Manuel de-Frutos-López, Óscar del-Ama-Esteban, Sergio Sanz-Rodríguez, Fernando Díaz-de-María

A two-level sliding-window VBR controller for real-time hierarchical video coding

Suggested Citation
de-Frutos-López, Manuel; del-Ama-Esteban, Óscar; Sanz-Rodríguez, Sergio; Díaz-de-María, Fernando: A two-level sliding-window VBR controller for real-time hierarchical video coding. - In: 2010 IEEE International Conference on Image Processing : ICIP. - New York, NY [u.a.] : IEEE, 2010. - ISBN: 978-1-4244-7992-4. - pp. 4217-4220. - DOI: 10.1109/ICIP.2010.5651371. *(Postprint version is cited, page numbers differ.)*

A TWO-LEVEL SLIDING-WINDOW VBR CONTROLLER FOR REAL-TIME HIERARCHICAL VIDEO CODING

Manuel de-Frutos-López, Óscar del-Ama-Esteban, Sergio Sanz-Rodríguez, Fernando Díaz-de-María

Department of Signal Theory and Communications
Universidad Carlos III, Leganés (Madrid), Spain

ABSTRACT

In this paper, a novel rate control algorithm for real-time VBR hierarchical video coding is proposed. The algorithm works at two levels that are called long- and short-term levels. The long-term level aims at ensuring that the bit count does not exceed the maximum allowed amount for a few-second long window. To this end, it considers a sliding window spanning several GOPs, which is shifted on a GOP basis. In doing so, it avoids the potentially sharp adjustments at the end of the GOP that usually happen in non-sliding approaches. The short-term level aims to provide a proper QP adaptation to fit the target bit budget, which is dictated by the long-term level. It also uses a sliding window, which in this case extends over one GOP.
The proposed algorithm has been assessed in realistic conditions for a variety of video sequences. It has been compared to both a constant quality and CBR hierarchical approaches, showing an excellent performance in terms of both rate-distortion and PSNR variation.

Index Terms— Rate control, VBR, sliding window, real-time

1. INTRODUCTION

Given the continuous development of new technologies for video information exchange, from the Internet to ad-hoc networks with wireless devices, as well as the traditional storage and editing applications, a number of video compression solutions, such as MPEG-2, MPEG-4 and H.264/AVC, have been developed during the last decades. Typically, the encoding techniques must be adapted to the available processing capabilities since the aforementioned scenarios set different complexity constraints. Aditionally, either the storage devices or the communication networks usually require a target output encoded bit rate. In order to deal with all these conditions, a rate control algorithm (RCA) is embedded in almost any video encoder, which generally aims to approach the target bit rate while maintaining quality changes that are as smooth as possible. In constant bit rate (CBR) scenarios, given the heterogeneous nature of video information, a leaky bucket paradigm is often used to bear the difference between the actual output bit rate and the nominal network rate. The RCA must control the buffer level in order to prevent overflow and underflow situations, as defined in the hypothetical reference decoder (HRD) description [1]. On the other hand, in variable bit rate (VBR) scenarios, a long-term bit rate adaptation is allowed to improve the visual quality consistency. In practice, VBR control has been applied in a wide range of applications with very different characteristics. For example, in video streaming for cell networks, the aim is to produce a consistent visual quality at the decoder since the network does not impose a constant bit rate, but rather a short-term limit to avoid network congestion. In contrast, in real-time storage applications, such as surveillance, the aim is to properly encode the whole sequence using the lowest possible bit rate and maintaining a consistent visual quality, without paying attention to instant bit rate limitations.
Not only does the quantizer play a key role in an RCA but also the group of pictures (GOP) pattern. For instance, hierarchical prediction structures can be used either to improve the coding efficiency in comparison to classical GOP patterns [2], or to provide temporal scalability in scalable video coding (SVC) applications, such as surveillance and transmission to heterogeneous clients with different display and computational capabilities [3]. Recently, some RCAs have been proposed for these hierarchical patterns [4, 5] and extended for SVC [6, 7]. Although they are capable of achieving a high coding efficiency by means of frame-wise bit allocation methods and rate-quantization (R-Q) models optimized for hierarchical B pictures, none of them has been designed to operate in a VBR environment, in which a higher perceptual quality is achieved.
In this paper, a novel two-level sliding-window RCA for real-time VBR scenarios is proposed, which combines both a short-term and a long-term control of rate variations, making it suitable for several different rate constraints. Although our proposal is designed for hierarchical video coding, it can also be applied to other classical GOP structures in a straightforward way.
The paper is organized as follows. In Section 2, a brief introduction to the relevant parameters in a VBR environment is provided. Section 3 describes the proposed algorithm. Results are shown and discussed in Section 4. Finally, some conclusions are drawn and future lines of research are outlined in Section 5.

