A novel rate estimation model for mode decision of H.264/AVC

Xin Zhao[*a], Jun Sun[b], Wen Gao[b]
[a]Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China;
[b]Institute of Digital Media, EECS, Peking University, Beijing, 100871, China

ABSTRACT

To acquire the optimal coding mode of each macroblock, the H.264/AVC encoder exhaustively calculates the rate-distortion cost for all available modes and chooses the minimum one as the best mode. Therefore, the mode decision process is very computationally demanding. To reduce the computation complexity of the rate-distortion cost, in this paper, we propose a novel rate estimation model for mode decision in H.264/AVC. By modeling the transform coefficients with Generalized Gaussian distributions (GGD), a direct relationship between the magnitude and the information bits of the quantized transform coefficients is deduced. Based on this deduction, the weighted sum of quantized transform coefficients is proposed as an efficient bit-rate estimator of the residual blocks. Extensive experiments show that the proposed algorithm can save up to 30% of total encoding time with ignorable degradation in coding performance for both inter- and intra-mode decision.

Keywords: Rate estimation, mode decision, rate-distortion optimization (RDO), Generalized Gaussian distributions (GGD), linear regression, H.264/AVC.

1. INTRODUCTION

To explore the coding efficiency of block-based hybrid video coding structure, coding strategies become more and more flexible during the development of international video coding standards such as MPEG-1, MPEG-2, MPEG-4, H.263 and the latest H.264/AVC. Lagrangian multiplier optimization technique is usually used to achieve the best coding mode decision in high-compression video coding. By using the optimization technique, all available modes are evaluated by rate-distortion (R-D) cost, and the one which minimizes the R-D cost is selected as the best mode. The minimization process of the R-D cost is well known as rate-distortion optimization (RDO). Although RDO can accurately choose the best mode for video coding, the computation complexity is very high.

To reduce the computational complexity of RDO in H.264/AVC, many fast mode-decision methods were proposed. Efforts were mainly dedicated to reduce the computation complexity in two ways. One is to explore the spatial characteristics of the pixels, with which the most probable Inter or Intra mode is predicted. Therefore unnecessary coding modes can be eliminated. The other is to estimate the coding rate or distortion by a certain rate model. In this way, the coding rate of a certain mode is estimated to avoid actual entropy coding which costs much computation time. In the second category, rate models observed from quantizer (Q)-domain in [1] and ρ-domain in [2] are established and theoretically justified, but both models are only considered for rate control schemes. To reduce the complexity of RDO, block-level rate estimation models were also proposed. In [3], a linear function of the number and the levels of nonzero quantized transform coefficients is used as an efficient rate estimator, but this model is only designed for inter-mode decision. Another block-level rate estimation model using five different tokens of CAVLC is proposed in [4]. This model is suitable for both inter-and intra-mode decision of H.264/AVC, but it is constrained in the CAVLC entropy coding method.

In this paper, we propose a novel transform-domain rate estimation model. This model is derived from the direct relationship between bit-rate and the magnitude of a single quantized transform coefficient. The model parameters are deduced from the GGD parameters of the transform coefficients and updated with linear regression during the encoding process.

The paper is organized as follows: Section 2 discusses the relationship between bit rate and the magnitude of a single quantized transform coefficient. The proposed rate estimation model is introduced in Section 3. Section 4 presents the experimental results. Finally, conclusion is summarized in Section 5.

