Single Frame Rate-Quantization Model for MPEG-4 AVC/H.264 Video Encoders

Tomasz Grajek and Marek Domański

Poznan University of Technology
Chair of Multimedia Telecommunication and Microelectronics
ul. Polanka 3, 60-965 Poznań, Poland
{tgrajek, domanski}@et.put.poznan.pl

Abstract. The paper describes a new model of MPEG-4 AVC/H.264 encoders. The model is proposed as a formula that expresses the number of bits per frame against the quantization step. Proposed are variants of the model for I-, P- and B-frames as well as for 3 intervals of quantization step values. The model has been derived from an extensive set of experimental data obtained by encoding numerous standard test video sequences. Experimental tests have proved good accuracy of the model proposed.

Keywords: compression, MPEG-4 AVC/H.264 standard, video encoding, video encoder modeling.

1 Introduction

The paper deals with modeling of advanced video encoders that are compliant with MPEG-4 AVC/H.264 standard [1]. In particular, the proposed models refer to encoders with compression performance being the same as that of the reference implementation of a video encoder of the standard MPEG-4 AVC/H.264 [13].

Modeling of video encoders is not only important for general research but it is also closely related to video encoder control. Despite of huge number of MPEG-4 AVC/H.264 video encoders working worldwide, designing efficient control algorithms is still a challenge. The basic parameter that is used to control an encoder is the quantization parameter *QP* that defines the quantization step size *Q* for transform coefficients. A typical goal of adjusting the quantization step size *Q* is to match the available channel bitrate. The *Q* parameter may be adjusted on the frame level, the slice level as well as on the macroblock level. Adjusting of *Q* value allows to control bitrate in a very wide range, e.g. from tens of megabits to tens of kilobits per second for standard definition video (i.e. 4CIF sequences).

In the informative part of MPEG-4 AVC/H.264 standard, a mode selection algorithm is described that exploits Rate-Distortion Optimization (RDO) [7]. At its input, this optimization algorithm needs a value of quantization parameter *QP* (directly related to *Q*), but this value can be determined from the mean absolute difference (MAD) between the original image and the predicted one. As the MAD value is unknown until an image is encoded, a linear model for MAD prediction has been proposed [7]. Then quadratic rate-quantization model is used to calculate *QP* value [2]. In fact, many existing rate control schemes use the above described approaches.

Different approach to rate control was presented by He and Mitra in [5]. They have proposed to use a linear ρ-domain source model, where ρ denotes percentage of zeros in quantized transform coefficients. This model has proved to be very accurate in source content estimation, therefore several new rate control models have been developed as its variants, e.g. [8, 9]. Unfortunately, direct implementation of this model inside MPEG-4 AVC/H.264 encoders is quite complicated [4, 9].

The above mentioned models describe relationship between Q values and texture bits only i.e. bits for quantized transform coefficients. However, so called non-textured bits constitute a substantial portion of the total number of bits in AVC/H.264 bitstream. This number of bits should not be neglected, especially for low bitrates (see Fig. 1) [10, 11]. Consequently, the encoder models should take both groups of bits into account. One solution is to apply separate models to texture and non-texture bits, at the price of higher complexity [6]. Another way is to develop a model for the total number of bits [10, 12]. In this paper we exploit this second approach.

