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

A Cauchy-density-based rate controller for H.264/AVC in low-delay environments

WISSEN IM ZENTRUM
UNIVERSITÄTSBIBLIOTHEK

Technische
Universität
Berlin

A CAUCHY-DENSITY-BASED RATE CONTROLLER FOR H.264/AVC IN LOW-DELAY ENVIRONMENTS

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

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

ABSTRACT

The accuracy of the Cauchy probability density function for modeling of the discrete cosine transform coefficient distribution has already been proved for the frame layer of the rate control subsystem of a hybrid video coder. Nevertheless, in some specific applications operating in real-time low-delay environments, a basic unit layer is recommended in order to provide a good trade-off between quality and delay control. In this paper, a novel basic unit bit allocation for H.264/AVC is proposed based on a simplified Cauchy probability density function source modeling. The experimental results show that the proposed algorithm improves the average peak signal-to-noise ratio in 0.28 and 0.35 dB with respect to two well-known rate control schemes, while maintaining similar peak signal-to-noise ratio standard deviation and buffer occupancy evolution.

Index Terms— H.264, low-delay, rate control, bit allocation, basic unit.

1. INTRODUCTION

The inherent variability of video information implies that the video encoder normally produces a variable output bit rate, which must be controlled in order to comply with the nominal network rate required by real-time communications through constant bit rate channels. Therefore, the inclusion of a rate control (RC) algorithm in video encoders is of paramount importance. Although it is not a normative tool for video coding standards, several schemes have been recommended during the different standardization processes, such as TM5 for MPEG-2 [1], VM8 for MPEG-4 [2], TMN8 for H.263 [3], and AVC-TM for H.264/AVC [4].

Almost in any RC algorithm, a virtual buffer is considered at the output of the encoder. This buffer aims at modeling the decoder buffer behavior and bears the difference between the nominal bit rate of the network and the variable source bit rate per picture. The RC algorithm must keep the buffer in secure levels, avoiding both the overflow and the underflow.

In order to maintain the output rate within the buffer limits without visual quality degradation, a RC scheme should assign, according to the buffer status and the picture complexity, the most appropriate amount of bits and the corresponding quantization step (Q) to each coding unit. Traditional RC schemes select the Q value using an analytical rate-quantization (R-Q) function. This function is derived by means of a source modeling of the discrete cosine transform (DCT) coefficients. For instance, using a Gaussian probability density function (PDF), a logarithmic R-Q function can be inferred [5]; while using a Laplacian PDF, different linear [6], quadratic [2] or ρ-domain-based [7] R-Q models have been proposed.

Kamaci *et al* demonstrated in [8] the accuracy of the Cauchy PDF for DCT coefficient modeling. Starting from this distribution, a simple exponential R-Q model is obtained which, applied to a RC algorithm for H.264/AVC, achieves a mean PSNR improvement when compared to schemes based on the Laplacian PDF.

As many others algorithms, the one reported in [8] employs a frame layer which assigns the Q value on a frame basis. However, in the case of low-delay applications, the buffer size is very restricted and a finer RC is desirable. Given the growing popularity of these scenarios, such as in videophone or videoconferencing, several RC algorithms have been proposed for different video coding standards, where an additional basic unit (BU) layer, defined as a group of macroblocks in raster scan order which share the same Q value, is employed. In [1],[3],[7] and [9] a single macroblock is proposed as BU size, while in [4] a greater size is recommended for a better trade-off between quality smoothness and target frame size adjustment.

In this paper, a Cauchy-PDF-based BU bit allocation algorithm is proposed, showing that the exponential R-Q and rate-distortion models work properly at the BU layer RC.

The paper is organized as follows. Section 2 describes the solutions adopted for group of pictures (GOP), picture and BU layer RC. Section 3 shows and analyzes the experimental results of the suggested method in comparison with other popular RC schemes. Finally, in Section 4 some conclusions are drawn.

2. PROPOSED CAUCHY-PDF-BASED RATE CONTROL

Since the main contribution of this paper involves a new BU layer, some state of the art algorithms have been studied and their solutions adopted for the GOP and picture layers of the proposed RC scheme. These solutions are briefly described in the following subsections as well as the BU layer itself.

2.1. GOP Layer

The most common coding pattern in low-delay applications is the IP...P structure, since no structural delay is involved. Moreover, since I pictures need a larger amount of bits than P or B pictures, a unique GOP is adopted to encode the whole sequence in order to reduce the buffer delay.

