
Використання детермінованих методів для локалізації людей в приміщеннях

Андрій Топольський

Кафедра автоматизації та інтелектуальних інформаційних технологій, Вінницький Національний Технічний Університет (ВНТУ), м.Вінниця, Україна
ORCID 0009-0007-7631-0872

Євген Паламарчук

Кафедра автоматизації та інтелектуальних інформаційних технологій, Вінницький Національний Технічний Університет (ВНТУ), м.Вінниця, Україна
ORCID 0000-0002-7443-099X

Анотація: У роботі досліджено застосування детермінованих методів для визначення місцезнаходження людей у приміщенні, а саме – метод визначення найменшої відстані від людини до центрів приміщень та метод порівняння координат людини з площею приміщень. Проаналізовано необхідні умови для роботи кожного методу. Показано, що застосування методу найменшої відстані є оптимальним в умовах однорідних форм та розмірів сусідніх приміщень, адже при їх неоднорідності людина поблизу спільних стін може виявитися ближчою до центру сусіднього приміщення. Для методу порівняння координат людини з площею приміщень необхідними умовами є достатня кількість (не менше трьох точок координат) і якість (розповсюдженість, різність) вхідних даних, адже при їх нестачі зменшиться точність визначення місцезнаходження людини. Для обох випадках критично важливо мати інформацію про поверх, де знаходиться людина, оскільки координати довготи і широти не враховують висоту, що призводить до колізій даних у багатоповерхових будівлях. Для вирішення цієї проблеми доцільно застосовувати методи машинного навчання на основі RSSI від WAP. Окрім умов роботи детермінованих методів, у роботі також досліджено вплив методів, що використовуються для обробки даних. Для отримання координат центрів приміщень було порівняно методи арифметичної центроїду та центроїди полігональної області. Показано, що арифметична центроїда зсувається до великих скупчень даних, тоді як полігональна центроїда більш стійка до нерівномірного розподілу даних. Для отримання відстані між двома точками було порівняно методи евклідової відстані та гаверсінуса. Показано, що вплив даних методів на точність методу найменшої відстані до центрів залежить від системи координат, в якій подано дані. Для глобальних географічних даних (WGS 84/EPGS:3857) кращий результат показав метод гаверсінуса, а для локальних – метод евклідової відстані. Показано, що при дотриманні вказаних умов обидва методи забезпечують достатньо високу точність локалізації людей в приміщенні.

Ключові слова: автоматизовані системи відвідуваності, електронні навчальні системи, локалізація людей в приміщенні, детерміновані методи визначення місцезнаходження, метод опуклої оболонки, методи визначення відстані.

1. Вступ

Однією із задач закладів освіти є контроль відвідуваності занять студентами. Найрозповсюдженішим методом для цього є ручний метод – викладач проводить опитування присутності студентів та записує результати у паперовий журнал. Цей метод має дві основні

проблеми – викладач може помилитися при заповненні журналу і процес опитування займає час.

Розвиток інформаційних технологій надає можливість реалізації різнопланових автоматизованих систем електронної освіти. Зокрема, такі системи дозволяють автоматизувати процес контролю відвідування студентами занять. Автоматизація дозволяє уникнути людського фактору у заповненні паперових журналів та дозволить використати час, який витрачався для здійснення цих процедур, на викладання дисципліни.

Основною задачею такої системи є визначення присутності студента на занятті з високою точністю. У контексті навчальних закладів достатньо великою точністю слід вважати 100% або величини близькі до 100%. Для прикладу, якщо розглянути заняття, на якому присутні 100 людей, то система, що має 95% точність визначення, помилково може зафіксувати присутність або відсутність до п'яти людей на занятті. У цьому випадку такі помилки є неприйнятними у навчальному процесі.

Дане дослідження зосереджено на використанні детермінованих методів визначення місцезнаходження об'єктів у приміщенні, за допомогою яких можливо отримати достатньо високу точність у випадку, коли умови експлуатації системи будуть відповідати певним умовам.

2. Об'єкт і предмет дослідження

Об'єктом дослідження є автоматизовані системи локалізації людей у приміщеннях. Предметом дослідження є математичні методи для визначення місцезнаходження людей або об'єктів у просторі.

3. Мета та задачі дослідження

Мета дослідження – визначення місцезнаходження об'єктів (у тому числі людей) з використанням детермінованих методів ідентифікації приміщення.

Гіпотеза дослідження – використання детермінованих методів прогнозування дозволить суттєво підвищити точність визначення місцезнаходження об'єктів, якщо умови експлуатації системи будуть відповідати певним параметрам.

Для досягнення мети необхідно розв'язати наступні задачі:

- дослідити методи визначення центрів приміщення на основі координат об'єктів у ньому;
- дослідити методи визначення відстані між двома точками координат;
- визначити основні умови для використання методів визначення місцезнаходження людей в приміщенні;
- експериментально перевірити досліджені методи на відкритих вибірках даних.

4. Аналіз літератури

Автоматизовані системи моніторингу присутності студентів на заняттях передбачають збір і обробку даних щодо їх просторової локалізації.

Для збору даних можуть бути використані різні технології. Wi-Fi Hotspot займалися М. Choi [1], B.S. Panca [2]. У даному випадку застосовується спеціалізований мобільний застосунок, який у фоновому режимі відстежує підключення смартфона до Wi-Fi мережі кампусу (або приміщення). У разі під'єднання, застосунок автоматично формує та надсилає на центральний сервер повідомлення з унікальною MAC-адресою пристрою. На сервері отримана MAC-адреса порівнюється з інформацією в базі даних студентів навчального закладу. Якщо дані збігаються, система фіксує студента як присутнього на відповідній парі чи лекції.

Системами на основі технологій Bluetooth займалися N.A. Azim [3], A. Puckdeevongs [4]. У цьому методі точність визначення місцезнаходження досягається завдяки обмеженій зоні дії Bluetooth. Встановлення в аудиторії сканерів дозволяє отримувати одночасно декілька сигналів від студентських пристроїв та обчислювати місцезнаходження за потужністю сигналу (RSSI). При знаходженні користувача в зоні дії як мінімум трьох сканерів, система фіксує MAC-адресу його пристрою і порівнює її з обліковими записами університету. Різниця рівнів сигналу дозволяє встановити, чи перебуває студент у потрібній аудиторії.

Системами на основі технологій RFID займалися W. A. Farag [5], J. Lin [6]. Студенти використовують спеціальні пристрої з вбудованими RFID-мітками. В RFID-мітках зберігаються унікальні ідентифікаційні коди студентів. В аудиторіях встановлюються RFID-зчитувачі, які сканують оточення. При скануванні RFID-мітки у систему навчального закладу передається ідентифікаційний код студента, після чого система порівнює його з реєстром студентів.

Системами на основі технологій QR-коду займалися A. Manori [7], F. Masalha [8]. Такі системи складаються з двох основних компонентів: серверної частини та мобільного застосунку. Сервер генерує QR-коди для кожного заняття та приймає запити на фіксацію присутності. Студентський мобільний застосунок використовується для входу в особистий акаунт і сканування згенерованого QR-коду. На занятті викладач отримує QR-код з серверу та демонструє його, наприклад, на проекторі. Далі студентам необхідно за допомогою мобільного застосунку відсканувати QR-код. При успішному скануванні додаток надсилає на сервер ідентифікаційні дані студента. Якщо відсканований QR-код відповідає заняттю з розкладу студента, то система фіксує, що студент перебуває на занятті.

Системами на основі технологій розпізнавання обличчя займалися D. Sunaryono [9], D.N. Singh [10]. У кожній аудиторії встановлюються камери, які безперервно ведуть відеозйомку. Спеціальне програмне забезпечення за допомогою алгоритмів розпізнавання обличчя аналізує кожен кадр відео. Якщо програма знаходить у кадрі обличчя, кадр обрізається до розмірів обличчя. Після цього, за допомогою алгоритмів розпізнавання обличчя, обрізані кадри порівнюються з фотографіями студентів у базі даних. При успішному розпізнаванні обличчя система фіксує присутність студента на занятті.

Системами на основі технологій сканування відбитку пальця займалися A. Basila [11], M. Alhothaily [12]. У кожній аудиторії встановлюється спеціальний сканер, який має інтерфейс для вибору дисципліни. Після того, як студент вказав дисципліну, йому необхідно прикласти палець до сканера. Зчитаний відбиток порівнюється з даними в базі даних навчального закладу. При успішного порівнянні відбитків пальців система фіксує присутність студента.

Обробкою даних за допомогою методів машинного навчання займалися N. Singh [13], Ji-Hyun Yoo [14]. Робота машинного навчання базується на побудові моделей, які навчаються з статистичних даних. Такі моделі можуть прогнозувати вихідні дані на основі вхідних даних. Перший етап роботи таких систем передбачає збір та попередню обробку масиву даних, таких як сигнали від різноманітних сенсорів (Wi-Fi модулі, Bluetooth-маяки, тощо) та координати.

