

RATE-DISTORTION OPTIMAL JOINT MACROBLOCK MODE SELECTION AND MOTION ESTIMATION FOR MPEG-LIKE VIDEO CODERS

Hyungjoon Kim and Yucel Altunbasak

Center for Signal and Image Processing
Georgia Institute of Technology
Atlanta, GA 30332-0250
hyung.yucel@ece.gatech.edu

ABSTRACT

Rate-Distortion optimization can significantly improve encoder performance in MPEG-like video coding applications especially when it is applied to coding mode selection and motion estimation. Nevertheless, it is not easy to implement Rate-Distortion optimization in real-time video encoders because of its prohibitive computational cost. In this paper, we propose a joint macroblock mode selection and motion estimation by employing a computationally-efficient, yet highly-effective rate-distortion model. Experimental results show that the proposed algorithm provides significant PSNR improvements with a small increase in overall computational load over the TM5-based encoder.

1. INTRODUCTION

Video coding standards define converting a raw video source into a specified bitstream. Although the bitstream syntax is specified by the standard, many of the encoder modules are left open to allow manufacturers to optimize the encoding performance or to tailor it to a particular application. Among the non-normative encoder modules, the motion compensated prediction module produces a motion vector and a prediction residual. Assuming that the number of bits needed to code the motion information is negligible, some existing hybrid video coding algorithms (*e.g.*, TM5 [9] for MPEG-2) have been designed to select a displacement vector that produces a minimal-energy residual. However, as the overall bit rate decreases, coding the motion information also needs a certain number of bits that becomes relatively substantial. Thus, a better strategy for the motion search is required to find a balance between the residual energy and the motion information. Another important block in the encoding process is macroblock (MB) mode selection. The distortion-based approach of TM5 for the coding mode selection can be improved by applying Lagrangian-based Rate-Distortion (R-D) optimization (RDO) as suggested in [7], [10]. However, it would be inefficient to calculate the actual rates and distortions by simulating the encoding process to minimize Lagrangian cost function associated with each possible encoding mode. In this paper, we propose a joint MB mode selection and motion estimation algorithm by employing a computationally-efficient, yet highly-effective rate-distortion model.

The contributions of this paper are as follows: (i) we introduce a simple, yet highly effective R-D model; (ii) we develop a computationally efficient R-D optimized mode selection and motion estimation framework using the proposed R-D model. Unlike the most R-D optimal mode selection algorithms in the literature,

the proposed one is *not* based on Lagrangian optimization; (iii) we effectively combine the non-Lagrangian MB mode selection with the Lagrangian motion estimation method; (iv) we show that coding gains of up to 3.17 dB are achievable with modest increase in the computational complexity of the encoder; and (v) a real-time hardware implementation using the proposed algorithm as well as all other MPEG-2 modules have been successfully implemented using DSP processors¹.

2. RATE-DISTORTION MODEL

In this work, we propose to use the rate distortion functions of the form:

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

where σ denotes the standard deviation of the source and β and γ are the model parameters. This form of R-D function is chosen based on information-theoretic analysis. As studied in [6], [11], distribution of the transform coefficients is best approximated by a generalized Gaussian distribution. In the following, we consider two special cases of the generalized Gaussian source :

- Laplacian source: According to Shannon's source coding theorem [4], for a Laplacian source, the minimum number of bits (R) needed to represent a symbol is given by

$$R = \log_2 \left(\frac{\sigma}{\sqrt{2D}} \right),$$

where σ is the standard deviation of the source and D is the distortion due to the quantization of the coefficients. For the distortion measure, mean absolute difference is employed. It immediately follows that D is given in terms of the rate to encode the coefficients as

$$D = \sigma e^{-\frac{3}{2} \ln(2)R}.$$

This equation suggests that, $\beta = 1$ and $\gamma = \frac{3}{2} \ln(2)$ for a perfect Laplacian source.

- Gaussian source: For a Gaussian source, the R-D function with squared error distortion (D) is

$$R = \frac{1}{2} \log_2 \left(\frac{\sigma^2}{D} \right),$$

¹The reader is referred to contact yucel@ece.gatech.edu if (s)he would like to get a software copy of the code.

where σ^2 is the source variance and R is a rate. The distortion is given in terms of the rate as

$$D = \sigma^2 e^{-2 \ln(2)R}.$$

For a perfect Gaussian source, this equation suggests that $\beta = 2$ and $\gamma = 2 \ln(2)$.