The GOP layer computes the QP values for the first picture of each type, (QP_0^I and QP_0^P). These values depend on the average number of target bits per pixel (see [4] for further details).

Before encoding the current j^{th} picture, the total number of bits to encode the remaining pictures, $B(j)$, and the buffer fullness $V(j)$ are calculated as follows:

$$B(j) = \begin{cases} \frac{J \cdot R}{f}, & \text{if } j=1 \\ B(j-1) - t(j-1), & \text{otherwise,} \end{cases} \quad (1)$$

1

$$V(j) = \begin{cases} 0, & \text{if } j=1 \\ V(j-1)+t(j-1)-\frac{R}{f}, & \text{otherwise,} \end{cases} \quad (2)$$

where J represents the number of pictures, R is the target bit rate, f is the frame rate and $t(j-1)$ is the total number of bits used to encode the last picture.

2.2. Picture Layer

At picture layer, the amount of target bits for the jth picture, $\hat{T}(j)$, is obtained as a portion of the remaining bits $B(j)$:

$$\hat{T}(j) = \frac{B(j)}{N_{p,r}}, \quad (3)$$

where $N_{p,r}$ represents the number of remaining pictures.

In low-delay environments an additional buffer-status aware bit allocation algorithm is highly recommended to achieve a tight buffer level control and compensate the extra bits needed for encoding the first I picture. This model obeys:

$$\tilde{T}(j) = \frac{R}{f} + \delta\big(S(j) - V(j)\big), \quad (4)$$

where δ is a constant which is set to 0.5 to achieve a fast enough adjustment to the target buffer level without increasing the QP fluctuation, and $S(j)$ is the linear model of the target buffer level (see [10] for details). Thus, the total number of target bits to encode the jth picture is calculated as a combination of (3) and (4):

$$T(j) = \beta\hat{T}(j) + (1-\beta)\tilde{T}(j), \quad (5)$$

where β is a constant which is set to 0.5 as a good trade-off between PSNR and buffer fluctuation. Finally, $T(j)$ is bounded to conform to hypothetical reference decoder (HRD) constraints ([4]).

2.3. Basic Unit Layer

As pointed out in [3] and [9], in real-time low-delay applications, where the overall delay between the transmitter and receiver is severely constrained, the picture layer RC schemes are generally not suitable approaches, since the misadjustments to the frame target bits could result in larger fluctuations of the buffer fullness and, therefore, the buffer overflow or underflow risk is higher.

A BU is a group of macroblocks in raster scan order which share the same quantization value. The number of macroblocks in a BU is set before beginning the encoding process and it is generally an entire fraction of the total number of macroblocks in a picture.

2.3.1. Bit Allocation

We assume that the jth picture is composed of K BUs ($K > 1$), and the previous $k-1$ BUs in the picture have already been encoded. We have a budget of $T_{r,k}(j)$ bits for the remaining BUs ($k, k+1, \ldots, K$), which is calculated as:

$$T_{r,k}(j) = \begin{cases} T(j), & \text{if } k=1 \\ T_{r,k}(j) - t_{k-1}(j), & \text{otherwise,} \end{cases} \quad (6)$$

where $t_{k-1}(j)$ represents the number of bits used to encode the previous BU. Then, we want to find the $\tilde{b}_l(j)$ target texture bits for the lth BU so that:

$$T_{r,k}(j) = \sum_{l=k}^{K}\big(\tilde{b}_l(j) + \tilde{h}_l(j)\big), \quad (7)$$

where $\tilde{h}_l(j)$ is a prediction of the number of header and motion data bits of the remaining BUs.

In order to maintain a consistent video quality within the frames, we propose that the proper bit allocation for the current BU must be subject to:

$$D_l(j) \approx D_m(j) \quad l \neq m, \quad (8)$$

where $D_l(j)$ represents the distortion of the lth BU in the jth picture. At frame level, it is assumed that the same QP value in consecutive pictures produces similar distortion [8], since the global variation between consecutive pictures is usually moderated. However, at BU level, this assumption is no longer valid because of the potentially large differences in complexity of neighboring BUs. Therefore, although we start from the same models, the theoretical approach to the bit allocation must be different from that used in [8]. Specifically, we propose to use the simplified R-D function derived from a Cauchy-type model of the DCT coefficient distribution, namely:

$$D(R) \approx cR^{-\gamma}, \quad (9)$$

where c and γ are the model parameters. Using this R-D model and the constant distortion (or quality) condition (8), the target texture bits for the lth BU, $\tilde{b}_l(j)$, can be obtained from those of the current one, $\tilde{b}_k(j)$, as follows:

$$\tilde{b}_l(j) = c_{l,j}^{\frac{1}{\gamma_l}} c_{k,j}^{-\frac{1}{\gamma_l}} \tilde{b}_k(j)^{\frac{\gamma_k}{\gamma_l}}. \quad (10)$$

