MPEG VIDEO BITRATE CONTROL USING CODER MODELING

Adam Łuczak, Marek Domański

Poznań University of Technology, Institute of Electronics and Telecommunication, Poznań, Poland
E-mail: { domanski, aluczak } @ et.put.poznan.pl

ABSTRACT

The problem considered is how to estimate the value of the quantizer scale factor Q that corresponds to a given number of bits B in a given frame of a sequence. The solution exploits two statistical models B(Q) of I-, P- and B-frames coded by an MPEG coder. These models B(Q) depend on some parameters related to video content. In the first model, there are very few such parameters that can to be estimated from a previous similar frame. In the second model, the content parameters are the AC coefficients of DCT. Accurate choice of the value of the quantizer scale factor Q can be made using these two models. The experimental results have been presented for MPEG-2 MP@ML coders but some other experiments show that similar approach is useful for other hybrid coders like H.263 and MPEG-4.

1. INTRODUCTION

In video coders, bitrate control remains a challenging problem with great importance for wide applications. The difficulties are related to the fact that there does not exist an universal quantitative mathematical model that allows for exact calculations of coder parameters from given bitrate and video quality. Recently, the problem became even more severe because of emerging video communication in networks with varying available throughput.

The basic parameter that can be used to control an encoder is the quantizer scale factor Q that influences the quantization of the DCT coefficients. A typical goal of adjusting the parameter Q is to match the available channel bitrate. The quantizer scale factor Q may be adjusted on frame level, slice level as well as on macroblock level. Here, we consider adjustment on the frame and slice levels only. This problem has been already considered by several authors [1-6]. Some approaches suffer from luck of models appropriate for exact and noniterative bitrate control.

The paper describes simple empirical models of bitstreams produced by MPEG-2 [7,8] video coders. The approximate numbers of bits that represent individual frames are expressed as functions of the quantizer scale factor Q. These models are used for global control of a video coder, i.e., for adjusting the value of the quantization scale parameter Q in frame headers. Simulation results prove that application of the model allows efficient bitrate control of a video coder both in the scenario of constant bitrate as well as for available bitrate rapidly varying due to network conditions.

The experimental results have been presented for the MPEG-2 MP@ML [7,8] coders implemented using standard software [9]. Some other experiments show that similar approach is useful for other hybrid video coders like H.263 and MPEG-4.

2. BITRATE CONTROL PROBLEM

The problem considered is how to estimate the value of the quantizer scale factor Q that corresponds to a given number of bits B in a given frame of a sequence encoded by an MPEG-2 MP@ML coder. A solution would need a coder model $B(Q)$ to be known. Such a model has to be defined individually for I-, P- and B-frames. Moreover, all three models have to use some parameters that express the influence of content.

A number of bits B allocated to an individual frame is a sum of the component B_{CONST} that does not depend from the quantizer scale factor Q and the component $B_{\mathrm{VAR}}(Q)$ that depends on Q [10], i.e.

$$B = B_{\mathrm{CONST}} + B_{\mathrm{VAR}}(Q) \ . \qquad (1)$$

The uncontrollable component, i.e. the value B_{CONST} is the number of bits needed for headers, intra-DC coefficients and motion vectors. The value B_{CONST} depends on frame type (I, P or B) and frame content [10]. As the coding mode (*intra-* or *inter*) is chosen independently from the current value of the quantizer scale factor Q, the value B_{CONST} can be calculated during the first stage of the frame encoding process, i.e. during those coding operations that do not depend on the quantizer scale factor Q.

Therefore the problem can be redefined as follows. For a given bit number B for a current frame, let us calculate the factor Q such that

$$B_{\mathrm{VAR}}(Q) = B - B_{\mathrm{CONST}}\,. \tag{2}$$

Such an approach needs a simple empirical model for the bitrate B_{VAR} as function of Q set for individual I-, P- and B-frames. The model should be valid for typical conditions of operation of the coder. For the sake of simplicity, absence of adjustment of Q in the individual macroblocks is assumed.