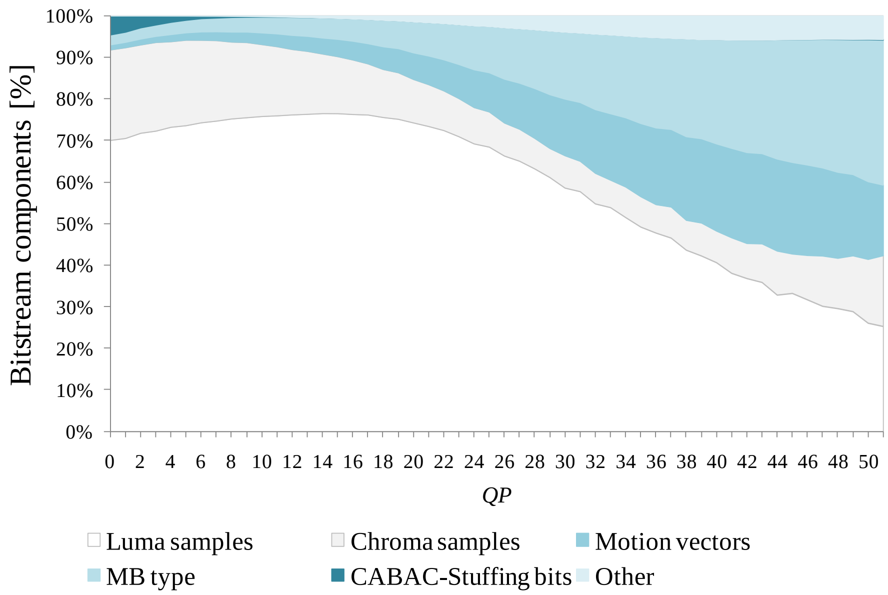

Fig. 1. Bitstream components for the reference MPEG-4 AVC/H.264 encoder averaged for all test sequences (encoder configuration, main profile, group of pictures: IBBPBBPBBPBBP, CABAC, deblocking filter switched on).

Our proposal is to treat a coder as a "black box" with one input (video sequence) and one output (bitstream), controlled by only one parameter – Q, and to find an appropriate mathematical description that allows to estimate the total number of bits B with high accuracy. Here, number of bits B is the number of bits in the representation of a single individual frame. Mostly, this number is quite similar for frames of the same type taken from the same shot, i.e. I-, P- or B-frames taken from the interval between two scene cuts.

The model considered can be expressed as

$$B = f(Q, \Phi)\,, \tag{1}$$

where Φ is a respective vector of parameters that depend on current video content and are obtainable by experimental data analysis.

In earlier papers (e.g. [2]), the authors have already proposed formulas for average number of all bits per frame within a single shot as well as for texture bits only in individual frames. In contrary to the above mentioned papers, here, we are going to derive a formula for all bits in individual frames.

2 Proposed Model

Our objective is to find the relationship between the number of bits B and quantization step size Q for a given frame type in a given sequence (Eq. 1). The relation will be established by analysis of experimental data.

Unfortunately, derivation of the model for the whole allowed range of Q values turned out to be inefficient (due to large approximation errors). Therefore, the entire allowed range of Q values has been divided into 3 intervals (Fig. 2, Table 1), and the model has been derived individually in each interval.

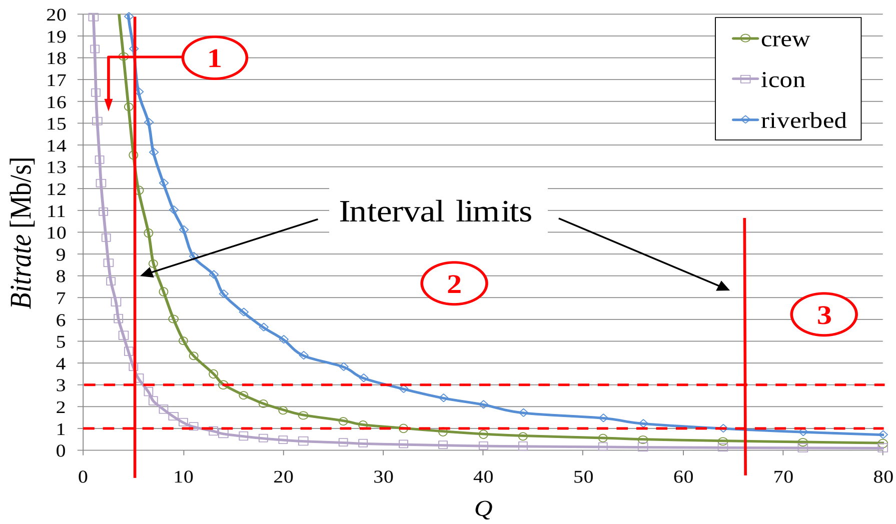

Fig. 2. Interval limits and experimental curves for 3 test sequences: *crew*, *icon*, *riverbed*. For the sake of clarity, range of Q and *Bitrate* have been clipped to 80 and 20 respectively.

The most interesting central interval (Interval 2) covers Q values corresponding to bitrates from about 1 to 3 Mb/s (Fig. 2). In practical applications, it is the most useful range of bitrates for standard-definition video, i.e. 4CIF video sequences. Each test video sequence has a different range of Q values corresponding to the above mentioned range of bitrates. Therefore, for Interval 2, its low end has been set as the minimum value of Q that corresponds to 1 Mb/s in all test video sequences (see Section 3). Similarly the high end of Interval 2 has been set as the maximum value of Q corresponding to 3 Mb/s.

