A RATE-QUANTIZATION MODEL FOR MPEG ENCODERS

Bo Tao[1] *Heidi A. Peterson*[2] *Bradley W. Dickinson*[1]

[1]EE Dept., Princeton Univ., Princeton, NJ 08544
botao, bradley@ee.princeton.edu
[2]Sarnoff Corporation, CN5300 Princeton, NJ 08543
heidi@sarnoff.com

ABSTRACT

We derive here a parametric rate-quantization model, based on traditional rate-distortion theory, for MPEG encoders. Given the bit budget for a picture, our model calculates a baseline quantization scale factor. We compare our approach to a technique using an ad-hoc rate-quantization model, and also to TM5. In both cases, experimental results demonstrate that our approach produces an actual encoded bitrate closer to the target bit budget for the picture, as well as an improved peak signal-to-noise ratio.

1. Introduction

The Moving Picture Experts Group (MPEG) standard specifies a bitstream syntax for compressed video; hence, the decoding method is specified. The quality of MPEG decoded video, therefore, depends largely on the encoding process. One key component of an MPEG encoder is rate control. Rate control is the algorithm used to monitor a virtual encoder buffer, and periodically adjust quantization parameters to prevent buffer overflow or underflow. In order for a rate control algorithm to make appropriate quantization parameter adjustments, an accurate estimate is needed of bitrate as a function of quantization parameters. In previously published work, this estimate has been provided by either an empirical rate-quantization model [1, 2], or via pre-coding [3, 4]. Empirical models often depend on training sequences, and may not generalize well to other sequences, while pre-coding entails extra computation, causing long encoding delays.

Alternatively, traditional rate-distortion theory can be applied to the problem of encoder quantization parameter adjustment. This approach has been used previously for still image coding [5, 6]. In this work we apply rate-distortion theory to MPEG video coding. We present here an MPEG rate control algorithm based on a new parametric rate-quantization model derived from rate-distortion theory. Comparisons are made between our rate control algorithm, the Test Model 5 (TM5) rate control algorithm (TM5 is a baseline encoder algorithm made available by the MPEG committee [7]), and a pre-coding-based rate control algorithm which uses an ad-hoc rate-quantization model proposed in [3].

2. Rate Control Based on a New Rate-Quantization Model

2.1. MPEG rate control

The MPEG standard has a layered structure, comprised of blocks, macroblocks, slices, pictures, and Groups of Pictures (GOPs). An MPEG-encoded picture can be intra-coded (I-picture), predictively-coded (P-picture) or interpolatively-coded (B-picture).

The MPEG bitstream specifies two quantization matrices (one for intra-coded blocks, and the other for non-intra-coded blocks), which we wlll assume are fixed for the entire sequence. Each matrix is composed of 64 elements; and the step size of the quantizer used to encode the ith DCT coefficient in a block is specified by the product of the appropriate quantization matrix entry, Q_i, and a scale factor, *mquant*.

The goal of MPEG rate control is to regulate DCT quantization, so as to maintain relative constancy in the encoded bitrate. Nearly all of the numerous MPEG rate control algorithms that have been proposed (including TM5) take a similar approach. First (at the GOP layer), a fixed number of bits is allocated for each GOP. Next (at the picture layer), the encoder buffer fullness is monitored after each picture is encoded, and the bit allocation for each picture type (I, P, and B) is updated. At the third level (the macroblock layer), the encoder buffer fullness is monitored after each macroblock is encoded, to obtain an individual *mquant* value for that macroblock. This *mquant* may also be additionally modulated based on local image content.

In the work presented here, we concentrate solely

0-8186-8183-7/97 $10.00 © 1997 IEEE

on one aspect of picture-level rate control: the translation of a target bitrate for a picture into a baseline *mquant* for that picture. We propose a new more accurate bitrate-*mquant* model. This model calculates the *mquant* which, when used for encoding every macroblock in a picture, will result in an encoded bitrate for the picture as close as possible to the target bitrate.

Note that, as in TM5, our baseline *mquant* could be further modulated based on local image content.

2.2. A new block-level rate-quantization model

Let $\{x_1, ..., x_N\}$ be a block of N input samples, and $\{X_1, ..., X_N\}$ be their DCT transform coefficients. (For an 8×8 DCT, as used in MPEG, $N = 64$.) As described in the previous section, the quantizer step size used to quantize X_i is given by $Q_i \times mquant$. Since every Q_i is assumed to be fixed for the sequence, rate control within a picture becomes the problem of choosing a suitable *mquant* for each macroblock in the picture.

