

УДК 004.75, 004.932

ВИКОРИСТАННЯ РОЗПОДІЛЕНИХ ОБЧИСЛЕНЬ ДЛЯ СЕГМЕНТАЦІЇ ДАНИХ ДИСТАНЦІЙНОГО ЗОНДУВАННЯ ЗЕМЛІ

Ю. Шийка, Р. Шувар

*Львівський національний університет імені Івана Франка
вул. Драгоманова, 50, 79005 Львів, Україна
shiyka@mail.lviv.ua*

Розроблено проект сегментації зображень дистанційного зондування Землі з використанням розподілених обчислень. Розглянуто систему управління розподіленими обчисленнями Condor. Описано принципи роботи системи та наведено особливості розробки програм опрацювання зображень для використання в проекті розподілених обчислень. Для сегментації зображень дистанційного зондування Землі використано модифікацію алгоритму Краскала з урахуванням кольорових і текстурних характеристик. Проаналізовано ефективність роботи реалізованого проекту розподіленого опрацювання зображень.

Ключові слова: розподілені обчислення, Condor, сегментація зображень, алгоритм Краскала, дистанційне зондування Землі.

Опрацювання великих обсягів растрових зображень дистанційного зондування Землі високої роздільної здатності зумовлює необхідність використання для цього високопродуктивних обчислювальних систем. Одним з можливих методів побудови високопродуктивної системи опрацювання даних дистанційного зондування Землі є організація розподілених обчислень. Системи розподілених обчислень дають змогу об'єднувати обчислювальні ресурси окремих вузлів для виконання спільного завдання, їх щораз частіше використовують у подібних задачах [1].

Виділення і пошук характерних областей і окремих об'єктів космічних знімків є одним з важливих завдань опрацювання даних дистанційного зондування Землі. Для цього застосовують алгоритми сегментації зображень за певними параметрами. Висока роздільна здатність доступних сьогодні космічних знімків дає змогу використовувати як критерії сегментації не тільки інформацію про колір пікселів, а й текстурні характеристики частин зображення [2]. Важливим завданням є розробка методів визначення текстурних характеристик та вибір оптимальних наборів характеристик і їхніх вагових коефіцієнтів для алгоритму сегментації.

Для організації розподілених обчислень ми використали систему Condor, яка дає змогу об'єднати ресурси обчислювальних вузлів завдяки розподілу й управлінню виконанням окремих завдань на обчислювальних вузлах. Розподілені обчислення полягають в

опрацюванні незалежних між собою підзавдань для отримання одного кінцевого результату.

Система Condor – це спеціалізована система управління навантаженням обчислювальних вузлів для ресурсноємних задач. Вона охоплює організацію черги завдань, систему обміну даними, систему моніторингу та управління ресурсами, дає змогу визначати політики планування та схеми пріоритетів завдань [3].

Для виконання розподілених обчислень під керуванням системи Condor потрібно:

- підготувати вхідні дані для окремих завдань (вручну чи за допомогою програми генератора завдань);
- вибрати середовище (набір умов), у якому потрібно виконувати завдання. Забезпечити доступ до вхідних і вихідних файлів завдань тим чи іншим способом (NFS, Condor FileTransfer тощо);
- створити виконавчий файл програми, який повинен виконуватися на робочих вузлах відповідно до вимог системи Condor та встановленої операційної системи;
- створити файл опису завдання Condor з параметрами виконання, заданими вхідними і вихідними файлами тощо;
- додати завдання в чергу Condor за допомогою команди `condor_submit`;
- після виконання завдань отримати вихідні файли (згідно зі способом виконання завдань) та в разі потреби виконати програму опрацювання результатів.

Відповідно до цих вимог реалізації розподілених обчислень під керуванням Condor, створено проект розподіленої сегментації растрових зображень даних дистанційного зондування Землі, який охоплює такі програмні модулі.

