Chapter 31
Mobile GPU Computing Based Filter Bank Convolution for Three-Dimensional Wavelet Transform

Di Zhao
Chinese Academy of Sciences, China

ABSTRACT

Mobile GPU computing, or System on Chip with embedded GPU (SoC GPU), becomes in great demand recently. Since these SoCs are designed for mobile devices with real-time applications such as image processing and video processing, high-efficient implementations of wavelet transform are essential for these chips. In this paper, the author develops two SoC GPU based DWT: signal based parallelization for discrete wavelet transform (sDWT) and coefficient based parallelization for discrete wavelet transform (cDWT), and the author evaluates the performance of three-dimensional wavelet transform on SoC GPU Tegra K1. Computational results show that, SoC GPU based DWT is significantly faster than SoC CPU based DWT. Computational results also show that, sDWT can generally satisfy the requirement of real-time processing (30 frames per second) with the image sizes of 352×288, 480×320, 720×480 and 1280×720, while cDWT can only obtain read-time processing with small image sizes of 352×288 and 480×320.

1. INTRODUCTION

System on Chip (SoC) is a tiny but complete computer, which consists almost all components of a computer such as CPU, GPU and memory. Because of SoC's advantage of low power consumption, these chips are wildly embedded into mobile systems. Recently mainstream SoCs include Atom from Intel, Tegra from Nvidia, Snapdragon from Qualcomm, A*x* from Apple, MT*x* from MediaTek, *etc*, where *x* means number(s). SoC runs on multiple operation systems such as Android (Alejandro Acosta & Francisc Almeida, 2014; Alejandro Acosta & Francisco Almeida, 2014a), Linux and Windows.

DOI: 10.4018/978-1-5225-0983-7.ch031

Existing research topics of SoC include performance analysis (A. Acosta & F. Almeida, 2014; Alejandro Acosta & Francisco Almeida, 2014b; Papadopoulos et al., 2014), power consumption (Grasso, Radojkovic, Rajovic, Gelado, & Ramirez, 2014; Papadopoulos et al., 2014; Zhan, Lung, & Srivastava, 2014), *etc*. Similar with regular GPU in desktop or notebook, SoC embedded GPU (SoC GPU) is responsible for graphics processing for SoC (Giles & Reguly, 2014). Companies develop different architectures for SoC GPU for example Apple A*x*'s PowerVR, Tegra K1's Kepler (Singh & Jain, 2014), *etc*.

Tegra K1 is one of Nvidia's latest SoCs which include 32-bit version and 64-bit version. Tegra K1 32-bit version is released in 2014, and Tegra K1 64-bit version is on developing. Tegra K1 32-bit is fabricated by 28nm HPM. Tegra is developed for applications such as rendering (Mobeen & Lin, 2012; Rodríguez & Alcocer, 2012; Q. Wang, Yu, Rasmussen, & Yu, 2014), ray tracing (Lee et al., 2013), optical flow (Plyer, Le Besnerais, & Champagnat, 2014), face recognition (Kwang-Ting & Yi-Chu, 2011; Y.-C. Wang, Donyanavard, & Cheng, 2012), object tracking (Růžička & Mašek, 2014), computational photography (Pulli & Troccoli, 2014) and sift detector (Rister, Guohui, Wu, & Cavallaro, 2013).

A wavelet is a mathematical function for decomposing a given function into different scale components, wavelet is applied to digital signal processing for decades. Wavelet transform generally includes two categories: continuous wavelet transform and discrete wavelet transform. Discrete Wavelet Transform (DWT) is a category of wavelet transform with discrete wavelet coefficients (Press, 2007), and there are two mainstream DWT implementation algorithms: filter bank convolution and the lifting scheme (Jung, Park, & Kim, 2005).

DWT is discretized in scale and continuous in time, and DWT is a time-consuming part of the fields such as signal processing, image processing, video processing and data compression. However, these applications of DWT generally have the real-time requirements, and acceleration techniques for DWT are developed for decades on the computational platforms from multi-core CPU, GPU computing and mobile GPU.

Three-dimensional DWT is the essential component for applications such as MPEG encodec, medical imaging processing, image compression, video watermarking and volume segmentation. While one-dimensional DWT handle one-dimensional signals, and two-dimensional DWT handles two-dimensional signals such as image, three-dimensional DWT handle three-dimensional signals such as video. Since DWT's applications generally hold real-time requirements, acceleration of DWT attracts researchers for years. Acceleration of DWT includes filter bank convolution by GPU (Changhe, Yunsong, Jie, & Jie, 2014; Joaquín Franco, Bernabé, Fernández, & Ujaldón, 2012; Galiano, López, Malumbres, & Migallón, 2013; Simek & Asn, 2008; Su, 2011; Tenllado, Setoain, Prieto, Pinuel, & Tirado, 2008; Tien-Tsin, Chi-Sing, Pheng-Ann, & Jianqing, 2007), filter bank convolution by multi-core CPU, lift scheme by GPU (Changhe et al., 2014; Chen et al., 2013; Joaquín Franco et al., 2012; Galiano, López, et al., 2013; Matela, 2009; Sharma & Vydyanathan, 2010; Simek & Asn, 2008; Tenllado et al., 2008; van der Laan, Jalba, & Roerdink, 2011; J. Wang, Lu Wen, Liu, & Jiang, 2011; Zhuo, Zilong, Yajuan, & Shengsheng, 2010), lifting scheme by multi-core CPU and other approaches (Dechevsky, Gundersen, & Bang, 2010; J. Franco, Bernabe, Fernandez, & Acacio, 2009; Galiano, López-Granado, Malumbres, & Migallón, 2013; Inoue, Kuroki, Kurosaki, Nagao, & Ochi, 2011; Jie, Xinxiang, Peng, & Weiwei, 2012).

However, although SoC GPU accommodates two important applications of image processing and video processing, there is almost no research of developing fast three-dimensional DWT on SoC GPU such as Tegra K1, the reasons are: (1) Tegra K1 is a brand new embedded system which is released in CES 2014, and Tegra K1 lacks supports for DWT; (2) before Tegra K1 with 192 CUDA core, the embedded systems generally lack of strong ability of parallel computing, but DWT is in low efficiency on