AN EFFICIENT CRITERION FOR MODE DECISION IN H.264/AVC

Yu-Kuang Tu[*], *Jar-Ferr Yang*[*], *and Ming-Ting Sun*[+]

[*]Institute of Computer and Communication Engineering, Department of Electrical Engineering,
National Cheng Kung University, Taiwan
[+]Department of Electrical Engineering, University of Washington, Seattle, USA

ABSTRACT

In this paper, an efficient cost function for mode decision in H.264/AVC is proposed. The proposed cost function is based on integer transform coefficients, where the rate and the distortion are jointly modeled by the number of nonzero quantized coefficients, the sum of absolute integer transformed differences (*SAITD*) and sum of squared integer transformed differences (*SSITD*). Comparing to the high-complexity cost function, which should be calculated from real bit-consumption and true reconstructed distortion for each coding mode, the proposed efficient cost function can achieve 79.93% and 22.61% time savings of computing rate-distortion cost and overall encoding, respectively, while introducing only slight degradation with 1.05% bit-rate increment and 0.049dB PSNR drop.

1. INTRODUCTION

To optimize the performance of video coders, the rate-distortion (*R-D*) trade-off should be taken into account very carefully. The Lagrangian technique gives a general way in rate-distortion optimized mode decision by minimizing the rate-distortion cost for video encoding [1]. In H.264/AVC [2], inter-modes include the partition of a macroblock into 16×16, 16×8, 8×16, and $P8\times 8$ modes. In the $P8\times 8$ mode, each 8×8 block could be further divided into 8×8, 8×4, 4×8, or 4×4 subblock mode. For intra prediction, 4 Intra_16×16 and 9 Intra_4×4 modes are used for predicting the content of a macroblock and a 4×4 block respectively, from the adjacent reconstructed pixels coded previously.

There are two optimization modes suggested in the H.264/AVC reference software [3] for mode decision: high-complexity mode and low-complexity mode. The optimization with high-complexity cost (HC) requires calculations of real distortions and bit-consumptions for all candidate modes through forward/inverse transforms, forward/inverse quantization, reconstruction, and entropy coding. The low-complexity cost (LC) simply consists of the prediction error and the penalty of the coding mode. Mode decision by minimizing a certain cost function among coding options is very computationally demanding and is one of the bottlenecks of applications of H.264/AVC. Several fast mode decision algorithms [4]-[5] have been developed to significantly reduce this computationally intensive part in the encoder. However, with the high-complexity cost, the rate-distortion computation is still a heavy computational load, even if fast decision algorithms are exploited since perfect prediction of a coding mode is a hard task and the computation of the costs of remaining modes are required. Unfortunately, with low-complexity cost, it introduces considerable degradation in the coding efficiency though it is much less complicated.

In this paper, we first analyze the rate and the distortion with respect to the transform coefficients. Thus, the cost function can be modeled by the coefficients with some simple operations. With predicted rate and distortion, the proposed method can reduce the complexity in calculating the cost to as low as that of the low-complexity cost, while maintaining the coding efficiency as high as the high-complexity one.

2. BRIEF REVIEW OF COST FUNCTIONS

To jointly consider the trade-off between the coding bit-rate and the distortion during encoding the pictures, the low-complexity cost function for mode decision suggested in the H.264/AVC reference software is

$$J_L = SAD + bias \quad \text{or} \quad J_L = SATD + bias \ , \tag{1}$$

where *SAD* is the *S*um of *A*bsolute *D*ifferences, *SATD* is the *S*um of *A*bsolute (Hadamard) *T*ransformed *D*ifferences, and "*bias*" is designed to favor the prediction mode that needs fewer bits [6]. The *bias* is a parameter representing bit usage times the Lagrange multiplier λ.

On the other hand, the H.264/AVC reference software also supports the high-complexity cost function, which should be minimized by computing the actual bit consumption R and the reconstruction distortion D (measured in sum of squared differences, SSD)

$$J_H = D + \lambda \cdot R \, . \tag{2}$$

In the integer-pel motion search stage, the LC optimization with *SAD* is used while *bias* is the cost of motion information because the motion estimation is the major computational load of the encoder, especially when there are 7 block-sizes supported in H.264/AVC. When refining

the motion vector obtained in integer-pel accuracy, the LC with SAD or SATD could be exploited in sub-pel accuracy motion search. When performing mode decision, the HC achieves the rate-distortion optimized mode selection, but it is inevitably complicated; while the LC approach does not achieve sufficient accuracy to fairly perform the rate-distortion trade-off, but it requires much less computational complexity.

