



Л. І. Мочурад, А. Б. Осідач

Національний університет "Львівська політехніка", Львів, Україна

ПАРАЛЕЛЬНИЙ МЕТОД RANSAC ДЛЯ ПОТОКОВОГО ОБРОБЛЕННЯ ДАНИХ СЕНСОРІВ LiDAR

Встановлено, що технологія LiDAR у кластеризації дає змогу підвищити ефективність потокового оброблення даних й отримати точніші результати. Проаналізовано, що проблема створення тривимірного контуру кімнати на підставі даних, зібраних сенсорами виявлення та визначення відстані LiDAR, є ключовим аспектом просторового аналізу, оскільки від цього залежить точність моделі довкілля простору. З'ясовано, що одним з ефективних підходів до вирішення цієї проблеми є використання методу консенсусу випадкової вибірки RANSAC, який дає змогу виділити і відокремити відхилення та випадкові помилки в даних, що отримані зі сенсорів. Запропоновано вдосконалити метод RANSAC шляхом використання технології паралельних обчислень. Це дало можливість розділити початкове завдання на менші частини та обробляти їх одночасно, що сприяє збільшенню швидкості оброблення даних і зменшенню тривалості виконання алгоритму. Оцінено продуктивність методу кластеризації у спосіб запуску програми на виконання 50 разів для послідовного та розпаралеленого алгоритмів, зберігаючи час кожного запуску для розрахунку середньої тривалості виконання. Наведено теоретичну апіорну оцінку показника пришвидшення для різної кількості паралельних потоків. Унаслідок цього здійснено порівняння фактичної ефективності з теоретичною оцінкою. Наведено результати, які вказують на масштабованість підходу та його потенціал для подальшого підвищення ефективності систем за використання більшої кількості обчислювальних ядер. Зменшено тривалість обчислення в 5,5 раза за використання 8 паралельних потоків порівняно з послідовним виконанням, що підкреслює значну перевагу паралельних обчислень під час оброблення великих даних, таких як хмари точок. Наведено результати оцінювання обчислювальної складності послідовного та вдосконаленого паралельного алгоритмів та проаналізовано показники ефективності паралельного. Подальші дослідження полягають у модифікації методів кластеризації на підставі LiDAR для різних застосувань, враховуючи робототехніку та геологічні дослідження, а також розроблення ефективних підходів до паралельного оброблення даних з інших типів сенсорів.

Ключові слова: кластеризація; сенсор; розпаралелення; тривимірні моделі; пришвидження.

Вступ / Introduction

Еволюція 3D-технологій та їх застосування в різних сферах є актуальною та перспективною тенденцією у сучасному світі. Однією з ключових особливостей цього розвитку є побудова контурів приміщень на підставі даних сканування LiDAR (англ. *Light Detection and Ranging*), що відзначається своєю новизною та потенційними застосуваннями у різних галузях, таких як будівництво та архітектура (проектування та реконструкція, будівельний контроль), геодезія та картографування (топографічне картографування, екологічний моніторинг), медицина (хірургічне планування, моделювання органів), автономні транспортні засоби (повітряні та водні безпілотні дрони, дорожні системи), віртуальна реальність (ігрова індустрія). Отже, формалізація та впровадження 3D-технологій, зокрема побудова контурів приміщень на підставі даних сенсора LiDAR, відкриває широкі можливості для інновацій у будівництві, геодезії, медицині, транспорті та розважальній індустрії.

Для технології LiDAR характерна висока точність отримання тривимірних даних, що забезпечує надійний

метод кластеризації об'єктів. Щоб поліпшити цей процес, можливо розглядати оптимізацію методів кластеризації для ще точнішого та швидшого визначення об'єктів у просторі. Додаткове вдосконалення може виникнути внаслідок поєднання лідару з іншими передовими технологіями, такими як штучний інтелект та машинне навчання. Це сприятиме покращенню адаптивності системи до різноманітних умов і завдань.

Хоча можливості технології LiDAR у розпізнаванні об'єктів і картографуванні навколишнього середовища були визнані [5], однією з невіршених головних проблем є потреба в ефективній під час оброблення великих обсягів даних 3D-хмари точок. Зокрема, завдання кластеризації об'єктів у тривимірній кімнатній хмарі традиційно було обчислювано затратним процесом [12]. Методи кластеризації дають змогу ідентифікувати та класифікувати окремі об'єкти в тривимірній хмарі точок, що може істотно підвищити ефективність архітектурного планування та розвитку. У цьому сенсі паралелізм цих процесів широко досліджують, і в попередніх дослідженнях спостерігаються перспективні результати [8, 13].

Інформація про авторів:

Мочурад Леся Ігорівна, канд. техн. наук, доцент, кафедра систем штучного інтелекту.

Email: lesia.i.mochurad@lpnu.ua; <https://orcid.org/0000-0002-4957-1512>

Осідач Андрій Богданович, студент, кафедра систем штучного інтелекту. Email: andrii.osidach.shi.2022@lpnu.ua

Цитування за ДСТУ: Мочурад Л. І., Осідач А. Б. Паралельний метод RANSAC для потокового оброблення даних сенсорів LiDAR. Науковий вісник НЛТУ України. 2024, т. 34, № 3. С. 110–116.