Другий етап – навчання моделі. Масив даних розбивається на дві вибірки - навчальну та тестову. Навчальна вибірка використовується для навчання моделі. На основі навчальної вибірки будується модель, яка прогнозує вихідні дані. Тестова вибірка використовується для тестування побудованої моделі. Це необхідно для того, щоб пересвідчитися, що побудована модель здатна робити прогнози для будь-яких даних з того ж процесу, а не лише для даних з навчальної вибірки.

Серед всіх розглянутих систем можна виділити дві основні проблеми. Перша проблема – необхідність додаткових активних дій з боку студента (використання застосунку на смартфоні, прикладання пальця до сканера). Друга проблема – присутність студента на занятті визначається з недостатньою точністю. У зв'язку з цим дане дослідження спрямоване на розроблення системи, яка зможе з достатньо високою ймовірністю зафіксувати присутність студента на занятті та водночас не вимагатиме від нього будь-яких додаткових активних дій.

5. Методи досліджень

5.1. Використані методи

Для визначення місцезнаходження людей в приміщенні було використано метод визначення найменшої відстані від людини до центрів приміщень та метод порівняння координат людини з площею приміщень. Для роботи обох методів необхідні статистичні дані, в яких точно визначено координати людини та приміщення, де вона знаходиться. Статистичні дані будуть використовуватися для порівняння з поточними координатами людини для визначення приміщення, де вона знаходиться.

Одним з методів отримання статистичних даних є проведення мануальних вимірювань у всіх приміщеннях, де будуть проводитися спостереження. Такий метод дозволяє отримати чітко визначену площу для кожного приміщення, яку можна використовувати для визначення місцезнаходження людей. Даний метод схематично зображено на рис. 1, де чорним позначено фактичну площу приміщення, червоним позначено мануально виміряну площу, кружками позначено людей у приміщенні:

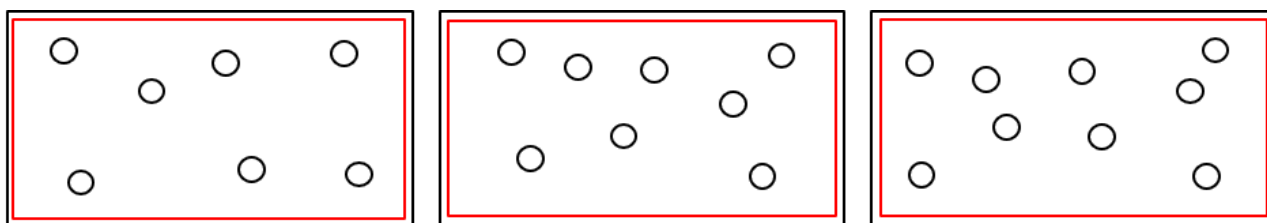


Рис. 1 Схематичне зображення мануального вимірювання площі приміщення.

Однак, в залежності від загальної кількості приміщень, де будуть проводитися спостереження, мануальні вимірювання можуть зайняти занадто багато часу та людського ресурсу. Окрім цього, такий підхід можна використати лише для задач, де спостережувана площа фіксована. Якщо розглядати задачу поза контексту моніторингу присутності студентів на занятті, то залежність від фіксованої спостережуваної площі є недоліком. До прикладу інших задач з не фіксованою спостережуваною площею, для яких можна застосувати дані методи, можна віднести спостереження за дикими тваринами у природі, моніторинг місцезнаходження тварин у сільському господарстві, аналіз логістичних маршрутів для середнього та крупного бізнесу, тощо.

Альтернативним варіантом отримання статистичних даних є використання даних користувачів, в даному випадку, студентів. До статистичних даних відносяться координати студентів та ідентифікатор приміщення, де вони знаходяться.

Для методу порівняння координат людини з площею приміщень необхідно для кожного приміщення виділити крайові координати та побудувати багатокутник, який буде описувати площу цього приміщення. У порівнянні з мануальним вимірюванням, такий підхід буде не ідеально описувати площу приміщення, але він дозволить отримати її аналітично. Даний метод схематично зображено на рис. 2, де червоними кружками позначено людей з крайовими координатами:

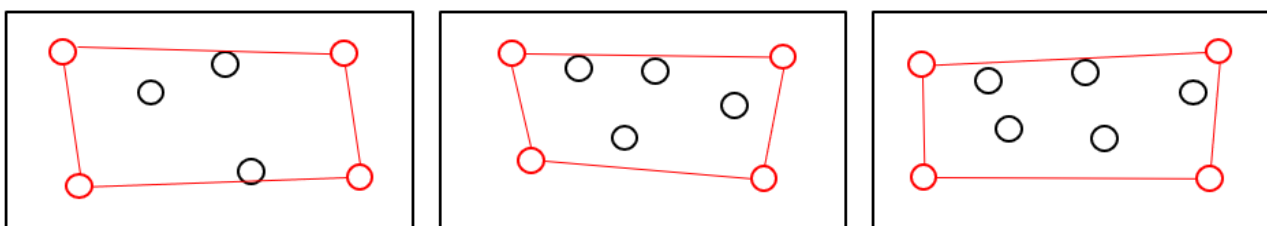


Рис. 2. Схематичне зображення методу порівняння координат людини з площею приміщень.

Необхідною умовою для роботи методу порівняння координат людини з площею приміщень є достатня кількість та розповсюдженість статистичних даних, щоб отримана площа максимально точно описувала площу приміщення. Якщо даних буде недостатньо, то побудована площа буде описувати лише частину площі, де можуть знаходитися студенти, що може зменшити точність роботи системи.

Іншою альтернативою є метод визначення найменшої відстані від людини до центрів приміщень. Для цього методу необхідно визначити точки координат, які будуть позначати центри всіх приміщень. Робота методу полягає у визначенні відстані від координат людини до кожного з центрів приміщень. Найменша з отриманих відстаней буде вказувати на приміщення, де знаходиться людина. Даний метод схематично зображено на рис. 3, де червоними кружками позначено центри приміщень, а пунктирними стрілками позначено відстані від людини до центрів:

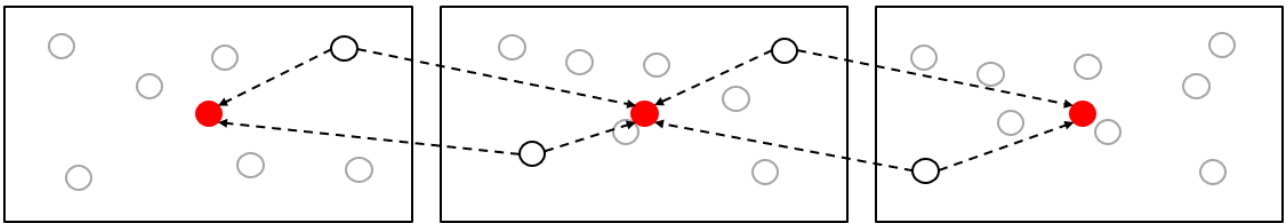


Рис. 3. Схематичне зображення методу визначення найменшої відстані від людини до центрів приміщень.

Необхідною умовою для роботи методу визначення найменшої відстані від людини до центрів приміщень є однорідність форм та розмірів сусідніх приміщень. Якщо сусідні приміщення будуть неоднорідні за формою або розміром, людина поблизу спільних стін може виявитися ближчою до центру сусіднього приміщення. Дану проблему схематично зображено на рис. 4:

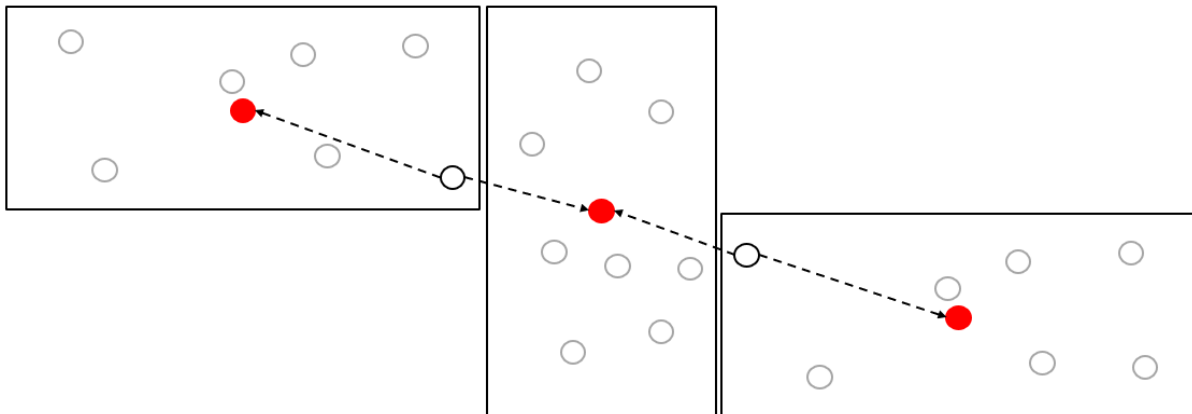


Рис. 4. Схематичне зображення проблеми неоднорідності форм сусідніх приміщень при визначенні найменшої відстані від людини їх центрів.