1. Модуль формування незалежних підзавдань (`work_generator`). Набір незалежних підзавдань формується з вхідного зображення нарізанням за принципом мозаїки на менші зображення підзавдань. Крім того, до кожного підзавдання створюють файл з інформацією про розміри і позицію зображення підзавдання у вхідному зображенні, а також параметри для виконавчої програми. Тобто для коректного опрацювання та наступного формування кінцевого результату для кожного підзавдання створюють два вхідні файли: перший – файл зображення підзадачі, другий – файл параметрів та індексації, який містить параметри для виконавчої програми, індекси зображення підзавдання (для визначення його місця у початковому зображенні) та параметри нарізання (розміри зображень підзавдання для коректного формування кінцевого результату).
2. Виконавча програма опрацювання незалежних завдань у розподіленому режимі (`distributed_executable`) (виконується на обчислювальних вузлах). Розроблена модульна структура проекту розподілених обчислень дає змогу виконувати різні завдання і використовувати різні алгоритми опрацювання зображень. Зокрема, реалізовано виконавчу програму сегментації зображень на підставі модифікації алгоритму Краскала. Унаслідок виконання програми за двома вхідними файлами (зображення і параметри) утворюються три вихідні файли: перший – сумарне сегментоване зображення підзавдання, другий – містить індекси та розміри зображень підзавдань, третій – звіт про сегментацію та детальну інформацію про утворені сегменти для постпроцесингу.
3. Модуль формування кінцевого результату (`assimilator`) після виконання всіх підзавдань утворює зображення, зібране з результатів виконання окремих підза-

вдань. Крім того, утворюється об'єднана інформація про сегменти для постпроцесингу.

4. Модуль постпроцесингу (postprocess). Він необхідний тому, що в зображенні результату на стиках зображень підзавдань сегменти, які мали б бути суцільними, часто є розбиті. Унаслідок постпроцесингу такі сегменти об'єднуються, використовуючи інформацію про характеристики сегментів з вихідних файлів підзавдань, утворених виконавчою програмою. Постпроцесинг відбувається за інформацією про властивості та характеристики суміжних сегментів, утворених унаслідок сегментації незалежних підзавдань.

5. Для ефективної роботи з файлами растрових зображень великих розмірів розроблено модуль читання, записування й інших необхідних операцій, який дає змогу працювати з окремими областями растрових зображень. Це допомагає опрацювати зображення необмежених розмірів, оскільки дає змогу використовувати обчислювальні ресурси (головно це стосується оперативної пам'яті) порційно.

Для додання набору підзавдань у чергу завдань Condor створено файл опису завдань Condor, у якому зазначено вибрані параметрами та середовище виконання завдань, шлях та вигляд вхідних і вихідних файлів, параметри доступу і пересилання файлів між обчислювальними вузлами тощо:

```

Executable =      condor_segment      //виконавчий файл завдань
Universe     =      vanilla             //середовище виконання завдань

Error        =      log/err.segment.%(Process)//звіти про помилки
Input        =      in/$(Process)_1     //вхідні файли завдань (stdin)
Output       =      out_$(Process)_1    //вихідні файли завдань (stdout)
Log          =      log/condor_segment.log //звіти про виконання завдань

Arguments    =      $(Process)          //аргументи для виконавчої програми
Should_transfer_files = YES             //параметри пересилання файлів
When_to_transfer_output = ON_EXIT
Transfer_input_files = in/$(Process)_0, in/$(Process)_1
Queue 162    //помістити завдання в чергу

```

У виконавчій програмі сегментації растрових зображень дистанційного зондування Землі ми використали модифікацію алгоритму Краскала, суть якого полягає в побудові мінімального остовного дерева заданого графа [4, 5] (об'єднання всіх вершин графа в одну область мінімальною сумарною довжиною ребер так, щоб існував шлях між усіма вершинами). Для сегментації растрових зображень застосовано модифікацію цього алгоритму, у якій на початковому етапі будують граф, у якому кожен піксель належить окремій області (сегменту) і з'єднаний з сусідніми пікселями (чотирма чи вісьмома) ребрами з довжиною, яка визначена як різниця властивостей (характеристик), а сегменти (класи) об'єднуються доти, доки не виконається умова розрізнення сегментів [4]. Умова розрізнення сегментів задана так, що на певному кроці сегменти об'єднуються в один тільки тоді, коли перепад на їхній межі (різниця характеристик цих сегментів) менший, ніж максимальний перепад усередині кожного з цих сегментів.

Як характеристики для сегментації використано комбінацію кольорових і текстурних характеристик [6] з можливістю задання вагових коефіцієнтів характеристик.