2. RATE CONTROL IN VBR ENVIRONMENTS

Basically, the VBR environment can be described through the following parameters as stated in [8]:

- **Target bit rate** ($\mathbf{R_T}$): It is the average bit rate for the whole video sequence, i.e., the total bit count divided by the total sequence duration in seconds.
- **Maximum bit rate** ($\mathbf{R_M}$): In transmission applications, this parameter is related to the maximum buffer size. In storage applications, it could be identified with the access speed limit of the storage device.
- **Maximum exceeded bit count** ($\mathbf{MEBC}$): In transmission applications, it can be seen as a long-term restriction for the produced average bit rate when R_T is exceeded. In storage applications, it is directly the percentage that the average bit rate can be exceeded after encoding the sequence.

The RCA should perform a double-level control over the output bit rate, keeping both the long-term rate near R_T (or at least under $MEBC$) and the short-term rate under R_M, as well as maintaining a consistent subjective quality along the whole sequence.

3. TWO-LEVEL SLIDING-WINDOW VBR CONTROLLER

In general, analytic R-Q models for the quantization parameter (QP) estimation turn out to be inaccurate for non-stationary sequences since their reaction to changes in the video complexity is not quick enough, sometimes producing large quality variations in the encoded sequence. Since our proposal focuses on maintaining a smooth quality variation, it avoids the use of frame bit allocation and R-Q modeling. Instead, the proposed method pays attention to the mismatch between the expected and the produced amount of bits on a long-term (LT) basis, using a window of a few seconds (since the end of the sequence is unknown), and it determines the way the QP value should be modified on a short-term (ST) basis. The functionality of both long- and short-term levels is described in the following subsections.

3.1. Long-term layer

The LT window covers N GOPs and, as shown in Fig.1, it is shifted on a GOP basis; thus, there is an overlap of $(N-1)$ GOPs between two consecutive LT windows.

3.1.1. Bit bucket update

After encoding the i^{th} GOP, the algorithm determines whether or not the amount of bits $t_W(i)$ generated by the i^{th} window is within a predetermined range given by a lower and upper thresholds, $L_{TH}(i)$ and $U_{TH}(i)$, respectively:

$$D(i)=\begin{cases} L_{TH}(i)-t_W(i) & \text{if } L_{TH}(i)>t_W(i) \\ U_{TH}(i)-t_W(i) & \text{if } U_{TH}(i)<t_W(i) \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where $D(i)$ represents the amount of bits outside the range, which can be negative or positive, depending on whether it exceeds or goes below the considered range, respectively. This difference of bits is then distributed in the next N GOPs by means of an array $S_W(i)$ of bit buckets, which is updated as:

$$S_W(i+k)=S_W(i+k)+\frac{D(i)}{N^2} \qquad k=1,2,\ldots,N. \quad (2)$$

Thus, the content of these bit storage buckets adds an offset to the available bits for the following windows. The rationale behind this updating equation is illustrated in Fig. 1, for $N=3$. Assuming that the i^{th} GOP produces either an excess or a shortage of $D(i)=d$ bits, with $d\neq 0$, two effects should be account for. First, since the LT window moves forward on a GOP basis, the following N sliding windows are affected. Second, to smooth visual quality variations, the bit mismatch is distributed in the following N buckets. As a result, $D(i)$ should be divided by N^2.

