DYNAMIC BASIC UNIT SIZE IN RATE CONTROL FOR REAL-TIME H.264 VIDEO CODING

Sergio Sanz-Rodríguez, Darío García-García, Manuel de-Frutos-López, Jesús Cid-Sueiro

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

ABSTRACT

Several rate control (RC) schemes include the basic unit (BU) layer, where the quantization parameter (QP) value can be modified within a picture to get a fine adjustment to the target bits. The BU is a group of macroblocks (MBs) which share the same QP value, and its size is set previously to the encoding process. This paper describes a RC algorithm capable of detecting the instants in the sequence encoding process where a small BU size works efficiently and, for the rest of cases, use a large one to enhance quality. Our experimental results show a great robustness in both quality and buffer control for different kind of sequences and target bit rates.

Index Terms— H.264, rate control, basic unit, real-time, scene-change

1. INTRODUCTION

The emerging H.264 standard has become a reality in video coding since the rising of new video applications, like mobile video streaming and video over IP, demanding both high quality and low bit rates. H.264 achieves about the half rate than previous coding standards with no noticeable quality reduction.
A RC scheme must be defined for these new tools in order to comply with network requirements. Since the RC is not part of the standard itself, its design has become a major research area for codec designers and network administrators. A RC has been proposed for H.264 [1]. Other works have been proposed for low-delay environments. In [2] a novel bit allocation algorithm is defined and alternative rate-quantization (R-Q) models are described, as in [3]. Although some RC schemes, such as [4], employ a frame as the minimum unit for QP assignment, the aforementioned works use a BU layer in order to modify the QP value intra-frame. Taking into account the properties of small and large BU sizes, we propose a RC scheme that is able to detect the scenarios in which each proposed size achieves a high performance.
The rest of the paper is organized as follows. The section 1 compares the most common BU sizes. In section 2, the proposed RC are described. The experiments results are provided in section 3. Finally, this paper comments the conclusions and future work in section 4.

2. ABOUT BASIC UNIT SIZE

A BU is a group of MBs in raster scan order which share the same QP. The RC algorithm described in [1] allocates the picture target bits among BUs according to a complexity map obtained from the last encoded picture of the same type, improving intra-plane bit allocation. The number of MBs in a BU is set before the encoding process and it must be an entire fraction of the total number of MBs in a picture. Therefore, a lot of BU sizes are available, from one MB to an entire frame.

2.1. Basic unit properties

In stationary situations, where the sequence complexity is almost constant, any R-Q regression model for QP selection works properly, so a high performance can be achieved with a large BU size. Nevertheless, in scenarios with time-varying complexity, the model should be updated quickly to adapt to the changes of the complexity. An inappropriate adaptation could imply a target bit rate missadjustment and, therefore, an increment of the buffer overflow or underflow risk. So a small BU size is recommended in these situations.
The benefits of using small BU sizes are reduced in stationary scenarios, due to a possible overfitting to the picture target bits. The QP values for the last encoded BUs can suffer abrupt changes given that they are calculated based on the remaining bits for the picture. A high number of BUs also implies a computational cost increase due to the RC operations needed to obtain the QP for each BU, and an increased overhead since the QP variations must also be sent to the decoder. Therefore, the bits dedicated to encode texture information are reduced.
A frame is proposed as large BU size to enhance the average peak signal-noise ratio (PSNR) in stationary situations. Nevertheless, in order to get a good trade-off between PSNR and buffer level control in non-stationary scenarios, a row of MBs is a suitable option for small BU size. It shows similar performance than smaller BUs proposed in other works [2][3] reducing computational cost.

