VLSI Design for a Memory Efficient Motion and Disparity Estimation of the Multiview Video Coding

[1]Felipe Sampaio, [1]Bruno Zatt, [1]Sergio Bampi, [2]Luciano Agostini
{felipe.sampaio, bzatt, bampi}@inf.ufrgs.br, agostini@inf.ufpel.edu.br

[1]Universidade Federal do Rio Grande do Sul - UFRGS
Programa de Pós-Graduação em Computação - PPGC

[2]Universidade Federal de Pelotas - UFPel
Grupo de Arquiteturas e Circuitos Integrados - GACI

Abstract

This paper presents a high throughput and low off-chip memory bandwidth Motion and Disparity Estimation architecture targeting the Multiview Video Coding requirements. The ME and DE modules are the critical paths in the multiview encoding process, corresponding to up to 80% of the encoding time. Besides, these two modules are responsible for more than 70% of the off-chip memory accesses. The goal of this work is to design a hardware architecture that deals with these two constraints. The design space exploration points the best balance between area and throughput. Besides, the Memory Hierarchy allows a reduction of 87% for memory accesses when compared to a solution without memory management. The synthesis results for the FPGA implementation show that the ME/DE architecture is able to process up to 5-view HD 1080p multiview videos in real time in a typical prediction structure with 2 reference frames (temporal and disparity neighbors). When compared to related works, this work presents the best efficiency in terms of off-chip memory access and maximum throughput at this data input.

1. Introduction

The Multiview Video Coding (MVC) is the H.264/AVC extension that deals with multiple views redundancies. Besides the temporal and the spatial redundancies which have already been explored by the traditional monoview encoders, the MVC encoders aim to increase the compression rates by 2 times exploring the inter-view redundancies [1]. The Motion Estimation (ME) is part of the Inter Frame Prediction and is responsible to exploit the temporal redundancies, i. e., objects that appear in two or more consecutive frames. The Disparity Estimation (DE), part of the Inter View Prediction MVC innovation, has the goal of reducing the disparity redundancy that is inserted by the multiple scene views.

The MVC Standard defines that the frame is divided in blocks with 16x16 pixels which are called macroblocks (MB). The MVC ME/DE search is based on a further division of this MB in variable-sized blocks (from 4x4 up to 16x16 pixels), called as current block. The block is searched in a delimited Search Window (SW) of one or more reference temporal or disparity neighbor frames and it is guided by a search algorithm. This work considers the Full Search (FS) algorithm, which performs all possible comparisons. The search is performed by a block matching approach by using some similarity metric between the current block and the candidate block. The most widely used metric is the SAD (Sum of Absolute Differences) [2].

Several issues must be considered in a design of MVC codecs, like the target throughput that is required to achieve real time processing (24~30 frames per second). Besides, another problem is related to how to delivery all the necessary data maintaining the desired throughput. In other words, the memory bandwidth must be efficiently used. In MVC processing, the critical modules in terms of complexity and memory bandwidth are the Motion and Disparity Estimation. These modules are the core of the Inter View and Inter Frame Predictions and are responsible for the highest gains in compression among all coding tools [1].

This work focuses on two main goals: (1) high throughput and (2) low off-chip memory bandwidth. The high throughput is achieved by exploring the FS inherent parallelism using a column based approach. Besides, the off-chip bandwidth was optimized using some well-known literature schemes (Level D data reuse) [6] and additional frame scheduling processing order (SWCS - Search Window Centric Scheduling).

The designed architecture is able to process two block searches in an interlaced way (ME and DE processing). It is based on a SWCS frame schedule approach and performs the FS of two current blocks using the same Search Window that is managed by a Memory Hierarchy. The architecture design was focused on a FPGA device. The two on-chip memories were designed to be mapped on Block RAMs, available in all recent FPGA families. Besides, the synthesis results and comparisons consider the architecture implementation on a Xilinx Virtex 5 xc5vlx30 FPGA device.

The paper is organized as follows: Section 2 explains all the architecture issues: memory organization, processing unities and buffers; Section 3 discusses the synthesis results and presents a comparison with the state-of-art works; finally, Section 4 concludes this work.

2. ME/DE Architecture Issues

The main goal of this work is to ally high throughput (required for real time processing) with reduced off-chip memory bandwidth. It means that the processing datapath must take advantage of each memory access and performs all possible operations while the data is not discarded.

Some design decisions were assumed based on some evaluations done with the H.264/AVC reference software using HD 1080p videos. Since this work is focused in HD 1080p videos, only the 16x16 block size is supported by the designed architecture.

