FAST SELECTIVE-INTRA MODE SEARCH ALGORITHM BASED ON MACRO-BLOCK TRACKING FOR INTER-FRAMES IN THE H.264/AVC VIDEO STANDARD

Byung-Gyu Kim†, Chang-Sik Cho†† and Tae-jeong Kim†††

Embedded S/W Research Division, Electronics and Telecommunications Research Institute (ETRI)
161 Gajeong-dong, Yuseong-gu, Daejeon 305-350, Rep. of Korea
E-mail: bg.kim@ieee.org†, cscho@etri.re.kr††, taejung@chungbuk.ac.kr†††

ABSTRACT
A fast intra mode determination algorithm based on the macro-block (MB) tracking scheme and rate-distortion (RD) cost is proposed for inter-frames in the H.264/AVC video standard. In addition to the inter mode search procedure with variable block size, an intra mode search causes a significant increase in the complexity and computational load for an inter frame. To reduce the computational load of the intra mode search at the inter frame, the rate-distortion (RD) cost of the tracked MB for the current MB are used and we propose adaptive thresholding algorithms for skipping the intra mode search. For the IPPP sequence type, the overall encoding time can be reduced up to 52% through comparative analysis of experimental results with JM reference software.

Index Terms— Image coding, Prediction methods, Video coding, Video processing, Image motion analysis

1. INTRODUCTION
A variable block size for inter-mode prediction maximizes the coding efficiency based on rate-distortion optimization (RDO) in the H.264/AVC coding standard [1]. The block sizes are 16 × 16, 16 × 8, 8 × 16, 8 × 8, 8 × 4, 4 × 8, and 4 × 4. In addition, intra-mode prediction (nine modes for a 4 × 4 luma block and four modes for 16 × 16 luma and four modes for 8 × 8 chroma blocks) follows inter-mode prediction to determine the best residual image [1]. Recently, nine intra-mode predictions for an 8 × 8 luma block were added for a high profile. As a result, the complexity and computational load increased dramatically for the inter-frame. Therefore, a fast mode decision scheme that can reduce the complexity of the H.264/AVC video encoder is needed.

Lee et al. proposed a selective intra-mode search method based on the transformed coefficients of the residual image and upper-row and left-column pixel values [2]. A directional field based approach has been reported by Pan et al. [3] where several directions are selected by using an edge direction texture histogram according to block types.

Kim et al. [4] suggested a method based on the rate-distortion (RD) costs of the neighborhood MBs for the current MB and proposed adaptive thresholding scheme. For intra prediction of 4 × 4 blocks, Cheng and Chang have presented a Three Step intra prediction algorithm using the correlation between the neighborhood directions [5].

As described above, most previous work was based on an I-Slice, except for Lee’s method [2]. For many fast mode decision schemes, the following rate-distortion (RD) optimization has also been used [2]-[5]:

\[
J_{RD} = SAD_{Mode} + \lambda \cdot \{ R(Header) + R(Residual) \},
\]

where \( J_{RD} \) is a bitrate-distortion value as a cost function, \( SAD_{Mode} \) is the sum of the absolute differences for the given mode, \( \lambda \) denotes the Lagrangian multiplier, \( R(x) \) is a bit amount for coding \( x \), \( Header \) provides header information, and \( Residual \) is the residual data for the given MB.

In this paper, we propose a fast intra-mode SKIP determination method on the basis of the rate-distortion (RD) cost of the tracked MB for the current MB in the reference frame. To do this, we introduce a simple MB tracking scheme.

2. THE PROPOSED FAST INTRA MODE SKIP DETECTION ALGORITHM
Two consecutive frames in a video sequence are highly correlated. With slow object motion in the video, the mode information for the same MB in the previous frame may still affect the mode determination process of the current MB because of the high temporal correlation. In this study, we propose an MB tracking scheme to make use of temporal correlation. Also an adaptive decision rule is suggested based on the proposed MB tracking scheme.

2.1. A Simple MB Tracking Strategy
To track an object is to locate the current object region in an adjacent frame. To do this, we need to define the desired object region in the image plane. We need a search procedure
for the desired object region with a predefined tracking criterion in the temporal domain.

