

УДК 004.75, 004.932

ВИКОРИСТАННЯ ПЛАТФОРМИ VOINC ДЛЯ ОПРАЦЮВАННЯ ДАНИХ ДИСТАНЦІЙНОГО ЗОНДУВАННЯ ЗЕМЛІ

Ю. Шийка, Р. Шувар

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

Розглянуто систему розподілених обчислень VOINC та реалізацію на її основі проекту розподіленого опрацювання растрових зображень дистанційного зондування Землі. Схарактеризовано принципи роботи системи та наведено особливості адаптації програми опрацювання зображень для використання в проекті розподілених обчислень. Для сегментації зображень дистанційного зондування Землі запропоновано модифікацію алгоритму Краскала. Рекомендовано використання текстурних характеристик на основі моментів для сегментації зображень. Проаналізовано ефективність роботи реалізованого проекту розподіленого опрацювання зображень.

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

У процесі обробки растрових зображень з метою виділення й підсилення інформативних частин зображень виникає потреба пошуку характерних областей і окремих об'єктів для їхнього подальшого опрацювання. Зокрема, в геоінформаційних системах для виділення об'єктів різних типів широко використовують алгоритми сегментації зображень дистанційного зондування Землі за певними параметрами. У цьому разі виникає проблема опрацювання зображень величезних розмірів. Для ефективного вирішення цієї проблеми запропоновано використовувати систему розподілених обчислень та алгоритм сегментації на основі графів, що дає змогу отримати набагато більші обчислювальні потужності. Системи розподілених обчислень сьогодні є особливо актуальними, їх щораз частіше використовують у подібних системах [1].

Доступні сьогодні зображення дистанційного зондування Землі високої роздільної здатності, що займають значні обсяги даних, зумовлюють необхідність використання великих обчислювальних потужностей для їхнього опрацювання. Висока роздільна здатність цих зображень дає змогу використовувати як критерії сегментації не тільки інформацію про колір пікселів, а й текстурні характеристики частин зображення [3], що додатково збільшує вимоги до обчислювальних потужностей, однак дає змогу отримувати ліпші результати сегментації. З огляду на це важливим завданням є розробка методів визначення текстурних характеристик та вибір оптимальних наборів характеристик і їхніх вагових коефіцієнтів для алгоритму сегментації, а також використання нових комп'ютерних технологій для опрацювання великих обсягів растрових даних.

Для організації розподілених обчислень використано грид-технологію, суть якої полягає в об'єднанні обчислювальних ресурсів клієнтів (вузлів) грид-мережі й координації потоків даних та обчислювальних завдань сервером. Одна з основних частин програмного забезпечення грид-інфраструктури – “Middleware” (проміжне програмне забезпечення), яке здебільшого складається з клієнтської частини та заготовки серверної частини з можливістю її налаштування під конкретний проект. Крім проміжного програмного забезпечення, необхідна розробка виконавчих програм (виконують клієнти), адаптованих до розподілених обчислень і серверних додатків формування завдань, збирання та звірки результатів, утворення кінцевих результатів, постпроцесингу тощо. На підставі аналізу можливостей різних грид-платформ та вимог, зумовлених особливостями задачі опрацювання зображень дистанційного зондування Землі, вибрано платформу з відкритим кодом BOINC (Berkeley Open Infrastructure for Network Computing) [2]. На рис. 1 показано загальну схему виконання розподілених обчислень у рамках проекту на базі платформи BOINC.

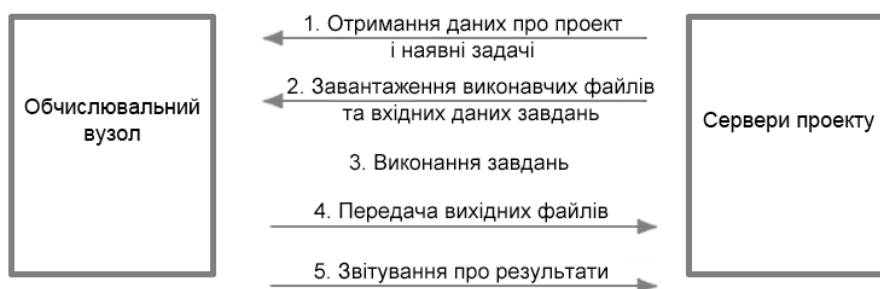


Рис. 1. Схема роботи проміжного ПЗ платформи BOINC.

