NEW MODEL OF MPEG-4 AVC/H.264 VIDEO ENCODERS

Tomasz Grajek and Marek Domański

Chair of Multimedia Telecommunications and Microelectronics
Poznań University of Technology, Poland, http://www.multimedia.edu.pl

ABSTRACT

For MPEG-4 AVC/H.264 video encoders, the paper describes a model that defines a relationship between the number of bits for texture information and the quantization step size Q. The model is derived individually for I-, P- and B-frames. The model is defined by formulas given individually for 3 intervals covering whole eligible range of Q values. Extensive experiments with numerous test video sequences have proved good accuracy of the model proposed. In particular the new model exhibits higher accuracy than that from the reference implementation of AVC.

Index Terms**—** compression, video encoding, MPEG-4 AVC/H.264, video encoder modeling.

1. INTRODUCTION

Despite of huge number of video encoders employed worldwide, development of efficient bitrate control algorithms is still a research problem that gains a lot of attention.

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 conventional goal of adjusting the value of quantization step size Q is to match available channel throughput. The standard approach is to set Q value for whole frame, and then possibly adjust the value of Q for individual slices and macroblocks.

The quantization step size Q strongly influences the number of texture bits B, i.e. the bits that represent transform coefficients in a frame. Moreover, the number of bits B in a frame depends also on video content features. Therefore the applicable approximate models may 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.

The very often used model for P-frames is that from the reference software of MPEG-4 AVC/H.264 [8, 15],

$$B(Q,a,b) = a \cdot \frac{MAD}{Q} + b \cdot \frac{MAD}{Q^2}, \tag{2}$$

where a and b are model parameters that depend on current video content. Mean Absolute Difference (MAD) is the mean of all absolute values of luma prediction errors from the whole frame. Because this model has some drawbacks, various improvements of the above mentioned model have been proposed, for example in [12-14].

In references, many other approaches that exploit some modeling of video encoders are described. For instance, He and Mitra in [5] proposed different approach to rate control by introducing a linear ρ-domain source model, where ρ denotes percentage of zeros in quantized transform coefficients. It turned out to be very accurate in source content estimation hence several new rate control models have been developed based on their observations e.g. [9, 10]. Unfortunately, direct implementation of this model in MPEG-4 AVC/H.264 encoders is complicated [4, 10].

Different approach has been presented in [7], were authors derived power model based on Cauchy-distribution of transform coefficients.

In this paper, the goal is to find a mathematical formula (see Eq. 1) that expresses the number of texture bits B per frame as a function of quantization step size Q and content-dependent parameters Φ. The function $f(Q, \Phi)$ together with its constants will be found by experimental data analysis. The constants of $f(Q, \Phi)$ will be estimated individually for each frame type (i.e. I, P and B).

In papers [2, 3], the authors have already proposed formulas for various numbers of all bits per frame within a shot. In contrary to the above mentioned papers, here, we are going to derive a formula for individual frames and for texture bits only.

2. PROPOSED MODEL

A simple global model of MPEG-2 bitstreams has been described in [1, 11], where bitrate was modeled with high accuracy using only one content-depended parameter. Existence of this model was the motivation for exploration of experimental data obtained from the state-of-the-art MPEG-4 AVC/H.264 encoder [6]. Unfortunately, derivation of the model for the whole allowed range of Q values turned out to be much more difficult that for the MPEG-2 data. Therefore, the entire eligible range of Q values has been divided into 3 intervals (Table 1), and the model has been derived individually in each interval.

Table. 1. Intervals for quantization step Q and the corresponding values of quantization index 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>

The most interesting central interval (Interval 2) covers Q values corresponding to bitrates from about 1 to 3 Mb/s (Fig. 1). In practical applications, it is the most useful range of bitrates for 4CIF sequences. Each test video sequence has a different range of Q values corresponding to the above mentioned range of bitrates. Therefore, lower Interval 2 limit has been set to minimum value of Q, corresponding to 1 Mb/s (for all test video sequences – see Section 3). Similarly the upper Interval 2 limit has been set to maximum value of Q corresponding to 3 Mb/s.

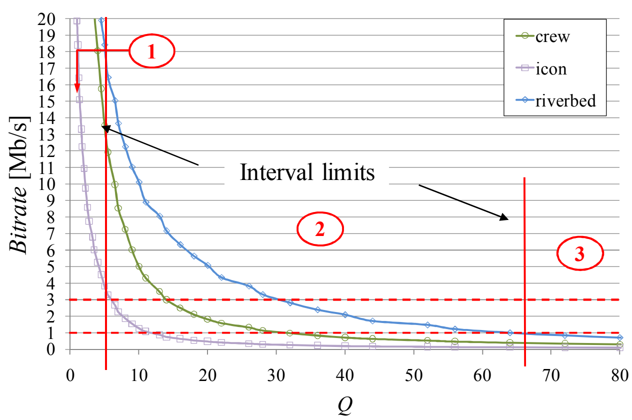

Fig. 1. Interval limits and experimental curves for 3 test sequences: *crew, icon, riverbed.*

Interval 2

For Interval 2, a hyperbolic model is proposed,

$$B(Q,\Phi) = \frac{a}{Q^b + c}, \tag{3}$$

where Φ=[a b c] is the vector of parameters that depend on sequence content and B is the number of texture bits per frame.

