

平成27年 3月31日

プログラム名：セレンディピティの計画的創出による新価値創造

PM名：合田圭介

プロジェクト名：細胞同定技術開発

委 託 研 究 開 発

実 施 状 況 報 告 書 (成 果)

平成26年度

研究開発課題名：

セレンディピターにおける光学系と信号処理の最適化

研究開発機関名：

国立大学法人大阪大学

研究開発責任者

堀崎 遼一

[当該年度における計画と成果]

1 . 当該年度の担当研究開発課題の目標と計画

平成 26 年度は、プロジェクト 3 が開発を進める細胞計測部の知見にもとづき、細胞同定法の選定・開発を進める。特にプロジェクト 3 が作成する細胞データベースを用いて細胞同定アルゴリズムの検討をおこなう。

細胞計測部からの出力は画像情報や分光情報から成る多次元の光学情報であることが想定される。多次元信号の認識や識別アルゴリズムは画像認識の分野を中心に多数提案されている。まず、既存の細胞同定アルゴリズムの比較検討をおこなう。選定したアルゴリズムの実装形態をプロジェクト 4 の他のチームと検討する。必要に応じてセレンディピター用に同定アルゴリズムのチューニングをおこなう。細胞同定アルゴリズムの選定後は、本チームは光学実装、他のチームは GPU や FPGA などの高速計算ハードウェア実装を主に担当する。

2 . 当該年度の担当研究開発課題の進捗状況と成果

2-1 進捗状況

セレンディピターを考えた場合、光学実装や高速計算ハードウェア実装にかかわらず並列に実装できることが望ましい。そこで最初の取っ掛かりとして機械学習にもとづく初歩的な識別アルゴリズムである AdaBoost を検討した [Y. Freund , R. E. Schapire, Journal of Computer and System Science, 55, 119-139 (1997)]。AdaBoost は対象データベースに対し異なる特性をもつ複数の弱い分類器を作成し、その多数決で対象の識別をおこなう。演算コストが低く、精度や汎化能力が高いことが知られ、近年は特に実応用環境下で使われることが多いアルゴリズムの一つである。各分類器を並列化できるため高速計算ハードウェアとの親和性が高い。また、分類器の種類によっては光学的な実装も可能である。

当初はプロジェクト 3 が用意する細胞データベースでシミュレーションを行う予定であったが、データベースの準備が遅れたため、物体認識アルゴリズムのベンチマークに広く使われる顔データベースを代用し顔 or 非顔の認識問題を通して AdaBoost の評価をおこなった。

2-2 成果

学習用の顔画像 1000 枚、非顔画像 1000 枚を用いてシミュレーションを行った。相関演算を用いた 100 個の分類器を用いアンサンブル学習をおこなった。相関演算は光学系による実装も可能である [J. W. Goodman, "Introduction to Fourier optics," MacGraw-Hill (1996)]. 2.6GHz の CPU をもつ計算機上で MATLAB を用いてシミュレーションをおこなったところ、学習に 8 分、顔 or 非顔の認識に 10 ミリ秒/画像かか

った。また学習用のデータベースとは別の顔画像 100 枚、非顔画像 100 枚を用いて評価をおこなったところ真陽率と真偽率を合わせた認識の正答率は 93%であった。この結果と光学実装や高速計算ハードウェア実装との親和性を踏まえると、AdaBoost はセレンディピターで用いられる細胞同定アルゴリズムの有力な候補の一つであるといえる。

2-3 新たな課題など

細胞データベースが準備できなかったため、今回は顔データベースでアルゴリズムの評価をおこなった。今後プロジェクト 3 で作られる細胞データベースにもとづきアルゴリズムの選定やチューニングをおこなう必要がある。後段の課題にも対応するがプロジェクト 3 と綿密に連絡を取り合い、細胞計測部を考慮して細胞同定アルゴリズムおよび実装法を最適化し、セレンディピター全体としてのシステムのボトルネック解消に努める必要がある。

今回は二次元の空間データのみでシミュレーションをおこなった。プロジェクト 3 では奥行きを含めた三次元空間データや、蛍光やラマン散乱を含む分光情報データを取得する細胞計測部を開発している。これらの多次元データをどのように取り扱うかも課題のひとつである。各次元の重みを変えることでより効果的に細胞同定をおこなえる可能性がある。

今回検討をおこなった AdaBoost の利用を進める場合は、光学実装や高速計算ハードウェア実装を具体的に詰める必要がある。AdaBoost の高速計算ハードウェア実装は先行研究もありそれほど難しくないと考える。一方、光学実装を用いる場合、例えば相関演算を用いた分類器の光学実装は可能であるが、分類器の数は少ない方が望ましい。一方で認識器の安易な削減は認識率の低下を招く。実装容易性と認識率のバランスを考える必要がある。

検討をおこなった AdaBoost 以外にも多数の認識や識別アルゴリズムが存在する。それらの評価をおこない細胞同定アルゴリズムの選定やチューニングを今後もおこなう必要がある。セレンディピターにおける細胞同定アルゴリズムでは特に認識率と処理速度が重要なファクターとなる。認識率の向上のためには多数の細胞同定アルゴリズムの比較や作り込みをおこなわなければならない。また、処理速度の向上のためにはアルゴリズムの実装形態をプロジェクト 4 の他チームに所属する GPU や FPGA の専門家を交えて議論する必要がある。

3 . アウトリーチ活動報告

なし .