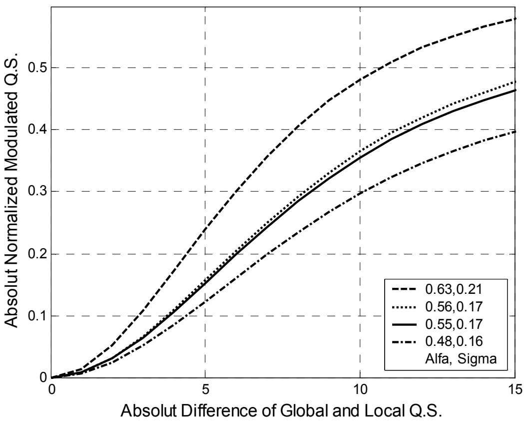

Fig. 1. Normalized modulation function.

in the final QSs. Furthermore, the modulation gain is small when the difference is very high to limit the final QS around the global QS and satisfy the buffer constraint.

We consider four different cases for the modulation function for two reasons. First, the global QS has a long term increase or decrease to derive the buffer fullness toward the reference points; while, the local QS has short term fluctuations around the global QS. It may increase or decrease independent of the global QS. We give a higher weight to the local QS when it varies in the direction of the global QS variation and also we assign a lower weight to the local QS when it varies in the opposite direction to the global QS variation. The second reason is the existence of nonlinearity in the relationship between rate and distortion or QS correspondingly. Considering the simple rate distortion model in (1), we have asymmetric variations in the rate when there are symmetric variations in the QS around a fixed operating point. In other words, the rate can increase easily by changing QS but it needs more changes in QS to decrease the rate at the same increased value. So we implement asymmetric variation in QS to compensate the nonlinearity relationship between the rate and QS and to provide a relatively symmetric local variation around the global bitrate. We give a lower weight to the local QS when it increases the rate compared to the case when it decreases it. Combinations of the above conditions produce four different cases.

To calculate the final QS for P-frames, we first smooth the Q_P^M with a filter equivalent to (11). Thereafter, we calculate

$$Q_P = Q_P^{SOF} + \beta \times Q_P^{MSA}, \tag{13}$$

where Q_P^{MSA} is the filtered version of Q_P^M. We use the β as a tunable factor that provides control over a wide range of applications from constant quality to constant rate. A bigger β steers the algorithm towards a constant rate; whereas, a smaller β results in a performance closer to constant quality. Value of 0.7 yielded good results for local recording and streaming applications. Generally, β can vary from 0.4 to 2 or even out of this range according to the application.

Using the R-D model, the quantization scale for I-frame is calculated by

$$Q_I = \frac{S_I}{R_I - H_I^{Avg}}, \tag{14}$$

where H_I^{Avg} is the average of I-frame header bits and S_I is the complexity of I-frame. The complexity of the I-frame in Intra-frequencies is estimated according to the complexity of the previous P-frame and the value of the relative complexity. The average value can be used for S_I in scene-cuts. After encoding each frame, all related average parameters are updated for future use.

IV. RESULTS

We have implemented the proposed real-time variable rate control algorithm on H.263 and MPEG-4 encoders. A large number (about 20) of video sequences, including those commonly used in standardization, were selected to evaluate the algorithm. We compared the results of our algorithm with those of two other algorithms: H.263 TMN8 in [10] and the Two-Pass rate control algorithm in [8].

Our algorithm performs significantly better than the mentioned anchors, for almost all sequences tested. When comparing to the Two-Pass algorithm, it provides 0.98 dB enhancement in average quality over all video sequences, 42% decrease in real buffer size in streaming and 20% decrease in required delay in streaming. When compared to TMN8, the use of our algorithm results in an 0.42 dB improvement in average quality. While this improvement is not very spectacular, it should be kept in mind that TMN8 drops a large number of frames, about 4%, whereas our algorithm was configured not to drop any frame. Moreover, the average of QS in our algorithm is 0.78 lower than in TMN8. Consequently, we believe that our algorithm performs a visual quality much better than TMN8.

Table 1 lists the results of encoding of ten video sequences with a target bit rate of 64 kb/s, utilizing an H.263 encoder. The rate control algorithms have been compared in terms of average signal to noise ration in luminance frames, average of quantization scales and number of dropped frames. As shown, the proposed algorithm offers a considerable improvement when compared to the TMN8 algorithm.

We have also compared our algorithm with constant QS case, which corresponds to constant quality. Figure 2 plots quality (measured in SNR), QS, and buffer levels over time. We selected the Glasgow sequence as an example, because this sequence has large variations in content activity and contains scene cuts, all making it a difficult target for a rate control algorithm. As one can see, the variations for our rate control in terms of SNR and QS are relatively low. As discussed before, the resulting constant quality is pleasant

TABLE 1: COMPARISON OF THREE RATE CONTROL ALGORITHM BY AVERAGE OF YSNR, AVERAGE OF QUANTIZATION SCALE AND NUMBER OF DROPPED FRAMES IN 64000 BIT/S.