3. ANALYSIS OF THE RATE-DISTORTION COST

Compared to the HC, the LC suggests that *SAD* or *SATD* represents the cost of quantized transform coefficients. However, in our experiments, the required bits for coefficients could be modeled by the weighted sum of number of nonzero quantized coefficients N_{nz} and their l_1-norm E_{QTC} as:

$$\hat{B}_{\text{COEFF}} = \alpha \cdot N_{nz} + \beta \cdot E_{QTC}\,, \tag{3}$$

where E_{QTC} requires a real quantization process to calculate the sum of quantized levels. From experiments, E_{QTC} can be replaced by the sum of absolute values of those transform coefficients which will not be quantized to zeros divided by the quantization step-size qs. As a result, the predicted bit-consumption of coefficients can be rewritten as

$$\hat{B}_{\text{COEFF}} = \alpha \cdot N_{nz} + \beta \cdot \left(SATD_{nz}/qs\right), \tag{4}$$

where

$$SATD_{nz} = \sum_{|W(u,v)|<TH(u,v)} \left|W(u,v)\right| \cdot E_f(u,v) \tag{5}$$

is the *S*um of *A*bsolute *T*ransformed *D*ifferences of those which will not quantized to zeros, $W(u, v)$ is the 4×4 integer transform coefficient, $\hat{W}(u,v)$ is the quantized coefficient, and $E_f(u, v)$ is the post scaling factor which is absorbed in the quantization process [7]. To avoid real quantization, the integer transform coefficient can be compared to a threshold to determine whether the coefficient will be quantized to zero:

$$\text{if } \left|W(u,v)\right| < TH(u,v) \;\; |\,\hat{W}(u,v)\,| = 0\,. \tag{6}$$

The quantization process in H.264/ AVC is

$$|\,\hat{W}(u,v)\,| = \left(\left|W(u,v)\right| MF(u,v;QP\%6) + f\right) >> qbits\,, \tag{7}$$

where $MF(u,v;QP\%6)$ is the multiplication factor, $qbits = 15 + \text{floor}(QP/6)$, and $f = \text{floor}(qbits/k)$ in which k = 6 and k = 3 for Inter and Intra blocks, respectively. According to (7), when the following condition is satisfied, $W(u, v)$ will be quantized to zero,

$$\left|W(u,v)\right| \cdot MF(u,v;QP\%6) + f < 2^{qbits}\,, \tag{8}$$

Therefore, the zero-quantized threshold $TH(u, v)$ under a certain QP is

$$TH(u,v;QP) = \left\lfloor \frac{2^{15+\text{floor}(QP/k)} - \text{floor}\left(2^{15+\text{floor}(QP/6)}/k\right)}{MF(u,v;QP\%6)} \right\rfloor, \tag{9}$$

and the threshold matrix $\boldsymbol{TH}$ can be predetermined when QP is known, e.g., for an Inter block,

$$\boldsymbol{TH}\big|_{QP=28} = \begin{bmatrix} 54 & 84 & 54 & 84 \\ 84 & 131 & 84 & 131 \\ 54 & 84 & 54 & 84 \\ 84 & 131 & 84 & 131 \end{bmatrix}. \tag{10}$$

As a result, N_{nz} and $SATD_{nz}$ can be computed by using comparisons and additions after the integer transform. Although the post scaling factor requires floating point multiplications, it can be absorbed into the weighted factors which will be discussed later.

When considering the distortion D, it consists of two parts, one is the distortion of nonzero quantized transform coefficients D_{nz}, and the other is the distortion of the zero quantized transform coefficients D_z:

$$D = D_{nz} + D_z\,. \tag{11}$$

D_z can be calculated simply by summing up the energy of the transform coefficients which will be quantized to zero,

$$D_z = SSTD_z = \sum_{|W(u,v)|<TH(u,v)} \left|W(u,v)\right| \cdot E_f(u,v), \tag{12}$$

where *SSTD* is the *S*um of *S*quared *T*ransformed *D*ifferences. On the other hand, D_{nz} can be computed by integer transform coefficients and their inverse quantized values by mathematical manipulations [8], which requires intensive computations. However, from our empirical results, D_{nz} can be modeled by N_{nz} and the quantization step-size,

$$D_{nz} = \gamma \cdot N_{nz} \cdot qs^2\,. \tag{13}$$

Therefore, the rate and distortion contributed by the quantized transform coefficients mainly depend on three important parameters: N_{nz}, $SATD_{nz}$, and $SSTD_z$.

