Journal of Systems Architecture 56 (2010) 305–316

Contents lists available at ScienceDirect

Journal of Systems Architecture

journal homepage: www.elsevier.com/locate/sysarc

Efficient architectures for 3D HWT using dynamic partial reconfiguration

A. Ahmad [a,b,*], B. Krill [a], A. Amira [c], H. Rabah [d]

[a] Department of Electronic and Computer Engineering, School of Engineering and Design, Brunel University, West London, UB83PH Uxbridge, United Kingdom
[b] Department of Computer Engineering, Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia (UTHM), 86400 Batu Pahat, Johor, Malaysia
[c] Nanotechnology and Integrated BioEngineering Centre (NIBEC), University of Ulster, Shore Road Newtownabbey, BT37 0QB Belfast, Northern Ireland
[d] Laboratoire d'Instrumentation, Electronique de Nancy, University Henri Poincare, 540003 Nancy, France

ARTICLE INFO

Article history:
Received 28 September 2009
Received in revised form 4 February 2010
Accepted 9 February 2010
Available online 20 February 2010

Keywords:
Dynamic partial reconfiguration (DPR)
Medical image processing
Haar wavelet transform (HWT)

ABSTRACT

This paper presents the design and implementation of three dimensional (3D) Haar wavelet transform (HWT) with transpose based computation and dynamic partial reconfiguration (DPR) mechanism on field programmable gate array (FPGA). Due to the separability property of the multi-dimensional HWT, the proposed architecture has been implemented using a cascade of three N-point one dimensional (1D) HWT and two transpose memories for a 3D volume of $N \times N \times N$ suitable for real-time 3D medical imaging applications. These applications require continuous hardware servicing, hence DPR has been introduced. Two architectures were synthesised using VHDL and implemented on Xilinx Virtex-5 FPGAs. Experimental results and comparisons between different configurations using partial and non-partial reconfiguration processes and a detailed performance analysis of the area, power consumption and maximum frequency are analysed in this paper.

1. Introduction

Real-time three dimensional (3D) medical image processing is a niche area concerned with the processing of real-time sequences of medical image data which represents a developing trend in medical industries. Due to significant advances in medical technology, there are various 3D medical imaging modalities such as magnetic resonance imaging (MRI), computed tomography (CT) and positron emission tomography (PET) that have been widely used especially for cancer screening and diagnosis.

With a more realistic manner of anatomic structure representation including the information about position, size, shape, and ease of visualisation, 3D images provide more accurate parameters for the assessment of pathological changes [1], and thus leading to great advantages for image based screening, diagnosis and treatments.

Despite their many advantages, most 3D medical imaging algorithms are computationally intensive with matrix transformation as the most fundamental operation involved in transform based methods. Computational complexity for various algorithms that have been employed in 3D real-time medical imaging [2] reveal that they are in the order from $O(N \times \log N)$ for fast Fourier transform (FFT) to $O(N^2 \times J)$ for the recently developed curvelet transform (where N is the transform size and J is the maximum transform resolution level) and hence are extremely computationally intensive for large medical volumes data [3].

Therefore, there is a need for high performance systems while keeping architectures flexible to allow for quick upgradeability to cope with real-time applications. Moreover, an efficient implementation for these operations are of significant importance and hence lead to efficient solutions for large medical volumes data in real-time medical imaging applications.

As diverse as the applications may seem, there is no doubt that the discrete wavelet transform (DWT) plays a significant role in image processing applications as an alternative to classical time frequency representation techniques such as the discrete Fourier transform (DFT) and discrete cosine transform (DCT) [4].

Owing to its multi-resolution characteristics and its capability to represent real life non-stationary signals such as image and speech, DWT has attracted a great deal of research and development. Haar wavelet transform (HWT) is the simplest of all wavelets and it has been applied in image multi-resolution due to the following advantages: conceptually simple, fast, memory-efficient, and it is exactly reversible without the edge effects which are a problem with other wavelet transforms [5,6].

Reconfigurable hardware, especially field programmable gate arrays (FPGAs) are widely used in real-time processing from simple low-resolution and low bandwidth applications to very high-resolution and high-bandwidth [7].

The advantages offered by FPGAs, such as massive parallelism capabilities, multimillion gate counts, and special low power packages contributes to minimise the amount of memory used, computational complexity and power consumption [8] and seems to be

* Corresponding author. Address: Department of Electronic and Computer Engineering, School of Engineering and Design, Brunel University, West London, UB83PH Uxbridge, United Kingdom.
E-mail address: Afandi.Ahmad@brunel.ac.uk (A. Ahmad).

1383-7621/$ - see front matter
doi:10.1016/j.sysarc.2010.02.001

an excellent hardware candidate for an efficient implementation of real-time 3D medical image processing compared to application specific integrated circuit (ASIC) and even digital signal processor (DSP) platforms [9].

The application of 3D real-time image processing uses several building blocks for its computationally intensive algorithms to perform matrix transformation operations. Moreover, complexity in addressing and accessing large medical volumes data and massive amount of data to be processed have resulted in vast challenges from a hardware implementation point of view.

In order to cope with these issues, FPGAs with efficient reconfigurability mechanisms should be employed to meet the requirements of these applications in terms of maximum speed, size, power consumption and throughput.

