SIMPLE GLOBAL MODEL OF AN MPEG-2 BITSTREAM

Marek Domański, Adam Łuczak

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

ABSTRACT

The paper describes a simple empirical model of bitstreams produced by MPEG-2 video coders. The number of bits that represent an I- or P-frame is expressed as a function of the quantization factor Q. This model is global in the sense that it describes the total number of bits in a frame thus ignoring local bitrate variation within a frame as well as changes of Q within a frame. This five-parameter model fits well to experimental data. Four of these parameters may be fixed thus obtaining less accurate one-parameter model that can be used for coder control. The experimental results are reported for some video test sequences in the BT.601/4CIF format encoded by an MPEG-2 MP@ML coder.

1. INTRODUCTION

In the past 15 years, several successful video compression standards have emerged, like ITU-T H.261, H.263/H.263+, ISO/IEC MPEG-1, MPEG-2 and MPEG-4. These well-known international standards are related to innumerable software and hardware implementations but they actually define mostly semantics and syntax of the compressed bitstream and not coder implementations, which are described in the informative parts of some standards as examples only.

Among many video coding standards, MPEG-2 [1,2] is extremely successful because of proliferation of digital storage media and digital transmission of television signals. Similarly as other standards and recommendations, MPEG-2 does not standardize any algorithm to set the coder parameters in order to obtain required bitrate and/or distortion level. The video coding part of MPEG-2 describes a buffer control algorithm which is only a part of the bitrate and distortion control mechanisms. The basic parameter that can be used to conrol an encoder is the quantization factor Q that influences the quantization of the DCT coefficients. Typical goals of adjusting the parameter Q is to preserve constant quality in the variable bitrate mode (VBR) or to preserve constant bitrate in the constant bitrate mode (CBR). In the communication systems, the video coders often operate in the latter mode because of the need to match the bitstream and communication channel throughput. Quality of the control algorithm deeply influences the overall coding performance of the video coding system. Unfortunately, despite of millions of the MPEG-2 coders and decoders working worldwide, defining efficient control algorithms is still an open problem that gains a lot of attention recently [3]. Some methods use Lagrangian multiplier techniques [e.g. 4-6] which require large computational complexities. Some schemes require large delays [7]. Other techniques try to exploit some picture models, e.g. [8].

In particular, the very interesting question is how to control the value of Q in order to maximize the video quality by a given bitrate constrain. The answer is related to the two following control problems:

- Global control, i.e. adjusting the Q value in a frame header,
- Local control, i.e. adjusting the Q value in the invidual macroblock headers.

The paper addresses the first problem. The goal is to establish a simple empirical model for a bitrate as function of Q set for individual frames. The model should be valid for typical conditions of operation of the coder. In the paper, the experimetal results are limited to I- and P-frames but similar models can be developed for B-frames also. The model can be used for global control in the constant bitrate mode of operation. For the sake of simplicity, absence of adjustment of Q in the individual macroblocks is assumed.

An empirical model is proposed in the paper. The model matches the experimental data gathered for an MPEG-2 MP@ML coder. All the experimental results are quoted for progressive video sequences rounded to 4CIF format.

2. BITSREAM STRUCTURE

A number of bits B allocated to an individual frame:

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

where B_{CONST} – the portion that does not depend on Q, $B_{\mathrm{VAR}}(Q)$ – the portion that depends on Q.

0-7803-7304-9/02/$17.00 C2002 IEEE

In an I-frame, there is

$$B_{CONST} = B_{CTR} + B_{YDC} + B_{CDC}\,, \qquad (2)$$

where B_{CTR} denotes the number of bits for the frame, slice and macroblock headers and B_{YDC} and B_{CDC} denote the number of bits for intra-DC coefficients in the luminance and the chrominance component, respectively. Similarly, in a P-frame, there is

$$B_{CONST} = B_{CTR} + B_{MV}, \qquad (3)$$

where B_{MV} denote the number of bits needed to encode motion vectors. For a given sequence, all the components of B_{CONST} are relatively similar for various frames both for I- and P-frames (Figs. 1 and 2, Table 1) .

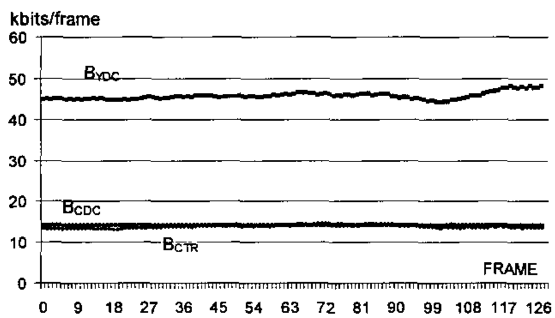

Fig. 1. The bitstream components that are independent from Q in I-frames of the test sequence *Basket*.

