Efficient Rate-Distortion Optimized Mode Selection of H.264/AVC Intra Coding

Mohammed Golam Sarwer, and Q.M. Jonathan Wu
Department of Electrical and Computer Engineering
University of Windsor
Windsor, ON, Canada
sarwer@uwindsor.ca, jwu@uwindsor.ca

***Abstract*— Rate-distortion optimization (*RDO*) can significantly improves encoder performance in H.264-like video coding applications. In H.264/AVC, rate-distortion optimization (*RDO*) method has to code the video by exhaustively trying all the mode combinations including the different intra- and inter-prediction modes. Therefore, the complexity and computation load of video coding in H.264/AVC increase drastically compared to any previous standards. This paper reviewed the conventional fast cost functions of the intra encoder and proposed an enhanced low complex cost function for H.264/AVC intra 4x4 mode selections. The enhanced cost function uses sum of absolute Hadamard-transformed differences (*SATD*) and mean absolute deviation of the residual block to estimate distortion part of the cost function. A threshold based large coefficients count is also used for estimating the bit-rate part. The proposed method improves the rate-distortion performance of the conventional fast cost functions while maintaining low complexity requirement. As the results, the encoding process can be significantly accelerated with use of the proposed cost function. Simulation results confirmed that the proposed method reduced about 85% of computation of original encoder with negligible rate-distortion performance degradation.**

I. Introduction

H.264/AVC [1] video coding standard is the latest international standard developed by ITU-T Video Coding Experts Group and the ISO/IEC Moving Picture Experts Group, which provides gains in compression efficiency up to 50% over a wide range of bit rates and video resolutions compared to previous standards [2]. The rate-distortion optimization (*RDO*) is one of the essential parts of the H.264/AVC encoder to achieve better coding performance. However, the computational complexity of the *RDO* technique is extremely high.

In order to reduce the computation of rate-distortion (*RD*) optimization process, several rate distortion models are proposed in literature [3-11]. To reduce the computation of intra encoding, an improved cost function for intra 4x4 mode decisions is proposed in [5]. In this cost function, sum of absolute integer transform differences (*SAITD*) is used in distortion part and a rate prediction algorithm is used in rate part. The major drawback of this cost function is that it requires performing the true integer transform. Even though fast transformation algorithm was proposed to perform *SAITD*, the overall complexity is still quite high. To reduce the complexity of rate-distortion cost computation, a fast bit rate estimation technique is presented to avoid the entropy coding method during mode decision [6, 9]. But this cost function still needs to calculate highly complex sum of squared differences (*SSD*). A bit rate estimation method based on standard deviation of transform coefficient is proposed in [8].

In this paper, an enhanced sum of absolute Hadamard transform differences (*SATD*) based cost function is proposed. It is well known that the block with high detail produces more distortion. Based on this assumption, in addition to *SATD*, the distortion part of the proposed cost function uses the mean absolute deviation of the residual block. If mean absolute deviation is higher, the distortion is also higher. Bit rate is predicted based on the number of large Hadamard transformed coefficients. A block with large number of quantized-Hadamard transformed coefficients produces a large bit rate. A threshold based large coefficient count is used for estimating the bit-rate part. The complexity of the proposed cost function is further reduced by considering only low frequency Hadamard Transform coefficients for distortion and bit rate estimation.

The remainder of this paper is organized as follows. Section II provides the review of rate distortion optimized cost functions of H.264/AVC intra 4x4 mode decision techniques. Section III introduces the proposed enhanced *SATD* based cost function. The simulation results of the proposed method are presented in section IV. Finally, section V concludes the paper.

II. Cost Functions of H.264/AVC

H.264/AVC intra prediction uses 9 prediction modes for a 4x4 luma block. The best mode is the one having minimum *RD* cost and this cost is expressed as

$$J_{RD} = SSD + \lambda \cdot R \tag{1}$$

978-1-61284-857-0/11/$26.00

where the *SSD* is the sum of squared difference between the original blocks ***S*** and the reconstructed block ***C***, and it is expressed by

$$SSD = \sum_{i=1}^{4}\sum_{j=1}^{4}(S_{ij} - C_{ij})^2 \quad (2)$$

where S_{ij} and C_{ij} are the *(i, j)*th elements of the current original block ***S*** and the reconstructed block ***C***. In (1), the *R* is the true bits needed to encode the block and λ is an exponential function of the quantization parameter (*QP*). But this *RDO* process bears extremely high computational load. Hence, the cost function will make H.264/AVC impossible to be realized in real-time applications.