2. GLOBAL MODEL

The global model $B_{\mathrm{VAR}}(Q)$ has been already described in [10]. Here, we report briefly this model that is expressed as a sum

$$B_{\mathrm{VAR}}(Q) = B_{\mathrm{YV}}(Q) + B_{\mathrm{CV}}(Q)\,, \tag{3}$$

where $B_{\mathrm{YV}}(Q)$ and $B_{\mathrm{CV}}(Q)$ denote the bits needed for encoding of the DCT coefficients (except the *Intra*-DC ones) in the luminance and the chrominance, respectively.

In order to match experimental data the following global model has been chosen for $B_{\mathrm{YV}}(Q)$ and $B_{\mathrm{CV}}(Q)$

$$B_x(Q) = \frac{a}{c(Q^b + e) + d}\ , \tag{4}$$

where a, b, c, d and e are real constants that depend on the frame coding mode and frame content and $B_x(Q)$ stands either for $B_{\mathrm{YV}}(Q)$ or for $B_{\mathrm{CV}}(Q)$ (for both chrominance components together).

The accuracy of the model has been tested for several test sequences and for I-, P- and B-frames [10]. The parameter values can be estimated by minimization of approximation error over the set of the allowed values of Q. The model with such parameters accurately approximates the experimental data (Fig. 1).

The content-dependent parameters a, b, c, d, e may be estimated in a process of some trials of encoding. Nevertheless, for bitrate control, such a technique would not be very useful. Fewer parameters depend on frame content, the more useful is the model for coder control. The set of five parameters can be successfully reduced to two or even one content-dependent parameter. Both cases have been studied and modeling accuracy was estimated for both cases.

For example, it can be assumed that only two parameters a and c depend on frame content, while the other three parameters b, d and e are fixed to general values that can be estimated by minimization of the approximation error ΔB over a set of frames from some set of training sequences. In the experiments, the minimization was performed independently for luminance and chrominance for 9 frames from different portions of each of the training sequences: *Basket*, *Bus* and *Mobile and Calendar*. Similar procedure lead to general values of four fixed parameters a, b, d and e when only one free parameter c is assumed.

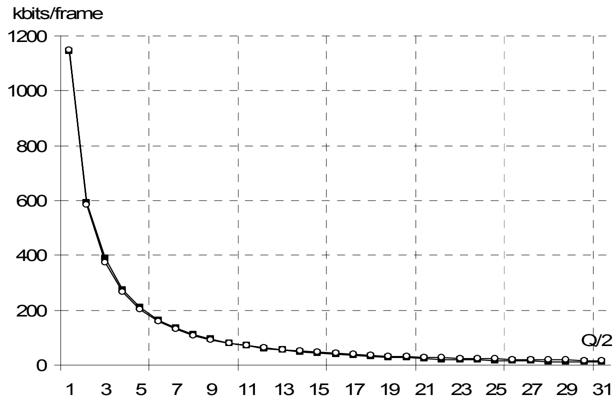


Fig. 1. Exemplary comparison of the experimental and modeled curves $B_{\mathrm{YV}}(Q)$ for an P-frame from the test sequence *Basket.*

In the latter case, the only free parameter c depends on frame content. Its value can be estimated from one coding experiments resulting in estimation of a point (Q_0, B_0) on a curve $B_x(Q)$ With fixed values of a, b, d and e and only one free parameter c, the precision of approximation is still quit good (Table 1). The average approximation error ΔB_{ave} and the maximum error ΔB_{max} have been calculated for a set of other test sequences different from that used to estimate the fixed parameters. The values of the approximation errors for the test sequences *Flower Garden, Cheer*, *Funfair* have been calculated relative to the total number of bits per frame (Table 1).

Moderate accuracy of the model can be accepted as only integer values of Q may be chosen.

