

ВІЗУАЛЬНО-ІНЕРЦІЙНИЙ SLAM З ВИКОРИСТАННЯМ РОЗШИРЕНОГО ФІЛЬТРА КАЛМАНА ДЛЯ АВТОНОМНОЇ НАВІГАЦІЇ

¹Вінницький національний технічний університет

Досліджено використання візуально-інерційного SLAM на основі розширеного фільтра Калмана (EKF) для задач автономної навігації. У дослідженні використано дані з KITTI-датасету, що включають зображення зі стереокамери, а також лінійні та кутові швидкості, отримані з інерційно-вимірювального пристрою. Для визначення візуальних орієнтирів запропоновано використовувати ознаки ORB, які характеризуються швидкістю обчислення та стійкістю до змін освітлення, обертання і масштабування. До того ж розроблено алгоритм вибору релевантних орієнтирів, що дозволяє підвищити точність та швидкість роботи SLAM.

Результати серії експериментів продемонстрували, що ефективність роботи системи значною мірою залежить від налаштування співвідношення шумів моделі руху (Q) та сенсорних вимірювань (R). Визначено, що оптимальний діапазон Q/R для досліджуваних наборів даних становить $0,001 \dots 0,00001$.

Для досягнення балансу між швидкістю та якістю роботи системи досліджено вплив кількості релевантних візуальних ознак у межах від 10 до 120. За результатами експериментів встановлено, що оптимальна кількість орієнтирів становить $50 \dots 80$, а найкращі результати досягнуто коли $Q/R = 0,001$ або $Q/R = 0,00001$ залежно від використаного набору даних. Оцінка точності траєкторії здійснювалася за допомогою метрики ATE (абсолютна похибка траєкторії). Для перевірки використовувалися GPS-дані як ground truth.

Результати підтвердили, що візуально-інерційний SLAM є ефективним інструментом для автономної навігації, особливо в умовах відсутності GPS. Використання EKF у візуально-інерційному SLAM зменшує обчислювальні витрати порівняно з іншими методами, через що він є оптимальним для пристроїв з обмеженими ресурсами, зокрема для мобільних роботів.

Далі планується дослідити оптимізацію вибору орієнтирів, вдосконалення моделювання шумів та адаптацію алгоритму для монокулярних камер з метою підвищення точності та ефективності SLAM. До того ж планується дослідження інваріантного розширеного фільтра Калмана (IEKF) для підвищення точності картографування та стійкості системи у складних і динамічних умовах.

Ключові слова: візуально-інерційний SLAM, розширений фільтр Калмана, інерційний вимірювальний пристрій, набір даних KITTI.

Вступ

Одночасна локалізація та картографування (SLAM) є технологією, яка дозволяє роботам створювати карти невідомих середовищ, використовуючи дані з різних датчиків, що постійно збирають інформацію про оточення. Основною метою SLAM є побудова точної карти. Це потребує ретельної обробки даних. У процесі роботи робот може зіткнутися з перешкодами, а сенсори — збирати зашумлені або неповні дані, які потребують фільтрації та корекції. Саме в таких ситуаціях візуально-інерційний SLAM демонструє свою ефективність, поєднуючи переваги камер та інерційних датчиків для створення стабільних та достовірних карт навіть у складних умовах [1]. Крім того, візуально-інерційний SLAM може виступити ефективною альтернативою GPS у середовищах, де його застосування є неможливим [2].

В умовах війни в Україні автономні роботи, оснащені технологією візуально-інерційного SLAM, є перспективними для використання у небезпечних середовищах. Завдяки доступності

обладнання та здатності адаптуватися до різних умов середовища ці роботи можуть виконувати завдання з розвідки, пошуку постраждалих, доставки медикаментів і перевірки безпеки територій. Їх застосування може підвищити ефективність рятувальних операцій та зменшити ризики для людей у зоні бойових дій.