3.1.2. Lower and upper threshold calculation

The lower threshold $L_{TH}(i)$ is set to the nominal amount of bits per LT window plus the bit mismatch produced by previously encoded LT windows. Additionally, a bound related to the maximum allowed bit budget is considered, leading to:

$$L_{TH}(i)=\min\left(\frac{NM}{f}R_M, \frac{NM}{f}R_T+\sum_{k=i-N-1}^{i} S_W(k)\right), \quad (3)$$

where M is the number of frames per GOP and f is the frame rate. Whenever the bound acts, the extra bit count with respect to the maximum allowed budget is distributed in the next N buckets.

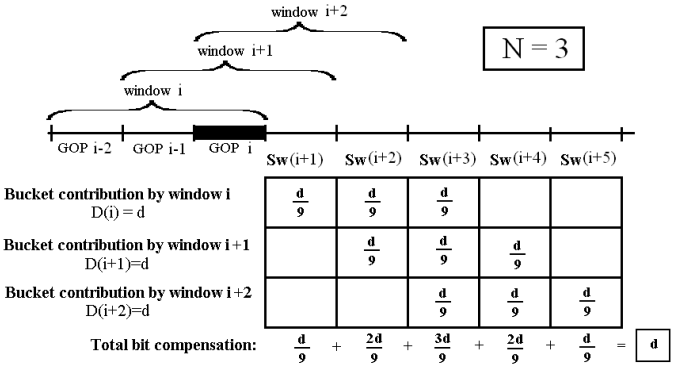

Fig. 1. Sliding window for long-term layer and bit bucket update.

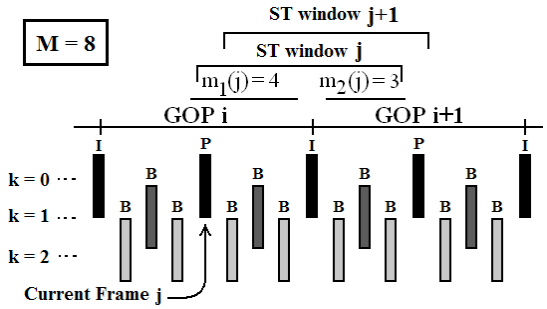

Fig. 2. Sliding window for short-term layer.

The upper threshold $U_{TH}(i)$ is determined from $L_{TH}(i)$ and $MEBC$ using the following expression:

$$U_{TH}(i)=\left(1+\frac{MEBC}{100}\right)L_{TH}(i). \quad (4)$$

Thus, since the end of the sequence is unknown, the upper threshold ensures that the $MEBC$ constraint is met at the end of each LT window.

3.2. Short-term layer

The ST layer focuses on predicting potential bit rate deviations with respect to R_T at GOP level. Neither bit allocation nor R-Q models are involved in the process. Similarly to the LT layer, a ST sliding window is used for the prediction, which shifts on a frame basis in order to evenly manage all the frames. As illustrated in Fig. 2, it begins at the current j^{th} picture and covers a total of M frames. The ST layer works in three phases: 1) sliding window bit budget calculation; 2) sliding window bit count prediction; 3) quantization parameter selection.

3.2.1. Sliding window bit budget calculation

The target bits $T_T(j)$ to encode all the frames in the j^{th} ST window depends on the average amount of bits to encode an inter picture in those GOPs spanned by the ST window, $T_{P/B}(i)$ and $T_{P/B}(i+1)$, respectively, as well as on a bit budget for the intra picture in the $(i+1)^{th}$ GOP, which is estimated by the number of bits generated by the last encoded intra picture, t_I. The final expression is as follows:

$$T_T(j)=T_{P/B}(i)m_1(j)+T_{P/B}(i+1)m_2(j)+t_I, \quad (5)$$

with

$$T_{P/B}(i) = \frac{R_T f^{-1} M + S_W(i) - t_I}{M-1}. \quad (6)$$

The parameters $m_1(j)$ and $m_2(j)$ are the number of inter pictures belonging to the respective GOPs covered by the j^{th} ST window which obviously sum $(M-1)$.

