One-pass Bitrate Control for MPEG-4 Scalable Video Coding using rho-domain

Yohann Pitrey, Marie Babel, Olivier Déforges

To cite this version:

Yohann Pitrey, Marie Babel, Olivier Déforges. One-pass Bitrate Control for MPEG-4 Scalable Video Coding using rho-domain. International Symposium on Broadband Multimedia Systems and Broadcasting, May 2009, Bilbao, Spain. pp.1-4, 2009. <hal-00392042>

HAL Id: hal-00392042

https://hal.archives-ouvertes.fr/hal-00392042

Submitted on 5 Jun 2009

One-pass bitrate control for MPEG-4 Scalable Video Coding using ρ-domain

Yohann Pitrey, Marie Babel, Olivier Déforges

Institute of Electronics and Telecommunications of Rennes (IETR), CNRS UMR 6164
Image and Remote Sensing Group
INSA de Rennes, 20 av. des Buttes de Coesmes, 35043 Rennes, FRANCE.
Email: {yohann.pitrey,marie.babel,olivier.deforges}@insa-rennes.fr

***Abstract*—This paper presents an attractive rate control scheme for the new MPEG-4 Scalable Video Coding standard. Our scheme enables us to control the bitrate at the output of the encoder on each video layer with great accuracy. Each frame is encoded only once, so that the computational complexity of the whole scheme is very low. The three spatial, temporal and quality scalabilities are handled correctly, as well as inter layer prediction and hierarchical B frames. A linear bitrate model is used to predict the output bitrate for a frame, based on a simple and effective framework called ρ-domain. A coding-complexity measure is also introduced to dispatch the available bits among the frames, in order to reach a constant quality throughout the encoded video stream. To attest the performances of our rate control scheme, we present comprehensive results on some representative scalable video set-ups.**

I. INTRODUCTION

Scalable Video Coding was designed as a response to the growing need for flexibility in video transmissions over current networks and channels. The recently finalized MPEG-4 Scalable Video Coding (SVC) [1] standard is based on MPEG-4 AVC/H.264 and provides spatial, temporal and quality scalabilities. Spatial scalability acts on the frame resolution, and addresses variable screen sizes. Temporal scalability increases the number of frames per second, improving the motion smoothness. Quality scalability increases the signal-to-noise-ratio (SNR), to provide adjustable quality in the decoded video stream. The standard also provides spatial and temporal inter layer prediction, to exploit the redundancies between layers and enhance the coding efficiency on the whole scalable encoded stream.

Rate control is a critical part of the encoding process, as it regulates the bitrate at the output of the encoder and alleviates the problems caused by bitrate fluctuations. Meanwhile rate control has been widely studied for conventional video coding such as MPEG-4 AVC/H.264 [2]–[4], only few propositions were made for scalable video coding. In [5], each frame is encoded twice, which makes the encoding process complexity grow dramatically. The method presented in [6] is quite attractive, as it exploits the dependencies between layers to perform more accurate rate control. Although, the algorithm remains quite complex and requires a lot of calculations. Besides, the tested configurations do not reflect practical SVC applications, such as presented in [7].

In this paper, we present a new rate control scheme for MPEG-4 SVC. A bitrate model based on the ρ-domain framework [4] is used to predict the bitrate before encoding a frame. This modeling framework is very accurate and has a quite low computational complexity. Besides, we choose to control the bitrate at the frame level to minimize the computation. Additionally, we use the statistics from the previous frame as a basis for the bitrate model, so that each frame is encoded only once. Thus, the computational complexity of the whole rate control process is extremely low. To get smooth quality variations in the decoded stream, a relative coding-complexity measure is also used to dispatch the available bitrate inside a group-of-pictures (GOP).

This paper is organized as follows. Section II presents the rate model used to predict the bitrate before encoding a frame and the rate control scheme that is built around it. Section III presents some experimental results on representative scalable configurations. Section IV concludes the paper.

