

Д. С. Ніколаєнко, Т. В. Філімончук, Г. В. Майстренко

Харківський національний університет радіоелектроніки, Харків, Україна

МОДЕЛЬ ОБРОБКИ СПЕКТРАЛЬНИХ ДАНИХ АСТРОНОМІЧНИХ ОБ'ЄКТІВ

Анотація. Актуальність. На даний час існує напрям, який пов'язано з оптимізацією обробки даних астрономічних об'єктів, зокрема спектральних даних, що отримуються за допомогою радіотелескопів. Об'єм даних, котрий необхідно обробити, представляє собою однорідний потік інформації, який можна обробляти за допомогою розподілених обчислень, таким чином збільшивши швидкість обробки на доволі значну величину. **Метою даної роботи** є побудова моделі обробки спектральних даних астрономічних об'єктів з використанням розподілених обчислень. **Об'єктом дослідження** є процес обробки спектральних даних, які було отримано від астрономічних об'єктів. **Предметом дослідження** є існуючі на даний час моделі та методи обробки спектральних даних. **Результати.** Запропонована модель обробки спектральних даних астрономічних об'єктів, яка рекомендує використовувати розподілені алгоритми обробки та спеціалізоване програмне забезпечення, а саме середовище CUDA та бібліотеку OpenCV для зменшення часу обробки великих масивів даних та більш раціонального використання потужностей обчислювальних ресурсів. **Висновок.** Використання запропонованого інструментарію у моделі дозволяє покращити та пришвидшити обробку спектральних даних астрономічних об'єктів за допомогою розподілених алгоритмів.

Ключові слова: спектральний аналіз, паралельні та розподілені обчислення, обробка великих даних, швидкодія.

Вступ

Людство завжди прагнуло дослідити та проаналізувати небо, виявити закономірності, явища небесних об'єктів, дослідити самі об'єкти. З плином часу прості дослідження неба стали повноцінною наукою – астрономією. Астрономія вже давно перестала бути виключно наукою оптичних спостережень. Минуле століття дало великий поштовх розвитку всіх наукових сфер, які використовуються в наш час як інструментарій для досліджень неба та його явищ: фізика, хімія, географія, комп'ютерні технології.

Створення радіотелескопу в середині 20-го століття дало великий поштовх розвитку астрономії, так як тепер науковці мають змогу бачити випромінювання об'єктів не тільки в оптичному діапазоні, а й в інших. Проте, науковці стикаються з іншою проблемою: обробкою даних, які було отримано. Одними з типів даних, які отримуються з радіотелескопів є спектральні дані. У зв'язку з цим, сучасна астрономія все більше спирається на передові комп'ютерні технології та інноваційні методи обробки даних [1, 2]. Використання графічних процесорів (GPU) стало революційним кроком у цьому напрямку. GPU дозволяють проводити обчислення з високою ступенем паралелізму, що кардинально змінює підходи до обробки великих обсягів даних, отриманих від небесних спостережень. На даний час для оптимізації процесу обробки спектральних даних використовуються комп'ютерні технології [3]: науковці досліджують існуючі моделі, програми та алгоритми обробки, намагаючись створити новий, більш оптимізований підхід до їх аналізу.

Метою роботи є побудова моделі обробки спектральних даних астрономічних об'єктів за допомогою розподілених обчислень.

Результати досліджень

Актуальні засоби обробки. На даний час найпопулярнішим середовищем для аналізу та обробки спектральних даних астрономічних об'єктів є середовище Image Reduction and Analysis Facility (IRAF)

[4], яке було створено у 80-х роках для швидкої та оптимальної обробки даних, отриманих від астрономічних об'єктів. Мовою програмування середовища IRAF є Python, а для математичних обчислень використовується бібліотека Astropy [5]. Середовище являє собою потужний програмний пакет, розроблений для використання в астрономії та астрофізиці, призначений для полегшення роботи з даними, та надає зручні інструменти для виконання широкого спектру задач: від базових обчислень та обробки зображень до більш складних операцій, таких як управління астрономічними координатами, часом та моделюванням.

