

令和元年6月13日現在

機関番号：12608

研究種目：挑戦的萌芽研究

研究期間：2016～2018

課題番号：16K13783

研究課題名（和文）GPUと深層学習を用いた広視野サーベイのための高精度・高速天体認識技術の開発

研究課題名（英文）Development of high-precision high-speed transient detection method in wide-field survey utilizing GPU and Deep Learning

研究代表者

谷津 陽一（YATSU, YOICHI）

東京工業大学・理学院・助教

研究者番号：40447545

交付決定額（研究期間全体）：（直接経費） 2,900,000円

研究成果の概要（和文）：重力波天文学に代表される時間領域天文学に必須となる天体観測装置のソフトウェア的な基盤技術の開発を行った。現在のロボット観測においてどうしても人間が介在せざるを得ない、観測スケジュールのアレンジと取得画像からの突発天体抽出の完全自動化のために、深層学習やGPUを用いた気象識別と全く新しい突発天体検出アルゴリズムを開発した。これらは、既存のハードウェアをそのまま応用して、多くの観測所で使用することが可能である。また、これらの処理に用いたGPUを従来のデータリダクションに用いることで、30倍の解析時間短縮を実現した。

研究成果の学術的意義や社会的意義

2015年にスタートした重力波観測はすでに多くの重力波イベントを検知しているが、これらから物理的な情報を引き出すためには可視光での追観測が必須である。現状、当番制で国内望遠鏡の監視を継続しているが、この人的な負担は極めて大きい。本研究はこの完全自動化を阻む主要な問題を解決するための要素技術を、安価な市販の計算機で実現するものである。この技術は容易に移植が可能であり、すぐさま重力波監視作業に適用することが国内外の重力波電磁波観測コミュニティから期待されている。

研究成果の概要（英文）：Generic technologies for automating the optical/IR telescopes for the upcoming Gravitational astronomy Era were developed. Currently the robotic telescopes still require humans help for assigning of observation schedule and the final confirmations of transient detections. To reduce the delay of the follow-up observation, we developed a weather recognition algorithm and a transient detection algorithm by using machine learning and image recognition technologies. In addition we also successfully speeded up the data reduction pipeline by a factor of 1/30 by utilizing GPU parallel computing technology.

研究分野：天文学

キーワード：機械学習 画像認識 深層学習 畳み込みニューラルネットワーク 重力波 時間領域天文学 ロボット望遠鏡 GPU

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

(1) 重力波天文学の幕開け

本計画を申請した2015年秋は、重力波天文台 LIGO がまさに観測を開始した頃であり、我々は「インジェクション」というフェイクの重力波アラートに踊らされつつ、半信半疑の中で電磁波対応天体探査の準備を進めている真っ最中であった。当時、重力波現象の発生頻度は全く分かっておらず、年に数回という楽観的なものから、数百年に一度という悲観的なものまで期待されるイベントレートには何桁もの開きがあり、かつ、重力波の到来方向決定精度は数百平方度と、手持ちの望遠鏡で観測を行うのはほとんど絶望的とも思える状況であった。

(2) 計算機技術による装置機能向上の必要性

我々2008年までに、HETE-II や Swift といったガンマ線バースト (GRB) 観測衛星に即応するロボット望遠鏡を開発していたが、重力波と GRB の決定的な違いはその予報位置精度である。Swift の位置決定精度は Burst Alert Telescope で数分角、X-Ray Telescope の情報が得られれば～数秒角の精度に達し、ピクセルスケールでの対応天体探査が可能であった。このため Swift 以降の時代、GRB 観測望遠鏡は視野が狭くても実用上問題が無かったのである。例えば、GW170817 の予報円は 100 平方度を軽く超えており、視野 30'x30' の MITSuME-明野 50cm 望遠鏡でサーベイする場合、およそ 400 ポイントのタイリング観測が必要となる。星の疎な天行きであっても一視野には数千個の恒星が写るため、もはや人間が目視で可視光対応天体を探すのは無理であることが明白であった。