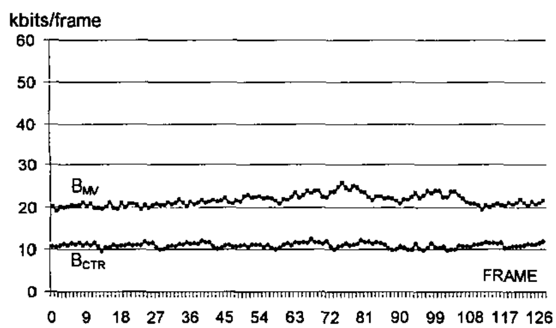

Fig.2. The bitstream components that are independent from Q in P-frames of the test sequence *Basket*.

For various frames of three test sequences, the simulation results are summarized in Table I for the bitstream components that are independent from Q. The standard deviation of the bitrate calculated for 128 frames remains small. Therefore, the control unit of the encoder is able to estimate B_{CONST} using the value of B_{CONST} calculated for a previous frame. In particular, the value of B_{CTR} is almost constant for I-frames.

Table 1. The bitstream components that are independent from Q in various frames of three test sequences.

	bits/frame				
Frame	I-frames			P-frames	
Flower garden					
	B_{CTR}	B_{YDC}	B_{CDC}	B_{CTR}	B_{MV}
1	14356	43895	14193	11120	9847
16	14355	45007	14837	10808	9887
32	14394	45694	14940	9784	9250
1-128	14364	45521	15279	10075	9650
Std.dev.	13	527	37	632	296
Cheer					
1	14383	43849	17010	7594	8310
16	14394	44196	17363	4848	7317
32	14395	43823	16936	7145	8198
1-128	14383	43714	16892	5674	7724
Std.dev	12	361	263	832	389
Stefan					
1	14364	37783	12536	8870	10801
16	14372	37626	12370	8915	12122
32	14377	38173	12583	8879	11127
1-128	14382	38330	12532	8842	11324
Std.dev.	11	439	146	162	698

3. GENERAL BITSTREAM MODEL

Let us consider the bitstream component that depends on the frame quantization factor Q.

$$B_{VAR}(Q) =, \quad B_{YV}(Q) + B_{CV}(Q) \qquad (4)$$

where $B_{YV}(Q)$ and $B_{CV}(Q)$ denote the bits needed for encoding of the DCT coefficients (except the intra-DC ones) in the luminance and the chrominance, respectively. In the MPEG-2 bitstream, the nonzero DCT coefficients are encoded using Huffman variable length codes. Nevertheless the average codeword length is almost constant when calculated over sufficiently large set of image blocks [8]. In the "codeword number per macroblock" and "bits per macroblock" coordinates, the two -dimensional histogram (Fig. 3) has its maxima along a line defined as

$$\textit{Number of bits} \approx 6.05 \cdot \textit{Number of codewords}\,. \qquad (5)$$

The above number of bits plus the number of bits needed for the EOB codes (codes for *End of block*) is equal to $B_{YV}(Q)$ or $B_{CV}(Q)$ 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 $B_{YV}(Q)$ and $B_{CV}(Q)$ can be modeled statistically as functions of Q. The reason for that is that a 4CIF frame contains a sufficiently large set of blocks

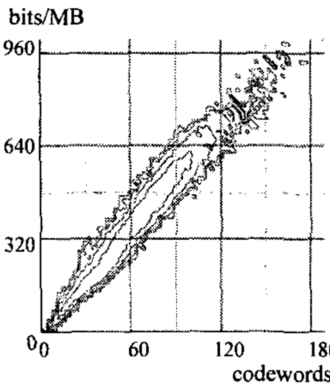

Fig. 3. Two-dimensional histogram "codewords – bits" for the test sequence *Basket*.

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

$$B_x(Q) = \frac{a}{c(Q^b + e) + d}\ , \tag{6}$$

where a, b, c, d and e are real constants that depend on the frame content. The parameter values can be estimated by approximation error minimization over the interval <1,31> of the allowed values of Q. Precision of approximation of the experimental data B_e is measured by

$$\varepsilon = \frac{|B_e - B_x|}{B_e} \cdot 100\%\ , \tag{7}$$

and ε_{max} means the maximum value of ε and ε_{ave} means the average value of ε. If the parameters are estimated for individual frames, the model $B_x(Q)$ approximates the experimental data $B_e(Q)$ so well that often it is difficult to distinguish the experimental and the modeled curve (Fig. 4).

