

ПРИКЛАДНІ АСПЕКТИ РЕАЛІЗАЦІЇ ПАРАЛЕЛЬНО-ІЄРАРХІЧНОГО ПЕРЕТВОРЕННЯ ІНФОРМАЦІЙНИХ СЕРЕДОВИЩ НА ОСНОВІ ГЕТЕРОГЕННОЇ КЛАСТЕРНОЇ СИСТЕМИ

А. А. Яровий, Ю. С. Богомолів

Вінницький національний технічний університет
Хмельницьке шосе, 95, Вінниця, 21021, Україна, E-Mail: axa@vinnitsa.com

Анотація

На основі проведених досліджень здійснено аналіз методологічних і прикладних аспектів реалізації паралельно-ієрархічного перетворення інформаційних середовищ на основі гетерогенної кластерної системи. Наведено попередні результати розробки кластерного обчислювального комплексу для виконання паралельно-ієрархічного перетворення інформаційних середовищ, зокрема зображень, а також перспективи подальшої розробки моделей інтелектуальних паралельно-ієрархічних та ієрарх-ієрархічних нейроподібних систем.

Ключові слова: паралельні обчислення, розподілені обчислення, кластерні системи, паралельно-ієрархічне перетворення, обробка зображень, системи штучного інтелекту.

Постановка проблеми. Аналіз підходів до реалізації паралельності при обробці надвеликих масивів даних

У наш час сфери задач, що потребують для свого вирішення застосування потужних обчислювальних ресурсів, постійно розширюються. Це пов'язано з тим, що відбулися фундаментальні зміни в самій організації наукових досліджень. Внаслідок широкого впровадження обчислювальної техніки, значно актуалізувався напрямок чисельного моделювання і чисельного експерименту. Чисельне моделювання, заповнюючи проміжок між фізичними експериментами і аналітичними підходами, дозволило вивчати явища, які є або занадто складними для дослідження аналітичними методами, або занадто дорогими, або небезпечними для експериментального вивчення. При цьому чисельний експеримент дозволив значно здешевити процес наукового та технологічного пошуку. Стало можливим моделювати в реальному часі процеси інтенсивних фізико-хімічних і ядерних реакцій, глобальні атмосферні процеси, процеси економічного та промислового розвитку регіонів і т.д. Очевидно, що вирішення таких масштабних завдань, що визначаються загальним поняттям "Grand challenges", потребує значних обчислювальних ресурсів [1-3].

Наведені вище задачі досить успішно вирішуються за допомогою паралельних та розподілених обчислень. Під розподіленими обчисленнями ми розуміємо спосіб розв'язання трудомістких обчислювальних завдань з використанням двох і більше комп'ютерів об'єднаних в мережу. Під паралельними обчисленнями – спосіб організації обчислень, в яких кілька дій проводяться одночасно. Варто зазначити, що при організації паралельних та розподілених обчислень не можливо обійтись без використання багатопроцесорних обчислювальних систем. У більшості випадків для таких задач використовуються суперкомп'ютери, але в окремих випадках непоганою альтернативою є кластерні обчислювальні системи [1-4].

В переважній більшості, кластери використовуються в обчислювальних цілях, зокрема в наукових дослідженнях. Для обчислювальних кластерів істотними показниками є висока продуктивність процесора в операціях над числами з плаваючою крапкою (flops) і низька латентність мережі, і менш істотними – швидкість операцій введення-виведення, яка більшою мірою важлива для баз даних і веб-сервісів. Обчислювальні кластери дозволяють зменшити час розрахунків, в порівнянні з одиночним комп'ютером, розбиваючи завдання на гілки, що виконуються паралельно, які обмінюються даними по локальній мережі. Мережева компонента кластеру може бути або звичайною локальною мережею, або бути побудована на основі спеціальних мережевих технологій, що забезпечують надшвидку передачу даних між вузлами кластера. Мережа кластеру призначена для інтеграції вузлів кластеру та, звичайно, відокремлена від зовнішньої мережі, через яку здійснюється доступ користувачів до кластеру. Спеціально виділяють високопродуктивні кластери (позначаються англ. аббревіатурою HPC Cluster – High-performance computing cluster) [1,3].