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

$$\gamma = \frac{1}{N} \sum_{i=1}^N \gamma_i = \frac{1}{N} \sum_{i=1}^N \frac{1}{R_i} \ln \left(\frac{\sigma_i^2}{D_i} \right),$$

where γ_i is the model parameter for the i^{th} MB, and D_i, R_i , and σ_i^2 are the actual distortion, rate and source variance for the i^{th} MB in the last coded frame, respectively.

3. MACROBLOCK MODE SELECTION AND MOTION ESTIMATION

3.1. R-D optimized Mode Selection

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

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

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

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

subject to

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

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 MBs, respectively. $D_i(m_i)$ denotes the distortion when the i^{th} MB is coded in the mode m_i . Similarly, $R_i(m_i)$ represents the rate of the MB in the mode m_i . R^{total} is the available total bit budget to encode the set of N MBs. The bit budget is shared to encode the DCT, the motion vector and the header information. So we can write

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

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} MB when it is coded in mode m_i . R^{misc} represents the rate for coding other information that is not related to the MBs, e.g., the sequence/picture/slice headers.

By assuming that current MB mode m_i is independent of any of the other MBs, we can modify the constrained minimization problem of Eq. 1 as

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

subject to

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

Further simplification is possible if we assume that the target total number of bits for the i^{th} MB (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 MB is then obtained by solving the following constrained minimization problem:

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

subject to

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

In the MPEG standard, for an Intra type MB, the distortion of a MB is composed of the distortion due to encoding the DC coefficient and the AC coefficients. This is because of the fact that encoding the DC coefficient of each block in a MB is different than the rest of the DCT coefficients. The DC coefficients of the MB 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 MB distortion as:²

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

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 MB, 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} is relatively smaller when compared with $D_i^{AC}(m_i)$. Thus, we can assume that

$$D_i(m_i) = D_i^{AC}(m_i). \quad (2)$$

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

$$D_i^{AC}(m_i) = \sigma_i^2(m_i) e^{-\gamma R_i^{AC}}, \quad (3)$$

where R_i^{AC} is a rate for AC components for a given mode m_i . In this equation, $\sigma_i^2(m_i)$ is the variance of the DCT coefficients that depends on the mode. We finalize the R-D formulation by combining Eq. 3 and Eq. 2 to get

$$\begin{aligned} D_i(m_i) &= \sigma_i^2(m_i) e^{-\gamma(R_i^T - R_i^{mv}(m_i) - R_i^{hdr}(m_i) - R_i^{DC}(m_i))} \\ &= \sigma_i^2(m_i) e^{-\gamma R_i^T} e^{\gamma(R_i^{mv}(m_i) + R_i^{hdr}(m_i) + R_i^{DC}(m_i))}, \end{aligned} \quad (4)$$

with $R_i^{AC}(m_i) = R_i^T(m_i) - R_i^{mv}(m_i) - R_i^{hdr}(m_i) - R_i^{DC}(m_i)$. $R_i^{DC}(m_i)$ is a number of bits needed to DC components which is

² $D^{DC} = 0$ for all non-intra modes

zero for non-intra modes. Using this model, in the optimization problem formulated in Eq. 2, we get:

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

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

Note that $e^{-\gamma R_i^T}$ is not related to any coding modes and it can be removed from the distortion minimization. Eq. 5 formulates the rule for choosing the best coding mode for the i^{th} MB.

3.2. Joint Mode Decision and Motion Estimation

In Section 3.1, we assumed that the motion vectors are separately determined by a motion estimation algorithm. For better encoding performance, we can combine motion estimation and mode decision in a joint algorithm.

For the i^{th} MB, we can state the R-D optimal motion estimation as estimating the motion vector that minimizes the mean absolute error, E_i , subject to a motion vector rate constraint. This constrained problem is converted to an unconstrained problem using a Lagrangian multiplier, thus, the R-D optimal motion vector is selected so as to minimize the Lagrangian cost function

