# Fast Mode Decision for H.264/AVC on SATD Value

Yuta Kuboyama* and Yoshimitsu Kuroki*
* Kurume National College of Technology
1-1-1, Komorino, Kurume-shi, Fukuoka 838-8555 Japan
Tel: +81-942-35-9392 Fax: +81-942-35-9398
E-mail: s42313yk@std.kurume-nct.ac.jp

***Abstract*—H.264/AVC is the state-of-the-art video coding standard, which has various functions to realize high compression performance. The codec prepares several modes in both intra- and inter-prediction, and chooses the best one by some criterion. Therefore, the encoder requires a heavy burden. This paper describes a fast mode decision method on Sum of Absolute Transformed Differences (SATD) criterion. The proposed method prunes candidates by projecting difference blocks onto the canonical bases without calculating transformed differences, and guarantees to choose the best mode. Experimental results show that the proposed method reduces computational time by 16% compared with the exhaustive calculation performed by Joint Model (JM) 14.0.**

## I. Introduction

The latest video coding standard H.264/AVC [1][2] has been developed by the Joint Video Team (JVT) established by the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). This standard decreases coding rate almost half compared with conventional standards such as MPEG-2 and MPEG-4 Simple Profile. In addition, JVT enhanced the standard called Fidelity Range Extension (FRExt) as a profile for high definition video to improve encoding efficiency.

H.264/AVC employs intra- and inter-prediction technologies to achieve high compression efficiency. These predictions prepare several modes, and the best mode is chosen in some criterion. The reference software of H.264/AVC, named JM14.0 [3], calculates values of a cost function between an original image and predicted images over all modes in blockwise, then the mode having the smallest cost value is selected. JM14.0 uses either "High Complexity Mode" or "Low Complexity Mode". The former implements Rate-Distortion Optimization (RDO), and the latter is a one pass encoder without obtaining coding bits. As for either method, a problem remains that the encoder requires a heavy burden. Therefore, various methods to speed up encoding have been examined [4][5][6]. However, these methods postulate a degradation of the compression performance.

The purpose of this study is to propose a fast mode decision method on SATD criterion without loss in SNR and bitrate. Our method achieves fast decision by pruning candidates without computing SATD values in "Low Complexity Mode".

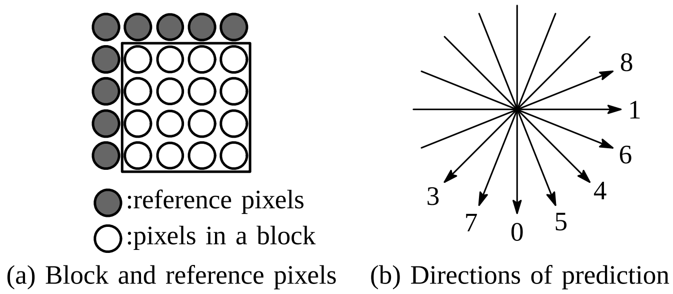


Fig. 1. Intra prediction of $4 \times 4$ pixels.

## II. Mode decision in H.264/AVC (JM14.0)

### A. Intra-prediction

The intra-prediction is a data-compression technique using a correlation of adjacent pixels in a frame. The best mode of the intra-prediction is assigned for each block. H.264/AVC prepares three kinds of blocks as follows; $4 \times 4$ pixels, $8 \times 8$ pixels, and $16 \times 16$ pixels.

We illustrate the intra-prediction modes of $4 \times 4$ blocks in Fig. 1, where (a) shows the positions of the current block and the reference pixels for prediction, and (b) shows nine directions of prediction modes except for the mode "2", which means a mean value computed by the reference pixels shown in Fig. 1 (a). The directions of the prediction are also used to predict the mode itself in order to reduce coding rate. The $8 \times 8$ block has the same nine modes as $4 \times 4$ blocks, whereas $16 \times 16$ blocks have four modes. We have to choose the best mode considering the various block sizes.

### B. Inter-prediction

The inter-prediction, frequently called motion-compensation prediction, contributes to reduce large amount of coding bits. In the inter-prediction, sets of vertical and horizontal differences of block positions, namely motion vectors, are encoded. Motion vectors also correlate with neighboring motion vectors; therefore, the motion vector is predicted by reference blocks. The inter-prediction in H.264/AVC defines various block sizes; $16 \times 16$ pixels, $16 \times 8$ pixels, $8 \times 16$ pixels, $8 \times 8$ pixels, $8 \times 4$ pixels, $4 \times 8$ pixels, and $4 \times 4$ pixels. In addition, H.264/AVC can refer to plural frames in order to accomplish effect prediction for occluded or uncovered regions.

Accordingly, H.264/AVC requires enormous computational load to realize precise prediction.