Разом з тим, паралельно-ієрархічне перетворення в поєднанні з можливостями систем розподіленої обробки інформації та інформаційної оптоелектроніки відкриває гарні перспективи поліпшення характеристик функціонування обчислювальних систем новаторської архітектури у сфері паралельної обробки надвеликих масивів даних [5-6].

Аналізуючи останні тенденції щодо повсюдного впровадження паралелізму обробки інформації, можна зробити висновок, що дана тематика є досить актуальною та затребуваною. Поєднання інноваційного паралельно-ієрархічного методу обробки інформації та розподілених кластерних обчислень можна

розцінювати як логічний та цілком доцільний крок, який дозволить виконувати ресурсоемний алгоритм паралельно-ієрархічного перетворення за прийнятний для користувача час [1,5].

Паралельно-ієрархічне перетворення застосовується для виділення характерних ознак зображень, їх кодування і скорочення розмірності при виконанні обчислень. Добра збіжність паралельно-ієрархічного перетворення використовується в структурах паралельної пам'яті, аналізу і розпізнавання зображень, при кодуванні і ущільненні даних, а також для обробки біомедичних сигналів [5].

Аналіз способів перетворення інформаційних середовищ

Обробка зображень має надзвичайно велике значення у багатьох сферах діяльності людини. Зокрема, розпізнавання зображень дозволяє розв'язувати складні інтелектуальні задачі класифікації, ідентифікації персони, автоматизованого пошуку схожих зображень, доведення автентичності зображення тощо [5,6].

Можна виділити три основних типи перетворень інформаційних полів [5]:

- Повне сприйняття характеризується тим, що або саме інформаційне поле (ІП) використовується в якості внутрішнього представлення інформації, або спосіб його перетворення такий, що дозволяє однозначно відновлювати вихідне поле. При цьому весь об'єм вхідних даних потрібно аналізувати тільки в тому випадку, якщо невідомо, яка їх частина достатня для виділення інформації, на основі якої можливе вирішення задачі кінцевого користування.
- При проблемно-орієнтованому сприйнятті використовується не взаємно-однозначне, а багатооднозначне відображення ІП у поле сприйняття, тобто інформація, яка аналізується системою, є узагальненням реальної ситуації. Прикладами обмежених внутрішніх представлень у системах цього класу є контурні препарати зображення, обмежений набір коефіцієнтів двовимірного перетворення тощо.
- Рефлекторні сприйняття характеризуються тим, що заздалегідь задана процедура перетворення ІП у кінцевий набір ознак. Подальша робота відбувається в просторі виділених ознак із використанням теорії рішень. Системи з таким типом сприйняття найбільш поширені в даний час, практично не мають переналадження і розробляються спеціально для визначеної задачі.

Отже, найбільш актуальними і трудомістким є пошук структур вторинних (перетворених) даних для проблемно-орієнтованих систем сприйняття, оскільки в системах повного сприйняття ця структура цілком визначається самим ІП, а в системах із рефлекторним сприйняттям – розв'язуваною задачею. У цьому зв'язку зробимо аналіз існуючих підходів до представлення ІП для організації даних у системах другого типу.

Виділяються чотири основні схеми [5]:

- 1) пряме представлення;
- 2) представлення на основі ортогональних перетворень;
- 3) структурне представлення;
- 4) представлення на основі нейроподібних методів перетворення.

Пряма схема припускає поелементне введення в пам'ять ЕОМ елементів ІП і роботу з ним як із матрицею чисел. Система опрацьовує елементи один за одним, виконуючи визначені операції [5].

Іншою широко використовуваною схемою подання є ортогональні перетворення (кодування з перетворенням) – Фур'є, Адамара, Карунена-Лоєва й ін., які описують ІП набором коефіцієнтів розкладання за деяким базисом. Важливою властивістю коефіцієнтів розкладання є те, що кожний із них характеризує все ІП в цілому, тобто є його інтегральною характеристикою. Інтегральність коефіцієнтів приводить до того, що вони виявляються інформативними при аналізі тільки тих класів ІП, що можна розглядати як "стаціонарні" сигнали [5].