5.2. Використані математичні моделі

Для визначення центру приміщення на основі координат об'єктів у цьому приміщенні було використано методи арифметичної центроїди та центроїди полігональної області.

Метод арифметичної центроїди використовується для знаходження центру мас множини точок у просторі. Нехай, задано скінченну множину точок $S = \{x_1, x_2, \dots, x_n\}$. Центроїдом цієї множини S є точка C , яка мінімізує суму квадратів евклідових відстаней до всіх точок

множини S [15]. Точка C визначається як середнє арифметичне координат всіх точок множини S :

$$C = \left(\frac{1}{n} \sum_{i=1}^n x_i, \frac{1}{n} \sum_{i=1}^n y_i \right). \quad (1)$$

Приклади центроїд різних багатокутників зображено на рис. 5:

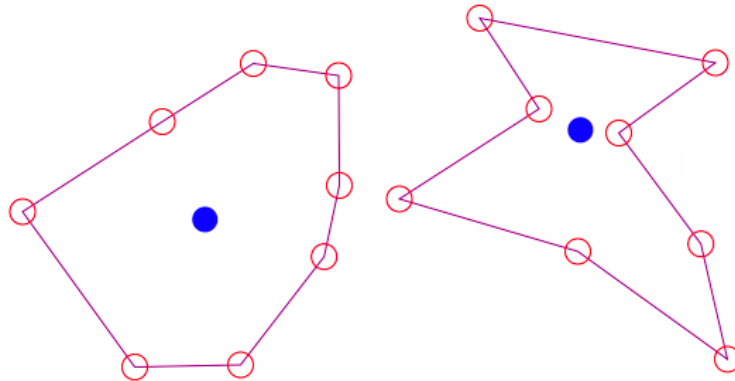


Рис. 5. Приклади центроїд різних багатокутників.

Опуклою оболонкою (convex hull) множини точок S у n -вимірному просторі називається перетин усіх опуклих множин, що містять точки S [16]. Іншими словами, опукла множина – найменший можливий багатокутник (для двовимірного простору) або поліедр (для тривимірного простору), який описує всю множину точок. Приклади опуклих оболонок зображено на рис. 6:

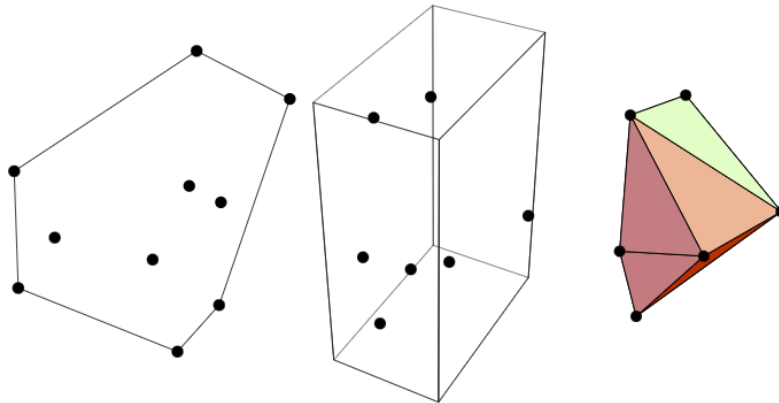


Рис. 6. Приклади опуклих оболонок для множин точок у двовимірному та тривимірному просторі.

Опуклу оболонку множини S з N точок $p_1, \dots, p_N$ можна виразити наступною формулою:

$$C = \left\{ \sum_{j=1}^N \lambda_j p_j; \lambda_j \geq 0 \text{ для всіх } j \text{ та } \sum_{j=1}^N \lambda_j = 1 \right\}, \quad (2)$$

де: λ – невід'ємний ваговий коефіцієнт точки p ;

p – точка множини S .

Використання опуклої оболонки дозволяє абстрагуватися від внутрішньої структури кластеру точок та використовувати лише крайні точки. Ці крайові точки формують

багатокутник, який описує площу, де знаходиться вся множина точок. В контексті автоматизованої системи моніторингу присутності студентів на занятті обчислення опуклої оболонки дозволяє аналітично отримати площу приміщення, де перебувають студенти. Також це дозволяє отримати центр цієї цього приміщення шляхом обчислення центроїди полігональної області на основі її вершин [17]:

$$A = \frac{1}{2} \sum_{i=0}^{n-1} (x_i y_{i+1} - x_{i+1} y_i), \quad (3)$$

$$C_x = \frac{1}{6A} \sum_{i=0}^{n-1} (x_i + x_{i+1}) (x_i y_{i+1} - x_{i+1} y_i), \quad (4)$$

$$C_y = \frac{1}{6A} \sum_{i=0}^{n-1} (y_i + y_{i+1}) (x_i y_{i+1} - x_{i+1} y_i), \quad (5)$$

де: A – площа багатокутника;

C_x та C_y – координати центроїда.

Для визначення відстані між двома точками було використано методи евклідової відстані та метод гаверсинуса.

Евклідова відстань – метод, який визначає найкоротшу відстань між двома точками в евклідовому просторі. Ця відстань відповідає довжині вектора, що з'єднує дві точки. Для двох точок $x = (x_1, x_2, \dots, x_n)$ та $y = (y_1, y_2, \dots, y_n)$ у n -вимірному просторі евклідова відстань визначається як [18]:

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, \quad (6)$$

де: d – евклідова відстань;

x_i, y_i – координати відповідних точок по i -тій осі.

Інший варіант метод для отримання відстані між двома точками – метод гаверсинуса. Метод гаверсинуса використовується для розрахунку відстані між двома точками на поверхні земної кулі по прямій лінії (рис.7):

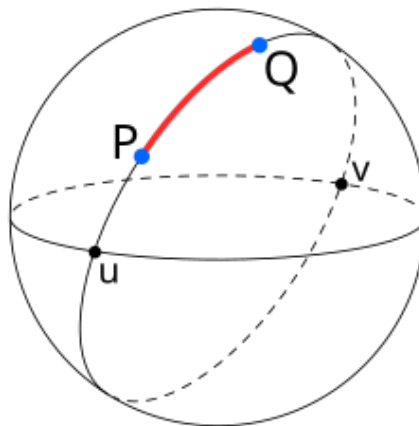


Рис. 7. Використання методу гаверсинуса для отримання відстані між двома точками на поверхні земної кулі.

В загальному вигляді метод гаверсинуса записується як [19]:

$$\alpha = \sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos\varphi_1 \cdot \cos\varphi_2 \cdot \sin^2\left(\frac{\Delta\lambda}{2}\right), \quad (7)$$

$$c = 2 \cdot \operatorname{atan2}\left(\sqrt{\alpha}, \sqrt{1-\alpha}\right), \quad (8)$$

$$d = R \cdot c, \quad (9)$$

де: φ – широта;

λ – довгота;

R – радіус Землі (6371 км);

d – відстань між двома точками.

Метод евклідової відстані базується на класичній евклідовій геометрії й передбачає, що простір є плоским і однорідним, тоді як метод гаверсинуса – на сферичній геометрії Землі, де поверхню розглядають як частину кулі. Таким чином, вибір між двома методами залежить від масштабу задачі. Для невеликих відстаней на обмежених плоских ділянках евклідова відстань може забезпечити достатню точність при мінімальних обчислювальних витратах. У ширших географічних масштабах доцільніше застосовувати метод гаверсинуса.

6. Результати досліджень

6.1. Використані набори даних

Для проведення експериментального дослідження було використано два набори даних: UjiIndoorLoc та SODIndoorLoc.

