

A Fast Mode Decision Algorithm and Its Hardware Design for H.264/AVC Intra Prediction

Wei Wang¹, Yuting Xie, Tao Lin, and Jie Hu

¹ College of Electronics Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
860290813@qq.com

Abstract. This paper presents an architecture of mode decision algorithm for H.264/AVC intra prediction. In the algorithm design, based on the inherent correlation existing in the spatial prediction modes, a significant computational savings can be achieved. In the hardware design, through efficient sharing of configurable units and parallel executions of different candidate prediction modes, a lower hardware utilization and higher execution speed can be achieved. Synthesis results show that the proposed architecture can process HDTV (1920×1080) video at 60 fps in FPGA platform and maximum frequency achieved is 184.8 MHz.

Keywords: H.264/AVC, intra prediction, spatial correlation, hardware verification and FPGA.

1 Introduction

With the wide use of technologies such as online video, digital television and video conference, video compression technology has become an inevitable component for storage and transmission. Currently, H.264/AVC is published as the new generate video compression coding standard which achieves the highest compression performance without sacrificing the quality of picture [1].

The high video compression efficiency of the H.264/AVC standard is achieved through a particular combination of a number of encoding tools but rather any single feature. Intra prediction algorithm is the important part of the encoder which occupies the most calculation and generates a prediction for a macroblock based on the spatial redundancy. However, the better compression efficiency comes with a high computation complexity which makes it different from meeting the demand of the real time [2].

The intra prediction algorithms with lower complexity and high-speed computing is important for the real-time requirements. Huang et al. [3] proposed an algorithm with better efficiency on mode decision, but it suffered from a higher computational complexity in determining the partitioning. Yu et al. [4] put forward a fast algorithm in determining the partitioning to reduce the high computational complexity, but their argument shows the insufficiency of efficiency on mode decision.

An efficient architecture for intra prediction is proposed based on the inherent correlation existing in the intra prediction modes which can reduce the candidate

prediction modes of luma blocks to less than half. With the hardware design, two metrics need to be better balanced for a higher throughput which represents the clock cycles and the critical path.

The essay is organized as follows: the implemented fast mode decision algorithm is proposed in section 2; the intra prediction hardware architecture is explained in Section 3; the synthesis results are presented in Section 4; and Section 5 presents the conclusion.

2 Implemented Fast Mode Decision Algorithm of Intra Prediction

This algorithm gives a fast way of calculating the best prediction mode of intra prediction, and which is proposed based on the inherent correlation between the intra prediction modes and blocks. In this way, the candidate prediction modes can reduce to less than half.

2.1 The Fast Decision of Partitioning in H.264 Intra Prediction

As mentioned above, the 4×4 intra prediction is preferred in the macroblock which has high texture complexity, otherwise, the 16×16 intra prediction is been chosen. Therefore, we can determine the partition by using the calculated texture complexity of a macroblock before intra prediction coding. The texture complexity of luma blocks on the directions of vertical, horizontal and oblique 135° are defined in the following way:

$$TC_{vertical} = \frac{1}{m(m-1)} \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} |P_{i,j+1} - P_{i,j}| \cdot \quad (1)$$

$$TC_{horizontal} = \frac{1}{m(m-1)} \sum_{j=0}^{m-1} \sum_{i=0}^{m-1} |P_{i+1,j} - P_{i,j}| \cdot \quad (2)$$

$$TC_{angle} = \frac{1}{(m-1)(m-1)} \sum_{j=0}^{m-1} \sum_{i=0}^m |P_{i-1,j+1} - P_{i,j}| \cdot \quad (3)$$

where m is the size of luma block, $P_{i,j}$ is the original pixel values located in the i row and j column of the luma blocks.

The mean value of the texture complexity is computed as:

$$M_{TC} = \frac{1}{3} (TC_{vertical} + TC_{horizontal} + TC_{angle}) \cdot \quad (4)$$

The threshold of texture complexity is computed as:

$$TC_{th} = \frac{1}{M_{TC}} (|TC_{vertical} - M_{TC}| + |TC_{horizontal} - M_{TC}| + |TC_{angle} - M_{TC}|) \cdot \quad (5)$$

A MB has high texture complexity while $TC_{th} > 1$, and I4MB is selected as the prediction block size; otherwise, I16MB is chosen.

2.2 Mode Decision for Chroma 8x8 Block Intra Prediction

In the intra prediction mode decision algorithm, the conventional calculation times of Rate-Distortion-Optimization is $4 \times (9 \times 16 + 4) = 592$, the prediction mode of chroma block is used as the large exterior loop. By reducing the number of chroma candidate prediction modes, a significant computational saving can be achieved.

The texture complexity of chroma blocks on the main prediction directions are calculated by above equations (1)~(5). And m is the size of chroma blocks. Based on the texture complexity calculation of the 8x8 chroma blocks, the main candidate prediction modes can be determined as shown in table 1.

