

Презентація магістерської
кваліфікаційної роботи:
«Інформаційна
технологія
управління
обчислювальним
кластером»»

Виконав: Беліченко Сергій

Науковий керівник: доцент Арсенюк І. Р.

- ▶ Об'єкт дослідження - процеси управління обчислювальними кластерними системами.
- ▶ Предмет дослідження - програмне забезпечення для виконання паралельних обчислень.
- ▶ Мета роботи - підвищення швидкодії процесу паралельних обчислень, а саме паралельно-ієрархічної обробки прямоподібних зображень на основі обчислювального кластеру.

Задачі дослідження, які потрібно виконати для досягнення мети роботи:

- ▶ 1) здійснити аналіз предметної області високопродуктивних паралельних обчислень;
- ▶ 2) навести класифікацію методів оброблення зображень;
- ▶ 3) здійснити аналіз математичних моделей паралельно-ієрархічного оброблення зображень
- ▶ 4) здійснити комп'ютерне моделювання процесу паралельно-ієрархічного оброблення зображень у кластерній системі;
- ▶ 5) розробити структурну організацію та здійснити програмну реалізацію паралельно-ієрархічного оброблення зображень у кластерній системі.

Наукова новизна

До нових наукових результатів, отриманих у магістерській кваліфікаційній роботі, відносяться:

1. Удосконалено моделі організації високопродуктивного обчислювального процесу паралельної обробки у кластерній системі, що забезпечило підвищення швидкодії до 96% в окремих випадках при застосуванні кластеру.
2. Використано моделі та структурну організацію високопродуктивної паралельно-ієрархічної системи на основі використання технології Message Passing interface (MPI), High Perfomance Cluster (HPC), що дозволило здійснювати обробку та класифікацію зображень середньої роздільної здатності з підвищеною продуктивністю.

Практичне значення отриманих результатів

Розроблена інформаційна технологія дозволяє здійснювати обробку та класифікацію зображень середньої роздільної здатності з підвищеною продуктивністю.

Удосконалено розподілення обчислень всередині кластеру, що дозволило отримати більш швидке розв'язання поставлених задач

Актуальність теми

- ▶ Останнім часом все більшого поширення набувають технології високопродуктивних обчислень. Це зумовлено тим, що в процесі розвитку інформаційних технологій виникає все більше задач, пов'язаних з обробкою надзвичайно великих об'ємів інформації. Вирішення таких задач потребує використання значних обчислювальних потужностей. Оскільки технічний прогрес розвивається не настільки швидко, наскільки швидко ростуть практичні потреби, актуальними стають підходи, спрямовані на більш ефективне і спільне використання наявних ресурсів. Одним з таких підходів і є технології паралельних обчислень.
- ▶ Застосування паралельних технологій також потребує створення відповідних паралельних методів і алгоритмів. Одним з таких підходів є пряме паралельно-ієрархічне (ПІ) перетворення за допомогою кластерних систем. ПІ обробка знаходить застосування у вирішенні ресурсоємних та складних задач у багатьох прикладних галузях, наприклад, у медицині. Оскільки, в таких задачах оброблення доводиться виконувати у реальному часі, виникає потреба у підвищенні швидкодії обчислень.

Аналіз предметної області

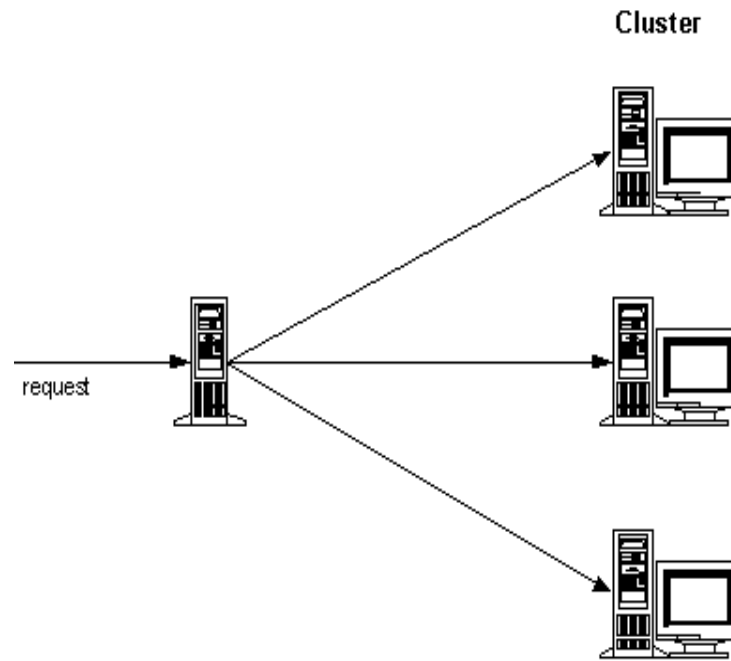
Кластер - це група комп'ютерів, об'єднаних високошвидкісними каналами зв'язку, яка представляє з точки зору користувача єдиний апаратний ресурс.

