A STUDY ON FAST RATE-DISTORTION OPTIMIZED CODING MODE DECISION FOR H.264

Akiyuki Tanizawa[I], *Shin-ichiro Koto*[I], *Takeshi Chujoh*[I] and *Yoshihiro Kikuchi*[II]
(I) Multimedia Laboratory, Corporate Research & Development Center
TOSHIBA Corporation, Kawasaki, Japan
(II) Core Technology Center, TOSHIBA Corporation, Ohme, Japan
akiyuki.tanizawa@toshiba.co.jp

ABSTRACT

The H.264 video coding standard can achieve considerably higher coding efficiency than any other existing standards by deciding the best mode among many prediction modes and various sizes of prediction blocks. Although the coding efficiency is improved by using Lagrange optimization for the mode decision, computational complexity increases significantly at the encoder. In this paper, we propose the Fast Rate-Distortion Optimization method for the hierarchical and adaptive coding mode decision in order to reduce the number of candidates for the best coding mode.

1. INTRODUCTION

H.264, a next-generation video coding standard [1], is ready for various applications, such as DVD, broadcasting and mobile telecommunication. H.264 offers an approximately twofold improvement in coding efficiency compared with conventional standards, such as MPEG-2 or MPEG-4 [2].

H.264 has many motion compensation block sizes and can use multiple reference frames for the inter prediction, and many prediction directions for the intra prediction. Selecting the best coding mode among many prediction modes contributes mainly to the high coding efficiency of H.264. In order to decide the best coding mode, the Rate-Distortion Optimization (RDO) method is used [3]. The method is also introduced to H.264 reference software, JM [4]. The RDO method based on the Lagrange optimization techniques can improve the coding efficiency significantly. However, the required computational complexity of RDO mode decision is very high at encoding.

In this paper, we propose a fast Rate-Distortion Optimization (Fast RDO) method in order to reduce the computational complexity for mode decision without decreasing coding efficiency. The proposed method is based on 2-step hierarchical mode decision. Furthermore, we propose a technique for a fast intra (Fast Intra) prediction mode decision for intra prediction in inter coding picture, such as P and B-picture.

In Section 2, the Rate-Distortion Optimization method implemented in JM is explained. In Section 3, the Fast RDO method is proposed. In Section 4, the Fast Intra prediction method for Intra4x4 and Intra16x16 is proposed. In Section 5, some experimental results are presented.

2. RATE DISTORION OPTIMIZATION (RDO)

It is well known that the RDO method improves coding efficiency compared with conventional mode decision. To select the best mode for a macroblock, the RDO method based on the Lagrange method is used.

Distortion D and rate R are dependent on source samples S and a coding mode I, respectively, for a given quantization value QP. To minimize the distortion for the source samples subject to a given rate constraint, Lagrange cost (R-D cost) J can be formulated as

$$J(S, I \mid \lambda) = D(S, I) + \lambda R(S, I), \quad (1)$$

where $\lambda\ (\geq 0)$ is the Lagrange parameter. The distortion D is measured as the sum of the squared differences (SSD) between the local decoded image samples and the original image samples in a current block. If a rate constraint Rc is given, the best mode is chosen from all possible coding modes that satisfy $Rc > R$. The Lagrange parameter λ is a constant value dependent on a quantization scale QP and is introduced by the next equation in JM.

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

Thus, when a QP is given, the Lagrange parameter λ is calculated according to (2) and the best coding mode I that gives minimum J is selected with (1).

R and D of all coding modes must be calculated by tentative coding in order to determine the minimum J. Because H.264 has many coding modes, huge computational burden is required in order to calculate J for each mode.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

3. FAST RATE-DISTORTION OPTIMIZATION (Fast RDO) METHOD

To reduce computational complexity, 2-step hierarchical mode decision is proposed. In the first step, a simple R-D cost without tentative coding is calculated. One or more coding candidates are selected using the R-D cost. The number of candidates varies depending on the quantization value.

In the second step, the conventional RDO method is applied to the candidates selected in the first step. The first step is explained below.

3.1. Simple R-D Cost

In JM, the RDO method and the RDO-off method are implemented. The coding mode decision of the latter method is performed using only a simple coding cost, which does not require the tentative coding. In the first step of the Fast RDO method proposed in this paper, the best coding candidates are determined using the simple R-D cost.