Table 1. Mode decision for Chroma 8x8 blocks

The Threshold of TC_{th}	The minimum value	Candidate Modes of Chroma 8x8 Block
$TC_{th} > 1$	TC vertical	mode 2 (Vertical)
	TC horizontal	mode 1 (Horizontal)
	TC angle	mode 3 (Plane)
	TC vertical	mode 2 ,0 (DC)
$TC_{th} < 1$	TC horizontal	mode 1 ,0 (DC)
	TC angle	mode 3 ,0 (DC)

2.3 Mode Decision for Luma 4x4 Block Intra Prediction

In the mode decision of intra prediction, a macroblock can be divided into one 16x16 block or sixteen 4x4 blocks or four 8x8 blocks, which are used for executing the mode decision. Therefore, the mode decision of 8x8 blocks are correlated with the 4x4 blocks, and the main candidate prediction modes of the 4x4 luma blocks can be determined depending on the prediction mode of 8x8 block as described in table 2.

Table 2. Intra prediction mode decision of luma 4x4 blocks

prediction mode of Chroma 8x8 block	Candidate modes of luma 4x4 macroblock
Mode 0	Main candidate modes 0, 1, 2, 4
Mode 1	Main candidate modes 1, 2, 6, 8
Mode 2	Main candidate modes 0, 2, 5, 7
Mode 3	Main candidate modes 0, 1, 2, 3

2.4 Mode Decision for Luma 16x16 Blocks

Due to the similarity of the prediction modes between luma 16x16 blocks and chroma 8x8 blocks, the 16x16 intra prediction mode decision are shown in table 3.

Table 3. Intra prediction mode decision of luma 16×16 blocks

prediction mode of Chroma 8×8 blocks	Main candidate modes of luma 16×16 block
Mode 0	Mode 2 (DC)
Mode 1	Mode 1 (Horizontal) + Mode 2 (DC)
Mode 2	Mode 0 (Vertical) + Mode 2 (DC)
Mode 3	Mode 3 (Plane) + Mode 2 (DC)

In order to cut down the number of the intra prediction modes and reduce high calculation burden, we choose the Simple Sum of Absolute Difference (SAD) calculation as the rate control model, which is used to determine the best intra prediction mode according to the principle of the minimum SAD value before coding.

$$SAD = \sum_{(x,y) \in MB_k} |original(x,y) - predict(x,y)| \quad (6)$$

While the position (x,y) represents the location of the luma pixels in the macro-block or sub-block, original(x,y) represents the original pixel value, predict(x,y) represents the prediction pixel value.

2.5 The Comparison between the New Algorithm and Original Algorithm of Intra Prediction

Combine with the original algorithms, a new mode decision algorithm for H.264 intra prediction is proposed. The computational complexity comparison of the two algorithms is shown below as table 4 gives.

Table 4. The computational complexity comparison of two algorithms

Modes	Chroma 8×8 intra prediction	luma 4×4 intra prediction	luma 16×16 intra prediction
New	1 or 2	2	1
Original	4	9	4

According to the above analysis, it is evident that the new mode decision algorithm can reduce the calculation times to $1 \times (16 \times 2) = 32$, $1 \times (16 \times 2 + 1) = 33$, $1 \times 1 = 1$, $2 \times (16 \times 2) = 64$, $2 \times (16 \times 2 + 1) = 66$ or $2 \times 1 = 2$. A significant operation efficiency improving can be achieved.

3 FPGA Design of H.264 Intra Prediction Hardware Architecture

From the above analysis, the hardware architecture of the proposed algorithm can be designed as shown in figure 1.

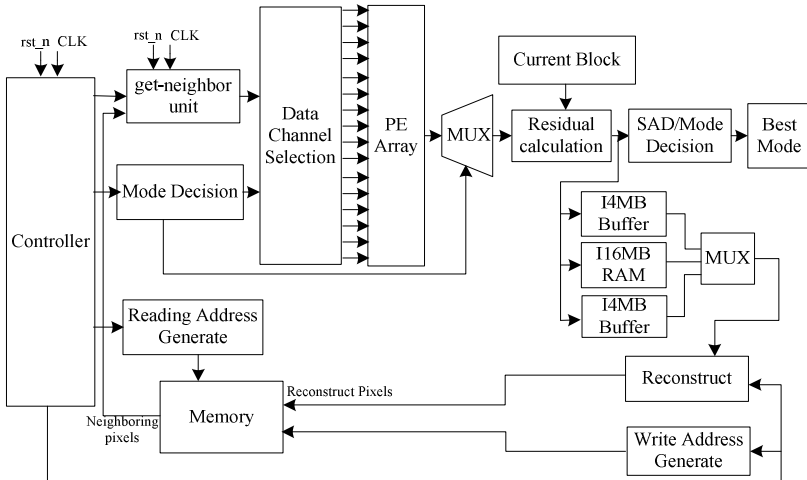


Fig. 1. Intra prediction hardware architecture

In the hardware design, several prediction modes are processed in parallel to generate the predict pixels for SAD calculation which is used for mode decision of intra prediction. After residual data generation and SAD calculation, the best prediction mode can be chosen, and the corresponding residual pixel value is input to the reconstruction loop for image reconstruction.