Intervals 1 and 3

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

$$B(Q,\Phi) = \begin{array}{l}(a_1 \cdot d + a_2)\cdot Q^3 + (b_1 \cdot d + b_2)\cdot Q^2 \\ + (c_1 \cdot d + c_2)\cdot Q + d\end{array}, \tag{4}$$

where Φ=[d] is the vector with one parameter that depends on video content. 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 Tables 2 and 3).

Table. 2. Values of universal model constants (Interval 1).

Const.	Frame type		
	I	P	B
a_1	-0.0070	-0.0101	-0.0094
a_2	-13965	-18937	-17898
b_1	0.0757	0.1053	0.0978
b_2	155972	195741	189803
c_1	-0.3300	-0.4274	-0.4101
c_2	-552877	-600594	-590877

Table. 3. Values of universal model constants (Interval 3).

Const.	Frame type		
	I	P	B
a_1	-0.00000010	-0.00000013	-0.00000017
a_2	0.00062513	0.00007609	-0.00000190
b_1	0.00005764	0.00007174	0.00009157
b_2	-0.18372429	-0.04031971	0.00037165
c_1	-0.01176460	-0.01387808	-0.01635265
c_2	10.99548793	7.00143862	0.07582038

For Interval 2, the proposed model has 3 parameters that depend on current video content. For Intervals 1 and 3, the cubic model has only one parameter d that is related to video content.

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

$$\min_{\Phi} \max_{Q} \varepsilon(Q,\Phi), \tag{5}$$

$$\varepsilon(Q,\Phi) = \frac{|B_X(Q) - B(Q,\Phi)|}{B_X(Q)} \cdot 100\%, \tag{6}$$

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

For both, model type choice and constants estimation, experimental data have been collected for 21 various 4CIF at 25Hz and 30Hz sequences (each of a length of 199 frames) with 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 [15] (main profile, CABAC and RDO enabled, GOP: IBBPBBPBBPBBP, constant Q mode). 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 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 constants are chosen correctly, i.e. model accuracy remains similar also for the test material from outside of the initial set of 21 sequences. For comparison, the accuracy of the model (Eq. 2) from the reference implementation of the MPEG-4 AVC/H.264 encoder [15] has been estimated for P-frames.

Interval 1

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

$$B(Q,\Phi) = a \cdot Q^3 + b \cdot Q^2 + c \cdot Q + d. \qquad (7)$$

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

Table. 4. Mean relative approximation error (Interval 1).

Model	Mean relative error [%]			
	max	min	mean	std. dev.
I frame				
cubic(4)	2.49	0.64	0.98	0.29
cubic(1)	102.82	0.90	4.58	11.61
P-frame				
cubic(4)	10.30	1.45	3.00	1.51
cubic(1)	167.21	1.83	8.21	16.23
ref(2)	29.60	1.58	8.63	5.42
B-frame				
cubic(4)	15.36	1.42	3.38	2.04
cubic(1)	199.62	2.01	10.09	20.33

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 5. For P-frames, for comparison with the model from the reference software of MPEG-4 AVC/H.264, the respective results are also shown as "P-ref". Additionally, Fig. 2 shows experimental and approximated $B(Q)$ lines for 2 exemplary P-frames from *bluesky* and *football* sequences for Q values from Interval 2.

Table. 5. Mean relative approximation error (Interval 2).

Frame type	Mean relative error [%]			
	max	min	mean	std. dev.
I	6.92	0.86	1.56	0.71
P	72.00	0.98	4.27	4.82
P-ref	101.10	2.45	15.87	9.43
B	95.29	1.34	13.61	16.62

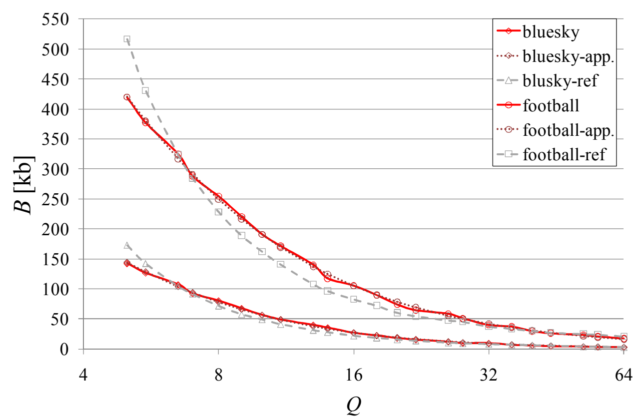

Fig. 2. Experimental and approximated (the proposed and the reference model) curves for 2 exemplary P-frames from *bluesky* and *football* sequences (interval 2): "bluesky" and "football" refer to experimental data while "bluesky-app." and "football-app." refer to approximation by the proposed model and "bluesky-ref." and "football-ref" refer to the model from the reference software.

