

LOW-COMPLEXITY RATE-DISTORTION OPTIMAL MACROBLOCK MODE SELECTION FOR MPEG-LIKE VIDEO CODERS

Hyungjoon Kim and Yucel Altunbasak

Center for Signal and Image Processing
Georgia Institute of Technology
Atlanta, GA 30332-0250

ABSTRACT

Constrained optimization techniques provide an effective solution to the rate-distortion optimized mode selection in hybrid motion compensation-DCT based video encoders. However, the minimization of the Lagrangian cost function often presents a computational burden in real-time applications. In this paper, we propose a new method for rate-distortion optimized coding mode selection. Instead of using the Lagrangian method, we employ a computationally efficient, yet highly effective rate-distortion model to minimize the overall distortion of a macroblock. Extensive experimental results show that the proposed method provides PSNR improvements of 2.79, 1.64, and 1.54 dB over TM5-based MPEG-2 encoder, at the bit-rates of 2, 4, and 6 Mbps, respectively, with negligible increase in computational complexity.

1. INTRODUCTION

A typical commercial video encoder includes such processing modules as preprocessing, motion estimation, coding mode selection, and rate control. Among these, coding mode selection is a non-normative processing block, which proves to be critical to efficiently encode scenes with widely varying content and motion.

In the Test Model 5 (TM5) [5] for MPEG-2, a mode is selected based on residual energy. However, since this mode selection strategy does not consider any rate constraints, the selected mode does not result in the optimal coding performance. To this end, Wiegand et al. [6] proposed a rate-distortion (R-D) optimal mode selection scheme that provides significant improvements in PSNR, especially in very low bit-rate video coding. However, this scheme has high computational complexity that makes it impractical in real-time implementations. Other researchers also proposed constrained optimization methods along the same lines, which

typically simulate computationally expensive operations such as DCT, quantization etc. In this paper, instead of using Lagrangian minimization, we employ a simple and yet highly effective rate-distortion model to develop a computationally efficient R-D optimized mode selection framework. Here, the model parameters and the minimum cost coding mode are easily calculated in terms of computational complexity. Experiment results show that coding gains of up to 2.79 dB are achievable with little increase in the computational complexity of the encoder. To the best of our knowledge, these are the best results reported to date at this level of complexity.

2. RATE-DISTORTION MODEL

In this work, we consider R-D functions of the form :

$$D(R) = \sigma^\beta e^{-\gamma R}, \quad (1)$$

where σ denotes the standard deviation of the source and β and γ are unknown model parameters. This form of R-D function is chosen based on information-theoretic analysis. As studied in [3], [7], distributions of the DCT transform coefficients are best approximated by a generalized Gaussian distribution. It can be shown that $\beta = 2$ and $\gamma = 2 \ln(2)$ for a perfect Gaussian source, and $\beta = 1$ and $\gamma = \frac{3}{2} \ln(2)$ for a perfect Laplacian source.

In our work, we consider only the $\beta = 2$ case and use γ as a coding parameter to control the model accuracy since the actual distributions of DCT coefficients are not exactly Laplacian or Gaussian. We can estimate the actual value of γ for a frame by using the encoding statistics of previous frames of the same picture type. The estimate of γ is :

$$\begin{aligned} \gamma &= \frac{1}{N} \sum_{i=1}^N \gamma_i \\ &= \frac{1}{N} \sum_{i=1}^N \frac{1}{R_i} \ln \left(\frac{\sigma_i^2}{D_i} \right) \end{aligned} \quad (2)$$

This work is supported by NSF under NSF CAREER award CCR-0133221.

where γ_i is the γ parameter associated with the i^{th} macroblock and D_i , R_i , and σ_i^2 are the actual distortion, rate, and source variance for the i^{th} macroblock in the last coded frame of the same type, respectively.

3. MACROBLOCK MODE SELECTION

Consider a group of N macroblocks to be encoded in a frame. Let m_i be the coding mode of the i^{th} macroblock, ($i = 1, 2, \dots, N$), and let M_N be the set of the modes of all macroblocks. That is,

$$M_N = \{m_1, m_2, \dots, m_N\},$$

where m_i is the mode of the i^{th} MB. The problem of finding the R-D optimal set of the modes (M_N^*) for the group of N macroblocks can be formulated as :

$$M_N^* = \arg \min_{M_N} D(M_N) = \arg \min_{M_N} \sum_{i=1}^N D_i(m_i), \quad (3)$$