$$J_i = E_i + \lambda R_i^{mv}, \quad (5)$$

where λ is the unknown Lagrangian multiplier and R_i^{mv} is the number of motion vector bits. λ can be viewed as a factor that determines the relative importance of the rate and the distortion. If $\lambda = 0$, then the rate constraint is ignored. Several methods were proposed for λ estimation [2], [3], [10]. Another approach is to use a set of possible λ values, $\lambda_1, \lambda_2, \dots, \lambda_M$, and select the motion vector that minimizes the cost J_i . That is, for each λ_j , $j = 1, 2, \dots, M$, we can find a number of bits required to encode motion vector, MV_j that minimizes the cost $J_{i,j} = E_{i,j} + \lambda_j R_{i,j}^{mv}$, where $R_{i,j}^{mv}$ is the number of bits required to encode MV_j for the i^{th} MB. This procedure results in a maximum of M candidate motion estimates. We can combine the mode decision and motion estimation by using all candidate motion vectors determined for each λ_j and jointly optimize the MB mode and the motion vectors by evaluating all possible cases by extending Eq. 5 as

$$(m_i^*, MV_i^*) = \arg \min_{m_i, MV_{i,j}} \left\{ \sigma_{i,j}^2(m_i) \cdot e^{\gamma(R_{i,j}^{mv}(m_i) + R_i^{hdr}(m_i) + R_i^{DC}(m_i))} \right\},$$

where $MV_{i,j}$ is the candidate motion vector(s) for the λ_j . Fig. 1 shows the overall procedure for R-D optimal mode selection and motion estimation.

4. COMPUTATIONAL COMPLEXITY

The proposed algorithm requires such additional computations as source variance and distortion calculations, and the determination of the rate terms. The computational complexity analysis is as follows:

- Calculation of the number of bits for the MVs, the DC coefficients, and the header : In the MPEG2 standard, MVs and DC coefficients are coded in a differential format. Thus,

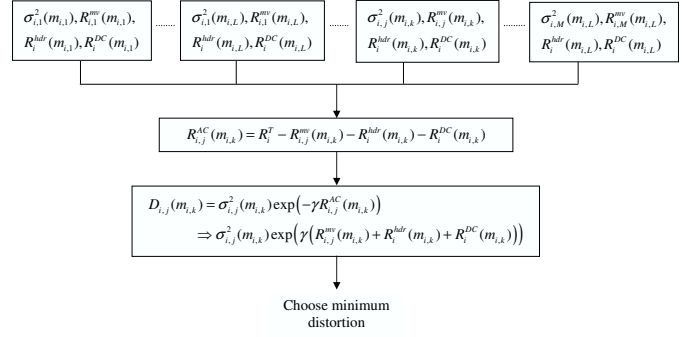


Fig. 1. R-D optimal MB mode selection and motion estimation procedure. For the i^{th} MB, $m_{i,k}$ denotes the k^{th} candidate mode and $R_{i,j}^{mv}$ is a number of bits for motion vectors associated with j^{th} λ value. L is a total number of possible MB modes and M is a total number of λ values.

the number of bits can be easily obtained from a lookup table if the differences are available. We can also easily calculate the number of header bits from the MPEG2 specification [8].

- Calculation of the variances : According to our analysis, to calculate the source variance of a 4:2:0-type MB, we need to perform a total of $384 \times (3M + 1)$ multiplications and $3M \times 762 + 378$ additions per MB for P-type frames, where M is number of λ values. Similarly, for B-type frames, there will be a total of $384 \times (6M + 1)$ multiplications and $6M \times 762 + 378$ additions per MB.
- Estimation of the model parameter γ : The model parameter γ is estimated using Eq. 1 after encoding a frame. This requires the knowledge of the rate, the variance and the distortion. Fortunately, we only need to calculate the distortion since the actual rates and variances (for the MBs) are available after coding a frame. To calculate the distortion (sum of squared difference), for a 4:2:0-type MB, we need 384 multiplications and 762 additions per MB. For the calculation of the γ , we need $2N + 1$ multiplication, $N - 1$ additions, and N logarithm operations.
- Computation of motion cost for multiple λ values : For a search point, M motion cost values for M λ s are calculated with M additions and M multiplications based on Eq. 5. This may become a large computational burden in motion estimation module in some situations. However, by replacing multiplications with table lookup and choosing moderate M values, e.g. 2 or 3, we can obtain acceptable results with very small increase in computation.

