H.264 FAST INTRA-PREDICTION MODE DECISION BASED ON FREQUENCY CHARACTERISTIC

Takeshi Tsukuba†, Isao Nagayoshi‡, Tsuyoshi Hanamura‡ and Hideyoshi Tominaga†

† GITS, Waseda University, 29-7 building 1-3-10 Nishi-Waseda, Shinjuku-ku, Tokyo 169-0051 Japan
‡ Media Glue Corp. Ramdax Bldg. 8th floow, 2-4-12 Okubo, Shinjuku-ku, Tokyo 169-0072 Japan
phone: 03-5286-3385, fax: 03-5273-7368
E-mail: {tsukuba, isao, hana, tominaga}@tom.comm.waseda.ac.jp

ABSTRACT

In this paper, we propose a mode decision scheme for 4×4 intra-prediction in H.264 encoder to reduce its complexity. We focus on the characteristics of each prediction mode in terms of reducing its power in DCT domain since it is considered that reducing frequency components with higher level power may improve the efficiency of prediction. The proposed method reduces the candidates of prediction modes by classifying frequency characteristics of 4×4 block, which are computed from its low-frequency components of DCT coefficients. We evaluate the proposed method in terms of PSNR, bitrate, and the encoder complexity. Experimental results show that the proposed scheme reduces the complexity by nearly 60 % while keeping PSNR and bitrate equally to JM.

1. INTRODUCTION

In the video coding standard H.264[1], the spatial domain prediction is used to improve the efficiency of intra video coding. The spatial domain prediction includes 4×4 block intra-prediction and 16×16 block intra-prediction. In the 4 ×4 block intra-prediction, there are nine different prediction modes designed in a directional manner, which are shown in Fig.1

In the Reference software "JM"[2], the Rate-Distortion Optimization Method[3] is applied to all of the 4×4 block intra-prediction modes in order to find the best one with smallest cost. Although this approach can achieve the optimal prediction mode decision, it is costly in the computational complexity. Hence, it is desirable to develop the 4×4 intra-prediction mode decision techniques with lower complexity.

There are following two approaches in the 4×4 block intra-prediction mode decision methods.

- The mode decision based on the simple cost calculation.
- The mode decision based on the signal characteristic.

The former employs a partial computation of the cost function[4]. The latter employs the idea that the optimal prediction mode of a block is strongly associated with the dominant edge direction within that block. In [5] and [6], a fast intra-prediction mode decision based on edge direction histogram is proposed. This method reduces the number of possible modes to achieve the low complexity. However, the computation of dominant edge direction needs to a lot of additional complexity to calculate each pixels' edge direction. In [7], another fast intra-prediction mode decision is proposed that employs the simpler computation of the dominant edge direction within a block.

In this paper, we propose a mode decision scheme for 4×4 intra-prediction to reduce the complexity of the H.264 encoder. Proposed method is categorized to the mode decision based on the signal characteristic. We focus on the characteristic of each prediction mode to reduce the power of input 4×4 pixels. According to that characteristic, we derive the parameters to classify the distribution of frequency component of input signal. The proposed method reduces the candidates of prediction modes by using these parameters.

First, we describe the parameters and the detailed algorithm of mode decision in section 2 and section 3, respectively. At last, we show the effectiveness of proposed method by simulations.

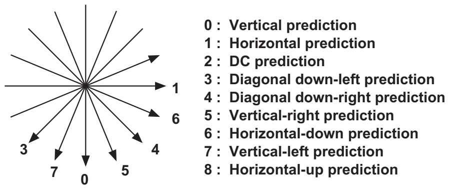

Figure 1: 4×4 block intra-prediction mode direction

2. THE FEATURES BASED ON FREQUENCY CHARACTERISTIC

2.1 Discussion about the characteristics of 4×4 intra-prediction

Here, we discuss about the characteristics of 4×4 intra-prediction. It is considered that reducing frequency components with higher level power in input signal may improve the efficiency of prediction. First, the spatial prediction process can be described as Eq.(1).