subject to

$$R(M_N) \leq R^{\text{total}},$$

where $D(M_N) = \sum_{i=1}^N D_i(m_i)$ and $R(M_N) = \sum_{i=1}^N R_i(m_i)$ represent the sum of the distortions and the rates of N macroblocks, respectively. $D_i(m_i)$ denotes the distortion when the i^{th} macroblock is coded in the mode m_i . Similarly, $R_i(m_i)$ represents the rate of the macroblock in the mode m_i . R^{total} is the available total bit budget to encode the set of N macroblocks. The bit budget is shared to encode the DCT, the motion vector, and the header information. So, we can write :

$$R(M_N) = \sum_{i=1}^N R_i^{mv}(m_i) + \sum_{i=1}^N R_i^{dct}(m_i) + \sum_{i=1}^N R_i^{hdr}(m_i) + R^{\text{misc}},$$

where $R_i^{mv}(m_i)$, $R_i^{dct}(m_i)$, and $R_i^{hdr}(m_i)$ denote the rates needed to encode motion vectors, DCT coefficients, and headers, respectively, associated with the i^{th} macroblock when it is coded in mode m_i . R^{misc} represents the rate for coding other information that is not related to the macroblocks, *e.g.*, the sequence/picture/slice headers.

Assuming that current macroblock mode m_i is independent of any of the other macroblocks, we can modify the constrained minimization problem of Eq. 3 as :

$$M_N^* = \sum_{i=1}^N \arg \min_{m_i} D_i(m_i),$$

subject to

$$R(M_N) \leq R^{\text{total}}.$$

Further simplification is possible if we assume that the target total number of bits for the i^{th} macroblock (R_i^T) is known. With this assumption, the rate constraint simplifies to :

$$R_i^{mv}(m_i) + R_i^{dct}(m_i) + R_i^{hdr}(m_i) \leq R_i^T, \quad \forall i = 1, \dots, N.$$

Coding mode (m_i^*) of each macroblock is obtained by solving the following constrained minimization problem :

$$m_i^* = \arg \min_{m_i} D_i(m_i), \quad (4)$$

subject to

$$R_i^{mv}(m_i) + R_i^{dct}(m_i) + R_i^{hdr}(m_i) \leq R_i^T.$$

In the MPEG standards, for an Intra type macroblock, the distortion of a macroblock is composed of the distortion due to encoding the DC coefficients and the AC coefficients. This is because of the fact that encoding the DC coefficients of each block in a macroblock is different than the rest of the DCT coefficients. The DC coefficients of the macroblock are quantized with a fixed step size (either 1, 2, or 4 depending on the DC precision determined by the user) and differentially encoded. Thus, we can write the macroblock distortion as :¹

$$D_i(m_i) = D_i^{DC}(m_i) + D_i^{AC}(m_i),$$

where $D_i^{DC}(m_i)$ and $D_i^{AC}(m_i)$ are the distortions of DC components and AC components, respectively. However, for an intra-coded macroblock, the value of D_i^{DC} is generally small since the DC coefficients are quantized with a small step size. Furthermore, the values of D_i^{DC} are relatively smaller when compared with $D_i^{AC}(m_i)$. Thus, we can assume that :

$$D_i(m_i) = D_i^{AC}(m_i). \quad (5)$$

For the distortion of the AC components, we use the R-D relation presented in Section 2 :

$$D_i^{AC}(m_i) = \sigma_i^2 e^{-\gamma R_i^{AC}}, \quad (6)$$

where R_i^{AC} is a rate for AC components for a given mode m_i . In this equation, σ_i^2 is the variance of the AC coefficients that depends on the mode. We finalize the R-D formulation by substituting Eq. 6 in Eq. 5 to get :

$$D_i(m_i) = \sigma_i^2(m_i) e^{-\gamma(R_i^T - R_i^{mv}(m_i) - R_i^{hdr}(m_i) - R_i^{DC}(m_i))} = \sigma_i^2(m_i) e^{-\gamma R_i^T} e^{\gamma(R_i^{mv}(m_i) + R_i^{hdr}(m_i) + R_i^{DC}(m_i))}, \quad (7)$$

¹ $D^{DC} = 0$ for all non-intra modes

with $R_i^{AC} = R_i^T - R_i^{mv} - R_i^{hdr} - R_i^{DC}$. R_i^{DC} is a number of bits needed to DC components which is zero for non-intra modes. Using this model, in the optimization problem formulated in Eq. 4, we get :

$$m_i^* = \arg \min_{m_i} \left\{ \sigma_i^2(m_i) e^{\gamma(R_i^{mv}(m_i) + R_i^{hdr}(m_i) + R_i^{DC}(m_i))} \right\}. \quad (8)$$

Note that $e^{-\gamma R_i^T}$ is not related to any coding modes and it can be removed from the distortion minimization. That is, R_i^T does not affect our proposed mode selection algorithm and is determined by a rate control scheme.