The difference between an original sample and a predictive sample, is given by the next equation for position (x,y).

$$Diff(x,y) = s(x,y) - p^I(x,y) \tag{3}$$

s corresponds to the original sample and p^I corresponds to the predictive sample for coding mode I. The sum of absolute differences (SAD) is given by

$$SAD = \sum_{(x,y)} |\, Diff(x,y)\,| \cdot \tag{4}$$

Because a positive correlation exists between the SAD and the generated bits, coding modes with minimum SAD are selected generally as candidates of the first step. In addition, it should be taken into consideration that a predictive error signal is coded by using 2-dimensional orthogonal transformation such as DCT. A simple cost is applied based on the sum of the absolute transformed differences ($SATD$) using Hadamard transform. The coefficient of transformation is given by 4 point Hadamard transform.

$$T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \tag{5}$$

This transform is performed horizontally and vertically, resulting in $Diff\mathrm{T}(x,y)$. Finally, $SATD$ of the present prediction mode for the block is given by

$$SATD = (\sum_{(x,y)} |\, Diff\mathrm{T}(x,y)\,|)/2 \cdot \tag{6}$$

Furthermore, the generated bit to code certain mode index is added to a simple coding cost as an overhead OH of coding mode. For example, in inter coding mode, OH corresponds to the sum of code of the reference frame and the associated motion vector. Using (6), a simple R-D cost is produced by

$$Cost = SATD + \lambda \times OH \,. \tag{7}$$

3.2. Adaptive Selection of Candidates

In this section, we reduce the number of candidates using the simple R-D cost which is calculated in (7). If the simple R-D costs among several coding modes are large, these coding modes should not be chosen as the best mode, because it is thought that prediction error is large or many bits are generated. We eliminate these coding modes in the first step. In the second step, since tentative coding is not needed for the eliminated modes, the computational burden can be reduced significantly.

We confirmed that there was strong correlation between the simple coding cost and the R-D cost by preliminary experiments using several images. In addition, this correlation changed not according to images but according to the quantization value. Therefore, the degradation of the coding efficiency can be suppressed by changing the number of the candidates corresponding to a quantization value adaptively in the first step. An example of selection of candidates is shown in Fig. 1.

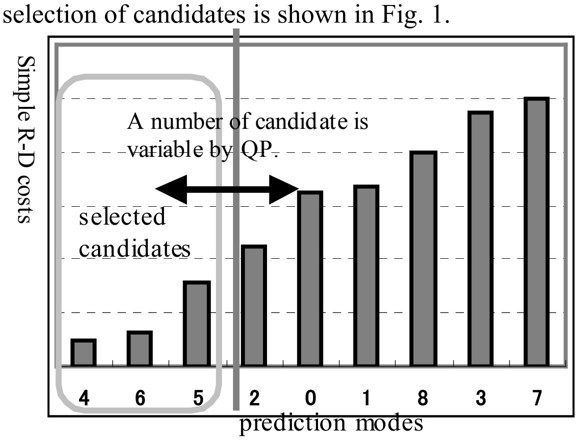

Fig. 1 An example of selection of candidates

The modes that give lower costs are selected according to the number of coding candidates. Since the mode that is not chosen here is not coded tentatively in the second mode decision step, the reduction of the computational burden is realized.

4. FAST INTRA PREDICTION METHOD

I-picture has slight computational complexity because it does not require motion estimation (ME). Degradation of the picture quality of I-picture decreases the coding efficiency for the succeeding predictive coded pictures, because I-picture will become a reference picture used as the basis of inter prediction. We apply the Fast RDO method, which was proposed in section 3, to I-picture. On the other hand, although intra predictive coding mode in P or B-picture is seldom chosen, all predicted samples must be generated. Therefore, we want to reduce computational-load as much as possible.

In this section, we explain the Fast Intra prediction method for Intra4x4 prediction and Intra16x16 prediction that is specialized in P and B-picture.

4.1. Hierarchical Prediction for Intra4x4

There are 9 prediction candidates in Intra4x4, which consist of 8 prediction directions and DC prediction. The Intra4x4 prediction directions are shown in Fig. 2.

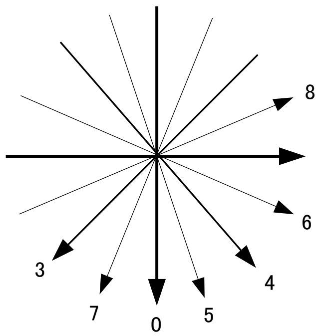

Fig. 2 Intra 4x4 prediction

Intra4x4 has each prediction angle of 22.5 degrees from prediction 0 to 8 except DC prediction 2. In the case of JM, since all these prediction modes are used as candidates, computational complexity to generate predicted samples increases. Therefore, we propose the following hierarchical method using the prediction direction of neighboring blocks and the most probable mode, which is calculated using a mode index of the top block and the left block for the current block.