As mentioned before, the complexity of neighboring BUs may be significantly different in those pictures with high spatial heterogeneity. Thus, the assumptions $c_{l,j} = c_{k,j}$ and $\gamma_{l,j} = \gamma_{k,j}$, with $l \neq k$, are no longer valid. Then, combining the expressions (7) and (10), the following BU bit allocation is obtained:

$$T_{r,k}(j) = \sum_{l=k}^{K}\left(c_{l,j}^{\frac{1}{\gamma_l}} c_{k,j}^{-\frac{1}{\gamma_l}} \tilde{b}_k(j)^{\frac{\gamma_k}{\gamma_l}} + \tilde{h}_l(j)\right). \quad (11)$$

Finally, assuming the same header and motion data bits for the N_r remaining BUs within the picture, the expression (11) can be rewritten as:

$$T_{r,k}(j) = \sum_{l=k}^{K}\left(c_{l,j}^{\frac{1}{\gamma_l}} c_{k,j}^{-\frac{1}{\gamma_l}} \tilde{b}_k(j)^{\frac{\gamma_k}{\gamma_l}}\right) + N_r\tilde{h}_k(j), \quad (12)$$

where $\tilde{h}_k(j)$ can be obtained as the average number of header and motion data bits used to encode the previous $k-1$ BUs. The Newton's method can be used to iteratively obtain a solution for $\tilde{b}_k(j)$.

2.3.2. Quantization Parameter Estimation

When the jth picture is to be encoded, the quantization parameter (QP) value for the kth BU, $QP_k(j)$, is obtained from the corresponding $Q_k(j)$, which is computed using the exponential R-Q model described in [8] adapted for BU level:

$$Q_k(j) = \left(\frac{\tilde{b}_k(j)}{a_{k,j}}\right)^{-\frac{1}{\alpha_k}}. \quad (13)$$

where $a_{k,j}$ and $\alpha_{k,j}$ are model parameters. Finally, the $QP_k(j)$ value is bounded at BU and picture levels as stated in [4] in order to avoid noticeable quality variations within the frame.

Table 1. α_k and γ_k coefficients update.

$B_{pp,k}$	α_k	$B_{pp,k}$	γ_k
< 0.05	1.60	< 0.07	0.50
$0.05 - 0.10$	1.40	$0.07 - 0.20$	0.70
> 0.10	1.20	> 0.20	1.00

2.3.3. Rate-Quantization and Rate-Distortion Models Update

After encoding the kth BU of the jth inter picture, the coefficients $a_{k,j+1}$ and $c_{k,j+1}$ are updated as follows:

$$a_{k,j+1} = b_k(j)\widetilde{Q}_k(j)^{\alpha_k}, \qquad (14)$$

$$c_{k,j+1} = d_k(j)b_k(j)^{\gamma_k}, \qquad (15)$$

where $b_k(j)$ is the number of texture bits used to encode the kth BU, and $d_k(j)$ is the sum of square error between the original and reconstructed luminance pixels in the BU. $\widetilde{Q}_k(j)$ is the Q value obtained from $QP_k(j)$, the QP value for the kth BU of the jth frame. After encoding the first inter picture in the sequence, Table 1 is employed to initialize the parameters α_k and γ_k according to the average number of texture bits to encode a pixel of the kth BU, $B_{pp,k}$. Thresholds used to set the initial α_k values have been obtained from [8].

3. EXPERIMENTS AND RESULTS

The proposed BU layer RC algorithm was implemented on the Joint Video Team (JVT) software version JM 10.2 [11]. In order to assess its performance, it was compared to two Laplacian-PDF-based schemes: the RC algorithm adopted by JVT [4] and the one proposed in [12]. CIF and QCIF common video sequences as well as linkings of them were used in the experiments.

