Enhanced SAITD Cost Function for H.264/AVC Intra 4×4 Mode Decision

Liping Wang, Lai-Man Po and Shenyuan Li, Ka-Ho Ng
Department of Electronic Engineering, City University of Hong Kong
Kowloon, Hong Kong SAR, China
lipiwang@student.cityu.edu.hk, eelmpo@cityu.edu.hk, shenyuli@cityu.edu.hk

Abstract—**In this paper, an enhanced sum of absolute integer transform difference (ESAITD) cost function with rate predictor "sum of non-zero coefficient frequencies" from CAVLC-based rate estimation is proposed for Intra_4x4 mode decision in H.264/AVC. This rate predictor is using simple but significant parameter to approximate the total encoded bitrate accurately. Experimental results show that ESAITD achieves much better rate-distortion (RD) performance than sum of absolute Hadamard transform difference (SATD) cost function. It also outperforms sum of absolute integer transform difference (SAITD) cost function in medium and high bitrate. Overall, ESAITD can encode the sequences with high quality.**

Keywords - H.264/AVC, intra prediction, RDO, mode decision, RD cost function

I. INTRODUCTION

Joint Video Team (JVT) of ITU-T VCEG and ISO MPEG have released the latest standard for video coding, which is known as H.264 or MPEG-4 Part 10 Advanced Video Coding (AVC) [1]-[2]. Compared with previous standards, H.264/AVC achieves up to 50% improvement in bitrate efficiency due to many new coding techniques. Lagrangian rate-distortion optimization (RDO) in mode decision is one of the important techniques to greatly increase the overall coding performance. In H.264/AVC, RDO is used to select the best mode in terms of maximum coding quality and minimum bitrate. But its drawback is it brings high computational load to the encoding process.

Many other cost functions have been proposed for fast intra mode decision to reduce the complexity of RDO. For example, fast mode decision cost functions such as sum of absolute difference (SAD) and sum of absolute Hadamard transform difference (SATD) were adopted in H.264/AVC reference software since version 6.1d [3]. However, their rate-distortion (RD) performances were degraded quite significantly. To achieve better performance, an advanced low complexity cost function that combined sum of absolute integer transform difference (SAITD) and the designed bitrate predictor was proposed in [4]. However, the bitrate estimation part was not accurate since it did not consider the total encoded bitrate. Recently, a new SATD-based cost function which used absolute transform coefficient variance to estimate the real bitrate was proposed in [5]. But calculating the standard deviation of transformed coefficients is very computational complexity.

On the other hand, a fast bitrate estimation based on the properties of context based adaptive variable length entropy coding (CAVLC) was proposed in [6]. Through skipping the entropy coding in Intra_4x4 mode decision, it reduced 60% complexity while maintaining similar performance as RDO-based mode decision. However, the rate estimation method is only used with sum of squared difference (SSD), which requires complicated implementation.

In this paper, a new cost function called enhanced sum of absolute integer transform difference (ESAITD) that incorporates SAITD with CAVLC-based rate estimator is proposed. In ESAITD, the term "sum of non-zero coefficient frequencies" is chosen to be the rate predictor from CAVLC-based rate estimator. Its performance approaches RDO while the complexity maintains similar as SATD.

The remainder of this paper is organized as follows. Intra prediction and cost functions for Intra_4x4 mode decision in H.264/AVC are reviewed in section II. A brief introduction of fast bitrate estimation based on CAVLC is presented in section III. The ESAITD cost function is proposed in section IV. Simulation results are presented in section V. Finally, conclusions are drawn in section VI.

II. INTRA PREDICTION OF H.264/AVC

Intra prediction modes in H.264 and the cost functions for Intra_4x4 mode decision will be reviewed in this section.

A. Intra Prediction Modes in H.264/AVC

Intra prediction of H.264/AVC is performed in a block-based manner. The block is predicted by referring to the neighboring samples of previously-decoded blocks which are left to and/or above it. Two primary types of luminance intra coding are supported: Intra_4x4 prediction is suitable for the parts with significant details, while Intra_16x16 is applied to the smooth areas. Intra_4x4 prediction includes 9 modes; while Intra_16x16 includes 4 modes.

B. Cost Functions for Intra Mode Decision in H.264/AVC

Cost functions are used to select the best mode by calculating RD cost of each candidate mode in H.264/AVC.