Візуально-інерційна навігація є однією з ключових технологій для створення точних систем локалізації та побудови карт у реальному часі. Вона поєднує дані з камер та інерційних сенсорів, що дозволяє компенсувати обмеження кожного з цих пристроїв: камери забезпечують високу роздільну здатність у текстурованих та статичних середовищах, проте вразливі до розмиття, оклюзій і змін освітлення, тоді як інерційні сенсори гарантують високу частоту оновлення, стабільність роботи та забезпечують масштаб руху, хоча й схильні до накопичення помилок. Інтеграція цих даних значно підвищує точність і стабільність системи, особливо у монокулярних системах, які без інерційних даних зазнають проблем з масштабом і дрейфом. Завдяки цій взаємодоповнюваності візуально-інерційні системи є перспективними для використання у складних середовищах, зокрема в галузях автономного водіння та робототехніки [1], [2].

Підходи SLAM, основані на використанні розширеного фільтра Калмана (EKF), продемонстрували високу ефективність, проте мають низку обмежень, пов'язаних з накопичуваними похибками. Основною проблемою є дрейф даних інерційного вимірювального пристрою (ІВП) та похибки, спричинені некоректною лінеаризацією, що призводить до поступової втрати узгодженості моделі з реальним середовищем у довготривалих або масштабних сценаріях. Попри те, що EKF дозволяє виконувати замикання петель та корекцію глобальних помилок, його обмеження у роботі з необоротними станами залишаються значною перешкодою для точного відстеження траєкторії у складних умовах [3], [4].

Для подолання цих обмежень розроблено удосконалені методи, такі як Unscented Kalman Filter (UKF) і Observability-Constrained Extended Kalman Filter (OC-EKF). Зокрема, OC-EKF враховує необоротні напрями руху, що дозволяє зберігати узгодженість навіть під час масштабних обчислень. Проте складність реалізації та чутливість до початкових умов обмежують їхнє застосування в умовах високої динаміки середовища [5]—[7].

Значний прорив у галузі забезпечив підхід Invariant EKF (IEKF), який базується на використанні симетрій систем для врахування необоротності станів і корекції дрейфу. IEKF-SLAM демонструє здатність точно оцінювати положення та орієнтацію навіть у складних сценаріях. Використання геометричної структури задачі дозволяє IEKF досягати високої стабільності й точності, що підтверджено як симуляціями, так і натурними експериментами. Алгоритм виявив свою ефективність у довготривалих місіях з високим рівнем шуму [6], [8].

Для детальнішого аналізу проведено порівняння EKF та IEKF за ключовими критеріями (табл. 1).

Таблиця 1

Порівняння методів SLAM

Критерій	EKF SLAM	IEKF SLAM
Складність реалізації	простий у реалізації	складніший, потребує більше обчислень
Точність в складних умовах	менш точний за значних нелінійностей	точніший за значних нелінійностей
Обчислювальні ресурси	економить ресурси, швидший	потребує більше ресурсів, повільніший
Тип середовища	придатний для простих середовищ	придатний для складних, шумних середовищ
Оптимальні умови застосування	простий рух, малі карти з помірними помилками	складні рухи, великі карти, високий рівень шуму

EKF SLAM є оптимальним для простих середовищ та невеликих карт, де нелінійність не має значного впливу. Через його низьку обчислювальну складність метод придатний для вбудованих пристроїв, зокрема мобільних роботів з обмеженими ресурсами. Натомість IEKF SLAM перевершує EKF у складних і шумних середовищах, а також під час роботи з великими картами, завдяки мінімізації помилок лінеаризації. Таким чином, вибір між EKF та IEKF залежить від складності середовища, рівня шуму, вимог до точності та доступних обчислювальних ресурсів.

Метою роботи є дослідження застосування візуально-інерційного SLAM з використанням розширеного фільтра Калмана для забезпечення автономної навігації.

Математичний опис підходу

Розглянемо основні елементи алгоритму візуально-інерційного SLAM такі як: стан системи, коваріаційна матриця, модель руху, модель спостереження та ініціалізація візуальних орієнтирів.