For individual quantization of each X_i, applying the rate-distortion function for a Gaussian random variable with squared distortion, we have the following expressions for rate and distortion in a block:

$$R = \sum_{i=1}^{N} \frac{1}{2} \log_2 \frac{\sigma_i^2}{d_i} \tag{1}$$

and

$$D = \sum_{i=1}^{N} d_i , \tag{2}$$

where σ_i^2 is the variance of X_i, and d_i is the power of the distortion induced by quantization of X_i.

In intra-coded MPEG macroblocks, the DC DCT coefficient (X_1) is treated separately from the AC coefficients (X_2, ..., X_{64}). In an intra-coded block, X_1 is quantized using a fixed quantizer step size. For this special case, we then write for rate and distortion:

$$R = r_1 + \sum_{i=2}^{N} \frac{1}{2} \log_2 \frac{\sigma_i^2}{d_i} \tag{3}$$

and

$$D = d_1 + \sum_{i=2}^{N} d_i, \tag{4}$$

where r_1 is the bit count for encoding X_1, and d_1 is the power of the distortion induced by quantization of X_1. Since the quantizer step size used for X_1 is constant throughout a sequence, we will assume that d_1 is constant. Our approach for handling r_1 will be discussed in Section 2.3..

At relatively high bitrates, DCT coefficient quantization induces quantization noise which is approximately uniformly distributed over the quantization interval, with quantization noise power $d_i = mquant^2 \times Q_i^2/12$. At low bitrates (e.g. below 1 bit/sample), the uniform distribution assumption on quantization noise does not hold, and we assume that $d_i \propto mquant^2$. Hence, we can represent the quantization noise power at all bitrates by

$$d_i = \gamma_i \times mquant^2 . \tag{5}$$

Since block content can differ widely across an image, one cannot assume all blocks to be realizations of the same random process, with equal transform coefficient variance. To estimate the transform coefficient variance one could use a classification algorithm [5], but the computational burden of this approach is high. We will instead assume that all pixels in a block are drawn from the same random process, and estimate the transform coefficient variance from the block pixel variance σ^2 (which is easily computed). This is accomplished via the parameter κ, which is defined for non-intra-coded blocks by

$$\kappa = \frac{\sigma^2}{\sqrt[N]{\prod_{i=1}^{N} \sigma_i^2/\gamma_i}},$$

and for intra-coded blocks, by

$$\kappa = \frac{\sigma^2}{\sqrt[N-1]{\prod_{i=2}^{N} \sigma_i^2/\gamma_i}}.$$

Then for both non-intra-coded and intra-coded blocks, we have:

$$R = \frac{\hat{N}}{2} \log_2 \frac{\sigma^2}{\kappa \times mquant^2} , \tag{6}$$

where $\hat{N} = N - 1$ for intra-coded blocks, and $\hat{N} = N$ otherwise. This is our block-level rate-quantization model.

2.3. The rate control algorithm

Given the picture bit budget R_f, we propose here a method, based on the block-level rate-quantization model introduced in the previous section, to select the baseline *mquant* to be used for all macroblocks in a picture.

Suppose a total of M blocks are to be quantized (including all color channels). Let σ_j, κ_j and $\hat{N}_j$ indicate, for block j, σ, κ and $\hat{N}$ as defined in section 2.2., and let the *coding parameter* α be defined such that

$$\sum_{j=1}^{M} \frac{\hat{N}_j}{2} \log \alpha = \sum_{j=1}^{M} \frac{\hat{N}_j}{2} \log(\kappa_j). \tag{7}$$

The rate control problem is then to choose the smallest *mquant* such that

$$\sum_{j=1}^{M} \frac{\hat{N}_j}{2} \times \log_2 \frac{\sigma_j^2}{\alpha \times mquant^2} \leq R_f. \qquad (8)$$

Because of the monotonicity of the rate-quantization relationship (eq. (6)), this minimization can be easily performed using a search algorithm such as divide-and-conquer.

However, before we can perform the above minimization, there are two issues remaining. Both will be resolved by appealing to the temporal stationarity typical in video. First, in our discussion above, we neglected the bits spent on coding DC coefficients in intra-coded blocks, as well as overhead bits such as those used to encode motion vectors. Fortunately, we have empirically found little variation in these bit counts for pictures of the same type within a sequence. Therefore, when encoding a picture, the number of overhead bits (including DC bits in intra-coded blocks) used in the previous picture of the same type is simply subtracted from the bit budget R_f in eq. (8).

