

GPU 上でのハーモニッククラスタリングを用いた 基本周波数解析の並列処理

Parallel Processing of Fundamental Frequency Analysis Using Harmonic Clustering on GPU

藁谷 拓也†
Takuya Waragai

吉田 明正†
Akimasa Yoshida

1 はじめに

本稿では、デジタル音響信号に対する短時間フーリエ変換とハーモニッククラスタリングによる基本周波数解析システムの GPGPU 上での並列処理手法を提案する。多重音の音響信号から音響の特徴量を推定するためには、高い性能の多重音解析システムの開発が必要である。また、近年マルチコア及び GPGPU を用いることにより計算機の演算処理能力は飛躍的に向上した。並列処理は解析精度の高いシステムを実用化させるために必要不可欠になる。

本研究で並列化を行うハーモニッククラスタリングは倍音比の情報を必要としない基本周波数推定法である。本稿では、NVIDIA Tesla S2050 上で CUDA を用いてハーモニッククラスタリングの並列化を実現し、GPU 上で並列処理による性能評価を行った。

2 音源数推定を含むハーモニッククラスタリング

基本周波数解析にはフーリエ変換によって得られたスペクトルをもとに亀岡らによるハーモニッククラスタリング (Harmonic Clustering)[2, 3] を用いる。ここでは窓関数の影響により左右に広がったスペクトルが観測され、その形状が正規分布で近似できると仮定する。ハーモニッククラスタリングは基本周波数とそれぞれの倍音周波数を中心とするクラスタ帯域、クラスタ重心を決定し、それぞれのクラスタに倍音構造としての拘束を与えて自由度 1 のクラスタリングを行うものである。このような拘束を与えられたクラスタ群を倍音クラスタ群と呼ぶ。また、このハーモニッククラスタリングでは倍音周波数は基本周波数の整数倍であると仮定している。音源数を判定するには、AIC (赤池情報量基準) [4] を導入し最適なモデルの選択基準とする。

この周波数軸上に広がったスペクトルから求めたパワースペクトル密度を微小エネルギーの度数分布と考えると、その微小エネルギーのクラスタ帯域に対する帰属確率は正規分布で最適近似するものである。

2.1 音源数推定のための準備

多重音のハーモニッククラスタリングは複数 (K 個) の倍音クラスタ群を用いて EM アルゴリズムにより解く。その際、倍音クラスタ群 k の n 倍音に対する重心を $\mu_k + \log n$ とし、クラスタ群 k において上限 (ナイキスト周波数の対数) まで取りえるクラスタ重心数を $N(k)$ とする。また、クラスタごとの重みを w_n^k とし、クラスタ帰属度を $p_n^k(x)$ 、尤度関数を $\varphi(x, \mu_k + \log n)$ 、観測スペクトル密度を $f(x)$ と定義することにより、クラスタリング評価関数 (式 (1)) を得る。

$$D = \sum_{k=1}^K \sum_{n=1}^{N(k)} \int_{-\infty}^{\infty} w_n^k \cdot \varphi(x, \mu_k + \log n) \cdot p_n^k(x) \cdot f(x) dx \quad (1)$$

この評価関数を局所最大化する μ_k, w_n^k を求めることにより、基本周波数が解析される。さらに、同時発音数推定では様々な倍音クラスタ群数の場合の多重音モデルが候補となるため、多数の倍音クラスタ群でのクラスタリングから開始する。ここでは、あるオクターブ範囲内の 12 平均律を基本周波数として倍音クラスタ群を作成するため、倍音クラスタ群の個数 K の初期値は 12 とする。また重み w_n^k は同一倍音クラスタ群において全て等しいこととする。

2.2 音源数推定のための AIC

AIC は $-2 \times (\text{モデルの最大対数尤度}) + 2 \times (\text{モデルの自由パラメータ数})$ で与えられ、この AIC が最小となるモデルを最良のモデルと考える。このとき、パラメータ数は倍音クラスタ群ごとに 2 つの自由パラメータ μ_k, w^k があるので、自由パラメータ総数は $2 \times K$ である。

