CONSTANT BIT-RATE CONTROL EFFICIENCY WITH FAST MOTION ESTIMATION IN H.264/AVC VIDEO CODING STANDARD

*Daniele Alfonso, Daniele Bagni, Luca Celetto and Simone Milani**

Advanced System Technology labs, STMicroelectronics, Agrate Brianza, Italy (Europe)
* Department of Information Engineering, University of Padova, Italy (Europe)
emails: daniele.alfonso@st.com, simone.milani@dei.unipd.it

ABSTRACT

The emerging H.264/AVC video coding standard provides significant enhancements in compression efficiency with respect to its ancestors in the MPEG and H.26x families. In this paper, we analyse the performance of two different Constant Bit-Rate control methods, suitable for H.264/AVC encoding at SD-TV resolution, where the motion estimation is performed by a fast proprietary predictive algorithm.

1. INTRODUCTION

The H.264/AVC video coding standard [1] greatly outperforms its ancestors in the MPEG and H.26x families, providing about 50% more compression at equal quality [2, 3] in all kinds of applications, from low-bitrate streaming to high-quality storage.

Until now, the performance of H.264/AVC was analysed mostly in an ideal environment, with pure variable bitrate (at fixed quantization values) and using the common Full-Search Block-Matching (FSBM) algorithm for motion estimation. In practical applications, however, we need to impose some constraints on the encoding process, i.e. on the number of bits generated per time unit in case of limited bandwidth, and on the overall computational complexity for real-time encoding. The former goal can be obtained by a rate control algorithm, which adapts the quantization of the residual coding in order to match a target bitrate, whereas the computational complexity can be reduced by adopting a fast algorithm to perform motion estimation, which is one of the most resource consuming tasks in the whole encoding process.

Indeed, the relationship between rate control and motion estimation is an important topic of investigation, because the two systems interact with each other, determining the overall performance of the encoder. In fact, if the motion estimation is not efficient, the prediction error will be greater, forcing the rate control to raise the quantization step, in the attempt to match the target bitrate. This will negatively affect both the achieved quality and the motion estimation of successive frames, as the motion search is performed referencing the reconstructed frames, which are less correlated to the current one due to the coarser quantization.

In this paper, we consider two Constant Bit-Rate (CBR) control algorithms, named JVT-D030/E069 and *ρ-domain*, while using a proprietary motion estimation technique that greatly reduces the computation with respect to FSBM without visible quality losses at SD-TV resolution.

The JVT-D030/E069 and *ρ-domain* algorithms are presented in sections 2 and 3 respectively, whereas section 4 introduces the motion estimation method. The numerical results are showed in section 5, followed by our conclusions in section 6.

2. JVT-D030/E069 RATE CONTROL

The JVT-D030/E069 [4, 5] is a CBR control method, developed from the widely known TM5 [6], originally conceived for MPEG-2 video coding. The main difference between the two algorithms consists in the definition of luminance macroblock activity, which is the Sum of Absolute Differences after prediction (either Intra or Inter) for JVT-D030/E069, and variance for TM5.

Given a target bitrate, the JVT algorithm determines the *QP* quantizer for each macroblock operating at three levels: Group-Of-Pictures (GOP), frame/field picture and macroblock. In comparison with the original version [4, 5], we changed the experimental constants, in order to optimally adapt the algorithm to SD-TV video formats.

2.1 Rate control at GOP level

The target number of bits for each GOP is

$$R = N \cdot \frac{BitRate}{PictureRate} + R_{prev} \tag{1}$$

where N is the number of frames in the GOP, i.e. the distance between two Intra coded pictures, and R_{prev} is the number of exceeding bits after the encoding of the previous GOP, which is initialised to zero for the first GOP of the sequence.

2.2 Rate control at frame/field level

At the beginning of each frame or field (respectively in case of progressive or interlaced source video), the target number of bits for each I, P and B picture is computed as in the following:

$$T_i = \max\left[R \Big/ \left(1 + \frac{N_p X_p}{K_p X_i} + \frac{N_b X_b}{K_b X_i} \right), \frac{BitRate}{8 \times PictureRate} \right]$$
$$T_p = \max\left[R \Big/ \left(N_p + \frac{N_b K_p X_b}{K_b X_p} \right), \frac{BitRate}{8 \times PictureRate} \right] \tag{2}$$

$$T_b = \max\left[R/\left(N_b + \frac{N_p K_b X_p}{K_p X_b} \right), \frac{BitRate}{8 \times PictureRate} \right]$$

where R is defined in (1) and N_p and N_b are the number of remaining P and B frames or fields in the GOP, respectively.