$$\mathbf{Y} \;=\; \mathbf{X}-\mathbf{X_P}, \tag{1}$$

where $\mathbf{X}$, $\mathbf{X_P}$, and $\mathbf{Y}$ denotes input 4×4 pixel signal, intra-prediction signal, and residual signal respectively. Then, the expression of Eq.(1) in DCT domain can be described as Eq.(2). Because DCT transform is linear transformation, we can rewrite Eq.(2) as Eq.(3).

$$\begin{aligned} T_{DCT}\{\mathbf{Y}\} &= T_{DCT}\{\mathbf{X}-\mathbf{X_P}\}, & (2)\\ &= T_{DCT}\{\mathbf{X}\}-T_{DCT}\{\mathbf{X_P}\}, & (3)\end{aligned}$$

where $T_{DCT}\{\cdot\}$ represents the 4×4 DCT used in H.264[8]. According to Eq.(1) and Eq.(3), spatial domain differences and DCT domain differences are equivalent. Take mode 0 (vertical prediction) shown in Fig.2

as an example, we can represent the prediction of the current block as:

$$\mathbf{X_P} \quad = \quad \begin{pmatrix} A & B & C & D \\ A & B & C & D \\ A & B & C & D \\ A & B & C & D \end{pmatrix}, \qquad (4)$$

The coefficients of the prediction can be described as Eq.(5).

$$T_{DCT}\{\mathbf{X_P}\} \quad = \quad \begin{pmatrix} S & T & U & V \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad (5)$$

$$\begin{aligned} S &= 4A + 4B + 4C + 4D, && (6) \\ T &= 8A + 4B - 4C - 8D, && (7) \\ U &= 4A - 4B - 4C + 4D, && (8) \\ V &= 4A - 8B + 8C - 4D, && (9) \end{aligned}$$

According to Eq.(5), its coefficients only exist on the top row. In this case, it is appropriate to apply mode 0 to the input signal which consists of horizontal-frequency components with higher level power. Same discussion will apply other prediction modes. The optimal prediction mode may be determined according to the power distribution of input signal, i.e.

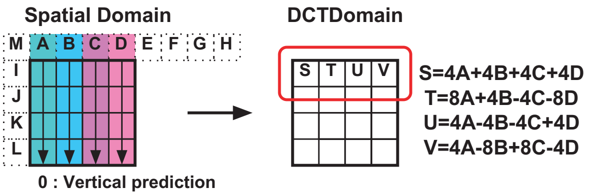

Figure 2: 4 × 4 block intra-prediction signal on the DCT domain

2.2 Definition of the parameters; C_h, C_v, E_h, E_v

First, We introduce the following parameters, C_h, C_v, E_h, E_v. Here, we define the DCT coefficients matrix of input signal $\mathbf{X}$ as $\mathbf{Z}$. $\mathbf{Z} = T_{DCT}\{\mathbf{X}\}$, i.e. C_h and C_v indicate the dominant gradient of the current block, which are defined as Eq.(10).

$$C_h \quad = \quad \lfloor Z_{01}/S \rfloor, C_v = \lfloor Z_{10}/S \rfloor, \qquad (10)$$

E_h and E_v indicate the bias condition of power, which are defined as Eq.(11).

$$E_h \quad = \quad \left\lfloor \sum_{j=1}^{3} Z_{0j}^2/S^2 \right\rfloor, E_v = \left\lfloor \sum_{i=1}^{3} Z_{i0}^2/S^2 \right\rfloor, \quad (11)$$

where S is a scaling factor, $\lfloor \cdot \rfloor$ represents the floor function.According to four parameters, the frequency distribution is categorized into some typical parts.

3. PROPOSED INTRA-PREDICTION MODE DECISION METHOD

3.1 Algorithm

Proposed method consists of the following three steps. The flowchart is shown Fig.3.

Step A. The mode selection based on the frequency characteristic

1. Transform the input 4×4 pixel signal.
2. Compute C_v, C_h, E_v, and E_h.
3. Categorize the frequency distribution into 8 cases by using Table 1.
4. If it is categorized to Case 2 or Case 5, intra-prediction mode is set to DC prediction(mode 2).Then terminate the process. Otherwise, go to Step A-4.
5. If it is categorized to Case 6, or Case 7, then go to Step B. Otherwise, go to Step C-1.