Зазвичай виділяють такі основні види кластерів:

- ▶ Відмовостійкі кластери (failover clusters).
- ▶ Кластери з балансуванням навантаження (Load balancing clusters).
- ▶ Обчислювальні кластери (High performance computing clusters, HPC).
- ▶ Системи розподілених обчислень (grid).

Кластери з балансуванням навантаження (Load balancing clusters)

- Принцип кластерів розподілення навантаження будується на розподіленні запитів через один або декілька вхідних вузлів, які перенаправляють їх на обробку в інші, обчислювальні вузли. Основна ціль таких кластерів - продуктивність, але в них часто використовуються також методи, які підвищують надійність.



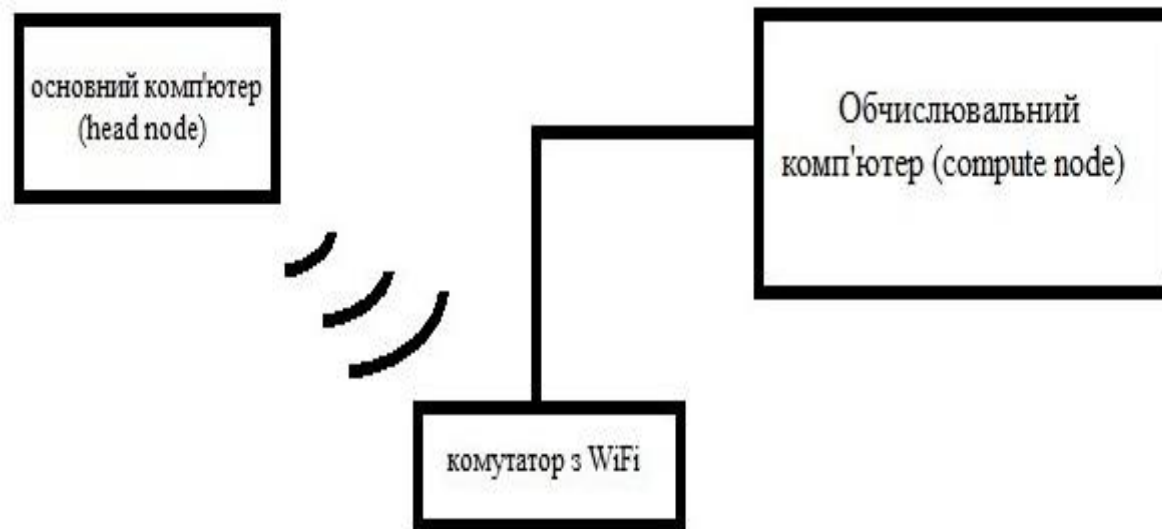
Обчислювальні кластери (High performance computing clusters, HPC)

- Обчислювальні кластери використовуються в обчислювальних цілях, частково у навчальних дослідженнях. Для обчислювальних кластерів найбільш важливими показниками є висока продуктивність процесора в операціях над числами з плаваючою точкою (flops) та низька затримка мережі, яка об'єднує кластер, та менш важливими - швидкість операцій введення-виведення, яка в більшій ступені важлива для баз даних та web-сервісів.
- Обчислювальні кластери дозволяють зменшити час розрахунків, порівняно з одним комп'ютером, розділяючи завдання на гілки, які паралельно виконуються, що обмінюються даними по мережі, яка об'єднує.