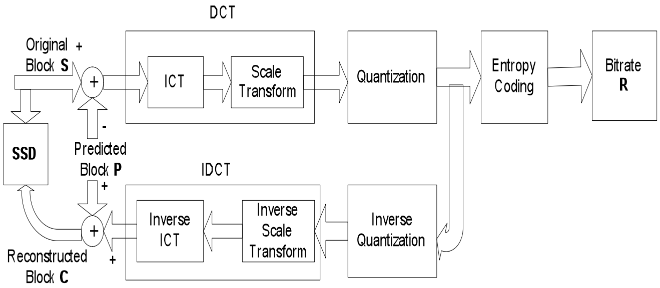

Fig. 1 Architecture of RDO computation.

Three kinds of cost functions are provided in H.264/AVC reference software: J_{RDO}, J_{SAD}, and J_{SATD}.

RDO cost function achieves the optimal mode decision performance. It goes through all the candidate modes exhaustively to select the best one which minimizes RDO cost function as formula (1):

$$J_{RDO} = SSD(S,C) + \lambda_0(QP) \cdot R \tag{1}$$

where $SSD(S,C)$ is the distortion part denotes sum of squared difference between the original block S and the reconstructed block C. R represents the encoded bits associated with the chosen mode and quantization parameter (QP). The Lagrange multiplier λ_0 is an exponential function of QP which balances the distortion and bitrate parts.

The architecture of RDO computation is shown in Fig. 1, in which we can see that RDO calculation is very complicated since it needs to undergo all the encoding and decoding processes to provide the reconstructed block C [7] for computing $SSD(S,C)$. To speed up the encoding process, the distortion part can be approximated by the difference between the original block S and the predicted block C that can avoid the transformation and quantization processes. The well-known example of this approach is the cost function J_{SAD} adopted in H.264/AVC, which is defined as

$$J_{SAD} = \begin{cases} SAD(S,P) + \lambda_1(QP) \cdot 4K & if\ Intra\_4 \times 4 \bmod e \\ SAD(S,P) & otherwise \end{cases} \tag{2}$$

where $SAD(S,P)$ is sum of absolute difference between the original block S and the predicted block P. The parameter K equals to 0 for the most probable mode (MPM) and 1 for other modes when it is used in Intra_4x4 mode decision [8]. λ_1 is also an approximate exponential function of QP. But the RD performance of this fast cost function is degraded quite significantly. To improve the performance of J_{SAD}, H.264/AVC also adopts an alternative SATD-based cost function, which is defined as

$$J_{SATD} = \begin{cases} SATD(S,P) + \lambda_1(Q_p) \cdot 4K & if\ Intra\_4 \times 4\ \bmod e \\ SATD(S,P) & otherwise \end{cases} \tag{3}$$

where $SATD(S,P)$ is sum of absolute Hadamard transformed (HT) difference between the original block S and the predicted block P. The parameters K and λ_1 are defined the same as in J_{SAD}. Experimental results show that J_{SATD} can achieve better performance than J_{SAD}, but it requires higher computation due to HT. However, the overall RD performance of J_{SATD} is still lower than the optimized J_{RDO}.

C. Sum of Absolute Integer Transform Difference (SAITD)-based Cost Function

From Fig. 1, it can be seen that the final DCT, quantization, and bitrate compression are mainly based on the result of integer cosine transform (ICT). Hence using ICT to replace HT in JSATD can improve the accuracy of mode decision [4]. The corresponding SAITD-based cost function for Intra_4x4 prediction is defined as

$$J_{SAITD} = SAITD(S,P) + \lambda_1(QP) \cdot R_e \tag{4}$$

where $SAITD(S,P)$ is sum of absolute ICT difference between the original block S and the predicted block P. The rate estimation part is approximated as

$$R_e = 4T_c - T_o + 4K \tag{5}$$

where T_c represents the number of non-zero coefficients and T_o represents the number of trailing ±1 in CAVLC.

SAD, SATD and SAITD are the prediction error in spatial, HT, and ICT domains, respectively. Experimental results show that SAITD outperforms SAD and SATD in Intra_4x4 mode decision. From [4], SAITD is the best choice as prediction error measurement in a low complexity cost function. In the following sections, a more accurate rate estimation method [6] is reviewed and then it is used to develop a new enhanced SAITD cost function.

