Homogeneity and Distortion-Based Intra Mode Decision Architecture for H.264/AVC

Guilherme Corrêa, Cláudio Diniz and Sergio Bampi
Informatics Institute (INF) – Federal University of Rio Grande do Sul, Porto Alegre, Brazil
grcorreia, cmdiniz, bampi}@inf.ufrgs.br

Daniel Palomino, Roger Porto and Luciano Agostini
Federal University of Pelotas
Pelotas, Brazil
danielp.ifm, rogerecp, agostini}@ufpel.edu.br

Abstract—In order to achieve the best coding performance, the H.264/AVC encoder must choose the best coding mode and the best block size in terms of bit-rate and distortion. The H.264/AVC reference software applies the Rate-Distortion Optimization (RDO) technique, which makes the encoding process a complex task in applications which require real-time operation. This paper presents a fast intra decision process and its architecture, where the mode and the block size decisions are performed based on the block distortion and homogeneity. The performed tests show that the method achieves PSNR results similar to the RDO technique and a low bit-rate increase. On the other hand, the gains in terms of complexity are near to 148 times when compared to RDO method. Also, the implemented architecture is capable of processing 1080p videos in real time.

I. INTRODUCTION

Intra-frame prediction is an important coding feature included in H.264/AVC [1], the latest video coding standard, to decrease the spatial redundancy present in each video frame. With two partition types (4x4 or 16x16) and 13 prediction modes, it provides a meaningful improvement in terms of video quality and coding efficiency for intra frames, e.g. an H.264/AVC encoder configured in intra-only mode is able to overcome the JPEG2000 for static images. The adjustment of quality and coding enhancement for intra prediction is controlled by the mode decision, which selects the coding mode that better improves output video quality for a fixed bit-rate or reduces the number of bits for a fixed video quality. The best coding performance is usually achieved with Lagrangian rate-distortion optimization (RDO), but it results in a very high computational complexity [2].

Some works [7]-[8] have proposed hardware architectures to improve intra prediction and mode decision performance for high resolution video coding. The work in [7] encodes all the intra prediction modes and do not propose any novel algorithm for mode decision. The work in [8] proposes a three-step algorithm for 4x4 mode selection but does not detail how the partition size is determined. This selection is very important to reduce the number of coded bits for homogeneous regions.

In this paper, we propose an algorithm and hardware architecture for the intra mode decision of H.264/AVC targeting high resolution video encoding. Prediction modes in the same partition type are selected using a distortion metric. Homogeneity-based mode decision is applied in the original video to determine which partition is best suited for each macroblock (MB).

Paper organization: Section II presents the intra prediction and mode decision concepts. Sections III and IV introduces the proposed mode decision and hardware architecture. Section V shows the heuristics and architecture results. Section VI concludes this work and indicates some future work.

II. INTRA PREDICTION AND MODE DECISION

The Intra Prediction module exploits the spatial redundancy in a frame, exploring similarities among neighbor blocks. The blocks can be predicted by the Intra Prediction Module in three different ways: (1) 4M (prediction over 4x4 luminance blocks), (2) 16MB (prediction over 16x16 luminance blocks) and chroma 8x8 (prediction over 8x8 chrominance blocks). The intra prediction process is based on copying neighbor pixels to form one block. There are nine different prediction
modes for 4x4 blocks and four prediction modes for 16x16 blocks, which are the same for chroma 8x8 blocks. The 16x16 block size is generally used in homogeneous areas, where the pixel values present little difference. On the other hand, the 4x4 block size is usually used to heterogeneous areas, i.e., highly detailed areas. After the computation of all these modes, the mode decision is responsible to choose which block size is used. This decision is extremely important, since it will affect directly the image quality and the bit-rate of the coded video.

In order to take full advantage of these modes and to select the best of the available modes, the H.264/AVC encoder selects the best mode using the rate distortion optimization (RDO) technique. RDO examines exhaustively all the modes for each 116MB and 14MB macroblocks and chooses the one which produces the minimum rate-distortion cost. The rate-distortion calculation is shown in equation (1), where $D$ is the distortion level of the coded image, $R$ is the bit-rate and $J$ is the rate-distortion cost. The $\lambda$ parameter is called Lagrangian factor and it relates distortion with bit-rate.

$$J = D + \lambda \cdot R$$  

(1)

This calculation is performed for each macroblock possible mode. It means that the complexity of the intra-mode decision is extremely high, since the transforms, quantization and entropy coding must be performed for every possible mode.

III. PROPOSED MODE DECISION

In this work we have divided the intra-frame mode decision aiming at the simplification and modularization of the whole process for a future hardware implementation. The steps are described in the following items:

- Decision of the best 4x4 mode for each luminance block;
- Decision of the best 16x16 mode for the luminance MB;
- Decision of the best 8x8 mode for the chrominance MB;
- Block-size decision (I4MB or I16MB) for the MB.

The next two sections explain the premises and heuristics used in this work to perform the fast intra-frame decision.

