$\rho-$DOMAIN RATE-DISTORTION OPTIMAL RATE CONTROL FOR DCT-BASED VIDEO CODERS

Nejat Kamaci and Yucel Altunbasak

Center for Signal and Image Processing
Georgia Institute of Technology
Atlanta, GA 30332
email:{kamaci,yucel}@ece.gatech.edu

ABSTRACT

Most real-time MPEG-like encoders are designed to perform in a constant bit rate (CBR) mode, in which large deviations from a target rate are avoided at any time instant because of video buffer regulations set at the decoder. A rate-control method at the encoder ensures that by intelligent selection of encoding parameters, the video buffer constraints are met, while at the same time a best possible video quality is achieved. Almost all reported rate-control methods model the rate and distortion in terms of the quantization step size. A recently suggested alternative is to use ρ, the fraction of zeros among the quantized DCT coefficients [5]. In this paper, we address the problems associated with the ρ-domain rate-control approaches in the literature. In addition, we propose a robust MPEG-2 rate-control algorithm that aims to minimize picture quality variations. Our extensive analysis with several video sequences and using MPEG-2 codec indicates that the proposed algorithm can achieve a more uniform visual quality, and provides a PSNR gain that ranges from 0.5 dB at 2 Mbps to 0.9 dB at 6 Mbs over TM5 on the average.

1. INTRODUCTION

The rate-control problem for video coding applications can be addressed at three layers. At the group of pictures (GOP) layer, the problem is to find the bit budget for each GOP. Similarly, at the picture layer, it involves finding the picture bit budget (for each frame within a GOP) from a given GOP bit budget. Finally, at the macroblock layer, the rate control problem reduces to finding the coding parameters for each macroblock, given the frame bit budget.

The output bit rate of a video source can be controlled by making several encoding/decoding passes to meet the target bit rate, but most applications demand a very fast encoder. Conventional rate-control methods use mathematical models to estimate the output bit rate and video quality (or distortion) to avoid multiple encoding passes. The performance of these rate control algorithms greatly depends on the accuracy of the models used. The performance of an encoder is measured by its bit rate R, and the distortion, D, that is introduced. The distortion is a measure of the actual or perceived difference between the original and coded video frame. Most R and D models reported in the literature [1, 2, 3, 4] try to relate these quantities to the quantization step size Q. He *et al.* have shown, however, that the fraction of zero coefficients among the quantized DCT coefficients is a simpler to compute quantity that can estimate R and D more robustly [5]. Thus, it is possible to improve rate-control performance by using ρ-domain models instead of Q-domain ones. For an MPEG-2 coder, we have observed that the ρ-domain rate model suggested by He *et al.* is accurate. However, we have also observed that the distortion model suggested by He *et al.* is less accurate and the frame bit allocation using this model (proposed in [6]) exhibits stability problems. In this paper, we present a robust rate-control algorithm for MPEG-2 based on new ρ-domain rate and distortion models, and a method for estimating the model parameters. We also consider a new framework in which achieving a uniform distortion (PSNR) is selected as the optimization criterion instead of the commonly-used criterion of minimizing (maximizing) the average distortion (PSNR).

2. ρ - DOMAIN OPTIMIZATION FOR RATE CONTROL

2.1. Rate and distortion models in ρ domain

Based on the assumption that the AC coefficients of a natural image have a distribution that is best approximated by a Laplacian distribution, He *et al.* showed that the number of bits per pixel required to encode one color component of a quantized DCT image is given by [5]

$$R = \log_2 \frac{1+\psi^a}{1+\psi^a-2\psi+(1-a)\left(1+\psi^a\right)\psi\ln\psi},$$

where $\psi = 1-\rho$ and a is the ratio of the quantization step size to the quantization dead zone width. He *et al.* suggested a linear approximation to the above formula given as

$$R = \theta\left(1-\rho\right) \quad \text{bits per pixel,} \tag{1}$$

where θ is a constant. Similarly, the distortion, defined as the mean square error introduced by quantization of the DCT coefficients is given by

$$D = E\left[|x-\hat{x}|^2\right] = \frac{1+\psi^a-2\psi+(1-a)\left(1+\psi^a\right)\psi\ln\psi}{\lambda\left(1+\psi^a\right)}.$$

As in the case for the rate, a simplified approximation to D is possible. In [6], He *et al.* suggested an exponential approximation as

$$D\left(\rho\right) = \sigma^2 e^{-\alpha(1-\rho)}, \tag{2}$$

where σ^2 is the picture variance and α is an unknown constant.