Table 1. Intervals for quantization step size Q and quantization parameter QP.

Interval	Range of Q	Range of QP
1	<0.625; 4.5>	<0; 17>
2	<5; 64>	<18; 40>
3	<72; 224>	<41; 51>

2.1 Interval 2

For Interval 2, a hyperbolic model is proposed for number of bits per frame B

$$B(Q, \Phi) = \frac{a}{Q^b + c}, \qquad \textbf{(2)}$$

where $\Phi[a\ b\ c]$ is the vector of parameters that depend on sequence content. For these parameters, their values may change significantly at scene cuts. Within a shot, their values remain similar for all frames of a given type.

2.2 Intervals 1 and 3

For Interval 1 as well as for Interval 3, a cubic model is proposed,

$$B(Q, \Phi) = (a_1 \cdot d + a_2) \cdot Q^3 + (b_1 \cdot d + b_2) \cdot Q^2 + (c_1 \cdot d + c_2) \cdot Q + d\,, \qquad \textbf{(3)}$$

where $\Phi[d]$ is the vector with one parameter that depends on video content. The parameters a_1, a_2, b_1, b_2, c_1 and c_2 are the model constants that exhibits different values for both intervals and for each frame type (see Table 2). The model constants have estimated experimentally using a set of test video sequences.

Table 2. Values of universal model constants for Intervals 1 and 3.

Constants	Interval 1			Interval 3		
	I-frame	P-frame	B-frame	I-frame	P-frame	B-frame
a_1	$-1.65\cdot10^{-2}$	$-2.39\cdot10^{-2}$	$-1.85\cdot10^{-2}$	$-7.06\cdot10^{-8}$	$-1.25\cdot10^{-7}$	$-1.49\cdot10^{-7}$
a_2	$4.58\cdot10^{3}$	$1.49\cdot10^{4}$	$3.96\cdot10^{3}$	$-2.11\cdot10^{-3}$	$8.50\cdot10^{-4}$	$5.08\cdot10^{-4}$
b_1	$1.66\cdot10^{-1}$	$2.20\cdot10^{-1}$	$1.73\cdot10^{-1}$	$4.58\cdot10^{-5}$	$7.02\cdot10^{-5}$	$8.11\cdot10^{-5}$
b_2	$-2.63\cdot10^{4}$	$-8.67\cdot10^{4}$	$1.07\cdot10^{4}$	$8.90\cdot10^{-1}$	$-4.22\cdot10^{-1}$	$-2.35\cdot10^{-1}$
c_1	$-5.79\cdot10^{-1}$	$-6.82\cdot10^{-1}$	$-5.70\cdot10^{-1}$	$-1.07\cdot10^{-2}$	$-1.36\cdot10^{-2}$	$-1.51\cdot10^{-2}$
c_2	$-2.94\cdot10^{4}$	$3.21\cdot10^{4}$	$-2.08\cdot10^{5}$	$-8.46\cdot10^{1}$	$-6.21\cdot10^{1}$	$-2.97\cdot10^{1}$

3 DERIVATION OF THE MODEL

The proposed model has been derived individually for each interval of Q and each frame type. Function fitting was applied to experimental data collected for a set of 21 test video sequences. In that way function type was derived.

For the same set of 21 test video sequences, the values of model constants (a_1, a_2, b_1, b_2, c_1, c_2) have been estimated individually in Interval 1 and Interval 3. These constants have been estimated by minimizing the maximum relative approximation error over the respective interval of Q values

$$\min_{\Phi} \max_{Q} \varepsilon(Q, \Phi)\,, \tag{4}$$

$$\varepsilon(Q, \Phi) = \frac{|B_x(Q) - B(Q, \Phi)|}{B_x(Q)} \cdot 100\%\,, \tag{5}$$

where $B_x(Q)$ denotes the measured number of bits per frame and $B(Q, \Phi$ denotes the value calculated from the model.