To apply this tracking scheme to block-based video coding, we consider each MB as a desired object in a mode decision procedure. As shown in Fig. 1, a \( P_{16 \times 16} \) block type for the current MB (at time \( t \)) is used to locate the region in the previous frame (at time \( t - 1 \)) with the highest correlation. This is an integer pel motion estimation procedure for the current MB based on a block matching technique.

![Fig. 1](image)

**Fig. 1.** A MB tracking scheme using \( P_{16 \times 16} \) block motion estimation.

Once the best motion vector and most highly correlated region are obtained, we determine the most highly correlated MB of the current MB in the previous frame, as follows:

\[
\text{Max}_{MB(k,l)}\{\text{Area A, Area B, Area C, Area D}\},
\]

where \( (k,l) \) denotes an index of MBs that contain the overlapped (correlated) region of the current MB. We use the above equation to determine that MB which has the maximum correlation with the current MB.

### 2.2. Adaptive Thresholding Scheme Based on MB Tracking

An H.264/AVC encoder performs the intra mode prediction for the intra frame and the inter frames. In fact, there are a few intra mode MBs in one inter frame (or slice). In a full intra mode search, all combinations of intra mode predictions are executed for the best intra mode prediction, as described in the previous section.

Figure 2 shows the occupation ratio of the intra mode MB as the final MB mode according to various \( QP \) values.

![Fig. 2](image)

**Fig. 2.** The occupation of intra mode MBs with various \( QP \) values.

scheme, we use the RD costs of the most correlated MB or image region for the current MB (\( MB_{kl} \)) using MB tracking. Since these tracked MB or image region in the reference frame is the most correlated with the current MB, we may get information that will allow skipping the intra-mode prediction process.

Based on Eq. (1), the following relationship can be satisfied for a given \( MB_{kl} \):

\[
J_{RD_{kl}} = SAD_{Mode}^{kl} + \lambda \cdot \{R(Header) + R(MV) + R(Residual image)\}.
\]

Here, if \( R(Header) \) can be usually negligible, then we can
approximate the Eq. (3) as the following:

$$ J_{RDk_1} \cong \frac{\text{SAD}_{\text{Mode}} + \lambda \cdot R(MV) + \lambda \cdot R(\text{Residual image})}{m \cdot \text{RD cost of the tracked MB}}, $$

$$ = m \cdot \text{cost}_{k_1} + \lambda \cdot R(\text{Residual image}). \quad (4) $$

In the above equation, the most right term is usually very smaller than a value of $m \cdot \text{cost}_{k_1}$. This means the RD cost of the current MB can be approximated as a value of $m \cdot \text{cost}_{k_1}$. Also, the current MB ($M_{B_{k_1}}$) is highly correlated with the tracked MB or image region in the reference picture.

Using these facts, we propose two adaptive thresholding algorithms for skipping intra mode search.

As described in prior, the MB tracking scheme can provide an approximated information for the current MB. Based on this tracking scheme, we suggest the following decision rule for extracting intra mode SKIP:

$$ J_{RDk_1}^{t-1} \geq m \cdot \text{cost}_{k_1}^{\text{Inter Best}}, \quad (5) $$

where $J_{RDk_1}^{t-1}$ is the RD cost of the tracked (most correlated) MB in the reference frame ($t-1$), and $m \cdot \text{cost}_{k_1}^{\text{Inter Best}}$ is defined as the motion cost of Eq. (4) when the best inter mode was determined. Especially, to get the most correlated MB, if the correlation ratio is more than some predefined threshold $\tau$, then we use the RD cost of the tracked MB. We have chosen as $\tau = 0.8 \sim 0.9$ to guarantee enough correlation between two MBs.

If $J_{RDk_1}^{t-1}$ is larger than the value of $m \cdot \text{cost}_{k_1}^{\text{Inter Best}}$, this means the selected inter-best mode is good enough for the current MB. So we can skip the intra mode search procedure. Otherwise, the intra mode search will be followed.