In between of scene cuts, the optimal values of the parameters a and c remain almost constant and the values from a previous frame can be used

Table 1. Typical bounds for average error ΔB_{ave} and maximum error ΔB_{max} of bitrate measured for test sequences.

Free parameters	Approximation error [%]			
		I-frame	P-frame	B-frame
5 parameters: *a, b, c, d, e*	ΔB_{max}	3.6	6.0	0.5
	ΔB_{ave}	2.0	3.0	0.3
2 parameters: *a, c*	ΔB_{max}	8.5	10.0	1.0
	ΔB_{ave}	4.0	5.0	0.7
1 parameter: *c*	ΔB_{max}	9.0	11.0	1.1
	ΔB_{ave}	4.0	6.0	0.8

4. MICROSCOPIC MODEL

The previous model was global in that sense that the number of bits, needed to represent DCT coefficients for luminance or chrominance in a frame, was expressed as a simple function of Q with one or two parameters depending on content. Here, another model is introduced. The number of bits $B_{\text{VAR}}(Q)$ per frame is estimated from the histograms of the DCT coefficients (excluding *Intra* DC coefficients). These DCT coefficients are known before the quantization parameter Q has to be chosen.

For each component, i.e. the luminance and two chrominances, a two-dimensional histogram $H_{ij}(|F_{ij}|)$ is calculated for each DCT coefficient F_{ij} rounded the closest integer. In this way, 64 histograms are calculated for each component in a P- or B-frame while 63 histograms are calculated for components from an I-frame. An $H_{ij}(|F_{ij}|)$ histogram expresses a probability distribution of values of the F_{ij} coefficient within a given frame and color component.

In the MPEG-2 bitstream, the nonzero DCT coefficients are encoded using Huffman variable length codes. The codes and their lengths are set in the MPEG-2 standard.

The number of bits for DCT coefficients (excluding *Intra*-DC ones) plus the number of bits needed for the EOB codes (codes for *End of block*) is equal to $B_{\text{YV}}(Q)$ or $B_{\text{CV}}(Q)$ for luminance and chrominance, respectively. The number of DCT-codewords is equal to the number of nonzero DCT coefficients. The latter one strongly depends on quantization factor Q. Therefore the values of $B_{\text{YV}}(Q)$ and $B_{\text{CV}}(Q)$ can be modeled statistically as functions of Q. The reason for that is that a BT.601/4CIF frame contains a sufficiently large set of blocks.

After quantization with a dead zone around zero, the coefficient F_{ij} remains nonzero if

$$F_{ij} \geq T_{ij} \quad \text{where threshold } T_{ij} = Q \cdot W_{ij} / 16 \qquad (5)$$

and W_{ij} is the (i,j)-th element of the respective quantization table. Number of bits $B_{\text{VAR}}(Q)$ can be estimated as

$$B_{\text{VAR}}(Q) \approx \cdot \sum_{i,j} B_{i,j}(Q), \qquad (6)$$

where summation is performed over DCT coefficients F_{ij} that exceed respective thresholds T_{ij} (Fig. 2).

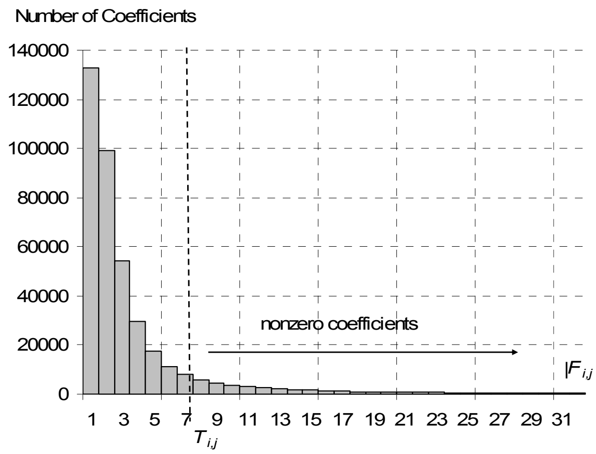


Fig. 2. A histogram $H_{ij}(|F_{ij}|)$ for luminance in the test sequence *Basket* (4CIF sequence).