2.3 音源数とクラスタ重心 μ_k の導出

前節までに述べたハーモニッククラスタリングと AIC を用いて、音源数とクラスタ重心を導出する手順を以下に示す。

ステップ 0:(初期値の設定) あるオクターブ範囲内における 12 平均律音階の基本周波数数値を基本クラスタ重心とする。つまり初期の K は 12 となる。

ステップ 1:($p_n^k(x)$ の更新) $p_n^k(x)$ を算出し、評価関数 D を求める。

ステップ 2:(μ_k と w_n^k の更新) パラメータを以下のように更新し、ステップ 1 に戻る。最尤パラメータが求まったらステップ 3 へ進む。但し、 $F = \int_{-\infty}^{\infty} f(x) dx$ とする。

$$\bar{\mu}_k = \frac{\sum_{n=1}^{N(k)} \int_{-\infty}^{\infty} (x - \log n) p_n^k(x) f(x) dx}{\sum_{n=1}^{N(k)} \int_{-\infty}^{\infty} p_n^k(x) f(x) dx} \quad (2)$$

$$\bar{w}^k = \frac{1}{FN(k)} \sum_{n=1}^{N(k)} \int_{-\infty}^{\infty} p_n^k(x) dx \quad (3)$$

ステップ 3:(AIC の算出) AIC を算出する。AIC が上昇した時点で終了する。

ステップ 4:(倍音クラスタ群数の削減) 重みの最も小さい倍音クラスタ群を削除する。また、2 つの基本クラスタ重心の距離が $(1/12) \log 2$ より小さい時、重みの小さいほうの倍音クラスタ群を削除する。残った K を新たな K としステップ 1 に戻る。

† 東邦大学理学部情報科学科

Department of Information Science, Toho University

3 GPU 上での基本周波数解析の並列処理

2 章で述べたハーモニッククラスタリングによる解析では 3 重ループが複数存在する。さらに AIC が最小となるまで反復計算を繰り返し行わなければならないので計算回数が増大になってしまう。そこで、本章ではこれらの計算を CUDA を用いて並列化し、GPU 上で高速化を図る手法を提案する。NVIDIA Tesla のような GPU では、MP と呼ぶマルチプロセッサの内部に複数の SP と呼ぶプロセッサが搭載され、計算処理を行う。GPU 上での CUDA によるプログラミングにおいては、ブロックとスレッドという単位に分割して並列計算を行う。その際 MP 上ではブロックの計算を行い、MP 内部の SP によってスレッドが実行される。

3.1 フレーム間・フレーム内の並列処理

ハーモニッククラスタリングでは各々のフレーム(短時間の音響信号)間に依存関係がないためフレーム間での並列処理を行うことができる。そこでフレーム間の並列処理ではフレームをブロックごとに分けることとした。ブロックとして取ることでできる最大数は 65535×65535 個であり、音源のサンプリング周波数が 44100Hz でフレーム長 4096、フレームシフトが 128 とした場合でも十分に長い時間の音響信号を扱える。フレーム内の並列処理では、前述の 3 重ループの最内ループに着目した。最内ループの計算回数はフレーム長の半分(ナイキスト周波数分)となる。今回はフレーム長を 4096 にしたため計算回数は 2048 となる。なお、異なる MP 間で同期をとるには処理を一度終了して CPU に制御を戻す必要があるため、性能が低下する可能性が高い[5]。そのため本研究では同一フレーム内での MP 間並列処理は行わないこととした。

3.2 GPU 上で使うメモリの考慮

GPU のデバイスメモリにはレジスタ、ローカルメモリ、シェアードメモリ、グローバルメモリ、コンスタントメモリ、テクスチャメモリが存在する[5]。本手法ではチップ外に置かれて、全体で扱えるグローバルメモリと、チップ上に置かれて、同一ブロック内で扱えるシェアードメモリとした。計算を行うときに最も使われるパワースペクトル $f(x)$ はシェアードメモリに載せた。また、ステップ 2 で計算に用いる $p_n^k(x)$ は、データサイズが大きくシェアードメモリでは扱えないためグローバルメモリに載せることとした。

