HYBRID VIDEO CODER MODELING USING ANALYSIS OF TRANSFORM COEFFICIENTS

Adam Łuczak, Marek Domański

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

ABSTRACT

The paper describes a model *B(Q)* of bitrate *B* as a function of the quantizer scale factor *Q*. The bitrate is calculated from the histograms of the AC coefficients of DCT. These coefficients may be known before the decision upon the value of *Q* has to be made. The model can be used to estimate the value of the quantizer scale factor *Q* that corresponds to an assumed number of bits B in a given frame of a sequence. The experimental results have been presented for MPEG-2 MP@ML coders. These results indicate that application of the proposed model improves control quality as compared to the widely used TM5 algorithm. Experimental results prove that the model can be used within control algorithms for video coders (not necessarily MPEG-2).

1. INTRODUCTION

Despite of millions of hybrid video coders working worldwide, defining efficient control algorithms is still an open problem that gains a lot of attention. Recently, the problem became even more severe because of emerging video communication in networks with varying available throughput. One of the major problems is related to lack of an universal quantitative mathematical model that allows for exact calculations of coder parameters from assumed bitrate and video quality.

In a hybrid video coder, bitrate is controlled by setting the quantizer scale factor *Q* that scales the quantization step for the DCT coefficients. A typical goal of adjusting the parameter *Q* is to match the available channel bitrate. Therefore modeling of bitrate *B* as a function of *Q* is an important but somewhat difficult problem as the function *B*(*Q*) strongly depends on video content. This problem is related to rate-distortion modeling for video coders, and has been already considered by several authors [1-8]. Several various approaches have been proposed including general "black box" purely statistical modeling [e.g. 14] and picture feature analysis. For the sake of brevity, the review of these solutions must be left beyond the scope of this paper. Nevertheless none of those references proposes a model of a hybrid coder similar to that proposed in this paper. Here, the model exploits analysis in the frequency domain.

This paper describes empirical model of bitstreams produced by MPEG-2 [9,10] video coders. The approximate numbers of bits that represent individual frames are expressed as functions of the quantizer scale factor *Q*. This model is used for global control of a video coder, i.e., for adjusting the value of the quantization scale parameter *Q* in frame and slice headers. The model may be used for the constant bitrate (CBR) mode of coder operation, i.e. with the goal of keeping constant bitrate of the bitstream fed into a communication channel. Moreover, the model can be helpful in the variable rate (VBR) mode as well.

The experimental results have been presented for the MPEG-2 MP@ML [9,10] coders implemented using standard software [11]. Some other experiments show that similar approach is useful for other hybrid video coders like H.263. [12] After some modifications, the model may be adopted even for AVC/H.264 [13] codecs but this problem will be left out of the scope of this paper.

2. BASIC BITSTREAM COMPONENTS

The number of bits *B* allocated to an individual frame is a sum of the component B_{CONST} that does not essentially depend on the quantizer scale factor *Q* and the component $B_{\mathrm{VAR}}(Q)$ that depends on *Q*,

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

The B_{CONST} part consists of the following items:

$$B_{CONST} = B_{CTRL} + B_{\mathrm{YDC}} + B_{\mathrm{CDC}} + B_{\mathrm{MV}} \tag{2}$$

where B_{CTRL} – number of bits needed for headers of pictures, slices and macroblocks, B_{YDC} and B_{CDC} – number of bits that represent *Intra DC* coefficients, B_{MV} – bits needed for motion vectors. In fact, bitstream B_{CTRL} depends slightly on *Q* as quantization may influence macroblock type. Nevertheless this influence to the total bitrate *B* is negligible and the number of bits B_{CTRL} can be estimated with sufficient accuracy. As the coding mode (Intra or Inter) is chosen independently from the current value of the quantizer scale factor *Q*, the values B_{YDC}, B_{CDC} and B_{MV} can be calculated exactly 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*. In this way, the value B_{CONST} can be estimated quit exactly before setting the value of quantizer scale factor *Q*.

3. CONTROLABLE PART OF THE BITSTREAM

Let us consider the bitstream component that directly depends on the frame quantization factor *Q*:

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

where $B_{YV}(Q)$ and $B_{CV}(Q)$ denote the bits needed for encoding of the DCT coefficients (except the Intra DC ones) for luminance and chrominance, respectively. The $B_{CBP}(Q)$ is the number of bits needed to encode *CodedBlockPatern* (CBP) field included into macroblock syntax. This field occurs only in P- and B-macroblocks and is encoded using VLC codes. The CBP code is known after quantization of DCT coefficients, i.e. after quantization scale factor Q is chosen. Therefore, the value of the $B_{CBP}(Q)$ has to be estimated using a bitrate model. For the CBP field, its length can be approximated by a linear function of Q.