II. PROPOSED RATE CONTROL SCHEME

The purpose of rate control is to regulate the bitrate at the output of the encoder so that it copes with a given communication channel bandwidth. In scalable video coding, each layer is generally intended to be transmitted on a specific channel. In our rate control scheme, we specify a bits-per-second constraint C_l for each scalable layer l. This bits-per-second constraint is first converted to a bits-per-GOP budget.

A. GOP budget allocation

Each GOP gets the same amount of bits, according to the bits-per-second constraint specified for the layer. The GOP budget G_l is processed such as

$$G_l = S_l \times \frac{C_l}{F_l} + E, \tag{1}$$

where S_l is the size of a GOP in layer l and F_l is the number of frames per second in layer l. E is a small feedback term to correct the errors from previous GOPs (*i.e.:* $E < 10\%$ of the entire GOP budget). The GOP budget is then dispatched among the frames.

B. Frame budget allocation

To provide fair user experience, we try to achieve a constant quality throughout each GOP. Based on our previous work [8], we use a relative coding complexity measure to take each type of frame into account. MPEG-4 SVC supports several types of frames (*i.e.:* I, P and hierarchical B-frames). Each GOP starts with an I or P frame, followed by a set of hierarchical B frames [9]. Up to eight successive temporal levels of B frames are encoded using a pyramidal pattern. Each type of frame has a specific coding efficiency because of the coding tools it uses. For example, more bits are needed for a P-frame to get the same quality as a B-frame. Besides, each temporal level of B frames, denoted $B_1 \dots B_8$, has a specific coding efficiency, and it can be considered as a particular type of frame. To get a constant quality inside a GOP, we need to dispatch the bits according to the coding efficiencies of each frame. In [8], we define the coding complexity measure for a frame f such as

$$K_{T_f,l} = 2^{q/6} \times b_f, \tag{2}$$

where T_f is the frame type of f in layer l, q is the quantization parameter (QP) used for the frame and b_f is the number of bits needed to encode it. For each frame f, a target bitrate R_f is processed such as

$$R_f = \frac{K_{T_f,l}}{\sum_{f_i \in GOP} K_{T_{f_i},l}} \times G_l, \tag{3}$$

where $\sum_{f_i \in GOP}$ is the sum of the coding complexities of all frames in the current GOP.

C. QP processing

Once each frame has a target bitrate, the rate control scheme must choose the optimal value of QP to encode the frame. One of the main issues about rate control is to find a relationship between the QP and the output bitrate in order to predict the QP value that produces the closest match to the target bitrate. To find this relationship, we use the ρ-domain bitrate modeling framework.

Let ρ be the percentage of zero coefficients in a frame after quantization. As displayed in Figure II-C, it has been demonstrated that the output bitrate linearly decreases when ρ increases [4], [8]. Therefore, the relationship between ρ and the bitrate R can be formulated as

$$\rho(R) = \frac{R_0 - R \times (1 - \rho_0)}{R_0}, \tag{4}$$

where R_0 and ρ_0 are two initial values to be determined [10]. There also exists a relationship between ρ and the QP. Considering the transform coefficients of a frame, it is quite straightforward to know how many of them will be lost (*i.e.:* coded as zeros) during the quantization step. For a frame f, the relationship between ρ and and the QP q is given by

$$\rho(q) \;=\; \frac{1}{M} \sum_{m \in f} \left(\sum_{(i,j) \in m} z(c^m_{ij}, i, j, q) \right), \tag{5}$$

where M is the number of coefficients in the frame and $z(c^m_{ij}, i, j, q)$ is a function that indicates if the coefficient c^m_{ij} at position (i, j) in the macroblock m is under the deadzone threshold of the quantization scheme using the QP q [10]. Therefore, using ρ as an intermediate, we can establish a relationship between the bitrate and the QP. Given a target bitrate R_f, the target QP q_f is processed such as

$$q_f \;=\; \underset{q \in [0;51]}{\arg\min} \; |\rho(q) - \rho(R_f)| \,. \tag{6}$$

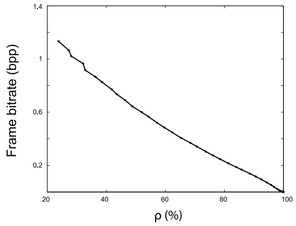

Fig. 1. Relationship between ρ and the bitrate.