III. FAST BITRATE ESTIMATION BASED ON CAVLC

In H.264/AVC, CAVLC is an entropy coding algorithm for encoding zig-zag scanned block of quantized transform coefficients. The block is encoded by CAVLC which is based on five syntax elements. They can be obtained as soon as the transform coefficients are quantized. In [6], a CAVLC-based rate estimation method is developed to calculate these elements for avoiding the entropy coding in RDO-based mode decision. These elements for rate estimation are described as follows:

Coeff_Token: It encodes both the number of non-zero coefficients (T_c) and the number of trailing ±1 (T_o):

$$R_{coff} = T_c - T_o \tag{6}$$

Sign of Trailing Ones: For each trailing ±1, a single bit encodes the sign (0=+, 1=-):

$$R_{trail1} = T_o. \tag{7}$$

Total_Zeros: It represents the total number of zero coefficients between DC and the last non-zero coefficient:

$$R_{zero} = T_z. \tag{8}$$

Level: The level (sign and magnitude) of all remaining non-zero coefficients is calculated as:

$$R_{level} = SAT_l = \sum_{k=1}^{T_c} |L_k| \tag{9}$$

where SAT_l is sum of absolute quantized transform coefficients in residual block. $|L_k|$ is the absolute value of kth non-zero coefficient.

Run_Before: It indicates the number of zeros preceding each non-zero coefficient in reverse order. The frequency of non-zero coefficient is defined as its position index. As we know, more zeros are found at the high-frequency position after DCT. So the value of *Run_Before* for the higher frequency non-zero coefficient is larger. The *Run_Before* of each non-zero coefficient can be estimated by:

$$R_{run_{(k)}} = 0.3\sum_{k=1}^{T_c} f_k, \quad 1 \le f_k \le 16 \tag{10}$$

where f_k is the frequency of kth non-zero coefficient in residual block.

Noted that the weightings of T_o from R_{coeff} and R_{trail1} are opposite. The final rate estimation equation only includes four different types symbols as follows:

$$R_{e(CAVLC)} = T_c + T_z + SAT_l + 0.3\sum_{k=1}^{Tc} f_k. \tag{11}$$

Through the new CAVLC based rate estimation method, it is found that the rate estimation equation (5) is not accurate. T_o is redundant and a more efficient and accurate bitrate predictor "sum of non-zero frequencies" will be proposed in next section.

IV. ENHANCED SAITD-BASED COST FUNCTION

As mentioned in section II, SAITD is regarded as the best measurement of prediction error in fast Intra_4x4 mode decision. The simulation results in [4] show that the performance of SAITD is better than SATD significantly in low bitrate. However, its performance in high bitrate is not as good as in low bitrate. It is because the bitrate estimation in SAITD is not accurate enough.

From the analysis in section III, it can be seen that the bitrate estimation of SAITD in equation (5) only considered the *coeff_token* bitrate consumption. It cannot represent the total encoded bits. Moreover, T_o should be absent in the whole bitrate estimation because the weightings of T_o from R_{coeff} and R_{trail1} are opposite [6]. It is necessary to find a more accurate bitrate predictor to replace the rate estimation part in J_{SAITD}.

From the definition of SAT_l, it is found that SAT_l is proportional to SAITD for a specific *QP*. Because SAITD is already available in the cost function, SAT_l can be omitted. Then the bitrate estimation function can be simplified as

$$R_{e(CAVLC)} \cong T_c + T_z + 0.3\sum_{k=1}^{Tc} f_k. \tag{12}$$

Intuitively, $\sum_{k=1}^{Tc} f_k$ will increase when T_c and/or T_z increases. It is shown that $\sum_{k=1}^{Tc} f_k$ is the most significant element in the whole bitrate estimation, which can represent the changes both in T_c and T_z simultaneously. Consequently, $\sum_{k=1}^{Tc} f_k$ is chosen as the rate predictor to remove the redundant terms from $R_{e(CAVLC)}$ and maintain its high precision. Based on the above analysis, the proposed SAITD based cost function for Intra_4x4 mode decision in H.264/AVC is written as

$$J_{SAITD,run} = SAITD(S,P) + \lambda_1(Q_p) \cdot R_{e,run}. \tag{13}$$

The estimated bitrate $R_{e,run}$ is given by

$$R_{e,run} = \beta\sum_{k=1}^{T_c} f_k + 4K \tag{14}$$

where β is used for adjusting $R_{e,run}$ to work with SAITD better. β=0.8 performs best for most sequences in simulations. Calculating the rate predictor $\sum_{k=1}^{Tc} f_k$ only needs simple threshold and accumulation operations on one parameter after transform. It also avoids the quantization process.