The shortest VLC code for the CBP consists of 3 bits, and the longest has 9 bits, in MPEG-2 standard. As the value of Q increases, more blocks are skipped out. Here, the proposed experimental model for $B_{CBP}(Q)$ is :

$$B_{CBP} = MN \cdot C_{FT} \cdot C_{max} \cdot \left[1 - \left(\frac{Q}{Q_{MAX}} \right) \right], \tag{4}$$

where

$$C_{FT} = \begin{cases} 0 & \text{for I} - \text{macroblocks}, \\ 0.5 & \text{for P} - \text{macroblocks}, \\ 1 & \text{for B} - \text{macroblocks}, \end{cases}$$

and MN is number of macroblocks in a frame, C_{max} is the length of the longest CBP Huffman codeword, Q_{MAX} denotes the maximum value allowed for the quantizer scale factor Q (e.g. 62 for the first mode of MPEG-2).

Experiments prove that accuracy of this approximation is sufficient for estimation of the total bitrate. The approximation error is always less then 0.5% due to the fact that the number of bits needed for encoding the CBP field is a small portion of the whole bitstream.

Let us consider the $B_{YV}(Q$ and $B_{CV}(Q)$ components of $B_{VAR}(Q)$ (see Eq. (3)).

The new approach is that the numbers of bits $B_{YV}(Q)$, $B_{CV}(Q)$ per frame are estimated from the histograms of the DCT coefficients. Of course, Intra DC coefficients are excluded from this analysis as they were already included into the term B_{CONST} that is not controlled by factor Q. Since DCT computation is the very first stage of the transform coding process, the DCT coefficients are known before the quantization scale factor Q is chosen. The frequency domain statistics may be used in order to estimate the function $B_{VAR}(Q)$, and than to estimate Q for a given available number of bits.

For each coefficient, the respective quantization weight w_{ij} together with the quantizer scale factor Q define exactly the quantizer. Quantization of the DCT coefficient F_{ij} may be viewed as an operation $F_{ij} \to l_{ij}$, where l_{ij} is quantized coefficient. In particular, a weight w_{ij} and factor Q define a threshold

$$T_{ij}(Q) = \left\lceil \frac{w_{ij} \cdot 4 \cdot Q - w_{ij} \cdot \left\lfloor \frac{(3 \cdot Q + 2)}{4} \right\rfloor - \frac{w_{ij}}{2}}{32} \right\rceil, \tag{5}$$

such that $l_{ij} = 0 \quad \text{if} \quad |F_{ij}| < T_{ij}$. (6)

Only nonzero quantized values l_{ij} are encoded. In order to estimate number of bits needed to encode the quantized transform coefficients, histograms $H_{ij}(|F_{ij}|)$ of not quantized transform coefficients are calculated.

There are 64 transform coefficients per (8×8)-block. From all luminance blocks in a picture, 64 histograms $H_{ij}(|F_{ij}|)$ are calculated for each DCT coefficient, respectively. Similarly, two other sets of 64 histograms $H_{ij}(|F_{ij}|)$ are calculated for both chrominance components.

For a given value of Q and certain coefficient F_{ij}, from a histogram $H_{ij}(|F_{ij}|)$, the number of nonzero coefficients l_{ij} can be calculated (Fig. 1). Moreover, using such a histogram, the numbers of nonzero coefficients with particular values l_{ij} = 1, 2, 3 etc. may be estimated as well.

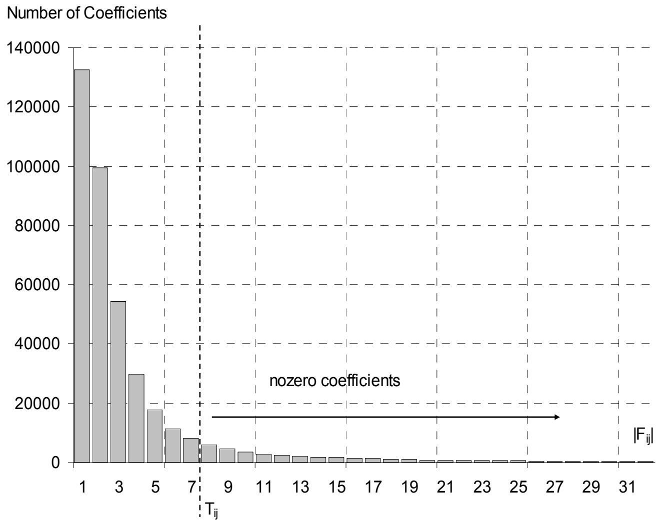

Fig. 1. Histogram of one AC DCT coefficient computed for the first frame of *Flower* test sequence (4CIF resolution).