A. Distortion-Based Mode Decision

A set of simulations using video sequences which are typically used for tests has shown that the difference between the RDO-based mode decision and a distortion-based mode decision is very small in terms of bit-rate and image quality when the dimensions of the block is fixed (i.e., when only 4x4 or only 16x16 blocks are used). This happens because the mode which results in minimal residual information (difference between the original block and the intra-predicted block) tends to generate fewer encoded information, once small residual values are generally transformed and quantized to zero. The decision of the best intra 4x4 mode and the best intra 16x16 mode can be thus performed taking into consideration only the distortion of the coded image.

The reference software computes the distortion between the reference and the reconstructed frame only after the whole encoding and decoding process is complete. As the decision is performed as a comparison between all possible modes, this process must be performed once for each mode and for each block. To minimize this problem, in this work the distortion is calculated between the original block and the predicted block (i.e., the output of the intra-frame prediction).

The tests were performed considering three different distortion metrics: Sum of Absolute Differences (SAD), Sum of Absolute Transformed Differences (SATD) and Sum of Squared Difference (SSD) [2]. While SAD is the simplest metric, SSD is the most complex one, since it uses a multiplier to obtain the square of the difference between the blocks. On the other hand, SAD generates the worst results in terms of encoded image quality. It is generated through simple subtractions between every co-located pixel of the two blocks and a final summation of each subtraction previously performed. The SATD performs the same subtractions of the SAD metric, followed by an integer adaptation of the Direct 2D 4x4 Hadamard Transform, defined by a set of 64 sums and subtractions proposed in [10]. The results of the transform are all summed together in order to generate a final SATD value. The SSD metric is also similar to SAD. However, the result of every subtraction is multiplied by itself (i.e., squared) before being summed.

This approach returned results which are comparable to the ones provided by the RDO technique with fixed block size. The PSNR loss of the proposed distortion-based mode decision varies between 0.09 dB and 0.16 dB, which is not relevant, when compared to the RDO technique. The bit-rate has increased 5.03% in the worst case, which is not significant when the enormous complexity decrease is taken into account.

B. Homogeneity-Based Block Size Decision

The higher the detailing (i.e., heterogeneity) of an image region, the higher tends to be the variability of the residual information generated after the intra-frame prediction. Applying a two dimensional transform over the residual data allow an evaluation of the homogeneity in the frequency domain. Heterogeneous areas generate high frequency components after transformation, increasing the amount of bits used to encode the video. Besides, higher values in the transformed blocks lead to higher losses during the quantization step. This way, smaller blocks are more suitable for video sequences which contain detailed image areas, since they allow the identification of local homogeneity in smaller regions. In this work, we propose a homogeneity-based block size decision which is performed from the original macroblock data. This process is not dependent on the distortion-based decision and can be performed at the same time, if necessary.

H.264/AVC uses the Hadamard Transform and the Discrete Cosine Transform (DCT) which concentrate the highest coefficient values at the low-frequency positions of the transformed block. This way, we propose a heuristic based on the calculation of the 16x16 DCT coefficients to identify the homogeneity of the macroblock. As the H.264/AVC standard uses only the 4x4 DCT transform, a 16x16 version of this transform was developed in this work in order to obtain the equations for the coefficients which are necessary to the block size decision.
The DCT transform is given in (2) and (3), where A is the transform matrix, X is the input data, Y is the transformed block, N is the number of input elements and i,j represents the position of the element inside the block [9].

\[ Y = AXA^T \]  \hspace{1cm} (2)

\[ A_y = C_i \cos \left( \frac{(2j + 1)i \pi}{2N} \right) \text{, where} \]
\[ C_i = \begin{cases} \frac{1}{\sqrt{N}} & \text{if } i = 0 \\ \frac{2}{\sqrt{N}} & \text{if } i > 0 \end{cases} \]  \hspace{1cm} (3)

After the possible simplifications and reductions to the first quadrant of the trigonometric circle, the 256 generated coefficients of matrix A are reduced to only 16. Also, in order to allow the implementation of the transform as a hardware architecture, the DCT was factorized and an integer approximation of it was created, just as the 4x4 DCT transform of the standard. Equation (4) shows the resulting DCT function, where C and C^T represent the factorized transform matrix and its transform, X represents the input values and E is the factor matrix generated after the factorization [9].

\[ Y = (C X C^T) \otimes E \]  \hspace{1cm} (4)

CXC^T is called the kernel of the transform. Table I shows the transform matrix C obtained after the factorization. It is an integer approximation of the original DCT, composed by 8 different values.

<table>
<thead>
<tr>
<th>i</th>
<th>j</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Despite the performed simplifications, the number of arithmetic operations involved in the calculation of one coefficient is still large. However, the homogeneity analysis proposed in this work is performed only over a sample of the transformed coefficients. The used sample is composed by the first line and the first column of coefficients, as proposed by Lee [2]. A set of simulations was also performed considering other groups of samples, such as the superior diagonal of the macroblock, but no significant gains were obtained. The homogeneity calculation is presented in (5), where \( AC_{\text{energy}} \) represents the level of homogeneity of the original block and Y is the transformed block. The lower the value of \( AC_{\text{energy}} \), the higher is the homogeneity of the original macroblock.