Ефективність розробленого проекту досліджено в лабораторії комп'ютерних інформаційних технологій факультету електроніки Львівського національного університету імені Івана Франка на навчальному кластері паралельних та розподілених обчислень. Кластер охоплює 14 обчислювальних вузлів (AMD Athlon II X2 245, 8 Гб ОЗП) під керуванням операційної системи Scientific Linux. Керування розподіленими обчисленнями реалізовано за допомогою системи Condor v. 7.8.0. Виконано дослідження швидкодії роботи проекту для різних розмірів вхідних растрових зображень ДЗЗ. З'ясовано, що використання розподілених обчислень на 14 обчислювальних вузлах для опрацювання зображень обсягом > 1 Гб дає змогу зменшити час виконання у вісім–десять разів порівняно з виконанням завдання на одному обчислювальному вузлі. Отже, середній коефіцієнт приросту обчислювальних потужностей на додатковий обчислювальний вузол становить 0,6–0,7. Визначено оптимальну стратегію та пріоритети формування підзадач і розподілу їх між клієнтськими частинами. Досліджено ефективність модуля постпроцесингу для формування кінцевого сегментованого зображення. Отримані результати дають змогу зробити висновок про високу ефективність систем розподілених обчислень для опрацювання великих обсягів растрових зображень. Розроблений проект використано на практиці для сегментації космічних знімків високої роздільної здатності поверхні Шацького національного природного парку.

1. *Kussul N., Shelestov A., Korbakov M. et al.* Grid Infrastructure for Satellite Data Processing in Ukraine // Information Technologies and Knowledge. – 2008. – Vol. 2. – N 1. – P. 69–76.
2. Progress in Pattern Recognition, Image Analysis and Applications 12th Iberoamerican Congress on Pattern Recognition / ed. Luis Rueda Domingo Mery Josef Kittler – CIARP, 2007.
3. Condor High Throughput Computing – [Електронний ресурс] — Режим доступу: <http://research.cs.wisc.edu/condor/>.
4. *Шийка Ю. Я., Шувар П.* Використання алгоритму Краскала для сегментації даних дистанційного зондування поверхні землі // Проблеми електроніки та інформаційні технології: зб. тез II Всеукр. наук.-практ. конф. – Львів, 2010. – С. А-13–А-14.
5. *Felzenszwalb P., Huttenlocher D.* Efficient Graph-Based Image Segmentation // International J. of Computer Vision. – 2004. – Vol. 59. – N 2.
6. *Haralick R., Shamnugam K., Dinstein I.* Textural Features for Image Classification // IEEE Transactions on Cybernetics. – 1973. – Vol. SMC-3. – N 2. – P. 610–621.

**SEGMENTATION OF REMOTE SENSING DATA USING
DISTRIBUTED COMPUTING****Y. Shyuka, R. Shuwar**

*Ivan Franko National University of Lviv,
50 Drahomanov St., UA-79005 Lviv, Ukraine
shyuka@mail.lviv.ua*

The project for distributed segmentation of remote sensing images is developed. The Condor system of distributed computing is described. The main principles of work of the system and approaches used while developing applications for use in the distributed computing project are elucidated. In order to segment the images of remote sensing, a modification of Kruscal algorithm is presented. Colour and textural features are used for the image segmentation. The efficiency of the project presented is analyzed.

Key words: distributed computing, Condor, image segmentation, Kruscal's algorithm, remote sensing.

**ИСПОЛЬЗОВАНИЕ РАСПРЕДЕЛЁННЫХ ВЫЧИСЛЕНИЙ
ДЛЯ СЕГМЕНТАЦИИ ДАННЫХ
ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ****Ю. Шийка, Я. Шувар**

*Львовский национальный университет имени Ивана Франко,
ул. Драгоманова, 50, 79005 Львов, Украина
shyuka@mail.lviv.ua*

Разработан проект сегментации изображений дистанционного зондирования Земли с использованием распределённых вычислений. Рассмотрена система управления распределёнными вычислениями Condor. Описаны принципы работы системы и представлены особенности разработки приложений обработки изображений для использования в проекте распределённых вычислений. Для сегментации изображений дистанционного зондирования Земли использована модификация алгоритма Краскала с учётом цветовых и текстурных характеристик. Проанализирована эффективность работы реализованного проекта распределённой обработки изображений.

Ключевые слова: распределённые вычисления, Condor, сегментация изображений, алгоритм Краскала, дистанционное зондирование Земли.

Стаття надійшла до редколегії 06.05.2012

Прийнята до друку 19.06.2012