Sequence	RCA	Dropped Frames	QS Mean	Bitrate kb/s	YSNR (dB)
Carphone	TMN8	8	17.41	61.99	30.73
	Proposed	0	15.37	62.26	31.02
Salesman	TMN8	8	6.37	61.49	34.70
	Proposed	0	6.55	61.44	35.70
Hall	TMN8	7	7.25	62.45	35.50
	Proposed	0	7.33	62.66	36.40
Foreman	TMN8	31	19.56	60.92	29.04
	Proposed	0	17.76	60.43	29.41
Container	TMN8	8	7.06	62.79	33.98
	Proposed	0	7.47	62.62	34.57
Silent	TMN8	8	10.44	62.75	32.92
	Proposed	0	9.21	62.55	33.31
Table Tennis	TMN8	7	16.82	61.99	30.11
	Proposed	0	15.71	62.00	30.55
Sailboat	TMN8	15	5.36	62.14	34.94
	Proposed	0	6.22	62.24	35.97
Terevor	TMN8	9	13.85	62.94	32.60
	Proposed	0	12.31	63.27	33.72
Newyork	TMN8	3	16.33	62.67	31.56
	Proposed	0	14.36	62.63	31.90

from a human perception point-of-view. The high correlation in buffer fullness between a constant QS and our algorithm indicates that the heuristics for the bit allocation are a close match to the complexity of the frames. Moreover, the buffer fullness graph shows how well the proposed rate control algorithm uses the available buffer space. In the experiment, a buffer size of 180 kbits has been employed, which represents less than three seconds of the encoded bit stream.

V. SUMMARY AND OUTLOOK

In this paper, we propose a real-time, simple video rate control algorithm which is optimized for streaming and local recording applications in mobile devices. In local recording and streaming applications on a mobile phone, we assume that no memory for storage of uncompressed video is available. Furthermore, the processing power and battery life are considered as constraints in mobile devices.

The described algorithm implements a variable bitrate (VBR) by controlling the quantization scale (QS) on a per picture basis. The QS is calculated based on two other QSs, which correspond to constant rate and constant quality rate controls. This structure provides control over a wide range of applications from constant quality to constant rate.

The algorithm utilizes the variable bitrate benefits as much as possible so as to minimize the variation of the QS scale, and to provide encoded video with high visual quality. Although all of the constraints are obeyed in the algorithm, the experimental results show that it allows encoded video at average quality levels significantly higher than reported in earlier works.

The scalability and error resiliency for streaming in mobile applications can be considered as constraints in design of rate controllers in future research works.

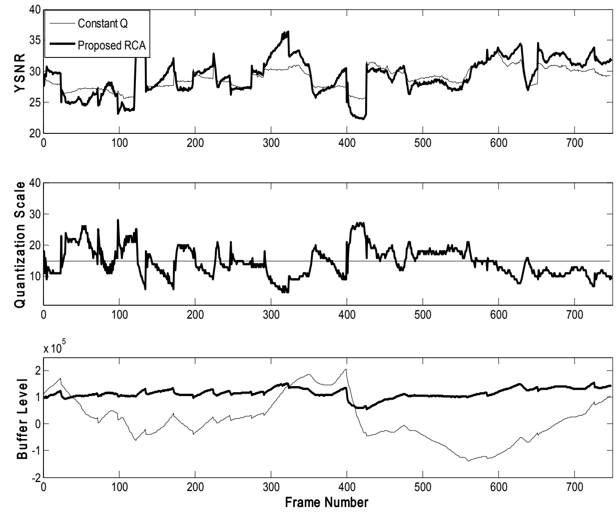

Fig. 2. Comparison of proposed rate control algorithm with constant QS as constant quality case.

REFERENCES

[1] S. Takamura, N. Kobayashi, "MPEG-2 one-pass variable bit rate control algorithm and its LSI implementation," 2002 *IEEE Int. Conf. Image Processing,* pp. 942–945.

[2] Xi Min Zhang, A. Vetro, Y. Q. Shi, Huifang Sun, "Constant Quality Constrained rate Allocation for FGS-Coded Video," *IEEE Trans. Circuits and Systems for Video Technology,* vol. 13, pp. 121–130, Feb. 2003.

[3] Ashish Jagmohan, Krishna Ratakonda, "Mpeg-4 One-Pass VBR Rate Control for Digital Storage," *IEEE Trans. Circuits and Systems for Video Technology,* vol. 13, pp. 447–452, May 2003.

[4] S. Kondo, H. Fukudaa, "A Real-time Variable Bit rate MPEG2 Video Coding Method for Digital Storage Media," *IEEE Trans. Consumer Electronics,* vol. 43, pp. 537–543, Aug. 1997.

[5] Byung Cheol Song, Kang Wook Chun, "A One-Pass Variable Bit-Rate Video Coding for Storage Media," *IEEE Trans. Consumer Electronics,* vol. 49, pp. 689–692, Aug. 2003.

[6] E. C. Reed, F. Dufaux, "Constrained Bit-Rate Control for Very Low Bit-Rate Streaming-Video Applications," *IEEE Trans. Circuits and Systems for Video Technology,* vol. 11, pp. 882–889, July 2001.

[7] A. G. Nguyen, Jenq-Neng Hwang, "SPEM online rate control for real time streaming video," 2002 *IEEE Int. Conf. Information Technology ,* pp. 65–70.

[8] V. Varsa, M. Karczewicz, "Long Window Rate Control for Video Streaming," 2001 *11th International Packet Video Workshop,* pp. 154–159.

[9] ITU-T, "Video Codec Test model near-term, Version 5 (TMN5)", H.263 Ad Hoc Group, 1995.

[10] ITU-T, "Video Codec Test Model, Near-Term, Version 8 (TMN8)", Video Coding Experts Group, 1997.