2.2. Dynamic basic unit

The BU size can be modified dynamically to take the advantages of both small and large proposed BU sizes: a row of MBs and a picture. This mechanism is named *dynamic BU size*. Although the large BU size works properly in most of situations, some scenarios in which a small BU size achieves a better performance have been identified: scene changes and buffer overflow and underflow risk.
In scene cuts with high complexity difference between both scenes, their appropriate QP ranges are also different. With small BU size, the necessary QP leap will be reached faster than with the largest one. On the other hand, the proximity to the overflow and underflow in the buffer occupancy is another cause of temporal QP heterogeneity. With small BU size, a fine adjustment to the target bits imposed by the bit allocation algorithm will be achieved.

3. PROPOSED RATE CONTROL ALGORITHM

The proposed RC based on dynamic BU size is shown in Fig. 1 for IP...P by simplicity, but it can be extended for any coding pattern. Its structure is similar to [1] where three levels are identified : GOP layer, picture layer and basic unit layer. In the following subsections these layers and other contributions, such as the scene change detector and frame skipping mechanism, will be described.

The j^{th} picture is encoded with ...	
small BU	**big BU**
$b_i(j) = \sum_{k=1}^{N_{BU}} b_{k,i}(j)$	$b_{k,i}(j) = \frac{b_i(j)}{N_{BU}}$
$\delta_i(j) = \frac{\sum_{k=1}^{N_{BU}} \delta_{k,i}(j)}{N_{BU}}$	$\delta_{k,i}(j) = \delta_i(j)$
$QP_i(j) = \left\lfloor \frac{\sum_{k=1}^{N_{BU}} QP_{k,i}(j)}{N_{BU}} + 0.5 \right\rfloor$	$QP_{k,i}(j) = QP_i(j)$ for all $k \in j$

Table 1. R-Q model update expressions for the non used BU size in the j^{th} picture.

3.1. General description

In the GOP layer, we need to update the encoder buffer level, $V_i(j)$, and the number of remaining bits, $T_r(j)$, for the rest of pictures in the i^{th} GOP after encoding the j^{th} picture. The quantization parameter for the intra (I) picture, $QP_i(1)$, is also obtained, and if a scene change is detected a new GOP begins (see subsection 3.2).
The picture layer computes the target bits for each P type picture, $T_i(j)$. The quantization parameter for the j^{th} picture in the i^{th} GOP, $QP_i(j)$, is then determined by using a quadratic R-Q model:

$$T_i(j) = c_1 \frac{\tilde{\delta}_i(j)}{Q_i(j)} + c_2 \frac{\tilde{\delta}_i(j)}{Q_i^2(j)} + H_i(j) \quad (1)$$

where c_1 and c_2 are coefficients, $Q_i(j)$ is the quantization step (related with $QP_i(j)$), $H_i(j)$ is the number of header and motion vector bits, and $\tilde{\delta}_i(j)$ is a prediction of the j^{th} mean absolute difference (MAD) between the original and prediction images in the i^{th} GOP by means of a linear model (see [1] for details).
After encoding each P picture, a linear regression is used to update the model coefficients. Nevertheless, in some RC algorithms [4][5] other R-Q models are used for I pictures as well.
The basic unit is an optional layer which allows selecting different QP values within a picture. Therefore, a fine adjustment to the target bits imposed by the aforementioned layer will be achieved. For the k^{th} BU in the j^{th} picture an amount $T_{k,i}(j)$ of target bits is estimated, and the $QP_{k,i}(j)$ value is obtained. Then, the R-Q model is updated when a BU is encoded.
In the proposed RC (see Fig. 1), the number of MBs in a picture is the default basic unit size, and its value will be changed to a smaller one using a QP variability criterion. The following situations have been considered:

- 1^{st} GOP of sequence ($i = 1$)
- 1^{st} GOP of each new scene
- $V_i(j)$ higher than Γ or lower than Θ