Eq. 8 formulates the rule for choosing the best coding mode for the i^{th} macroblock. To solve Eq. 8 for the optimal macroblock mode, we need variance for each mode, and the number of bits for encoding the MVs, header, and the DC coefficients (for Intra mode only). We also need to estimate the model parameters. At this point, we assume that the motion estimation is already done so that the motion vector and the residual is determined. Using the encoding tables we can calculate the number of bits for the motion vector(s) and the header. We can also determine the rate for the DC coefficients since DC quantization does not depend on the choice of the quantization parameter. Estimation of γ was already discussed in Section 2.

4. COMPUTATIONAL COMPLEXITY

In this section, we discuss the complexity of the proposed algorithm. The proposed algorithm requires additional computations such as computation of source variance and distortion, and determination of the rate terms. The analysis of the computational complexity is as follows :

- Calculation of number of bits for MVs, DC coefficients, and the header : In the MPEG-2 standard, MVs and DC coefficients are coded in a differential format. The number of bits can be easily obtained from a look-up table if the differences are available. We can also easily calculate the number of bits for the header from the specification of MPEG-2 [4]. Thus, the computational complexity for these calculations can be ignored.
- Calculation of variances : For each candidate mode, we need to compute the variance of the DCT coefficients. However, since the DCT coefficients are not available before the mode selection, we need to approximate the variance in the pixel-domain. For the calculation of the approximated variance of an 8×8 block, we use :

$$\sigma^2 \approx \sum_{i=0}^{63} x_i^2,$$

where x_i is the pixel value or the pixel difference value for intra mode and non-intra modes, respectively. To perform this computation, we need 64 multiplications and 63 additions per block for the intra mode. In the case of a non-intra mode, we need an extra 64 additions to compute the pixel differences, therefore the total numbers of operations per block for one candidate mode are 64 multiplications and 127 additions for the non-intra modes. For a 4:2:0-type macroblock there are six blocks, thus, the numbers of operations required per macroblock for one mode are 64×6 multiplications and 63×6 additions for the intra mode and 64×6 multiplications and 127×6 additions for the non-intra modes. For a P-type frame, we consider three non-intra modes and one intra mode, so the total numbers of operations per macroblock are $64 \times 6 \times 4$ multiplications and $127 \times 6 \times 3 + 63 \times 6$ additions. For a B-type frame, there are six non-intra modes and one intra mode, so the total numbers of operations per macroblock are $64 \times 6 \times 7$ multiplications and $127 \times 6 \times 6 + 63 \times 6$ additions.

- Calculation of distortions and choosing a mode : To calculate the distortion of a candidate mode using Eq. 7, we need two multiplications, two addition and one exponentiation. Thus, we need totally 5×4 and 5×7 operations per macroblock for P-type frame and B-type frame, respectively. These numbers are relatively small when compared to the number of operations required for computing the variances of each candidate macroblock mode.
- Estimation of the model parameter γ : The model parameter γ is estimated using Eq. 2 after encoding a frame. This requires the rate, variance, and distortion. We need to calculate only the distortion since the actual rates are available after coding a frame and the variances are already calculated in the mode selection module. To calculate the distortion (sum of squared difference), for a 4:2:0-type macroblock, we need 64×6 multiplications and 127×6 additions per macroblock. For the calculation of the γ , we need $2N + 1$ multiplication, $N - 1$ additions, and N logarithm operations.

We observe that the majority of the overall computational cost comes from the macroblock variance computation for each candidate mode. However, this computational complexity is significantly less when compared to the complexity of the DCT, quantization, and VLC operations. A state-of-the-art hardware or software processor can easily realize this level of computational cost. ²

²The proposed algorithm has been successfully implemented using DSP processors. The reader is referred to contact yucel@ece.gatech.edu if (s)he would like to get a copy of the code.