4. DESIGNATION OF AN EFFICIENT METRIC

After the analysis of the rate and the distortion in the previous section, the cost of the quantized transform coefficients can be formed as

$$\begin{aligned} J_{\text{COEFF}} &= \hat{D}_{\text{COEFF}} + \lambda \cdot \hat{R}_{\text{COEFF}} \\ &= SSTD_z + \gamma \cdot N_{nz} \cdot qs^2 \\ &\quad + \lambda \cdot \left(\alpha \cdot N_{nz} + \beta \cdot \left(SATD_{nz}/qs\right)\right). \end{aligned} \tag{14}$$

In the reference software, the Lagrange multiplier is suggested as $\lambda = 0.85 \cdot 2^{(QP-12)/3}$. qs and QP in H.264/AVC possess a nonlinear relationship as $qs = 2^{(QP-4)/6}$. Therefore, λ is proportional to qs^2. By substituting $\lambda = \varepsilon \cdot qs^2$, where ε is a constant into (14),

$$J_{\text{COEFF}} = a_1 N_{nz} qs^2 + a_2 qs \cdot SATD_{nz} + SSTD_z, \tag{15}$$

where a_1 and a_2 are weighted factors which can be predefined by empirical results and they will be affected by QP or qs. After absorbing qs into a_1 or a_2, a new cost function is established as:

$$\begin{aligned} J_E &= J_{\text{COEFF}} + J_{\text{HEADER}} \\ &= a_1(QP) \cdot N_{nz} + a_2(QP) \cdot SATD_{nz} \\ &\quad + SSTD_z + \lambda \cdot B_{\text{HEADER}}, \end{aligned} \tag{16}$$

where B_{HEADER} is the number of coding bits of the side information including the macroblock or sub-macroblock type, motion information (reference index and motion vector data), delta *QP*, coded block pattern (CBP), etc. When (15) is exploited as the cost measure for mode decision, it requires performing integer transform of the residual block after prediction and accumulating values of N_{nz}, $SATD_{nz}$ and $SSTD_z$ by checking whether the coefficient will be quantized to zero. To simply compute $SATD_{nz}$ and $SSTD_z$, the integer transform coefficients could be grouped into three groups according to the position (u, v), i.e., post-scaling factor: 1) G_1, those with u and v are even; 2) G_2, those with u and v are odd; and 3) G_3, otherwise. As a result, the cost measure can be rewritten explicitly as

$$J_E = a_1(QP)\cdot N_{nz} + \sum_{i=1}^{3} a_2^{G_i}(QP)\cdot SAITD_{nz}^{G_i} + \sum_{i=1}^{3} E_f^{G_i} SSITD_z^{G_i} + \lambda \cdot B_{\text{HEADER}}, \quad (17)$$

where *SAITD* and *SSITD* are the *S*um of *A*bsolute *I*nteger *T*ransformed *D*ifferences and the *S*um of *S*quared *I*nteger *T*ransformed *D*ifferences, respectively, and $a_2^{G_i}$ is the weighted factor which belongs to G_i.

5. EXPERIMENTAL RESULTS

To evaluate the new cost function, we use it for the mode decision in H.264/AVC encoding. It should be noted that the additional complexity cannot increase too much to justify the better coding efficiency comparing to the LC. On the other hand, when comparing to the HC, we hope the new cost measure would not sacrifice too much coding performance, while the computational complexity can be significantly reduced. We use (17) mainly for the inter mode decision, and Fig. 1 shows that the predicted cost is very close to the actual cost.