In summary, the architecture is composed by several modules: prediction generator, system control module, residual calculation module, SAD calculation module, SAD comparator module and mode decision module.

3.1 System Control Module

With the system control module, the best prediction mode and the partitioning of the current block can be determined. In addition, the scan sequence of the macroblock, the prediction sequence of the sub-blocks and the generation of reconstruction pixels are also generated by the system controller.

In the finite-state machine (FSM) design, the candidate prediction modes of intra prediction are specified in a certain order depending on the texture complexity value. When the best prediction mode of chroma 8×8 block is mode1, the FSM of luma 4×4 candidate prediction modes are described in figure 2.

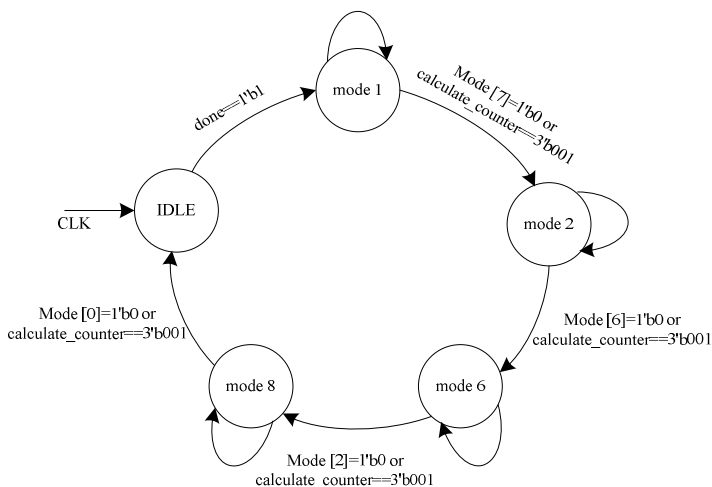


Fig. 2. State transition diagram of Control Unit

3.2 Predicted Generator Module

According to the fundamentals of intra prediction, it is proved that there are many common calculation parameters for different prediction modes, and many calculation formulas can be achieved in a same configurable calculation unit. In conclusion, we can obtain a configurable architecture for prediction generator module, and the processing element (PE) array is designed to generate concurrently 16 prediction pixels as shown in figure3.

(1) Parallel processing unit

Considering the clock frequency and the effective use of middle result registers, we proposed a parallel configurable and pipelining processing units to achieve the prediction calculation. Each PE generates one prediction pixel by selecting the right required reference pixel using multiplexers, and selects the right signal by a special logically controlled by FSM. However, the circuit will become a large scale, and requires a large capacity memory to match in the subsequent processing.

In the configurable architecture, each prediction element is composed of 3 components: the sum operation of the reference pixels, round value and shift value.

(2) Reconstructed neighbor samples memories and prediction memories

In intra prediction, the prediction processing for the current macroblock is existed when the reconstructed samples belonging to the neighbor blocks are available. For the real time processing of 60 fps of HD 1080p resolution, 486000 macroblocks should be processed per second, which result in high external memory bandwidth. This is why this work of research proposes a scheme, that a line of pixels in a frame is stored into the FPGA internal memory (BRAM).

For the luma 16×16 and Chroma 8×8 prediction mode decision is not executed in 4×4 block level, the re-processing result in a high number of clock cycles. In order to prevent the shortcomings, the architecture is taken into advantage of BRAMs in FPGA to store the predicted luma 16×16 blocks and Chroma 8×8 blocks.

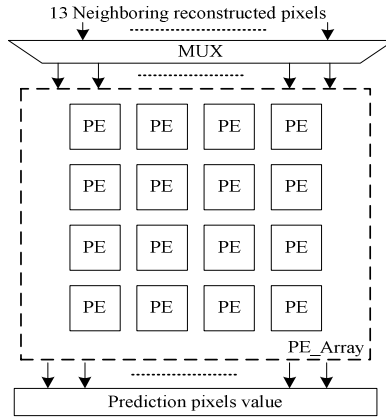


Fig. 3. Hardware architecture of PE array

4 Results and Comparisons with Related Work

The proposed algorithm is described in Verilog HDL and verified using Modelsim 6.5 SE. And then the Verilog RTL is synthesized to a Xilinx Virtex-5 FPGA using Synplify. The maximum clock frequency can be achieved at 184.8 MHz. In the simulation example, when the best prediction mode of chroma 8x8 block is mode1, the simulation waveform of prediction mode decision is shown in figure4. In the simulation waveform, the mincost port represents the minimum SAD value. The best mode port represents the best prediction mode corresponding to the minimum SAD value, which is mode 1 in this algorithm design.