In a video coder, the quantized DCT coefficients are scanned and coded as (r,l)-pairs, where r is the number of zero-valued coefficients preceding a nonzero coefficient with value l. The (r,l)-pairs are coded using Huffman codes. In MPEG-2 standard, a codelength is independent from sign of l.

The Huffman codes are not defined for all the (r,l)-pairs. Pairs with no defined Huffman code are encoded using 24-bit constant-length *ESCAPE* coding. Probability for such codes is mostly much lower than for Huffman codes (Fig. 2).

The number of bits needed for encoding luminance component is 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*).

$$B_{YV}(Q) \approx \cdot \sum_{i,j} B_{ij}(Q) + 2 \cdot N_B \,, \tag{7}$$

where N_B stands for the number of coded blocks and expresses the number of bits for the *End of block* codes. The number of bits needed for (i,j)-th transform coefficient is given by

$$B_{ij}(Q) \approx \sum_{l=|F_{ij}|=T_{ij}+1}^{40} C_l \cdot H_{ij}(|F_{ij}|) + 24 \cdot \sum_{|F_{ij}|=T_{i,j}+Esc}^{2048} H_{i,j}(|F_{ij}|) \tag{8}$$

where T_{ij} denotes threshold for (i,j)-th histogram, Esc denotes such value of l that no Huffman code is defined for $l \geq Esc$ (Esc=41 for MPEG-2), and C_l is the average codelength for nonzero coefficient with quantized value l.

The number of bits for chrominance $B_{CV}(Q)$ is to be estimated in the same way.

Run	Level 1	2	3	4	5	6	7	8	9	10
0	70614	26910	14658	12215	8314	6035	4617	3834	3809	2931
1	13160	2295	742	452	224	141	104	64	61	31
2	4676	385	81	32	14	12	4	3	2	3
3	2206	88	7	1	1	3	1	0	0	0
4	1229	30	10	2	0	0	0	0	0	0
5	660	15	7	0	0	3	0	0	0	0
6	333	4	1	0	0	1	1	0	0	0
7	276	3	0	0	0	13	5	1	0	0
8	158	2	0	0	0	0	0	0	0	0
9	80	0	0	0	0	0	0	0	0	0
10	52	0	0	0	0	0	0	0	0	0
11	43	9	0	0	0	0	0	0	0	0
12	10	0	0	0	0	0	0	0	0	0
13	6	0	0	0	0	0	0	0	0	0
14	9	0	0	0	0	0	0	0	0	0
15	4	0	0	0	0	0	0	0	0	0
16	3	0	0	0	0	0	0	0	0	0
17	1	0	0	0	0	0	0	0	0	0
18	1	0	0	0	0	0	0	0	0	0
19	2	0	0	0	0	0	0	0	0	0

Fig. 2 An exemplary histogram of (r,l)-pairs for luminance of first frame in the test sequence *Basket* (4CIF resolution) for Intra quantization mode for Q=8 (MPEG-2 system). Huffman codes are defined for shaded (r,l)-pairs only.

The proposed model is based on observation of the properties of the probability $p(r,l)$ of occurrence of (r,l) of pairs. The majority of the most significant values of $p(r,l)$ are along $r = 0$ and $l = 0$ axes (Fig. 2). Therefore the C_l values for l>4 (Fig.2) are close to the codelengths for the $(0,l)$-pairs. The values of C_l for l≤4 have been estimated using probability distributions for pairs (r,l). These distributions are relatively stable for various pictures (Fig. 3).