[*] xzhao@jdl.ac.cn; www.jdl.ac.cn

2. RELATIONSHIP BETWEEN BIT RATE AND QUANTIZED TRANSFORM COEFFICIENTS

The basic idea of the proposed rate estimation model originates from the observation that the same magnitude of different frequency components can result in different amount of information bits. For example, let F be the 4×4 quantized transform coefficients matrix in H.264/AVC, and a single frequency component is represented by *F(u,v)*. Then the information bits from "F(3,3)=20" is always more than the information bits from "F(0,0)=20". This is because *F(u,v)* conforms to different distributions for different u or v, and the first event always happens with less probability than the second. The matrixes shown in Fig.1 are two actual quantized transform block of Foreman with CIF format. The l_1-norms of the 2 blocks are both 20. But the left block results in a coding rate of 62 bits, while the right block of which the power distributes to low frequency components results in a coding rate of only 28 bits. This demonstrates that even when the l_1-norms of two quantized transform block are the same, the actual coding bits can be still very different. And the difference in the number of coding bits is due to the different distributions of power in the transform block. Therefore, the quantized transform coefficients should be weighted before being used to estimate the bit rate.

$$F_0=\begin{bmatrix}2&1&2&1\\2&1&0&0\\2&1&2&0\\3&1&2&0\end{bmatrix},F_1=\begin{bmatrix}16&2&1&1\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix}$$

Fig. 1. Two actually coded quantized transform block from Foreman in CIF format with l_1-norm both equal to 20.

To get the exact weighting form, we model the transform coefficients F_{uv} with a Generalized Gaussian distribution given by

$$f_{uv}(x)=\frac{\eta\alpha_{uv}(\eta)}{2\sigma_{uv}\Gamma(1/\eta)}\exp\left\{-[\alpha_{uv}(\eta)\,|\frac{x}{\sigma_{uv}}|]^{\eta}\right\}, \quad (1)$$

where $\Gamma(\cdot)$ is the gamma function, $\alpha_{uv}(\eta)=\sqrt{\Gamma(3/\eta)/\Gamma(1/\eta)}$, η and σ are positive real valued distribution parameters. The quantization process suggested in H.264/AVC is represented by

$$|\hat{F}_{uv}|=(|F_{uv}|\bullet Q+f\bullet 2^{qbits})>>qbits, \quad (2)$$

where Q is the multiplication factor, f is the rounding control parameter. With (2), the probability of F_{uv} being quantized as $\hat{x}$ is computed as

$$\mathrm{P}\left\{\hat{F}_{uv}=\hat{x}\right\}=\begin{cases}2\int_0^{(1-f)\bullet Q_{step}}f_{uv}(x)dx, & \hat{x}=0\\ \int_{(|\hat{x}|-f)\bullet Q_{step}}^{(|\hat{x}|+1-f)\bullet Q_{step}}f_{uv}(x)dx, & \hat{x}\neq 0\end{cases}, \quad (3)$$

where Q_{step} is the quantizer step size equal to $2^{qbits}/Q$. The information bits from "$\hat{F}_{uv}=\hat{x}$" is

$$r_{uv}=-\log_2\mathrm{P}\left\{\hat{F}_{uv}=\hat{x}\right\}. \quad (4)$$

According to [6], the rounding control parameter f in (2) can be adaptively adjusted during encoding to better locate the expectation value of $\hat{F}_{uv}$ inside a quantization interval. Therefore, the probability of F_{uv} being quantized as $\hat{x}$ can be approximated by

$$\mathrm{P}\left\{\hat{F}_{uv}=\hat{x}\right\}\doteq\begin{cases}2(1-f)Q_{step}\bullet f_{uv}(f\bullet Q_{step}) & \hat{x}=0\\ Q_{step}\bullet f_{uv}(\hat{x}\bullet Q_{step}) & \hat{x}\neq 0\end{cases}. \quad (5)$$