Dynamic partial reconfiguration (DPR) is a promising mechanism for reducing the hardware required for implementing an efficient design for 3D medical image processing application as well as improving the performance, speed and power consumption of the system.

With this technique, the design can be divided into sub-designs that fit into the available hardware resources and can be uploaded into the reconfigurable hardware when needed. As the first configuration finishes its operations, the next configuration can be uploaded to the hardware [10]. In short, by timesharing, large designs can be implemented on limited hardware resources.

From designer and FPGA technology perspective, partial reconfiguration enables designers to realise the implementation of multiple modules in complex system on one FPGA device where a static area must be defined for components communication, while full reconfiguration implementation treat the entire FPGA as one module [10].

Xilinx provides many FPGA devices that support active partial reconfiguration, ranging from the Spartan-3 device to the latest Virtex-5 device. The active or dynamic reconfiguration allows the modification of a part of the FPGA while the remaining part of the device is active [11]. Thanks to the internal configuration access port (ICAP) the reconfiguration management can be implemented in the FPGA so that the final system is self-reconfigurable [12,13].

Basically, a dynamically and partially reconfigurable system based on commercial FPGAs is composed of a static part and different dynamic parts. The communication between these parts requires the use of bus macros. In Virtex-II family this bus macros are implemented using tri-state buffers which is not very effective. A new alternative based on slice macro in form of predefined routed macro has been introduced in more recent devices (Virtex-4, 5 and 6) [14]. These macros are placed at the edge of boundaries separating dynamic and/or static region modules and allowing the communication between different regions of the FPGAs. Further explanation will be discussed in Section 4 for the implementation of 3D HWT with DPR mechanism.

The aim of this paper is to develop efficient architectures for 3D HWT with transpose based computation using DPR technique. The proposed architectures will be deployed in a reconfigurable environment for adaptive medical image compression. An in depth evaluation of these architectures in terms of area, power consumption and maximum frequency is also carried out. Influence of the transform size on the hardware performance for both proposed architectures is addressed.

The rest of the paper is organised as follows. An overview of the related work of 3D HWT and DPR are given in Section 2. Section 3 explains the mathematical background for HWT and matrix transposition. Section 4 exposes the proposed architecture of 3D HWT with DPR mechanism. Experimental results, comparison and analysis are presented in Section 5. Finally, concluding remarks and further potential ideas to be explored are given in Section 6.

2. Related work

In this section, existing 3D DWT architectures that have been implemented on FPGAs and a broad idea of DPR implementation technique in various fields are reviewed.

2.1. 3D DWT

Despite its complexity, there has recently been an interest in 3D DWT implementation on various platforms. However, a survey of existing literature indicates that the research is still in its infancy as demonstrated by the limited contributions proposed and it can be classified into three categories: architecture development [15–17], architecture with FPGA implementation [18–22] and finally architecture that has been implemented on other silicon platforms [23]. Since the aim and contribution of this paper is on the implementation of partial and non-partial reconfiguration on FPGA, the reviews and discussion will focus on previously published FPGA implementation only.

With regard to FPGA implementation, in Ref. [18] they proposed a memory-efficient real-time architecture that has been implemented with optimised memory requirement to the order $O(KN^2 + (K-2) \times N)$. Moreover, the proposed architecture also is low power and can run at higher frequency which fully utilises the advantages of multiplierless, parallelism and pipelined structures for filter design. The proposed architecture has been implemented on Xilinx FPGA devices and operating at an operating frequency of 75 MHz. However, it is noted that the drawback of this architecture is the reduction in compression of video sequences and modifications are required for better optimisation in performance.

Another significant implementation in [19,20], focuses on area-efficient high-throughput architecture. The architecture is based on distributed arithmetic (DA) and suitable for real-time medical imaging applications. It was implemented in VHDL and synthesised on Xilinx Virtex-E FPGAs. With the advantages offered by the DA design technique, the architecture has been proven as area-efficient high-throughput with the processor running up to 85 MHz and capable to process five levels DWT analysis of a $128 \times 128 \times 128$ functional magnetic resonance imaging (fMRI) volume image in 20 ms.

Ismail et al. presented another hardware design and FPGA implementation of a new efficient 3D DWT for video compression application [21]. The design has been implemented on a device from Altera. It exhibits low memory demands and lower latencies for the compression and decompression processes. The advantages of their design is that it is generic to any wavelet filter coefficients and scalable to fit for any frame size of the video sequence.

The recent FPGA implementation of 3D wavelet has been highlighted for video segmentation in traffic monitoring by Salem et al. Empirically, the design has a slice utilisation of 63% and the maximum frequency allows a 100 MHz system clock [22].

Table 1 summarises the previously published architectures for FPGA implementation of 3D DWT [18–22]. Due to the fact that the implementation of 3D DWT on FPGA is still immature, in order to evaluate and conduct a comparative study, the following justifications have been considered:

1. Proposed architectures in Refs. [18–22] were targeted for different FPGA platforms, applications and design techniques. With a different FPGA technology available on each devices such as process technology and circuit topology, it draw inconsistent comparison and performance evaluation.