Структурні методи представлення ґрунтуються на опису зображення за допомогою набору деяких стандартних елементів (наприклад, відрізка, кута, перетинання відрізків і зв'язків або відношень між ними). Перевагою синтаксичного опису є можливість його лінгвістичної організації, опис зображення в термінах укрупнених, семантично значущих одиниць, що відповідає особливостям людського сприйняття. При автоматичному створенні опису виникають ті ж труднощі, що і при використанні прямого подання, — у реальних задачах ЕОМ виявляється не в змозі самостійно виділити істотні елементи, узагальнені характеристики [5].

Нейроподібні методи подання інформації засновані на використанні чутливості нейронів у зоровій системі до паралельно-ієрархічного аналізу орієнтації стимулу в поєднанні з роздільною (пофрагментною) пірамідальною обробкою даних. Метод паралельно-ієрархічного перетворення належить саме до класу нейроподібних методів подання інформації [5].

Таким чином, у якості базового методу обробки цифрових зображень оберемо метод прямого паралельно-ієрархічного перетворення як найбільш перспективний та гнучкий з точки зору можливостей паралелізації.

Математична модель прямого паралельно-ієрархічного перетворення в порівнянні з відомими числовими методами перетворення (наприклад, розкладання в математичні ряди) простими операціями типу додавання забезпечують складну функціональну обробку сигналів у реальному масштабі часу, а також однозначність і оборотність із гарною збіжністю обчислювального процесу. Причому мережний метод обробки приводить до швидкого ущільнення вхідних масивів інформації [5].

У даному випадку алгоритм обробки зображень за допомогою паралельно-ієрархічного перетворення дозволяє обробляти зображення по частинах, вікнами певної розмірності. При цьому різні вікна обробляються незалежно, що дозволяє виконувати таке перетворення паралельно. Для обчислень було вирішено використовувати обчислювальний кластер – групу комп'ютерів, об'єднаних високошвидкісними каналами зв'язку, яка представляє з точки зору користувача єдиний апаратний ресурс [5,7].

Галузі застосування мережного методу ПІ перетворення різноманітні. Це мережні структури: паралельної пам'яті, цифрових систем прийому-передачі інформації, цифрових пристроїв ущільнення інформації, систем порівняння багатоградацийних зображень, у тому числі кореляційного порівняння, цифрових пристроїв попередньої обробки зображень, сегментації, кодування, формування ознак для розпізнавання, обробки біомедичної інформації [5,6].

Реалізація паралельно-ієрархічного перетворення інформаційних середовищ на основі гетерогенної кластерної системи

Опишемо узагальнений алгоритм функціонування програмного комплексу. Для цього будемо відштовхуватись від бажаного результату — набору векторів хвостових елементів, які однозначно характеризуватимуть вхідне зображення [7,8]. Отож, загальний алгоритм функціонування програмного комплексу повинен мати вигляд:

а) користувач завантажує вхідне зображення, яке необхідно обробити, а також вказує множину вузлів у локальній мережі, на яких необхідно проводити обчислення;

б) після ініціалізації початку обчислень комплекс автоматично формує пакет завдань для кожного обчислювального вузла та запускає їх на виконання;

в) після завершення обчислень комплекс збирає результати окремих вузлів на головному (користувацькому) сервері та демонструє результати користувачу.

Для реалізації даного алгоритму розіб'ємо структуру обчислювального комплексу на окремі складові:

1. Консоль керування — графічний додаток MS Windows, за допомогою якого виконується управління роботою кластеру. Повинна забезпечувати:

а) можливість завантаження зображення у будь-якому форматі;

б) можливість вказати множини вузлів, які складають обчислювальний кластер, або надати автоматизовану можливість просканувати локальну мережу на предмет виявлення множини обчислювальних вузлів;

в) можливість запустити на виконання обробку зображення на обраній множині вузлів та отримати результати обробки.

2. Хост-програма, яка запускається на кожному обчислювальному вузлі та забезпечує наступну функціональність:

а) отримання набору завдань по локальній мережі;

б) виконання над кожним набором даних із пакету завдань, які базуються на моделі прямого паралельно-ієрархічного перетворення;

в) формування набору векторів хвостових елементів та передача його на вузол-ініціатор обчислень.