We experimented with the rate model of (1) and the distortion model of (2) for ρ-domain optimum bit allocation as described in [6]. For a wide range of video sequences and various encoding parameters, we had a number of observations:

- The rate model is accurate (for the purposes of rate control).
- The distortion model is not always accurate, particularly for B-frames.

- The frame bit allocation algorithm using the ρ-domain distortion model occasionally becomes unstable and fails to achieve the target bit rate.

Observe that according to the distortion model of (2), we always have $D \leq \sigma^2$. However, the case where $D > \sigma^2$ is frequently observed when the source video is encoded using an MPEG-2 encoder. This is due to the distortion component coming from the DC coefficient quantization that becomes significant especially when σ^2 is small. In the extreme case where all coefficients are quantized to zero, the distortion is equal to the sum of the squares of the pixel values which is larger than the variance. As a result of the fact that $D > \sigma^2$ is possible, the estimate of the model parameter α can assume negative values and the frame bit allocation algorithm can return negative values for the number of target bits for the consecutive frames. To alleviate this problem, we propose a new distortion model:

$$D = \kappa \sigma^2 e^{-2\theta(1-\rho)},$$

where $\kappa > 0$ is a model parameter. Note that the rate R is defined as the number of bits per pixel. We will use this model to develop a new rate-control algorithm for the MPEG-2 encoder. As discussed in Section 1, we will develop our solution to the rate-control problem at three layers in a hierarchical order.

2.2. GOP layer bit allocation

One of the fundamental problems in rate control is the allocation of the available bits for a group of pictures (GOP). For constant bit rate encoding, the bit budget for each group of pictures is a constant, which is determined using the target bit rate of the encoded video bitstream. That is, the target number of bits for the GOP, R_{gop}, is computed using the target rate R, frame rate F, and the GOP size G as:

$$R_{gop} = \frac{R \times G}{F}.$$

2.3. Picture layer bit allocation

The overall picture bit budget is shared by the encoding of the motion vectors, the quantized DCT coefficients and the headers. The sum of the required bits for encoding the motion vectors and the headers will be referred to as the *overhead* bits. We have the two formulas for the distortion and DCT bit rate mentioned earlier, $R(\rho) = \theta(1-\rho)$, and $D(\rho) = \kappa\sigma^2 e^{-2\theta(1-\rho)}$.

The optimal bit allocation in the sense of achieving uniform distortion within the GOP can be formulated as follows. Define

B_{gop} : total GOP bits,
B_{dct} : GOP bits for encoding the DCT coefficients,
B_{oh} : GOP overhead bits.

B_{gop} can be computed using the target gop rate, R_{gop}, and assume that B_{oh} is estimated from the previous GOP. At the start of the sequence, B_{oh} must be initialized. Then

$$B_{dct} = B_{gop} - B_{oh}.$$

Assume we are about to encode the k^{th} picture of the current GOP. Let $\{T_i\}_{i=k}^{G}$ be the set of bit budgets for encoding the DCT coefficients for pictures $i = k, \cdots, G$, where G is the GOP size. The set of T_i's that would yield uniform distortion throughout the GOP satisfies

$$\kappa_i \sigma_i^2 e^{-2\frac{T_i}{K}} = \kappa_j \sigma_j^2 e^{-2\frac{T_j}{K}}, \quad \forall\, i \neq j, \;\; j = k, \cdots, G$$
$$\text{subject to} \sum_{i=k}^{G} T_i \leq B_{dct}, \tag{3}$$

where K is the number of pixels in a frame (*e.g.*, if the video is in $4:2:0$ format and has a resolution of 720×480 pixels, then $K = 720 \times 480 \times 1.5$). The solution to (3) is given by

$$T_i = \frac{K}{2}\ln 2\kappa_i\sigma_i^2 + \frac{\left(B_{dct} - \sum_{j=k}^{G} \frac{K}{2}\ln 2\kappa_j\sigma_j^2\right)}{G-k+1}, \tag{4}$$

where $i = k, \cdots, G$. Note that the variance σ_k^2 can be computed for the current picture, however it must be estimated for the future frames ($i > k$). Similarly, the model parameter κ_i needs to be estimated for the current and future frames.