Table 1: Distribution pattern and prediction modes according four frequency features

Case	The relationship among features	Candidates-modes
0	$E_v = E_h = 0$	2
1	$E_h = 0$, $E_v > 0$	0,2
2	$E_v = 0$, $E_h > 0$	1,2
3	$E_v = E_h > 0$, $C_v \times C_h < 0$	3,2
4	$E_v = E_h > 0$, $C_v \times C_h > 0$	4,2
5	$E_v = E_h > 0$, $C_v \times C_h = 0$	2
6	$E_v > E_h > 0$	0,3,4,5,7,2
7	$E_h > E_v > 0$	1,3,4,6,8,2

Step B. The mode selection based on the simple cost function

1. Compute the cost of the each mode defined in Table 1(Case 6,7) by SATD[1].
2. Sort the cost of each mode in small order.
3. Choose N modes with small cost from candidate-modes.
4. Choose 1∼N modes from N modes by the threshold scheme and go to Step C-1.

Step C. The mode decision by the Rate-Distortion Optimization Method

1. Compute the cost of the each mode defined in Table 1(Case 1,2,3,4) or selected in Step B(Case 6,7) by the Rate-Distortion Optimization Method[2].
2. Choose the mode with the smallest cost.

In the following section, we describe the details of Step B and Step C.

3.2 Step B. Mode selection based on the simple cost function

If Case 6, or Case 7 is selected in Step A, cost evaluation based on SATD in Step B is performed to reduce the number of the possible modes to 1 ∼ N modes.

First, SATD cost of each mode shown in Table 1 is computed by using Eq.(12). SATD is defined as Eq.(13).

$$\begin{aligned} Cost(m) &= SATD(m) + Const., && (12) \\ SATD(m) &= \sum_{i=0}^{3}\sum_{j=0}^{3} |T_H\{X_{ij} - P_{ij}(m)\}|, && (13) \end{aligned}$$

---

[1] the sum of the absolute transform differences

$$\mathcal{M}_1 \quad = \quad \begin{cases} \{0,3,4,5,7,2\} & (Case6), \\ \{1,3,4,6,8,2\} & (Case7), \end{cases} \qquad (14)$$

where $\mathcal{M}_1$ is the candidate-modes set(Case6, or Case7), m is the prediction mode number that meets $m \in \mathcal{M}_1$, $P_{ij}(m)$ is the element of the prediction signal $\mathbf{X_P}$, and $T_H\{\cdot\}$ shows the Walsh-Hadamard Transform. *Const.* is taken into account the cost of the header information, which is given by Eq.(15).

$$Const. \quad = \quad \begin{cases} 0 & (m = mpm) \\ 4 \times QUANT(QP) & (m \neq mpm) \end{cases} (15)$$

where mpm represents the most probable mode that is a smaller number between the upper block and the left block[1], and $QUANT(\cdot)$ is the quantization step corresponding to QP.

Secondly, candidate-modes in $\mathcal{M}_1$ are sorted in small order of the cost value. Here, we redefine the sorted candidate-mode number as m_k. That is, m_k means the mode with k th smaller cost. Then, we can rewrite Eq.(14) as Eq.(16).

$$\mathcal{M}_2 \quad = \quad \{m_1, m_2, m_3, m_4, m_5, m_6\}, \qquad (16)$$

Thirdly, N modes are chosen from candidate-modes in Eq.(16).

$$\mathcal{M}_3 \quad = \quad \{m_k | k \leq N, k = 1, 2, \cdots, 6\}, \qquad (17)$$

Finally, the mode that meets Eq.(18) is determined as candidate-modes for Step C. Then the number of candidate-modes vary from 1 to N.

$$\mathcal{M}_{StepC} = \{m | Cost(m) < min\_cost + TH, m \in \mathcal{M}_3\}, (18)$$

where min_cost is the smallest cost value in candidate-modes $\mathcal{M}$. N and TH are thresholds.