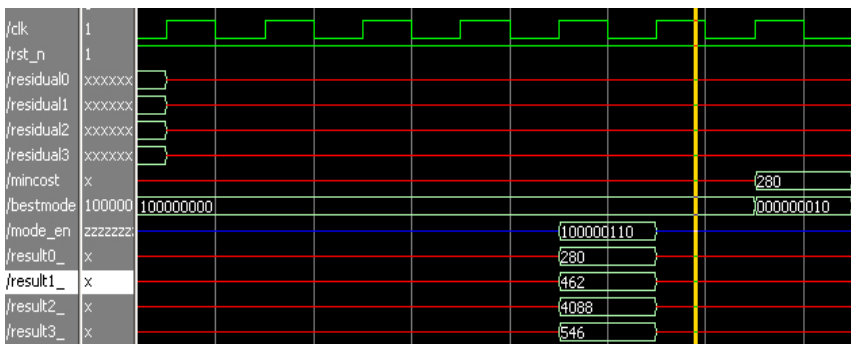


Fig. 4. The simulation waveform of intra prediction mode decision

The synthesis results of the architecture are shown in Table 5; and table 6 shows the comparisons with related work.

Table 5. Synthesis Result Of Intra Prediction

FPGA Device	Xilinx Virtex 5
Pixel Parallelism	16 pixels
36Kb BRAMs	10
Max.freq. (MHz)	184.8
LUTs	4699
Throughout (MB/s)	2.888 M

Table 6. Contrasts With Related Work

	[5]	[6]	[7]	This work
FPGA Device	Altera Stratix II	Altera Stratix II	Xilinx Virtex 2	Xilinx Virtex 5
Cycles/MB	36	--	--	<160
Max.freq/MHz	98.43	153	120	184.8
LUTs	3267	--	16546	4699

In the proposed architecture, with pipelining and parallel architecture, we can obtain 16 prediction pixels concurrently. Moreover, the throughput achieved by the proposed architecture is higher than the published results in [5,6,7], which make it possible to satisfy the requirement for the real time encoding.

5 Conclusion

An effective architecture is designed for intra prediction algorithm. By exploiting the inherent spatial correlation existing in the neighbor pixels and prediction modes, a significant computational saving can be achieved. With the hardware design, a parallel and configurable architecture is adopted to speed up the encoding time and at the same time it allows to reduce the computational complexity without any coding performance loss. The maximum clock frequency of the proposed hardware architecture can achieve 184.8 MHz in the Xilinx Virtex-5 FPGA. The experimental results confirm that the architecture can completely satisfy the real-time requirement for HDTV (1920×1080) video at 60 fps.

References

1. Wiegand, T., Sullivan, G.J., Bjøntegaard, G.: Overview of the H. 264/AVC Video Coding Standard. *Circuits and Systems for Video Technology* 13(7), 560–576 (2003)
2. Sahin, E., Hamzaoglu, I.: An Efficient Hardware Architecture for H.264 Intra Prediction Algorithm. In: *Design, Automation & Test in Europe Conference & Exhibition*, pp. 1–6. Nice (2007)

3. Huang, Y.H., Ou, T.S., Chen, H.H.: Fast Decision of Block Size, Prediction Mode, and Intra Block for H.264 Intra prediction. *Transactions on Circuits and Systems for Video Technology* 20(8), 1122–1132 (2010)
4. Yu, Y., Wang, L.: A fast Intra Mode Selection Method for H.264 High Profile. In: *International Conference on Acoustics, Speech and Signal Processing, Las Vegas*, pp. 681–684 (2008)
5. Palomino, D., Corrêa, G., Diniz, C., Bampi, S.: Algorithm and Hardware Design of a Fast Intra-frame Mode Decision Module for H.264/AVC encoders. In: *SBCCI 2011 Proceedings of the 24th Symposium on Integrated Circuits and Systems Design, New York*, pp. 143–148 (2011)
6. Shrivastava, V.K., Muralidhar, P., Rama Rao, C.B.: Architecture for H.264 Intra Prediction Fast Mode Decision Algorithm. *International Journal of Computer Applications* 68(7), 1–6 (2013)
7. Li, X.Y., Ji, F.: A Parallel H.264 Intra-Frame Prediction Decision Architecture Based on FPGA. In: *International Conference on Computational and Information Sciences (ICCIS), Shiyang*, pp. 1611–1615 (2013)