Саме такий вид кластера використовується у кваліфікаційній роботі

Структура кластерної системи

Усього в кластері є 2 ноди (вузли): один нод є сервером (менеджером), інший комп'ютер є обчислювальним нодом. Сервер розподіляє задачу між усіма комп'ютерами (сервер також входить в число обчислювальних вузлів і приймає участь в обчисленнях).



Для реалізації інформаційної технології управління обчислювальним кластером потрібно:

- 1) Обрати операційну систему, в якій є можливість з'єднати декілька комп'ютерів в один кластер.
- 2) Встановити обрану операційну систему на комп'ютери, які будуть згодом об'єднуватись у кластер, встановити потрібні додатки та додаткові бібліотеки.
- 3) Об'єднати комп'ютери в один кластер, протестувати його працездатність та продуктивність.
- 4) Виконати тестове завдання.
- 5) Зафіксувати результати.

Комп'ютерне моделювання кластерної обробки інформації

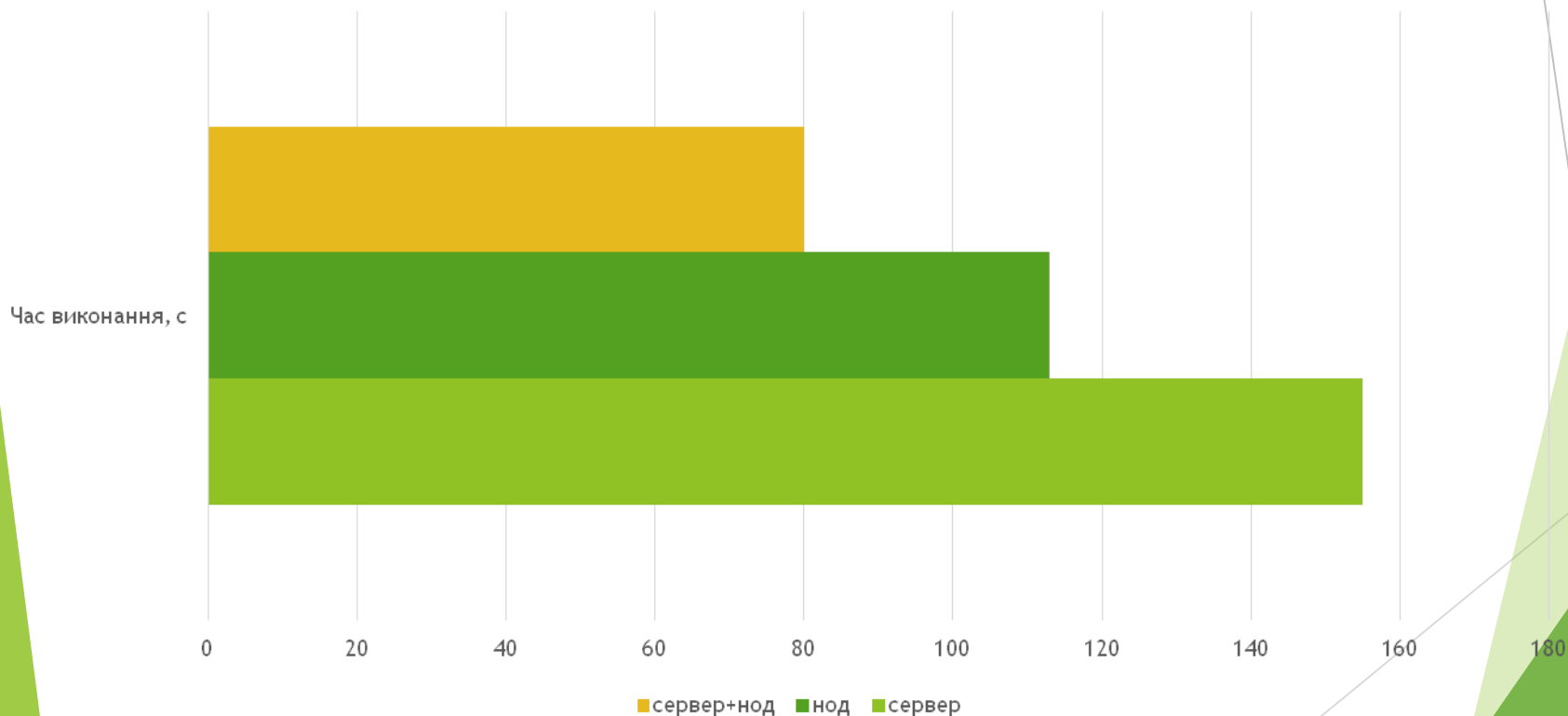
Тестовий приклад - обробка набору з 16 високорозмірних зображень, представлених у вигляді масивів розміром 4096x4096 елементів. Над кожним масивом виконується 1000 операцій, що відповідає багатоступеневому алгоритму обробки. Число процесів можна змінювати.