For both, model type choice and model constant estimation, experimental data have been collected for 21 various 4CIF at 25Hz and 30Hz sequences. Each test video sequence consisted of 199 frames. The test video sequences had different motion characteristics. The following sequences have been used: *basket*, *bluesky*, *bus*, *cheer*, *city*, *crew*, *flow*, *football*, *harbour*, *ice*, *icon*, *pedestrian*, *riverbed*, *rushhour*, *soccer*, *station2*, *stefan*, *sunflower*, *tractor*, *universal* and *warner*. All sequences have been encoded using MPEG-4 AVC/H.264 reference software version JM_13.2 [13] (main profile, CABAC and RDO enabled, GOP: IBBPBBPBBPBBP). Each sequence has been encoded with all eligible values of Q. Of course, each value of Q was set indirectly by properly setting the value of quantization parameter QP. Sequences *bluesky*, *pedestrian*, *riverbed*, *rushhour*, *station2*, *sunflower* and *tractor* have been cropped to 4CIF resolution from their original 720p format.

4 MODEL ACCURACY

In order to measure the accuracy of the model, the mean relative approximation error has been calculated separately for each encoded frame and for each interval of Q values. The experiments with additional test video sequences have shown that the model constants are chosen correctly, i.e. model accuracy remains similar also for the test material from outside of the initial set of 21 test sequences.

Interval 1

In Table 3, the values of relative approximation error have been shown for Interval 1. The number in brackets indicates how many parameters the model has, e.g. “cubic (4)” refers to a version of the model from Eq. 3 with 4 content-dependent parameters

$$B(Q, \Phi) = a \cdot Q^3 + b \cdot Q^2 + c \cdot Q + d\,, \qquad (6)$$

Here, *a, b, c* and *d* are video-content-dependent parameters.

Table 3. Mean (i.e. averaged over all Q values) relative approximation error for Interval 1.

Model	Relative error averaged for all Q values for a given frame [%]			
	max over all frames	min over all frames	mean over all frames	std. deviation over all frames
I frame				
cubic(4)	2.27	0.80	1.52	0.36
cubic(1)	70.02	1.16	3.84	7.98
P-frame				
cubic(4)	9.11	1.77	3.67	1.00
cubic(1)	152.16	2.79	7.93	14.50
B-frame				
cubic(4)	11.59	1.60	3.29	1.39
cubic(1)	197.15	2.43	9.29	20.00

The average relative approximation error for the simplified cubic model (with one content dependent parameter *d*) is reasonable.

Interval 2

The characteristic parameters of the mean relative approximation error calculated individually for I-, P- and B-frames have been presented in Table 4.

Table 4. Mean (i.e. averaged over all Q values) relative approximation error for Interval 2.

Frame type	Relative error averaged for all Q values for a given frame [%]			
	max over all frames	min over all frames	mean over all frames	std. deviation over all frames
I	7.36	0.81	1.30	0.55
P	35.13	0.92	3.11	3.03
B	55.12	0.89	7.60	6.77

The average relative approximation error is about 1.3%, 3.1% and 7.6% for I-, P- and B-frames, respectively. This proves usefulness and very good accuracy of the proposed model for encoders in the mostly used range of bitrates and quantization steps.

Interval 3

For simplified cubic models, for P- and B-frames, model accuracy is poor in Interval 3 (Table 5), in contrary to the two previous intervals. In Interval 3, estimation of the number of bits is more difficult (especially for P- and B-frames) because for large quantization steps, numbers of bits are quite low. Therefore even small absolute errors in the number of bits, yield large relative approximation errors, especially for larger values of Q. In average, the approximation errors are probably acceptable but there exist some frames with really huge approximation errors.

Table 5. Mean (i.e. averaged over all Q values) relative approximation error for Interval 3.

Model	Relative error averaged for all Q values for a given frame [%]			
	max over all frames	min over all frames	mean over all frames	std. deviation over all frames
I frame				
cubic(4)	4.40	0.73	1.62	0.49
cubic(1)	38.80	1.15	6.06	4.97
P-frame				
cubic(4)	34.52	0.88	3.20	2.57
cubic(1)	380.36	1.57	14.70	31.38
B-frame				
cubic(4)	42.25	0.00	6.54	4.62
cubic(1)	164.15	2.78	38.32	35.64

5 Conclusions

In this paper, a new model for MPEG-4 AVC/H.264 encoders has been described. This model is given by as a function $B(Q)$, where B is the total number of bits in a frame. This is the difference with respect to other model described in the references that have proposed models for transform coefficient bits.