3.3 Step C. The RDO mode decision scheme

In Step C, the best prediction mode is selected from candidate-modes defined in Table 1(Case 0,1,3,4), or $\mathcal{M}_{StepC}$ applying Rate-Distortion Optimization(RDO) Method. The RDO mode decision method finds the optimal prediction mode in the view of rate distortion. This method computes RD cost based on the actual rate and distortion after successive processes, transform, quantization, entropy coding, and reconstruction. The RD cost J is defined as Eq.(19).

$$J \quad = \quad SSD + \lambda \cdot R, \qquad (19)$$

$$\lambda \quad = \quad 0.85 \times 2^{(QP-12)/3}, \qquad (20)$$

where λ is the Lagrangian multiplier, SSD means the sum of the squared differences between the input 4×4 luminance block and its reconstruction signal, and R represents the number of actual bits associated with the chosen mode. The mode that minimize J is chosen as the best mode.

4. EXPERIMENTS

We implemented the proposed methods on the JM 9.2. The system platform is the Intel Pentium 4 processor 3.0 GHz CPU, 1024MB DDR RAM, and Vine Linux 2.6 CR(kernel 2.4.22). For experiments, we have tested four sequences, such as "Foreman(QCIF:176x144)", "News(QCIF)", "Mobile & Calendar(CIF:352x288)", and "Tempete(CIF)". The total number of frames is 150 for each sequences. Table 2 shows simulation conditions. For evaluation, we compared to the JM method[2]. Table 3 shows comparison methods.

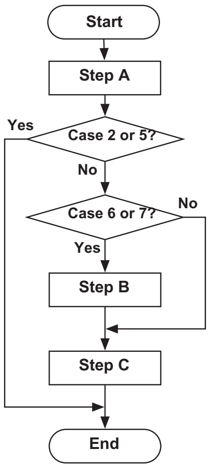

Figure 3: Flowchart

Table 4 shows the difference of average data bits(ΔBits), the difference of average PSNR(ΔPSNR), and the difference of average coding time(ΔTime) from JM[2]. It can be seen the that the Method 2 achieves about 60% time saving, while the Method 1 achieves 55% time saving. This means that the time spent for encoding of the Method 2 is smaller than that of the Method 1.

The degradation of PSNR of the Method 1 is about 0.1 ∼ 0.2dB, while that of the Method 2 is about 0.15∼0.25dB. Compared with Method 3, the performance loss is small.

Equivalently, the increase of bitrate of the Method 1 is about 0.80 ∼ 2.0%, while that of the Method 2 is about 1.0 ∼ 2.0 %. Figure 4, and 5 show the rate-distortion curves of the sequences "Foreman" and "Mobile & Calendar". It can be seen that each RD curve of Method 1 shows better than that of Method 2. This means that Method 1 has better performance in R-D sense. In general, our proposed methods reduce the encoder complexity by nearly 55∼60% with a little R-D performance loss in any quantization parameters, and in any sequences, compared to JM.

Table 2: Simulation conditions

Symbol Mode	CAVLC
RD Optimization	On
GOP Structure	All Intra-frame
Frame Rate	15[fps](QCIF), 30[fps](CIF)
Intra-prediction	intra 4 × 4 prediction only
QP	16,18,· · ·,32(const.)

Table 4: Results of sequences

(a) Foreman

Method	Δ Time(%)	Δ PSNR(dB)	Δ bitrate(%)
Method 1	-54.73	-0.084	1.12
Method 2	-61.66	-0.117	1.25

(b) News

Method	Δ Time(%)	ΔPSNR(dB)	Δbitrate(%)
Method 1	-56.30	-0.102	1.95
Method 2	-62.57	-0.165	2.06

(c) Mobile & Calendar

Method	ΔTime(%)	ΔPSNR(dB)	Δbitrate(%)
Method 1	-54.83	-0.199	0.775
Method 2	-61.51	-0.263	0.921

(d) Tempete

Method	ΔTime(%)	ΔPSNR(dB)	Δbitrate(%)
Method 1	-52.98	-0.172	0.889
Method 2	-60.69	-0.256	1.03