Задача розв'язувалась у декількох режимах: сервер (8 та 16 потоків), нод (8 та 16 потоків) та сервер і нод разом (8 і 16 потоків).

	1 процес	2 процеси	4 процеси	8 процесів	16 процесів
Сервер				2 хв 35 с	2хв 36 с
Нод				1 хв 53 с	1хв 52 с
Сервер + нод				1 хв 20 с	1 хв 19 с

Графічне представлення комп'ютерного моделювання

Комп'ютерне моделювання обробки набору з 16 зображень, представлених у виді масивів розміром 4096x4096 елементів



Тестування програмного продукту та аналіз результатів

У результаті виконання магістерської кваліфікаційної роботи створена інформаційна технологія управління обчислювальним кластером, за допомогою якого здійснюється ПІ перетворення зображень, що використовує при обчисленнях декілька обчислювальних нодів одночасно. Дана технологія реалізована на програмній платформі HPC Cluster Manager з використанням бібліотеки MPI та OpenCV.

Програмний модуль ПІ-перетворення зображень тестувався у чотирьох режимах:

1) ПІ перетворення набору з 50 зображень формату bmp різної розмірності (128x128, 1024x1024) за допомогою OpenCV та MPI на одному комп'ютері (сервер) з різною кількістю потоків;

	1 процес	2 процесів	4 процесів	8 процесів	16 процесів
128x128	1,4 с	0,77 с	0,61 с	0,65 с	0,8 с
1024x1024	58 с	36 с	26 с	26 с	28 с

- 2) ПІ перетворення набору з 50 зображень формату bmp різної розмірності (128x128, 1024x1024) за допомогою OpenCV та MPI на одному комп'ютері (ноді) з різною кількістю потоків;

	1 процес	2 процеси	4 процеси	8 процесів	16 процесів
128x128	4 с	2 с	1,12 с	1,4 с	1,9 с
1024x1024	60 с	34 с	29 с	15 с	15 с