Table 3. Parameters for the luminance in I-frames.

Frame	Parameters					ε_{max}	ε_{ave}
	$a/10^6$	b	c	d	e		
Flower garden							
0	5.7	0.91	2.06	0.11	0.87	7.9	2.9
32	6.4	0.91	2.05	0.11	0.88	6.8	2.9
64	6.5	0.91	2.05	0.11	0.89	7.1	2.6
96	3.3	0.86	2.28	0.10	0.86	6.2	2.3
Cheer							
0	1.4	0.95	0.61	0.06	0.71	7.6	2.8
32	1.3	0.93	0.63	0.05	0.68	8.2	2.8
64	1.4	0.95	0.64	0.05	0.67	8.0	2.9
96	1.3	0.94	0.63	0.05	0.67	8.3	2.9
Stefan							
0	1.9	0.89	0.59	0.05	0.38	7.0	2.3
32	1.1	0.88	0.66	0.05	0.44	5.8	2.0
64	1.1	0.86	0.67	0.05	0.38	6.2	2.2
96	1.1	0.85	0.67	0.05	0.37	5.5	1.9

Table 4. Parameters for the luminance in P-frames.

Frame	Parameters					ε_{max}	ε_{ave}
	$a/10^6$	b	c	d	e		
Flower garden							
0	2.7	1.07	0.90	0.53	0.56	13.8	4.9
32	2.1	1.31	0.47	0.77	0.49	15.6	8.2
64	1.2	1.20	0.31	0.39	0.03	13.7	8.5
96	2.2	1.18	0.62	0.62	0.10	14.7	8.7
Cheer							
0	2.3	1.24	0.56	0.56	0.47	7.3	2.8
32	2.1	1.14	0.63	0.66	0.41	15.8	4.2
64	2.0	1.26	0.58	0.70	0.42	9.9	4.1
96	2.0	1.13	0.61	0.64	0.41	12.6	4.6
Stefan							
0	1.7	1.11	0.59	0.67	0.42	4.2	2.2
32	1.9	1.09	0.64	0.76	0.41	10.6	5.6
64	1.9	1.22	0.73	0.68	0.41	13.8	5.6
96	2.1	1.23	0.64	0.92	0.42	8.6	4.3

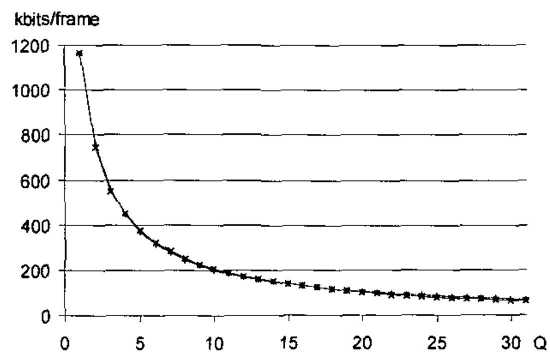

Fig. 4. The experimental and modeled curves for an I-frame from the test sequence *Basket*.

4. BITSTREAM MODEL WITH ONE FREE PARAMETER

The less parameters depending on frame content, the more useful becomes the model for coder control. Therefore it was assumed that only the parameter c depends on frame content, while the other parameters assume general values that can be found by minimization of the approximation error $|B_e - B_x|$ over a set of frames from some training sequences. 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*. The optimization was performed in a window $Q_0 \pm dQ$ for various Q_0 and dQ (Fig. 5). The precision of approximation was measured using the parameter ε calculated for a given frame. This precision has been measured for a set of test sequences being different of that used to train the model (Tables 5 and 6).

Table 5. Precision of approximation for a general model with one free parameter for I-frames.