(3) 第三次 AI ブーム・GPU 利用プログラム開発環境の一般化

2010 年以降、SNS 上の画像解析等を中心に深層学習といった機械学習が注目を集め、計算機の性能向上と相まって、顔認識等の物体検知・認識の自動化が実用レベルに達した。これらの技術は開発プラットフォームがハードウェアベンダーから無料で公開されたことで、誰にでも利用できるようになった。この様な、劇的な技術革新の一方で、早くから計算機が利用されてきた天体データの解析には、いまだに IRAF が広く使われている。IRAF は 30 年前の計算機でも動く様に設計されており、要求リソースが極めて小さい反面、現在の計算機の性能をフル活用することができない。実際、IRAF はすでにほとんどメンテナンスがなされておらず、時間領域天文学を標榜する大規模なプロジェクトでは、専用の解析プログラムを独自に開発している。

2. 研究の目的

我々は、計算機・ネットワーク・IoT 技術が切り開いた新たな発見領域である「時間領域天文学」に挑むため、既存の観測システムのうち、計算機とソフトウェアの努力により来るべき重力波天文学の時代に対応することを目標とした。高コストな望遠鏡本体はできるだけ既存の装置を流用し、ムーアの法則に従って劇的な性能向上を続ける最新の計算機技術を取り込むことで、費用対効果の高い、汎用性のある技術開発を行うこととした。

解決すべき喫緊の課題は、「重力波天文学への対応」であるが、これはすなわち即時観測・即時解析の実現である。現状の観測システムにおいて、観測効率を著しく制限しているのは人間の介在であり、天候などを加味した観測計画の立案 (ターゲットのアレンジ) や、最終的な画像チェックはどうしても人間に頼らざるを得ないために、これらの作業が多大な労力を要求する上に、世界の自動望遠鏡群に出遅れる要因にもなっている。また、大量のデータを処理する処理時間そのものも本格的な観測が始まれば問題になる。これらを解決するために、我々は以下 3 つの具体的な課題を立てた：

(1) 気象識別器の開発

J-GEM の様に、日本全国の望遠鏡が協力して重力波対応天体を観測する際、どこで何を観測するかが重要である。日本は特に気象状態が安定しないため、観測条件の程度を評価できているだけでも、観測効率を劇的に改善することができる。本研究では、全国の観測所にすでに設置されている気象モニターの可視光・カラー画像をそのまま使える気象識別器 (雲の識別器) を開発し、観測計画の自動立案と高効率化を行う。

(2) 突発天体の自動検知

遠方の天体は基本的に点光源として見えるが、この識別が案外難しく、最終的には人間が確認している状態である。これを可能な限り省力化して最低でも人間の判断をサポートできるような識別技術を確認する。先行研究として すばる HSC の機械学習による超新星サーベイなどの研究があるが、本研究では、すばるほど完璧なメンテナンスのされていない中小望遠鏡に適用でき、参照画像がカタログデータであっても機能するような汎用の技術を目指す。

(3) GPGPU を用いたデータ処理の高速化

上記の画像解析でも用いる画像演算装置をデータリダクションに活用し、10 倍以上の高速処理を実現する。

3. 研究の方法

(1) 気象識別器

従来の気象識別の方法としては、中間赤外線照度計もしくは中間赤外線撮像装置をもちいて直接雲からの放射を検知し雲量を計測してしまう方法挙げられる。照度計の場合、空間分解能がないため、局所的な雲には対応できない。一方、撮像装置の場合は特殊な光学系が必要であ

り、極めて高額になる。カラー画像を用いた方法では、色情報による識別が試みられているが、月齢、気象条件、時間などでも空の色が変化するため、機械的な閾値設定は困難であった。そこで、本研究では画像認識で高い精度を達成している、Deep Learning を応用して、気象識別を実現することとした。

(2) 突発天体の自動検知

先行研究(すばる HSC の例)では、点源の検出を Source Extractor で行い、明るさはサイズなど十数個のパラメータ特徴量として Deep Learning で識別するという構成となっていた。また、情報量を削減するために事前に取得した画像と新規に取得した画像の差分を用いる方法も用いられる。この差分計算は2枚の画像の PSF と湾曲収差などが完全に一致している必要があり、同一望遠鏡での比較でない場合、なかなか成功しないという問題がある。世界的に広く用いられている最もよくできたソフトウェアの一つが HOTPANTS であるが、このソフトウェアでは PSF を再現するカーネル関数として二次元ガウシアン関数を仮定しているが、光学系の収差や重力によるたわみ、回折パターンの影響など、実際の観測ででてくるさまざまな状況には完全に対処できていない。