### C. Mode decision on SATD value

As mentioned above, there are many modes in the intra- and the inter- prediction, and it is necessary to choose the most suitable mode. JM14.0 defines three cost functions and chooses the mode having the smallest value on a employed function as the most suitable mode. The functions are as follows; Sum of Squared Differences (SSD), Sum of Absolute Differences (SAD), and SATD. We would like to narrow the discussion down to the "Low Complexity mode" on SATD criterion. In this case the cost function appears as

$$J_{SATD} = SATD + SATD_0. \tag{1}$$

The first term $SATD$ estimates the cost of degradation calculated by a current block $\mathbf{S} = (s_{ij})$ $(i, j = 1, 2, \ldots, H)$ and $k$-th prediction block $\mathbf{P}_k = (p_{ij})_k$ as follows:

$$SATD(\mathbf{S}, \mathbf{P}_k) = \sum_{(i,j) \in Block} |h_{ij}|, \tag{2}$$

where the $h_{ij}$'s are elements of the transformed matrix $\mathbf{H}$. The elements indicate the transformed differences between $\mathbf{S}$ and $\mathbf{P}_k$ in Hadamard bases. Here, $\mathbf{H}$ is obtained by using the $H$-dimensional $(H = 4, 8, \ldots)$ transform matrix $\mathbf{T}_H$ as follows;

$$\mathbf{H} = \mathbf{T}_H(\mathbf{P}_k - \mathbf{S})\mathbf{T}_H^T. \tag{3}$$

Let $H = 2$: then

$$\mathbf{T}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{4}$$

We can extend $\mathbf{T}_2$ to $H$-dimensions with tensorial product shown as;

$$\mathbf{T}_H = \mathbf{T}_2 \otimes \mathbf{T}_{\frac{H}{2}}. \tag{5}$$

The second term $SATD_0$ indicates the cost of coding bits defined by

$$SATD_0 = QP2Quant(QP) \times Headerbit. \tag{6}$$

$QP2Quant$ is transformative equation from quantization parameter (QP) to quantization scale. JM14.0 calculates a cost function for all modes and chooses the smallest one.

## III. Proposed Method

We would like to speed up the mode choice maintaining compression rate and the quality of the image. We explain the proposed method on a two-dimensional coordinate using Fig. 2.

### A. Basic method

The proposed method chooses the best mode by using the relation between the Hadamard bases and the canonical bases. Fig. 2 shows that the block $\mathbf{S}$ of the original image as the origin, and the block $\mathbf{P}'_k$ $(k = 1, 2, 3)$ are difference blocks between $\mathbf{S}$ and prediction block $\mathbf{P}_k$. Further, $\mathbf{e}_1$, $\mathbf{e}_2$ are the canonical bases, and $\mathbf{u}_1$, $\mathbf{u}_2$ are the hadamard bases. Then the procedure for choosing the mode is advanced as follows.

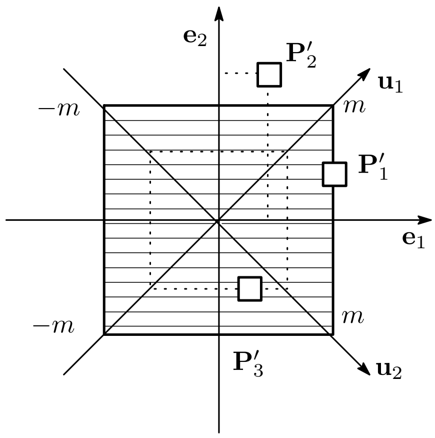


Fig. 2. Proposed method (2 dimensions).

1) Compute the $J_{SATD}$ value between $\mathbf{S}$ and $\mathbf{P}_1$. For the temporal smallest $J_{SATD}$ value, denoted by $m$, the area $SATD < m$ is shaded by horizontal lines in Fig. 2.
2) Project $\mathbf{P}'_2$ to the canonical bases.
3) If the absolute values of projection are greater than $\frac{m}{\sqrt{2}}$, $\mathbf{P}'_2$ is removed from the candidates because $SATD_0$ is a positive number. $\mathbf{P}'_2$ is pruned by the prediction value on $\mathbf{e}_2$.
4) Next, Project $\mathbf{P}'_3$ to the canonical bases.
5) In Fig. 2, $\mathbf{P}'_3$ can be the best mode, because the projection values are smaller than $\frac{m}{\sqrt{2}}$.
6) Compute the $J_{SATD}$ value between $\mathbf{S}$ and $\mathbf{P}_3$. If $J_{SATD}$ is smaller than $m$, the value $m$ is replaced, and the area $SATD < m$ is illustrated by the dotted line in Fig. 2.

As described above, the proposed method selects the best mode by making a projection to the canonical bases. That is to say, the proposed method selects the best mode by comparing an absolute cost values with an absolute values of the differential block's elements. The reason why the encoding time may be smaller than JM14.0 is that the proposed method do not have to compute $J_{SATD}$ for all modes. The mode selected by JM14.0 is corresponding to the mode selected by the proposed method.