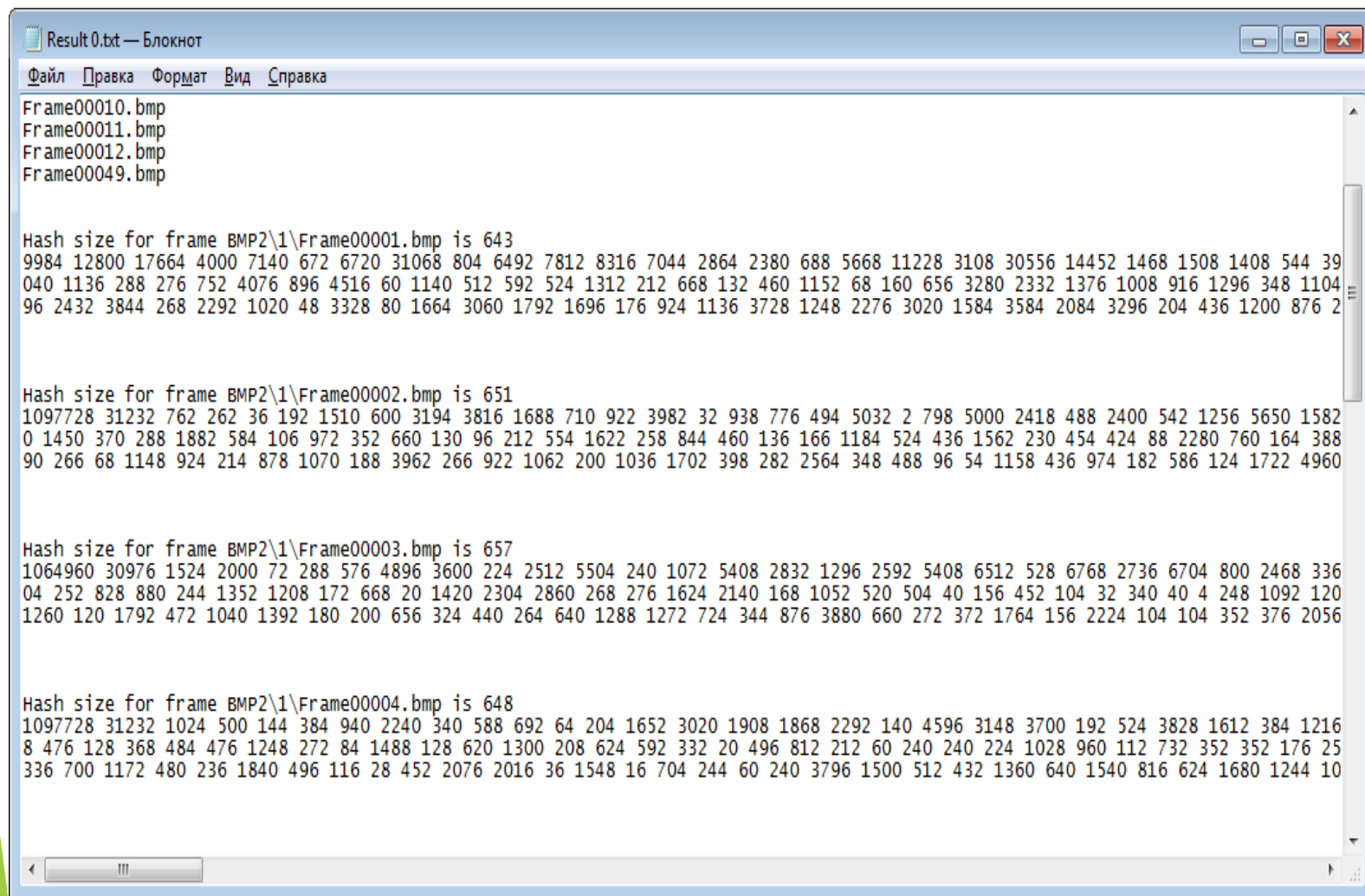
3) ПІ перетворення набору з 50 зображень формату bmp різної розмірності (128x128, 1024x1024) за допомогою OpenCV та MPI на двох комп'ютерах (сервер та нод) з різною кількістю потоків;

	1 процес	2 процеси	4 процеси	8 процесів	16 процесів
128x128	1,6 с	0,86 с	0,57 с	1,3 с	1,6 с
1024x1024	58 с	36 с	28 с	16 с	16 с

4) ПІ перетворення набору з 50 зображень формату bmp різної розмірності (128x128, 1024x1024) за допомогою OpenCV та MPI на одному комп'ютері (нод) з різною кількістю потоків та завантаженням зображень через загальний доступ (через мережу).

	1 процес	2 процеси	4 процеси	8 процесів	16 процесів
128x128			6 с.		
1024x1024			1 м 26 с		