2.4. Macroblock layer quantizer adaptation

The macroblock-level rate-control algorithm is adopted from [6], with slight modifications. At this layer, the goal is to find the set of quantization parameters (Q) so that the picture bit budget is met, while the quality variation is minimized within the frame. Suppose we are given the DCT coefficient bit budget T_k for the current k^{th} frame to be encoded. Let $R(m)$ be the number of remaining bits for encoding the DCT values of the unencoded macroblocks before encoding the macroblock with index m. Let θ be the rate parameter of the frame. Let D_k^0 and D_k^1 be histograms of the DCT coefficients for intra and non-intra cases, respectively. Initially, $m = 0$ and $R(0) = T_k$. The $\rho - Q$ mappings are determined using histograms of the DCT coefficients of macroblocks as

$$\rho(Q) = \frac{1}{K}\sum_{|z|\leq Q} D_k^0(z) + \frac{1}{K}\sum_{|z|\leq Q+\delta} D_k^1(z), \tag{5}$$

where $D_k(z)$ is the number of the DCT coefficients that has a value of z, and $\delta + Q$ is the width of the dead zone of the non-uniform quantizer of the non-intra macroblocks. After encoding each macroblock, the remaining bit budget is updated. That is, after encoding $m-1$ macroblocks, we update $R(m)$, the remaining bit budget for encoding the remaining picture macroblocks, by

$$R(m) = T_k - (\text{DCT bits of first } (m-1) \text{ macroblocks}).$$

We also update the histograms D_k^0 and D_k^1 of the DCT coefficients so that the histograms show the DCT coefficient distributions of the remaining macroblocks to be encoded. To compute Q for the current macroblock, the target ρ should be computed using the residual picture budget as

$$\rho(m) = 1 - \frac{R(m)}{\theta \times N(m)},$$

where $N(m)$ is the number of the remaining macroblocks to be encoded. The quantization parameter Q is determined using the mapping between ρ and Q given in (5).

2.5. Estimation of model parameters θ, σ^2, κ

After a frame is encoded, the actual number of DCT bits (S_{dct}) and the actual fraction of zeros (ρ_{act}) are recorded. The value of θ depends on the picture type even though the video content does not vary. As a result, we use separate estimators for different picture types, θ^I, θ^P and θ^B. Let $\mathcal{T}(k)$ be the type of the k^{th} frame ($\mathcal{T}(k) \in \{I, P, B\}$). For the k^{th} picture we estimate $\theta^{\mathcal{T}(k)}$ as

$$\theta^{\mathcal{T}(k)} = \frac{S_{dct}}{(1-\rho_{act}) \times K}, \qquad \mathcal{T}(k) \in \{I, P, B\}.$$

In a single pass environment, the actual frame variance for the current picture is available, while that of future frames are not. Therefore they need to be estimated so that we can use in (4). We consider using the variances of previous frames of the same type for future pictures. Therefore, we use separate estimators for each picture type, namely V^I, V^P, and V^B. The actual frame variance for the current picture is assigned as the estimate for the future picture of the same type as

$$V^{\mathcal{T}(k)} = \sigma_k^2.$$

The distortion parameter κ relates the fraction of zeros and the residual picture variance to the picture distortion. It depends on the picture type. Therefore as in the case of the bit production parameter θ, we propose using separate estimators for each picture type. Let κ^I, κ^P, and κ^B be the estimators for I-, P-, and B-type pictures, respectively. After a frame is encoded, the distortion D is computed as the mean square error (MSE) between the original and reconstructed pictures (X and $\hat{X}$), *i.e.,* $D = E[(X - \hat{X})^2]$. Using the distortion, the distortion model parameter κ for the k^{th} picture is updated as follows:

$$\kappa^{\mathcal{T}(k)} = \frac{V^{\mathcal{T}(k)}}{D} e^{2S_{dct}}.$$

The current values of the estimators (V^I, V^P, V^B) and ($\kappa^I, \kappa^P, \kappa^B$) are used as the (V, κ) values associated with the future frames in (4).

3. RESULTS

The proposed rate-control algorithm was implemented with the Berkeley MPEG-2 codec and tested for various video sequences. All of the test sequences were in the $4:2:0$ sampling format with 720×480 resolution and a 30 fps frame rate. The GOP size was 15 and the P picture distance was 3. The search range was selected as 16×16.

We compare the coding efficiency of the proposed algorithm with that of TM5 and a ρ-domain R-D optimal rate control algorithm that uses the distortion model of Eq. (2). We will denote this algorithm by RC1. We compare the average PSNR values at three different rates, and with six different sequences.

