

論文審査の要旨

博士の専攻分野の名称	博 士 (工 学)	氏名	戸倉 宏樹
学位授与の要件	学位規則第4条第1・2項該当		
論 文 題 目 A Study on Efficient GPU Implementations for Many Small Problems (大量の小さな問題のための効率的な GPU 実装に関する研究)			
論文審査担当者			
主 査	教 授	中野 浩嗣	印
審査委員	教 授	藤田 聡	印
審査委員	准教授	伊藤 靖朗	印
〔論文審査の要旨〕			
<p>GPU(Graphics Processing Unit)は、内部に数千ものコアを搭載した画像処理に特化したハードウェアである。近年では、汎用演算の高速化のために GPU を用いる研究が盛んにおこなわれている。本論文では、GPU 上での大量の小さな問題に対する高速計算に着目しており、行列の column-wise prefix-sum に対する GPU 実装と大量の小さな行列の固有値計算に対する GPU 実装を提案している。本論文の構成は以下の通りである。</p> <p>第一章では、研究背景、研究成果及び本論文の構成について述べている。</p> <p>第二章では、GPU アーキテクチャ及び CUDA について説明している。</p> <p>第三章では、行列の column-wise prefix-sum に対する GPU 実装について説明している。行列の column-wise prefix-sum は入力行列の列ごとに対して prefix-sum を計算する。GPU 上で高速計算を行うためには、十分なスレッドを用いること、コアレスドアクセスによりグローバルメモリにアクセスする必要がある。単純な実装(naive)では、十分なスレッド用いることができず、高速に計算することができない。CUB は GPU 上における一次元配列に対する prefix-sum を大量のスレッドを用いて高速に計算できるアルゴリズムである。CUB を各列に対して割り当て、column-wise prefix-sum を計算するアルゴリズム(column-wise CUB)では、グローバルメモリに対してコアレスドアクセスを行えないため、高速に計算できない。提案手法では、行列を部分行列に分割し、各部分行列に対して CUDA ブロックを割り当て、計算を行う。そのため、十分なスレッドを用いることができ、さらにコアレスドアクセスを行うため、高速に計算を行える。また、Column-wise prefix-sum 計算時間の下限値は行列の複製の実行時間である。Naïve では行列の複製に対して 80%以上のオーバーヘッドがあり、column-wise CUB では 200%以上のオーバーヘッドを持っている。提案手法は行列の複製に対して 6%以下のオーバーヘッドで計算できた。そのため提案手法はほとんど最適といえる。</p> <p>第四章では、大量の小さな行列の固有値計算に対する GPU 実装について説明している。</p>			

制御器の設計手法の一つであるパラメータ空間設計法では、大量の小さな行列の固有値計算を行う必要があるが、その計算時間がボトルネックとなっており、高速化が求められている。固有値はヘッセンベルグリダクション、ダブルシフト QR スイープの流れで計算を行う。GPU において大量の行列の固有値計算を行うために、2つのスレッドの割り当て手法 Single-Warp Based(SWB)と Multiple-Warp Based(MWB)を提案している。SWB では1 ワープが1 つ以上の行列の固有値計算を行い、MWB では1 つ以上のワープが32 行列の固有値計算を行う。ヘッセンベルグリダクションでは、 5×5 と 10×10 の行列では MWB が、 15×15 から 30×30 の行列では SWB が高速に計算した。ダブルシフト QR スイープでは、すべての行列サイズで MWB が SWB よりも高速に計算を行った。各ステップにおいて最適な手法を用いることで提案実装はライブラリの CPU 並列計算に対して5倍以上の高速化を達成した。

第五章では、本論文の結論を述べている。

以上、審査の結果、本論文の著者は博士（工学）の学位を授与される十分な資格があるものと認められる。