Citation APA: Mochurad, L. I., & Osidach, A. B. (2024). Parallel RANSAC method for streaming data from LiDAR sensors. *Scientific Bulletin of UNFU*, 34(3), 110–116. <https://doi.org/10.36930/40340314>

Використання паралельних обчислень під час оброблення даних технології LiDAR є сучасною сферою досліджень, яка прагне пришвидшити аналіз й оброблення великих масивів даних хмари точок. Розподіляючи завдання між кількома процесорами, стає можливим обробляти більші набори даних, значно скорочуючи тривалість обчислення. Це особливо важливо, враховуючи збільшення міських просторів і подальший попит на складніший, широкомасштабний аналіз просторових даних [7].

Об'єкт дослідження – використання паралельних обчислень та кластеризації об'єктів у тривимірних хмарах точок, отриманих з LiDAR.

Предмет дослідження – ефективність та результативність синтезу кластеризації об'єктів у тривимірних хмарах точок, що базуються на даних LiDAR, за допомогою методів паралельних обчислень.

Мета роботи – дослідити та розробити ефективний метод кластеризації об'єктів у тривимірних хмарах точок, які представляють дані про простір кімнати, отримані з використанням результатів сканування LiDAR та методів паралельних обчислень, що дасть змогу істотно покращити розуміння та аналіз простору кімнати для роботизованих систем, таких як роботи-пилососи, роботи-доставки та інших автономних або напівавтономних пристроїв.

Для досягнення зазначеної мети визначено такі основні завдання дослідження:

- проаналізувати стан проблеми кластеризації об'єктів у тривимірних хмарах точок, що дасть змогу виокремити недоліки у попередніх дослідженнях та підкреслити переваги запропонованого підходу;
- удосконалити метод кластеризації об'єктів у просторі на підставі технологій паралельних обчислень, що дасть змогу зменшити обчислювальну складність методу та пришвидшити аналіз й оброблення великих масивів даних хмари точок;
- дослідити вплив застосування паралельних обчислень на швидкість та ефективність оброблення великих обсягів даних LiDAR, що забезпечить можливість прийняття рішень у режимі реального часу;
- провести порівняння апіорного теоретичного оцінювання показника пришвидшення та фактичного на підставі запропонованого паралельного чисельного методу кластеризації об'єктів у тривимірних хмарах точок.

Аналіз останніх досліджень та публікацій. Багато дослідників визнали цінність даних технології LiDAR для створення точних 3D-зображень кімнат та інших просторів [11]. Проте ще потрібно розв'язувати ключові проблеми, зокрема щодо методів оброблення даних, які це дослідження має на меті вивчити та зробити свій внесок у їх вирішення. Однією з найактуальніших проблем, пов'язаних з даними LiDAR, є величезний обсяг згенерованих даних хмар точок. Оскільки технологія LiDAR продовжує розвиватися, то обсяг і складність даних на її основі постійно зростає, створюючи великі проблеми для їх ефективного оброблення та аналізу. Наявні методи оброблення цих даних, зокрема з метою кластеризації об'єктів, часто вимагають значних обчислювальних ресурсів і часу [6].

Стосовно завдання кластеризації об'єктів у кімнатних хмарах у літературі запропоновано кілька алгоритмів, таких як K-середніх [18], DBSCAN [3] та ієрархічна кластеризація [15]. Однак ці алгоритми мають проблеми з масштабованістю та продуктивністю під час об-

роблення великих обсягів даних високої роздільної здатності. Окрім цього, вони часто вимагають ретельного налаштування параметрів для досягнення задовільних результатів, що може бути непрактичним для нефакхівців або в автоматизованих процесах.

Паралельні обчислення були визначені як перспективний підхід для вирішення цих проблем. Одночасне виконання кількох підзадач значно пришвидшує оброблення великих наборів даних [9]. Однак, попри потенціал, застосування паралельних обчислень у кластеризації об'єктів для тривимірних кімнатних хмар не було достатньо вивчено в літературі, особливо щодо архітектурного та будівельного контекстів. Існують дослідження, де використовують паралельні обчислення для оброблення даних LiDAR. Більшість з них зосереджені на моделюванні рельєфу та географічних застосуваннях, проте не враховують внутрішнє середовище та архітектурної роботи, розроблення та оптимізація паралельних обчислювальних підходів для кластеризації об'єктів у тривимірних кімнатних хмарах точок є новою та необхідною сферою дослідження.