UjiIndoorLoc [20] – це відкритий набір даних для досліджень у сфері локалізації людей в приміщенні на основі Wi-Fi, зібраний у трьох корпусах Jaume I University (Іспанія) площею майже 110 000 м². У ньому міститься понад 19 000 сигналів від 520 різних точок доступу, отриманих більше ніж від двадцяти користувачів на 25 моделях Android-пристроїв, а також точні координати (широта, довгота), ідентифікатори будівлі, поверху та приміщення (SpaceID). Опис даних, з яких складається датасет, наведено у таблиці 1:

Таблиця 1. Опис набору даних UjiIndoorLoc

Назва	Опис даних	Формат даних
WAP001-WAP520	Значення інтенсивності сигналу (RSSI) для WAP (Wireless Access Point)	Дискретні, $-100 \dots 0$ dBm (100 = WAP не виявлено)
Longitude	Координати смартфона під час сканування, довгота	Неперервні
Latitude	Координати смартфона під час сканування, ширина	Неперервні
Floor	Номер поверху у межах будівлі	Дискретні, $0 \dots 4$
BuildingID	Ідентифікатор будівлі, вимірювання проводились у трьох різних будівлях	Дискретні, $0 \dots 2$

Продовження таблиці 1

SpaceID	Внутрішній ідентифікаційний номер для ідентифікації приміщення, в якому було здійснено сканування	Дискретні
UserID	Ідентифікатор користувача	Дискретні

Приклад записів, з яких складається набір даних, зображено в таблиці 2:

Таблиця 2. Приклад записів з набору даних UjiIndoorLoc

№	WAP001	...	WAP520	Longitude	Latitude	Floor	BuildingID	SpaceID
1	100	...	100	-7541.2643	4864921	2	1	106
2	100	...	100	-7536.6212	4864934	2	1	106
3	100	...	100	-7519.1524	4864950	2	1	103
4	100	...	100	-7524.5704	4864934	2	1	102
5	100	...	100	-7632.1436	4864982	0	0	122

SODIndoorLoc [21] – це відкритий набір даних для досліджень у сфері локалізації людей в приміщенні на основі Wi-Fi сигналів. Він охоплює три різні будівлі з багатьма поверхами на загальній площі близько 8000 м², та містить понад 1800 маркерних точок з середньою відстанню між ними менше 1,2 м. Опис даних, з яких складається датасет, наведено у таблиці 3:

Таблиця 3. Опис набору даних SODIndoorLoc

Назва	Опис даних	Формат даних
MAC1-MACN	Значення інтенсивності сигналу (RSSI) для N виявлених WAP (Wireless Access Point)	Дискретні, -100... 0 dBm (100 = WAP не виявлено)
ECoord	Локальна X-координата точки вимірювання на плані поверху	Неперервні, метри
NCoord	Локальна Y-координата точки вимірювання на плані поверху	Неперервні, метри
FloorID	Номер поверху у межах будівлі	Дискретні, цілі 1...4
BuildingID	Ідентифікатор будівлі (1 – CETC331; 2 – HСХУ; 3 – SYL)	Дискретні, цілі 1...3
SceneID	Тип простору, де виконано вимірювання (1 – коридор; 2 – офіс; 3 – конференц-зала)	Дискретні, цілі 1...3

Приклад записів, з яких складається набір даних, зображено в таблиці 4:

Таблиця 4. Приклад записів з набору даних SODIndoorLoc

№	MAC1	...	MAC52	ECoord	NCoord	FloorID	BuildingID	SceneID
1	-44	...	100	45	17.5	1	1	1
2	-37	...	100	45.5	17.5	1	1	1
3	-41	...	100	46	17.5	1	1	1

Продовження таблиці 4

4	-38	...	100	46.5	17.5	1	1	1
5	-35	...	100	46.9	17.5	1	1	1

Обидва набори даних UjiIndoorLoc та SODIndoorLoc зібрані на території Jaume I University. Основна їх різниця полягає у кількості і щільності даних. UjiIndoorLoc має більше даних SpaceID – 254 унікальні приміщення, тоді як у SODIndoorLoc їх лише 7. Незважаючи на те, що у SODIndoorLoc значно менше унікальних приміщень, кількість унікальних записів в ньому (24000) перевищує кількість унікальних записів в UjiIndoorLoc (20000). Приклади розподілу даних зображено на рис. 8-9:

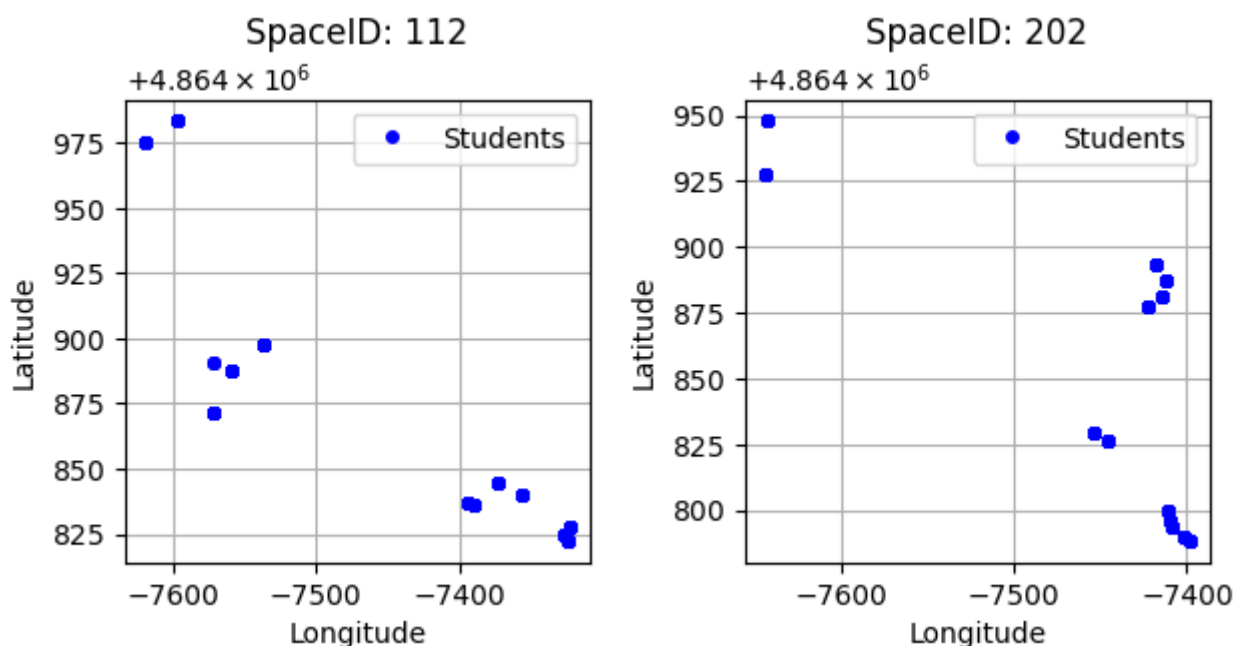


Рис. 8. Розподіл даних в приміщеннях з UjiIndoorLoc.

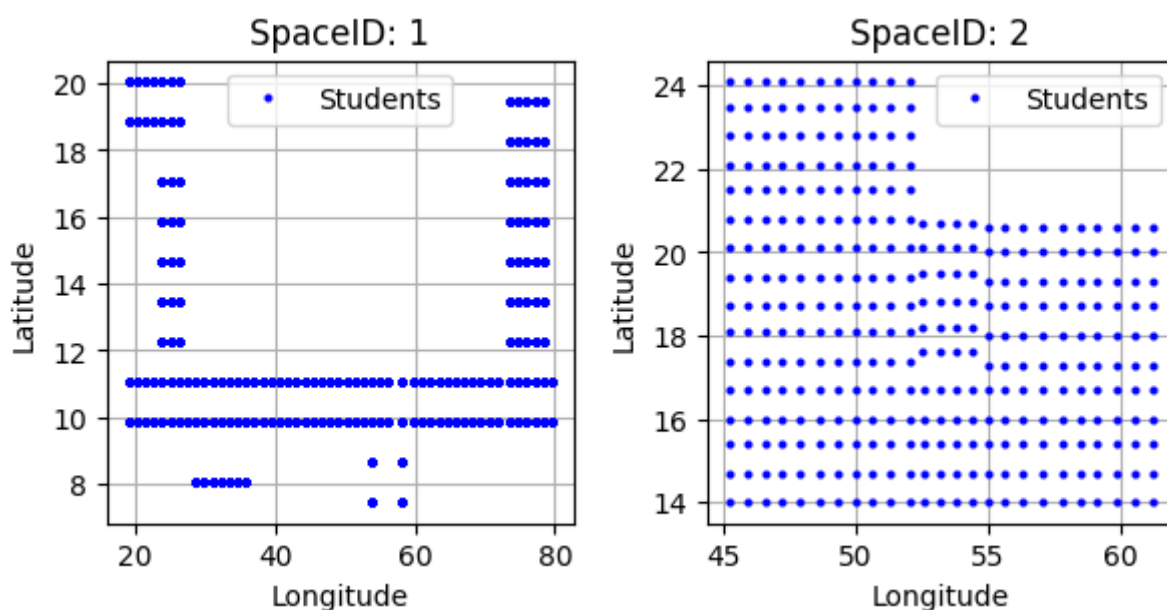


Рис. 9. Розподіл даних в приміщеннях з SODIndoorLoc.

Як видно з рис.8 та рис.9, UjiIndoorLoc має дані, які описують багато різних SpaceID, але з малою кількістю записів для кожного приміщення. І навпаки – SODIndoorLoc описує значно менше SpaceID, але з значно більшою кількістю записів для кожного приміщення. Окрім цього видно, що дані у UjiIndoorLoc хаотично розподілені, тоді як дані у SODIndoorLoc – рівномірно.