Fortunately, the model can be transformed into a more accurate version. Here, we explain the version for an MPEG-2 MP@ML encoder. The improvement is based on observation of the properties of the probability density $p(r,l)$ of pairs (r,l) = (*run*, *level*). The most of significant values of $p(r,l)$ are along $r = 0$ and $l = 0$ axes. Therefore the number of bits in a color component (luminance or chrominance) can be estimated as

$$B_{\text{ij}}(Q) \approx \sum_{l=|F_{ij}|=T_{ij}+1}^{40} C_l \cdot H_{ij}\left(\left|F_{ij}\right|\right) + 24 \cdot \sum_{|F_{ij}|=T_{i,j}+41}^{2048} H_{i,j}\left(\left|F_{ij}\right|\right) + 2 \cdot N_B , \qquad (8)$$

where the values of constants C_l for l>4 are close to the code lengths for (r,l)=(0,l) and N_B stands for the number of coded blocks. The first term expresses bits for nonzero DCT coefficients encoded by Huffman codes, the second term stands for bits related to *Escape* codes and respective numerical values of r and l, the third term expresses bits for the *End of block* codes. The values of C_l for l<5 have been estimated from probability distributions for pairs *(r,l)* (Fig.3). The values of constant C_l are listed in Table 2.

Table 2. Values of the C_l parameter.

l	C_l	l	C_l	l	C_l	l	C_l
1	4.5	11	13.2	21	15.0	31	16.0
2	6.0	12	14.1	22	15.0	32	16.0
3	6.7	13	14.1	23	15.0	33	16.0
4	8.5	14	14.1	24	15.0	34	16.0
5	9.5	15	15.0	25	15.0	35	16.0
6	9.5	16	15.0	26	15.0	36	16.0
7	11.5	17	15.0	27	15.0	37	16.0
8	13.2	18	15.0	28	15.0	38	16.0
9	13.2	19	15.0	29	15.0	39	16.0
10	13.2	20	15.0	30	15.0	40	16.0

The model (8) is quite accurate in the whole range of Q, i.e. <1,112>. The approximation error ΔB was always below 3% in the whole range of Q. The values of constants C_l have been estimated for a set of training video sequences for the encoder obtained from [9] with default quantization tables and the first set of Huffman codes.

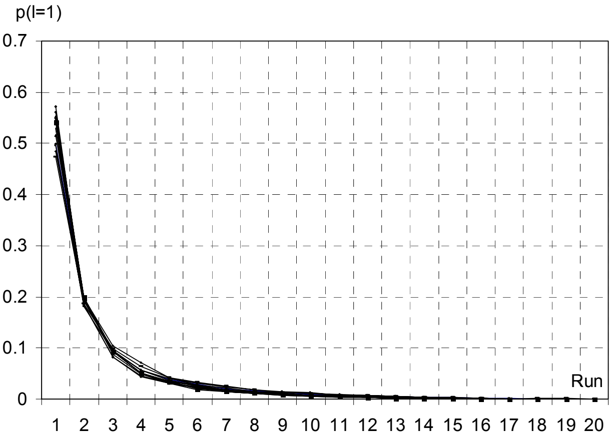


Fig. 3. Conditional probability function *p(r,l|l=1)* for luminance (Q = 2).

Note that there exits two sets of model parameters C_l: one for I-frames and the second for P- and B-frames. This model is applicable for whole frames as well as for individual slices.

As shown in Figure 4, it is necessary to estimate the C_l parameter only for *level*=1, 2, 3 and 4. In other cases parameter C_l is constant. Gray fields marks pairs *(r,l)* which are encoded using Huffman codes, the other are encoded with *Escape* codes.