Стан системи

Загальний стан системи визначається так:

$$X_t = \begin{bmatrix} X_t^{robot} \\ X_t^{landmarks} \end{bmatrix}, \quad (1)$$

де X_t^{robot} — стан роботи, який складається з x, y, z — координати роботи в глобальній системі координат; ϕ, θ, ψ — кути обертання (roll, pitch, yaw), що описують орієнтацію робота.

$X_t^{landmarks}$ — координати орієнтирів; $l_i = [x_i, y_i, z_i]$ — положення i -го орієнтира в світовій системі координат.

Коваріаційна матриця P_t

Коваріаційна матриця відображає невизначеність стану системи та складається з блоків, що описують взаємодію між роботом і орієнтирами. Вона має вигляд

$$P_t = \begin{bmatrix} P_{robot} & P_{robot,landmarks} \\ P_{landmarks,robot} & P_{landmarks} \end{bmatrix}, \quad (2)$$

де P_{robot} — це матриця 6×6 , яка описує невизначеність стану робота (лінійні та кутові швидкості, положення); $P_{robot,landmarks}$ — взаємна коваріація між роботом і орієнтирами; $P_{landmarks}$ — матриця $3M \times 3M$, яка описує невизначеність координат M орієнтирів.

Модуль руху (етап передбачення)

Стан робота оновлюється за даними ІВП згідно з рівнянням

$$X_{t+1}^{robot} = f(X_t^{robot}, u_t, \Delta t), \quad (3)$$

де $u_t = [v_x, v_y, v_z, w_x, w_y, w_z]$; v_x, v_y, v_z — лінійні швидкості; w_x, w_y, w_z — кутові швидкості; Δt — часовий інтервал.

Оновлення коваріаційної матриці стану робота виконується за формулою

$$P_{t+1}^{robot} = F_t P_t^{robot} F_t^T + Q_t, \quad (4)$$

де F_t — Якоб'ян моделі руху, що визначає вплив зміни стану на наступний стан, Q_t — шум моделі руху (похибки ІВП) розмірністю 6×6 .

Модель спостереження (етап оновлення)

Спостереження орієнтирів через камеру моделюється як:

$$z_t = h(X_t^{robot}, X_t^{landmarks}), \quad (5)$$

де z_t — координати спостережень у системі камери (піксельні координати), $h(\cdot)$ — функція, яка виконує проєкцію 3D-точок у 2D-простір

$$h(l_i) = K \frac{R_c (X_i - X_{robot}) + t_c}{z}, \quad (6)$$

де K — калібрувальна матриця камери, R_c, t_c — обертання та зсув камери щодо робота, z — глибина точки.

Коефіцієнт посилення Калмана обчислюється за формулою

$$K_t = P_t H_t^T (H_t P_t H_t^T + R_t)^{-1}, \quad (7)$$

де H_t — Якоб'ян функції спостереження, R_t — шум вимірювань камери, розмірність $4N \times 4N$, де N — кількість спостережень.

Оновлення стану системи виконується за формулою

$$X_t = X_t + K_t (z_t - h(X_t)). \quad (8)$$

Оновлення коваріаційної матриці виконується так:

$$P_t = (I - K_t H_t) P_t. \quad (9)$$

Ініціалізація нових орієнтирів

Коли спостерігаються нові орієнтири, їхні координати визначаються за допомогою стереотріангуляції. Глибина орієнтира Z визначається за формулою

$$Z = \frac{f \cdot b}{z_t^{\text{left}} - z_t^{\text{right}}}, \quad (10)$$

де f — фокусна відстань камери, b — базова відстань між камерами (стереобазис), $z_t^{\text{left}}, z_t^{\text{right}}$ — піксельні координати орієнтира, отримані від лівої та правої камер.

Координати візуального орієнтира у світовій системі визначаються так:

$$l_i = T_{\text{world}} \cdot K^{-1} \cdot [u_L, v_L, 1] \cdot Z, \quad (11)$$

де T_{world} — матриця трансформації від системи координат камери до світової системи координат, u_L, v_L — координати орієнтира у лівій камері, K — калібрувальна матриця камери.