Some important works in literature, like [7], assume an external value to determine the start point for DE searches based on the already know camera disparity of the multiview video. This information is expressed by a vector that is commonly referred as GDV (Global Disparity Vector). This work assumes this information to improve the coding quality results and to reduce the computational complexity. The Search Window was sized as [-8,8) in this work, using the GDV. Fig. 1 presents the overall block diagram of the designed ME/DE architecture with its main modules.

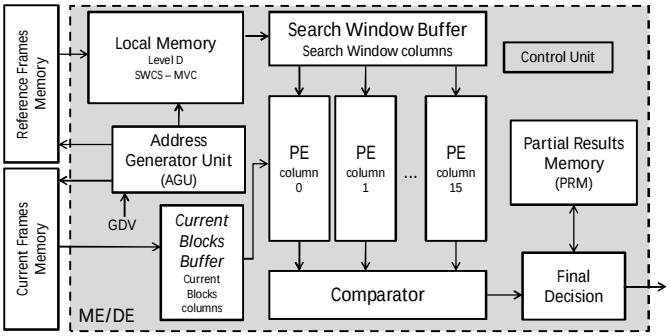

Fig. 1 - ME/DE Architecture Block Diagram.

The architecture is composed by three main parts: (1) the Memory Hierarchy that implements the Level D - SWCS data reuse scheme, (2) the datapath that synchronizes and processes all data flow and (3) the control modules (AGU and Control Unit in Fig. 1). Further details are described in the next sections.

2.1. Datapath Architecture

The datapath is composed by a Processing Element Array (PE Array) which is formed by sixteen SAD tree calculator that are able to calculate the partial SAD for one column of 16 pixels. Also, the PE stores the partial SAD in accumulator structures. After 16 clock cycles, the PE deliveries one complete SAD calculation for one candidate block. Each PE was allocated to process a specific column of blocks in the Search Window. In order to synchronize the correct data flow of the Search Window and the Current Blocks, two buffers were designed and sized to store the required information until they are not necessary anymore. The Final Comparator joint all SAD results of each PE column and deliveries the best one to the Final Decision.

2.2. Memory Hierarchy

The Memory Hierarchy is composed by two on-chip memories. The Local Memory (as shown in Fig. 1) is responsible to store the Search Window samples that are currently scanned by the ME/DE engine. This memory is organized as a circular buffer and it was implemented to support the Level D data reuse strategy. When the ME/DE is executed for adjacent current blocks, their Search Windows share a region of the reference frame where both searches must access. These samples must be accessed a few times in successive ME/DE executions. It implies in redundant memory accesses. The Level D strategy stores locally all frame width [6]. Since the MBs are processed in the raster scan order, when the last MB in a line is finally processed by the ME/DE and the first MB of the next line is initialized, just a few samples must be read from the off-chip memory.

The Local Memory output is connected to the Search Window Buffer and, at each clock cycle, two samples are accessed and organized in the buffer to be correctly passed to the PE array. The Current Block Buffer organization intercalates ME and DE current block columns. This way, the ME and DE columns are passed to the PE array in an intercalated way. Search Window Buffer organization allows the storage of two Search Window columns of 32 pixels.Besides, the SWCS was used to improve the off-chip memory bandwidth reduction. The SWCS (Search Window Centric Scheduling) is an alternative way to view the video coding order. Instead of considering the current block as the center of the coding process (traditional CBCS schedule – Current Block Centric Scheduling), now the target is to scan one specific reference frame and perform all the ME and DE searches with this Search Window.

The Partial Results Memory (PRM) is required by the out-of-order frame processing required by the SWCS data reuse frame scheduling. Its function is to temporally store the last SAD result of a given block until they are finally decided. When it is decided, then the memory position can be erased and it is available for a partial result of other non-decided block. For each block, the PRM must be able to store: (a) the best SAD distortion information found until that moment, (b) its related motion (or disparity) vector and (c) its reference frame index.

2.3. Pipeline Schedule

The architectural design combines two main ideas: (1) column based Search Window scan and (2) SWCS frame scheduling. One goal of this work is to efficiently use the off-chip memory bandwidth. Then, the processing order of the overall architecture was defined in order to achieve this requirement. Fig. 2 presents a simple scenario with the configurations that were used in the architectural design.