Thresholds Γ and Θ are set to $0.8 \times BS$, where BS is the buffer size, and an initial GOP target buffer level, TBL_0, respectively.
Note that in our scheme two different R-Q models like (1) are considered, one per BU size. After encoding the j^{th} picture, both models should be updated in order to be ready for a potential change of size in the next picture. Therefore, for the non used BU size, the amount $b_i(j)$ of generated texture bits, MAD and QP values will be computed as in Table 1, where N_{BU} is the total number of basic units in a picture. With these values the points used for R-Q regression are estimated, so the resulting coefficients could not be optimum. For this reason, continuous changes of BU size should be avoided and, therefore, we propose the following switching rules, based on an hysteresis criterion, for the aforementioned thresholds Γ and Θ:

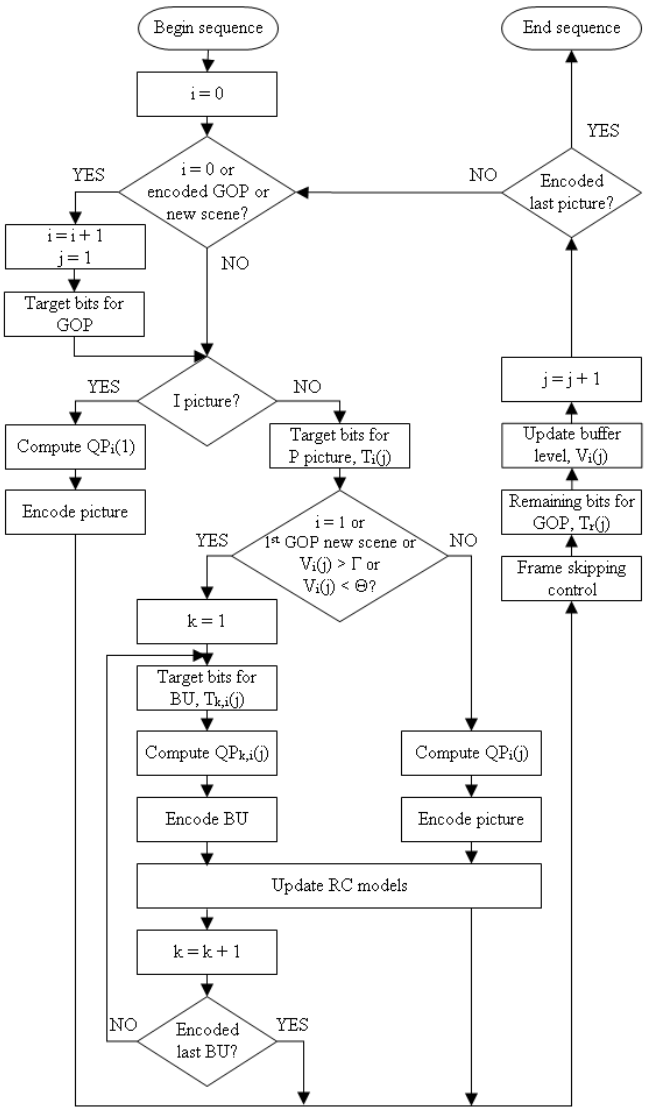

Fig. 1. Rate control scheme based on dynamic basic unit size.

- Case $V_i(j) > \Gamma$ and large BU → small BU
- Case $V_i(j) < (\Gamma - 0.05 \times BS)$ and small BU → large BU
- Case $V_i(j) < \Theta$ and large BU → small BU
- Case $V_i(j) > (\Theta + 0.05 \times BS)$ and small BU → large BU
- Otherwise → do not switch BU size

3.2. Scene change detection

Generally, a new scene has different texture and motion information than the previous one. If the j^{th} picture is the first one of the new scene to be encoded, there will not be any temporal correlation with its references and most of MBs will be encoded as intra mode. This supposes a serious drawback in real-time environments, since a lot of motion estimations are wasted.
In order to reduce computational cost, a scene change (SC) detector is proposed. It is based on the absolute difference between the luminance histograms of two consecutive pictures, H_{j-1} and H_j, and