A row of macroblocks was employed as BU size to achieve a good trade-off among quality, buffer fluctuation and computational cost. The complete configuration of our tests is stated here:

- **QCIF sequences:** "MobileCalendar", "MotherDaughter", "News", "Salesman", "Silent" and "Paris-Football"

- **CIF sequences:** "Coastguard", "Container", "Football", "Foreman", "Paris" and "Bus-Foreman"

- **Profile:** baseline

- **Number of pictures:** 300

- **Frame rate:** $f = 25$ f/s

- **Target rates (QCIF/CIF):** $R = 32/128, 64/256, 96/512, 128/1024$ kbits/s

- **Buffer size:** 500 ms

- **R-D optimization:** disabled

- **Symbol mode:** CAVLC

Luminance peak signal-to-noise ratio (PSNR), in terms of mean and standard deviation, and output bit rate are shown in Tables 2 and 3 for some QCIF and CIF sequences, respectively. The proposed algorithm achieves an average PSNR improvement of 0.35 dB over the RC scheme reported in [12], since the latter does not take into account the complexities of BUs for better target bit assignment. Furthermore, we achieve a 0.28 dB average PSNR improvement over the default RC algorithm of JM 10.2 [4], which employs a BU bit allocation algorithm based on a complexity ratio that takes into account the remaining BUs of a picture. Moreover, the results in terms

Table 2. PSNR (mean and standard deviation) and output rate for several QCIF sequences and target rates.

Sequence Target Rate (kbits/s)	Algorithm on JM 10.2	Mean PSNR (dB)	PSNR Deviation (dB)	Output Rate (kbits/s)
News **64**	**[12]**	36.22	0.82	64.16
	[4]	36.41	0.81	64.12
	Proposed	36.91	0.95	64.05
News **128**	**[12]**	40.95	1.08	128.24
	[4]	41.02	0.98	128.17
	Proposed	41.42	1.10	128.14
Paris-Football **64**	**[12]**	27.29	2.82	64.05
	[4]	27.27	2.81	63.13
	Proposed	27.68	2.95	64.03
Paris-Football **128**	**[12]**	31.55	3.76	127.98
	[4]	31.52	3.67	128.12
	Proposed	31.89	3.91	127.99

Table 3. PSNR (mean and standard deviation) and output rate for several CIF sequences and target rates.

Sequence Target Rate (kbits/s)	Algorithm on JM 10.2	Mean PSNR (dB)	PSNR Deviation (dB)	Output Rate (kbits/s)
Bus-Foreman **256**	**[12]**	31.32	3.92	256.29
	[4]	31.48	3.78	256.19
	Proposed	31.49	3.75	256.08
Bus-Foreman **1024**	**[12]**	37.65	3.58	1024.79
	[4]	37.92	3.31	1024.35
	Proposed	37.91	3.31	1024.19
Paris **256**	**[12]**	31.76	0.59	256.18
	[4]	31.95	0.64	256.09
	Proposed	32.57	0.81	256.07
Paris **1024**	**[12]**	40.29	0.69	1024.71
	[4]	40.91	0.80	1024.21
	Proposed	41.19	0.89	1023.95

of PSNR standard deviation and output bit rate are similar to those of the reference schemes.

Finally, the PSNR and encoder buffer occupancy versus picture number are plotted in Figs 1 and 2, respectively. It can be seen that the proposed method achieves a good quality performance and similar buffer occupancy evolution when compared to [12] and [4]. Furthermore, in sequences with scene cuts, such as "Paris-Football" and "Bus-Foreman", a faster adaptation to the complexity of a new scene is reached with our proposal.

4. CONCLUSIONS

In this paper, a BU layer RC algorithm has been proposed for H.264/AVC. Starting from a simplified Cauchy PDF for DCT coefficient distribution modeling, a novel bit allocation has been developed and the exponential R-Q model has been applied on a BU basis. In order to assess its performance, the RC was integrated on an H.264/AVC encoder, which was configured according to a
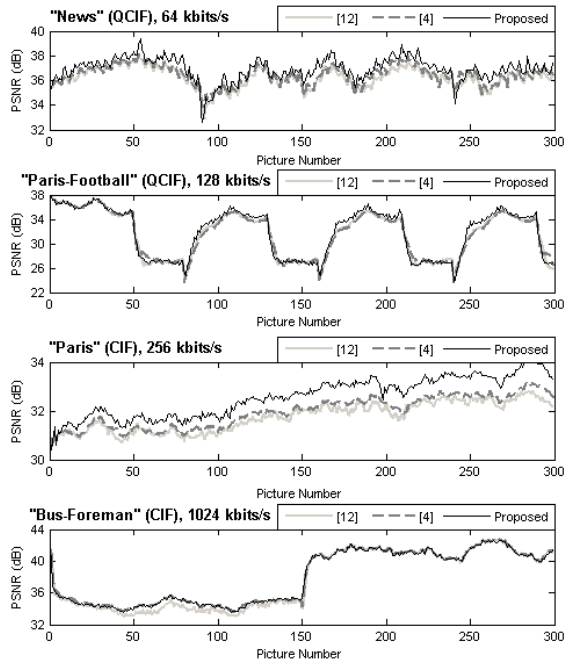
Fig. 1. PSNR evolution for several video sequences and target rates.

