LOW-COMPLEXITY MACROBLOCK MODE SELECTION FOR H.264/AVC ENCODERS

Hyungjoon Kim and Yucel Altunbasak

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

ABSTRACT

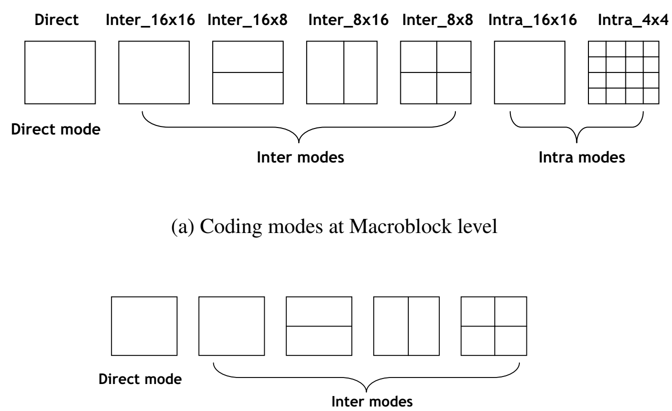
Lagrangian minimization is one of the most powerful tools for Rate-Distortion optimal coding mode selection. However, in the latest video compression standard H.264/AVC, the total calculation for costs of all candidate modes may become huge since the new standard supports a large number of coding modes. In this paper, we propose an algorithm for fast coding mode selection in H.264/AVC encoders by reducing the number of candidate modes. We also present a rate estimation method for further reduction of Lagrangian cost calculation. Extensive experiments show that the proposed methods can reduce the execution time of mode selection in H.264/AVC reference software by 85% on the average while average PSNR loss is only 0.07 dB.

1. INTRODUCTION

The newest video compression standard H.264/AVC introduces various coding modes [1]. All modes at macroblock (MB) level for luma components are illustrated in Fig. 1(a). There are two intra prediction modes which are denoted as Intra_16x16 and Intra_4x4. The Intra_16x16 does spatial predictions of 16x16 luma block and the Intra_4x4 consists of sixteen 4x4 luma blocks that are separately predicted. For inter-frame prediction, each MB mode corresponds to a specific partition of the MB. For 8x8 inter prediction mode which is denoted as Inter_8x8, each of the four 8x8 blocks is split further in four ways. Fig. 1(b) shows five candidate modes for a 8x8 block in B-type frames.

In general, selecting a mode with a large partition size means that a small number of bits for motion information is required, however, motion estimation may not be accurate resulting in generating a large number of bits for sending transform coefficients. On the other hand, selecting a mode with a small partition size may require a small number of bits needed to signal residual information but produce a large number of bits for motion vectors and side information. Therefore, the choice of coding mode has a significant impact on compression efficiency.

In order to select an optimal mode in Rate-Distortion (R-D) sense, Lagrangian minimization is successfully applied to mode selection problem by Wiegand [2]. A general form of cost function used in Lagrangian R-D optimized (RDO) mode selection method is



(a) Coding modes at Macroblock level

(b) Coding modes for a 8x8 block in Inter_8x8

Fig. 1. Coding modes in H.264

$$J = D + \lambda R,$$

where J, D and R are R-D cost, distortion and rates of a mode, respectively and λ is a Lagrangian multiplier. In H.264/AVC, calculation of D is much easier than other standards because of integer transform and quantization that require only integer operations [3]. Calculation of R also can be implemented in a efficient way using a table lookups [4]. However, although Lagrangian R-D cost calculation for a single mode may have low-complexity, computation for R-D costs of all modes becomes huge because a large number of candidate modes are provided in H.264/AVC.

Based on our analysis, we concluded that total computation for mode selection can be reduced significantly if the cost calculations for less probable modes, e.g. intra modes in very low motion videos, can be skipped. Also the computation can be reduced further if R is estimated in cost calculation. In this paper, we present a low-complexity mode

0-7803-8554-3/04/$20.00 ©2004 IEEE.