A set of simulations was performed in order to obtain the \( AC_{\text{energy}} \) values for macroblocks of different video sequences. The simulations allowed the identification of a threshold value that could be used to decide the block size. The best results were achieved when the threshold value was set to 2800. In this case, the number of decisions similar to the ones performed by the RDO technique was the highest, leading to the best results.

\[ AC_{\text{energy}} = \sum_{i=1}^{15} Y(u,0) \left| \sum_{i=1}^{15} Y(0,v) \right| \]  \hspace{1cm} (5)

IV. DESIGNED ARCHITECTURE

The designed architecture implements the whole decision for intra macroblocks, including both the best mode and block size decisions. The decision of the best 4x4 and 16x16 mode is performed using SAD. The best chrominance mode is chosen based on the SAD calculation for the 16x16 modes. Fig. 2 shows the whole intra decision architecture. The SAD 16x16 decision block is composed by 4 SAD units and the SAD 4x4 decision block is composed by 9 SAD units. Also, each one of these modules contains a comparator which chooses the block with the smaller SAD value.

For the block size decision, the top and the left border coefficients generated after the DCT transform calculation are used. The 16x16 DCT transform architecture was designed focusing on the samples that are used in the homogeneity-based decision described in Section III. It means that the architecture does not perform the whole 16x16 DCT transform, but only calculates coefficients which are necessary. The operators were grouped by patterns in order to reuse them, decreasing the hardware consumption. Fig. 3 shows the block diagram of the DCT-based decision architecture.

The transform calculations are performed by the SM (Sum Module) and the COEFFICIENTS SUM modules. As the DCT transform is an integer approximation of the real DCT, all the calculations in these modules could be solved using only adders and shifters, making the architecture simpler. Besides, each SM module was designed following one different pattern, i.e., the similar operations among the input samples were grouped in one MS module, while the COEFFICIENTS SUM module perform operations which are similar for all coefficients. The control unit is responsible to synchronize the input samples and send it to the right SM module.

The architecture takes 32 clock cycles to delivery all the 24 coefficients, 12 at each 16 clock cycles. The control unit synchronizes these coefficients in the output buffer. After that, all these coefficients are summed to generate the value that is
compared to the threshold in order to decide which intra block size (4x4 or 16x16) will be used.

![DCT-based decision diagram](image)

Figure 3. Partial 16x16 DCT transform architecture

V. RESULTS

The two heuristics proposed in this work were combined in order to perform a complete and consistent intra mode decision. Several tests were performed aiming at the evaluation of the method. Table II shows a comparison between the proposed method and the RDO-based decision. The first line of results from Table II shows the results of the complete RDO-based intra mode decision and the second line shows the results of the proposed method.

The results show a little degradation of 0.28 dB in the average video quality (PSNR loss) and an increase in the bitrate of 12%. These losses are acceptable in comparison with the expressive gains in terms of complexity reduction that the proposed method can offer. The RDO-based intra mode decision takes 148 iterations of the whole encoding process to decide the best mode, since all the encoder operations must be performed over all residual blocks considering all the possible modes. With the method presented in this paper, these operations must be performed only for one block, in one encoding iteration. This way, it is possible to conclude that an acceleration of circa 148 times is obtained through the proposed heuristic. A little additional operations are included to support the early decision.

The designed architecture was described in VHDL and synthesized for the TSMC 0.18 μm technology using the Leonardo Spectrum tool. Table III shows the synthesis results for the three modules which compose the complete decision process for intra frames.

The three modules presented in Table III can operate in parallel, since they do not dependent on one another. This way, the critical delay is defined by the best mode decision module (4.11 ns, which means a frequency of 243 MHz), while the block size decision is the module which takes more cycles to complete the decision (32 cycles) for a macroblock. For real time applications using 1080p HD videos (composed by 8100 macroblocks) the decision process takes 259,200 clock cycles per frame, totaling 7,776,000 clock cycles per second. This means 7.7 MHz, which is 31 times lower than the obtained frequency and let us conclude that our architecture is capable of processing 930 frames per second.

<table>
<thead>
<tr>
<th>Block size / Best mode decision</th>
<th>PSNR (dB)</th>
<th>Bit-rate</th>
<th>PSNR loss (dB)</th>
<th>Bit-rate increase (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDO / RDO</td>
<td>40.56</td>
<td>6,981,128</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Partial DCT / SAD</td>
<td>40.39</td>
<td>7,887,157</td>
<td>0.28</td>
<td>12.03</td>
</tr>
</tbody>
</table>

VI. CONCLUSIONS

This paper has presented a homogeneity and distortion-based intra-mode decision architecture for H.264/AVC. The architecture was described in VHDL and synthesized for TSMC 0.18μm standard cell technology. The results show a small loss in PSNR and a small increase in the bit-rate when compared with RDO technique. However, the homogeneity and distortion-based decision proposed presents a computational complexity about 148 times lower than the RDO technique, since only SAD and 16x16 DCT calculations are performed to make the best mode decision, instead of a loop over the whole encoding and frame recreation process.

As future works we plan to design a whole mode-decision architecture, which will be able to choose among all intra and inter-frame modes.

REFERENCES