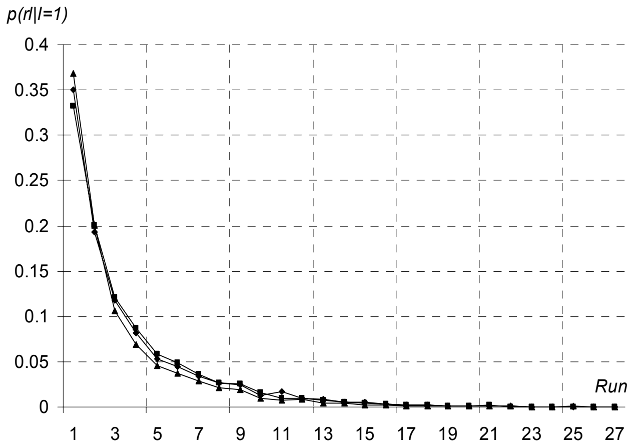

Fig. 3. Probability distributions of pairs (r,l) for various I-frames of sequence *Basket* for l=1, (MPEG-2 system).

The values of constants C_l have been estimated for a set of training video sequences and for MPEG-2 encoder with default quantization matrices and with the first set of Huffman codes. For P- and B-frames, the values of C_l are the following:

$l = 1$, $C_l =$ 5.0 ,
$l = 2$, $C_l =$ 6.3 ,
$l = 3$, $C_l =$ 6.8 .

The numbers of bits for $l > 3$ are given in Table 1. These values are similar for all picture types.

Therefore, for each value of Q, the thresholds T_{ij} may be estimated and then the respective numbers of bits can be calculated using the histograms $H_{ij}(|F_{ij}|)$ and Eqs. (4), (7) and (8). The calculations are to be repeated for both chrominance components as well.

Table 1. The average codelengths C_l for all $l < Esc$ for I-frames within an MPEG-2 codec.

l	C_l	l	C_l	l	C_l	l	C_l
1	4.0	11	13.2	21	15.0	31	16.0
2	5.6	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

4. ACCURACY OF THE MODEL

The obtained C_l values have been used for checking estimation accuracy of the bitstream value $B_{VAR}(Q)$. The estimation error is defined as follows:

$$\varepsilon B(Q) = \frac{|B_e(Q) - B_x(Q)|}{B_x(Q)} \cdot 100\%, \tag{9}$$

where $B_x(Q)$ is the measured value of bitstream resulting from encoding of DCT coefficients (excluding Intra DC), and $B_e(Q)$ is the estimated value of bitstream.

Table 2 Maximum and average estimation error computed for exemplary three sequences (Basket, Cheer and Warner) for three values of quantization scale factor Q.

Frame	*Q=16*		Q=32		Q=48	
	aver.	*max.*	*aver.*	*max.*	*aver.*	*max.*
sequence Basket						
I	1.21	7.98	1.03	8.19	1.12	8.78
P	1.84	6.67	1.41	7.53	1.52	7.02
B	2.01	6.45	1.86	7.69	1.92	8.59
sequence Cheer						
I	1.89	6.65	1.72	5.18	1.60	7.25
P	2.21	8.16	2.56	6.54	2.01	6.72
B	2.64	8.51	2.98	7.93	2.32	8.83
sequence Warner						
I	2.31	5.47	2.29	4.83	1.70	8.92
P	2.98	6.52	2.64	5.22	1.89	9.45
B	3.34	6.10	3.17	8.35	2.03	10.0

The maximum error is always below 7% in the whole range of Q for all test video sequences. Moreover, the average error is below 3.5% and is smaller for larger Q values, and for those values does not exceed 2.5% (Fig. 5). The model is quite accurate in the whole range of quantization scale factor Q. Figure 4 shows results of AC DCT bitstream estimation for luminance and chrominance.

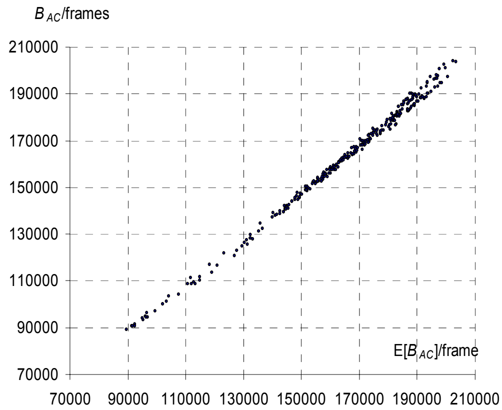

Fig. 4. Measured bitrate B_{AC} of the AC luminance coefficients versus estimated bitrate $E[B_{AC}]$ for the test sequence *Basket* in the 4CIF resolution.