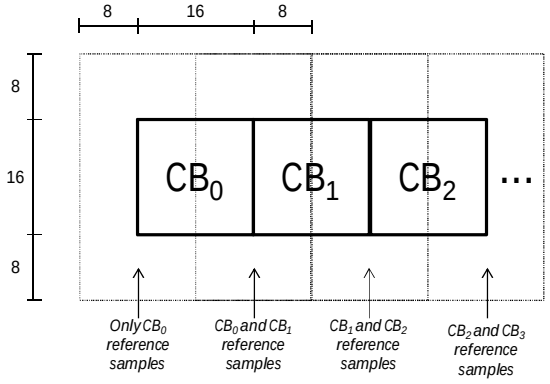

Fig. 2 - Current blocks and Search Windows typical scenario.

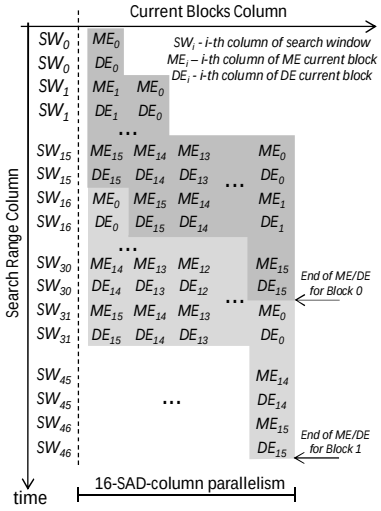

Fig. 3 - Pipeline Schedule.

The pipeline flow is guided by the Search Window columns. At the first moment, the first column of the Search Windows for the CB_0 is read from the memory. With these samples in the Search Window Buffer, all the possible partial SAD calculations are performed. When there are not more calculations to be performed with these samples, they are discarded and a new column is accessed. In an advanced stage, when the current Search Window column of samples belongs to two Search Windows (the overlap region between CB_0 and CB_1 Search Windows), the architecture must calculate partial SAD for all necessary data of these two current blocks.

The pipeline schedule considers that each Search Window column will be evaluated with all necessary current block columns of the ME and DE in an interlaced way. Fig. 3 shows this pipeline schedule. For example, in the two first time slots, the first Search Window column is matched with the first columns of ME and DE current blocks. This interlaced processing is repeated for all columns.Each time slot, with 16 clock cycles, represents the number of cycles available to the PE array to calculate all partial SAD operations. Fig. 3 shows the pipeline filling process and, after the column SW_{15} is accessed, all PE columns are fully processed until the end of the line of the frame, following the raster scan order

It takes for the ME/DE architecture to perform the FS for the first current blocks (ME and DE current blocks) $32 \times 16 \times 2 = 1024$ clock cycles. Furthermore, because of the pipeline approach, the next two adjacent ME/DE current block searches will be done in $16 \times 16 \times 2 = 512$ cycles. Considering the two target resolutions, XGA and HD 1080p, the architecture spend 33,280 and 61,952 cycles, respectively, to process one entire line of blocks of each ME and DE current frame. This way, a whole reference frame is processed by the ME/DE architecture in 1,597,440 cycles for XGA resolution and in 2,246,400 cycles for HD 1080 videos.

3. Results

The goal of this work is to design a ME/DE architecture that deals with the MVC performance constraints. Besides, data reuse techniques were employed to reduce the off-chip memory bandwidth. Tab. 1 presents some technology independent and dependent metrics that result from the design space exploration. All results are related to a prediction structure using two reference frames (temporal and disparity neighbors).

Tab. 1 - ME/DE Architecture Results

Specification		XGA (1024x768)	HD 1080p (1920x1080)
5views @30fps	**Freq. (MHz)**	228.5	321.4
	Off-chip BW (Mbytes/s)	114.6	299.1
6views @30fps	**Freq. (MHz)**	274.2	385.6
	Off-chip BW (Mbytes/s)	137.1	358.9
Local Memory (Kbytes)		31	58.1
PRM (Kbytes)		51	143.5
FPGA Implementation (Virtex5)	**#SliceLUTs**	19,661(28%)	
	#SliceRegisters	12,345(17%)	
	#LUT-FF Pairs	9,775 (43%)	
	Memory Bits	1,612.8	
	Freq. (MHz)	369.5	

The maximum frequency result from the synthesis allows the 8-view processing for a XGA multiview video in real time. For the high definition resolution, the architecture is able to real-time process 5 views of HD 1080p. At this operation frequency, the required bandwidth is up to 300 Mbytes per second.