After the encoding of each frame, R is updated as

$$R = R - S$$

where S is the number of bits used to encode the current frame, and N_p and N_b are decremented by one if the current frame or field was of P- or B-type respectively.

X_i, X_p and X_b define the content complexity of the different picture types. They are initialised with

$$X_i = (155 \cdot BitRate) / 115$$
$$X_p = (100 \cdot BitRate) / 115$$
$$X_b = 0.9 \cdot X_p$$

and after the encoding of each frame they are updated as

$$X_{\{i,p,b\}} = \frac{1}{2} \cdot S \cdot QP_{avg}$$

where QP_{avg} is the average quantizer for the current picture.

Finally, K_p=1.1 and K_b=1.4 define the relative complexity of I pictures with respect to P and B ones.

2.3 Rate control at macroblock level

Before encoding each macroblock, an initial quantizer is chosen according to the following formula

$$Q_m = \left(\frac{d_m^n \times 31}{r} \right) + dq$$

where r is a constant called *reaction parameter*, defined as

$$r = 10 \cdot \frac{bit\_rate}{frame\_rate}$$

dq is named *delta parameter* and it depends from the activity of the current macroblock in the following way

$$dq = \begin{cases} -floor(AvgAct / act_m - 1), & 0 < act_m / AvgAct <= 1/2 \\ 0, & 1/2 < act_m / AvgAct < 2 \\ floor(act_m / AvgAct) - 1, & act_m / AvgAct >= 2 \end{cases}$$

$AvgAct$ is the average activity for the current picture and act_m the activity of the current *m-th* macroblock. At the beginning of the sequence we set $AvgAct_I$=2000, $AvgAct_P$=1500 and $AvgAct_B$=800.

We would remark that the Intra/Inter prediction could be performed twice for each macroblock in JVT-D030/E069. In a first step, we derive the activity using as temporary QP the one of the previously encoded macroblock. Hence, we encode the macroblock with the final QP computed by the rate control, when it differs from the temporary one. Experimentally we found that every macroblock is Intra/Inter predicted about 1.25 times, on the average, for target bitrates in the ranges from 2 to 7 Mbit/s at SD-TV resolution.

The d_m parameter indicates the occupation of the virtual buffer for picture type n=i,p,b, and it is specified as

$$d_m^n = d_0^n + B_{m-1} - T_n \times (m-1) / MB\_CNT$$

where:

- d_m^i, d_m^p, and d_m^b represent the fullness of virtual buffers at macroblock m for each picture type. The final fullness is used as initial value d_0^i, d_0^p, and d_0^b for the next frame;
- d_0^n is the initial virtual buffer occupancy, respectively set as $d_0^i = 20 \cdot r/31$, $d_0^p = K_p \cdot d_0^i$ and $d_0^b = K_b \cdot d_0^i$ for I, P and B pictures at the beginning of the encoding process;
- B_{m-1} is the number of bits generated by encoding the first *m-1* macroblocks in the picture (composed of *MB_CNT* macroblocks in total).

3. ρ-DOMAIN RATE CONTROL

Traditional rate control algorithms, as also the JVT-D030/E069, operate in the so-called *q-domain*, where rate and distortion characteristic curves of the encoder are determined as a function of the quantization step. However, an alternative is to consider them as functions of the percentage of null quantized transform coefficients, indicated by ρ. This is named *ρ-domain* analysis, and an efficient rate control method exploiting this theory was presented in [7] and properly adapted to H.264/AVC in [8].