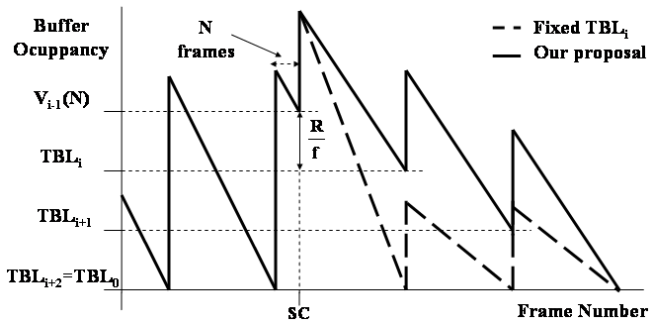

Fig. 2. Buffer occupancy variation for two options of TBL_i.

obeys the following expression:

$$D_j = \frac{1}{LB}\sum_{z=1}^{B}|H_{j-1}(z) - H_j(z)| \quad (2)$$

where L is the number of pixels in a frame, and B is the number of bins in each histogram, set to 128. If $D'_j \geq 0.08$, where D'_j is the first derivative of D_j, a scene cut is detected. This threshold value achieves a good trade-off between false alarm and hit rates.
Similar to [5], the GOP layer must carry out the following tasks after detecting the scene cut (see Fig. 1):

1. **Set the j^{th} picture as I type, $j = 1$:** For IP...P patterns, the IDR type can be used to eliminate the reference images and thus reduce motion estimations in following P pictures.
2. **Begin a new GOP, $i = i + 1$, and allocate target bits:** The total pictures in the GOP are the same than previous GOPs.
3. **Assign a QP for the j^{th} picture, $QP_i(1)$:** No information about the last GOP should be considered, due to the possible differences of complexity between both scenes. So the rule proposed by [1] to obtain $QP_1(1)$ can be used.
4. **Update TBL_{i+n}:** As it is shown in Fig. 2, after a SC the QP values for the i^{th} GOP would have to be increased to reach the TBL_0 value (see bold dashed line), set to zero in the experiments. For n future GOPs, a better visual quality is achieved decreasing TBL_{i+n} in a linear way as:

$$TBL_{i+n} = MAX\left[V_{i-1}(N) - (n+1)\frac{R}{f}, TBL_0\right] \quad (3)$$

where R is the target bit rate, f is the frame rate, and N is the number of encoded pictures in the $(i-1)^{th}$ GOP.

3.3. Frame skipping

When the picture buffer cannot accommodate an encoded frame, it must be discarded in order to prevent a buffer overflow. This situation is referred to as a *frame skipping*. Furthermore, the RC should be able to prevent posterior discards, so the following QP assignment is proposed for the j^{th} picture in the i^{th} GOP after skipping:

$$QP_i(j) = QP_i(j-1) + 4 \quad \text{if large BU size} \quad (4)$$

$$QP_{k,i}(j) = QP_{ave}(j-1) + 4 \quad \text{if small BU size} \quad (5)$$

where $QP_{ave}(j-1)$ is the average QP of the skipped $(j-1)^{th}$ picture. If the dropped picture is an I type, the next one will be encoded as intra instead of inter (the i^{th} GOP is delayed one frame) in order to improve the quality, since the posterior inter pictures will have better references.

Skipped frames vs. Target rate (kbits/s)					
Sequence	**BU (MBs)**	**128**	**256**	**512**	**1024**
Fb	**22**	0	0	2	0
	396	0	0	7	0
	Dynamic	0	0	5	0
Fm	**22**	1	0	0	0
	396	1	1	0	0
	Dynamic	1	0	0	0
PrFb	**22**	1	5	12	0
	396	0	7	22	2
	Dynamic	0	6	16	0
GdSfFb	**22**	0	0	4	0
	396	1	0	8	2
	Dynamic	0	0	6	0

Table 2. Skipped frames comparison among basic unit sizes.