To accelerate the coding process, H.264/AVC provides a fast *SAD*-based cost function:

$$J_{SAD} = SAD + \lambda_1 \cdot 4P \quad (3)$$

where *SAD* is sum of absolute difference between the original block ***S*** and the predicted block ***P***. The λ_1 is almost the square of λ, and the *P* equal to 0 for the most probable mode and 1 for the other modes. The *SAD* is expressed by

$$SAD = \sum_{i=1}^{4}\sum_{j=1}^{4}\left|S_{ij} - P_{ij}\right| \quad (4)$$

where S_{ij} and P_{ij} are the *(i, j)*th elements of the current original block ***S*** and the predicted block ***P***, respectively. This *SAD*-based cost function could save a lot of computations but the expense of the computation reduction usually comes with quite significant degradation of coding efficiency. To achieve better *RD* performance, H.264/AVC also provided an alternative *SATD*-based cost function:

$$J_{SATD} = SATD + \lambda_1 \cdot 4P \quad (5)$$

where *SATD* is sum of absolute Hadamard-transformed difference between the original block ***S*** and the predicted block ***P***, which is given by

$$SATD = \sum_{i=1}^{4}\sum_{j=1}^{4}\left|h_{ij}\right| \quad (6)$$

where h_{ij} are the (*i, j*)th element of the Hadamard transformed image block ***H***.

III. Proposed Cost Function

It is reasonable to say that a complex block produces a large distortion value compared to a simple block. In other words, residual block with high details has larger distortion value than homogeneous block. Let us consider two residual blocks E_A and E_B.

$$\boldsymbol{E_A} = \begin{bmatrix} 0 & 10 & 8 & 10 \\ 9 & 7 & 4 & 10 \\ 1 & 10 & 11 & 4 \\ 19 & 6 & 15 & 7 \end{bmatrix} \text{ and } \boldsymbol{E_B} = \begin{bmatrix} 22 & 22 & 22 & 22 \\ 22 & 22 & 22 & 22 \\ 20 & 20 & 20 & 20 \\ 22 & 22 & 22 & 22 \end{bmatrix}$$

Assume *QP=24*. If we compute *SSD*, it is found that $SSD_A = 173$ and $SSD_B = 48$. That means distortion of block ***A*** is higher than that of block ***B***. This is understandable because block ***A*** contains larger detail than block ***B***. But if we calculate SATD of ***A*** and ***B*** by (6), we found that both of the residual blocks produce same *SATD* value $SATD_A = SATD_B = 368$. This means only *SATD* is not enough to measure the distortion.

Based on these observations, the distortion part of the proposed cost function is estimated as follows

$$ESATD = SATD + \alpha \times \sigma \quad (7)$$

where *ESATD* is enhanced *SATD* and α is a constant value and σ is the variance of the residual block ***E*** which represent the variation of pixel values. σ is estimated by following equation

$$\sigma \approx \frac{1}{16}\sum_{i=1}^{4}\sum_{j=1}^{4}\left|E_{ij} - \mu\right| \quad (8)$$

with $\mu = \frac{1}{16}\sum_{i=1}^{4}\sum_{j=1}^{4} E_{ij} = \frac{h_{11}}{16} = (h_{11} >> 4)$

and $E_{ij} = S_{ij} - P_{ij}$

Where h_{11} is the (*1,1*)th co-efficient of Hadamard transformed matrix ***H*** which is already computed for *SATD* calculation.