Bit rates	Algorithms	SEQ-1	SEQ-2	SEQ-3
2 Mbps	TM5	28.35	26.38	30.13
	Mod. TM5	29.36	27.13	29.92
	Proposed	31.64	28.49	33.09
4 Mbps	TM5	33.35	30.17	35.34
	Mod. TM5	34.72	31.28	35.90
	Proposed	35.32	31.69	36.76
6 Mbps	TM5	35.44	32.15	37.61
	Mod. TM5	37.04	33.53	38.27
	Proposed	37.44	33.82	38.57

Table 1. The average PSNR values for (i) TM5, (ii) Modified TM5, and (iii) the proposed mode selection algorithm

5. EXPERIMENTAL RESULTS

We implemented the ρ -domain rate control scheme [2] into the software MPEG-2 encoder that was developed at the University of California, Berkeley [1]. The proposed mode selection algorithm is incorporated into the software encoder with ρ -domain rate control which we will refer to as “Modified TM5”. In the test set, we used three sequences with CCIR-601 parameters: CAROUSEL (SEQ-1), FLOWER GARDEN (SEQ-2), and FOOTBALL (SEQ-3). In all experiments, the number of encoded frames was 103. We chose a GOP size of 15. We considered three bit rates of 2, 4, and 6 Mbps. The search range was selected as 63×63 for both P- and B-type frames, and the alternate scan option was turned on for all cases.

Table 1 shows the PSNR performance comparison for (i) TM5, (ii) Modified TM5, and (iii) proposed mode selection method. With the proposed method, we achieve 2.79, 1.64, and 1.54 dB PSNR gains over TM5 on the average, at the bit rates 2, 4, and 6 Mbps, respectively. The average PSNR gains over the modified TM5 encoder are 2.27, 0.62, and 0.33 dB at bit rates 2, 4, and 6 Mbps, respectively. Especially at 2 Mbps, the PSNR improvement of the proposed algorithm is quite substantial. The average of 2.27 dB gain at 2 Mbps corresponds to at least 20% bit rate savings. Overall, these results justify that the proposed R-D optimization can provide significant benefits, especially at low rates. Fig. 1 illustrates the PSNR values for the CAROUSEL at 2Mbps.

6. CONCLUSION

In this paper, using an accurate rate-distortion model we proposed a computationally efficient R-D optimized mode selection method for MPEG-like video encoders. Based on our extensive tests and computational analysis, we demonstrated that the proposed algorithm can achieve significant

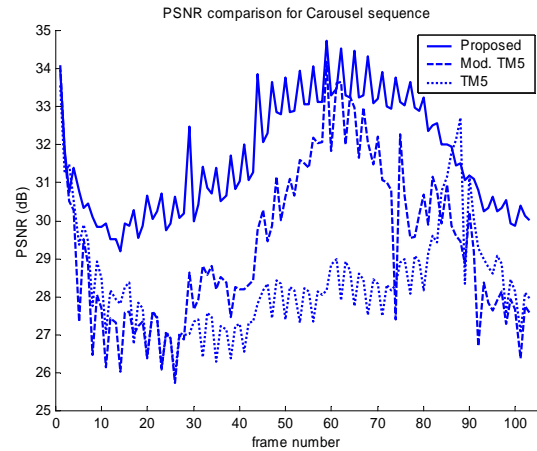


Fig. 1. The PSNR comparison for Carousel sequence. The bit rate is 2 Mbps.

coding gains and bit rate savings while keeping the computational complexity manageable.

7. REFERENCES

- [1] MPEG codec, <http://bmrc.berkeley.edu/projects/mpeg/>
- [2] Z. He, Y. K. Kim, and S. K. Mitra “Low delay rate control for DCT video coding via ρ -domain source modeling,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 8, Aug. 2001.
- [3] E. Y. Lam, and J. W. Goodman, “A mathematical analysis of the DCT coefficient distributions for images,” in *IEEE Trans. Image Proc.*, vol. 9, pp. 1661-1666, Oct. 2000.
- [4] “The MPEG-2 international standard,” ISO/IEC, Reference number ISO/IEC 13818-2, 1996.
- [5] “Coded representation of picture and audio information - MPEG-2 test model 5,” ISO-IEC AVC-491, Apr. 1993.
- [6] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell, and S. K. Mitra, “Rate-distortion optimized mode selection for very low bit rate video coding and the emerging H.263 standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 449-458, Apr. 1997.
- [7] G. S. Yovanof, and S. Liu, “Statistical analysis of the DCT coefficients and their quantization error,” *Conf. Rec. 30th Asilomar Conf. Signals, Systems, Computers*, vol. 1, pp. 601-605, 1997.