Бібліотека Astropy надає розширені можливості для аналізу спектральних даних, що є невід'ємною частиною сучасних астрономічних досліджень. Вона дозволяє вченим легко обробляти, аналізувати та інтерпретувати спектральну інформацію, яка була отримана в наслідок астрономічних спостережень.

Важливим фактором використання IRAF та Astropy є можливість зчитувати та обробляти формат Flexible Image Transport System (FITS), який є найбільш популярним для збереження характеристик сигналу з радіотелескопів для подальшого опрацювання. FITS являє собою відкритий стандарт, який визначає цифровий формат файлу корисний для зберігання, передачі та обробки даних: у форматі багатовимірних масивів (наприклад, 2D-зображення) або таблицях. Слід зазначити, що FITS є найбільш поширеним форматом файлів, який використовується в астрономії і містить такі дані: опис фотометричної та просторової калібрувальної інформації разом з оригінальними метаданими зображення.

Формат IRAF підтримується до сих пір, проте він використовує послідовні алгоритми обробки на процесорі. Такий підхід накладає значні обмеження на швидкодію обробки даних, особливо в контексті “любительської” астрономії. Велика потужність потребується від комп'ютерної системи, щоб обробляти такий великий потік даних в реальному часі. Обробка одного файлу FITS на доволі посередніх комп'ютерах може зайняти більше кількох хвилин,

що є доволі довгим процесом, враховуючи те, що один такий файл містить дані про об'єкт в конкретний момент часу. Головною задачею моделі, що пропонується, є прискорення швидкості обробки даних у форматі FITS за рахунок максимально швидкого проходження по масиву спектральних даних з застосуванням до кожного елементу математичні функції (аналогічно бібліотеці *Astropy*) з використанням розподілених обчислень.

Технологічна база. Складність обчислювань вимагає різкого збільшення ресурсів та швидкодії комп'ютерів. Найбільш перспективним напрямком збільшення швидкості розв'язання є широке впровадження ідей паралелізму у роботу обчислювальних систем. Сьогодні спроектовано та випробувано сотні різних комп'ютерів, що використовують у своїй архітектурі той чи інший вид паралельної обробки даних. Основна складність при проектуванні паралельних програм – забезпечення правильної послідовності взаємодій між різними обчислювальними процесами, а також координація ресурсів, що розділяються між ними. Для прискорення обробки масиву спектральних даних було обрано найпопулярнішу програмну архітектуру від *Nvidia CUDA* [6]. *CUDA* – це програмно-апаратна архітектура паралельних обчислень, яка дозволяє істотно збільшити обчислювальну продуктивність завдяки використанню графічних процесорів *NVIDIA*. При використанні цієї технології слід орієнтуватися на таку термінологію:

- пристрій (*device*) – сама відеокарта, графічний процесор (*GPU*), який виконує команди центрального процесора;

- хост (*host*) – центральний процесор (*CPU*), який запускає різноманітні завдання на пристрої, виділяє пам'ять;

- ядро (*kernel*) – завдання, що буде виконуватися на *GPU*.

CUDA дозволяє програмістам реалізувати на спеціальному спрощеному діалекті мови програмування *C* алгоритми, що виконуються на графічних процесорах *Nvidia*, і включати спеціальні функції в текст програми на *C*. Архітектура *CUDA* дає розробнику можливість на свій розсуд організувати доступ до набору інструкцій *GPU* й управляти його пам'яттю. Дана технологія підтримує декілька мов програмування. Серед них є *Java*, *Python*, *C/C++*.

Для запуску розподіленого алгоритму обробки спектральних даних слід використовувати наступні інструменти:

- бібліотеку *OpenCV* відкритого коду, яка призначена для надання спільної інфраструктури для застосунків комп'ютерного зору та прискорення використання перцепції машин у комерційних продуктах;

- середовище *Microsoft Visual Studio*;

- мову програмування *C++*.

Бібліотека *OpenCV* містить в собі алгоритми обробки даних, які орієнтовано на розподілені обчислення з подальшим використанням їх у середовищі *CUDA*. Також слід зазначити, що бібліотека містить математичні утиліти, аналогічні *Astropy*.

