

# ОСОБЛИВОСТІ ЗАСТОСУВАННЯ ПАРАЛЕЛЬНО-ІЄРАРХІЧНОГО ОБРОБЛЕННЯ ЗОБРАЖЕНЬ У ГЕТЕРОГЕННОМУ ПРОГРАМНО-АПАРАТНОМУ СЕРЕДОВИЩІ

Вінницький національний технічний університет

## *Анотація*

*Розглянуто особливості застосування методу прямого паралельно-ієрархічного перетворення у гетерогенному програмно-апаратному середовищі з точки зору підвищення швидкодії процесу оброблення зображень.*

**Ключові слова:** паралельні обчислення, багатопотоковість, GPGPU, паралельно-ієрархічне перетворення, багаторівневий паралелізм.

## *Abstract*

*Features of usage of parallel-hierarchical transformation method in heterogeneous software and hardware environment are analyzed from the process of images processing's performance improvement point of view.*

**Keywords:** parallel computing, multithreading, GPGPU, parallel-hierarchical transformation, multi-level parallelism.

## Вступ

На сучасному етапі розвитку технологій паралельних обчислень існує широкий спектр підходів до їх реалізації: багатопотокові обчислення, гетерогенні обчислення, розподілені обчислення на основі обміну повідомленнями тощо. Кожен з цих підходів є орієнтованим на певне коло задач, має свої характерні особливості застосування та обмеження [1]. Проте, у випадку застосування паралельних технологій у комплексних програмних системах, що зазвичай працюють з різними типами даних та використовують складні багатокрокові алгоритми, такий стан речей призводить до цілого ряду обмежень, що пов'язані у тому числі з різною обчислювальною складністю різних етапів алгоритмів. Одним із можливих рішень є комплексне застосування декількох технологій в межах однієї системи, тобто багаторівневий паралелізм.

Метою роботи є аналіз особливостей застосування методу прямого паралельно-ієрархічного перетворення у гетерогенному програмно-апаратному середовищі для підвищення швидкодії процесу оброблення зображень.

## Результати дослідження

Яскравим прикладом комплексних систем вищезгаданого типу є спеціалізовані системи для обробки зображень, зокрема системи розпізнання та класифікації, системи штучного зору та інші. Такі системи, як правило, повинні працювати з різними типами зображень, що мають різну розмірність, різну кольорову модель, різну ступінь спотворення тощо [2]. Окрім того, важливу роль відіграє швидкодія системи, оскільки оброблення зображень високої розмірності потребує виконання значної кількості обчислень. Це зумовлює доцільність використання технологій паралельних обчислень, але актуальним в цьому випадку стає питання обрання технології. Технології гетерогенних обчислень надають значний приріст швидкодії у випадку зображень великої та надвеликої розмірності, але у випадку зображень малої розмірності накладні витрати на організацію обчислень можуть перевищувати потенційний приріст швидкодії [3]. Концепція багатопотоковості зазвичай виявляється доцільною для зображень малої та середньої розмірності, але при надвеликій розмірності даються взнаки технологічні обмеження CPU [4]. Також, важливу роль відіграють характеристики методу оброблення зображень.

В якості прикладу доцільно розглянути метод прямого паралельно-ієрархічного (ПІ) перетворення інформації. ПІ перетворення можливо визначити як принцип паралельного оброблення інформації, метою якого є досягнення максимально можливої алгоритмічної та схематехнічної швидкодії при перетворенні інформації, зокрема зображень. З алгоритмічної точки зору пряме ПІ перетворення складається з трьох головних операцій: зсув, транспонування та G-перетворення. Ці три операції ітераційно виконуються над матрицею чисел, що відповідає вхідному зображенню. Кінцевим результатом є матриця хвостових елементів, що може використовуватися при класифікації зображень. Окрім того, цим операціям передують етапи попередньої обробки: зчитування зображення та попередня класифікація (на основі ступеня спотворення в порівнянні з еталоном) [5,6].

ПІ перетворення за своєю природою є паралельно орієнтованим методом і тому може бути ефективно реалізованим на основі технології NVIDIA CUDA із застосуванням додаткової бібліотеки CUDA Thrust [7]. Одним із можливих підходів до подальшого розвитку цієї концепції є одночасне застосування декількох GPU (технологія Multi-GPU Programming) за принципом паралелізму даних для оброблення декількох зображень одночасно. Проте, перешкодою на цьому шляху виявилась синхронність багатьох функцій бібліотеки CUDA Thrust, що не дозволяло в повній мірі використати інструментарій Multi-GPU Programming.

Вирішенням цієї проблеми стало комплексне застосування технологій OpenMP та NVIDIA CUDA. За допомогою засобів OpenMP було створено декілька паралельних потоків на CPU (у відповідності з кількістю графічних карт у системі), кожен з яких отримав свою графічну карту і свою частину набору зображень для оброблення. За таким підходом кожен потік  $i$ , відповідно, кожна графічна карта працювали незалежно один від одного, що дозволило обійти обмеження бібліотеки CUDA Thrust та отримати додатковий приріст швидкодії.

Варто відзначити, що приріст швидкодії від застосування технологій паралельних обчислень є різним для кожного з етапів процесу оброблення зображень [8]. Так, використання NVIDIA CUDA для задач зчитування та попередньої класифікації зображень виявилось недоцільним через накладні витрати, а застосування для тих самих задач технології OpenMP дозволило отримати приріст швидкодії в 1,22-2,75 разів (в залежності від розмірності зображення) при зчитуванні зображень та в 1,2-2,08 разів (в залежності від розмірності зображення) при попередній класифікації зображень. Отримані результати наведені в таблиці 1.