本研究では、PSF を説明するカーネル関数を人間が指定することなく、CNN に直接推定させてしまうことにした。これにより、二次元ガウシアン関数では説明のできないイレギュラーな PSF/星像パターンに対応できるようにした。

(3) GPGPU を用いたデータ処理の高速化

明野望遠鏡でもっとも処理時間が長いのはリダクションの処理、なかでも FLAT フレームの作成など、ファイル枚数の多い処理である。まずは、現行のシステムがどのような部分で時間を消費しているのかを明確にし、効率的に処理時間を短縮する方法を検討し、その結果を反映して現行のハードウェアの性能を最大限活かす様なテストプログラムを作成し、性能検証を行うことにした。

4. 研究成果

(1) 気象識別器

【既存のアルゴリズムを用いた試作】

最初のステップとして、民生品のカメラで撮像した可視カラー画像を使って気象識別機の開発を行った。識別クラスは、晴天域、厚雲域、薄雲域、月の4クラスとし、明野観測所に設置した全天カメラの過去数年分のデータから無作為に抽出した 1000 枚画像に対して目視にてクラス情報を付与した教師データを作成した。最初に試した識別器は 25 層(1 億 3 千万パラメータ)の CRF-RNN であり、2016 年当時、画像識別のコンペティションにおいて最高性能を達成したネットワーク構成である。上記教師データを用いた最適化により、月や水滴等の特定条件を除いて 80%を超える識別精度を達成した。しかしながら、上記の試作アルゴリズムは、高い精度と引き換えにハイエンドの GPU と数十ギガバイトのビデオメモリが要求され、容易に移植できるものではなかった。

【雲検知に機能を絞った高速・軽量版】

より実用的にするため、年度後半ではこれと同程度の性能を保った軽量の識別器の設計・評価を行った。提案した手法では CNN の層数を 5 層(650 万パラメータ)としたが、同一の学習データを用いた最適化により、25 層の識別器と同等以上の精度で気象を識別することに成功した。最終的な識別精度は 93%に達し、人間と同程度の精度が得られたと結論する。これにより、目標として掲げた晴天域の自動識別は達成できたといえる。複雑なモデルの精度が簡略モデルに劣る理由については、パラメータ数に対して学習データが十分ではなく、局所解に陥っていたためだと予想している。また、現状の気象識別精度の上限は我々が作成した学習データ側に制限されていると考えている(図 4.1)。この学習データは複数の作業者が作成したものであり、定量的な判定論理に従って作成されたものではない。このため、薄雲などの透過性かつ濃淡のある曖昧な被写体については、人間であっても統一的な解釈がむずかしく、判断にばらつきが発生していた。将来的には気象識別機と連動した天体観測を実施することで、より精度の高い