With (4) and (5), when $\hat{x}\neq 0$, the information bits from "$\hat{F}_{uv}=\hat{x}$" is approximated by

$$r_{uv}\doteq-\log_2\left\{Q_{step}\bullet\frac{\eta\alpha_{uv}(\eta)}{2\sigma_{uv}\Gamma(1/\eta)}\exp\left\{-[\alpha_{uv}(\eta)\,|\frac{Q_{step}\bullet\hat{x}}{\sigma_{uv}}|]^{\eta}\right\}\right\}=\mathrm{a}_{uv}\bullet|\hat{x}|^{\eta}+\mathrm{b}_{uv}, \quad (6)$$

where $\mathrm{a}_{uv}=\log_2(e)\cdot[Q_{step}\cdot\alpha_{uv}(\eta)/\sigma_{uv}]^{\eta}$ and $\mathrm{b}_{uv}=-\log_2[Q_{step}\cdot\eta\alpha_{uv}(\eta)/(2\sigma_{uv}(\Gamma(1/\eta))]$. The derivation of r_{uv} when $\hat{x}=0$ is similar to (6), but a_{uv} and b_{uv} have different values. Then for transform coefficients having Generalized Gaussian distributions, the amount of information bits is a power function of their magnitudes. Observe that when η=1, the GGD becomes a Laplacian distribution, and (6) becomes a simple linear function. Because $1/\sigma_{uv}$ is always larger with a larger index of u and v, and $\mathrm{a}_{uv}=\log_2(e)\cdot[Q_{step}\cdot 2^{1/2}/\sigma_{uv}]$, it can be observed that the high frequency coefficients with larger u and v can affect the bit rate more than the low frequency coefficients. This result verifies the effect that different frequency components of quantized transform coefficients can affect the bit rate differently.

3. PROPOSED RATE ESTIMATION METHOD

According to information theory, the number of information bits from a composite signal is equal to the sum of information bits of its independent components. And for the good decorrelation ability of discrete cosine transform, with (6), the amount of information bits of a quantized transform block can be approximately computed by

$$r_B \doteq \sum_u \sum_v (\mathrm{a}_{uv} \bullet |\hat{x}_{uv}|^{\eta} + \mathrm{b}_{uv}), \tag{7}$$

and we propose to estimate the actual coding bits of a single block with the information bits in (7) as

$$R_B = \alpha \bullet r_B + \beta, \tag{8}$$

where α and β are the parameters of the proposed rate estimation model, R_B is the estimated coding bits of a single block.

3.1 Implementation of the proposed rate estimation algorithm

Based on (6), (7) and (8), the proposed rate estimation algorithm is composed of the following four steps:
1. Estimate the GGD parameters of current frame by previous frame of the same type.
2. Compute the information bits of current quantized transform block coefficients by (7).
3. Compute the estimated bit rate of current quantized transform block coefficients by (8).
4. Update the model parameters α and β after the mode decision process of current block.

The quantized transform block coefficients used in the above 4 steps were recorded in the transform module. And the entropy coding is replaced by the proposed rate estimation algorithm during mode decision process.

3.2 Updating of model parameters

To make the model adaptive to variously changing frame statistics, we update the model parameters α and β with linear regression. The linear regression process is expressed by

$$R_i = \alpha \bullet r_i + \beta + \varepsilon_i, \tag{9}$$

where i=0,1,…,n-1, R_i represents the actual coding bits by the entropy coder, r_i represents the information bits computed by (7), ε_i is the prediction residual error. The least square estimates of model parameters α and β are given by

$$\alpha = \frac{n\sum_i R_i r_i - \sum_i R_i \sum_i r_i}{n(\sum_i r_i)^2 - \sum_i r_i^2}, \beta = \frac{n(\sum_i r_i)^2 \sum_i R_i - \sum_i R_i r_i \sum_i r_i}{n(\sum_i r_i)^2 - \sum_i r_i^2} \tag{10}$$

4. EXPERIMENTAL RESULTS

We integrated the algorithm into a recent version of H.264/AVC reference software JM13.2. Experiments are performed at different QP values ranging from 24 to 39, and some important coding parameters are set as: all available inter- and intra-modes in reference software are enabled; fast motion estimation algorithm "Simplified UMHexagon" is used; motion search range is 33 by 33; the number of reference frames is 1 for IPP coding type and 2 for IBP coding type; IntraPeriod is set as 8; CABAC entropy coding method is used; Fast chroma intra mode decision is turned off.