Step1: The most probable mode is calculated from the mode information of the neighboring 4x4 blocks.

Step2: The prediction samples in the most probable mode are generated and the simple coding cost is computed. If the obtained cost is smaller than a threshold *Th1*, go to step7. Otherwise step3.

Step3: Vertical 0, horizontal 1, diagonal-down-left 3 and diagonal-down-right 4 are predicted and these simple coding costs are computed. The mode that gives the minimum cost is chosen. If the differences of the cost in each obtained mode are smaller than a threshold *Th2*, go to step5. Otherwise step4.

Step4: 2 neighboring modes of the mode chosen by step3 (for example, if the selected mode is 1, the neighboring modes are 6 and 8) are evaluated, and go to step6.

Step5: DC prediction is performed and its cost is computed.

Step6: The mode that gives the minimum cost in the selected modes is chosen as the best mode.

Step7: It returns to step1 and the succeeding Intra4x4 blocks are processed.

This method performs Intra4x4 prediction once in the best case and 7 times in the worst case. The threshold *Th1* takes the value that is changed according to *QP* and is applied to an early termination of Intra4x4 prediction according to the accuracy of Intra4x4. *Th2* is a threshold to select DC prediction when no predominant direction exists. By the hierarchical mode decision, the number of operations needed to generate predicted samples can be reduced considerably with little decrease of coding efficiency.

4.2. Adaptive Mode Decision for Intra16x16

Intra16x16 prediction is performed after all Intra4x4 modes in the current MB are decided using the above method. The best mode of Intra16x16 prediction is determined from the variance and the weighted frequency of 16 modes' indices acquired by Intra4x4 prediction. The weighted frequency is given by

$$H = \sum_{k}^{k=16} \varphi(M_k),\quad \varphi(M_k) = \begin{cases} 4 & \text{if } (M_k = 1) \\ 2 & \text{if } (M_k = 6 \text{ or } M_k = 8) \\ 1 & \text{if } (M_k = 3 \text{ or } M_k = 4) \\ 0 & \text{others} \end{cases}, \qquad (8)$$

$$V = \sum_{k}^{k=16} \phi(M_k),\quad \phi(M_k) = \begin{cases} 4 & \text{if } (M_k = 0) \\ 2 & \text{if } (M_k = 5 \text{ or } M_k = 7) \\ 1 & \text{if } (M_k = 3 \text{ or } M_k = 4) \\ 0 & \text{others} \end{cases}. \qquad (9)$$

M_k shows a prediction direction of Intra4x4 to block *k* in Fig.2 and a vertical and a horizontal weighted frequency are shown in *H* and *V*, respectively. Although Intra4x4 modes have prediction angles that show directions of 22.5 degrees, Intra16x16 modes have only vertical, horizontal, DC and plane prediction. Adaptive selection of the best mode in Intra16x16 is explained in the following steps.

Step1: When *H* and/or *V* are larger than a threshold *Th3*, horizontal and/or vertical prediction samples are generated, and the simple R-D cost is computed.

Step2: The plane prediction is also performed when the variance of mode indices for Intra4x4 is the smaller than a threshold *Th4* and the simple coding cost is computed.

Step3: If step1 and step2 are not selected, then DC prediction is performed and the simple coding cost is computed.

Step4: The mode giving the minimum coding cost is selected as the best coding mode.

The best simple R-D cost is compared with the combination of best coding cost for Intra4x4 and smaller one is selected. The best coding mode for intra is determined.

5. EXPERIMENTAL RESULTS

The conventional RDO method, the RDO-off method and the proposed method (combination of the Fast RDO method and the Fast Intra method) are simulated on QCIF sequences, Foreman, Container and Stefan and SDTV sequences of ITE standards videos, 17SD, 20SD, 30SD and 37SD. Experimental conditions are shown in Table 1.

We calculated the averaged PSNR gain (ΔPSNR), the bitrate increase (ΔBR) and speed ratio (SR) based on the quantization values. ΔPSNR1, ΔBR1 and SR1 correspond to the difference between the RDO method and the RDO-off method. ΔPSNR2, ΔBR2 and SR2 correspond to the difference between the RDO method and the proposed method. The speed ratio is computed except ME time, because the Fast RDO method is independent of the implementation of ME.