Ще одна різниця між двома наборами даних – формат, у якому вказано координати. У наборі даних UjiIndoorLoc координати подано у форматі широти та довготи, що є стандартним форматом для визначення місцезнаходження на земній кулі. У наборі даних SODIndoorLoc також використані довгота та широта, але замість географічних координат використовуються локальні координати, які вказують на точне місце всередині конкретної будівлі.

Різниця між двома наборами даних обумовили вибір саме цих наборів даних для експериментального дослідження. UjiIndoorLoc дозволить перевірити роботу алгоритмів в умовах хаотичних та недостатніх даних, а SODIndoorLoc в умовах упорядкованих та щільних даних.

6.2. Результати моделювання

До даних з наборів даних UjiIndoorLoc та SODIndoorLoc було застосовано метод визначення найменшої відстані від людини до центрів приміщень. Для визначення центру кожного приміщення було застосовано методи арифметичної центроїди (1) та центроїди полігональної області (3-5). Для визначення відстані між людиною та центрами кожного приміщення було використано методи евклідової відстані (6) та гаверсінуса (7-9). Для отримання полігональної області було використано метод опуклої оболонки (2). Результати роботи методів для UjiIndoorLoc та SODIndoorLoc (у трьох будівлях CETC331, HCXY та SYL) зображено на рис. 10-13:

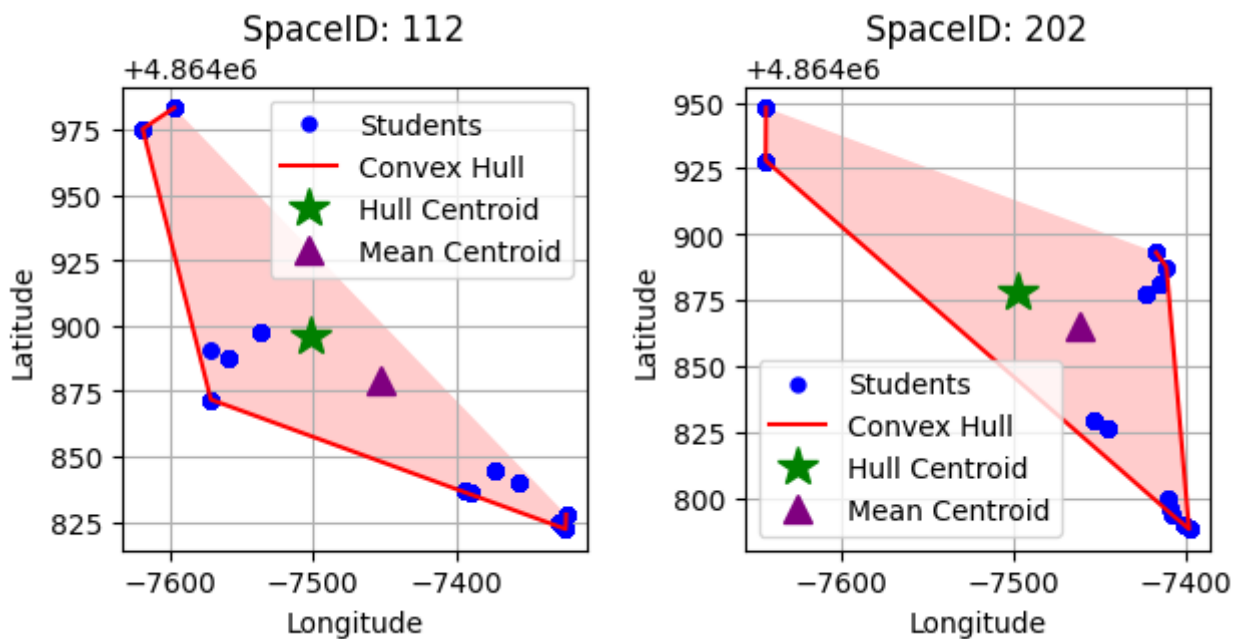


Рис. 10. Результат роботи методів арифметичної центроїди та центроїди полігональної області для UjiIndoorLoc.

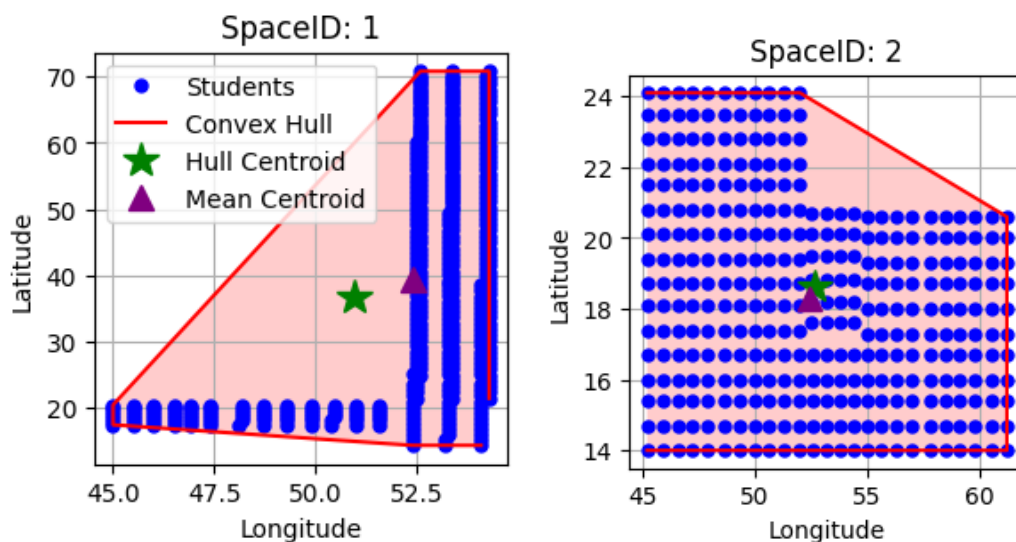


Рис. 11. Результат роботи методів арифметичної центроїди та центроїди полігональної області для SODIndoorLoc (CETC331).

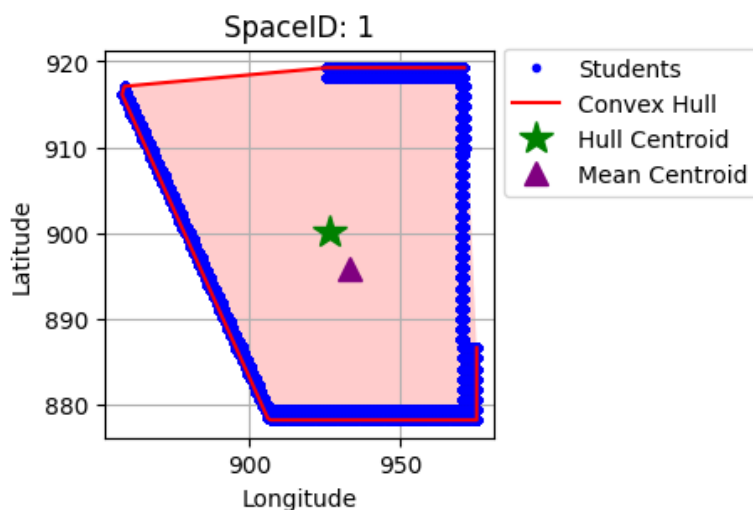


Рис. 12. Результат роботи методів арифметичної центроїди та центроїди полігональної області для SODIndoorLoc (HCXY).

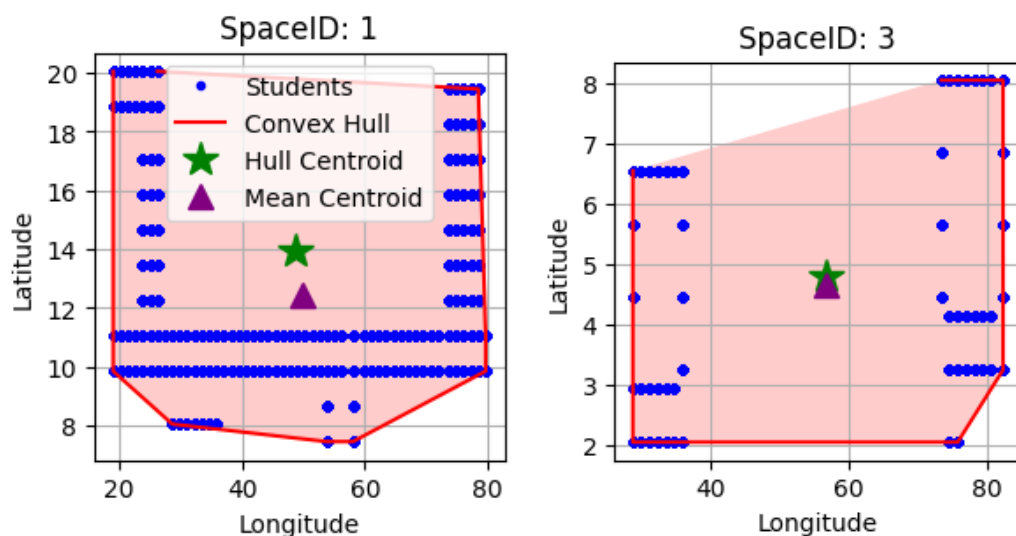


Рис. 13. Результат роботи методів арифметичної центроїди та центроїди полігональної області для SODIndoorLoc (SYL).