In the simulations, the reference software is JM 8.2 and the test conditions are: 1) baseline profile, 2) Hadamard transform is used, 3) search range is 16, 4) number of reference frame is 1, and 5) fast motion estimation is enabled. In addition, a fast inter mode decision algorithm [4] is also integrated into the software. There are four measures to evaluate the performance: time saving of rate-distortion cost computation ΔT_{RDCost} [9],

$$\Delta T_{\text{RDCost}} = \frac{T_{HC} - T_{\text{Proposed}}}{T_{HC} - T_{LC}} \times 100\%, \quad (18)$$

time saving of total encoding ΔT_{Total},

$$\Delta T_{\text{Total}} = \frac{T_{HC} - T_{\text{Proposed}}}{T_{HC}} \times 100\%, \quad (19)$$

average PSNR difference ΔPSNR in dB over the whole range of bit-rates, and average bit-rate difference ΔBit-rate in % over the whole range of PSNR [10].

From Tables 1 to 6, it can be seen that the proposed cost function can reduce 79.93% computation time of calculating the rate-distortion cost, and 22.61% saving of total encoding time compared to the original H.264/AVC with the HC optimization. The performance degradation is insignificant since the bit-rate increment and the PSNR degradation are only 1.05% and 0.049dB, respectively.

6. CONCLUSION

In this paper, we proposed an efficient cost function for the mode decision in H.264/AVC. The rate and distortion for transform coefficients are first analyzed and the total cost is then modeled by the weighted sum of the number of nonzero quantized coefficients, and the absolute sum and the squared sum of those integer transform coefficients which will be quantized to nonzero levels and zeros, respectively. Based on the new metric, only simple operations are required after the integer transform, and it can achieve comparable performance with only 1.05% bit-rate increment and 0.049dB PSNR degradation compared to the high-complexity cost function. Our proposed criterion can save about 79.93% time of computing rate-distortion cost, and about 22.61% overall encoding time.

7. REFERENCES

[1] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 688–703, July 2003.

[2] T. Wiegand, G. J. Sullivan, and A. Luthra, "Draft ITU-T Recommendation H.264 and Final Draft International Standard 14496-10 Advanced Video Coding," Joint Video Team of ISO/IEC JTC1/SC29/WG11 and ITU-T SG16/Q.6, Doc. JVT-G050rl, Geneva, Switzerland, May 2003.

[3] Joint Video Team (JVT) Reference Software [Online]. Available: http://bs.hhi.de/~suehring/tml/download/

[4] K. P. Lim, S. Wu, D. J. Wu, S. Rahardja, X. Lin, F. Pan, and Z. G. Li, "Fast inter mode selection," in Joint Video Team (JVT) Doc. JVT-I020, Sept. 2003.

[5] B. Jeon and J. Lee, "Fast mode decision for H.264," in Joint Video Team (JVT) Doc. JVT-J033, Dec. 2003.

[6] G. Sullivan, T. Wiegand, and K.-P. Lim, "Joint model reference encoding methods and decoding concealment methods," in Joint Video Team (JVT) Doc. JVT-I049, Sept. 2003.

[7] A. Hallapuro and M. Karczewicz, "Low complexity transform and quantization," in Joint Video Team (JVT) Docs. JVT-B038 and JVT-B039, Jan. 2002.

[8] Y.-K. Tu, J.-F. Yang, and M.-T. Sun, "Rate-distortion estimation for H.264/AVC coders," in *Proc. IEEE Int. Conf. on Multimedia and Expo* (ICME '05), Amsterdam, The Netherlands, July 2005, pp. 554-557.

[9] Q. Chen and Y. He, "A fast bits estimation method for rate-distortion optimization in H.264/AVC," in *Proc. Picture Coding Symposium* (*PCS 2004*), San Francisco,

CA, Dec. 2004, pp. 133-137.

[10] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," ITU-T Q.6/SG16 VCEG 13th Meeting, Doc. VCEG-M33, Austin, TX, Apr. 2001.

Table 1. Time saving of computing *R-D* cost ΔT_{RDCost} for sequences in CIF @ 30 fps

Sequence	Quantization Parameter, *QP*				Avg.
	28	32	36	40	
Basket	81.99%	79.19%	76.86%	75.27%	79.05%
Coastguard	81.88%	79.53%	77.37%	76.60%	
Foreman	80.22%	78.36%	77.63%	75.95%	
Mobile	83.82%	80.15%	77.49%	76.94%	
Stefan	83.33%	79.73%	77.31%	75.06%	
Table Tennis	82.50%	80.65%	79.90%	79.41%	