Based on comprehensive simulations upon various types of video sequences, it is found that $J_{SAITD,run}$ can achieve better performance than J_{SAITD} at low and medium *QP* (or high to medium bitrate). J_{SAITD} can achieve better performance only when *QP* is very high. Based on this characteristic, the enhanced SAITD (ESAITD) cost function is defined as:

$$J_{ESAITD} = \begin{cases} SAITD(S,P) + \lambda_1(Q_p)\cdot(\beta\sum_{k=1}^{T_c} f_k + 4K) & \text{low and medium } QP \\ SAITD(S,P) + \lambda_1(Q_p)\cdot(4T_c - T_o + 4K) & \text{high } QP \end{cases} \tag{15}$$

V. EXPERIMENTAL RESULTS

The proposed ESAITD cost function was implemented on H.264/AVC reference software JM10.2 [10]. The proposed cost function J_{ESAITD} was applied to substitute J_{SATD} in Intra_4x4 prediction in the source codes. The first 100 frames of three QCIF sequences (Mobile, Stefan, and Foreman), four CIF sequences (Stefan, Bus, Mobile, and Coastguard), and two 4CIF sequences (Soccer and Harbour) were used in our experiments. In order to investigate the characteristics of J_{ESAITD} from high bitrate to medium bitrate, the *QP* is set to *QP*=20, 23, 26, 29, 32, 35, 38, and 41. For high *QP*, J_{ESAITD} works exactly the same as J_{SAITD}. Therefore, the simulation results of high *QP* are not shown. The test conditions are as follows:

(a) RD Optimization is off and use Hadamard is on;
(b) Only Intra_4x4 is used;
(c) Intra_16x16 is disabled;
(d) Entropy coding method is CAVLC;
(e) GOP structure is all Intra frames.

Since both J_{ESAITD} and J_{SAITD} belong to low complexity cost functions, their performances are compared with SATD scheme of H.264/AVC. The average improvements of J_{SAITD} and J_{ESAITD} compared with J_{SATD} at high bitrate (*QP*=20, 23, 26, 29) and medium bitrate (*QP*=32, 35, 38, 41) are listed in Table I and Table II. The average PSNR increase (dB) is calculated when the bitrates are equal while the average bitrate decrease (percentage) is calculated when the PSNRs are the same [11].

TABLE I. SIMULATION RESULTS WITH QP= {20, 23, 26, 29}.

QP={20,23,26,29}		J_{SAITD}		J_{ESAITD}	
Resolution	**Sequence**	**ΔPSNR (dB)**	**Δbitrate (%)**	**ΔPSNR (dB)**	**Δbitrate (%)**
QCIF	**Mobile**	0.02	-0.15	0.07	-0.52
	Stefan	0.05	-0.39	0.10	-0.77
	Foreman	0.02	-0.28	0.07	-0.79
CIF	**Stefan**	0.09	-0.82	0.14	-1.20
	Bus	-0.00	0.04	0.06	-0.54
	Mobile	0.03	-0.24	0.08	-0.63
	Coastguard	0.01	-0.10	0.06	-0.66
4CIF	**Soccer**	0.08	-1.06	0.14	-1.87
	Harbour	0.01	-0.12	0.07	-0.82
Average		**0.03**	**-0.35**	**0.09**	**-0.87**
Average improvement of J_{ESAITD} compared with J_{SAITD}				**0.05**	**-0.52**

TABLE II. SIMULATION RESULTS WITH QP= {32, 35, 38, 41}.

QP={32,35,38,41}		J_{SAITD}		J_{ESAITD}	
Resolution	**Sequence**	**ΔPSNR (dB)**	**Δbitrate (%)**	**ΔPSNR (dB)**	**Δbitrate (%)**
QCIF	**Mobile**	0.02	-0.31	0.07	-0.88
	Stefan	0.09	-1.16	0.11	-1.50
	Foreman	0.13	-1.54	0.14	-1.76
CIF	**Stefan**	0.13	-1.58	0.14	-1.73
	Bus	-0.00	0.05	0.05	-0.71
	Mobile	0.05	-0.66	0.08	-1.03
	Coastguard	0.05	-0.81	0.12	-1.90
4CIF	**Soccer**	0.23	-3.53	0.24	-3.86
	Harbour	-0.02	0.32	0.09	-1.18
Average		**0.08**	**-1.02**	**0.12**	**-1.62**
Average improvement of J_{ESAITD} compared with J_{SAITD}				**0.04**	**-0.59**