4. EXPERIMENTS AND RESULTS

Our scheme based on dynamic BU size has been implemented over the RC [1] adopted by JVT software version JM 10.2 [6].
Two set of sequences in CIF format have been defined. In the first one, typical sequences used in video experiments are included, such as "Paris" (Pr), "Highway" (Hw), "Foreman" (Fm), "Coastguard" (Cg), "Container" (Ct) and "Football" (Fb). In the second group, linkings of other sequences have been used to get scene changes: "Paris-Football" (PrFb), "News-Garden" (NwGd), "Garden-Stefan-Football" (GdSfFb), and "Bus-Foreman" (BsFm). All of them have been encoded with the following configurations:

- **Profile:** main
- **Number of pictures:** 300
- **Frame rate:** $f = 25$ f/s
- **GOP:** IP...P and an intra picture each second
- **Target rates:** $R = 128, 256, 512, 1024$ kbits/s
- **Buffer size:** $BS = 0.5 \times R$ (half a second)
- **R-D optimization:** disabled (for real-time applications)
- **Symbol mode:** CAVLC

In order to prove the performance of our proposal, the BU size switching criterion has been disabled and other two RC schemes have been defined as references, each one with different BU size. Therefore, a last item in the configurations is presented:

- **BU sizes:** 22 (row) MBs, Dynamic BU, 396 (frame) MBs

Three performance measurements have been used. One of them is the number of skipped frames. The others represent the quality improvement of a basic unit, X_{BU}, with respect to a reference, X_{ref}, and obey:

$$\Delta X_{BU} = X_{BU} - X_{ref} \quad X = \mu, \sigma \quad (6)$$

where X can be the average PSNR, μ, or PSNR standard deviation, σ, and the reference is the worst of the three proposed BU sizes, for a given sequence and target rate.
The experimental results show that the dynamic BU scheme provides a good trade-off between μ and σ in a variety of sequences (see Figs. 3 and 4) and target rates (see Fig. 5). As it is shown in Fig. 6, the BU size changes when necessary and, therefore, a good PSNR performance is achieved. Nevertheless, in some cases the number of skipped frames is high (see Table 2), due to the suboptimal choice of the initial QP for the scene cuts and first GOP of sequence.

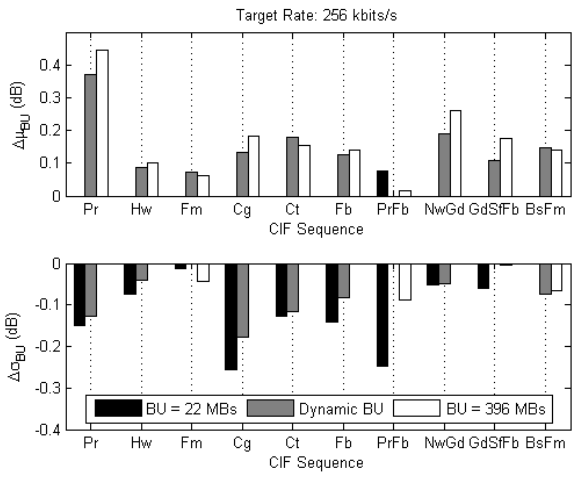

Fig. 3. PSNR improvement vs. CIF sequence. 256 kbits/s.

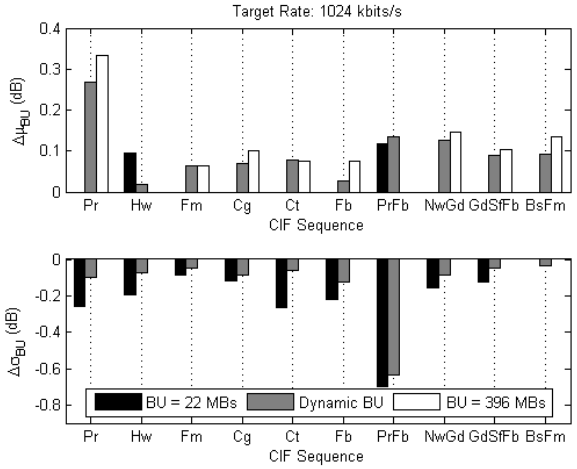

Fig. 4. PSNR improvement vs. CIF sequence. 1024 kbits/s.