The number of bits R for each picture can be expressed as a function of ρ through the relation

$$R(\rho) = \lambda_1 \cdot \rho + \lambda_2 \quad (3)$$

With the approximation that no bits are sent if all coefficients are null, equation (3) becomes

$$R(\rho) = \lambda \cdot (1 - \rho) \quad (4)$$

Given a target bit budget T_n for the n-th frame, the corresponding ρ_n can be determined from (4) as

$$\rho_n = 1 - \frac{T_n}{\lambda} \quad (5)$$

The parameter λ can be estimated from the results of the encoding of previous frames, whereas the percentage of null coefficients ρ can be written as a function of the quantizer QP in the following way

$$\rho(\Delta) = \sum_{|QP|<\Delta} p_x(QP) \quad (6)$$

where $p_x(q)$ is the probability distribution of transform coefficients x and [-Δ,+Δ] is the quantization step interval associated with the null reconstructed value. For the probability distribution, we empirically have chosen a laplacian-impulsive function defined as

$$p_x(q) = \alpha \cdot \delta(q) + e^{-\frac{2}{\beta}|q|} \quad (7)$$

being $\delta(q)$ the Dirac impulse.

Writing (6) in integral form and considering (7) we obtain

$$\rho(\Delta) = \frac{\alpha}{1+\alpha} + \frac{1}{1+\alpha} \cdot \left(1 - e^{-\frac{2}{\beta}\Delta} \right)$$

from which we have

$$\Delta = -\frac{\beta}{2} \ln\left(1 + \alpha - \rho_n (1+\alpha)\right) \quad (8)$$

The coefficients α and β can be accurately approximated by relating them to the average activity of the picture in the polynomial form:

$$\alpha(AvgAct) = \alpha_0 + \alpha_1 \cdot \log(AvgAct) + \alpha_2 \cdot \log\log(AvgAct)$$
$$\beta(AvgAct) = \beta_0 + \beta_1 \cdot AvgAct + \beta_2 \cdot (AvgAct)^2$$

and then tabulated.

As also the JVT-D030/E069 algorithm, the *ρ-domain* rate control operates at GOP, picture and macroblock levels.

At GOP level, the behaviour is identical to what explained in section 2.1, i.e. a certain number of bits is assigned as a budget to the current GOP, taking into account the exceeding bits spent for the previous one.

At picture level, a different amount of bits T_n is assigned in a way similar to (2), depending on relative complexities defined for I, P and B images. Given the target bits, equation (5) is used to determinate the corresponding ρ, which is then substituted in (8), thus obtaining a target average quantization step for the current frame (or field).

At macroblock level, the quantization step is corrected considering that, after coding the m-th macroblock of the n-th frame, the percentage of null quantized coefficients in the previous coded m macroblocks is ρ_m^P and the number of bits used to code the picture is B_m^P. According to the given target, $B_m^R = T_n - B_m^P$ bits are left to code the remaining blocks, and the required percentage of null coefficients is

$$\rho_m^R = 1 - \frac{B_m^R}{\lambda} \cdot \frac{MB\_CNT}{MB\_CNT - m}$$

From that, it is possible to determinate the ratio $k = \rho_m^R / \rho_m^P$ and to compute QP_{m+1}, as proved in [8]. The value QP_{m+1} can be clipped depending on specific thresholds and on the previous QP_m value, to avoid different coding quality between adjacent macroblocks, which would produce unpleasant visual effects.

4. FAST PREDICTIVE MOTION ESTIMATION

The classical FSBM method exhaustively evaluates all possible motion vector candidates within a predefined search window area. For a practical, real-time implementation, it is far too complex and expensive, therefore faster methods are absolutely needed. The class of spatial-temporal correlation techniques proved its great efficiency in our previous experience with MPEG video compression [9, 10] and it can be considered an excellent alternative to the FSBM [11].

The proposed predictive-recursive algorithm is expressively conceived for high-quality H.264/AVC video coding and operates in two steps. The first one is the identification of the best motion vector within a set of candidates, chosen among the available spatially and temporally correlated motion vectors, exploiting the results of the motion estimation already performed on previous macroblocks. To reduce the computation, only the nearest spatial and temporal vectors are tested, as shown in Figure 1. Once the best vector is selected, a refinement step is then applied by adding fixed updates in search for the optimal motion vector. These updates depend adaptively on the motion activity of the encoded video sequence. Furthermore, the total number of vectors actually tested is fixed and predetermined for each macroblock, regardless of the search window range, which can therefore be set as large as the entire picture. On the other hand, either software or hardware implementations of FSBM require limiting its search area, respectively to reduce simulation times or silicon area to reasonable, cost-effective levels.