Серед особливостей платформи BOINC назвемо такі:

- автономність проектів – різні BOINC проекти є незалежними, кожен з них керує власними серверами та базами даних. Немає центрального каталогу або процесу затвердження проектів;
- гнучкість для учасників – учасники можуть брати участь у декількох проектах, вони контролюють, у яких проектах беруть участь і як їхні ресурси розподілені між цими проектами. Коли проект не працює або не має завдань, ресурси розподіляються між іншими проектами. Всі завдання проектів клієнти виконують у фоновому режимі, використовуючи тільки незадіяні ресурси і жодним чином не перешкоджаючи нормальній роботі комп'ютера;
- гнучка структура програм – наявні програми написані різними мовами (C, C++, Fortran), можуть працювати як додатки BOINC практично без змін. Додаток може складатися з декількох файлів (наприклад, декілька програм і об'єднувальний скрипт). Нові версії програм можуть бути впроваджені без необхідності будь-яких дій учасників проекту;
- безпека – BOINC захищений від атак. Наприклад, використання цифрового підпису в ході розповсюдження виконавчих файлів забезпечує захист від поширення зловмисниками вірусів під виглядом задач BOINC;

- продуктивність сервера і масштабованість – один сервер середнього класу може відправляти задачі й опрацьовувати результати мільйонів робочих місць за день. Серверна архітектура також масштабована, що дає змогу легко збільшити потужність сервера завдяки додаванню нових машин;
- доступність вихідного коду – платформа VOINC поширюється вільно з відкритим кодом на умовах Lesser General Public License. Проте VOINC додатки не обов'язково повинні бути з відкритим вихідним кодом;
- незалежність від платформи учасника – клієнт VOINC доступний для більшості поширених платформ (Mac OS X, Windows, Linux та інших Unix-подібних систем). Клієнт може використовувати кілька процесорів, у тому числі GPU.

Для багатьох задач опрацювання зображень, у тому числі сегментації зображень дистанційного зондування Землі, розподілені обчислення можна організувати шляхом розбиття великої задачі на незалежні підзадачі та формування кінцевого результату за результатами опрацювання окремих підзадач. Розбиття зображення на підзадачі відбувається шляхом нарізання, а формування кінцевого результату – шляхом об'єднання результатів підзадач за принципом мозаїки. У випадку сегментації зображень необхідний також додатковий постпроцесинг.

На основі платформи VOINC у міжкафедральній лабораторії комп'ютерних інформаційних технологій факультету електроніки реалізовано проект розподіленого опрацювання растрових зображень. У рамках проекту встановлено та налаштовано серверну частину грид-платформи VOINC як віртуальну машину на сервері Xeon X3440 з 16 Гб оперативної пам'яті. Клієнтськими обчислювальними вузлами є 14 вузлів кластера паралельних та розподілених обчислень на базі AMD Athlon II X2 з 8 Гб оперативної пам'яті.

Для поділу зображення на сегменти (класи) використано алгоритм Краскала, суть якого полягає в побудові мінімального кістякового дерева заданого графа [4] (об'єднання всіх вершини графа в одну область мінімальною сумарною довжиною ребер так, щоб існував шлях між усіма вершинами). Принцип роботи алгоритму Краскала показано на рис. 2. Для сегментації растрових зображень за допомогою алгоритму Краскала запропоновано використовувати 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, яку визначають як різницю властивостей (характеристик), а сегменти (класи) об'єднують доти, поки не виконується умова розрізнення сегментів.