3. Програмна бібліотека, яка забезпечуватиме мережеву взаємодію між вищевказаними компонентами:

а) отримання вхідного зображення, над яким необхідно здійснити перетворення;

б) можливість просканувати локальну мережу на предмет наявності обчислювальних вузлів, або можливість задати їх множину вручну;

в) розбиття вказаного зображення на набір завдань відповідно до множини обчислювальних вузлів, та призначення завдання кожному вузлу;

г) функціональність для ініціалізації початку обробки та отримання результатів.

У загальному вигляді, схему взаємодії компонентів програмного комплексу зображено на рисунку 1.

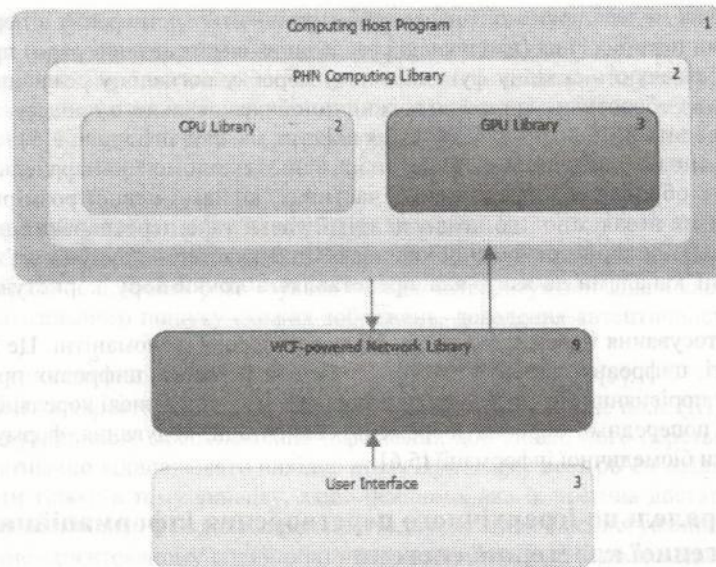


Рис.1. Схема взаємодії компонентів

Алгоритм кластерної обробки зображень методом прямого паралельно-ієрархічного перетворення можна описати так:

— Вхідне зображення розділяється на множину завдань для кожного із обчислювальних вузлів наступним чином:

а) зображення розділяється на набір вікон обробки розмірністю 128×128 пікселів кожне;

б) перше вікно обробки призначається першому обчислювальному вузлу, друге – другому вузлу, вікно $N+1$ (де N – кількість обчислювальних вузлів) знову призначається першому вузлу і т.д., із урахуванням того, що кількістю обчислювальних вузлів у загальному випадку значно менша за кількість завдань.

— Кожен обчислювальний вузол проводить обробку власного набору завдань методом прямого паралельно-ієрархічного перетворення, та формує множину наборів хвостових елементів як результатів перетворення.

— Результати обробки із кожного вузла збираються на головному (керуючому) вузлі та формується загальний результат обробки зображення.

Проектування програмного комплексу із урахуванням вищезазначеної схеми взаємодії компонентів відбувалось за такими кроками:

— На найзагальнішому метарівні: програмний комплекс «DirectPHT::Cluster» повинен завантажувати та ідентифікувати зображення довільної розмірності за допомогою прямого паралельно-ієрархічного перетворення, яке виконується на обчислювальному кластері.

— На другому рівні деталізації: програмний комплекс «DirectPHT::Cluster» повинен автоматично виявляти топологію обчислювального кластера, виявляти доступні обчислювальні вузли, назначати їм індивідуальні завдання та збирати результат. Також необхідно визначити та видати відповідний пакет завдань кожному обчислювальному вузлу.

— На третьому рівні деталізації: необхідно емпірично визначити оптимальний розмір частин, на які буде розбиватись вхідне зображення для створення пакету завдань для кожного обчислювального вузла. Також необхідно визначити програмну структуру сервера контролю та обчислювальних вузлів. По можливості програмний комплекс повинен бути незалежним від топології кластеру.