Техніка воксельного забарвлення, описана в роботі [19], справді є революційним підходом до 3D-реконструкції в реальному часі. Перекладаючи проблему 3D-реконструкції на узгодженість кольорів у вокселях, алгоритм може створювати детальні моделі з кількох зображень. Однак їх метод обмежений роздільною здатністю вхідних зображень, і він за своєю суттю не підтримує інтеграцію даних LiDAR, які можуть забезпечити додаткову геометричну деталізацію та точність. З іншого боку, автори у роботі [4] спробували вирішувати проблему побудови детальних 3D-моделей способом інтеграції даних LiDAR із фотограмметрією з багатьох джерел. Хоча цей підхід використовує переваги обох джерел даних, він потребує складних методів калібрування та об'єднання даних. Ефективність їх процесу для більших наборів даних чітко не продемонстрована. Нещодавні досягнення у глибокому навчанні також значно вплинули на оброблення хмари точок. У роботі [17] автори досліджували цей спосіб, переглядаючи різні алгоритми глибокого навчання для сегментації набору даних хмари точок. Однак, їхній аналіз виявив, що хоча ці методи можуть запропонувати виняткові результати, вони потребують значних обчислювальних ресурсів і часто мають проблеми з великомасштабними наборами даних.

У роботі [1] автори досліджували паралельні обчислення, щоб вирішити проблему оброблення великих наборів даних хмар точок. Запропонований підхід є перспективним у покращенні ефективності оброблення, проте він здебільшого теоретичний і не вистачає практичного застосування чи варіантів використання, особливо в архітектурному контексті.

Отже, попри те, що технологія LiDAR і алгоритми кластеризації об'єктів зазнали прогресу, залишається нагальна потреба в ефективніших методах оброблення даних. Застосування паралельних обчислень у цьому контексті виглядає перспективним, але наразі недостатньо вивченим, що підкреслює актуальність зосередженості цього дослідження на розробленні та оптимізації підходів паралельних обчислень для кластеризації об'єктів на підставі даних LiDAR.

Результати дослідження та їх обговорення / Research results and their discussion

Метод RANSAC для досліджуваної задачі. Цей метод спрямований на ефективну кластеризацію об'єктів у тривимірній кімнатній хмарі, отриманій від сканування LiDAR, і його головною метою є істотне зменшення обчислювальної складності, що зазвичай пов'язана з цим завданням, для чого використовують RANSAC (англ. *RANdom Sample Consensus*) [2], який є надійним і універсальним методом підлаштування моделі та розпізнавання об'єктів у даній хмарі точок.

Метод RANSAC розглядають як ітеративний процес з двома кроками:

1. На першому кроці вибірково підмножина, що складається з мінімальної кількості елементів даних, вибирається випадково із вхідного набору даних. Параметри моделі обчислюються з використанням елементів цієї вибіркової підмножини. Властивості підмножини вибірки, такі як кількість даних, визначають параметри моделі.
2. На другому кроці метод перевіряє, які елементи всього набору даних узгоджуються з моделлю, параметри якої були оцінені на першому кроці. Елемент вважається викидом, якщо він не відповідає моделі в межах заданого порогового значення помилки, яке визначає максимальне відхилення даних внутрішніх елементів. Елементи, що перевищують це відхилення, визнаються викидами.

Набір точок, які задовольняють модель, отриманий для підлаштування моделі, називають консенсусним набором. Метод RANSAC буде ітеративно повторювати два зазначені вище кроки, доки отриманий консенсус-набір у певній ітерації не матиме достатньо кількості точок. У роботі запропоновано метод, розроблений для роботи з даними LiDAR для створення 3D-контурів кімнати. Алгоритм використовуватиме RANSAC для пошуку площин з даних хмари точок, які відповідають стінам, стелі та підлозі. Ось основні кроки:

- 1) Збір даних: давач LiDAR збирає дані хмари точок кімнати. Кожна точка в хмарі визначається її декартовими координатами (x, y, z) .
- 2) Генерація моделі площини: для кожної ітерації в методі RANSAC вибиратиметься випадкова підмножина з трьох точок із хмари точок. Ці три точки визначають площину за допомогою рівняння площини $Ax + By + Cz + D = 0$. Тут (A, B, C) – вектор нормалі до площини, D – зміщення площини від початку координат, а (x, y, z) – точки на площині.
- 3) Консенсусний підрахунок: для згенерованої моделі площини підраховуватиметься кількість точок у межах певного порогового значення відстані (t). Ці точки вважаються внутрішніми, тоді як точки за межами цього порогу – викидами.
- 4) Оцінювання та оновлення моделі: найкраща модель оновлюється, якщо кількість консенсусних точок є найбільшою, зареєстрованою на цей час.
- 5) Послідовність: кроки 2-4 повторюються для заздалегідь визначеної кількості ітерацій (k).
- 6) Покращення: коли знайдено найкращу модель, то виконують її покращення, переоцінивши параметри площини з використанням усіх внутрішніх елементів цієї моделі.
- 7) Сегментація: розмежування хмари точок на підставі знайдених моделей площини, у такий спосіб позначивши стіни, підлогу, стелю та інші елементи кімнати.

Рівняння площини $Ax + By + Cz + D = 0$ використовують для створення моделей площини. Кількість консенсусних точок – i для даної моделі площини – P_i , об-

числюють підрахунком точок, для яких абсолютне значення їх відстані зі знаком до площини менше t (поріг):

$$I(P) = \left\{ \left| \frac{Ax + By + Cz + D}{\sqrt{A^2 + B^2 + C^2}} \right| < t \right\},$$

де: (x, y, z) – декартові координати точки; (A, B, C) – коефіцієнти вектора нормалі до площини; D – зміщення площини від початку координат; t – поріг відстані для визначення консенсусних точок; i – індекс ітерацій RANSAC; P_i – i -та модель площини в RANSAC.