Формування наборів даних для досліджень

Для досліджень вибрано датасет KITTI, який широко використовується в задачах комп'ютерного зору, зокрема у візуальній одометрії та автономній навігації [9]. Датасет містить великий набір реальних даних, зібраних у різних умовах освітлення та складності траєкторій. Дані отримані з автомобіля, оснащеного сучасними сенсорами, серед яких: GPS/ІВІ-система, Velodyne LiDAR, камери перспективи та стереокамери «риб'яче око». KITTI включає стереопари зображень, кутові та лінійні швидкості, а також високоточні позиційні дані, отримані з GPS/ІВІ. Однією з ключових особливостей цього датасету є синхронізація всіх сенсорних даних, що дозволяє ефективно поєднувати інформацію для задач SLAM та інших алгоритмів автономної навігації [10].

У роботі використано два набори даних з KITTI-датасету: 2011_09_26_drive_0022 та 2011_09_26_drive_0027, які далі позначаються як kitti_22 і kitti_27 відповідно.

Крім даних, наданих у датасеті KITTI, для реалізації візуально-інерційного SLAM необхідно отримати візуальні ознаки з послідовностей зображень. З цією метою для кожного зображення з обох камер застосовано дескриптор ORB (Oriented FAST and Rotated BRIEF) [11].

ORB дозволяє виявляти ефективні та стійкі візуальні ознаки, забезпечуючи інваріантність до масштабування, обертання та змін освітлення. Завдяки низьким обчислювальним витратам і швидкості роботи ORB є оптимальним для реальних застосувань у SLAM [2], [11]. Для кожного зображення використовуються лише координати x, y розташування ознак, які в цій роботі слугують візуальними орієнтирами для обчислень у задачі SLAM.

Для реалізації візуально-інерційного SLAM будуть застосовані такі дані з кожного набору:

- калібрувальна матриця камери: використовується для точного проектування тривимірних точок у двовимірний простір зображення.
- базова відстань між двома камерами (стереобазис): необхідна для розрахунку глибини шляхом використання різниці положень об'єктів на зображеннях з лівої та правої камер.
- стереопари зображень: послідовності кадрів з лівої та правої камер, синхронізовані за часом.
- лінійна швидкість: значення, отримані з ІВІ для кожної часової мітки, які описують поступальний рух.
- кутова швидкість: дані ІВІ для кожної часової мітки, які характеризують обертання.
- матриця трансформації між ІВІ та камерою: використовується для коректної інтеграції даних від різних сенсорів у єдину координатну систему.
- часові мітки для кожного зображення: забезпечують узгодженість між даними від різних сенсорів у часі.

Приклад зміщення візуальних орієнтирів між двома сусідніми зображеннями у відео послідовності набору даних kitti_27 для лівої камери подано на рис. 1.