Key to the rate control algorithm described above is the coding parameter α. Since macroblocks of different types are assumed to rise from different stationary processes, a different coding parameter value is used for each picture type: α_k, $k = I, P, B$. One way to estimate the α_k parameters is to use pre-coding passes to generate an (R, D) pair for each, but this introduces extra computation and hence encoding delay. However, we have found that the α_k parameters exhibit little variation from picture to picture, as is demonstrated in Fig. 1. Therefore, when encoding a picture of type k, we will use the α_k computed after encoding the previous picture of the same type. The update rule we use to compute the coding parameter for the lth picture of type k is given by:

$$\alpha_k(l) = \alpha_k(l-1) \times 4^{(R_e(l) - R_a(l))/\sum_{i=1}^{M} \hat{N}_i} \qquad (9)$$

where $R_e(l)$ is the estimated number of encoded bits for picture l, and $R_a(l)$ is the actual number of encoded bits for picture l, for the *mquant* determined from eq. (8). This update rule requires initial values $\alpha_I(0)$, $\alpha_P(0)$, and $\alpha_B(0)$. These are obtained by performing, at the beginning of a sequence (or after a scene change), three pre-coding passes, one for the first picture of each picture type.

The rate control algorithm presented here is computationally efficient. It requires the variance of each block, which requires one multiplication per pixel. Using eq. (8) for *mquant* selection, requires $M+1$ multiplications and logarithm operations, $2M-1$ additions,

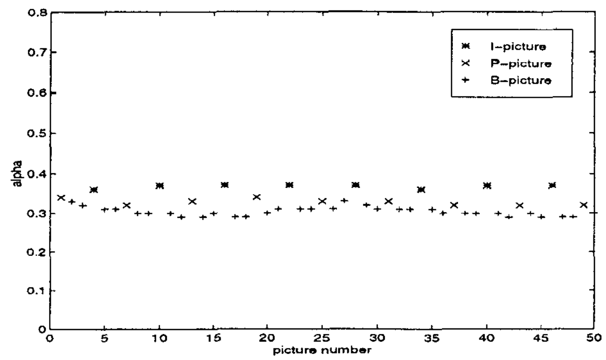

Figure 1. Coding parameter α.

and N_s multiplications and logarithm operations, where N_s is the number of search steps. Since there are only 31 legal *mquant* values in MPEG, N_s is upperbounded by 31. In practice, by starting from the *mquant* value used by the previous picture of the same type, and employing a divide-and-conquer search algorithm, we found three search steps are typically enough.

3. Experimental Results and Conclusions

We have applied the rate control algorithm described above to MPEG2 encoding, and compared it to the TM5 rate control algorithm, and to a pre-coding-based rate control algorithm. All three algorithms use the same GOP and picture-level bit allocation (that used in TM5), but each uses a different technique for macroblock-level *mquant* selection. Our algorithm and the pre-coding algorithm both use one constant *mquant* value to encode all the macroblocks in a picture, and do not use any macroblock-level modulation of *mquant*. The TM5 algorithm employs macroblock-level *mquant* modulation based on buffer fullness.

The pre-coding rate control technique we have used for comparison is based on the rate-quantization model proposed in [3]: $R = \alpha + \beta/mquant^\gamma$, where γ was chosen to be 1. Two pre-codings are used to solve this equation for α and β; and the resulting model is used to translate the desired bitrate for a picture into an *mquant* value. (We only use this model for illustrative purposes. The actual algorithm in [3] specifies *mquant* up to once per slice.)

The TM5 rate control algorithm calculates a picture-level *mquant* using the empirical formula: $mquant = d \times 31/r$, where r is a reaction parameter, and d is the virtual buffer fullness. After a macroblock is en-