Як видно з рисунків 10-13, точність визначення центру приміщення залежить від розповсюдження даних. При рівномірному розповсюдженні (як по площині, так і по кількості) обидва методи визначають центр майже однаково. Але у випадках, коли дані розповсюджені нерівномірно, метод центроїди полігональної області краще визначає центр приміщення. Це пов'язано з тим, що метод арифметичної центроїди, як видно з формули (1), фактично є середнім арифметичним усіх точок на площині. При нерівномірному розподілі це призводить до того, що центроїда зміщується у напрямку до великих скупчень точок.

Далі на основі отриманих центроїд було розраховано відстані від людей до центрів приміщень. Для цього було використано методи евклідової відстані та гаверсинуса. Для кожної точки на площині розраховувалася відстань від цієї точки до центрів кожного з приміщень. Далі ці відстані порівнювалися для знаходження мінімальної відстані. Мінімальна відстань вказувала на номер приміщення, де знаходиться людина.

Для перевірки результатів було виконано порівняння з еталонним SpaceID, взятим з наборів даних, та SpaceID, визначений алгоритмом. Так як використаний алгоритм може мати лише True Positive та False Positive результати, для оцінки точності було використано метрику точності (accuracy) [21]:

$$\text{Accuracy} = \frac{TP}{TP + FP}, \quad (10)$$

де: TP – true positive, кількість правильних визначень місцезнаходження людини;

FP – false positive, кількість помилкових визначень місцезнаходження людини.

У SODIndoorLoc дані наведено окремими наборами для кожної будівлі (CETC331, HСХУ та SYL). Точність визначення місцезнаходження людини методом для обох наборів даних наведено у таб.5:

Таблиця 5. Точність визначення SpaceID методом найменшої відстані до їх центрів

Набір даних	Метод визначення центроїди	Метод визначення відстані	Ассурасу
UjiIndoorLoc	Арифметична центроїда	Евклідова відстань	2.43%
UjiIndoorLoc	Арифметична центроїда	Гаверсинус	2.56%
UjiIndoorLoc	Центроїд полігональної області	Евклідова відстань	2.51%
UjiIndoorLoc	Центроїд полігональної області	Гаверсинус	2.61%
SOD (CETC331)	Арифметична центроїда	Евклідова відстань	74.14%
SOD (CETC331)	Арифметична центроїда	Гаверсинус	72.98%
SOD (CETC331)	Центроїд полігональної області	Евклідова відстань	77.8%
SOD (CETC331)	Центроїд полігональної області	Гаверсинус	76.02%
SOD (HСХУ)	Арифметична центроїда	Евклідова відстань	100%
SOD (HСХУ)	Арифметична центроїда	Гаверсинус	100%
SOD (HСХУ)	Центроїд полігональної області	Евклідова відстань	100%
SOD (HСХУ)	Центроїд полігональної області	Гаверсинус	100%
SOD (SYL)	Арифметична центроїда	Евклідова відстань	42.91%
SOD (SYL)	Арифметична центроїда	Гаверсинус	42.91%
SOD (SYL)	Центроїд полігональної області	Евклідова відстань	42.23%
SOD (SYL)	Центроїд полігональної області	Гаверсинус	42.57%

Як видно з таблиці 5, точність визначення SpaceID сильно відрізняється для різних наборів даних. Було зроблено припущення, що причиною є приміщення, які знаходяться за однаковими координатами, але на різних поверххах. Даний алгоритм не враховує висоту координат, тому відбулася колізія даних, яка знизилася точність. 100% точність у наборі даних

SODIndoorLoc для будівлі HCXY обумовлена тим, що цей набір містить дані лише одного приміщення. У зв'язку з цим було вирішено не використовувати будівлю HCXY в подальших дослідженнях.

Для перевірки припущення впливу різних поверхів на точність прогнозування дані були вручну відфільтровані по будівлі і поверху, після чого методи були повторно використані. В наборі даних UjiIndoorLoc є дані з трьох різних будівель, в двох з яких є чотири різних поверхи, а в третій – п'ять поверхів. У наборі даних CETC331 є дані з трьох різних поверхи. У наборі даних SYL є дані лише з одного поверху, тому для нього фільтрація даних не змінить точність.

Для визначення SpaceID було використано методи центроїди полігональної області і метод гаверсинуса. Результати наведено у таб.6:

Таблиця 6. Точність визначення SpaceID з фільтрацією даних по поверхам

Набір даних	Будівля	Поверх	Accuracy
UjiIndoorLoc	0	0	100%
UjiIndoorLoc	0	1	100%
UjiIndoorLoc	0	2	98.6%
UjiIndoorLoc	0	3	100%
UjiIndoorLoc	1	0	91.3%
UjiIndoorLoc	1	1	98.6%
UjiIndoorLoc	1	2	94%
UjiIndoorLoc	1	3	87%
UjiIndoorLoc	2	0	95%
UjiIndoorLoc	2	1	89%
UjiIndoorLoc	2	2	81.27%
UjiIndoorLoc	2	3	92.07%
UjiIndoorLoc	2	4	99.09%
SOD (CETC331)	1	1	100%
SOD (CETC331)	1	2	73.82%
SOD (CETC331)	1	3	100%

Як видно з таблиць 5 і 6, точність значно підвищилася, в деяких випадках досягнувши 100%. Це підтверджує припущення, що для даного алгоритму номер поверху є критично важливою інформацією. Але при цьому більшість результатів все ще мають незадовільну точність. Було зроблено припущення, що така різниця між точностями свідчить про невиконання основної умови методу – однорідності приміщень у формах та розмірах.

Окрім методу визначення найменшої відстані від людини до центрів приміщень до наборів даних також було використано метод порівняння координат людини з площею приміщень. Площа визначалася методом опуклої оболонки, після чого координати людей порівнювалися з отриманою площею. Результати наведено у таб. 7:

Таблиця 7. Точність визначення SpaceID методом порівняння координат людини з площею приміщень

Набір даних	Accuracy
UjiIndoorLoc	15.49%
SOD (CETC331)	82.83%
SOD (SYL)	100%

Як видно з таблиці 7, точність для будівлі SYL досягла 100%, на відміну від попереднього результату 42.91%, що вже підтверджує припущення впливу форм та розміру сусідніх приміщень на точність, адже вони ніяк не впливають на роботу даного методу. Щодо UjiIndoorLoc та CETC331, вже відомо, що в них відбувається колізія даних для різних

поверхів. Для перевірки дані також було відфільтровано по поверхам. Результати наведено у таб. 8:

Таблиця 8. Точність визначення SpaceID методом порівняння координат людини з площею приміщень з фільтрацією даних по поверху

Набір даних	Будівля	Поверх	Accuracy
UjiIndoorLoc	0	0	0%
UjiIndoorLoc	0	1	0%
UjiIndoorLoc	0	2	0%
UjiIndoorLoc	0	3	0%
UjiIndoorLoc	1	0	8.04%
UjiIndoorLoc	1	1	22.17%
UjiIndoorLoc	1	2	23.50%
UjiIndoorLoc	1	3	34.28%
UjiIndoorLoc	2	0	10.2%
UjiIndoorLoc	2	1	3.42%
UjiIndoorLoc	2	2	0%
UjiIndoorLoc	2	3	0%
UjiIndoorLoc	2	4	0%
SOD (CETC331)	1	1	100%
SOD (CETC331)	1	2	100%
SOD (CETC331)	1	3	100%

Як видно з таблиці 8, точність для CETC331 досягла 100%, але точність для UjiIndoorLoc критично низька. Після аналізу набору даних було виявлено причину – недостатня кількість та якість даних. Багато даних з UjiIndoorLoc містять лише декілька записів для кожного приміщення. Для роботи методу опуклої оболонки необхідно мати мінімум 3 різних точки координат. Окрім цього, ці точки повинні бути достатньо розосереджені по площі приміщення. Для багатьох приміщень ці умови не виконувалися, тому метод опуклої оболонки не міг достатньо повно описати площу приміщення. Для прикладу на рис. 14 наведено порівняння даних з набору даних UjiIndoorLoc та CETC331:

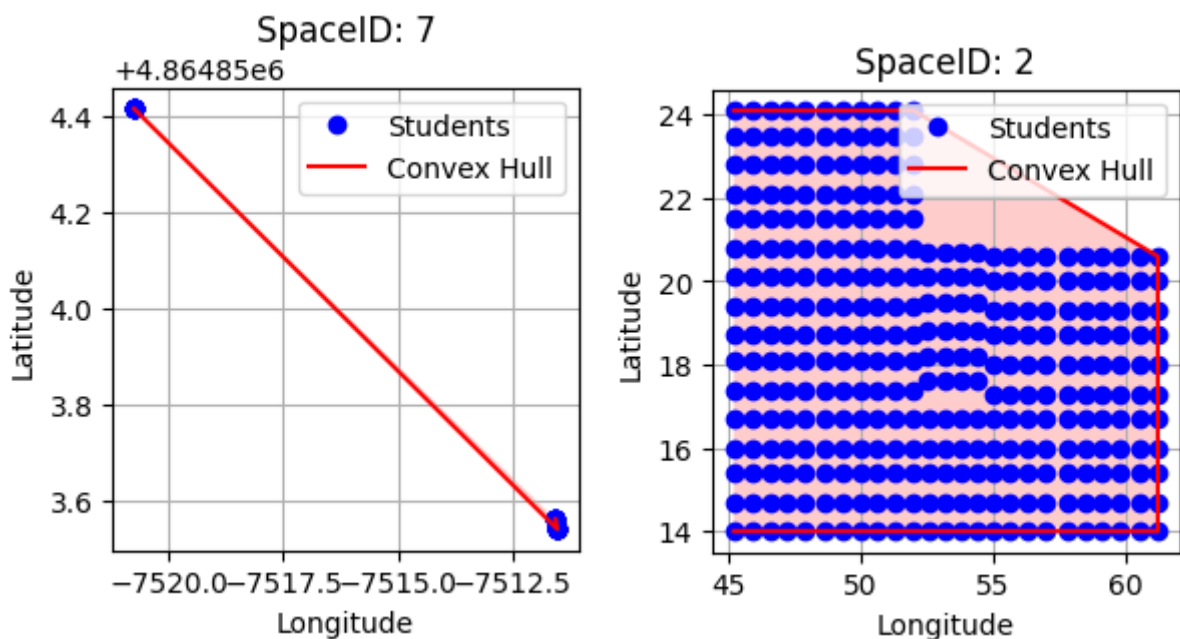


Рис. 14. Порівняння даних з набору даних UjiIndoorLoc та SODIndoorLoc (CETC331).

Як видно з рис. 14, дані суттєво відрізняються. З розташуванням точок, як у UjiIndoorLoc, неможливо побудувати опуклу оболонку, яка буде адекватно описувати площу приміщення. Точність визначення місцезнаходження для UjiIndoorLoc показує, що основною умовою для роботи методу порівняння координат людини з площею приміщень є достатня кількість і якість даних. Наприклад, якщо точок координат достатньо для побудови опуклої оболонки, але їх розповсюдження неповно описує форму приміщення, то точність визначення приміщення буде нестабільною. Причиною є висока ймовірність того, що нові дані (нові записи координат студентів) будуть в зовсім іншій частині приміщення і не будуть описані минулою опуклою оболонкою. І навпаки – чим більше отримано статистичних даних, тим менша ймовірність, що нові дані будуть поза межами опуклої оболонки.

Повертаючись до проблеми впливу номеру поверху на точність, можливим вирішенням проблеми буде використання машинного навчання для прогнозування номеру будівлі та поверху. Метод класифікації RandomForestClassifier дозволить прогнозувати будівлю з точністю до 99.7% і поверх з точністю до 99.6% [23]. Так як в такому випадку буде робитися три послідовних прогнозування, загальна точність розраховується як добуток трьох точностей:

$$\text{Accuracy} = \text{Accuracy}_B \cdot \text{Accuracy}_F \cdot \text{Accuracy}_S, \quad (11)$$

де: Accuracy_B – точність прогнозування будівлі машинним навчанням;

Accuracy_F – точність прогнозування поверху машинним навчанням;

Accuracy_S – точність визначення приміщень методом найменшої відстані до їх центрів або методом порівняння координат людини з площею приміщень.

Найкращі отримані точності з урахуванням фільтрації даних по поверхах та точністю прогнозування цих даних наведено у таб. 9 (точність для даних з різними поверхами було усереднено):

Таблиця 9. Найкращі отримані точності визначення SpaceID

Набір даних	Метод визначення	Accuracy_B	Accuracy_F	Accuracy_S	Accuracy
UjiIndoorLoc	Найменша відстань до центрів приміщень	99.7%	99.6%	94.3%	93.64%
SOD (CETC331)	Порівняння координат з опуклою оболонкою	99.7%	99.6%	100%	99.3%
SOD (HCXY)	Порівняння координат з опуклою оболонкою	99.7%	99.6%	100%	99.3%
SOD (SYL)	Порівняння координат з опуклою оболонкою	99.7%	99.6%	100%	99.3%

Результати, наведені у таб.9 показують, що детерміновані методи визначення місцезнаходження людей в приміщенні потенційно спроможні досягати майже ідеальної точності. Через те, що дані з UjiIndoorLoc не відповідають необхідним умовам для застосування методів порівняння координат людини з площею приміщень та визначення найменшої відстані до їх центрів, використаними методами не вдалося досягнути задовільної точності. Але низька точність лише вказує на важливість дотримання основних умов при використанні даних методів. Точність з набору даних SODIndoorLoc показує, що при

дотриманні основних умов дані методи можуть визначати місцезнаходження люди з достатньо високою точністю.

7. Перспективи подальшого розвитку досліджень

Подальший розвиток досліджень з даної теми може продовжуватися шляхом підвищення точності прогнозування будівлі та поверху, де знаходиться людина, машинним навчанням. Іншим перспективним напрямком є знаходження альтернативного методу для отримання даної інформації.

8. Висновки

В даному дослідженні було розглянуто задачу визначення місцезнаходження людей у приміщенні. Для виконання цієї задачі було використано детерміновані методи з метою отримання високої точності визначення приміщення.

Основними методами для визначення місцезнаходження людей у приміщенні були визначення найменшої відстані до центрів приміщень та порівняння координат людини з площею приміщень. Для роботи обох методів було використано координати людей у просторі (як географічні, так і локальні). Було встановлено, що у кожного з методів є свої умови для використання, які обмежують їх область застосування.

Основною умовою для роботи методу визначення найменшої відстані до центрів приміщень є однорідність сусідніх приміщень по формі та розміру. Якщо форми та розмір сусідніх приміщень суттєво відрізняються між собою, то перебуваючи біля спільних стін, людина може бути ближче до центру сусіднього приміщення, ніж до центру того де вона знаходиться. Дослідження показало, що даний фактор суттєво впливає на точність визначення приміщення.

Основною умовою для роботи методу порівняння координат людини з площею приміщень є достатня кількість та якість даних. Для побудови полігональної площі, що описує всю множину даних для певного приміщення, необхідно мати мінімум три різні точки координат. Точність визначення приміщення даним методом прямим чином залежить від наявних даних – чим більш повно вони описують площу приміщення (наприклад, побудована площа утворила багатокутник, периметр якого майже збігається зі стінами приміщення), тим більша буде точність. І навпаки – чим гірше дані описують приміщення (наприклад, побудована площа утворила трикутник, площа якого описує менше половини площі приміщення), тим менша буде точність. Найкращі дані для даного методу можливо отримати мануальним вимірюванням площі приміщення, для чого достатнім буде виміряти чотири точки у кожному з кутів приміщення.

Основною умовою для роботи обох методів є їх використання для одного поверху будівлі. Обидва методи використовують лише координати довготи та широти і не враховують висоту координат. Наявність приміщень з однаковими координатами, але різними поверхами, значно зменшує точність визначення приміщення, адже це приводить до колізії даних. У зв'язку з цим критично важливо фільтрувати дані по поверху, що створює нову задачу – визначення поверху приміщення де знаходиться людина. Для вирішення цієї задачі оптимально використовувати методи машинного навчання, які в якості даних використовують значення інтенсивності сигналу (RSSI) для WAP (Wireless Access Point).

Окрім цього, в даному дослідженні було розглянуто вплив різних математичних методів на точність для методу визначення найменшої відстані до центрів приміщень. Для цього методу необхідно виконати дві задачі – отримати координати центрів кожного приміщення та отримати відстані між координатами людини та центрами приміщень.

Для отримання координат центрів приміщень було використано методи арифметичної центроїди та центроїди полігональної області. Дослідження показало, що метод центроїди

полігональної області забезпечує більшу точність, ніж метод арифметичної центроїди. Це обумовлено тим, що на метод арифметичної центроїди впливає розповсюдження даних по площині – якщо дані розповсюджено нерівномірно, наприклад, сконцентровано біля кута приміщення, то арифметична центроїда зсувається до цього кута. Метод центроїди полігональної області враховує лише вершини отриманого багатокутника, а не всі точки з множини, тому він виявився стійким до нерівномірно розповсюджених даних.