The simulation results are shown in Table 2 for QCIF and in Table 3 for SDTV. In the RDO-off method, although speed ratio improves, coding efficiency decreases considerably. From the QCIF results, we observed that the maximum ΔBR2 increase is 1.7% and the maximum ΔPSNR2 decrease is -0.06 dB, and the proposed method is more than 2.6 times as fast as the conventional RDO. From the SDTV results, we observed that the maximum ΔBR2 increase is 1.3% and the maximum ΔPSNR2 decrease is -0.05 dB, and SR2 is more than 4.0. Fig. 3 shows R-D curves of SDTV results. The R-D curves show that the proposed method works well in a wide range of bitrate. We confirmed that the proposed method could realize almost the same coding efficiency with significant complexity reduction.

The proposed method for mode decision is independent of a complexity reduction technique for ME. If it is combined with our proposal, encoding speed will be improved additionally.

Table 1. Experimental Conditions

Number of References Frame	1
Frame structure	IPPP…
Quantization parameter	20,24,28,32(fix)
Entropy Coding	CABAC(SD), CAVLC(QCIF)
Number of Frames	300
Frame rate	30(SD),15(QCIF)
MV search range	±16 pixel

Table 2. Experimental Results (QCIF)

	ΔBR1	ΔBR2	ΔPSNR1 [dB]	ΔPSNR2 [dB]	SR1	SR2
foreman	8.29%	1.19%	-0.31	-0.05	3.30	2.69
container	10.49%	1.66%	-0.50	-0.06	3.15	2.62
stefan	7.33%	-1.51%	-0.37	0.08	3.31	2.72

Table 3. Experimental Results (SDTV)

	ΔBR1	ΔBR2	ΔPSNR1 [dB]	ΔPSNR2 [dB]	SR1	SR2
17SD	7.69%	0.49%	-0.39	-0.03	7.12	4.06
20SD	3.89%	0.51%	-0.15	-0.02	6.67	4.26
30SD	10.95%	1.24%	-0.40	-0.05	5.78	4.12
37SD	7.10%	1.21%	-0.26	-0.05	5.84	4.01

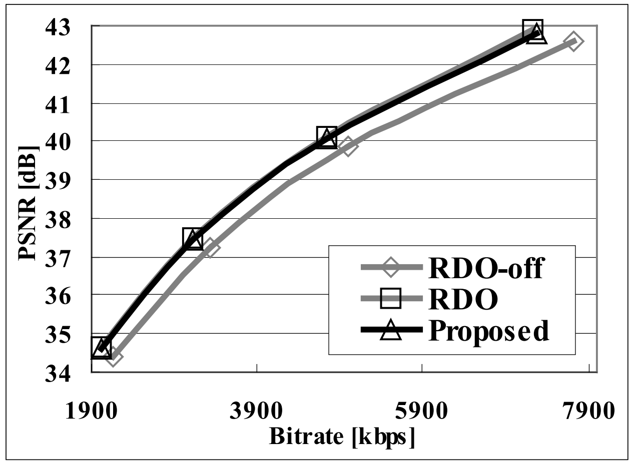

Fig 3. R-D curves (30SD-Crowded_Crosswalk)

6. CONCLUSION

In this paper, a fast RDO method is proposed in order to reduce the computational complexity for mode decision. The proposed method consists of the hierarchical mode decision with adaptive candidate reduction. In addition, a fast intra prediction method for inter predictive coding pictures is introduced using the hierarchical prediction. Experimental results show that the proposed methods can reduce the computational complexity significantly without decreasing the coding efficiency.

7. REFERENCES

[1] H.264, *Draft ITU-T Recommendation and Final Draft International Standard*, Pattaya, Thailand, 2003

[2] T. Wiegand, Rate-Constrained Coder Control and Comparison of Video Coding Standard, IEEE Trans. on Circuits and Systems for Video Technology, vol 13, No 7, July 2003

[3] T. Wiegand and B. Girod, Multi-frame motion compensated prediction for video transmission, Kluwer Academic Publishers 2001

[4] JVT Reference Software http://bs.hhi.de/~suehring/tml/download/

[5] B. Meng, Efficient Intra-Prediction Mode Selection for 4x4 Blocks in H.264, IEEE Trans. On Circuits and Systems for video technology, vol. III-521 ICME 2003