Table 1. Coding performance of the proposed algorithm (Compared to RDO in H.264/AVC for QCIF sequences).

Sequence	Type	ΔPSNR(dB)[5]	ΔTotal(%)	Sequence	Type	ΔPSNR(dB)[5]	ΔTotal(%)
Foreman	IPP..	-0.0467	12.90	Container	IPP..	-0.0693	13.85
	IBP..	-0.0491	11.53		IBP..	-0.0794	11.81
Bus	IPP..	-0.0300	18.48	Hall_monitor	IPP..	-0.0623	14.92
	IBP..	-0.0299	15.01		IBP..	-0.0766	13.06
Football	IPP..	-0.0559	16.63	Mother_daughter	IPP..	-0.0498	9.32
	IBP..	-0.0470	12.69		IBP..	-0.0447	10.10
Tempete	IPP..	-0.0347	19.56	Silent	IPP..	-0.0543	18.62
	IBP..	-0.0406	17.96		IBP..	-0.0455	10.89
Coastguard	IPP..	-0.0364	14.65	Stefan(352x240)	IPP..	-0.0434	16.98
	IBP..	-0.0377	12.99		IBP..	-0.0266	15.53

To verify the robustness of our proposed rate estimation algorithm, extensive experiments were performed on standard test sequences with QCIF and CIF format. In the experiment, we mainly compare the complexity reduction and the coding performance loss [5] of the proposed algorithm. When evaluating the complexity reduction, we use ΔT in [4] defined as $\Delta T = [(T_{RDO} - T_{Proposed})/T_{RDO}] \times 100\%$, where T_{RDO} and $T_{Proposed}$ represent the average computation time of the encoder with the original RDO and with the proposed algorithm, respectively. The experimental results listed in Table.1 show that the average PSNR loss [5] compared with original RDO is ignorable, while the proposed algorithm can achieve 5% to 30% of total encoding time reduction depending on QP value ranging from 24 to 39, and on average about

14.4% of total encoding time reduction. To verify the accuracy of the proposed algorithm, the actual and the estimated bit rate of randomly selected 100 blocks in Foreman with CIF format is shown in Fig.2. From Fig.2 it can be seen that our proposed algorithm can estimate the rate efficiently. The R-D performance of the proposed algorithm compared with RDO turned on and without RDO is also shown in Fig.3. Compared with the rate model in [3], our proposed model is efficient for both inter- and intra-mode decision in H.264/AVC. Different from the rate model in [4], the proposed model is not constrained in the properties of CAVLC entropy coding. Although only experimental results based on CABAC are shown here, experiments show that the proposed mode is also efficient for CAVLC entropy coding.

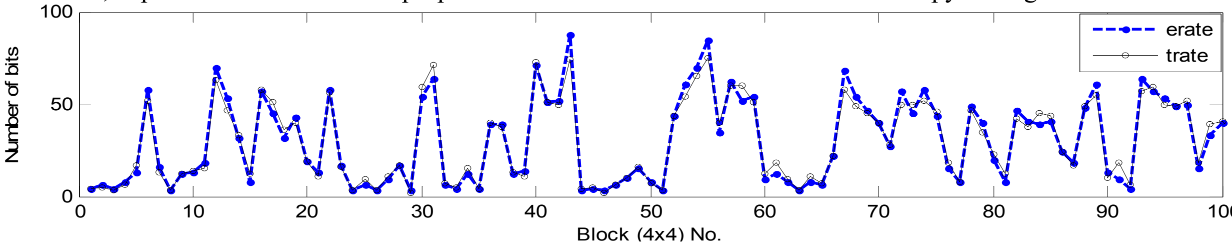

Fig. 2. Comparisons of the actual and the estimated block rate of consecutive 100 blocks in Foreman with CIF format.