### B. Order of comparison

Here, we consider the order of comparison to prune a candidate at an earlier stage. For example, $\mathbf{P}'_2$ in Fig.2 can

| | | | |
|---|---|---|---|
| 13 | 14 | 15 | 16 |
| 12 | 11 | 10 | 9 |
| 5 | 6 | 7 | 8 |
| 4 | 3 | 2 | 1 |

(a) mode 0/3/5/7

| | | | |
|---|---|---|---|
| 13 | 12 | 5 | 4 |
| 14 | 11 | 6 | 3 |
| 15 | 10 | 7 | 2 |
| 16 | 9 | 8 | 1 |

(b) mode 1/6/8

| | | | |
|---|---|---|---|
| 16 | 14 | 13 | 7 |
| 15 | 12 | 8 | 6 |
| 11 | 9 | 5 | 2 |
| 10 | 4 | 3 | 1 |

(c) mode 2/4

Fig. 3. Order of comparison for the intra-prediction in $4 \times 4$ pixels.

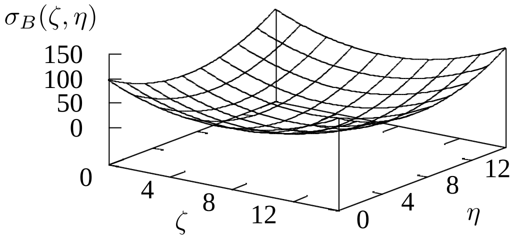


Fig. 4. Distribution of variance in $16 \times 16$ pixels.

| | | | |
|---|---|---|---|
| 1 | 5 | 6 | 2 |
| 12 | 13 | 14 | 7 |
| 11 | 16 | 15 | 8 |
| 4 | 10 | 9 | 3 |

Fig. 5. Order of comparison for the inter-prediction of $4 \times 4$ pixels.

be reduced without projection (comparison) to $\mathbf{e}_1$ by previous projection (comparison) to $\mathbf{e}_2$. In this process, the order of comparison is important to speed up. Therefore, we consider the variance of each canonical bases to select the order of comparison.

In the case of the intra-prediction, we use the property that the variance increases as the distance from the reference increases. Fig. 3 shows the order of the comparison in each mode of $4 \times 4$ pixels. The comparison order for the blocks having $8 \times 8$ pixels or $16 \times 16$ pixels are similar; then, we omit their depiction.

In the inter-prediction, the study published by Zheng, etc. tells that the farther pixels lie from the center of blocks, the greater variance the pixels indicate [7]. This attribute seems to be useful for the intra-prediction. From the theoretical approach, [7] indicates that the variance $\sigma_B$ of the inter-prediction block ($\zeta$,$\eta$) in $M \times N$ difference block appears as

$$\sigma_B(\zeta,\eta) = \alpha\left\{\left(\zeta-\frac{M-1}{2}\right)^2+s^2\left(\eta-\frac{N-1}{2}\right)^2\right\} +\sigma_n^2+\sigma_{me}^2. \quad (7)$$

Here $\alpha$ is a value depending on a motion vector and a pixel value, and $s$ shows the aspect ratio of a pixel. Moreover, $\sigma_n^2$ is variance of noise, and $\sigma_{me}^2$ is variance that originates in the presumption error margin of the motion vector. An example of (7) is shown in Fig. 4 at $M = N = 16$, $\alpha = s = 1$ and $\sigma_n^2 = \sigma_{me}^2 = 0$.

Then, the order of comparison for $4 \times 4$ pixels is determined as Fig. 5. The comparison orders of the other block size are similar; then, we omit their depiction.

TABLE I
CODING CONDITION FOR THE INTRA PREDICTION

| | |
|---|---|
| Profile | High 4:2:2 |
| Symbol mode | CABAC |
| Frame rate | 30.0 |
| GOP Structure | All intra-frame |
| RD Optimization | Off |
| Error metric | Hadamard transform |
| Frame skip | One frame |
| Quantization parameter | 28,32 |
| Total frames | 50 |

TABLE II
EXPERIMENTAL RESULT FOR THE INTRA PREDICTION

| | QP | Encoding time(s) | | | PSNR |
|---|---|---|---|---|---|
| | | JM14.0 | Proposed Method | rate (%) [a] | (dB) |
| mobile | 28 | 17.2 | 18.0 | △ 4.65 | 35.33 |
| | 32 | 19.8 | 22.8 | △ 15.2 | 31.94 |
| football | 28 | 18.8 | 18.2 | ▽ 3.19 | 37.55 |
| | 32 | 19.4 | 19.9 | △ 2.58 | 35.02 |
| | | | average | △ 4.78 | |

[a] △ is increase, and ▽ is decrease.