Умову розрізнення сегментів задають 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.

Запропоновано застосовувати як характеристики для сегментації комбінацію характеристик, які визначають різними способами. Сумарну характеристику визначають так:

$$F = \sum_i k_i \cdot F_i, \quad (1)$$

де k_i – ваговий коефіцієнт i -ї характеристики; F_i – значення i -ї характеристики.

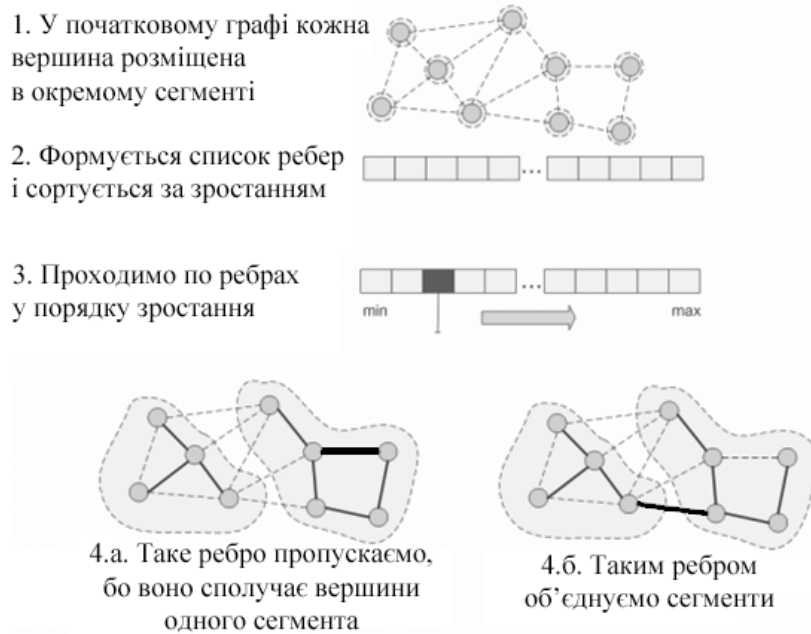


Рис. 2. Алгоритм Краскала.

Як характеристики для сегментації використано різницю кольорів пікселів і текстурні характеристики на основі моментів [5].

Для можливості сегментації великого зображення шляхом розподіленого опрацювання розроблено серверну програму формування незалежних підзадач – work_generator (генератор завдань). Параметрами цієї програми є параметри сегментації, вхідний файл великого зображення та максимальні розміри зображень підзадач. Зображення підзадач формують шляхом нарізання великого зображення та індексування зображень підзадач. Програма формує та реєструє задачі в базі даних проекту, після чого службові програми та скрипти проекту готові до розподілу задач між клієнтами, а після завершення виконання підзадач клієнтами утворюють кінцевий результат. З метою коректного опрацювання та наступного формування кінцевого результату для кожної підзадачі створюють два вхідні файли: перший – файл зображення підзадачі, другий – файл параметрів та індексації, який містить параметри сегментації для виконавчої програми, індекси зображення підзадачі (для визначення його місця в початковому зображенні) та параметри нарізання (розміри зображень підзадач для коректного формування кінцевого результату).

Виконавча програма, подібно до генератора завдань, унаслідок роботи за двома вхідними файлами створює два вихідні файли: перший – сумарне сегментоване зображення підзадачі, другий містить індекси та розміри зображень підзадач, а також дані про сегменти для постпроцесингу.

У разі отримання результатів виконання окремої підзадачі службові скрипти виконують перевірку на достовірність результату. Цей проект налаштовано так, що результат вважається достовірним, якщо отримано два ідентичні результати виконання однієї пі-

дзадачі на двох різних комп'ютерах за час, не менший від мінімально необхідного для виконання цієї підзадачі.