Розпаралелення методу RANSAC. Для реалізації запропонованого паралельного методу обрано Python як мову програмування завдяки її простоті у використанні, універсальності, великій кількості бібліотек і модулів, а також широкому визнанню в науковій спільноті. Multiprocessing модуль Python спеціально вибрано для розпаралелювання запропонованого методу. Він надає потужні функціональні можливості для паралельного виконання незалежних завдань, оптимально використовуючи декілька процесорів у комп'ютері. Серед цих функцій була обрана функція `Pool.starmap()` через її засосовність до запропонованого методу.

Запропонований підхід передбачає поділ початкових даних – хмари точок на кілька частин. Потім ці підплощини обробляються окремо за допомогою методу RANSAC через `multiprocessing.Pool.starmap()`. Ця функція створює `pool` робочих процесів і відображає ітерацію аргументів у функцію, даючи можливість їй виконуватися паралельно. Вона схожа на вбудовану функцію Python `map()`, але підтримує кілька аргументів і допускає паралельне виконання.

Розпаралелюючи кластеризацію цих підплощин хмари точок використовуючи метод RANSAC, можемо ефективно використовувати наявні обчислювальні ресурси. Кожен робочий процес може зосереджуватися на одній частині, у такий спосіб значно скорочуючи загальну тривалість обчислення. Це особливо важливо під час роботи з великими масивами даних 3D-хмари точок, оскільки дає змогу швидше обробляти дані та кластеризувати об'єкти. Після роздільної кластеризації підплощин хмари точок, програма перевіряє крайні точки (точки перетину хмари точок після поділу) на приналежність до одного кластера. Цей етап постоброблення є вирішальним для забезпечення узгодженості та цілісності загальної моделі.

Отже, багатопроцесорний модуль Python `i`, зокрема, функція `Pool.starmap()` є ідеальним рішенням для розпаралелювання запропонованої методології кластеризації об'єктів на підставі RANSAC. Ці інструменти дають можливість нам зосередитися на алгоритмічній логіці та оптимізації, тоді як програмне забезпечення впроваджується з тонкощами паралельних обчислень. Цей підхід оптимізує використання обчислювальних ресурсів, істотно скорочуючи тривалість оброблення, і дає змогу нам ефективно обробляти великомасштабні тривимірні дані хмари точок.

Обчислювальна складність та показники ефективності запропонованого алгоритму. Гнучкість методу RANSAC дає змогу обробляти дані реального світу, які часто містять шуми та викиди. Оцінюючи обчислювальну складність, гарантуємо, що цей метод можна реалізувати з наявними обчислювальними ресурсами. Обчислювальна складність запропонованого методу RANSAC передусім залежить від кількості ітерацій (k),

розміру вхідних даних (N) і складності підлаштування моделі до даних. Для кожної ітерації метод має генерувати плоску модель і підраховувати консенсусні точки, що потребує одноразової перевірки кожної точки.

- Для кожної ітерації підлаштування моделі до мінімальної підмножини точок маємо складність $O(M)$, де M є складністю методу підлаштування.
- Оцінювання відповідності моделі відносно решти точок даних вимагає обчислення відстані для кожної точки, що призводить до складності $O(N)$.
- Алгоритм завершується або після максимальної кількості ітерацій (k), або коли знайдено модель з достатньою кількістю консенсусних точок.

Враховуючи ці чинники, загальну обчислювальну складність методу RANSAC можна приблизно оцінити як $O(k \cdot (M + N))$, де: M – складність алгоритму підлаштування; N – кількість точок у даних LiDAR; k – кількість ітерацій методу RANSAC. Очевидно, що використовуючи запропонований вище підхід для розпаралелення досліджуваного методу отримаємо обчислювальну складність $O(k \cdot (M + N/p))$, де p – кількість потоків. Показник пришвидшення при цьому можна обчислити за формулою $S_p(N) = T_1(N)/T_p(N)$, де: $T_1(N)$ – часова складність послідовного виконання методу; $T_p(N)$ – часова складність паралельного виконання методу для p потоків.

Дані, які було використано, містять хмару точок, згенеровану за допомогою сканування кухонної зони, використовуючи LiDAR. Хмара точок – це набір із 511026 окремих точок, кожна з яких має певне місце в 3D-просторі сканованої області. Представлення кухні було обране як середовище сканування через її складність і різноманітність об'єктів. Кухні зазвичай містять великі стаціонарні об'єкти (наприклад, кухонні меблі та побутова техніка) і менші рухомі об'єкти (наприклад, посуд й інші кухонні предмети). Цей різноманітний діапазон об'єктів становить складний сценарій для запропонованого методу кластеризації та дає змогу перевірити та продемонструвати його надійність та ефективність.