The new model is applicable for bitrate control. Moreover, it can be considered as a mathematical model of a video encoder that may be used in studies in video compression.

In this paper, the model parameters have been estimated for standard-definition video. Nevertheless, similar relations hold for high-definition video as well.

The new model has been obtained by curve fitting that minimized the maximum approximation error of experimental data. These experimental data were obtained from extensive experiments with 21 video test sequences. The analysis of this huge set of experimental data resulted in proposal of the function type that is able to fit well the experimental data in individual intervals of quantization step size, and for different picture types. For two intervals, only one model parameter is needed that depends on sequence content. For these intervals, other parameters have been estimated as universal constants.

This model can be used to set a value of the quantization parameter QP for a given number of bits for an I- P- or B-frames. Tests proved that the model mostly fits experimental data very well in a wide range of bitrates. For the most useful range of bitrates for 4CIF sequences (Interval 2), the relative approximation error is about 1.3% for I-frames, 3.1% for P-frames and 7.6% for B-frames. For Intervals 1 and Interval 3, the relative approximation errors are higher when only one content-dependent parameter is used.

We conclude that the proposed model is very accurate for moderate bitrates, i.e. for bitrates that are mostly used for video broadcasting, e.g. in digital television.

Acknowledgements. This work was supported by the public funds as a research project.

References

1. ISO/IEC 14496-10 (MPEG-4AVC) / ITU-T Rec. H.264: Advanced Video Coding for Generic Audiovisual Services, International Organization for Standardization (2007)
2. Chiang, T., Zhang, Y.-Q.: A New Rate Control Scheme Using Quadratic Rate Distortion Model, IEEE Transactions on Circuits and Systems for Video Technology, vol. 7, pp. 246–250 (1997)
3. Grajek, T., Domański, M.: A Global Model of AVC/H.264 Video Encoders, 27th Picture Coding Symposium, PCS, Chicago, Illinois, USA (2009) also IEEE *Xplore* Digital Library, http://ieeexplore.ieee.org
4. He, Z., Chen, T.: Linear Rate Control for JVT Video Coding, International Conference on Information Technology: Research and Education, ITRE, pp. 65–68, Newark, USA (2003)
5. He, Z., Mitra, S.K.: Optimum Bit Allocation and Accurate Rate Control for Video Coding via ρDomain Source Modeling, IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 10, pp. 840–849 (2002)
6. Kwon, D.-K., Shen, M.-Y., Kuo, C.-C. J.: Rate Control for H.264 Video with Enhanced Rate and Distortion Models, IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 5, pp. 517–529 (2007)
7. Li, Z.G., Pan, F., Lim, K.P., Feng, G., Lin, X., Rahardja, S.: Adaptive Basic Unit Layer Rate Control for JVT, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), document JVT-G012r1, Pattaya, Thailand (2003)
8. Lim, S.C., Na, H.R., Lee, Y.L.: Rate Control Based on Linear Regression for H.264/MPEG-4 AVC, Signal Processing: Image Communication, Volume 22, Issue 1, pp. 39–58, Elsevier (2007)
9. Lin, G., Zheng, S., Hu, J.: A Two-Stage ρDomain Rate Control Scheme for H.264 Encoder, IEEE International Conference on Multimedia and Expo ICME'2008, pp. 713–716, Hannover, Germany (2008)
10. Liu, Y., Li, Z.G., Soh, Y.C.: A Novel Rate Control Scheme for Low Delay Video Communication of H.264/AVC Standard, IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 1, pp. 68–78 (2007)
11. Xu, J., He, Y.: A Novel Rate Control for H.264, IEEE International Symposium on Circuits and Systems, ISCAS, vol. 3, pp. 809–812, Vancouver, Canada (2004)
12. Zhuo, L., Gao, X., Wang, Z., Feng, D.D., Shen, L.: A Novel Rate-Quality Model Based H.264/AVC Frame Layer Rate Control Method, 6th International Conference on Information, Communications & Signal Processing, ICICS, pp. 1–5, Singapore (2007)
13. Fraunhofer Institute for Telecommunications Heinrich-Hertz-Institut Image Processing Department, H.264/AVC Software Coordination, http://iphome.hhi.de/suehring/tml/index.htm