

令和元年度 情報工学コース卒業研究報告要旨

片桐 研究室	氏 名	杉 浦 拓 未
卒業研究題目	医用画像処理における LDDMM の GPU 高速化	
<p>医用画像処理において、臓器の統計形状モデルは学習用 3 次元画像中の臓器領域データ群より構築される。臓器の統計形状モデルを求める手法のうち最も広く知られている手法は Active Shape Model (ASM) あるが、ASM によって構築されたモデルは表現性能が十分でなく、臓器として適切ではない形状が含まれる可能性がある。妥当な形状のみを表現するモデルの構築方法としては、臓器領域間の非線形な微分同相写像 (1 対 1 かつ順写像も逆写像も滑らかな写像) によるモデル化が知られている。2 つの臓器間の微分同相写像を求める手法の中でも Large Deformation Diffeomorphic Metric Mapping (LDDMM) が広く採用されている。しかし、LDDMM は莫大な演算を必要とする手法である。それにも関わらず統計形状モデルを構築するためにはこの手法を多数の学習データに対して多数回適用しなければならない。そのため LDDMM の高速化への要求は大きい。</p> <p>微分同相写像を求める LDDMM コードはすでにマルチコア CPU 向けにハイブリッド MPI/OpenMP 並列化されているが、その実行時間のうちの多くは浮動小数点演算で占められていることが分かっている。したがって、高い浮動小数点演算性能を持つ GPU を活用することでさらなる高速化が期待される。よって本研究では LDDMM コードをハイブリッド MPI/OpenMP 並列化したものをベースとして、複数の GPU 上で並列に実行可能な LDDMM コードを実装することによって高速化を試みた。</p> <p>本研究では LDDMM コードを GPU 上で実行するため OpenACC を用いた。OpenACC では既存の C,C++/Fortran ソースコードにディレクティブを挿入することで処理の一部を GPU へオフロードさせて実行できる。また、GPU を複数用いて並列に実行するために CUDA-Aware MPI または NCCL をそれぞれ OpenACC と組み合わせて用いた。CUDA-Aware MPI は複数 GPU 間で通信を行うために広く使われている方法である。特に GPUDirect を利用可能な環境においては GPU 側のメモリのデータをメインメモリにコピーすることなく GPU 間で通信できるため、高速な通信が可能である。NCCL (NVIDIA Collective Communications Library) は NVIDIA により開発された GPU 向けの集団通信ライブラリである。NCCL は NVLink を用いた高速な GPU 間集団通信関数を提供しており、MPI プログラムに対して少しの修正で利用することができる。本研究ではベースとなるハイブリッド MPI/OpenMP 並列化コードのうち OpenMP ディレクティブによる並列化部分を OpenACC ディレクティブによる並列化に書き換え、MPI 通信部分を CUDA-Aware MPI または NCCL による GPU 間通信に書き換えることでマルチ GPU 向けの実装を行った。</p> <p>実験は名古屋大学情報基盤センターの GPU サーバ sx40 を用いて行った。GPU の数 (1 枚~4 枚) や GPU 間の通信方法 (CUDA-Aware MPI または NCCL) を変えていくつかの条件での実行時間を比較したところ、NCCL/OpenACC による実装を 4 枚の GPU を用いて実行する場合が最も高速であった。結果として GPU サーバの CPU のみによる逐次実行に対して 4 枚の GPU を用いた場合は約 318 倍高速となった。</p>		