The proposed cost function also uses a rate-predictor for estimating the rate for encoding the residual block instead of just using a penalty cost (*4P*) for the unfeasible modes. The rate-predictor is based on the property of the context-based adaptive variable length coding (*CAVLC*) entropy coder. To avoid *DCT*, quantization and *CAVLC* encoding processes during *RDO*, the proposed rate-predictor only uses the total number of the non-zero quantized Hadamard transform coefficients (T_{bc}) which can be obtained by performing simple threshold and counter operations. Based on the property of the *CAVLC*'s *VLC* tables, the larger T_{bc} will produce more encoded bits. With the penalty cost (*4P*) from J_{SAD}, the overall estimated rate R_e is defined as

$$R_e = 3T_{bc} + 4P \quad (9)$$

The total number of the non-zero quantized Hadamard Transform coefficients (T_{bc}) is calculated as follows:

Step1: $T_{bc} = 0$;
Step 2: For *i=1* to *4*
    For *j=1* to *4*
      If ($|h_{ij}| \geq Q_{step}$) then $T_{bc} = T_{bc} + 1$
    End for
End for

Q_{step} is the quantization step size used in H.264/AVC encoder. Values of Q_{step} with seven different *QP*s are given in Table I. Q_{step} doubles in size for every increment of 6 in QP [12]. Therefore the proposed cost function becomes

$$J_{ESATD} = ESATD + \lambda_1 R_e \qquad (10)$$

TABLE I. QUANTIZATION STEP SIZES IN H.264/AVC

QP	Q_{step}
0	0.625
1	0.6875
2	0.8125
3	0.875
4	1
5	1.125
6	1.25

By putting value of $ESATD$ and R_e, the cost function becomes

$$J_{ESATD} = SATD + \alpha \times \sigma(\boldsymbol{E}) + \lambda_1 (3T_{bc} + 4P) \qquad (11)$$

In order to find the value of α, we have done some simulations with different types of sequences with different QPs and better results were found at $\alpha = 1.25$. The cause of the SSD is due to the quantization error of the DCT transformed coefficients of the residual block $\boldsymbol{E}$, which can be estimated by sum of absolute coefficients of $\boldsymbol{H}$ in $SATD$. It is well understandable that high frequency coefficients of Hadamard transformed coefficients are insignificant and bears low values.

Fig. 2 shows the zig-zag scan and corresponding value of frequency of Hadamard Transformed block $\boldsymbol{H}$. $\boldsymbol{H}$ can be redefined in terms of frequency as follows.

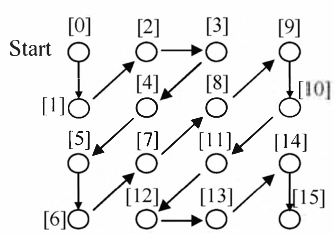

Fig. 2 Zigzag scan and corresponding frequency of $\boldsymbol{H}$

$$\boldsymbol{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \\ h_{41} & h_{42} & h_{43} & h_{44} \end{bmatrix} = \begin{bmatrix} I_0 & I_2 & I_3 & I_9 \\ I_1 & I_4 & I_8 & I_{10} \\ I_5 & I_7 & I_{11} & I_{14} \\ I_6 & I_{12} & I_{13} & I_{15} \end{bmatrix} \qquad (12)$$

where I_f is the Hadamard transformed coefficient of frequency f. After Hadamard transform, the high frequency coefficient usually has small energy. For this reason, the proposed cost function calculates only 10 low frequency Hadamard co-efficients instead of calculating all 16 co-efficient. Therefore the proposed cost function becomes,

$$J_{ESATD} = SATD' + \alpha \times \sigma(\boldsymbol{E}) + \lambda_1 (3T'_{bc} + 4P) \qquad (13)$$

$$\text{with, } SATD' = \sum_{f=0}^{9} |I_f|$$

Where T'_{bc} is the number of low frequency ($f \le 9$) non-zero Hadamard transform coefficients.

IV. SIMULATION RESULTS

The performance of proposed cost function was tested using the first 100 frames from nine different video sequences. The experiment was carried out in the JVT JM 12.4 [13] encoder and the test parameters are (a) $CAVLC$ is enabled; (b) Frame rate is 30; (c) All frames are I frame; (d) Intra 4x4 prediction only; (e) QPs are 30/36/42/48; and (f) Number of frames: 100.