Seq	PSNR loss (dB)	Exe.time (%)
Carphone	0.46	7.5
Claire	0.53	7.3
Container	0.39	6.5
Foreman	0.42	7.3
Grandma	0.52	7.5
Highway	0.30	7.2
MotherDa	0.50	6.8
News	0.51	6.7
Salesman	0.54	6.6
Silent	0.56	7.6
Average	0.47	7.1

Table 1. PSNR loss (dB) and execution time (%) of SATD-based mode selection

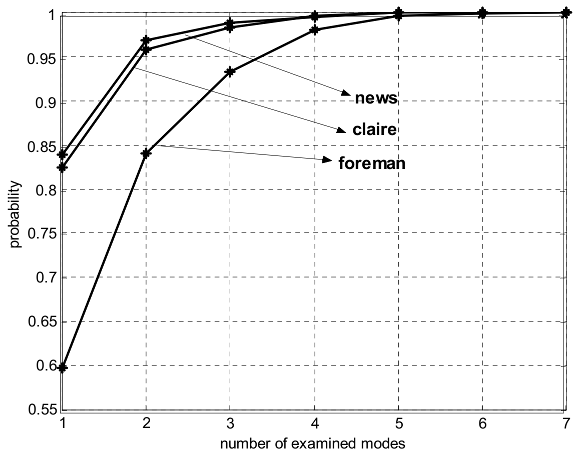


Fig. 2. Probability vs. number of most probable candidate modes

selection method by reducing number of candidate modes and estimating rates in Lagrangian cost function. The rest of this paper is organized as follows. In section 2, we will explain on skipping less probable modes. In section 3, we will describe how to estimate a number of bits for quantized transform coefficients. Simulation results will be shown in section 4. Finally, we will give a conclusion in section 5.

2. REDUCING THE NUMBER OF CANDIDATE MODES

H.264/AVC supports two intra prediction modes and various inter modes based on motion compensation block sizes. There are four spatial prediction modes in Intra_16x16. In Intra_4x4, there are nine prediction modes for each 4x4 block, which means that we need 144 (9x16) cost calculations for 4x4 block modes. For inter MB modes, five candidate modes are available including direct mode. In Inter_8x8, a MB consists of four 8x8 blocks that have separate 8x8 block modes as shown in Fig. 1(b). Therefore, 20 (5x4) cost calculations for 8x8 block modes are required to construct an Inter_8x8 MB. In general, exhaustive cost calculation for all possible modes is required to find a R-D optimal mode. However, in some cases, we can ignore some modes that are almost not used because they have a very little effect on encoding efficiency. For example, it is much less probable that intra modes are used in B-type frames of a low motion video. In this case, we can just skip calculations for costs of intra modes assuming that they have much larger cost values than that of inter modes. Exploiting this fact, we can reduce cost calculation time by skipping some less probable modes.

In H.264/AVC reference software, two mode selection methods are provided, one is RDO mode selection described in section 1, the other one is based on sum of absolute transformed difference (SATD) and estimated rates. The cost function used in the second method is

$$J_{SATD} = SATD + \lambda R_{est},$$

where *SATD* denotes a sum of absolute Hadamard transform coefficients difference and R_{est} is a estimated number of bits for motion vector and header information. SATD-based mode selection method is very fast because SATD is calculated with only integer additions and the number of bits can be estimated by simple tables. According to our experiments, the average execution time of SATD-based method is only 7% of that of RDO method. However, SATD-based method loses an average of 0.47 dB in PSNR compared to RDO mode selection. PSNR loss and computation time which is expressed as a percentage of average mode selection time of RDO method are shown in Table 1.

In our experiments, a MB mode selected by SATD-based method can be a good indicator that the candidate mode is highly probable that it also can be selected by RDO method. This means that there exists a strong relationship between SATD-based cost J_{SATD} and R-D cost J. The relationship between J_{SATD} and J can be analyzed with Fig. 2. A value in Y-axis means a probability that a minimum cost J can be found among N candidate MB modes which have N lowest J_{SATD} values where N is a value in X-axis. For example, for claire sequence, a R-D optimal mode can be found with probability of 0.98 among three most probable modes that have lowest three costs J_{SATD}. Based on our observation, we concluded that a R-D optimal MB mode can be selected with high probability by examining only some of most probable candidate modes. Similarly, this approach also can be applied to mode selections in Inter_8x8 and Intra_4x4.