— На четвертому рівні деталізації: обирається технологія комунікації між вузлами та сервером. У даному випадку був вибір між технологією Microsoft .NET Windows Communication Foundation та функціональною мовою програмування Erlang із пакетом бібліотек OTP.NET. Вибір було зроблено на користь першої технології через зручність її інтеграції із вже наявним проектом реалізації прямого паралельно-ієрархічного перетворення на одному обчислювальному вузлі [8].

— На п'ятому рівні деталізації: визначається структура сервісу взаємодії та формат повідомлень, якими будуть обмінюватись сервер та обчислювальні вузли. Відповідно до обраної

технології було визначено формат так званого контракту даних — тобто, формат повідомлень, якими будуть обмінюватись сервер та обчислювальні вузли кластеру. Оскільки кожен обчислювальний вузол матиме власний набір завдань, відповідно, час обробки у кожного вузла буде різний, тому було вирішено зробити обмін даними між вузлами та серверною програмною бібліотекою асинхронним. Крім того, виклик основного методу даної бібліотеки із головного додатку — консолі керування — теж було зроблено асинхронним: бібліотека сигналізуватиме про завершення обчислень головному додатку лише тоді, коли кількість результатів, отриманих від обчислювальних вузлів, буде рівною кількості початкових завдань на обчислення.

На шостому рівні деталізації: відбувається кодування зроблених рішень із урахуванням технології комунікації та можливостей мови програмування C#.

Згідно результатів аналізу та із використанням обраної технології міжкомпонентної мережевої взаємодії написано пакет програм для управління обчислювальним кластером.

Розроблений пакет програм протестовано на обчислювальному кластері Microsoft IT Academy ВНТУ. Результати тестування дозволяють зробити висновок про достовірність комп'ютерного моделювання, а також правильність та оптимальність проектування.

Література:

- [1] Хьюз К. Параллельное и распределенное программирование на C++. / Хьюз К., Хьюз, Т. [Пер. с англ.] – М.: ИД «Вильямс», 2004. – 672 с.
- [2] Воеводин В.В. Параллельные вычисления. / В.В.Воеводин, Вл.В.Воеводин – СПб.: БХВ-Петербург, 2002. – 608 с.
- [3] Розподілені комп'ютерні системи як складові інформаційних інфраструктур / В.П. Горбулін, О.Г. Додонов, О.С. Горбачик, М.Г. Кузнєцова // Реєстрація, зберігання і обробка даних. – 2008. – Т. 10, № 4. – С. 19-24. – ISSN 1560-9189.
- [4] Бройнль Т. Паралельне програмування: Навч. посібник / Бройнль Т. [Пер. з нім. В.А. Святного] – К.: Вища школа, 1997. – 358 с.
- [5] Паралельно-ієрархічне перетворення як системна модель оптико-електронних засобів штучного інтелекту : [Монографія.] / В.П. Кожем'яко, Ю.Ф. Кутасєв, С.В. Свечніков, Л.І. Тимченко, А.А. Яровий – Вінниця: УНІВЕРСУМ-Вінниця, 2003. – 324 с.
- [6] Образний відео-комп'ютер око-процесорного типу : [Монографія.] / Кожем'яко В.П., Лисенко Г.Л., Яровий А.А., Кожем'яко А.В. – Вінниця: Універсум-Вінниця, 2008. – 215 с. – ISBN 978-966-641-261-7.
- [7] Яровой А.А. Прикладные аспекты программно-аппаратной реализации нейроподобных параллельно-иерархических систем : Сборник научных трудов. В 2 частях. Ч. 2. [Научная сессия МИФИ – 2009. XI Всероссийская научно-техническая конференция «Нейроинформатика-2009»], (Москва, 27-30 января 2009 г.) – М.: МИФИ, 2009. – С. 39-48.
- [8] Свідоцтво про реєстрацію авторського права на твір № 33758. Комп'ютерна програма для реалізації паралельно-ієрархічного перетворення інформаційних середовищ на основі CPU (Direct PNT::CPU) / Яровий А.А., Богомолів Ю.С. Дата реєстрації Державним Департаментом інтелектуальної власності України 17.06.2010.