Fig. 2. Encoder buffer occupancy evolution for several video sequences and target rates.

real-time low-delay application and compared to two Laplacian-PDF-based BU layer RC schemes. The experimental results show an average PSNR improvement over the default RC algorithm of JM 10.2 [4] and the algorithm described in [12], with similar buffer occupancy evolutions. Furthermore, our proposal achieves similar PSNR standard deviation and output bit rate.

Another characteristic of the proposed RC algorithm is the low computational requirement to update the coefficients of the exponential R-Q and R-D models associated to each BU in the picture, a_k and c_k with the linear expressions (14) and (15). In contrast to this, the Laplacian-PDF-based method used in the reference schemes involves updating parameters on both the quadratic R-Q and the complexity prediction models by means of a regression analysis of previously encoded BUs, with the consequent overhead.

5. REFERENCES

[1] Test Model 5, "http://www.mpeg.org/MPEG/MSSG/tm5," .

[2] Tihao Chiang and Ya-Qin Zhang, "A new rate control scheme using quadratic rate distortion model," in *Image Processing, 1996. Proceedings., International Conference on*, 1996, vol. 1, pp. 73–76 vol.2.

[3] J. Ribas-Corbera and Shawmin Lei, "Rate control in DCT video coding for low-delay communications," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 9, no. 1, pp. 172–185, 1999.

[4] S. Ma, Z. Li, and F. We, "Proposed draft of adaptive rate control," *JVT-H017, 8th JVT Meeting*, Geneva, Switzerland, May 2003.

[5] Bo Tao, B.W. Dickinson, and H.A. Peterson, "Adaptive model-driven bit allocation for MPEG video coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 10, no. 1, pp. 147–157, Feb 2000.
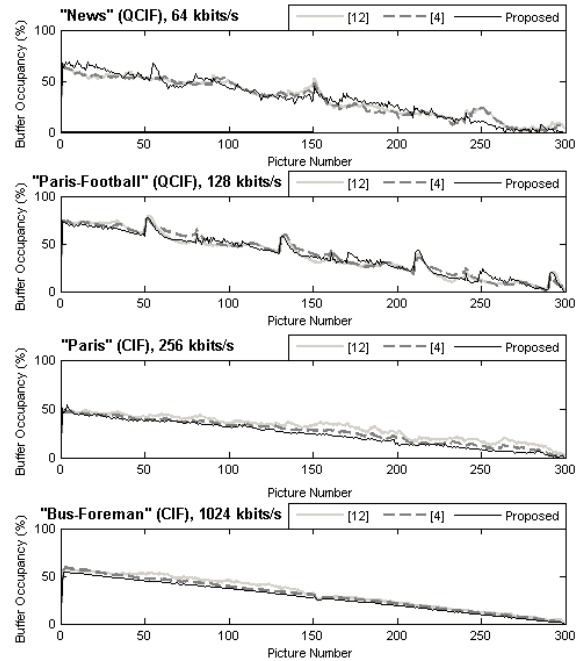
[6] S. Ma, Wen Gao, and Yan Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 15, no. 12, pp. 1533–1544, 2005.

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

[8] 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," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 15, no. 8, pp. 994–1006, 2005.

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

[10] Siwei Ma, Wen Gao, Peng Gao, and Yan Lu, "Rate control for advance video coding (avc) standard," in *Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on*, 2003, vol. 2, pp. II–892–II–895 vol.2.

[11] JM 10.2, "http://iphome.hhi.de/suehring/tml/download/old_jm/," .

[12] Z. Li, W. Gao, F. Pan, S. Ma, , G.N. Feng, K.P. Lim, X. Lin, S. Rahardja, and H.Q. Lu, "Adaptive rate control with HRD consideration," *JVT-H014, 8th JVT Meeting*, Geneva, Switzerland, May 2003.