Для отримання відстані між координатами людини та центрами приміщень було використано методи евклідової відстані та гаверсінуса. Дослідження показало, що їх вплив на точність методу визначення найменшої відстані до центрів приміщень залежить від формату даних, в яких представлено координати людей. У наборі даних UjiIndoorLoc координати були представлені у глобальній системі координат (WGS 84 / EPSG:3857), а у наборі даних SODIndoorLoc – у локальній системі координат із початком відліку в точці (0,0) на території Jaume I University (Іспанія). Було виявлено, що для локальної системи координат краща точність отримується методом евклідової відстані, а для географічної системи координат – методом гаверсінуса.

Таким чином, дане дослідження показало, що вибір детермінованих методів для визначення місцезнаходження об'єктів залежить від використовуваних даних та умов експлуатації. Як метод визначення найменшої відстані до центрів приміщень, так і метод порівняння координат людини з площею приміщень здатні визначати місцезнаходження об'єктів у приміщенні з достатньо високою точністю. Але для досягнення цього необхідно, щоб були виконані певні умови – визначення поверху об'єкту для обох методів, однорідність форм та розмірів сусідніх приміщень для методу визначення найменшої відстані до їх центрів і достатня кількість та якість даних для методу порівняння координат людини з площею приміщень.

Список літератури:

- 1) Choi, M., Park, J.-H., & Yi, G. (2015). Attendance check system and implementation for Wi-Fi networks supporting unlimited number of concurrent connections. *International Journal of Distributed Sensor Networks*, 11(7), 508698. <https://doi.org/10.1155/2015/508698>
- 2) Panca, B. S., Paulus, Y., & Karnalim, O. (2020). An automated attendance system in academia with implicit authentication via mobile devices and Wi-Fi connection. *International Journal of Online and Biomedical Engineering (iJOE)*, 16(07), 86. <https://doi.org/10.3991/ijoe.v16i07.13869>
- 3) Azim, N. A. (2020). Bluetooth smart based automated attendance system. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(3), 4137–4140. <https://doi.org/10.30534/ijatcse/2020/242932020>
- 4) Puckdeevongs, A., Tripathi, N. K., Witayangkurn, A., & Saengudomlert, P. (2020). Classroom attendance systems based on bluetooth low energy indoor positioning technology for smart campus. *Information*, 11(6), 329. <https://doi.org/10.3390/info11060329>
- 5) Farag, W. A., & Abouelela, M. (2022). Low-Cost active monitoring of attendance using passive RFID technology. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, 8(4), 552. <https://doi.org/10.26555/jiteki.v8i4.25168>
- 6) Lin, J., & Wen, S. (2022). Physical education attendance based on wireless and passive RFID technology. *Journal of Sensors*, 2022, 1–12. <https://doi.org/10.1155/2022/8485428>
- 7) Manori, A., Devnath, N., Pasi, N., & Kumar, V. (2017). QR code based smart attendance system. *International Journal of Smart Business and Technology*, 5(1), 1–10. <https://doi.org/10.21742/ijstb.2017.5.1.01>
- 8) Masalha, F., & Hirzallah, N. (2014). A students attendance system using QR code. *International Journal of Advanced Computer Science and Applications*, 5(3). <https://doi.org/10.14569/ijacsa.2014.050310>

- 9) Sunaryono, D., Siswantoro, J., & Anggoro, R. (2019). An android based course attendance system using face recognition. *Journal of King Saud University - Computer and Information Sciences*. <https://doi.org/10.1016/j.jksuci.2019.01.006>
- 10) Singh, D. N., Sri, M. K., & Mounika, K. (2019). IOT based automated attendance with face recognition system. *International Journal of Innovative Technology and Exploring Engineering*, 8(6S4). <https://doi.org/10.35940/ijitee.f1093.0486s419>
- 11) Basila, A., & Danladi, A. (2021). Design, simulate and construct a fingerprints attendance system with data logging. *Nigerian Journal of Technology*, 40(4), 703–712. <https://doi.org/10.4314/njt.v40i4.17>
- 12) Alhothaily, M., Alradaey, M., Oqbah, M., & El-Kustaban, A. (2015). Fingerprint attendance system for educational institutes. *Journal of Science and Technology*, 20(1), 34–44. <https://doi.org/10.20428/jst.v20i1.842>
- 13) Singh, N., Choe, S., & Punmiya, R. (2021). Machine learning based indoor localization using wi-fi RSSI fingerprints: An overview. *IEEE Access*, 1. <https://doi.org/10.1109/access.2021.3111083>
- 14) Yoo, J.-H. (2019). Study on Prediction of Attendance Using Machine Learning. *Journal of IKEEE*, 23(4), 1243–1249. <https://doi.org/10.7471/IKEEE.2019.23.4.1243>
- 15) Mammana, M. F., Pennisi, M., & Micale, B. (2008). On the centroids of polygons and polyhedra. *Forum Geometricorum*, 8, 121–130.
- 16) Weisstein, E. (2006). Convex hull -- from Wolfram MathWorld. A Wolfram Web Resource. <https://mathworld.wolfram.com/ConvexHull.html>
- 17) Weisstein, E. (2022). Polygon Centroid -- from Wolfram MathWorld. A Wolfram Web Resource. <https://mathworld.wolfram.com/PolygonCentroid.html>
- 18) Liberti, L., Lavor, C., Maculan, N., & Mucherino, A. (2014). Euclidean distance geometry and applications. *SIAM Review*, 56(1), 3–69. <https://doi.org/10.1137/120875909>
- 19) Prasetya, D. A., Phong Thanh Nguyen, Faizullin, R., & Iswanto, I. (2020). Resolving the shortest path problem using the haversine algorithm. *Journal of Critical Reviews*, 7(1), 62–64.
- 20) Torres-Sospedra, J., Montoliu, R., Martínez-Usó, A., Arnau, T. J., Avariento, J. P., Bénédicto-Bordonau, M., & Huerta, J. (2014). UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems. In *Proceedings of the Fifth International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE. <https://doi.org/10.24432/C5MS59>
- 21) Bi, J., Wang, Y., Yu, B., Cao, H., Shi, T., & Huang, L. (2022). Supplementary open dataset for WiFi indoor localization based on received signal strength. *Satellite Navigation*, 3(1). <https://doi.org/10.1186/s43020-022-00086-y>
- 22) Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond accuracy, f-score and ROC: A family of discriminant measures for performance evaluation. *Lecture notes in computer science* (c. 1015–1021). Springer Berlin Heidelberg. https://doi.org/10.1007/11941439_114
- 23) Топольський, А. І., & Паламарчук, Є. А. (2025). Використання машинного навчання для визначення місцезнаходження людей у приміщеннях. *Вісник Вінницького політехнічного інституту*, (1), 92–103. <https://doi.org/10.31649/1997-9266-2025-178-1-92-103>

Application of deterministic methods for indoor people localization

Andriy Topolskiy

Department of Automation and Intelligent Information Technologies, Vinnytsia National Technical University (VNTU), Vinnytsia, Ukraine
ORCID 0009-0007-7631-0872

Yevhen Palamarchuk

Department of Automation and Intelligent Information Technologies, Vinnytsia National Technical University (VNTU), Vinnytsia, Ukraine
ORCID 0000-0002-7443-099X

Abstract: This paper investigates the application of deterministic methods for indoor people localization, namely, the method of determining the smallest distance from a person to the centers of the rooms and the method of comparing the coordinates of a person with the area of the rooms. The necessary conditions for the operation of each method were analyzed. It is shown that the application of the method of the smallest distance is optimal when adjacent rooms have homogeneous shape and size, because with heterogeneous rooms, a person near a shared wall may be closer to an adjacent room's center. For the method of comparing the coordinates of a person with the area of the rooms, the necessary conditions are a sufficient number (at least three coordinate points) and quality (adequate distribution and diversity of input points) of input data, because if they are not sufficient, the accuracy of determining the person's location will decrease. For both cases, it is critical to have information about the floor where the person is located, since geodetic coordinates (latitude/longitude) omit elevation, causing ambiguity in multistory buildings. A practical solution is to apply machine-learning models trained on RSSI measurements from WAPs. In addition to the necessary conditions of deterministic methods, the paper also investigates the influence of the methods used for data processing. The arithmetic centroid and polygonal area centroid methods were compared to obtain the coordinates of the room centers. It is shown that the arithmetic centroid shifts to large data clusters, while the polygonal centroid is more resistant to uneven data distribution. The Euclidean distance and the haversine methods were compared to obtain the distance between two points. It is shown that the influence of these methods on the accuracy of the method of the least distance to the centers depends on the coordinate system in which the data are represented. For the global coordinate system (WGS 84/EPSG:3857), the best result was shown by the haversine method, and for the local coordinate system - by the Euclidean distance method. It is shown that, subject to the specified conditions, both methods provide a sufficiently high accuracy of localization of people in the room.

Keywords: automated attendance systems, e-learning systems, indoor people localization, deterministic location methods, convex hull method, distance determination methods.