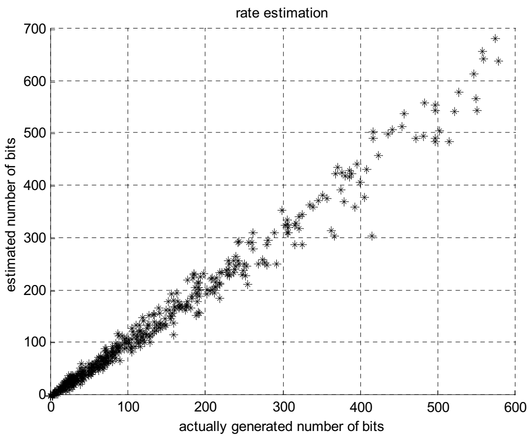


Fig. 3. Estimated rates vs, Actual rates

Seq	PSNR loss	Exe.time
Carphone	0.04	52.5
Claire	0.02	52.3
Container	0.01	50.1
Foreman	0.03	53.1
Grandma	0.02	52.7
Highway	0.01	53.3
MotherDa	0.01	51.8
News	0.05	50.3
Salesman	0.02	49.7
Silent	0.04	50.5
Average	0.03	51.6

Table 2. PSNR loss (dB) and execution time (%) with rate estimation method

3. RATE ESTIMATION IN LAGRANGIAN COST CALCULATION

In order to obtain Lagrangian R-D cost of a mode, first we need to calculate distortion by performing transform, quantization, inverse quantization and inverse transform. Then actual rates are calculated based on information on quantized transform coefficients, motion vector, header etc. Exploiting the fact that the number of zero quantized transform coefficients can be easily obtained in the process of calculating distortion, we use the information on the number of zeros to estimate the number of bits required for sending quantized transform coefficients.

A rate model called ρ-domain model which is based on the number of zero quantized transform coefficients is proposed by He [5]. The model was successfully used in rate control and bit allocation problems[5, 6]. The model is expressed as

$$R = \theta(1 - \rho),$$

Encoder options	values
Coding structure	IBBPBBPBB
Maximum search range	16
Quantization Parameters	22, 24,..., 40
Entropy coding	CABAC
SP frame	Not used
RD optimization	Used

Table 3. Encoder options used in the experiments

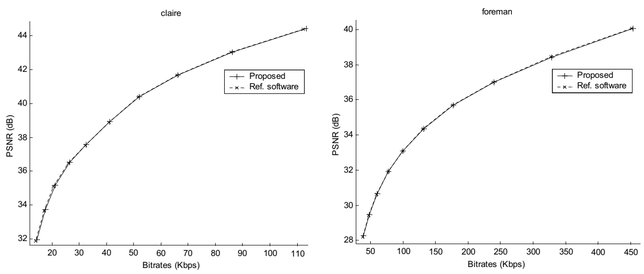
Seq.	Reduc. mod.		Reduc. mod. +Rate est	
	PSNR loss	Ex.time	PSNR loss	Ex. time
Carphone	0.05	17.9	0.08	15.3
Claire	0.08	16.8	0.09	15.2
Container	0.05	15.6	0.06	13.7
Foreman	0.04	17.5	0.07	15.1
Grandma	0.06	16.7	0.06	14.4
Highway	0.04	17.9	0.05	15.8
MotherDa	0.05	16.8	0.06	14.6
News	0.06	15.8	0.09	13.9
Salesman	0.05	15.8	0.07	14.1
Silent	0.06	16.5	0.09	14.5
Average	0.05	16.73	0.07	14.66