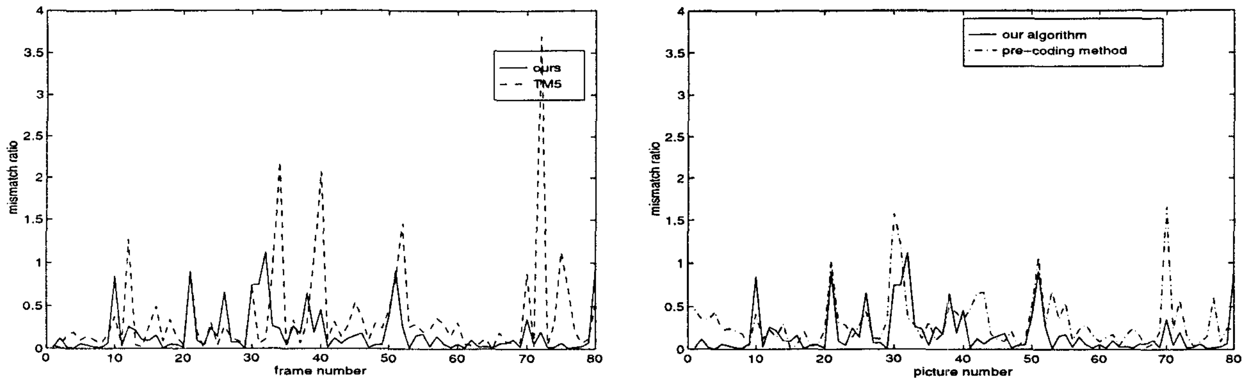

Figure 2. Bit budget mismatch comparisons.

coded, TM5 further adjusts this *mquant* value, based on buffer fullness. To make the comparisons here fair, the macro-block level, local-variance-based *mquant* modulation technique in TM5 was disabled.

For our experiments we used the Mobile & Calendar video test sequence (80-pictures, 4:2:0 chromatic sampling, 720 $\times$ 480 pixels), encoded at 6×10^6 bits/s, with GOP configuration IBBPBB. Shown in Fig. 2 are the results of encoding this sequence using each of the three rate control algorithms. These plots show, for each picture in the encoded sequence, the mismatch ratio between the target bitrate for that picture, and the actual number of bits used to encode the picture, using the *mquant* values calculated by each algorithm. (The mismatch ratio for l_{th} picture is defined as $|R_a(l) - R_f(l)|/R_f(l)$.) The total percentage mismatch over the entire 80-picture sequence was 9.97% for our algorithm, 15.07% for the pre-coding algorithm, and 18.71% for the TM5 algorithm. (The total percentage mismatch is defined as $\frac{\sum_l |R_a(l) - R_f(l)|}{\sum_l R_f(l)}$.) The comparison of peak signal-to-noise ratios of our algorithm and the TM5 algorithm is shown in Fig. 3. On average, our algorithm is 0.9 dB better than the TM5 algorithm, and 0.7 dB better than the pre-coding method. Similar results were obtained using the Flower Garden sequence.

These experimental data clearly demonstrate the improvement in bitrate estimate accuracy and peak signal-to-noise ratio gained by using our rate-quantization model. This is especially notable, considering that our algorithm uses no pre-coding, and no macroblock-level *mquant* adjustment.

Acknowledgements

We thank Dr. Jungwoo Lee at Sarnoff Corporation for helpful discussions on video coding.

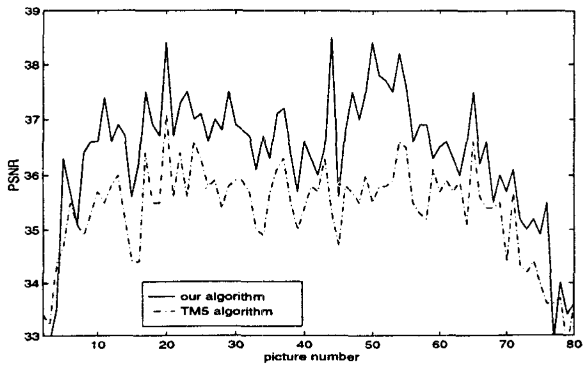

Figure 3. PSNR comparison.

References

[1] C. Chen and A. Wong. A self-governing rate buffer control strategy for pseudoconstant bit rate video coding. *IEEE Trans. Image Processing*, 2(1), 1990.

[2] A. Puri and R. Aravind. Motion-compensated video with adaptive perceptual quantization. *IEEE Trans. Cir. Sys. Video Tech.*, 1(4), 1991.

[3] W. Ding and B. Liu. Rate control of MPEG video coding and recording by rate-quantization modeling. *IEEE Trans. Cir. Sys. Video Tech.*, 6(1), 1996.

[4] L. Wang. Rate control for MPEG video coding. In *Visual Comm. and Image Processing, SPIE vol. 2501*, 1995.

[5] W. Chen and C. Smith. Adaptive coding of monochrome and color images. *IEEE Trans. Comm.*, 25(11), 1977.

[6] J. Huang and P. Shultheiss. Block quantization of correlated Gaussian random variables. *IEEE Trans. Comm.*, 11(9), 1963.

[7] ISO-IEC/JTC1/SC29/WG11. *Coded representation of pictures and audio information, Test Model 5*, 1993.