Для реалізації безпосередньо алгоритму прямого ПІ перетворення було використано засоби NVIDIA CUDA та додаткової бібліотеки CUDA Thrust, при чому приріст швидкодії також на пряму залежав від розмірності зображення. Більш того, виявилось, що час виконання кожної ітерації 3 головних операцій також залежить від вхідних даних і обраних технологій.

В результаті при обробленні зображень розмірності 1024x1024 пікселів із застосуванням 4 графічних карт досягається загальний приріст швидкодії 303% в порівнянні з послідовною реалізацією на CPU та 273% в порівнянні з реалізацією на основі одного GPU. При обробленні зображень розмірності 2048x2048 пікселів із застосуванням 4 графічних карт досягається приріст швидкодії 303% в порівнянні з послідовною реалізацією на CPU та 247% в порівнянні з реалізацією на основі одного GPU. При обробленні зображень розмірності 4096x4096 пікселів із застосуванням 4 графічних карт досягається приріст швидкодії 223% в порівнянні з послідовною реалізацією на CPU та 196% в порівнянні з реалізацією на основі одного GPU.

Таблиця 1 – Результати роботи програмного продукту

Час, с.				
Розмірність зображення (масиву)	Модуль зчитування зображень		Модуль попередньої класифікації	
	Без OpenMP	З OpenMP	Без OpenMP	З OpenMP
128x128	0,0077	0,0028	0,008	0,0066
1024x1024	0,356	0,167	0,365	0,175
2048x2048	1,358	0,616	1,494	0,729
4096x4096	5,468	4,458	6,006	4,602

## Висновки

Отримані результати свідчать про доцільність застосування концепції багаторівневого паралелізму на основі паралелізму даних у системах оброблення зображень, зокрема у системах на основі методу прямого ПП перетворення. Так, комбіноване застосування NVIDIA CUDA та OpenMP в системі з 4 GPU дозволило отримати приріст швидкодії від 196% до 273% в порівнянні з реалізацією на основі одного GPU (в залежності від розмірності зображень) та від 223% до 303% порівняно з реалізацією на основі CPU (в залежності від розмірності зображень).

Окрім того, згідно отриманих результатів приріст швидкодії на кожному окремому етапі обробки зображень залежить як від вхідних даних, так і від обраної технології або комбінації технологій паралельних обчислень, що вказує на доцільність пошуку найбільш оптимальних комбінацій технологій паралельних обчислень для кожного етапу обробки зображень.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Гергель В.П. Высокопроизводительные вычисления для многоядерных многопроцессорных систем. / Гергель В.П. – Н.: ННГУ им. Н.И.Лобачевского, 2010. – 421 с.
2. Форсайт Д. Компьютерное зрение. Современный подход / Форсайт Д., Понс Д. – М.: Мир, 2004. – 610с.
3. NVIDIA – WORLD LEADER IN VISUAL COMPUTING TECHNOLOGIES [Електронний Ресурс] – Режим Доступу: <http://www.nvidia.ru/page/home.html>.
4. Левін М. А. Параллельное программирование с использованием OpenMP [Електронний ресурс] / Левін М. А. // Режим доступу: <http://www.intuit.ru/studies/courses/1112/232/info>.
5. Яровий А. А. Методи та засоби організації високопродуктивних паралельно-ієрархічних обчислювальних систем із рекурсивною архітектурою : монографія / А. А. Яровий. – Вінниця: ВНТУ, 2016. – 363 с.
6. Яровий А. А. Паралельно-ієрархічне перетворення плямоподібних зображень на основі Multi-GPU систем / Яровий А. А., Кулик О. О., Кокряцька Н. І. // Інформаційні технології та комп'ютерна інженерія. – 2015. – №3(34). – С. 72-80.
7. A.A. Yarovy Parallel-hierarchical processing and classification of laser beam profile images based on the GPU-oriented architecture / A.A. Yarovy, L.I. Timchenko, V.P. Kozhemiako, N.I. Kokriatskaia, R.R. Hamdi, T.O. Savchuk, O.O. Kulyk, W. Surtel, Ye. Amirgaliyev, G. Kashaganova : [Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments], (August 07, 2017), Proceedings SPIE 10445. – 2017. – P. 10445 – 10445-10. – <http://dx.doi.org/10.1117/12.2280975>.
8. Яровий А.А. Комп'ютерне моделювання процесу паралельного оброблення зображень на основі технологій OpenMP та NVIDIA CUDA / А. А. Яровий, О. О. Кулик, І. Р. Арсенюк // Оптико-електронні інформаційно-енергетичні технології. – 2015. – №2(30). – С. 33-39.

**Кулик Олександр Олександрович** — аспірант кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця, Хмельницьке шосе, 95, e-mail: o.kulyk@vntu.edu.ua.

**Яровий Андрій Анатолійович** — д.т.н., професор, завідувач кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця, Хмельницьке шосе, 95, e-mail: a.yarovy@vntu.edu.ua.

**Olexandr O. Kulik** — Postgraduate Student of Computer Science Department, Vinnytsia National Technical University, Vinnytsia, Khmelnytske Shose, 95, e-mail: o.kulyk@vntu.edu.ua.

**Andrii A. Yarovy** — Doctor of Science (Eng.), Professor, Head of Computer Science Department, Vinnytsia National Technical University, Vinnytsia, Khmelnytske shose, 95, e-mail: a.yarovy@vntu.edu.ua.