Table 1 shows the average PSNR results for the six sequences at 4.0 Mbps. The second column shows the PSNR performance of the TM5 rate-control algorithm. In the third column, the same sequences are encoded using the RC1 method. The next column shows the proposed R-D optimum ρ-rate control algorithm. The proposed algorithm achieves as much as 1.0 dB PSNR gain over TM5. RC1 occasionally achieves very low average PSNR. For these cases, RC1 fails to achieve the target bit rate. Table 2 and 3 depict the average PSNR results at the encoding rates of 2 Mbps and 6 Mbps, respectively. These results show that at lower bit rates, the PSNR improvement is less when compared to higher bit rates. Figs. 1, 2, 3 and 4 show PSNR curves for some of the encoding tests. These results show that the proposed algorithm achieves consistently better picture quality. In all four cases, it achieves substantially less PSNR variation between the frames, which is desirable for video coding applications.

Table 4 shows the rate-control parameter statistics. The average values of the distortion parameter κ and the bit production parameter θ are different for different picture types. The value of ρ is around 0.85 for I pictures, around 0.95 for P-pictures and around 0.99 for B-pictures. The statistics indicate that the model parameters depend on the picture type and the characteristics of the video in general.

Sequence	TM5	RC1	Proposed RC
CAROUSEL	33.8	29.6	34.8
FLOWER GARDEN	30.4	26.1	31.1
FOOTBALL	32.2	28.9	32.6
MANHATTAN	39.0	32.4	40.0
MOBILE AND CALENDAR	27.0	24.2	28.0
SUZIE	40.5	34.3	40.7

Table 1. PSNR comparisons for six test sequences (in dB). The picture format is CCIR-601, with $4:2:0$ sampling, a GOP size of 15, and a P-picture distance $L = 3$. The target bit rate is 6 Mbps.

Sequence	TM5	RC1	Proposed RC
CAROUSEL	30.0	29.5	30.4
FLOWER GARDEN	26.8	26.1	27.7
FOOTBALL	28.8	28.5	28.9
MANHATTAN	35.8	32.4	36.6
MOBILE AND CALENDAR	24.2	24.1	24.6
SUZIE	38.5	34.1	38.7

Table 2. PSNR comparisons for six test sequences at 2 Mbps.

Sequence	TM5	RC1	Proposed RC
CAROUSEL	35.8	37.1	37.2
FLOWER GARDEN	32.2	26.1	33.2
FOOTBALL	34.2	28.5	34.7
MANHATTAN	40.7	41.5	41.7
MOBILE AND CALENDAR	28.9	24.3	30.0
SUZIE	41.7	34.8	42.0

Table 3. PSNR comparisons for six test sequences at 6 Mbps.

Sequence		$\bar{\rho}_{act}$	$\bar{\kappa}$	$\bar{\theta}$
CAROUSEL	I	0.893	0.37	6.94
	P	0.956	0.70	6.67
	B	0.986	1.20	6.80
FLOWER GARDEN	I	0.819	0.44	6.25
	P	0.957	1.10	8.76
	B	0.993	1.38	12.67
FOOTBALL	I	0.882	0.51	6.74
	P	0.956	0.82	6.85
	B	0.988	1.26	7.01
MANHATTAN	I	0.856	0.60	6.38
	P	0.962	0.81	6.55
	B	0.989	1.31	7.78
MOBILE AND CALENDAR	I	0.836	0.80	5.90
	P	0.966	0.98	7.20
	B	0.989	1.16	8.71
SUZIE	I	0.820	0.56	5.75
	P	0.960	0.93	6.81
	B	0.993	1.19	8.41

Table 4. Proposed ρ-domain rate control parameter statistics for six sequences. The coding parameters are the same as for the experiments in Table 1.

4. SUMMARY AND CONCLUSIONS

In this paper, we addressed the problems associated with the ρ-domain rate-control approaches proposed in [6]. We proposed to use a new distortion model in the ρ-domain to alleviate the stability problems. By using the new distortion model, we proposed a robust MPEG-2 rate-control algorithm that aims to minimize picture quality variations instead of maximizing the overall quality. By extensive analysis with several video sequences, we showed that the proposed algorithm can achieve a more uniform visual quality. We also demonstrated that the proposed algorithm provides an average PSNR gain that ranges from 0.5 dB at 2 Mbps to 0.9 dB at 6 Mbs over TM5. A future research direction is to apply the proposed rate and distortion models for other video encoders besides MPEG-2.