5. APPLICATION OF THE MODEL TO BITRATE CONTROL

The goal is to estimate Q in order to match assumed bitrate for a given picture. The control algorithm is the following:

- First step: Computation of B_{CONST} and histograms of DCT coefficients.
- Second step: Estimation of $B_{VAR}(Q)$ for certain quantization scale factor Q values.
- Third step: On the basis of calculated function $B(Q)$, choice the quantization scale factor Q value such that assumed bitrate is matched.

In order to assess experimentally the proposed bitrate control algorithm, a set of 11 test video sequences (both progressive and interlaced) was used. In order to compare the proposed algorithm and standard TM5 control scheme, the proposed control algorithm has been implemented within the reference software of Test Model 5 MPEG-2 [11].

Both default and proposed control algorithm achieves required bitrate with the same accuracy. However, coder with the new control algorithm obtains higher average PSNR of encoded video sequences. This gain is about 0.2÷0.5 dB. Moreover, variance of the PSNR is slightly lower then that for standard control. For variable channel capacity, the proposed algorithm proves to react well to rapid variations of available bitrate.

6. CONCLUSIONS

An universal bitstream model $B(Q)$ has been described for hybrid video coders. The model has been considered for MPEG-2 codecs but other experiments prove its applicability to H.263 codecs [15]. The model exhibit accuracy being high enough for bitrate control tasks. The computational cost of the model is quit low. It is estimated to be about 0.1% of the total computational effort needed for encoding.

The model can be used for setting a value of the quantizer scale factor Q for a given number of bits for a frame and can be also used to set a value of the quantizer scale factor Q for individual slices. The model is easy to use and its application allows for quit precise bitrate control. The prospective applications include low-delay video bitrate control as the coder can safely work with a small buffer.

ACKNOWLEDGEMENT

The work has been supported by Polish Committee for Scientific Research.

REFERENCES

[1] W. Ding, B. Liu, “Rate Control of MPEG-2 Video Coding and Recording by Rate-Quantization Modeling”, *IEEE Trans. Circ. Systems Video Techn.*, vol. 6, pp. 12-20, 1996.

[2] W. Y. Lee, J. B. Ra, “Fast Algorithm for Optimal Bit Allocation in Rate-Distortion Sense”, *Electronics Letters*, vol. 32, pp. 1871-1873, 1996.

[3] J. Streit, L. Hanzo, “Fixed-rate Video Codecs for Mobile Radio Systems,” *European Transactions Telecomm. (ETT)*, vol. 8, no. 6, pp. 551-572, 1997.

[4] K. Stuhlmuller, N. Farber, M. Link, B. Girod, “Analysis of Video Transmission over Lossy Channels”, *IEEE J. Selected Areas in Comm.,* vol. 18, no. 6, pp. 1012- 1032, 2000.

[5] G. Siemek, “The Coefficient Based Rate Distortion Model for the Low Bit Rate Video Coding”, *Computer Analysis of Images and Patterns Conference*, LNCS 2124, pp. 118-124, 2001.

[6] Y. Kim, Z. He, K. Mitra, “A Linear Source Model and a Unified Rate Control Algorithm for H.263/MPEG-2/MPEG-4”, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP*), May 2001.

[7] L.-J. Lin, A. Ortega, “Bit-Rate Control Using Piecewise Approximated Rate-Distortion Characteristics”, *IEEE Trans. Circuits Systems Video Techn.* , vol. 8, no. 4, pp. 446-459 , 1998.

[8] T. Chiang, Y. Q. Zhang, “A New Rate Control Scheme Using Quadratic Rate Distortion Model”, *IEEE Trans. Circ. Systems Video Techn.*, vol. 7, no. 1, pp. 246-250, 1997.

[9] ISO/IEC IS 13818-2, “Information Technology - Generic Coding of Moving Pictures and Associated Audio Information – Part 2: Video”.

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

[11] MPEG Software Simulation Group, MPEG-2 Encoder ver. 1.2, 1996, ftp:// ftp.mpeg.org/pub/mpeg/mssg.

[12] ITU-T Rec. H.263, “Video Coding for Low Bit Rate Communication.”

[13] ISO/IEC/SC29/WG11/ N5555, ISO/IEC FDIS 14496-10 │ ITU-T Rec. H.264, Information Technology – Coding of Audio-Visual Objects – Part 10: Advanced Video Coding, Pattaya, March 2003.

[14] 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.

[15] A. Łuczak, “Digital Video Bitrate Control Using Coder Models”, PhD dissertation, Poznań Univ. Technol., Poznań 2003.