In these experiments, four cost functions (J_{RD}, J_{SAD}, J_{SATD}, and J_{ESATD}) are simulated. The $PSNR$ and bit rate comparisons of the proposed cost function are tabulated in Table II. $PSNR$ and bit rate differences ($\Delta R\%$) are calculated according to the numerical averages between RD curves [14]. The positive values mean increments whereas negative values mean decrements. In case of J_{SATD} the average $PSNR$ reduction is about 0.31 dB and average bit rate increment is about 7.10%. Whereas in our proposed method, the average $PSNR$ reduction is about 0.13 dB and average bit rate increment is about 3.62%.

TABLE II. PSNR AND BIT RATE COMPARISON

Sequence	$\Delta PSNR$ in dB			$\Delta R\%$		
	J_{SAD}	J_{SATD}	J_{ESATD}	J_{SAD}	J_{SATD}	J_{ESATD}
Akiyo QCIF	-0.39	-0.37	-0.12	7.49	6.86	2.71
Foreman QCIF	-0.37	-0.21	-0.07	8.46	4.76	2.34
Container QCIF	-0.36	-0.30	-0.18	8.83	7.26	4.29
Claire QCIF	-0.33	-0.31	-0.01	5.96	5.64	1.52
Stefan QCIF	-0.60	-0.47	-0.27	16.84	12.99	7.43
News QCIF	-0.49	-0.42	-0.19	10.79	9.11	4.82
Mother_daughter QCIF	-0.29	-0.22	-0.07	6.90	4.86	2.13
Silent CIF	-0.28	-0.21	-0.12	8.90	6.20	4.82
Hall CIF	-0.36	-0.35	-0.14	6.41	6.29	2.59
Average	-0.38	-0.31	-0.13	8.95	7.10	3.62

It is clear that our proposed cost function always perform better than J_{SATD} and J_{SAD}. The worst case is encoding of *Stefan* video sequence. For *Stefan*, the proposed method increase the bit rate of about 7.43% whereas J_{SATD} generate around 12.99% of bit rate increment.

The complexity reduction, ΔT_1 (%) is defined as follows

$$\Delta T_1 = \frac{T_{original} - T_{proposed}}{T_{original}} \times 100\% \qquad (14)$$

where, $T_{original}$ and $T_{proposed}$ denote the total encoding time of the JM 12.4 encoder with J_{RD} and with proposed cost function, respectively.

TABLE III. COMPLEXITY COMPARISON

Sequence	ΔT_1 %		
	J_{SAD}	J_{SATD}	J_{ESATD}
Akiyo (QCIF)	86.76	85.92	85.76
Foreman (QCIF)	87.33	86.40	86.17
Container QCIF	85.49	84.65	83.99
Claire QCIF	84.03	83.15	82.49
Stefan QCIF	88.88	88.17	87.36
News QCIF	85.66	84.99	84.30
Mother_daughter QCIF	84.17	83.38	82.57
Silent CIF	87.43	86.99	86.67
Hall CIF	90.80	86.39	85.78
Average	86.73	85.56	85.01

TABLE IV. COMPARISON WITH J_{SATD}[5]

Sequence	$\Delta PSNR$	ΔR%	ΔT_2 %
Akiyo (QCIF)	0.10	-1.93	16.24
Foreman (QCIF)	0.04	-0.89	18.69
Container QCIF	0.05	-1.48	14.23
Claire QCIF	0.11	-1.85	12.77
Stefan QCIF	0.08	-1.99	11.60
News QCIF	0.12	-2.66	13.62
Mother_daughter QCIF	0.09	-1.87	12.98
Silent CIF	0.04	-1.22	21.35
Hall CIF	0.06	-1.15	22.59
Average	0.10	-1.67	16.00

The complexity reductions of J_{SAD}, J_{SATD}, and J_{ESATD} are tabulated in Table III. The proposed algorithm reduced about 85.01% of total encoding time compared to J_{RD}. The computational reduction of proposed method is almost similar with J_{SATD}. However, the rate-distortion performance of proposed method is much better as compared with J_{SATD}.

Table IV shows the comparison of the proposed method with J_{SAITD}. In Table IV, *PSNR* and bit rate performances are calculated based on [14] and complexity reduction is calculated as follows:

$$\Delta T_2\,\% = \frac{T_{ref[5]} - T_p}{T_{ref[5]}} \times 100\% \tag{15}$$

Where $T_{ref[6]}$ and T_p are the total encoding time of the method presented in [5] and proposed method, respectively. From the comparison result, it is shown that our proposed method reduced the bit rate of about 1.67% and increases the *PSNR* of about 0.10 dB on average. The proposed cost function not only improves the *RD* performance but also about 16% faster than that of J_{SAITD}.