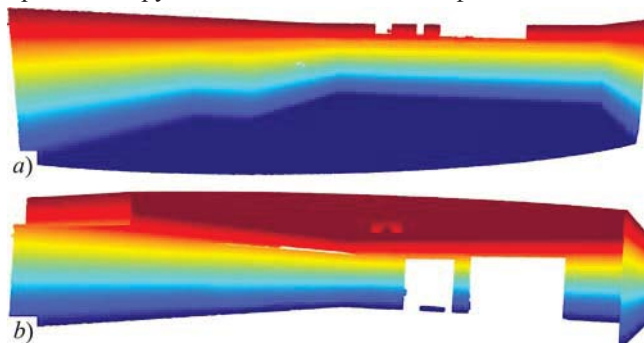


Рис. 1. Розмежування хмари точок на дві площини по осі OZ / Splitting a point cloud into two planes along the OZ axis

Запропонований метод передбачає поділ початкових даних – хмари точок уздовж осей OZ , OY та OX (на підставі кількості підкластерів, які нам потрібні), розділяючи їх на дві/чотири/шість/вісім окремих підчастинок, відповідно. Потім ці підчастини обробляються окремо, кожна з яких утворює незалежне завдання для окремого робочого процесу в запропонованій моделі паралельних обчислень. Далі, не зменшуючи загальності, наведемо результати запропонованого паралельного методу на двох паралельних потоках. Так, на рис. 1 зображено візуалізацію початкових даних хмари точок, розділених на дві підплощини вздовж осі OZ . Ця візуалізація забез-

печує чітке представлення запропонованого розподілу роботи між двома робочими процесами. На рис. 2 можемо побачити візуалізацію розподілених між кластерами даних – хмари точок після паралельного застосування методу RANSAC. Візуалізація висвітлює окремі кластери в кожній частині, ефективно демонструючи результати запропонованого підходу до розпаралелювання кластеризації об'єктів.

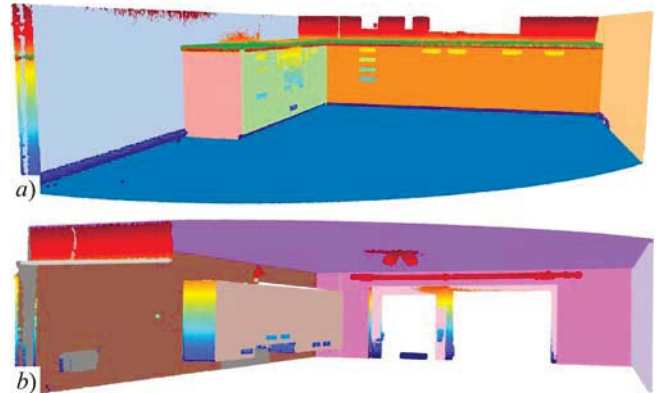


Рис. 2. Дві підплощини після паралельної кластеризації на двох потоках / Two subplanes after parallel clustering on two threads

Після застосування методу RANSAC кожна підплощина хмари точок трансформується. Тепер кожна точка в хмарі пов'язана з певним кластером, який представляє окремий об'єкт або поверхню в кухонній зоні. Після роздільної кластеризації підчастинок хмари точок перевіряємо крайні точки (точки перетину хмари точок після поділу) на приналежність до одного кластера. Цей етап після оброблення є вирішальним для забезпечення узгодженості та цілісності загальної моделі. З результатом об'єднання під-площин в єдину хмару точок можна ознайомитись на рис. 3.



Рис. 3. Хмара точок після об'єднання кластеризованих двох підплощин / Point cloud after merging the clustered two subplanes

У таблиці наведено тривалість виконання послідовного та паралельного алгоритмів за різної кількості підзадач (кількість підзадач відповідає кількості потоків).

Таблиця. Тривалість виконання запропонованого алгоритму та отримане пришвидшення / Execution time of the proposed algorithm and resulting speedup

Кількість підзадач	$T_p(N)$, с	$S_p(N)$
2	1.11	1.7
4	0.63	2.99
6	0.49	3.87
8	0.34	5.5

Підсумовуючи згадані вище експерименти, результати чітко ілюструють переваги паралельних обчислень для завдання кластеризації об'єктів у 3D-хмарі точок. Розпаралелена реалізація методу RANSAC істотно скоротила час обчислень, забезпечивши покращення ефек-

тивності, що зростало пропорційно кількості робочих потоків. Це демонструє масштабованість запропонованого підходу та його потенціал для ще більшого підвищення ефективності багатоядерних систем. Далі для перевірки достовірності отриманих результатів проведемо теоретичне оцінювання можливості пришвидшення обчислень для різної кількості паралельних потоків $p = \{2, 4, 6, 8\}$, які було використано для кластеризації початкової хмари точок $N = 511026$. Спочатку розрахуємо обчислювальні відповідні складності:

$$\begin{aligned} T_1(N) &= O(k \cdot (M + N)) = O(16 \cdot (1 + 511026)) = 8176432; \\ T_2(N) &= O(k \cdot (M + N/p)) = O(16 \cdot (1 + 511026/2)) = 4088224; \\ T_4(N) &= O(k \cdot (M + N/p)) = O(16 \cdot (1 + 511026/4)) = 2044120; \\ T_6(N) &= O(k \cdot (M + N/p)) = O(16 \cdot (1 + 511026/6)) = 1362752; \\ T_8(N) &= O(k \cdot (M + N/p)) = O(16 \cdot (1 + 511026/8)) = 1022068. \end{aligned}$$