Both candidate selection steps are part of two interleaved motion searches, called Coarse and Fine search. The former proceeds in the picture display order and its results are exploited by the latter, which proceeds in the usual picture coding order. In this way, the proposed algorithm is able to obtain the same image quality (measured in Peak Signal to Noise Ratio, PSNR) of the FSBM with maximum loss in compression efficiency of 3% and by using only 1% of the computation required by FSBM for a typical search range of ±64 pixels, at any bitrate. Table 1 concisely shows a performance comparison between FSBM and our approach.

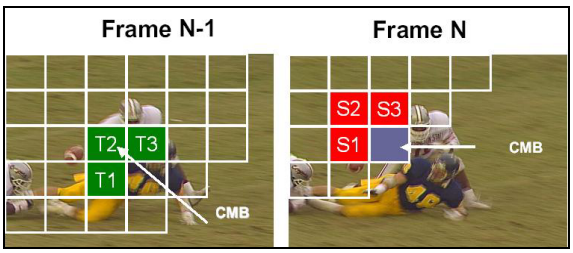

Figure 1: spatial (S1, S2, S3) and temporal (T1, T2, T3) motion vector candidates in the Coarse search step. CMB is the current macroblock under estimation.

Sequence name	FSBM stream size [kB]	Bitrate increase	Y PSNR loss [dB]
calendar	2 076	0.434%	-0.030
fball	3 390	2.697%	-0.015
renata	1 894	1.153%	-0.018
stefan	4 540	-2.372%	-0.045

Table 1: performance of proposed motion estimation algorithm with respect to FSBM at fixed QP = 31, for well-known SD-TV sequences. We used the same FSBM of the JVT reference SW encoder version 6.1e, with Hadamard transform, Loop Filter, CABAC, Multi-frame prediction with three references, no Rate-Distortion optimization.

5. EXPERIMENTAL RESULTS

The performance of JVT-D030/E069 and *ρ-domain* rate controls with the proposed fast motion estimation algorithm was evaluated by a proprietary H.264/AVC encoder, developed by the authors and compatible with the JVT reference software decoder version 6.1e.

We considered many different SD-TV sequences, in both NTSC (720×480 pixels, 30Hz) and PAL (720×576 pixels, 25Hz) interlaced formats, imposing target bitrates from 2 Mbit/s to 7 Mbit/s, with a step-size of 1 Mbit/s. The following parameters are common to all simulations: GOP length of 12 (PAL) or 15 (NTSC), 2 B pictures, Hadamard transform applied in motion estimation, Rate-Distortion optimisation and Multi-frame prediction not applied, search range of ±32 pixels, CABAC entropy coding.

Figure 2, 3 and 4 show the rate-distortion curves for three SD-TV sequences. Both rate controllers exhibit good behaviour in terms of quality and accuracy (shown in Table 2), which induce us to consider them as good references for evaluating other new rate control methods. The JVT D030/E069 is very reliable, but sometimes requires encoding twice the same macroblock, which makes expensive its im-

plementation for real-time, consumer-electronics devices such as DVD recorders. On the other hand, the *ρ-domain* offers nearly the same accuracy of JVT-D030/E069 with reduced computational complexity (from 15% to 25% lower in our experiments) and therefore it is more suitable for real-time application, although its dependence on pre-computed coefficients α and β can limit its reliability for particular conditions.

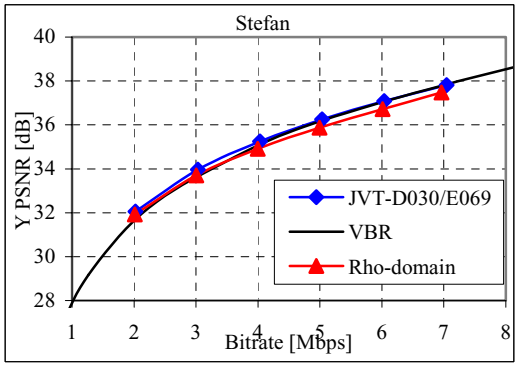

Figure 2: rate-distortion diagram for the "Stefan" sequence.

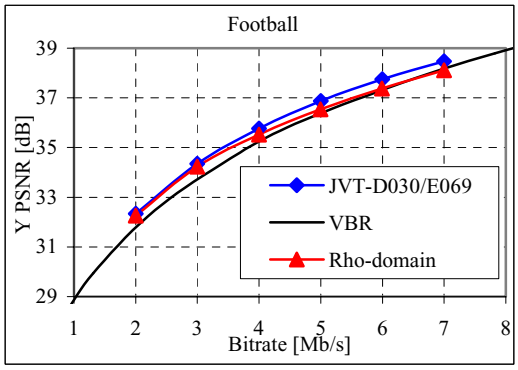

Figure 3: rate-distortion diagram for the "Football" sequence.