5. REFERENCES

[1] W. Ding and B. Liu, "Rate control of MPEG video coding and recording by rate-quantization modeling," *IEEE Transactions on Circuit and Systems for Video Technology,* vol. 6, no. 1, pp. 12-20, Feb. 1996

[2] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Transactions on Circuit and Systems for Video Technology,* vol. 9, pp. 172-185, Feb. 1999.

[3] T. Chiang and Y. -Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Transactions on Circuit and Systems for Video Technology,* vol. 7, pp. 246-250, Feb. 1997.

[4] H. -M. Hang and J. -J. Chen, "Source model for transform video coder and its application," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 7, pp. 287-298, Apr. 1997.

[5] Z. He and S. K. Mitra "A linear source model and a unified rate control algorithm for DCT video coding," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 12, no. 11, Nov. 2002.

[6] Z. He and S. K. Mitra "Optimum bit allocation and accurate rate control for video coding via ρ-domain source modeling," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 12, no. 10, Oct. 2002.

6. FIGURES

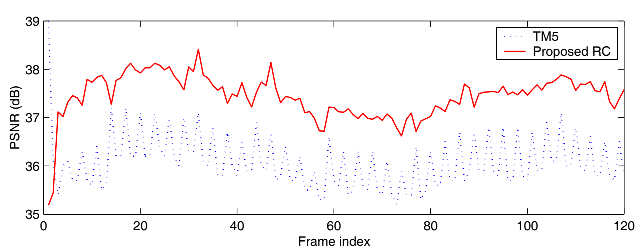

Fig. 1. Comparison of the frame PSNR values for CAROUSEL sequence. The coding parameters are the same as for the experiments in Table 3. The dotted line shows the PSNR values achieved with the macroblock level ρ-rate control. The solid line shows the PSNR values achieved with the proposed ρ-rate control algorithm.

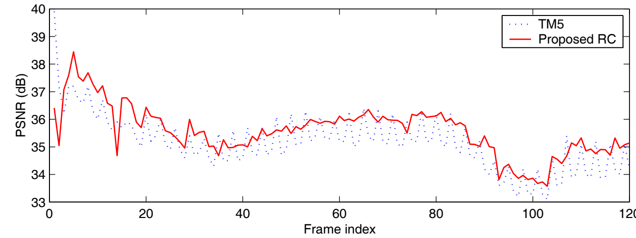

Fig. 2. Comparison of the frame PSNR values for FOOTBALL sequence. The coding parameters are the same as for the experiments in Table 3. The dotted line shows the PSNR values achieved with the macroblock level ρ-rate control. The solid line shows the PSNR values achieved with the proposed ρ-rate control algorithm.

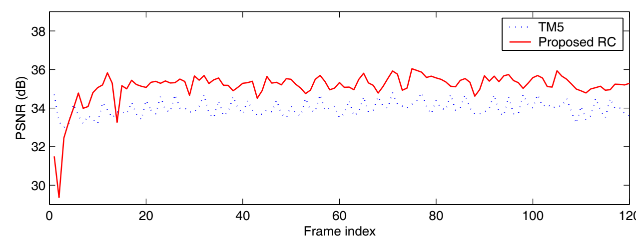

Fig. 3. Comparison of the frame PSNR values for FLOWER GARDEN sequence. The coding parameters are the same as for the experiments in Table 3. The dotted line shows the PSNR values achieved with the macroblock level ρ-rate control. The solid line shows the PSNR values achieved with the proposed ρ-rate control algorithm.

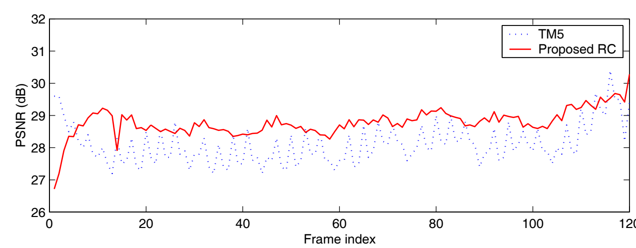

Fig. 4. Comparison of the frame PSNR values for MOBILE AND CALENDAR sequence. The coding parameters are the same as for the experiments in Table 3. The dotted line shows the PSNR values achieved with the macroblock level ρ-rate control. The solid line shows the PSNR values achieved with the proposed ρ-rate control algorithm.