Модель обробки спектральних даних. На даний час існує модель обробки спектральних даних,

яка може бути описана за допомогою кортежу (1), де *D* – датасет спектральних даних, *S* – програмне забезпечення, яке приймає участь в обробці датасету даних, *SA* – послідовний алгоритм обробки датасету. Наведена модель має низку недоліків і головними із них – це використання застарілого програмного забезпечення та бібліотек котрі не містять в собі змогу обробляти дані розподілено.

$$MP_{SD} = \{D, S [IRAF, Astropy], SA\} \quad (1)$$

Щоб уникнути зазначених недоліків, запропоновано нову модель (2), яка рекомендує використовувати розподілені алгоритми обробки (*DA*) та спеціалізоване програмне забезпечення, а саме середовище *CUDA* та бібліотеку *OpenCV*, які дозволяють значно покращити обробку спектральних даних астрономічних об'єктів за рахунок використання розподілених обчислень.

$$MP_{SD} = \{D, S [CUDA, OpenCV], DA\} \quad (2)$$

У порівнянні з послідовним алгоритмом, де дані обробляються послідовно один за одним, розподілений дозволяє реалізувати одночасну обробку, знижуючи загальний час обчислень та підвищуючи продуктивність. Таким чином, нова модель забезпечує більш ефективне використання обчислювальних ресурсів, відкриваючи шлях до більш глибокої та швидкої обробки астрономічних даних.

Алгоритм обробки датасету спектральних даних складається з таких кроків (рис. 1):

- 1) в проєкт реалізації необхідно додати *OpenCV* бібліотеку;

- 2) модель отримує дані для обробки;

- 3) використовуючи паралельні та розподілені обчислення здійснюється обробка спектральних даних об'єкта (бібліотека *OpenCV* містить методи для роботи з форматом *FITS*);

- 4) частину результатів доцільно наводити у вигляді графіків (частотний спектр) іншу слід подати у текстовому форматі (характеристика хвиль);

- 5) запис даних у будь-який текстовий документ чи створення власного формату збереження для відтворення у програмі.

Алгоритм роботи моделі, яка орієнтована на розподілені обчислення, використовує такі методи бібліотеки *OpenCV*:

- методи перетворення Фур'є (*FFT*): дозволяють аналізувати спектральні компоненти даних. Використання *FFT* в *CUDA* може значно прискорити обчислення за рахунок паралельної обробки даних;

- фільтрація Гауса (*Gaussian Blurring*): використовується для згладжування зображення шляхом застосування Гаусового ядра. Цей метод можна за-



Рис. 1. Алгоритм роботи моделі, яка орієнтована на розподілені обчислення

стосувати для попередньої обробки спектральних даних, щоб зменшити шум та підготувати дані для подальшого аналізу (реалізація Гаусового згладжування в CUDA може значно покращити швидкість обробки великих наборів даних);

- краєве виявлення (Edge Detection): методи, такі як оператор Собеля, можуть використовуватися для виявлення країв астрономічних об'єктів в спектральних даних (паралельна обробка даних за допомогою CUDA може забезпечити швидке виявлення країв в великих наборах даних);

- морфологічні операції: включають дилатацію та ерозію, які можуть використовуватися для корекції форми астрономічних об'єктів у спектральних даних (ці методи можуть допомогти покращити виділення та аналіз структур, роблячи їх більш виразними для подальшого дослідження);

- нормалізація: вона допомагає згрупувати дані з подібним діапазоном значень, що є критично важливим для порівняльного аналізу різних спектральних зображень);

- паралельна обробка даних: розбиття даних на менші блоки для паралельної обробки дозволяє максимально використовувати обчислювальні ресурси GPU. Важливо розробити алгоритм таким чином, щоб він міг ефективно масштабуватися залежно від кількості доступних потоків.

Під час реалізації алгоритмів обробки даних на CUDA важливо ефективно використовувати пам'ять GPU, щоб мінімізувати затримки доступу до даних. Використання різних рівнів кешування та розумне управління пам'яттю може значно підвищити продуктивність.