To process equation (4) and (5), the rate model must be initialized. We need to know the values of ρ_0 and R_0 and to access the transform coefficients of the frame. In our previous work [8], we pre-encoded each frame to initialize the rate model. Unfortunately, this causes a substantial increase of the encoder complexity as each frame has to be encoded twice. In this paper, we use the information from the previous frame to initialize the rate model. Indeed, the correlations between the statistics of two adjacent frames are very high, which allows us to use the information from one frame as a basis to perform rate control on the next frame. In equation (4), the values of ρ_0 and R_0 are obtained from the last encoded frame f_p that has the same type as frame f in the same layer. Then, equation (5) is processed using the transform coefficients from f_p.

In the next section, we analyze the performances of our rate control scheme on some representative scalable configurations.

III. EXPERIMENTAL RESULTS

We will now demonstrate the accuracy of our rate control scheme on spatial, quality and combined temporal-quality scalabilities. Table I sums up the tested configurations. The encoded streams contain three layers using inter layer prediction. All our tests were performed using the JSVM Reference Software version 8.6 [11].

Table II reports the mean frame bitrate error for each layer in each type of scalability. As we can see, the error is below 5% of the allocated budget for all the tested configurations. Our rate control scheme thus matches the target bitrate very closely for each type of scalability. This is confirmed by figure 2, which shows the achieved bitrates per GOP on the three test sequences. It is important to notice that our rate control scheme represents less than 5% of the total encoding time. Thus, its impact on the encoding process is negligible in terms of computational complexity. Although, it allows us to control the bitrate with great accuracy on each type of scalability.

TABLE I
TEST SCENARIOS FOR EACH TYPE OF SCALABILITY.

		frame size	frame rate	budget (kbps)	GOP size
SPATIAL	base layer	QCIF	30	100	16
	enh. layer 1	CIF	30	400	16
	enh. layer 2	4CIF	30	1600	16
QUALITY	base layer	CIF	30	400	16
	enh. layer 1	CIF	30	800	16
	enh. layer 2	CIF	30	1200	16
TEMP+QUAL	base layer	CIF	15	200	4
	enh. layer 1	CIF	30	400	8
	enh. layer 2	CIF	60	800	16

The behavior of our frame budget allocation policy is illustrated by Figure 3. P frames are granted more bits than B frames, and B frames from low temporal levels get more bits than B frames from high temporal levels. As a result, the PSNR variations are decreased between frames. Figure 4 displays the frame PSNR for each type of scalability. We observe some small variations remaining below $2dB$, which is not quite noticeable. Moreover, the visual quality is greatly smoothed. Without our dispatching policy, we can observe unpleasant quality oscillations in the reconstructed video. We manage to attenuate these oscillations to a level that is hardly perceptible.

IV. CONCLUSION

In this paper, we introduce a new rate control scheme for MPEG-4 Scalable Video Coding. The proposed scheme is based on the ρ-domain model to predict the output bitrate for a frame, and uses the information from the previous frame so that each frame is encoded only once. We define a frame relative coding complexity measure to dispatch the available bits so that the PSNR variations are smooth. Our tests show that our scheme achieves very accurate rate control on each type of scalability with inter layer prediction. The mean bitrate error is below 5% and graphical results attest that the output bitrate matches the target bitrate very closely. The computational complexity of the entire scheme is very low as it is kept under 5% of the total encoding time. Our future work will focus on exploiting further the correlations between layers, by trying to predict the output bitrate in the enhancement layers from one frame in the base layer rather than from the previous frame.

REFERENCES

[1] J. Reichel, H. Schwarz, and M. Wien, “Scalable video coding - joint draft 4,” Joint Video Team, Doc. JVT-Q201, Tech. Rep., 2005.

[2] S. W. Ma, W. Gao, Y. Lu, and H. Q. Lu, “Proposed draft description of rate control on JVT standard,” Joint Video Team, Doc. JVT-F086, Tech. Rep., 2002.