4 NVIDIA Tesla S2050 上での性能評価

本性能評価では、CUDA GPU アーキテクチャ「Fermi」を元に設計された NVIDIA TESLA C2050 を 4 基搭載している Tesla S2050 を用いる。Tesla C2050 は CUDA コアと呼ばれるプロセッサが $14 \times 32=448$ 基搭載されており、各プロセッサは 1.15GHz の動作周波数で、メモリクロックは 1.5GHz、メモリ転送帯域は 144GB/s となっている。プログラムの開発環境は CUDA を用いた。入力信号はモノラル音響信号を用い、スペクトル解析はサンプリング周波数 44.1kHz、フレーム長を 4096(93ms)、フレームシフト 128(3ms)とし、窓関数は Hamming 窓関数を用いた。STFT の際の FFT は CUDA の標準ライブラリである CUFFT を用いた。フレーム内並列処理の性能評価ではフレーム数を 1、スレッド数を 1, 8, 16, 32, 64, 128, 256, 512, 1024 とした。Tesla S2050 上での実行結果は、図 1 に示すように、1024 スレッドでは 1 スレッドの約 290 倍の速度向上が得られた。次に、フレーム内・間並列処理の性能評価では処理するフレーム数を 28、スレッド数は 1024、使用する SM 数を 1, 2, 4, 7, 14 とした。実行結果は図 2 に示すように、14SM(SM 内は 1024 スレッド)では 1SM(SM 内は

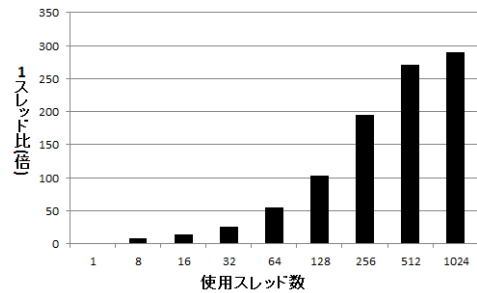


図 1 1SM 上でのフレーム内並列処理 (1 フレーム)。

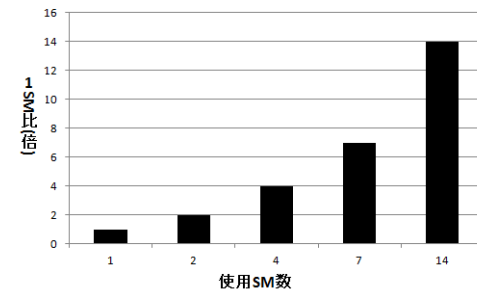


図 2 14SM 上でのフレーム内・間並列処理 (28 フレーム)。

1024 スレッド)と比べて約 13.8 倍の速度向上が得られた。この結果、ハーモニッククラスタリングのフレーム内・間並列処理の有効性が確かめられた。

5 おわりに

本稿では、GPU 上でのハーモニッククラスタリングを用いた基本周波数解析の並列処理手法を提案した。本手法では、ハーモニッククラスタリングを用いた基本周波数解析にフレーム内、フレーム間での並列処理を適用しており、Tesla S2050 上での性能評価から有効性が確認された。今後の課題としては、複数 GPU の有効利用及びメモリ階層やストリームの利用などが挙げられる。

参考文献

- [1] Curtis Roads, 青柳龍也, 小阪直敏, 平田圭二, 堀内靖雄. コンピュータ音楽. 東京電機大学出版局, 2001.
- [2] 亀岡弘和, 西本卓也, 嵯峨山茂樹. ハーモニッククラスタリングによる多重音信号高音抽出における音源数とオクターブ位置推定. 情報処理学会論文誌, No.48, pp.27-32, 2003.
- [3] 亀岡弘和, 西本卓也, 篠田浩一, 嵯峨山茂樹. ハーモニッククラスタリングによる多重音の基本周波数推定. 情報処理学会論文誌, No.82, pp.29-34, 2003.
- [4] 赤池弘次, 甘利俊一, 北川源四郎, 樺島祥介, 下平英寿, 室田一雄, 土屋隆. 赤池情報量基準 AIC-モデリング・予測・知識発見-. 共立出版, 2007.
- [5] NVIDIA CUDA C Programming Guide Version4.2. NVIDIA Corporation, 2012.