Отже, показники пришвидшення обчислень при $N = 511026$ і різних p потоках матимуть такі значення:

$$S_2(N) = 2,0; S_4(N) = 4,0; S_6(N) = 5,9999; S_8(N) = 7,9999.$$

Усі чисельні експерименти проводили на ноутбучі з 4-ядерним процесором, який підтримує технологію Hyper Threading, яка, водночас, надає можливість отримати додаткові 4 обчислювальні потоки (віртуальні потоки). Тому порівнюючи теоретичні оцінювання пришвидшення з експериментальними (див. таблицю) можемо зробити висновок про достовірність отриманих результатів, адже, хоч і можливо використовувати 8 паралельних потоків для кластеризації, навантаження на процесор значно збільшується, що призводить до втрати ефективності проведення обчислень. Тому за збільшення кількості паралельних потоків, які використовуємо для кластеризації, можемо побачити, що фактичне оцінювання пришвидшення значно відрізняється від теоретичного, також фактична кількість ітерацій кластеризації за допомогою RANSAC методу має зменшуватись зі збільшенням кількості поділу початкової хмари точок на підплощини, що водночас має пришвидшувати задачу кластеризації кожної підплощини).

Обговорення результатів дослідження. Автори роботи [8] розглянули проблему визначення позиції технології LiDAR на площині з оптимальною точністю, що важливо в робототехніці, навігації та інших галузях. Автори пропонують паралельний алгоритм на базі технології OpenMP для вирішення цієї проблеми. Вони розробили ефективний метод, який дає змогу істотно пришвидшити обчислювальний процес та досягти високої точності. Експерименти показали, що показник пришвидшення прямує до кількості ядер, а ефективність – до одиниці. Автори також вказують на перспективи розвитку методу, враховуючи максимальну оптимізацію за допомогою інших технологій розподілених і паралельних обчислень та розширення досліджень на складніші просторові області.

Робота [15] зосереджена на проблемі кластеризації для сегментації точкових хмар технології LiDAR у міських середовищах. Вона пропонує ієрархічний алгоритм кластеризації, який використовує класифікацію форм і перерозподіл викидів для ефективної ідентифікації різних форм структур. Запропонований метод поєднує грубу сегментацію за допомогою методу Gaussian mapping та DBSCAN з подальшою дрібною сегментацією за допомогою моделей BIP, що дає змогу досягти точної екстракції геометричних примітивів. Результати експериментів показують високу точність та якість

сегментації порівняно з іншими сучасними методами, що підкреслює актуальність роботи в контексті кластеризації в просторі.

У роботі [14] розглянуто проблему точності та ефективності фільтрації ґрунту в точкових хмарах технології LiDAR для виявлення тривимірних об'єктів. Способом аналізу принципів методу RANSAC та вдосконалення їх, автори пропонують новий метод фільтрації ґрунту, який демонструє вищу точність та ефективність порівняно з наявними методами. Вони також виявили, що їх метод може покращити точність виявлення об'єктів у точкових хмарах, не ушкоджуючи швидкодії алгоритму. Унаслідок експериментів на даних наборах ISPRS та KITTI було підтверджено ефективність запропонованого методу.

Оцінювання продуктивності сенсорів LiDAR для автономного водіння та систем допомоги водієві у реальних умовах експлуатації наведено у роботі [10]. Автори зазначають, що найбільшою проблемою є дослідження можливості та надійності сенсора LiDAR за різних погодних умов, що є важливим для забезпечення автомобільного руху. Запропоновано метод тестування, який використовує алгоритм просторово-часової сегментації точок для аналізу продуктивності сенсора LiDAR у динамічних сценаріях. Цей алгоритм може розглядатися як форма кластеризації, де точки, отримані від сенсора LiDAR, групуються в різні кластери залежно від їх просторового та часового розташування. Такий підхід дає змогу виділити рухомі об'єкти (наприклад, автомобілі) від статичного фону, що може бути корисним для оцінювання продуктивності сенсора LiDAR у реальних умовах. Отже, хоча саме дослідження не вирішує прямо проблему кластеризації, воно використовує підходи, які мають відношення до цієї теми, зокрема алгоритм просторово-часової сегментації, що може розглядатися як одна з форм кластеризації даних.

У дослідженні [16] розглянуто проблему миттєвого виявлення рухомих подій у динамічних середовищах з використанням камер, призначених для реєстрації подій. Однак використання сенсорів LiDAR для цієї задачі не було достатньо досліджено. Наявні підходи зазвичай полягають у нагромадженні точок сенсорів LiDAR у кадри та виявленні рухів на рівні об'єктів, що призводить до затримок від десятків до сотень мілісекунд. У роботі пропонують новий підхід, названий M-detector, який визначає, чи є точка рухомою безпосередньо після її надходження, що призводить до виявлення точки за точкою зі затримкою всього кілька мікросекунд. M-detector спроектований на підставі принципів оцінювання процедури заслонування та його можна використати в різних середовищах з різними типами давачів LiDAR. Експерименти показали ефективність M-detector на різних наборах даних і застосуваннях, демонструючи його переваги у точності, обчислювальній ефективності, затримці виявлення та загальній універсальності.