The average relative approximation error is about 1.6%, 4.3% and 13.6% for I-, P- and B-frames, respectively. The proposed model clearly outperforms the model from the reference implementation of MPEG-4 AVC/H.264 encoder (see Fig. 2). The accuracy of the proposed model is about 11% higher than that of the model from the reference software. This proves usefulness of the proposed model for encoders in the mostly used range of bitrates and quantization steps.

Interval 3

For all models, for P- and B-frames, model accuracy is poor in Interval 3 (Table 6), in contrary to the two previous intervals. The proposed model with one parameter d is not useful in Interval 3 but the model with 4 parameters (see Interval 1) yields the results comparable to those obtained for the reference model. In Interval 3, estimation of the number of texture bits is more difficult (especially for P- and B-frames) because for large quantization steps very small numbers of texture bits may be registered. Therefore even small absolute errors in the number of bits, especially for larger values of *Q,* yield large relative approximation errors.

Table. 6. Mean relative approximation error (Interval 3).

Model	Mean relative error [%]			
	max	min	mean	std. dev.
I frame				
cubic(4)	7.22	0.60	1.86	0.70
cubic(1)	27.16	1.23	8.33	6.33
P-frame				
cubic(4)	583.79	0.81	10.98	33.68
cubic(1)	15443.92	2.43	121.09	903.64
ref(2)	99.16	1.26	11.04	12.42
B-frame				
cubic(4)	1875.40	0.00	140.98	204.92
cubic(1)	864.64	11.12	237.94	246.41

5. CONCLUSIONS

A new quantitative model for MPEG-4 AVC/H.264 encoders has been described. The research started with careful analysis of experimental data gathered for 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 texture 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 1.6% for I-frames, 4.3% for P-frames and 13.6% for B-frames. Therefore, the proposed model outperforms the model from the reference implementation of MPEG-4 AVC/H.264 encoder. For Intervals 1 and Interval 3, the relative approximation errors are higher when only one content-dependent parameter is used. Nevertheless, proposed is a very accurate model that fits experimental data very well for the most useful range of bitrates.

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

REFERENCES

[1] Domański, M., Łuczak, A.: "Simple global model of an MPEG-2 bitstream", *IEEE International Conference on Multimedia and Expo*, ICME, pp. 645 – 648, Lausanne, Switzerland (2002).

[2] Grajek, T., Domański, M.: A global model of AVC/H.264 video encoders, 27th Picture Coding Symposium, PCS, Chicago, Illinois, USA (2009).

[3] Grajek, T., Domański, M.: Single Frame Rate-Quantization Model for MPEG-4 AVC/H.264 Video Encoders" in L. Bolc, R. Tadeusiewicz, and L.J. Chmielewski (eds.): Computer Vision and Graphics, Proc. ICCVG 2010, Lecture Notes in Computer Science, Springer-Verlag (2010) – in print.

[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 ρ-domain source modeling", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 10, pp. 840-849 (2002).

[6] ISO/IEC FDIS 14496-10: *Inform. Tech. – Coding of Audio-Visual Objects – Part 10: Advanced Video Coding* (2007).

[7] Kamaci, N., Altunbasak, Y., Mersereau, R.M.: "Frame bit allocation for the H.264 AVC video coder via Cauchy-density-based rate and distortion models", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 8, pp. 994 – 1006 (2005).

[8] 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, document JVT-G012r1*, Pattaya, Thailand (2003).

[9] 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*, vol. 22, no. 1, pp. 39-58, Elsevier (2007).

[10] Lin, G., Zheng, S., Hu, J.: "A two-stage ρ-domain Rate control scheme for H.264 encoder", *IEEE International Conference on Multimedia and Expo,* ICME, pp. 713-716, Hannover, Germany (2008).

[11] Łuczak, A.: *Digital video bitrate control using coder models*, Doctoral Dissertation, Poznań (2003).

[12] Wan, S., Izquierdo, E.: "A new rate distortion model for rate control of video coding", *7th International Workshop on Image Analysis for Multimedia Interactive Services*, WIAMIS, Incheon, S.Korea (2006).

[13] Xie, Z., Liu, J.: "Simple and efficient rate control for H.264/AVC", *International Symposium on Intelligent Signal Processing and Communication Systems*, ISPACS, pp. 862 – 865, Xiamen, China (2007).

[14] 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).

[15] http://iphome.hhi.de/suehring/tml/index.htm