There are not FPGA-based architectural implementations targeting low memory access for ME/DE processing. So, the comparison will be performed with ASIC implementations. The works [3] and [4] focus in the MVC in different ways: in [3] a fast ME/DE datapath is inserted in parallel with the usual ME in order to accelerate the MB processing for low motion and low disparity blocks, while the work [4] proposes a new algorithm for ME/DE in order to increase the cache hit of the proposal architecture. Both works admit some PSNR and bitrate losses in order to reduce complexity and memory access rates. Besides, the work [5]is in the comparison by the design similarities, since it also implements architecture for Multiple Reference Frames. In this case, all reference frames are temporal neighbors, since multiview videos are not allowed.

Tab. 2 – Related Works Comparison

Criteria	ThisWork	[3]	[4]	[5]
Technology	Virtex 5FPGA	IBM 65nm LPeLowK	TSMC 90nm	TSMC 180nm
Data ReuseScheme	Level DSWCS	Level A and Level C	Cache Based	Multilevel C+
Reference Frames	2	4	4	4
Max. Cap.	5 views1080p @30fps	4 views 1080p @30fps	2 views 1080p @30fps	1 view720p @56fps
Off-chip Mem. BW (MB/s)	300	N.I.	N.I.	204.4
On-chipMemory (KB)	201.6	92.1	7.8	3.96
Efficiency(Throughput/ Off-chip BW)	1,012.5	-	-	132.1

The works [3] and [4] do not clearly inform the off-chip memory bandwidth due the target specifications. This way, complete comparison with these two works was not possible. Besides, as these two works do not use the FS as the target search algorithm, neither a performance evaluation would be fair. It is important to notice that both [3] and [4] aim to achieve high throughput rates by reducing the ME/DE complexity. In [3], it was designed additional processing datapath for Fast ME/DE in order to save computation for low motion/disparity blocks. The work [4] proposes a cache efficient ME/DE algorithm in order to exploit cache hit and misses to direct the search algorithm. However, both works have PSNR and bitrate penalties.This work applies another design decision: exploit the FS parallelism and regularity in order to increase the throughput without any PSNR drop. The area overhead could not be calculated because of the different target technologies. The work [5] achieves a better off-chip bandwidth result than our work. However, the achieved throughput is considerably worst. It means that each off-chip memory access is better efficiently used in this work than in [5]. The last row of Tab. 2 presents the efficiency metric that express this trade-off. Considering this aspect, the memory efficiency of this work surpasses [5] in 7.6 times.

4. Conclusions

This paper presented the FPGA implementation of a ME/DE hardware architecture. The goal was deal with some important issues in the Multiview Video Coding: (a) high computational complexity and (b) high off-chip memory bandwidth. The high throughput was achieved by exploiting the available parallelism in the Full Search algorithm execution. The Memory Hierarchy was designed in order to locally exploit the spatial and temporal data locality. This way, important results in off-chip memory bandwidth reduction were achieved. As main results, the architectural FPGA implementation in a Xilinx Virtex 5 device allows real time processing when using 5-view 1080p HD. The memory access reduction is up to 87% when the Memory Hierarchy is used. The comparison with related works showed that the designed ME/DE architecture has the best efficiency, i. e., the best relation between throughput and required off-chip bandwidth.

5. References

[1] P. Merkle, et al. “Efficient Prediction Structures for Multiview Video Coding.” In: IEEE TCSVT, v. 17, n. 11, pp. 1461-1473, nov. 2007.

[2] T. Wiegand, et al. “Overview of the H.264/AVC Video Coding Standard”. In: IEEE TCSVT, v. 13, n. 7, pp. 560- 576, jul. 2003.

[3] B. Zatt et al. “Run-time adaptive energy-aware Motion and Disparity Estimation in Multiview Video Coding”. In: 48th DAC, pp.

[4] P.-K. Tsung, et al. “Cache-Based Integer Motion/Disparity Estimation for Quad-HD H.264/AVC and HD Multiview Video Coding”. In: ICASSP, . Taipei: pp. 2013-2016, 2009.

[5] M. Grellert et al. “A multilevel data reuse scheme for Motion Estimation and its VLSI design”. In: IEEE ISCAS, pp. 583-586, 2011.

[6] C.-Y. Chen, et al. “Level C+ Data Reuse Scheme for Motion Estimation With Corresponding Coding Orders.” In: TCSVT, v. 16, n. 4, p. 553-558, april. 2006.

[7] T.-Y. Kuo, et al. “A novel method for global disparity vector estimation in multiview video coding”. In: IEEE ISCAS 2009.