

論文審査の要旨

博士の専攻分野の名称	博 士 (工 学)	氏名	笠置 明彦
学位授与の要件	学位規則第4条第1・2項該当		
論 文 題 目 Memory Access Optimal Algorithms for the GPU (GPU用のメモリアクセスの最適なアルゴリズム)			
論文審査担当者			
主 査	教授	中野 浩嗣	
審査委員	教授	岩本 宙造	
審査委員	教授	中西 透	
審査委員	准教授	伊藤 靖朗	
〔論文審査の要旨〕			
<p>近年、画像生成や画像処理の計算を高速化するために設計された専用回路である GPU (Graphics Processing Unit) を汎用演算に利用する GPGPU (General Purpose computing on GPU) が広く利用されている。本論文は、GPU の特徴を捉えた並列メモリマシンモデルを導入し、そのモデル上で動作するいくつかの並列アルゴリズムの提案と、その GPU 実装に関する研究についてまとめたものである。</p> <p>第1章は、GPU に関する研究背景と各章の概要について紹介する。GPU は多数のコアを搭載したアーキテクチャであり、主に画像処理に用いられる。ここでは高い計算能力を持った GPU を汎用計算に用いる GPGPU について紹介する。</p> <p>第2章は、NVIDIA 社が提供する GPU アーキテクチャと統合開発環境である CUDA について紹介している。ここでは CUDA が提供する階層スレッドグループによるプログラミングモデルを説明し、GPU アーキテクチャは高速化の鍵となる高いレイテンシを持つグローバルメモリと低いレイテンシを持つシェアードメモリに焦点を当て、各メモリの特徴を紹介している。</p> <p>第3章は、GPU のグローバルメモリやシェアードメモリの特徴を反映した理論並列メモリマシンモデルを紹介している。この理論モデルは実際の GPU と同じく SIMD 方式で実行し、GPU と同様のメモリ空間を持っており、本研究でのアルゴリズムの詳しい解析に用いている。具体的には、GPU の構造を2つの並列メモリマシンモデルである Discrete Memory Machine (DMM) と Unified Memory Machine (UMM) を組み合わせた階層的なメモリマシンモデル (Hierarchical Memory Machine (HMM)) を提案している。</p> <p>第4章は、オンチップメモリであるシェアードメモリ上での小規模なオフラインパーミ</p>			

ューテーションのアルゴリズムを提案し、GPU に実装し評価を行った。シェアードメモリ上のデータへのアクセスは、パフォーマンスが低下するバンクコンフリクトに注意する必要があり、提案するアルゴリズムはデータへのアクセスをスケジューリングすることで、バンクコンフリクトを発生させず高速に処理を行う。評価実験では任意のパーミュテーションに対して一定の時間で処理可能であることを示し、バンクコンフリクトが発生する自明なアルゴリズムと比べて最大で 10.1 倍の高速化を達成した。

第 5 章は、オフチップメモリであるグローバルメモリ上での大規模なオフラインパーミュテーションのアルゴリズムを提案し、GPU に実装し評価を行った。提案するアルゴリズムは 3 つのステップから構成されており、各ステップではグローバルメモリ上のデータを取得し、シェアードメモリ上にコピーを行う。シェアードメモリ上では第 4 章で提案したアルゴリズムを用いて高速にパーミュテーションを行い、グローバルメモリへと書き戻しを行う。評価実験ではあらゆるパーミュテーションに対して提案するアルゴリズムは一定の時間で処理を行い、自明なアルゴリズムと比べて最大 3.6 倍の高速化を達成した。

第 6 章では、画像処理で用いられる範囲総和テーブルを GPU で高速に求めるアルゴリズムを提案する。範囲総和テーブルは各行と各列に対する累積和を計算することで求められる。本章ではグローバルメモリへのアクセス回数が最も少ないアルゴリズムを提案する。また、提案手法と既存手法と組み合わせたハイブリッドな手法を提案し、各手法を GPU に実装した。評価実験では単一コアの CPU の逐次処理と比べて 100 倍以上の高速化を達成した。

第 7 章では、第 6 章までの成果を要約し、結論としてまとめている。

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。

備考：審査の要旨は、1,500 字以内とする。