3.2.2. Sliding window bit count prediction

In this stage, a bit count prediction is computed for the M pictures belonging to the j^{th} ST window. Given the differences among temporal layers in rate-distortion (R-D) terms when hierarchical B pictures are employed for GOP encoding, a layered bit count prediction approach is proposed as follows:

$$\widetilde{T}_{SW}(j) = \widetilde{T}_I^0(j) + m_1^0(j)\widetilde{T}_P^0(j) + \sum_{k=1}^{K-1}\left(m_1^k(j)\widetilde{T}_B^k(j)\right) + \left(m_2^0(j)\widetilde{T}_P^0(j) + \sum_{k=1}^{K-1} m_2^k(j)\widetilde{T}_B^k(j)\right)\cdot\left(\frac{T_{P/B}(i+1)}{T_{P/B}(i)}\right). \quad (7)$$

The terms $m_1^k(j)$ and $m_2^k(j)$ represent the number of inter pictures belonging to the k^{th} temporal layer in the first and second GOPs, respectively. $\widetilde{T}_I^0(j)$ and $\widetilde{T}_P^0(j)$ are picture size predictors expressed as the number of bits corresponding to the lowest temporal layer ($k=0$). $\widetilde{T}_I^0(j)$ is set to t_I, while $\widetilde{T}_P^0(j)$ is calculated as an exponential average of the number of bits generated by the previously encoded P pictures. The predictor associated to the k^{th} temporal layer, $\widetilde{T}_B^k(j)$, is computed as an arithmetic average of the number of bits generated by a certain number of previously encoded pictures belonging to the same temporal layer. Finally, the ratio between $T_{P/B}(i+1)$ and $T_{P/B}(i)$ aims to consider the difference between target bits for consecutive GOPs.

3.2.3. Quantization parameter selection

The bit rate adaptation is carried out by means of a proper QP selection on a frame basis. Specifically, if the current j^{th} frame belonging to the k^{th} temporal layer is an inter picture, the ratio Φ between the bit count prediction for the j^{th} sliding window, $\widetilde{T}_{SW}(j)$, and its bit budget, $T_T(j)$, is used to update the corresponding QP value, $QP_{P/B}^k(j)$, as follows:

$$QP_{P/B}^k(j) = QP_{ref}^k + \Delta QP(\Phi). \quad (8)$$

The reference QP value, QP_{ref}^k, is set to the average QP of the $(k-1)^{th}$ layer in the current GOP. For $k=0$, QP_{ref}^0 is initialised to the QP value of last encoded P picture.

In order to determine an appropriate QP increment, $\Delta QP(\Phi)$, the ratio Φ is compared to a set of thresholds, given in Fig. 3, which has been experimentally designed by considering the influence of short-term QP variations on the average bit rate. Finally, an one-unit offset is added to $QP_{P/B}^k(j)$ for non-stored pictures.

Finally, in the case of intra pictures, the proposed VBR controller uses linear regression to obtain a first QP estimation, as suggested by the linear relation between QP and luminance peak signal-to-noise ratio (PSNR) [1]. Then, up to three more encodings of the picture around this value are tried, selecting the one keeping a more consistent quality and slightly enhancing the PSNR with respect to the last encoded frames. It should be noted that, in general, multi-encoding an intra frame is not computationally expensive when compared to inter pictures.

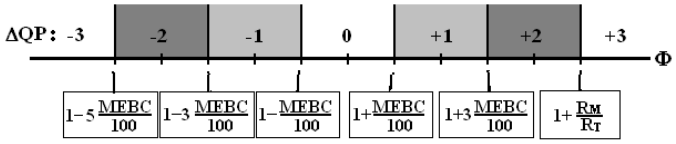

Fig. 3. Set of thresholds and corresponding QP increments.

4. ADDITIONAL IMPROVEMENTS

In order that the proposed VBR algorithm can work properly with real video sequences of different spatial resolutions, some additional improvements have been developed:

- **Scene cut detector:** the LT and ST windows are restarted when a scene cut happens and the bit budget is recalculated taking into account the amount of frames actually encoded. A simple scene cut detector based on the histogram difference of the original luminance component is used. Furthermore, the GOP structure is restarted when a scene cut is found.
- **Initial QP value:** the first QP value in a sequence as well as the first QP value after a scene cut are selected in a similar way than that described in the H.264/AVC rate control in [5], considering the target amount of bits per pixel. Furthermore, a set of new QP values has been experimentally found to improve the performance for standard definition (SD) resolution.