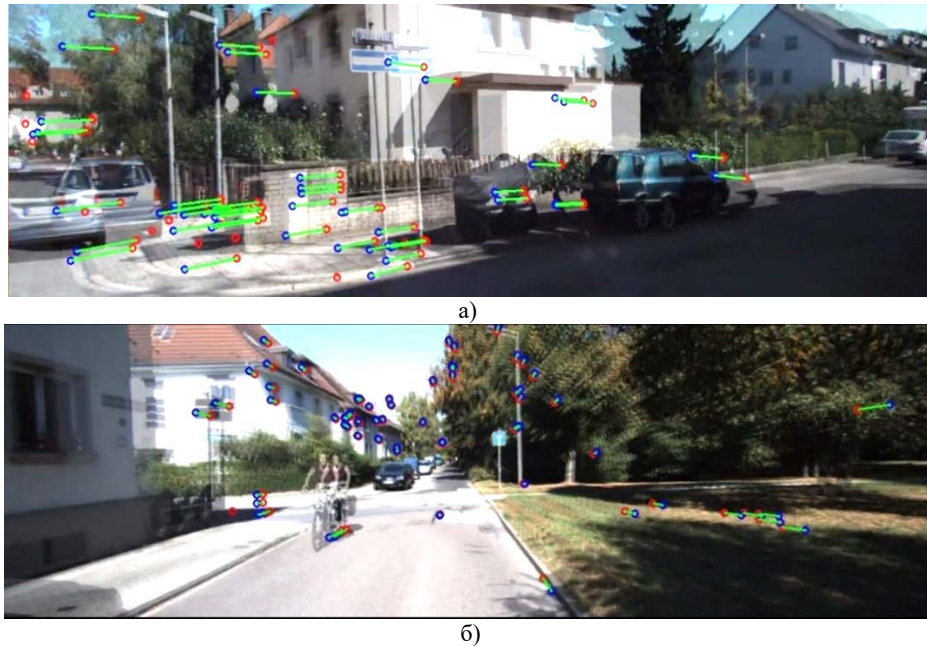


Рис. 1. Зміщення візуальних орієнтирів між двома сусідніми зображеннями відео послідовності kitti_27:
а) під час повороту; б) у разі прямого руху

Як видно з рис. 1, поворот автомобіля спричиняє значні зміщення візуальних орієнтирів між сусідніми кадрами, тоді як під час прямолінійного руху такі зміщення спостерігаються переважно для орієнтирів, розташованих близько до камери. У візуально-інерційному SLAM ці зміщення, разом з інформацією про кутові та лінійні швидкості з інерційного сенсора, дозволяють оцінити траєкторію руху.

Результати експериментальних досліджень

У ході досліджень для оцінювання якості SLAM використовувалася метрика абсолютна похибка траєкторії (ATE). Ця метрика визначає середнє відхилення між оціненою траєкторією та еталонною (Ground Truth) у просторі. Метрика ATE визначається за формулою

$$e_{ATE} = \sqrt{\frac{1}{K} \sum_{i=1}^K |G_i^{SLAM} - G_i^{GT}|^2}, \quad (12)$$

де G_i^{SLAM} — точка траєкторії SLAM, G_i^{GT} — точка траєкторії Ground Truth, K — загальна кількість точок траєкторії.

Для оптимізації обчислень і забезпечення актуальності карти запропоновано використовувати не всі візуальні орієнтири, а лише найрелевантніші. Релевантні орієнтири пропонується визначати на основі часу останнього спостереження. Тобто, орієнтири сортуються у порядку спадання їхньої релевантності, що обчислюється за часом останнього спостереження, і вибирається фіксована кількість N_{max} , які найбільше впливають на оновлення стану. Цей підхід дозволяє досягти балансу між точністю і швидкістю роботи системи.

Врахування шуму сенсорів у системах SLAM є критично важливим для забезпечення точності та стабільності роботи алгоритму в реальних умовах. Шум вимірювань, спричинений похибками сенсорів, таких як стереокамери чи інерційні модулі, впливає на визначення положення орієнтирів та робота [2]. Налаштування співвідношення Q/R у розширеному фільтрі Калмана, де Q відображає невизначеність моделі руху, а R моделює сенсорні похибки, визначає баланс між довірою до передбачення (dead reckoning) і корекцією на основі вимірювань. Неправильне моделювання шуму або некоректне налаштування Q/R може спричинити розбіжність фільтра, підвищену чутливість до шумів або накопичення помилок, що знижує ефективність системи. Оптимізація цих параметрів дозволяє мінімізувати ATE і забезпечує високу точність навігації автономних роботів у складних умовах.