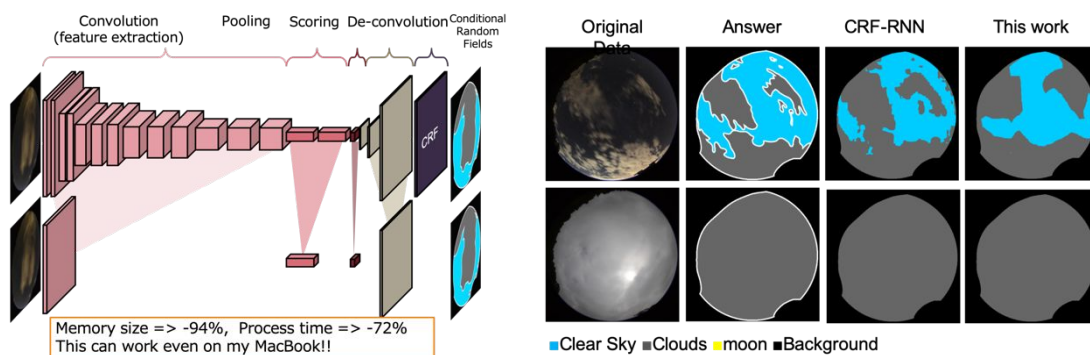


図 4.1: (左)開発したニューラルネットワークの概略図。上が CRF-RNN 下が新規に開発した単純化したネットワーク。(右)識別結果

学習データを自動生成できると考えている。

【他観測への移植性の検証】

また、この識別器の移植性については MITSuME 明野望遠鏡の他、木曽観測所の全天カメラやパロマ山天文台の全天画像等でも適用を試みた。結果として、この気象識別器は色情報に極めて敏感に反応していると考えられ、異なるカラーバランスを持つカメラを使ったり、周辺環境光が異なったりする場合において、正確な識別が困難になることが判明した。したがって、この識別器を他の観測所に移植する場合は、それなりの数のトレーニングデータを用いて最適化を行う必要がある。[1]

(2) 突発天体の自動検出

これまで、時間変動する天体の抽出のために、撮像画像とテンプレートの引き算画像に対して点源探索を行っていたが、望遠鏡の姿勢などに起因する光学特性の変化により、この減算処理で引き残しが生じることが多々あった。デファクトスタンダードとして利用されている Hotpants に対し、我々は PSF の評価に畳み込みニューラルネットワーク(CNN)を利用することにした。入力として CNN に渡す情報は、突発天体発生前のテンプレート画像 (template) と最新の画像(science)であり、突発天体の周辺領域のみ 20x20pixel 領域を抜き出し、20x20x2 のテンソルとして入力される。組み合わせる CNN は畳み込み 2 層、プーリング 2 層、全結合 2 層とし、最終出力は入力された画像セットが突発天体であるか否かの確率である (図 4.2)。

学習画像として、我々は MITSuME 明野望遠鏡で取得した実画像を 5074 枚 (うち Positive 2488 枚、Negative 2586 枚) 用いてニューラルネットワークの最適化を行った。最適化はおよそ 60epoch で収束した。この CNN を用いた画像比較の性能を示したのが図 4.3 である。比較対象として HOTPANTS による差引き結果と、我々の開発した CNN 突発天体識別器の認識している差引き結果を示す (この識別器は変動天体が居るかどうかを確率で出力するため、閾値を 0.35 とし、それ以下の場合は対象のある領域を背景レベルで補完することで、HOTPANTS との比較としている)。この識別成功率は当然ターゲットの明るさに依存し、S/N>10 つまり測光エラーで 0.1 等級程度の明るいターゲットに対しては 90%以上の精度に到達する。[2]

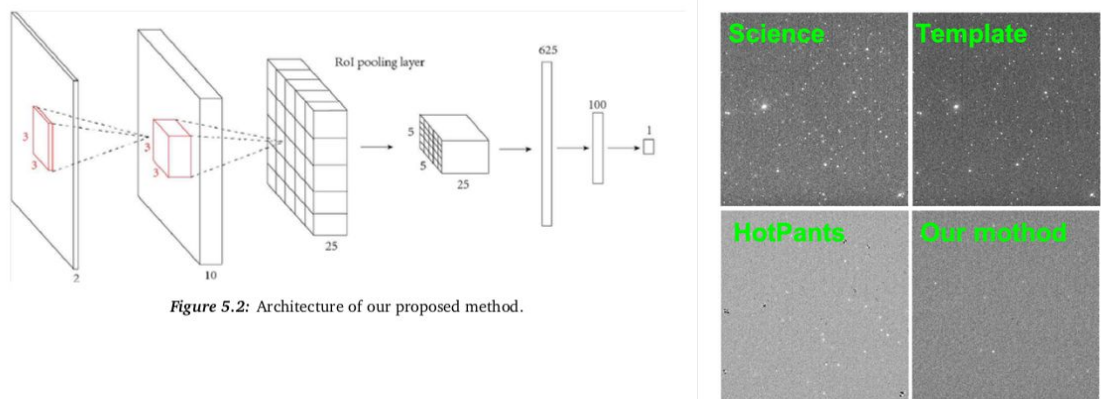


図 4.2: (左) 開発した識別器のネットワーク構成。(右) HOTPANTS との性能比較。

(3) GPGPU を用いた処理高速化

現行の一次処理パイプラインの処理の流れを図 4.3 (左) に示す。IRAF は RAM の使用率を抑えるためか、処理後の結果を一時ファイルとして保持する。これを処理ステップの回数だけ繰り返して行くため、ファイル I/O が全体の処理時間の大部分を占めていることがわかった。現在となっては無駄になってしまったこの処理を廃止し、代替する新しい一次処理スクリプトを作成した。これは IRAF を利用しないことで余計な I/O を省略した。さらに、ピクセル対ピクセルのような単純な演算を繰り返す処理については CuPy を利用して GPU 上で並列演算するようにして、高速化を図った。現行の処理プログラムは Python で実装しているため、まだ