[3] Z. Li, F. Pan, K. Lim, G. Feng, and X. Lin, “Adaptive basic unit layer rate control for JVT,” Joint Video Team, Doc. JVT-G012, Tech. Rep., 2003.

[4] Z. He and T. Chen, “Linear rate control for JVT video coding,” *Information Technology: Research and Education, International Conf. on*, pp. 65–68, 2003.

[5] L. Xu, S. Ma, D. Zhao, and W. Gao, “Rate control for scalable video model,” in *Visual Communications and Image Processing.*, vol. 5960, 2005, pp. 525–534.

[6] Y. Liu, Z. G. Li, and Y. C. Soh, “Rate control of h.264/AVC scalable extension,” *Circuits and Systems for Video Tech., IEEE Trans. on*, vol. 18, no. 1, pp. 116–121, 2008.

[7] ISO/IEC JTC1/SC29/WG11 MPEG2007/N9189, “Svc verification test plan, version 1,” Joint Video Team, Tech. Rep., 2007.

[8] Y. Pitrey, M. Babel, O. Déforges, and J. Vieron, “ρ-domain based rate control scheme for spatial, temporal and quality scalable video coding,” to be published in VCIP, SPIE Electronic Imaging.

[9] H. Schwarz, D. Marpe, and T. Wiegand, “Hierarchical b pictures,” Joint Video Team, Doc. JVT-P014, Tech. Rep., 2002.

[10] I. Shin, Y. Lee, and H. Park, “Rate control using linear rate-ρ model for h.264,” *Signal Processing - Image Communication*, vol. 4, pp. 341–352, 2004.

[11] JSVM Reference Software version 8.6, “http://ip.hhi.de/imagecom_G1/savce/downloads/.”

TABLE II
MEAN FRAME BITRATE ERROR FOR EACH TYPE OF SCALABILITY.

		SOCCER	**SOCCER**	**HOCKEY**
SPATIAL	base layer	0.28%	2.42%	0.31%
	enh. layer 1	0.43%	3.18%	0.23%
	enh. layer 2	0.78%	4.32%	1.28%
QUALITY	base layer	0.44%	3.66%	0.69%
	enh. layer 1	1.97%	3.96%	1.54%
	enh. layer 2	2.69%	4.51%	4.79%
	base layer	0.90%	3.15%	4.50%
	enh. layer 1	4.87%	1.80%	2.36%
	enh. layer 2	1.69%	4.66%	1.22%