Припустимо, що маємо дві моделі, одна з яких робить послідовну обробку, інша – розподілену та має 10 ядер, доступних для обробки. Тоді, якщо на

обробку одного файлу уходить одна хвилина, то послідовна модель буде обробляти один файл за хвилину, коли паралельна – 10 файлів за хвилину.

Висновки

В результаті проведених авторами досліджень було розглянуто актуальні методи та інструменти для аналізу спектральних даних астрономічних об'єктів, приділено увагу використанню середовища IRAF та бібліотеки Astropy, які залишаються важливими інструментами в для ефективного аналізу даних. Також було наголошено про актуальність паралельних та розподілених обчислень на платформі CUDA, що відіграють критичну роль в оптимізації обробки великих обсягів даних, отриманих з радіотелескопів.

На основі аналізу існуючих підходів та інструментів було обґрунтовано необхідність використання інструментів, які реалізують можливості паралельних обчислень. Запропонована модель обробки спектральних даних відкриває нові горизонти для підвищення ефективності обробки спектральних даних, які отримано з радіотелескопів та збережено у форматі FITS. Використання паралельних обчислень дозволяє значно зменшити час, необхідний для обробки великих даних, що є ключовим для прискорення наукових досліджень. Рекомендована модель, що ґрунтується на паралельних обчисленнях, пропонує ефективний спосіб вирішення цієї задачі. Вона не тільки сприяє підвищенню швидкості обробки даних, але й відкриває нові можливості для глибшого розуміння структури Всесвіту та динаміки небесних явищ.

Майбутні дослідження в цій галузі можуть включати подальше вдосконалення алгоритмів паралельної обробки, розширення функціоналу інструментів, що використовуються, а також адаптацію моделі для нових областей застосування в астрономії.

СПИСОК ЛІТЕРАТУРИ

1. Подорожняк А.О., Гриб Р.М., Домнін С.В. (2013), "Морфологічна обробка цифрових зображень з телескопів", Сучасна спеціальна техніка, № 1(32), С. 34-39.
2. Бандурка О.І., Свинчук О.В. (2022), "Метод ідентифікації космічних знімків для прогнозування лісових пожеж", Системи управління, навігації та зв'язку. Збірник наукових праць, Т. 1 (67). С. 13-18. doi:https://doi.org/10.26906/SUNZ.2022.1.013.
3. Анисенко О.В. (2017), "Розвиток дистанційного зондування землі в Україні", Агросвіт, № 7, С. 52-59.
4. IRAF 2.17.1. IRAF Community Distribution. URL: <https://iraf-community.github.io>
5. Astropy. URL: <https://www.astropy.org>
6. CUDA Toolkit - Free Tools and Training. NVIDIA Developer. URL: <https://developer.nvidia.com/cuda-toolkit>

Received (Надійшла) 26.02.2024

Accepted for publication (Прийнята до друку) 24.04.2024

MODEL FOR PROCESSING SPECTRAL DATA OF ASTRONOMICAL OBJECTS

Dmytro Nikolaienko, Tetiana Filimonchuk, Halyna Maistrenko

Abstract. Topicality. Currently, there is a direction that is related to the optimization of data processing of astronomical objects, in particular, spectral data obtained with the help of radio telescopes. The amount of data that needs to be processed is a homogeneous flow of information that can be processed using distributed computing, thus increasing the speed of processing by a fairly significant amount. **The purpose of this work** is to build a model for processing spectral data of astronomical objects using distributed computing. **The object of research** is the process of processing spectral data obtained from astronomical objects. **The subject of the study** is currently existing models and methods of spectral data processing. **Results.** A model for processing spectral data of astronomical objects is proposed, which recommends the use of distributed processing algorithms and specialized software, namely the CUDA environment and the OpenCV library to reduce the processing time of large data sets and more rational use of computing resources. **Conclusion.** The use of the proposed tool in the model allows to improve and speed up the processing of spectral data of astronomical objects using distributed algorithms.

Keywords: spectral analysis, parallel and distributed computing, big data processing, high-speed code.