6. CONCLUSIONS

In this paper, we showed two CBR control methods, JVT-D030/E069 and *ρ-domain*, applied to H.264/AVC compression at SD-TV resolution. They exhibit good performance in terms of rate-distortion analysis when the video encoder applies a novel fast motion estimation algorithm.

Future activities concern the development of new CBR and VBR control algorithms, jointly to the proposed fast predictive motion estimation, for DVD and PVR applications at SD-TV and HD-TV resolution.

Sequence	JVT-D030/E069	ρ-domain
calendar	-0.203%	0.773%
fball	-0.002%	-0.038%
renata	-0.005%	-0.001%
stefan	0.806%	0.028%

Table 2: target bitrate accuracy error for JVT-D030/E069 and *ρ-domain* algorithms, averaged on all experimented bitrates.

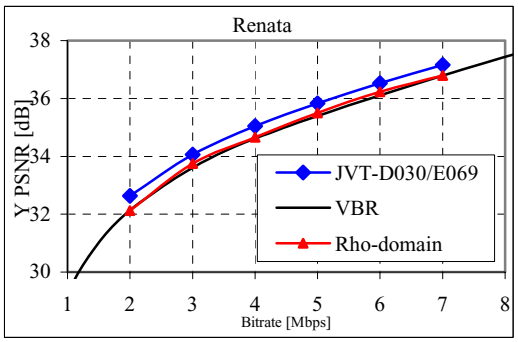

Figure 4: rate-distortion diagram for the "Renata" sequence.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Luca Pezzoni from STMicroelectronics and professor Gian Antonio Mian from University of Padova for their valuable contributions. The work was partially carried out within the FIRB Project of the Italian Ministry of Education, University and Research.

REFERENCES

[1] T. Wiegand, G. J. Sullivan, "Overview of the H.264/AVC video coding standard", *IEEE Tr. on C.S.V.T.*, pp. 560-576, vol. 13, n. 7, July 2003.

[2] D. Alfonso, D. Bagni, A. Chimienti, D. Pau, "A performance analysis of H.264 video coding standard", in *Proc. PCS 2003,* Saint-Malo, France, April 2003.

[3] D. Alfonso, D. Bagni, L. Celetto, L. Pezzoni, "Detailed rate-distortion analysis of H.264 video coding standard and comparison to MPEG-2/4", in *Proc. VCIP 2003,* Lugano, Switzerland, July 2003.

[4] S. Ma, W. Gao, Y. Lu, "Rate control in JVT standard", *JVT-D030*, July 15, 2002.

[5] S. Ma, W. Gao, Y. Lu, D. Zhao, "Improved rate control algorithm", *JVT-E069*, October 3, 2002.

[6] ISO/IEC JTC1/SC29/WG11, Test Model 5, April 1993.

[7] Z. He, Y. K. Kim, S. K. Mitra, "Low-delay rate control for DCT video coding via *ρ*-domain source modeling", *IEEE Tr. on C.S.V.T.,* vol. 11, n. 8, August 2001.

[8] S. Milani, L. Celetto, G. A. Mian, "A rate control algorithm for the H.264 encoder", in *Proc. Sixth Baiona Workshop on Signal Processing in Communications*, Baiona, Spain, September 2003.

[9] F. Rovati, D. Pau, E. Piccinelli, L. Pezzoni, J-M. Bard, "An innovative, high quality and search window independent motion estimation algorithm and architecture for MPEG-2 encoding", *IEEE Tr. on C.E.*, vol. 46, no. 3, pp 697-705, August 2000.

[10] D. Alfonso, F. Rovati, L. Celetto, D. Pau, "An innovative, programmable architecture for ultra-low power motion estimation in reduced memory MPEG-4 encoder", *IEEE Tr. on C.E.*, vol. 48, no. 3, pp.702-708, Aug. 2002.

[11] H. Y. C. Tourapis, A. M. Tourapis, P. Topiwala, "Fast motion estimation within the JVT codec", *JVT-E023*, October 9, 2002.