LOW-COMPLEXITY INTRA/INTER MODE-DECISION FOR H.264/AVC VIDEO CODER

Yi J. Liang and Khaled El-Maleh

Qualcomm CDMA Technologies, Qualcomm Inc.
5775 Morehouse Drive, San Diego, CA 92121, USA
{yliang, kmaleh}@qualcomm.com

ABSTRACT

In this work we present a novel mode decision scheme with which a portion of the macroblocks' Intra/Inter coding mode is determined without performing the exhaustive spatial estimation and motion estimation searches. The coding modes of such macroblocks are determined utilizing the correlation across spatially and temporally close macroblocks' prediction modes. By bypassing a large amount of spatial estimation and motion estimation searches, the scheme can reduce the computational complexity substantially. Simulation results show that the coding performance using the proposed mode decision scheme is perceptually as good as using traditional mode decision. The proposed scheme can be applied to any modern video codec that employs predictive coding, including H.263, MPEG-4, and H.264/AVC.

1. INTRODUCTION

Most modern video codecs employ motion-compensated Inter-frame prediction to exploit the temporal redundancy across video frames to achieve high compression. More advanced codecs, such as H.264/AVC, also employ spatial estimation for Intra-frame prediction to reduce the spatial redundancy within an Intra-coded frame [1], [2]. In such codecs, motion estimation (ME) and/or spatial estimation (SE) count for the majority of the total encoding time1. Predictive coding involving ME and SE is very computationally intensive in modern codecs because they employ exhaustive search through multiple coding modes before an optimal mode is selected that achieves the best compression efficiency. In H.264, for instance, nine coding modes for Intra 4 x 4 block type, and four modes for Intra 16 x 16 block type are supported in Intra-frame prediction; four coding types, including block sizes of 16 x 16, 16 x 8, 8 x 16, 8 x 8, are supported in Inter-frame prediction, among which the 8 x 8 block can be further partitioned [1].

In coding an I-frame, the best Intra-frame prediction mode has to be selected for each macroblock (MB) in H.264. In coding a P-frame, the best Intra-frame or Inter-frame coding mode has to be determined for each MB. Current mode decision (MD) algorithms rely on the measured distortion such as the sum-of-absolute-difference (SAD), as well as the number of bits spent on coding motion vectors and residuals, in order to find the coding mode that yields minimum cost (or rate-distortion optimized [3], [4]). This is performed after the exhaustive search of every coding mode allowed. We refer to this decision as the post-ME/SE mode decision as it is performed after the ME and/or SE searches.

The conventional post-ME/SE mode decision achieves high compression by selecting the optimal coding mode. However, it requires intensive computation in coding each MB, especially for P-frames. The computation intensity, i.e. hardware power consumption or DSP MIPS, pose great challenges for encoders that have real-time requirement as well as those built in light devices, such as those in PDAs, mobile phones, video telephony applications and etc. In this work, we propose a mode decision scheme that pre-determines the coding modes for a portion of the MBs without performing the exhaustive SE and ME searches, by utilizing the correlation across spatially and temporally close MBs' prediction modes. By bypassing a large amount of SE and ME searches, the scheme can reduce the computational load as well as hardware power significantly, without perceptible degradation of the coding quality. The proposed scheme also applies to earlier video codecs such as MPEG-4 and H.263, where ME and/or SE are employed.

This paper is structured as follows. In section 2, we describe the conventional mode decision that most codecs use today. In section 3, we present our proposed scheme of mode decision that reduces the computational complexity. We present the experimental results on computation savings and coding performance in Section 4.

2. CODING MODE DECISION

During predictive coding of a particular MB, SE and/or ME are performed for every coding mode allowed. In
H.264, nine coding modes (vertical prediction, horizontal prediction, DC prediction, diagonal-down-left, diagonal-down-right, vertical-right, horizontal-down, vertical-left, and horizontal-up predictions) for Intra Luma 4 × 4 block type, and four modes (vertical prediction, horizontal prediction, DC prediction, and plane prediction) for Intra Luma 16 × 16 block type may be used in Intra-frame prediction, and thereafter the best sub-modes are selected and the lowest SAD are obtained after prediction, for Luma 16 × 16 and Luma 4 × 4 respectively.

In Inter-frame prediction, four coding types associated with block sizes of 16 × 16, 16 × 8, 8 × 16 and 8 × 8 may be used, and thereafter the best motion search as well as the lowest SAD can be found through motion estimation. For the best Intra-coding modes and the best Inter-coding search result, a more complex cost that includes SAD and bit usage, is obtained, and a final winner is selected that achieves the overall minimum cost.

In coding every MB, this post-ME/SE mode decision process searches through all the possible coding modes exhaustively, and a selection is made after all predictions. The best mode determined in this way is optimal in the rate-distortion sense, however, the process is very complex in computation. In the next section, we will describe how to skip some of the exhaustive post-ME/SE mode decision process and reduce the complexity.

3. PROPOSED HYBRID MODE DECISION

3.1. Correlation between Coding Modes

Natural video sequences exhibit strong correlation within a particular frame (spatially) and between successive frames (temporally), which is the underlying basis for spatial estimation and motion estimation based prediction. It is found that the best prediction mode determined by the above mentioned mode decision is also highly correlated; due to the correlation of the video content itself. In other words, a MB tends to select a coding mode that is the same as its neighboring MBs in the same frame, or the co-located MB in the prior frame. For example, a relatively flat area in a frame may result in a block of MBs using the same Intra mode, which may even spread across multiple frames if the motion is low. Figure 1 illustrates an example of highly correlated coding modes across neighboring MBs.