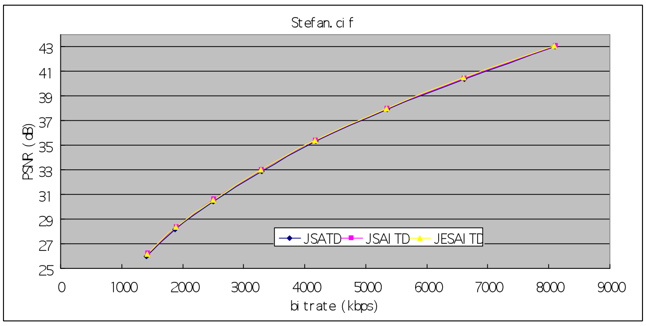

Fig. 2. The RD curves for Stefan sequence.

From Table I, it can be seen that J_{SAITD} achieves an average PSNR increase of 0.03dB as compared with J_{SATD}. Meanwhile the proposed J_{ESAITD} can achieve 0.09 dB improvement as compared with J_{SATD}, which indicates J_{ESAITD} improves the performance of J_{SATD} more significantly than J_{SAITD}. Similar results can also be obtained in the case of medium bitrate as shown in Table II. The RD performance comparisons of J_{SATD}, J_{SAITD}, and J_{ESAITD} in sequence *Stefan* are shown in Fig. 2.

Furthermore, we can see that J_{ESAITD} performs better in the sequences with complicated content, such as high motion sequences. The overall simulation results suggest J_{ESAITD} has better performance when coding complicated sequences in high and medium bitrate. Complicated content and high bitrate imply high quality coding requirement. The accuracy of the cost function affects the coding performance greatly when high coding quality is required. J_{ESAITD} performs better to meet high quality requirement indicates it has higher prediction accuracy than J_{SAITD} in intra mode decision.

VI. CONCLUSIONS

In this paper, an ESAITD cost function with simple rate predictor is proposed for fast Intra_4x4 mode decision in H.264/AVC. The rate predictor is inherited from the CAVLC-based rate estimation. The major advantage of this rate predictor is that it can accurately approximate the total coding bitrate with simple computation. The simulation results show that the performance of ESAITD is better than other conventional cost functions. ESAITD is especially suitable to encode sequences with high quality requirement.

REFERENCES

[1] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, “Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264/ISO/IEC 11496-10 AVC)”, JVT-G050, March 2003.

[2] Thomas Wiegand, Gray J. Sullivan, Gisle Bjontegaard, and Ajay Luthra, “Overview of the H.264/AVC video coding standard,” IEEE Trans. Circuits Syst. Video Technol., vol. 13, no.7, pp.560-576, July 2003.

[3] H.264 Joint Video Team (JVT) Reference Software version 6.1d, http://iphome.hhi.de/suehring/tml/download/

[4] Chao-Hsuing Tseng, Hung-Ming Wang and Jar-Ferr Yang, “Enhanced intra-4x4 mode decision for H.264/AVC coders,” IEEE Transactions on Circuits System Video Technology, vol. 16, no. 8, pp.1027-1032, August 2006.

[5] Yu Ting Sun and Yinyi Lin, “SATD-based intra mode decision for H.264/AVC video coding,” in Proc. IEEE ICIP, pp.765-768, 2004.

[6] Mohammed Golam Sarwer and Lai-Man Po, “Fast bit rate estimation for mode decision of H.264/AVC,” IEEE Transactions on Circuits and Systems for Video Technology, vol.17, no. 10, pp. 1402-1407, October 2007.

[7] Lai-Man Po and Kai Guo, “Transform-domain fast sum of the squared difference computation for H.264/AVC rate-distortion optimization,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 6, pp. 765-773, June 2007.

[8] Iain E. G. Richardson, H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia, Wiley, England, 2003.

[9] Iain E. G. Richardson, “H.264/MPEG-4 part 10 white paper: transform and quantization,” March 2003.

[10] H.264 Joint Video Team (JVT) Reference Software version 10.2, http://iphome.hhi.de/suehring/tml/.

[11] Gisle Bjontegaard, “Calculation of average PSNR differences between RD-curves,” ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, VCEG 13th Meeting, VCEG-M33, Austin, Texas, USA, April 2001.