Результат работы программного продукта



```
Result 0.txt — Блокнот
Файл  Правка  Формат  Вид  Справка

Frame00010.bmp
Frame00011.bmp
Frame00012.bmp
Frame00049.bmp

Hash size for frame BMP2\1\Frame00001.bmp is 643
9984 12800 17664 4000 7140 672 6720 31068 804 6492 7812 8316 7044 2864 2380 688 5668 11228 3108 30556 14452 1468 1508 1408 544 39
040 1136 288 276 752 4076 896 4516 60 1140 512 592 524 1312 212 668 132 460 1152 68 160 656 3280 2332 1376 1008 916 1296 348 1104
96 2432 3844 268 2292 1020 48 3328 80 1664 3060 1792 1696 176 924 1136 3728 1248 2276 3020 1584 3584 2084 3296 204 436 1200 876 2

Hash size for frame BMP2\1\Frame00002.bmp is 651
1097728 31232 762 262 36 192 1510 600 3194 3816 1688 710 922 3982 32 938 776 494 5032 2 798 5000 2418 488 2400 542 1256 5650 1582
0 1450 370 288 1882 584 106 972 352 660 130 96 212 554 1622 258 844 460 136 166 1184 524 436 1562 230 454 424 88 2280 760 164 388
90 266 68 1148 924 214 878 1070 188 3962 266 922 1062 200 1036 1702 398 282 2564 348 488 96 54 1158 436 974 182 586 124 1722 4960

Hash size for frame BMP2\1\Frame00003.bmp is 657
1064960 30976 1524 2000 72 288 576 4896 3600 224 2512 5504 240 1072 5408 2832 1296 2592 5408 6512 528 6768 2736 6704 800 2468 336
04 252 828 880 244 1352 1208 172 668 20 1420 2304 2860 268 276 1624 2140 168 1052 520 504 40 156 452 104 32 340 40 4 248 1092 120
1260 120 1792 472 1040 1392 180 200 656 324 440 264 640 1288 1272 724 344 876 3880 660 272 372 1764 156 2224 104 104 352 376 2056

Hash size for frame BMP2\1\Frame00004.bmp is 648
1097728 31232 1024 500 144 384 940 2240 340 588 692 64 204 1652 3020 1908 1868 2292 140 4596 3148 3700 192 524 3828 1612 384 1216
8 476 128 368 484 476 1248 272 84 1488 128 620 1300 208 624 592 332 20 496 812 212 60 240 240 224 1028 960 112 732 352 352 176 25
336 700 1172 480 236 1840 496 116 28 452 2076 2016 36 1548 16 704 244 60 240 3796 1500 512 432 1360 640 1540 816 624 1680 1244 10
```

Висновки

- ▶ Здійснено аналіз предметної області високопродуктивних паралельних обчислень та виявлено перспективність їх застосування в галузях, пов'язаних з обчисленням величин великої розмірності.
- ▶ Розглянуто технології паралельних обчислень на прикладі SIMD та SISD, виділено їх характерні особливості та сфери застосування. На основі порівняння теоретичних потужностей одного та декількох разом працюючих комп'ютерів доведено доцільність застосування обчислень на основі обчислювального кластеру при розв'язанні задач, які можуть бути виконані паралельно.
- ▶ Проаналізовано об'єкт проектування, наведено постановку задачі паралельно-ієрархічного перетворення оброблення зображень, вказано основні вимоги до апаратного і програмного забезпечення, головною з яких є наявність додатку HPC та бібліотек OpenCV, MPI на всіх використовуваних комп'ютерах, та вимоги до вхідних зображень, а саме: розширення bmp, формат кольору Grayscale, мінімальна допустима розмірність 128x128 пікселів та максимально допустима розмірність 1024x1024 пікселів. Доведено актуальність використання і подальшого розвитку алгоритму паралельно-ієрархічного перетворення зображень.
- ▶ Також розглянуто класифікацію методів обробки зображень, виділено місце паралельно-ієрархічної обробки зображень серед існуючих методів.
- ▶ З метою обрання найбільш ефективного з точки зору швидкодії підходу до організації паралельних обчислень при обробленні зображень було проведено комп'ютерне моделювання паралельно-ієрархічного оброблення зображень у кластерній системі. У якості прикладу було виконання елементарних арифметичних операцій над масивами. Згідно отриманих результатів комп'ютерного моделювання застосування технологій паралельних обчислень дозволило отримати приріст швидкодії майже в усіх випадках.
- ▶ Застосування кластеру виявилось найбільш ефективним при обробці зображень малої розмірності (128x128 пікселів). В такому випадку приріст швидкодії складає приблизно 13%. При обробці зображень великої розмірності переваги кластеру не мінімізуються, але в нашому випадку спостерігається спад продуктивності через велику різницю в потужностях обчислювальних нодів, але при наявності приблизно однакових за потужності нодів спостерігався би відчутний приріст. Вищенаведені результати підтверджують доцільність використання обраного методу при реалізації обробки зображень взагалі та ПІ оброблення зокрема.

ДЯКУЮ ЗА
УВАГУ!