Table 2. Time saving of total encoding ΔT_{Total} for sequences in CIF @ 30 fps

Sequence	Quantization Parameter, *QP*				Avg.
	28	32	36	40	
Basket	28.01%	24.01%	21.27%	19.62%	22.31%
Coastguard	26.23%	22.55%	19.76%	18.98%	
Foreman	22.28%	19.89%	18.98%	18.42%	
Mobile	30.92%	25.45%	21.98%	20.48%	
Stefan	28.10%	23.52%	20.87%	19.43%	
Table Tennis	23.65%	21.33%	19.96%	19.72%	

Table 3. Time saving of computing *R-D* cost ΔT_{RDCost} for sequences in QCIF @ 15 fps

Sequence	Quantization Parameter, *QP*				Avg.
	28	32	36	40	
Carphone	81.81%	80.09%	76.65%	77.34%	80.82%
Coastguard	84.97%	83.68%	71.18%	79.49%	
Foreman	78.83%	76.39%	78.76%	78.25%	
Mobile	85.92%	82.15%	79.30%	78.92%	
Stefan	86.05%	84.75%	82.52%	79.47%	
Table Tennis	86.55%	83.14%	80.12%	83.35%	

Table 4. Time saving of total encoding ΔT_{Total} for sequences in QCIF @ 15 fps

Sequence	Quantization Parameter, *QP*				Avg.
	28	32	36	40	
Carphone	24.44%	21.86%	20.03%	19.94%	22.91%
Coastguard	27.30%	23.15%	19.85%	19.82%	
Foreman	23.05%	20.13%	19.70%	19.63%	
Mobile	31.83%	26.44%	22.77%	21.23%	
Stefan	31.11%	26.67%	22.94%	20.58%	
Table Tennis	24.70%	22.22%	20.14%	20.31%	

Table 5. Results of proposed cost comparing with that of actual *R-D* cost in terms of ΔPSNR (dB) and ΔBit-rate (%) for sequences in CIF @ 30fps

Sequence	ΔPSNR (dB)	ΔBit-rate (%)
Basket	-0.057	1.000
Coastguard	-0.040	1.104
Foreman	-0.032	0.852
Mobile	-0.071	1.470
Stefan	-0.048	0.854
Table Tennis	-0.057	1.338
Average	-0.051	1.103

Table 6. Results of proposed cost comparing with that of actual *R-D* cost in terms of ΔPSNR (dB) and ΔBit-rate (%) for sequences in QCIF @ 15fps

Sequence	ΔPSNR (dB)	ΔBit-rate (%)
Carphone	-0.035	0.711
Coastguard	-0.044	1.255
Foreman	-0.013	0.200
Mobile	-0.063	1.266
Stefan	-0.038	0.659
Table Tennis	-0.097	1.833
Average	-0.048	0.987

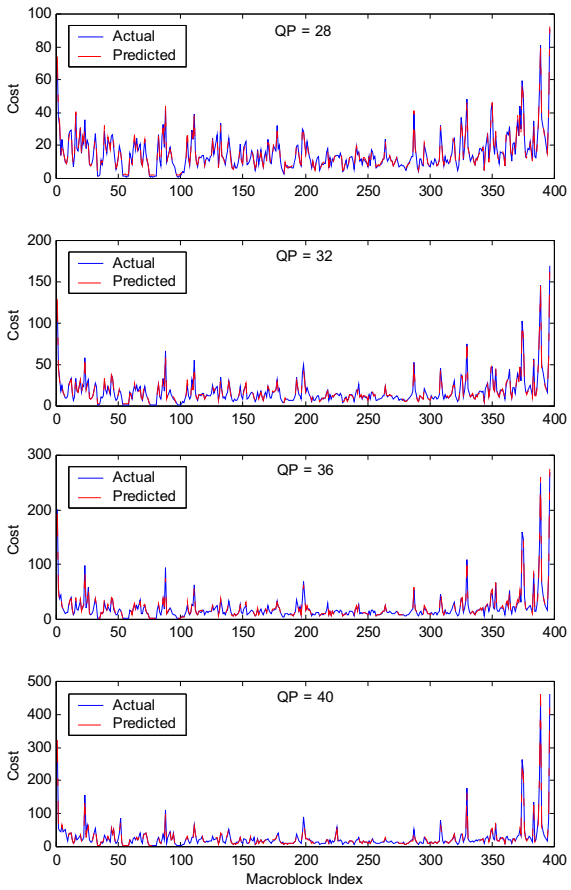

Fig. 1. MB-by-MB actual cost and predicted cost plots.