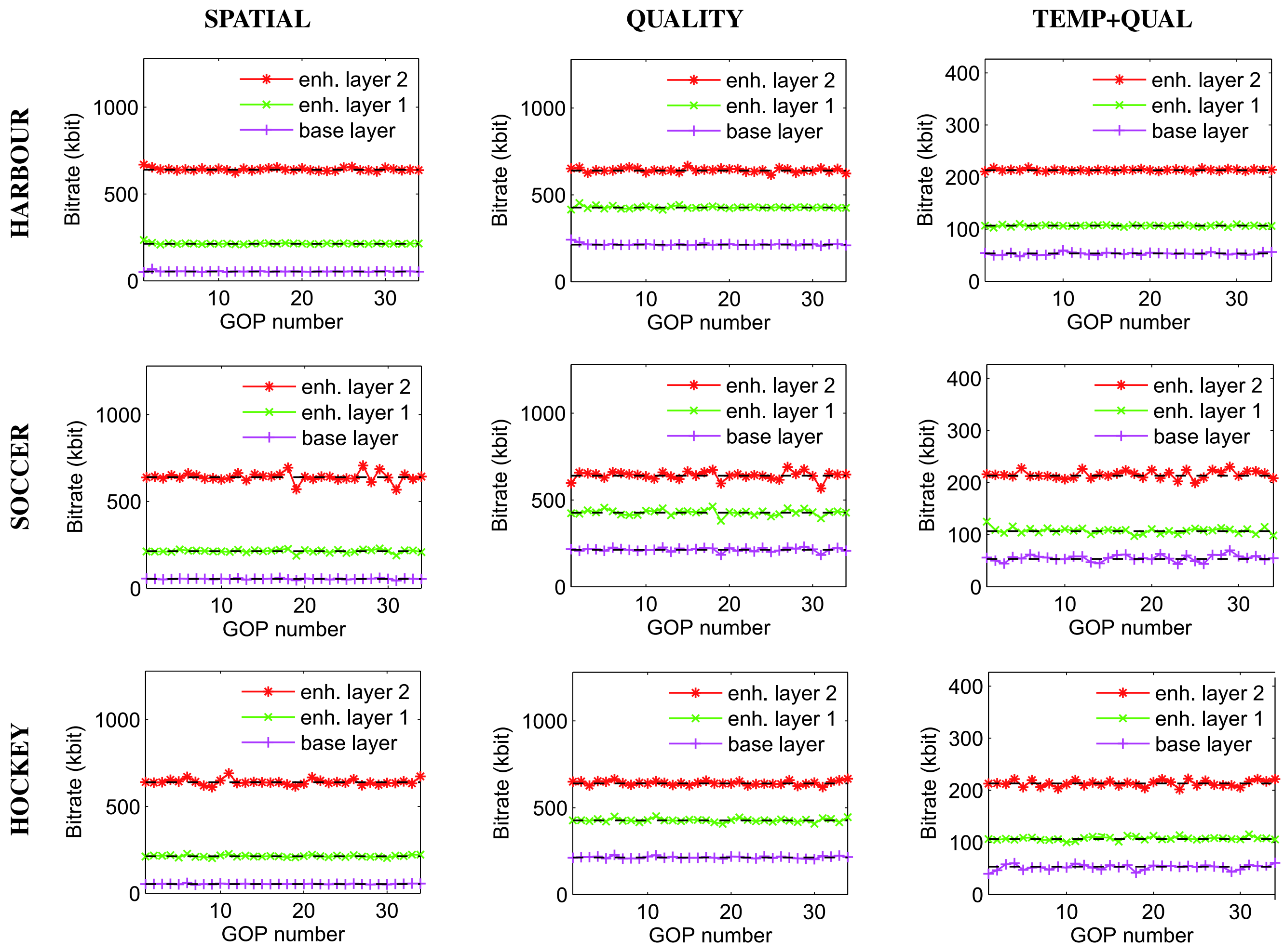

Fig. 2. Bitrates per GOP on each type of scalability.