V. CONCLUSIONS

In this paper, a simple and fast cost function based on *SATD* is proposed for intra mode decision of H.264/AVC. The distortion part of the cost function is estimated based on the *SATD* and mean absolute deviation of residual block. If mean absolute deviation is higher, the distortion is also higher. Bit rate is predicted based on the number of large Hadamard transformed coefficients. A block with large number of quantized-Hadamard transformed coefficients produces large bit rate. The experimental results verified that the proposed technique is very suitable for intra mode decision of H.264/AVC. With the proposed scheme, DCT, quantization, inverse-quantization, inverse DCT operations can be skipped during the mode decision process. The proposed technique reduces encoding time of JM 12.4 intra encoding by 85.01% with acceptable performance degradation. The RD performance and computational complexity of this algorithm is also better than fast *SAITD* based cost function.

REFERENCES

[1] Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC), Documents: JVT-G050, 7th Meeting, Pattaya, Thailand, 7-14 March, 2003

[2] J. Ostermann, , J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, and T. Wedi, "Video Coding with H.264/AVC: Tools, Performance, and Complexity", IEEE Circuits and Systems Magazine, vol. 4, 2004, pp.7-28.

[3] T. Chiang , and Y. Q. Zhang, "A new rate control scheme using quadratic rate distortion model", IEEE Trans. Circuits Syst. Video Tech., vol. 7, no. 1, Feb. 1997, pp. 246-250.

[4] J. R. Corbera, and S. Lei, " Rate control of DCT video coding for low-delay communication", IEEE Trans. Circuits Syst. Video Tech., vol. 9, no. 1, Feb. 1999, pp. 172-185.

[5] C. H. Tseng, H.M. Wang, and J. F. Yang, "Enhanced Intra 4x4 Mode Decision for H.264/AVC Coders", IEEE Trans. Circuits Syst. Video Tech., vol. 16, no. 8, August 2006, 1027-1032.

[6] M. G. Sarwer, and L. M. Po, "Fast bit rate estimation for mode decision of H.264/AVC", IEEE Trans. Circuits Syst. Video Tech., vol. 17, no. 10, Oct. 2007, pp. 1402-1407.

[7] Y. Kuang, J. F. Yang, and M. T. Sun, "Efficient rate-distortion estimation for H.264/AVC coders", IEEE Trans. Circuits Syst. Video Tech., vol. 16, no. 5, May 2006, pp. 600-611.

[8] Y. M. Lee, Y. T. Sun, and Y. Lin, "SATD-Based Intra Mode Decision for H.264/AVC Video Coding", IEEE Trans. Circuits Syst. Video Tech., vol. 20, no. 3, March 2010, pp. 463-469.

[9] M. G. Sarwer, and L. M. Po, " Bit Rate estimation for cost function of 4x4 intra mode decision of H.264/AVC", Proc. of IEEE International Conference on Multimedia and Expo (ICME 2007), Beijing, China, July 2-5, 2007, pp. 1579-1582.

[10] O. Wang, D. Zhao, W. Gao, and S. Ma "Low complexity RDO mode decision based on a fast coding-bits estimation model for H.264/AVC", Proc. IEEE International Symposium on Circuits and Systems, 2005, Kobe, Japan, May 23-26, 2005, vol. 4, pp. 3467 – 3470.

[11] J. Hahm, and C. M. Kyung, "Efficient CABAC Rate Estimation for H.264/AVC Mode Decision", IEEE Trans. Circuits Syst. Video Tech., vol. 20, no. 2, Feb. 2010, pp. 310-316.

[12] I. E. G. Richardson, H.264 and MPEG-4 Video Compression – video coding for next generation multimedia. John Wily & Sons, 2003, Ch. 6, pp. 191-192.

[13] Joint Video Team (JVT) reference software version 12.4, available online: http://iphome.hhi.de/suehring/tml/download/old_jm/

[14] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves", presented at the 13th VCEG-M33 Meeting, Austin, TX, 2001.