3.2. Hybrid Mode Decision

Based on the correlation of the prediction modes, we propose the mode decision scheme that does not depend on any up-front SE or ME, and apply it to a portion of selected MBs. This scheme is referred to as the pre-ME/SE mode decision, since the decision is made before any SE or ME is performed. The coding mode is selected according to the modes of its neighboring MBs that are spatially and/or temporally close. Since this approach bypasses the unnecessary and exhaustive up-front SE and ME-based prediction searches, it reduces the computational complexity.

In practice, to prevent inaccurate mode decisions from propagating, pre-ME/SE mode decision is used in combination with post-ME/SE mode decision, and this scheme is hence referred to as the hybrid mode decision. Some of the MBs are applied with the post-ME/SE decision for high coding performance, while for the others the coding modes are pre-determined to reduce complexity.

The pre-ME/SE mode decision may be performed at different levels with various trade-offs:

1) Level I: The pre-ME/SE mode decision only determines whether a MB should be either Intra or Inter coded, and then leave the submode decision to further searches. A later search will examine thirteen submodes for Intra, or four for Inter;

2) Level II: The pre-ME/SE mode decision first determines one of the following modes: Intra 4 × 4, Intra 16 × 16, Inter 16 × 16, Inter 16 × 8, Inter 8 × 16 and Inter 8 × 8. Then further searches are performed over the nine submodes if Intra 4 × 4 is selected, or the four submodes if Intra 16 × 16 is selected, or for the best motion estimation for the Inter-frame prediction;

3) Level III: The pre-ME/SE mode decision first determines a particular Intra submodes, such as Intra 4 × 4 - vertical prediction, or a particular Inter submodes. Further SE or ME only has to be performed for one particular mode that is already pre-determined.

From Level I to III, the mode determined by pre-ME/SE decision is more specific, and the computation load for a later SE and/or ME search is reduced. However, as the mode determined becomes more specific, the correlation between neighboring MB modes decreases, so that the accuracy of pre-ME/SE mode decision decreases, resulting in degradation in coding performance. In this work, we implement the
4. EXPERIMENTAL RESULTS

In this section, we present experimental results comparing the performance of the proposed hybrid (pre/post) mode decision with the conventional post-ME/SE mode decision. We are interested in how much savings in computation we are able to obtain using the proposed method, as well as the coding performance.

We have tested encoding four QCIF sequences including Foreman, Mother-Daughter, Car-Phone, Coast-Guard, and a CIF sequence, Table-Tennis, using H.264 test model JM 7.3. For each sequence we encode at 15 fps for the first 140 frames, with the first frame being an I-frame and the rest P-frames.

Table 1 shows the result of computation time savings in prediction, applying different percentages of pre-decision compared with the conventional post-decision. With half of the MBs applied with pre-ME/SE mode decision, the time savings is at least 46% for different sequences. When we increase the percentage of pre-ME/SE mode decision to 67% or even 75%, the time savings can reach 61% and 69% respectively.

<table>
<thead>
<tr>
<th>MBs applied with pre-decision</th>
<th>50%</th>
<th>67%</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreman</td>
<td>46</td>
<td>61</td>
<td>69</td>
</tr>
<tr>
<td>Mother-Daughter</td>
<td>47</td>
<td>61</td>
<td>69</td>
</tr>
<tr>
<td>Car-Phone</td>
<td>47</td>
<td>61</td>
<td>69</td>
</tr>
<tr>
<td>Coast-Guard</td>
<td>46</td>
<td>61</td>
<td>69</td>
</tr>
<tr>
<td>Table-Tennis</td>
<td>46</td>
<td>61</td>
<td>69</td>
</tr>
</tbody>
</table>

Table 1. Time savings in prediction (including all SE and ME) using the proposed pre-ME/SE mode decision compared with the conventional post-ME/SE mode decision.

Figures 3 - 5 show the coding performance using the proposed hybrid mode decision with different pre-decision percentages versus the conventional post-ME/SE mode decision. The sequence-averaged peak signal-to-noise ratio (PSNR) of the luminance signal is plotted versus the average data rate. It is observed that for various sequences, the performance degradation is below 0.5dB even at high bit rate, when half of the MBs are applied with pre-decision. This difference is not considered as visually perceptible. In practice, considering the substantial savings in computation, the proposed half-MB pre-decision performs with good satisfaction.

To study the effect of more aggressive pre-ME/SE mode decision, we also increase the percentage of MBs applied to 67% and 75% respectively. The gap in coding performance increases at higher data rate. At a bit rate of lower than 64Kbps, however, the degradation is below 1dB even when 75% of the MBs are applied with pre-ME/SE mode.
decision. This indicates that schemes using more aggressive pre-ME/SE mode decisions are still useful for applications that encode at a low bit rate, but has stringent real-time requirement that desire even lower computational complexity, such as video telephony.

5. CONCLUSIONS

We have presented a novel hybrid mode decision scheme that combines both the conventional post-ME/SE mode decision and a pre-ME/SE mode decision. Pre-ME/SE mode decision is applied onto a portion of the MBs without performing the exhaustive SE and ME searches, by utilizing the correlation across spatially and temporally close MBs’ prediction modes. By bypassing a large amount of SE and ME searches, the scheme can reduce the computational complexity as well as hardware power consumption substantially. Simulation results show that with half of the MBs applied with pre-ME/SE mode decision, the computational savings are up to 46% using the hybrid mode decision scheme, while the difference in coding performance is within 0.5 dB compared with a traditional post-ME/SE mode decision scheme. The percentage of the MBs applied with pre-decisions can be parameterized to trade off between coding performance and computational complexity reduction, and the scheme can be tailored to different applications with various complexity and quality requirements. The proposed scheme can be applied to any modern codec that employs predictive coding, including H.263, MPEG-4, and H.264.

6. REFERENCES