Для оцінки впливу співвідношення Q/R на точність роботи візуально-інерційного SLAM прове-

дено серію експериментів з варіюванням рівнів шуму у моделі процесу (Q) та сенсорних вимірюваннях (R). Значення Q/R досліджувалися в діапазоні $Q/R \in [10^{-6}, 10^6]$ у логарифмічній шкалі. Для кожного значення Q/R виконувалася оцінка траєкторії SLAM та IMU (Dead Reckoning) з використанням метрики АТЕ, порівнюючи їх з еталонною GPS-траєкторією (Ground Truth). У дослідженні використовувалася середня кількість візуальних орієнтирів ($N_{\max} = 60$) для досягнення компромісу між точністю та швидкістю роботи системи, а також з огляду на зниження чутливості до шумів при меншій кількості орієнтирів.

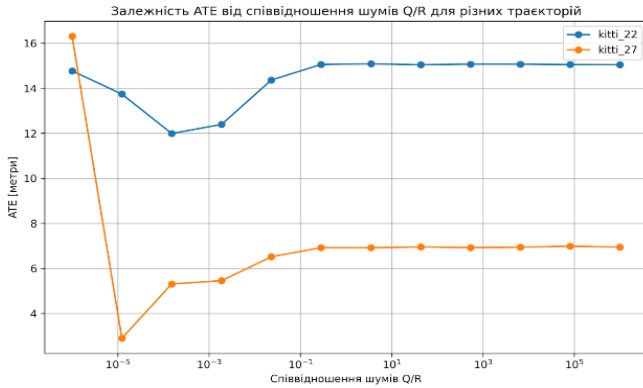


Рис. 2 Оцінка впливу співвідношення Q/R на точність роботи візуально-інерційного SLAM

релевантних орієнтирів варіюється від 10 до 120 з кроком 10, а значення Q/R набувають три фіксовані величини: 0,001, 0,0001 і 0,00001. Результати досліджень для наборів даних kitti_22 та kitti_27 показано на рис. 3.

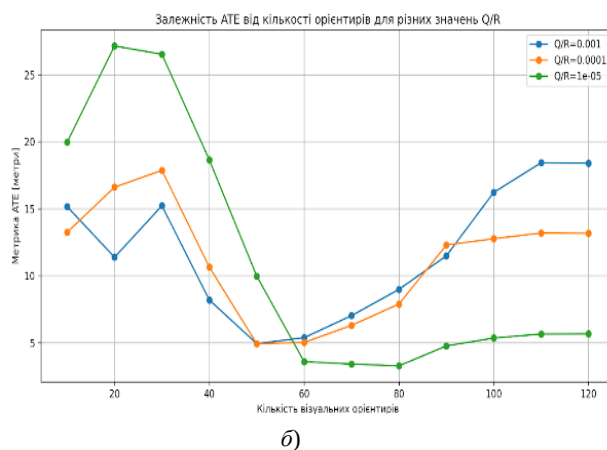
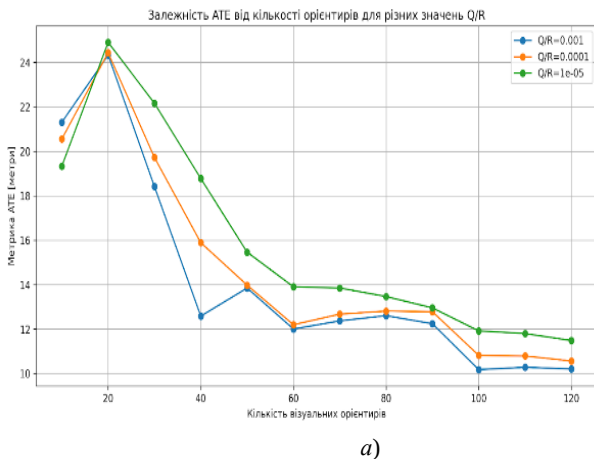


Рис. 3. Залежність показника АТЕ від кількості візуальних орієнтирів для наборів даних: а — kitti_22; б — kitti_27

За незначної кількості орієнтирів (до прикладу, 20 для kitti_27) високі значення Q/R (0,001) можуть показувати кращий результат (ATE = 12 м), оскільки система більше покладається на прогноз моделі руху, компенсуючи недоліки вимірювань від обмеженої кількості орієнтирів.