5. EXPERIMENTS AND RESULTS

The proposed VBR control algorithm was implemented on the Joint Video Team (JVT) reference software version JM 12.2 [9]. Its performance was tested using the following video sequences extracted from high-quality DVD disks: "Spiderman" (SP), "The Last Samurai" (LS), "The Patriot" (TP), "James Bond" (JB), and "Master and Commander" (MC). The resolution is 720×576 pixels and the length of the sequences is 1000 frames at 25 frames per second.

First, the reference software encoder was run for all the sequences and four different QP values: 25, 30, 35 and 40. The bit rates obtained were employed as target bit rates, R_T, by the CBR controller adopted by JVT [5] and the proposed VBR control algorithm. The CBR scheme used was RC_MODE_2, instead of RC_MODE_3, since this last one requires some "a priori" knowledge about the video content, which is not considered in this work. Furthermore, the RCA [5] was operated with a large enough buffer (some seconds). The rest of parameters and options are listed here:

- **GOP:** $M=24$ frames and $K=3$ temporal layers:1 I and 5 P frames ($k=0$); 6 B frames ($k=1$); 12 B frames ($k=2$).
- **R-D optimization:** enabled
- **Symbol mode:** CABAC
- **Motion estimation:** 5 reference frames and EPZS algorithm
- **VBR Parameters:** $R_M = 1.5 \times R_T$ kbps, $MEBC = 10\%$ and $N=10$ GOPs.

The average PSNR gain, $\overline{\Delta\mu_{PSNR}}$, for both CBR and VBR control schemes with respect to the fixed QP implementation, as well as the average PSNR standard deviation, $\overline{\sigma_{PSNR}}$, were some of the metrics employed to compare the algorithms. $\overline{\Delta\mu_{PSNR}}$ was obtained by interpolating the R-D curves for several bit rates and averaging the μ_{PSNR} differences. As shown by the performance curves in

Table 1. Average of the sequence PSNR gain with respect to fixed QP encoding, average PSNR standard deviation and MOVIE index.

Sequence	$\overline{\Delta\mu_{PSNR}}$ (dB)		$\overline{\sigma_{\mathbf{PSNR}}}$ (dB)			**MOVIE** ($\times\mathbf{10^3}$)		
	[5]	Prop.	F. QP	[5]	Prop.	F. QP	[5]	Prop.
SP	-0.19	-0.03	1.85	3.03	3.08	0.06	0.08	0.06
LS	-0.69	-0.17	1.98	4.17	3.60	0.86	0.66	0.59
TP	-0.35	0.44	2.01	5.36	3.87	0.20	0.40	0.32
JB	-0.04	0.36	1.42	4.02	3.63	0.31	0.40	0.31
MC	-1.54	0.38	2.81	5.31	5.71	0.68	0.58	0.72
Average	-0.56	0.19	2.01	4.37	3.98	0.42	0.43	0.40

Figs. 4 and 5, and the values in Table 1, the proposed RCA achieves a better performance in terms of μ_{PSNR} than both fixed QP and [5] solutions while providing an intermediate score between them in terms of quality consistency. The bit rate mismatch has an average value of 4.16%, higher than the obtained with [5], but it still fulfills the MEBC constraint. Furthermore, for each input sequence, the encoded bitstream with μ_{PSNR} close to 38 dB, our considered medium quality range, was selected for MOVIE index [10] calculation as an estimation of the subjective distortion. This measurement is based on multiple spatial-temporal consistency parameters taking into account the properties of the human visual system. The obtained values are also listed in Table 1 (a lower value means higher quality) and show that the proposed VBR controller achieves a good subjective performance when compared to the reference schemes.

6. CONCLUSIONS AND FURTHER WORK

A novel sliding-window-based VBR controller for hierarchical GOP patterns is described in this paper, which aims to produce a consistent visual quality in real-time applications. The proposed algorithm works at two different levels. The long-term level focuses on monitoring the MEBC constraint, acting on the bit budget for each GOP; while the short-term layer tries to prevent bit rate deviations from this budget keeping a smooth variation of QP. Our experimental results show that the proposed method improves the mean PSNR when compared to fixed QP encoding and the hierarchical CBR control algorithm [5] available in the reference software [9]. The good visual quality of this approach has also been guaranteed by the MOVIE index. Furthermore, the proposed RCA is flexible enough to work with other non-hierarchical GOP patterns or to be implemented on any video coding standard.
In order to take advantage of the good performance achieved by our proposal, a future work could be its adaptation and implementation on the SVC extension of the H.264/AVC standard.