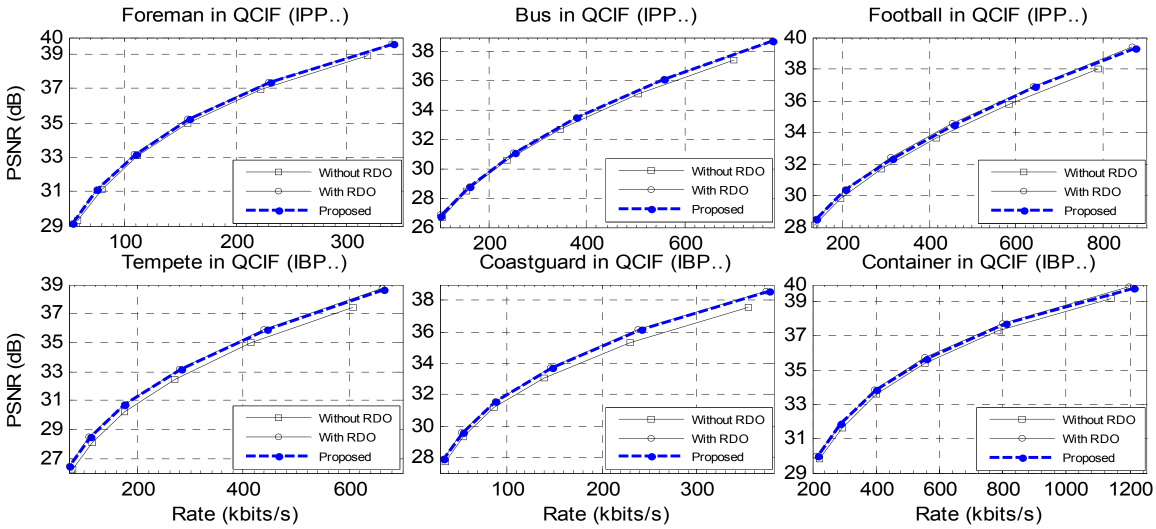

Fig. 3. Performance comparisons of different mode decision algorithms in H.264/AVC reference software JM13.2.

5. CONCLUSION

In this paper, we propose an adaptive transform domain bit-rate estimation model for mode decision in H.264/AVC. Extensive experimental results show that, for both intra- and inter-mode decision, the proposed algorithm can estimate the rate efficiently without actually performing CABAC entropy coding. Ignorable degradation of coding performance compared with original RDO is achieved and up to 30% of total encoding time can be saved.

REFERENCES

1. Tihao Chiang and Ya-Qin Zhang, “A new rate control scheme using quadratic rate distortion model,” *Circuits and Systems for Video Technology, IEEE Transaction on*, vol. 7, pp. 246-250, Feb 1997.
2. Zhihai He and S.K Mitra, “A linear source model and a unified control algorithm for DCT video coding,” *Circuits and Systems for Video Technology, IEEE Transaction on*, vol. 12, pp. 970-982, Nov 2002.
3. Yu-Kuang Tu, Jar-Ferr Yang, and Ming-Ting Sun, “Efficient rate-distortion estimation for H.264/AVC coders,” *Circuits and Systems for Video Technology, IEEE Transaction on*, vol. 16, pp. 600-611, May 2006.
4. M.G. Sarwer and Lai-Man Po, “Fast bit rate estimation for mode decision of H.264/AVC,” *Circuits and Systems for Video Technology, IEEE Transaction on*, vol. 17, pp. 1402-1407, Oct 2007.
5. Gisle Bjontegaard, “Calculation of average PSNR differences between RD-curves,” in *Proc ITU-T Q.6/SG16 VCEG 13th Meeting*, Austin, TX, Apr. 2001, Doc.VCEG-M33.
6. T. Wedi and S. Wittmann, “Adaptive quantization encoding technique using an equal expected-value rule,” JVT-N011, *JVT of ISO/IEC MPEG & ITU-T VCEG, 14th Meeting*, Hong Kong, China, 18-21 Jan, 2005.