Table 4. PSNR loss (dB) and execution time (%) of the proposed methods

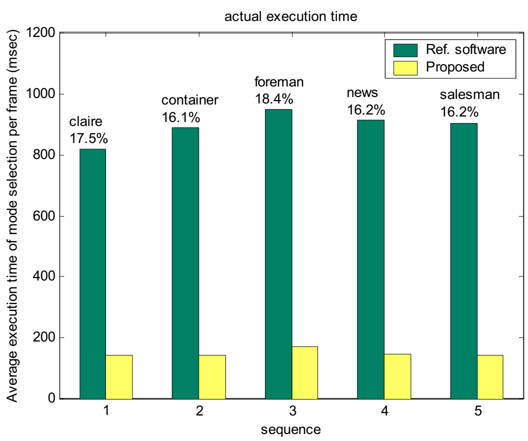
where R is a number of bits for quantized transform coefficients, θ is a model parameter and ρ is a number of zero quantized coefficients. The actual number of bits and the estimated number of bits for MBs are shown in Fig. 3. Although estimation error is large for some MBs, it did not cause significant PSNR loss in the simulations. PSNR loss and computation time shown in Table 2. In this experiments, we did not use the method described in section 2 i.e. we calculated R-D costs for all possible modes with estimated rates. In the table, execution time is expressed as a percentage of average mode selection time of RDO method. Computation time can be reduced by 48.4% on the average with only an average of 0.03 dB loss in PSNR. This method can be used independently with the method of skipping less probable modes. The results of combined methods will be shown in next section.

4. SIMULATION RESULTS

In the experiments, ten QCIF size videos are used and 300 to 400 frames are encoded for each video. The major coding options used in the experiments are shown in Table 3. Our implementation is based on reference software JM7.3 and



(a) PSNR vs. Bitrates



(b) Computation time (msec)

Fig. 4. PSNR vs. rates and Computation time

simulations were carried out on a 933MHz Intel Pentium machine with 256 MBytes memory.

PSNR loss and computation time for all sequences are shown in Table 4. We lose only 0.05 dB on the average in PSNR with an average of 17% of computation time of the reference software when only candidate mode skipping method is used. The average computation time is reduced to 15% when rate estimation method is combined while increasing PSNR loss by 0.02. For claire and foreman sequences, plots of PSNR vs. Bitrate are shown in Fig. 4(a). The PSNR curves are almost same at all bitrates for both sequences. Actual computation time for five test sequences are illustrated in Fig. 4(b).

In this experiments, we used a set of fixed numbers of most probable modes to be examined for MB candidate modes, 8x8 block candidate modes in Inter_8x8 and 4x4 block candidate modes in Intra_4x4. Let N_1 be the number of most probable modes for candidate MB modes, N_2 and N_3 be the number of most probable modes in Intra_4x4 and Inter_8x8, respectively. We used $N_1 = 4$, $N_2 = 3$ (1 for B-type frames, 7 for I-type frames) and $N_3 = 3$ (2 for B-type frames) for all tests. However, with our encoder system, we can easily control total computation of mode selection by changing N_1, N_2 and N_3. For example, if $N_1 = 7$, $N_2 = 9$ and $N_3 = 5$, then R-D costs of all possible modes will be calculated. On the other hand, if $N_1 = 1$, $N_2 = 1$ and $N_3 = 1$, then there will be no calculation for R-D costs and the results will be same as that of SATD-based method.

5. CONCLUSION

In this paper, a fast algorithm for coding mode selection in H.264/AVC encoders by both reducing the number of candidate modes and estimating rates is proposed. Our simulation results show that the execution time of mode selection can be reduced substantially with little loss in PSNR. As a future work, we will examine if the numbers of most probable modes, N_1, N_2 and N_3, which are currently fixed, can be adaptive depending on input sequences. We expect that computation time can be reduced further for some sequences without any PSNR loss if N_1, N_2 and N_3 can vary intelligently based on the characteristics of motion or contents of videos.

6. REFERENCES

[1] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Tech.*, July 2003.

[2] T. Wiegand, M. Lightstone, 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 Tech.*, vol. 6, no. 2, pp. 182–190, Apr. 1996.

[3] H. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, "Low-complexity transform and quantization in H.264/AVC," *IEEE Trans. Circuits Syst. Video Tech.*, July 2003.

[4] D. Marpe, H. Schwarz, and T. Wiegand, "Context-based adaptive binary arithmetic coding in the H.264/AVC video coding compression standard," *IEEE Trans. Circuits Syst. Video Tech.*, July 2003.

[5] Z. He and S. K. Mitra, "A linear source model and a unified rate control algorithm for dct video coding," *IEEE Trans. Circuits Syst. Video Tech.*, Nov. 2002.

[6] Z. He and S. K. Mitra, "Optimum bit allocation and accurate rate control for video coding via p-domain source modeling," *IEEE Trans. Circuits Syst. Video Tech.*, Oct. 2002.