За середньої кількості орієнтирів (50...80) низькі значення Q/R (1e-05) забезпечують найкращі результати для kitti_27, оскільки система отримує достатньо інформації для точного оновлення стану та уникає надмірної залежності від моделі руху. Для kitti_22 кращі результати загалом показує $Q/R = 0,001$. Для обох наборів компромісом між точністю та швидкістю роботи є $N_{\max} = 60$, для kitti_27 АТЕ = 3,59 м, для kitti_22 АТЕ = 12,01 м.

У разі великої кількості орієнтирів (100...120) для kitti_27 всі три значення Q/R демонструють гірші результати, оскільки накопичуються похибки від менш релевантних або шумних орієнтирів. Натомість для kitti_22 збільшення кількості орієнтирів загалом покращує результати для всіх Q/R , досягаючи найнижчого АТЕ = 10 м з $Q/R = 0,001$ та $N_{\max} = 100$. Відмінність результатів пояснюється простішою траєкторією та динамікою руху у наборі kitti_22.

Розглянемо особливості побудови траєкторій SLAM, за крайніх випадків кількості візуальних орієнтирів, малій (20) та великій (110) на прикладі набору даних kitti_27 (рис. 4).

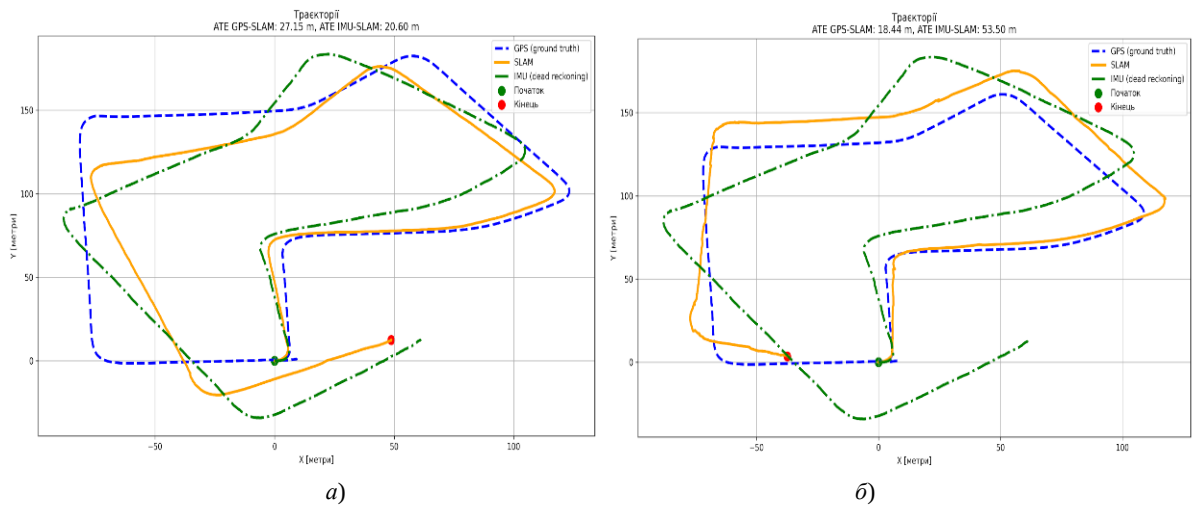


Рис. 4. Дослідження крайних випадків SLAM: *a* — 20 орієнтирів, $Q/R = 1e-05$; *б* — 110 орієнтирів, $Q/R = 0,001$