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

Bit rates	Algorithms	SEQ-1	SEQ-2	SEQ-3
2 Mbps	TM5	28.35	26.38	30.13
	ρ -RC	29.36	27.13	29.92
	Prop. $M = 10$	32.29	28.97	33.67
	Prop. $M = 2$	32.09	28.78	33.25
	Prop. $M = 1$	31.64	28.49	33.09
4 Mbps	TM5	33.35	30.17	35.34
	ρ -RC	34.72	31.28	35.90
	Prop. $M = 10$	35.64	32.00	36.96
	Prop. $M = 2$	35.51	31.84	36.76
	Prop. $M = 1$	35.32	31.69	36.76
6 Mbps	TM5	35.44	32.15	37.61
	ρ -RC	37.04	33.53	38.27
	Prop. $M = 10$	37.67	34.07	38.71
	Prop. $M = 2$	37.56	33.93	38.57
	Prop. $M = 1$	37.44	33.82	38.57

Table 1. The average PSNR values for (i) TM5, (ii) ρ -RC, (iii) the proposed algorithm with ten λ values, (iv) the proposed algorithm with two λ values, and (v) the proposed algorithm with one λ value.

5. EXPERIMENTAL RESULTS

We implemented ρ -domain rate control³ method proposed in [5] using the MPEG-2 encoder software that was developed at the University of California, Berkeley [1]. We refer to this implementation as " ρ -RC" in Table 1. The proposed mode selection and motion estimation algorithm is also incorporated into the software encoder with ρ -domain rate control. 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) ρ -RC, (iii) the proposed algorithm with ten λ values, (iv) proposed algorithm with two λ values, (v) the proposed algorithm with one λ value. For the case $M = 10$, the proposed algorithm achieves an average of 2.34 and 1.43 dB PSNR gain over the TM5 and the ρ -RC, respectively. Especially at 2 Mbps, the PSNR improvement of the proposed algorithm is quite substantial. Overall, these results demonstrates that the proposed R-D optimization can provide significant benefits, especially at low rates. Fig. 2 illustrates the PSNR values for the CAROUSEL at 2 Mbps.

Increasing the number of λ values will result in better encoding performance in terms of the picture quality, at the cost of an increased computational complexity. In our experiments, for the $M = 10$ case, we used $\{0, 10, 20, \dots, 80, 90\}$ and for the $M = 2$ case, $\{0, 90\}$ are used. When $M = 1$, λ is set to 0, which means that R-D optimized motion estimation is not used.

6. REFERENCES

[1] MPEG codec, <http://bmrc.berkeley.edu/projects/mpeg/>

³The ρ -domain rate control is one of the best performing rate control algorithms in the literature.

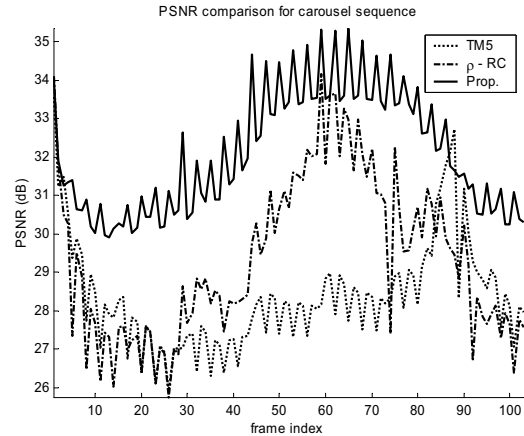


Fig. 2. The PSNR comparison for Carousel sequence. $M = 10$ and the bit rate is 2 Mbps.

- [2] M. C. Chen, and A. N. Willson, "Rate-distortion optimal motion estimation algorithms for motion-compensated transform video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, pp. 147-158, 1997.
- [3] W. C. Chung, F. Kossentini, and M. J. T. Smith, "An efficient motion estimation technique based on a rate-distortion criterion," *IEEE ICASSP-96*, vol. 4, pp. 1926-1929, 1996.
- [4] T. M. Cover and J. A. Thomas, "Elements of Information Theory," Wiley, New York, NY, 1991.
- [5] 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, Aug. 2001.
- [6] 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.
- [7] Y. W. Lee, F. Kossentini, and R. Ward, "Efficient RD optimized macroblock coding mode selection for MPEG-2 video encoding," in *Proc. Int. Conf. Image Processing*, vol. 2, pp. 803-806, 1997. 803-806
- [8] "The MPEG-2 international standard," ISO/IEC, Reference number ISO/IEC 13818-2, 1996.
- [9] "Coded representation of picture and audio information - MPEG-2 test model 5," ISO-IEC AVC-491, Apr. 1993.
- [10] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Proc. Magazine*, vol. 15, pp. 74-90, Apr. 1998.
- [11] 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.