

Fast filter bank convolution for three-dimensional wavelet transform by shared memory on mobile GPU computing

Di Zhao¹

© Springer Science+Business Media New York 2015

Abstract Mobile GPU applications usually constrain by the real-time requirement. However, FLOPS of mobile GPU is limited by the size and power supply of the SoC systems. Same to desktop GPUs, the mobile GPU consists of an on-chip memory hierarchy, and proper usage of memory hierarchy accelerates mobile GPU applications such as Discrete Wavelet Transform (DWT) to satisfy the real-time requirement. In this paper, by taking advantage of GPU shared memory in Tegra K1, a mobile GPU from Nvidia, we develop Bank Conflict Free Shared Memory Parallel DWT for mobile GPU applications. Computational results show that, with the display resolution of 640×350 (EGA), Bank Conflict Free Shared Memory Parallel DWT is significantly faster than SoC CPU-based DWT. Computational results also show that, with the display resolution of 320×200 (CGA), 640×480 (VGA), 800×600 (SVGA) and 1024×768 (XGA), Bank Conflict Free Shared Memory Parallel DWT can generally satisfy the real-time requirement.

Keywords One-level DWT · Three-dimensional DWT · Mobile GPU computing · Video processing · Pixel parallelization · Shared memory · Bank conflict

1 Introduction

General-purpose GPU computing performs scientific computation on graphics cards, and traditionally the general-purpose computing is handled by CPU [1]. Currently, one of mainstream open GPU computing languages is Apple's OpenCL, which is designed for GPUs of multiple companies. Nvidia develops the general-purpose GPU

✉ Di Zhao
zhaodi@sccas.cn

¹ Computer Network Information Center, Chinese Academy of Sciences, Beijing 100094, China