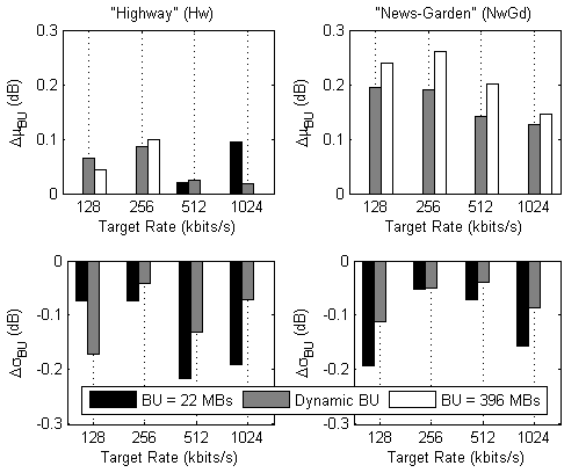

Fig. 5. PSNR improvement vs. Target rate. "Highway" and "News-Garden"' (CIF).

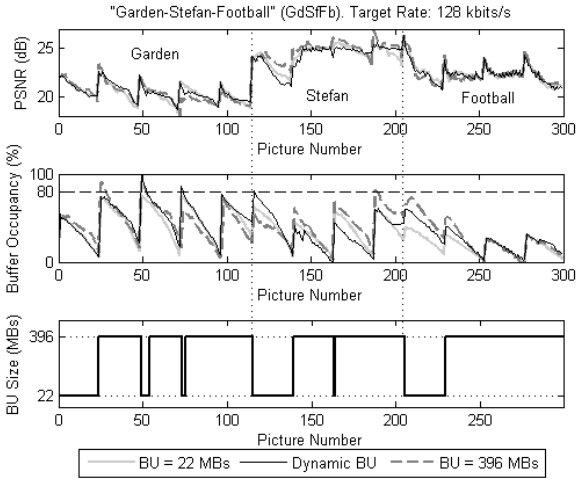

Fig. 6. PSNR, buffer occupancy and BU size variation. "Garden-Stefan-Football" (CIF). 128 kbits/s.

5. CONCLUSIONS AND FURTHER WORK

In this paper a robust rate control scheme based on dynamic basic unit size has been defined. A good performance in a lot of sequences and target rates has been achieved, identifying temporal heterogeneity scenarios to switch the BU size. Furthermore, our proposal is extensible to GOPs with B pictures, and can be implemented over any RC algorithm which includes the basic unit layer.

Nevertheless, the experiments show that the small BU size works better in a few stationary sequences (see Fig. 5). So this RC scheme could be generalized to an arbitrary BU size selection by means of the spatial heterogeneity analysis of the complexity. A better initial QP estimator which takes into account complexity information, as in [5], could be another future line.

6. REFERENCES

[1] S. Ma, Z. Li, and F. Wu, "Proposed draft of adaptive rate control," *JVT-H017*, Geneva, Switzerland, May 2003.

[2] M. Jiang and N. Ling, "Low-delay rate control for real-time H.264/AVC video coding," *IEEE Transactions on Multimedia*, vol. 8, no. 3, pp. 467–477, 2006.

[3] Z. He, Y.K. Kim, and S.K. Mitra, "Low-delay rate control for DCT video coding via ρ-domain source modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 8, pp. 928–940, 2001.

[4] N. Kamaci, Y. Altunbasak, and R.M. Mersereau, "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.

[5] J. Lee, I. Shin, and H. Park, "Adaptive intra-frame assignment and bit-rate estimation for variable GOP length in H.264," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 10, pp. 1271–1279, 2006.

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