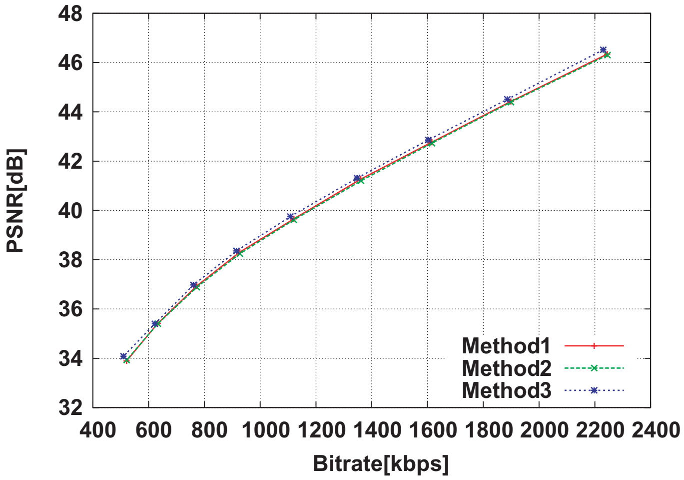

Figure 4: The R-D performance of the QCIF sequence "Foreman"

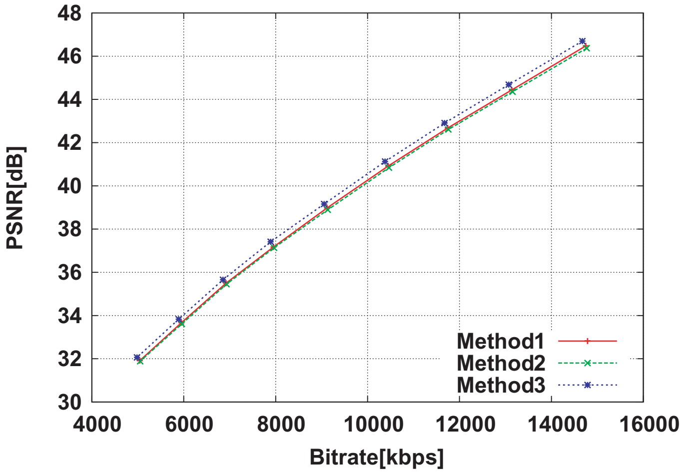

Figure 5: The R-D performance of the CIF sequence "Mobile & Calendar"

Table 3: Comparison methods

Method	Algorithm	Parameter sets
Method 1	Proposed Method	S=16, N=2, TH=256
Method 2	Proposed Method	S=16, N=3, TH=256
Method 3	JM	-

5. CONCLUSION

In this paper, we propose an efficient mode decision scheme for intra 4×4 predection in H.264. Its feature is reducing the candidates of prediction modes by classifying frequency characteristics, which are computed from its low-frequency components of DCT coefficients. The proposed method can reduce the encoder complexity by nearly 60 % while keeping PSNR and bitrate.

REFERENCES

[1] ITU-T Recommendation H.264, "Advanced video coding for generic audiovisual services, " 2003.

[2] Joint Video Team, "Reference Software JM9.2, " http://iphome.hhi.de/suehring/tml/

[3] G. Sullivan, T. Wiegand, "Rate-distortion optimization for video compression, " IEEE Signal Proc. Magazine, 1998

[4] Meng,Wong, Lam,"Efficient intra-prediction mode selection for 4x4 blocks in H.264, " IEEE Int. Conf. Multimedia & Expo, 2003

[5] F. Pan, X. Lin, R. Susanto, "Fast Mode Decision for Intra Prediction, " JVT-G013, 7th Meeting: Pattaya II,Thailand, 7-14 March, 2003

[6] F. PAN, "A DIRECTIONAL FIELD BASED FAST INTRA MODE DECISION ALGORITHM FOR H.264 VIDEO CODING, " IEEE Int. Conf. Multimedia & Expo, 2004

[7] Z. Yong-dong, D. Feng, L. Shou-xum, "Fast 4× 4 Intra-prediction Mode Selection for H.264, " IEEE Int. Conf. Multimedia & Expo, 2004

[8] Iain E. G. Richardson, "H.264/MPEG-4 Part 10 White Peper: Transform & Quantization, " www.vcodex.com, May 2003