Sequence	ε_{max} [%]			ε_{ave} [%]		
	dQ=2	dQ=4	dQ=8	dQ=2	dQ=4	dQ=8
Flower garden – luminance						
Q_0=10	0.9	1.8	5.9	0.4	0.8	2.0
Q_0=16	0.6	1.1	2.6	0.4	0.4	0.8
Q_0=22	0.6	1.0	1.6	0.3	0.4	0.5
Cheer– luminance						
Q_0=10	0.9	3.6	9.4	0.5	1.3	3.3
Q_0=16	0.9	1.7	5.3	0.5	0.6	1.6
Q_0=22	0.8	0.9	2.3	0.3	0.3	0.6
Stefan– luminance						
Q_0=10	1.0	3.3	4.9	0.6	1.0	1.8
Q_0=16	1.0	2.2	4.9	0.4	0.6	0.9
Q_0=22	1.2	2.6	5.0	0.7	1.1	1.4
Flower garden - chrominance						
Q_0=10	0.9	3.6	10.9	0.4	1.1	3.9
Q_0=16	2.0	2.6	3.4	0.9	1.0	1.8
Q_0=22	2.4	1.7	3.07	1.1	0.7	0.9
Cheer - chrominance						
Q_0=10	4.0	7.6	21.3	1.9	2.5	3.6
Q_0=16	2.8	5.6	8.2	1.2	2.1	3.6
Q_0=22	4.9	3.7	7.7	2.3	1.7	3.5
Stefan - chrominance						
Q_0=10	6.0	10.2	22.9	2.3	2.5	5.2
Q_0=16	4.0	7.3	18.2	1.3	2.0	4.3
Q_0=22	2.8	5.0	10.6	1.0	1.8	2.2

Table 6. Precision of approximation for a general model with one free parameter for P-frames.

Sequance	ε_{max} [%]			ε_{ave} [%]		
	dQ=2	dQ=4	dQ=8	dQ=2	dQ=4	dQ=8
Flower garden - luminance						
Q_0=10	8.2	8.5	28.9	1.5	2.3	4.9
Q_0=16	6.9	8.1	16.9	1.3	3.1	4.9
Q_0=22	3.3	6.3	7.9	0.8	1.35	2.4
Cheer- luminance						
Q_0=10	7.5	19.3	29.9	1.7	2.9	5.5
Q_0=16	8.6	17.9	36.7	2.6	4.8	8.4
Q_0=22	8.5	13.6	19.9	2.4	3.9	5.9
Stefan- luminance						
Q_0=10	8.8	8.9	21.0	1.7	2.2	5.8
Q_0=16	10.9	9.3	16.6	2.5	2.8	4.0
Q_0=22	11.9	16.5	24.4	1.52	3.0	5.1
Flower garden - chrominance						
Q_0=10	14.4	29.5	63.6	4.9	5.5	8.8
Q_0=16	23.0	27.8	36.7	6.6	6.2	10.9
Q_0=22	15.6	24.2	40.9	3.6	8.1	11.8
Cheer- chrominance						
Q_0=10	.19.7	22.3	43.2	4.9	6.3	8.9
Q_0=16	20.5	25.4	34.8	3.7	6.5	8.7
Q_0=22	14.7	20.8	26.2	4.4	5.9	7.1
Stefan- chrominance						
Q_0=10	19.0	33.25	101.8	6.9	9.4	16.3
Q_0=16	39.8	33.7	43.2	7.1	8	12.9
Q_0=22	26.7	23.6	68.7	5.4	6.8	14.7

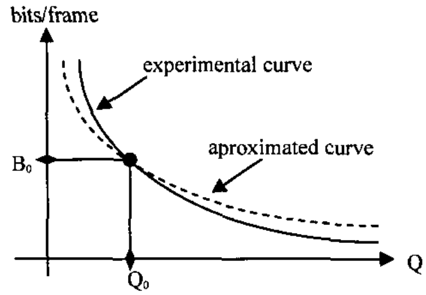

Fig. 5. Approximation in the vicinity of Q_0.

For a given frame, the value of Q that matches the bitrate can be easily estimated using the model with one free parameter c. The number of bits B_0 for an initial value Q_0 can be estimated by simple count of nonzero DCT coefficients. Than, the requested value of Q can be estimated from the model.

5. CONCLUSIONS

The paper presents a simple global model that well approximates the experimental values of bitrate as a function of the quantization factor Q. It was shown that this model with one free parameter can be efficiently used for bitrate control.

REFERENCES

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

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

[3] G.J. Sullivan, T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Magazine*, vol. 15, pp. 74-90, Nov. 1998.

[4] J.H. Choi, D.C. Park, "A stable feedback control of the buffer state using Lagrange multiplier method" IEEE Trans. Image Processing, vol. 3, pp. 546-558, Sept. 1994.

[5] W.Y. Lee, J.B. Ra, "Fast algorithm for optimal bit allocation in rate-distortion sense, Electronics Letters, vol. 32, pp. 1871-1873, Sept. 1996.

[6] T. Wiegand, B. Girod, "Lagrange multiplier selection in hybrid video coder control," *Proc. IEEE Int. Conf. Image Proc.*, pp. 542-545, Oct. 2001.

[7] 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.

[8] 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.