	Level								
Run	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0	10450	4647	3336	2171	1474	1155	852	792	512
1	3820	1097	556	261	170	102	69	58	31
2	1959	418	191	63	39	16	15	17	6
3	1192	213	66	20	14	7	5	6	2
4	850	122	40	8	3	4	3	1	0
5	666	104	36	4	7	3	1	1	1
6	511	78	13	15	3	3	3	1	0
7	386	60	19	7	2	1	0	0	0
8	319	57	6	1	2	0	1	0	0
9	252	31	10	2	2	0	0	0	0
10	158	41	6	2	1	0	1	0	0
11	162	23	5	0	0	0	0	0	0
12	117	15	0	0	0	0	0	0	0
13	66	4	0	0	0	0	0	0	0
14	55	6	0	0	0	0	0	0	0
15	26	5	0	0	0	0	0	0	0
16	14	2	0	0	0	0	0	0	0
17	12	3	0	0	0	0	0	0	0
18	16	4	0	0	0	0	0	0	0
19	13	3	0	0	0	0	0	0	0
20	12	5	0	0	0	0	0	0	0

Fig. 4. An exemplary histogram of (r,l) pairs for luminance in the test sequence *Basket* (4CIF sequence) for Q=2.

The values of constants C_l have been estimated from the experimental data for 12 frames taken out from six sequences. This set contains well known sequences *Basket, Cheer, Bus* and sequences from DVD: *Icon, Universal* and *Warner*.

5. BITRATE CONTROL

Two types of models have been proposed by the authors:

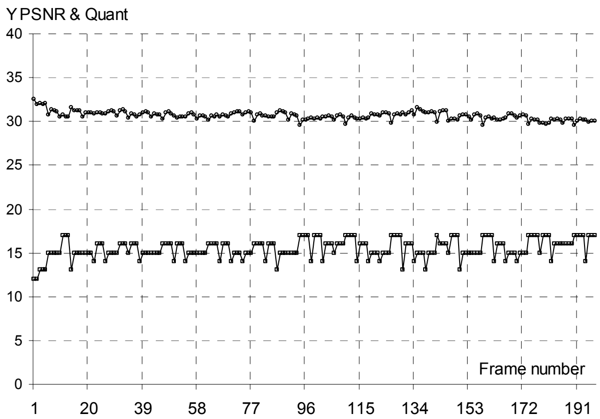
- The global model with one free parameter that depend on frame content.
- The microscopic model with parameters that can be estimated from the histograms $H_{ij}(|F_{ij}|)$.

Application of the global model needs extremely few computations and is efficient when consecutive frames are similar (no rapid scene change, the same frame coding mode, similar value of Q). The only free parameter c can be estimated from the data gathered during encoding of the previously encoded frame, i.e. previous P- or B-frame, respectively to the type of the currently encoded frame. The values are also obtainable from the second model that is more accurate in the wide range of Q and does not exploit the data from previous frame. This model is applicable even for rapid changes of Q and after scene cuts. The models may be combined, e.g. more exact microscopic model (8) may be applied to the luminance component while the less accurate and simpler global model may be applied to the smaller bitrates of chrominance components that may be estimated with less accuracy.

The models $B(Q)$ may be used for bitrate control. A value of B is calculated from the available bitrate. The value of Q may be set using the models described above. Application of this technique leads to stable quality of video sequence (Fig.5) and buffer occupancy (Fig. 6).

The technique using microscopic models is applicable even for rapid changes of bitrates and after scene cuts.

a)



b)