### 2.3. Refinement Stage

We consider a simple refinement stage for extracting additional MBs that can be omitted. In case of P8 × 8 sub-block mode, its RD cost value is typically so large. In this case, it is desirable to select one among intra modes as the best coding mode. However when using the averaged RD cost of P8 × 8 sub-block MBs only, there may be a small problem. As mentioned in the above, this value is usually large. Thus, there may be an excessive number of MBs that are skipped if the motion cost value of the current MB is compared with this threshold value. This may cause undesirable loss of quality and bit increment. Therefore we consider another term for coping with this problem in this refinement stage.

$I4MB$ mode is usually dominant when the intra mode is selected as the best coding mode. We utilize the average RD cost value of $I4MB$ MBs that were pre-encoded for our algorithm. Based on these facts, we can select a averaged threshold value as the following adaptively:

$$ J_{RDk_1}^{TH} = \frac{1}{n+1}(J_{RD}^{TH} + n \cdot J_{RD}^{TH}), \quad (6) $$

where $J_{RD}^{TH}$ is the RD cost value of MB that is encoded as P8 × 8 sub-block or $I4MB$ mode and $n$ is the number of MBs which are encoded as P8 × 8 sub-block or $I4MB$ mode until now (to the current MB). From the above equation, we can update the average RD value of $J_{RD}^{TH}$ easily.

If the value of $J_{RD}^{TH}$ is larger than the value of $m \cdot \text{cost}_{k_1}^{\text{Inter Best}}$, we consider that the selected inter-best mode is good enough for the current MB. So we can skip the intra mode search procedure. Otherwise, we need the intra mode search for a better coding efficiency.

### 3. RESULTS AND DISCUSSIONS

![Fig. 3. The RD curves for (a) the Paris and (b) Mobile sequences.](image)

To verify the performance of the proposed algorithms, various MPEG standard sequences were used with CIF and QCIF sizes. Analyses were performed with encoding frames=200, RD optimization enabled, $QP = 20, 24, 28, 32$, and 36 sequence types of IPPP in the Main profile, using CABAC, with a search range of MV = ±16, and the number of the reference frames = 1. The Hadamard transform option was turned on also.

JM 11.0 reference software of the JVT (joint video team) was used as a reference code for evaluation of the encoding performance. We defined three measures for evaluating the encoding performance, including average $\Delta PSNR$, average $\Delta Bits$, and an encoding-time saving factor, $\Delta T$. The average $\Delta PSNR$ is the difference in (dB) between the average $PSNR$ of the proposed method and the corresponding value of another method. As performance improves, this criterion becomes larger. The average $\Delta Bits$ is the bit-rate difference as a percentage (%) between the compared methods. Lastly, the encoding-time saving factor $\Delta T$ is defined for complexity comparison as:

$$ \Delta T = \frac{T_{ref} - T_{proposed}}{T_{ref}} \times 100\%, \quad (7) $$

3504
Table 1: Performance comparison of the proposed algorithm on the JM 11.0 reference encoder for IPPP sequences.

<table>
<thead>
<tr>
<th>Contents</th>
<th>QF=24</th>
<th>QF=28</th>
<th>QF=32</th>
<th>Average values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ΔPSNR</td>
<td>ΔBits</td>
<td>ΔT</td>
<td>ΔPSNR</td>
</tr>
<tr>
<td>Stefan(CIF)</td>
<td>Lee’s</td>
<td>-0.006</td>
<td>-0.034</td>
<td>41.54</td>
</tr>
<tr>
<td></td>
<td>Kim’s</td>
<td>-0.014</td>
<td>-0.009</td>
<td>38.14</td>
</tr>
<tr>
<td></td>
<td>Proposed</td>
<td>0.001</td>
<td>-0.409</td>
<td>48.40</td>
</tr>
<tr>
<td>Flower(CIF)</td>
<td>Lee’s</td>
<td>-0.009</td>
<td>0.029</td>
<td>55.62</td>
</tr>
<tr>
<td></td>
<td>Kim’s</td>
<td>0.004</td>
<td>0.046</td>
<td>45.50</td>
</tr>
<tr>
<td></td>
<td>Proposed</td>
<td>0.006</td>
<td>-1.39</td>
<td>45.19</td>
</tr>
<tr>
<td>Mobile(CIF)</td>
<td>Lee’s</td>
<td>-0.006</td>
<td>-0.061</td>
<td>32.42</td>
</tr>
<tr>
<td></td>
<td>Kim’s</td>
<td>0.002</td>
<td>-0.080</td>
<td>32.80</td>
</tr>
<tr>
<td></td>
<td>Proposed</td>
<td>0.000</td>
<td>-0.150</td>
<td>32.17</td>
</tr>
<tr>
<td>Paris(CIF)</td>
<td>Lee’s</td>
<td>0.001</td>
<td>-0.082</td>
<td>55.84</td>
</tr>
<tr>
<td></td>
<td>Kim’s</td>
<td>0.006</td>
<td>-0.237</td>
<td>57.79</td>
</tr>
<tr>
<td></td>
<td>Proposed</td>
<td>0.003</td>
<td>-0.435</td>
<td>51.72</td>
</tr>
<tr>
<td>Foreman(CIF)</td>
<td>Lee’s</td>
<td>-0.015</td>
<td>0.297</td>
<td>31.14</td>
</tr>
<tr>
<td></td>
<td>Kim’s</td>
<td>-0.011</td>
<td>0.170</td>
<td>30.57</td>
</tr>
<tr>
<td></td>
<td>Proposed</td>
<td>-0.021</td>
<td>0.407</td>
<td>43.26</td>
</tr>
<tr>
<td>Football(SIF)</td>
<td>Lee’s</td>
<td>0.001</td>
<td>0.009</td>
<td>25.54</td>
</tr>
<tr>
<td></td>
<td>Kim’s</td>
<td>-0.005</td>
<td>0.119</td>
<td>31.93</td>
</tr>
<tr>
<td></td>
<td>Proposed</td>
<td>-0.012</td>
<td>-1.286</td>
<td>37.93</td>
</tr>
</tbody>
</table>

under the condition that the full mode search (FMS) is optimum (reference). As this value increases, the performance speed is increased. Also, it must be noted that positive values for the PSNR and ΔBits indicate increments, and negative values indicate decrements. We used two methods for an objective comparison of the encoding performance. These are Lee’s [2] and Kim’s [4] methods.

Figure 3 illustrates the rate-distortion (RD) curves for two sequences. From these results, we can see that the proposed decision algorithms have a RDO performance similar to the JM 11.0 encoder with the full intra-mode search. We can observe that our method get better quality slightly at the higher bit rate for the Paris sequence.

Table 1 shows the results of all algorithms for the IPPP sequence type only for QF=24, 28, and 32 because of the limitation of space. The proposed algorithm achieves a better speedup factor with a minimal loss of image quality and a minimal bit increment. In sequences with stationary or very slow motion, Lee’s method yields better speed-up factor by the factor of almost 5% than Kim’s and the proposed algorithm. Also, we can see that Kim’s algorithm is superior to Lee’s method for sequences with very fast object’s motion or large global motion (Stefan, Mobile, Football) as they mentioned. The proposed algorithm show similar improvement of the encoding speed for sequences such as the Paris and Foreman. For sequences with very fast object’s motion or large global motion (Stefan, Mobile, Football), we can observe that our algorithm is faster than Kim’s and Lee’s methods by a factor of up to 25% in the average encoding time. For QF=24 of the Mobile sequence, results show that the encoding speed can be improved up to 52.17% with a decrease of 0.150% in the bits. We also observed that the proposed algorithm had similar performance with τ=0.75~0.95.

4. CONCLUSIONS

We have proposed an efficient intra mode SKIP detection algorithm based on the macro-block (MB) tracking scheme and rate-distortion (RD) cost for inter frames in H.264/AVC video coding. Through comparative analysis, a speed-up factor of 45%~52% was verified with a negligible bit increment and a minimal loss of image quality.

5. REFERENCES