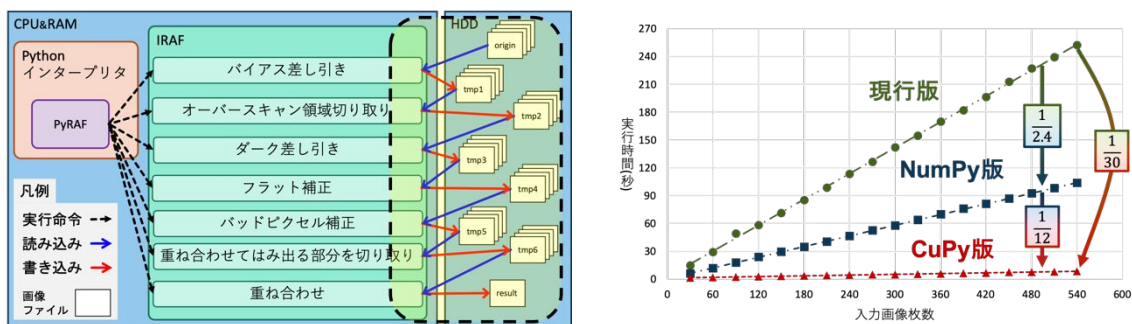


図 4.3: (左) 1次処理パイプラインの流れ。(右) 処理時間の短縮効果。

まだ改善の余地はあるが、結果として IRAF の 30 倍以上の高速化を達成した。この効果は絶大で、これまでフラットの作成にはおよそ 10 分程度の時間が掛かっており、測光精度の要となるにもかかわらず再解析がためらわれたが、この高速化により 20 秒で完了できるようになった。[3]

5. 主な発表論文等

[雑誌論文](計 0 件)

[学会発表](計 15 件)

[1] Y.Yatsu, 「Development of a cloud detection system utilizing image recognition technology」, 5th workshop on robotic autonomous observatories (2017 年 10 月)

[2] Y.Yatsu, 「Image subtraction and Transient Detection with CNN」, Growth Conference 2018@IITB, (2018 年 12 月)

[3] 谷津陽一、庭野聖史ほか, 「突発天体検知のための MITSuME 望遠鏡用自動解析パイプラインの GPU を用いた高速化」, 日本天文学会 2019 春季年会@法政大学 (2019 年 3 月)

[産業財産権]

出願状況(計 2 件)

名称: 画像処理装置

発明者: 谷津陽一他

権利者: 国立大学法人東京工業大学

種類: 特許

番号: 特願 2018-139611

出願年: 2018

国内外の別: 国内

名称: 画像処理装置

発明者: 谷津陽一

権利者: 国立大学法人東京工業大学

種類: 特許

番号: 特願 2018-139620

出願年: 2018

国内外の別: 国内

[その他]

ホームページ等

6. 研究組織

(1) 研究分担者

研究分担者氏名: 篠田 浩一

ローマ字氏名: SHINODA KOICHI

所属研究機関名: 東京工業大学

部局名: 情報理工学院

職名: 教授

研究者番号(8桁): 10343097

研究分担者氏名: 井上 中順

ローマ字氏名: INOUE NAKAMASA

所属研究機関名: 東京工業大学

部局名: 情報理工学院

職名: 助教

研究者番号(8桁): 10733397

研究分担者氏名: 下川辺 隆史

ローマ字氏名: SHIMOKAWABE TAKASHI

所属研究機関名: 東京大学

部局名: 情報基盤センター

職名: 准教授

研究者番号(8桁): 4063049

(2) 研究協力者

研究協力者氏名: 河合 誠之

ローマ字氏名: KAWAI NOBUYUKI

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。