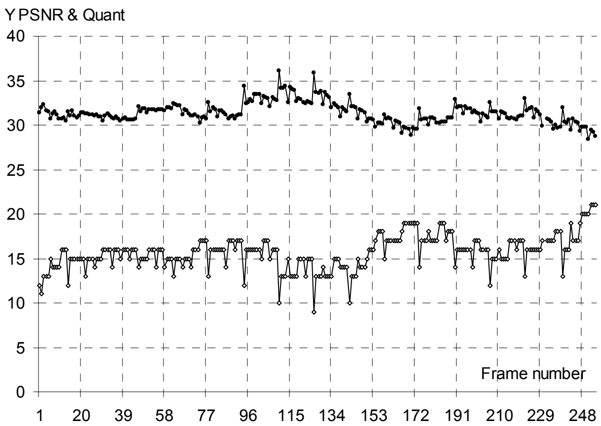


Fig. 5. Luminance PSNR (upper curve) and quantizer scale Q (lower curve) for the *Flower Garden* (a) and *Basket* (b) video test sequences encoded at about 5 Mbitps.

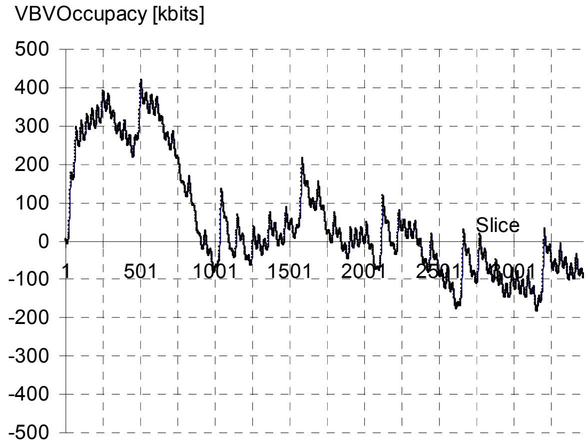


Fig. 6. Buffer occupancy for the *Basket* video test sequence encoded at about 5 Mbitps (0 – means half of the buffer).

6. CONCLUSIONS

Two coder models $B(Q)$ have been described. These two models can be used for setting a value of the quantizer scale factor Q for a given number of bits for a frame. The microscopic model can be also used to set a value of the quantizer scale factor Q for individual slices.

The models are easy to use and their application leads to stable quality of a sequence. The bitrate is mostly very stable with no overshoots. Therefore the prospective applications include low-delay video bitrate control as the coder can safely work with small buffer.

ACKNOWLEDGEMENT

The work has been supported by the Polish National Committee for Scientific Research under Grant 4 T11D 00722

REFERENCES

[1] T. Kim, B. Roh, J. Kim, “An accurate bit-rate control for real-time MPEG video encoder,” *Signal Processing: Image Communication,* vol. 15, pp. 479-492, 2000.

[2] L. Wang “Rate control for MPEG video coding”, *Signal Processing: Image Communication* 15, pp. 493-511, 2000.

[3] Y. Saw, M. Grant, J. Hannh, B. Mulgrew “Video rate control using a radial basis function estimator for constant bit-rate MPEG coders”, *Signal Processing: Image Communication* 13, pages 183-199, 1998.

[4] J. Ribas-Corbera, S. Lei, “Rate control in DCT video coding for low-delay communications”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp.172-185, 1999.

[5] Y. Yu, J. Zhou, Y. Wang, Ch. Chen, “A novel two-pass VBR coding algorithm for fixed-size storage application”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp.345-356, 2001.

[6] W. Ding, B. Liu, rate control of MPEG-2 video coding and recording by rate-quantization modeling,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 12-20, June 1996.

[7] ISO/IEC IS 13818, “Information Technology - Generic Coding of Moving Pictures and Associated Audio Information”.

[8] B. Haskell, A. Puri, A. N. Netravali, *Digital Video: An Introduction to MPEG-2*, New York: Chapman & Hall, 1997.

[9] MPEG Software Simulation Group, MPEG-2 Encoder Eecoder version 1.2, July 19, 1996, ftp://ftp.mpeg.org/pub/mpeg/mssg.

[10] M. Domański, A. Łuczak "Simple Global Model of an MPEG-2 Bitstream" Proc. IEEE Int. Conf. Multimedia and Expo ICME'2002, August 2002, Lausanne, Switzerland, vol. I, pp. 645-648