У міру надходження достовірних результатів опрацювання зображення розроблена програма *assimilator* (збирач) формує кінцевий результат. Утворюється зображення-заготовка і кожне зображення-результат виконання підзадачі вставляється в нього на відповідне місце (використовуючи дані про індекси та параметри нарізання). Для вставляння зображення підзадачі на відповідне місце без переписування цілого великого файлу зображення розроблено модуль розширеної роботи з файлами зображень.

Після формування результату (коли всі зображення підзадач зібрані у велике зображення) необхідно виконати постпроцесинг. Це необхідно тому, що в зображенні-результаті на стиках зображень підзадач сегменти, які мали б бути суцільними, часто розбиваються. Унаслідок постпроцесингу такі сегменти об'єднуються з використанням інформації про характеристики сегментів з вихідних файлів підзадач, утворених виконавчою програмою.

З використанням описаної вище методики практично реалізовано проект розподіленої сегментації растрових зображень дистанційного зондування Землі за кольоровими та текстурними ознаками. Для цього розроблено програму формування підзадач та розподілу їх між клієнтами (серверна частина), клієнтську програму сегментації растрових зображень, адаптовану до розподілених обчислень, та серверну програму формування кінцевого результату шляхом об'єднання та постпроцесингу отриманих результатів підзадач.

Досліджено ефективність роботи проекту. Визначено оптимальну стратегію та пріоритети формування підзадач і розподілу їх між клієнтськими частинами. В рамках проекту на чотирнадцяти вузлах досягнуто середньої обчислювальної потужності 114 GFlops. Розроблену систему використано на практиці для опрацювання зображень дистанційного зондування поверхні Шацького національного природного парку.

-
1. *Kussul N., Shelestov A., Korbakov M., Kravchenko O., Skakun S., Ilin M.* Grid Infrastructure for Satellite Data Processing in Ukraine // *International Journal Information Technologies and Knowledge*, Volume 2 / 2008.
 2. BOINC - Программное обеспечение с открытым исходным кодом для организации добровольных распределённых вычислений и распределённых вычислений в сети. Електронний ресурс. Режим доступу: <http://boinc.berkeley.edu/>
 3. *Конущин А.* Семантическая сегментация изображений, *Computer Science E-days / А. Конущин.* – Екатеринбург, 2010.
 4. *Felzenszwalb P.* Efficient Graph-Based Image Segmentation / *P. Felzenszwalb, D. Huttenlocher // International J. of Computer Vision.* – 2004. – Vol. 59. –N 2.
 5. *Progress in Pattern Recognition, Image Analysis and Applications*, ed. by Luis Rueda Domingo Mery Josef Kittler 12th Iberoamerican Congress on Pattern Recognition, CIARP, 2007.

PROCESSING OF REMOTE SENSING IMAGES USING BOINC PLATFORM**Y. Shyika, R. Shuwar**

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

The BOINC system of distributed computing and BOINC project of processing of raster images of remote sensing is described. Main principles of the work of BOINC system and approaches of adaptation of applications for use in distributed computing project. For segmentation of images of remote sensing modification of Kruskal algorithm is presented. Textural features based on moments are used for image segmentation. Efficiency of presented project is analyzed.

Key words: distributed computing, BOINC, image segmentation, Kruskal's algorithm, remote sensing.

**ИСПОЛЬЗОВАНИЕ ПЛАТФОРМЫ BOINC ДЛЯ ОБРАБОТКИ ДАННЫХ
ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ****Ю. Шийка, Я. Шувар**

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

Рассмотрено систему распределённых вычислений BOINC и реализацию на её основе проекта обработки растровых изображений дистанционного зондирования Земли. Описано принципы работы системы и представлено особенности адаптации приложений обработки изображений для использования в проекте распределённых вычислений. Для сегментации изображений дистанционного зондирования Земли предложено модификацию алгоритма Краскала. Рекомендовано использование текстурных характеристик на основе моментов для сегментации изображений. Проанализировано эффективность работы реализованного проекта распределённой обработки изображений.

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

Стаття надійшла до редколегії 29.06.2011

Прийнята до друку 30.08.2011