## IV. EXPERIMENTAL RESULTS

We compared the JM14.0 to the proposed method to verify the effectiveness. Moreover, we evaluate the intra-prediction and the inter-prediction independently, and measure the encoding time and Peak-Signal to Noise Ratio (PSNR). For experiments, we have tested two sequences; "mobile" and "football". These are YUV (4:2:2) format and $720 \times 480$ pixels in test sequence which Video Quality Experts Group (VQEG) offers [8]. The system platform is the Intel Xeon processor 2.33 GHz Dual-CPU, 2048MB DDR2 RAM, and CentOS release 5.0 (kernel 2.6.18).

### *A. Results of the intra-prediction*

We tested in the picture type as I-I-I-... under the condition specified in Table I to confirm the effectiveness for the intra-prediction. The encoding time and PSNR in the JM14.0 and the proposed method are indicated in Table II. In the encoding time, the difference between JM14.0 and the proposed method is insignificant. Therefore, the proposed method is not a effective method in intra-prediction.

As a probable cause, few modes seem to be reduced. Therefore, we counted the number of modes reduced by the proposed method. The result tells that the number of reduced mode is 0.2% on an average.

### *B. Results of the inter-prediction*

We tested in the picture type as I-B-P-B-P-... under the condition specified in Table III to confirm the effectiveness for the intra-prediction. The encoding time and PSNR in the JM14.0 and the proposed method are indicated in Table IV. There is not the performance loss at all. The proposed method reduced the encoding time by average 16%.

TABLE III
CODING CONDITION FOR THE INTER PREDICTION

| | | |
|---|---|---|
| Common | Profile | High 4:2:2 |
| | Symbol mode | CABAC |
| | Frame rate | 30.0 |
| | GOP structure | IBPBP structure |
| | RD optimization | Off |
| | Frame skip | One frame |
| | Quantization parameter | 28,32 |
| | Total frames | 50 |
| | Search range | 32 |
| | Number reference frame | 5 |
| | Error metric | Hadamard transform |
| P-frame | P slice List 0 reference override | Off |
| B-frame | B slice List 0 reference override | Off |
| | B slice List 1 reference override | One frame |

## V. CONCLUSION

In this paper, we propose a fast mode decision method on SATD value in H.264/AVC. The proposed method uses the relation of canonical bases and Hadamard bases, and prune candidates comparing the temporal smallest $J_{SATD}$ value to element blocks without calculation Hadamard transform. Furthermore, we discussed the orders of the comparison in the intra- and inter-predictions. The proposed method reduces the encoding time by average 16% while keeping coding performance.

The future direction of this study will be application RDO.

TABLE IV
EXPERIMENTAL RESULT FOR THE INTER PREDICTION

| | QP | Encoding time(s) | | | PSNR |
|---|---|---|---|---|---|
| | | JM14.0 | Proposed Method | rate (%) [a] | (dB) |
| mobile | 28 | 31321 | 25411 | ▽ 18.87 | 34.36 |
| | 32 | 29365 | 24415 | ▽ 16.86 | 31.16 |
| football | 28 | 34885 | 29961 | ▽ 14.11 | 36.48 |
| | 32 | 32032 | 27143 | ▽ 15.26 | 33.86 |
| | | | average | ▽ 16.28 | |

[a] △ is increase, and ▽ is decrease.

## REFERENCES


[1] ITU-T Recommendation H.264: "Advanced Video Coding for generic audiovisual services", 2003.

[2] ISO/IEC 14496-10: "Coding of audiovisual object - Part 10: Advanced Video Coding", 2003.

[3] Joint Video Team (JVT), "Reference Software. ver 14.0", http://iphome.hhi.de/syehring/tml/

[4] Akiyuki Tanizawa, Shinichiro Koto, and Takeshi Chujoh, "A Fast Rate-Distortion Optimized Coding Mode Decision for H.264", IEICE transactions on information and systems, vol.J89-D, No.1, pp.27-39, Jan. 2006 (Japanese).

[5] Jeyun Lee and Byeungwoo Jeon "Fast mode decision for H.264", Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on, Vol.2, pp.1131-1134, June 2004.

[6] Lai-Man Po, Kai Guo, "Transform-Domain Fast Sum of the squared Difference Computation for H.246/AVC Rate-Distortion Optimization", IEEE Transactions Circuits and Systems for Video Technology, Vol.17, No.6, pp.765-773, June 2007.

[7] Wentao Zheng, Yoshiaki Shishikui, Yasuaki Kanatsugu, Yutaka Tanaka, and Susumu Ito, "Analysis of space-dependent characteristics of motion-compensated frame differences based on a statistic motion distribution model", IEICE Transactions, vol.J84-D-2, No.9, pp.2001-2010, Sep. 2001.

[8] Video Quality Experts Group (VQEG), "Test Sequences", http://media.xiph.org/