Отже, за результатами виконаної роботи можна сформулювати такі наукову новизну та практичну значущість результатів дослідження.

Наукова новизна отриманих результатів дослідження – розроблено паралельний метод кластеризації об'єктів у тривимірних хмарах точок, що дає змогу істотно підвищити ефективність та результативність аналізу тривимірних просторових даних і, відповідно, зменшити обчислювальну складність та пришвидшити

оброблення великомасштабних тривимірних даних хмари точок.

Практична значущість результатів дослідження – запропонований підхід дає змогу покращити навігацію роботів у різноманітних середовищах, є важливим для робототехніки, оскільки використання точних тривимірних зображень кімнат та об'єктів, отриманих із сенсора LiDAR, сприяє підвищенню безпеки та ефективності роботів у завданнях, таких як уникнення перешкод та взаємодія з оточенням, а паралельні обчислення водночас дають змогу оптимізувати тривалість оброблення великих обсягів даних, що є критичним для реального часу реакції роботів на зміни у середовищі.

Висновки / Conclusions

Досліджено та розроблено ефективний метод класифікації об'єктів у тривимірних хмарах точок, які представляють дані про простір кімнати, отримані з використанням результатів сканування сенсора LiDAR та методів паралельних обчислень. За результатами дослідження можна зробити такі основні висновки.

1. Проаналізовано сучасний стан проблеми кластеризації об'єктів у тривимірних хмарах точок. Систематизовано вирішено проблему кластеризації об'єктів у 3D-хмарах точок, яка виникла під час аналізу хмари точок, отриманої за допомогою LiDAR сканування кухонної зони. Встановлено, що точна кластеризація має вирішальне значення для архітектурного проектування, автономної навігації робототехніки та розширеної реальності, сприяючи ефективнішому розпізнаванню об'єктів та підвищуючи ефективність застосувань.
2. Удосконалено метод кластеризації об'єктів RANSAC у просторі з використанням розпаралелення підзадач, що дало змогу істотно пришвидшити обчислення та підвищити їх ефективність.
3. Досліджено вплив застосування паралельних обчислень на швидкість та ефективність оброблення великих обсягів даних LiDAR. З'ясовано, що паралельний алгоритм дає змогу зменшити тривалість обчислення у 5,5 раза при використанні 8 паралельних потоків порівняно з послідовним виконанням, що підкреслює перевагу паралельних обчислень під час оброблення великих даних, таких як хмари точок.
4. Проведено порівняння апріорного теоретичного оцінювання показника пришвидшення та фактичного, отриманого на підставі низки чисельних експериментів. Унаслідок такого порівняння доведено достовірність отриманих результатів.

Отже, у роботі проведено дослідження, що має важливе значення у галузі наукових досліджень, спрямованих на паралельні обчислення та кластеризацію 3D-об'єктів, відкриваючи перспективний напрям для подальших досліджень.