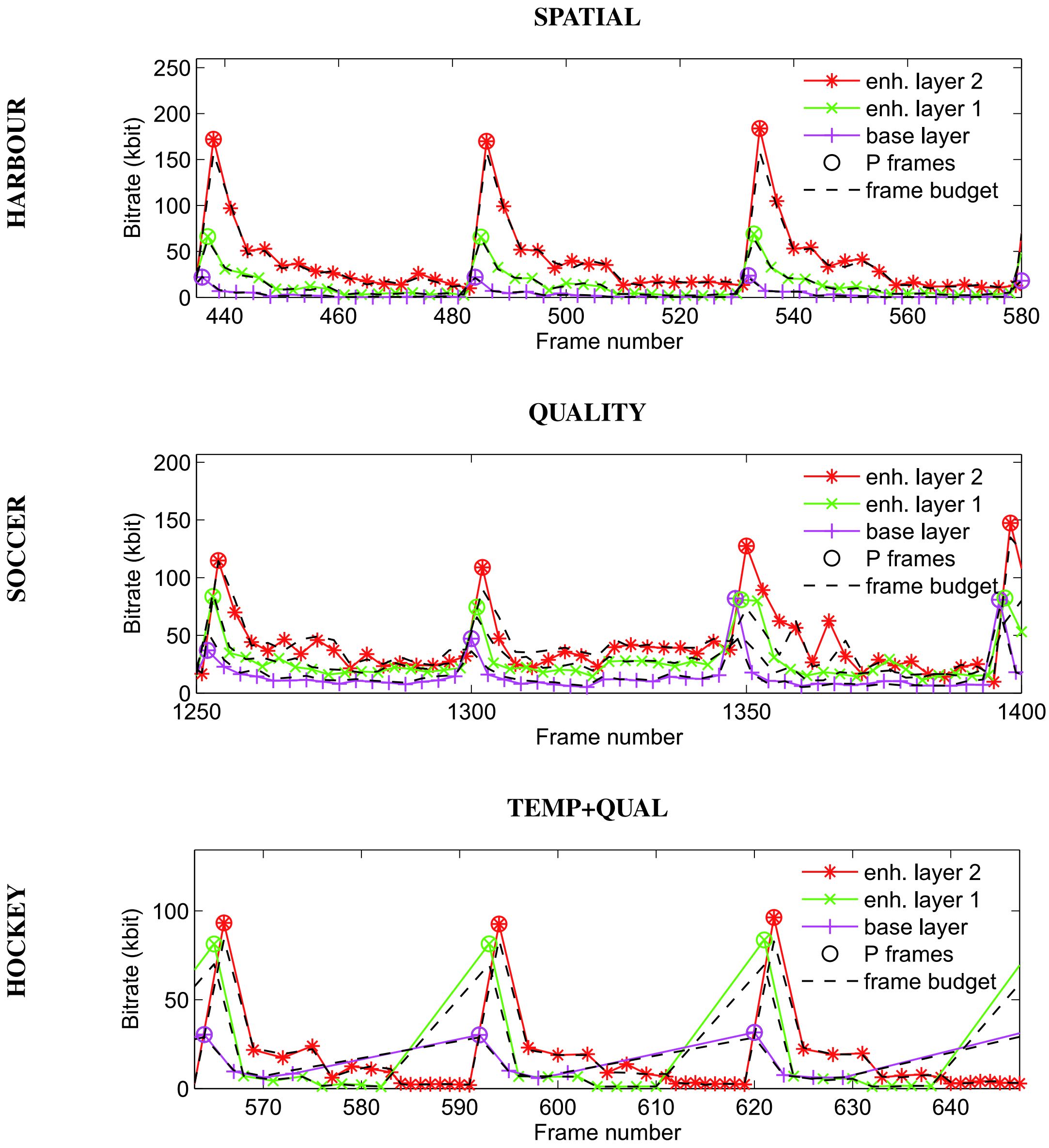

Fig. 3. Frame bitrates on each type of scalability.

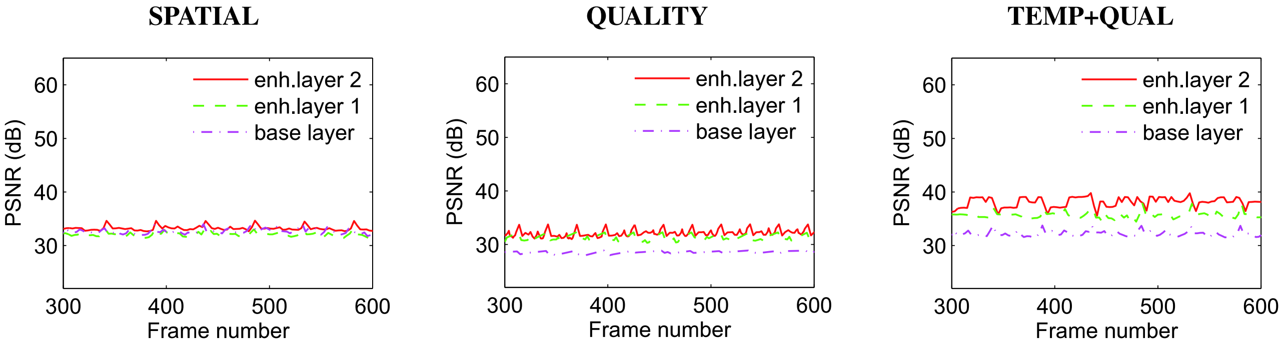

Fig. 4. Frame PSNR on each type of scalability using our bitrate dispatching policy for sequence HARBOUR.