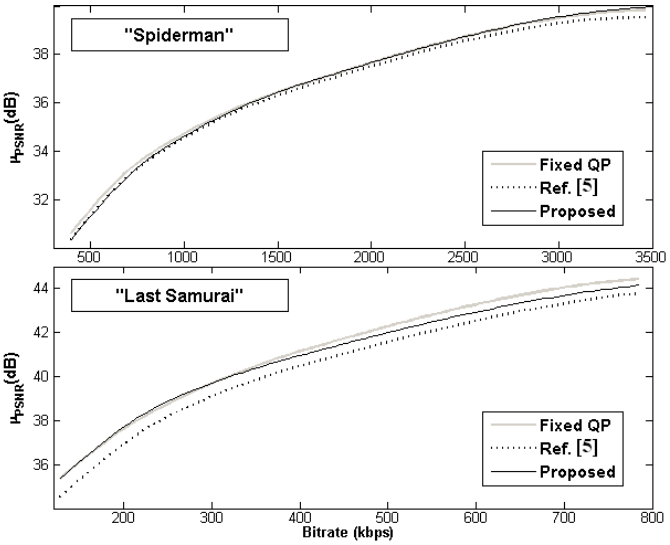

Fig. 4. PSNR vs. Bitrate curves for some sequences.

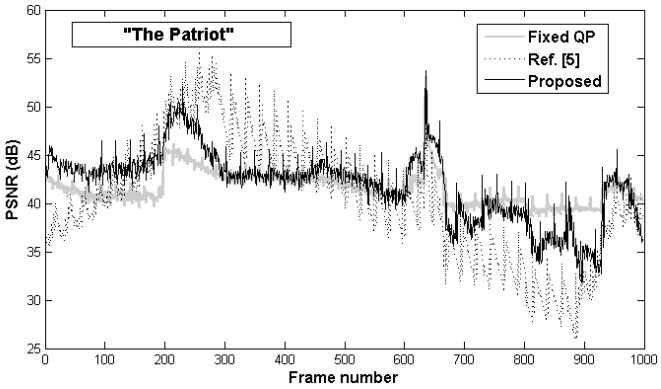

Fig. 5. PSNR evolution example.

7. REFERENCES

[1] S. Ma, Wen Gao, and Yan Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 15, no. 12, pp. 1533–1544, 2005.

[2] H. Schwarz, D. Marpe, and T. Wiegand, "Hierarchical B pictures," *JVT-P014, 16th JVT Meeting*, Poznan, Poland, Jul. 2005.

[3] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 9, pp. 1103–1120, Sept. 2007.

[4] L. Xu, W. Gao, X. Ji, and D. Zhao, "Rate control for hierarchical B-picture coding with scaling-factors," in *Circuits and Systems, 2007. ISCAS 2007*, 2007, pp. 49–52.

[5] A. Leontaris and A. Tourapis, "Rate control reorganization in the Joint Model (JM) reference software," *JVT-W042, 23th JVT Meeting*, San Jose, California, USA, Apr. 2007.

[6] L. Xu, W. Gao, X. Ji, D. Zhao, and S. Ma, "Rate control for spatial scalable coding in SVC," in *Picture Coding Symposium, 2007. PCS 2007*, Nov. 2007.

[7] Y. Liu, Z. G. Li, and Y. C. Soh, "Rate control of H.264/AVC scalable extension," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 18, no. 1, pp. 116–121, Jan. 2008.

[8] Y. Yokoyama and Y. Ooi, "A scene-adaptive one-pass variable bit rate video coding method for storage media," in *Image Processing, 1999. ICIP 99*, 1999, vol. 3, pp. 827–831 vol.3.

[9] Karsten Sühring, H.264/AVC software coordination, "http://iphome.hhi.de/suehring/tml/download/old_jm/," .

[10] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *Image Processing, IEEE Transactions on*, vol. 19, no. 2, pp. 335–350, Feb. 2010.