References

1. Alis, C., Boehm, J., & Liu, K. (2016). Parallel processing of big point clouds using Z-Order-based partitioning. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-ISPRS Archives*, 41, 71–77. International Society of Photogrammetry and Remote Sensing (ISPRS). <https://doi.org/10.5194/isprs-archives-XLI-B2-71-2016>
2. Béjar-Martos, J. A., Rueda-Ruiz, A. J., Ogayar-Anguita, C. J., Segura-Sánchez, R. J., & López-Ruiz, A. (2022). Strategies for the Storage of Large LiDAR Datasets – A Performance Comparison. *Remote Sensing*, 14(11). <https://doi.org/10.3390/rs14112623>
3. Chen, X., An, Q., Han, X., Ban, Y., & Li, L. (2023). Control of distributed segmentation of indoor point cloud via homogenization clustering network. *Journal of the Franklin Institute*, 360(12), 8704–8739. <https://doi.org/10.1016/j.jfranklin.2021.12.001>
4. Cheng, L., Tong, L., Chen, Y., Zhang, W., Shan, J., Liu, Y., & Li, M. (2013). Integration of LiDAR data and optical multi-view images for 3D reconstruction of building roofs. *Optics and Lasers in Engineering*, 51(4), 493–502. <https://doi.org/10.1016/j.optlaseng.2012.10.010>
5. Eltner, A., Kaiser, A., Castillo, C., Rock, G., Neugirg, F., & Abellán, A. (2016). Image-based surface reconstruction in geomorphometry – merits, limits and developments. *Earth Surface Dynamics*, 4(2), 359–389. <https://doi.org/10.5194/esurf-4-359-2016>
6. Jung, J., Che, E., Olsen, M. J., & Parrish, C. (2019). Efficient and robust lane marking extraction from mobile lidar point clouds. *ISPRS journal of photogrammetry and remote sensing*, 147, 1–18. <https://doi.org/10.1016/j.isprsjprs.2018.11.012>
7. Mochurad, L. I., & Mamchur, M. V. (2023). Parallel and distributed computing technologies for autonomous vehicle navigation. *Radio Electronics, Computer Science, Control*, 4, 111. <https://doi.org/10.15588/1607-3274-2023-4-11>
8. Mochurad, L., & Kryvinska, N. (2021). Parallelization of finding the current coordinates of the lidar based on the genetic algorithm and OpenMP technology. *Symmetry*, 13(4). <https://doi.org/10.3390/sym13040666>
9. Mochurad, L., Sydor, A., & Ratinskiy, O. (2023). A fast parallelized DBSCAN algorithm based on OpenMp for detection of criminals on streaming services. *Frontiers in big Data*, 6. <https://doi.org/10.3389/fdata.2023.1292923>
10. Park, J., Cho, J., Lee, S., Bak, S., & Kim, Y. (2023). An Automotive LiDAR Performance Test Method in Dynamic Driving Conditions. *Sensors*, 23(8). <https://doi.org/10.3390/s23083892>
11. Pu, S., & Vosselman, G. (2009). Knowledge based reconstruction of building models from terrestrial laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(6), 575–584. <https://doi.org/10.1016/j.isprsjprs.2009.04.001>
12. Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3 d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 652–660. <https://doi.org/10.1109/CVPR.2017.16>
13. Shen, J., Liu, N., & Sun, H. (2021). Vehicle detection in aerial images based on lightweight deep convolutional network. *IET Image Processing*, 15(2), 479–491. <https://doi.org/10.1049/ipr2.12038>
14. Wang, B., Lan, J., & Gao, J. (2022). LiDAR Filtering in 3D Object Detection Based on Improved RANSAC. *Remote Sensing*, 14(9). <https://doi.org/10.3390/rs14092110>
15. Wang, F., Zhou, G., Xie, J., Fu, B., You, H., Chen, J., & Zhou, B. (2023). An automatic hierarchical clustering method for the LiDAR point cloud segmentation of buildings via shape classification and outliers reassignment. *Remote Sensing*, 15(9), 2432.
16. Wu, H., Li, Y., Xu, W., et al. (2024). Moving event detection from LiDAR point streams. *Nature Communications*, 15, 345. <https://doi.org/10.1038/s41467-023-44554-8>
17. Zhang, J., Zhao, X., Chen, Z., & Lu, Z. (2019). A review of deep learning-based semantic segmentation for point cloud. *IEEE access*, 7, 179118–179133. <https://doi.org/10.1109/ACCESS.2019.2958671>
18. Zhao, H. (2022). Design and implementation of an improved K-means clustering algorithm. *Mobile Information Systems. Special Issue*, Vol. 2022, Article ID 6041484. <https://doi.org/10.1155/2022/6041484>
19. Zhao, Y., Zhang, X., & Huang, X. (2021). A technical survey and evaluation of traditional point cloud clustering methods for lidar panoptic segmentation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2464–2473. <https://doi.org/10.48550/arXiv.2108.09522>

PARALLEL RANSAC METHOD FOR STREAMING DATA FROM LIDAR SENSORS

LiDAR technology finds applications in diverse fields such as topographic mapping, urban planning, forestry management, disaster response, infrastructure inspection, autonomous vehicles, archaeology, mining, meteorology, and environmental monitoring. Its integration within clustering methodologies offers improved data processing efficiency and result accuracy. A critical aspect of spatial analysis involves delineating three-dimensional room contours based on data acquired from LiDAR detection and ranging sensors, as the fidelity of the surrounding space model hinges upon this process. Investigation reveals that leveraging the RANSAC (Random Sample Consensus) algorithm proves to be an effective strategy for tackling this challenge, facilitating the identification and segregation of deviations and random errors within sensor data. This study proposes an enhancement to the RANSAC algorithm through the integration of parallel computing technologies. This advancement enables the subdivision of the original problem into smaller segments, allowing for concurrent processing and consequent acceleration of data processing speed alongside a reduction in algorithm execution duration. The efficacy of the clustering algorithm is evaluated by executing the program 50 times in both sequential and parallel modes while maintaining consistent run times for each iteration to compute average execution durations. Theoretical speedup estimates are computed for varying numbers of parallel threads, subsequently juxtaposed against actual efficiency metrics. Findings indicate scalability and potential for further efficiency enhancements within systems employing additional computing cores. Notably, employing 8 parallel threads yields a reduction in computation time by a factor of 5.5 compared to sequential execution, underscoring the potency of parallel computing in handling vast datasets like point clouds. Additionally, computational complexity is assessed, and the performance of the proposed approach is analyzed. Future research endeavors aim to optimize LiDAR-based clustering algorithms for diverse applications encompassing robotics and geology. Furthermore, emphasis will be placed on devising efficient strategies for parallel processing across different data types and sensor modalities.

Keywords: clustering; sensor; parallelization; three-dimensional models; acceleration.