На рис. 4*a* показано результати експерименту з 20 орієнтирами та співвідношенням $Q/R = 1e-05$, де траєкторія SLAM наближена до Dead Reckoning. За показником ATE, відстань між SLAM і GPS становить 27,15 м, а між SLAM і Dead Reckoning — 20,6 м. Це пояснюється двома факторами: по-перше, мала кількість орієнтирів обмежує можливості системи для корекції стану, що зменшує ефективність компенсування помилок, накопичених моделлю руху. По-друге, низьке значення Q/R вказує на високу довіру до моделі руху, однак недостатня кількість орієнтирів не дозволяє сенсорним даним забезпечити точне коригування. У результаті система переважно покладається на Dead Reckoning, що і призводить до схожості між траєкторіями. На рис. 4*б* показано результати експерименту зі 110 орієнтирами та $Q/R = 0,001$, де траєкторія SLAM демонструє відхилення кінцевої точки та незамикання петлі. Відстань між SLAM і GPS становить 18,44 м, а між SLAM і Dead Reckoning — 53,5 м. Незважаючи на велику кількість орієнтирів, їхнє нерівномірне розташування у ключових зонах, а також накопичення помилок від менш релевантних або шумних орієнтирів обмежили ефективність корекції стану. Низьке значення Q/R , що зумовлює високу довіру до сенсорних даних, у поєднанні з тривалим рухом і накопиченням помилок моделі руху, виявилось недостатнім для замикання петлі.

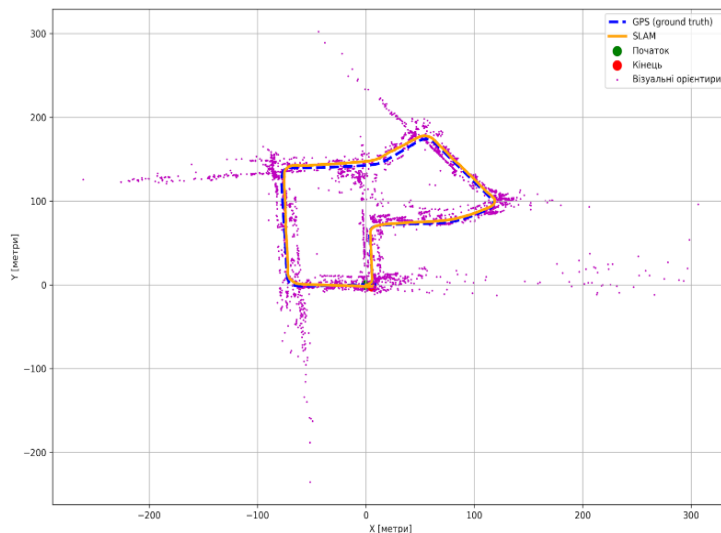


Рис. 5. Двовимірна реконструкція карти за допомогою візуально-інерційного SLAM набору даних kitti_27

Візуалізація реконструкції карти для набору даних kitti_27 за параметрів з найменшим ATE = 3,28 метри, коли $Q/R = 1e-5$ та 80 релевантних ознаках, показана на рис. 5.

На рис. 5 зображено побудовану траєкторію, отриману запропонованим методом візуально-інерційного SLAM, а також GPS-траєкторію та візуальні орієнтири, використані під час обчислень. Орієнтири дають змогу сформувати розріджену двовимірну карту з високою точністю, оскільки відхилення траєкторії SLAM від ground truth за показником ATE становить 3,28 м. Окрім низького рівня ATE, коректність результатів підтверджує також замикання петлі.

Висновки

Дослідження підтвердило ефективність використання візуально-інерційного підходу на основі розширеного фільтра Калмана для задач автономної навігації. Аналіз та експерименти, проведені на основі датасету KITTI, продемонстрували, що ефективність алгоритму SLAM значною мірою залежить від кількості використаних візуальних орієнтирів та співвідношення шумів Q/R , яке ви-

значає баланс між довірою до моделі руху (на основі ІВП) та корекцією на основі сенсорних вимірювань (візуальних орієнтирів).

У дослідженні для визначення візуальних орієнтирів запропоновано використовувати ознаки ORB, які відзначаються швидкістю обчислення та стійкістю до змін освітлення, обертання та масштабування. Це забезпечує ефективність роботи алгоритму навіть у складних умовах. До того ж у роботі запропоновано алгоритм вибору релевантних орієнтирів, який ґрунтується на часі останнього спостереження. Такий підхід дозволяє зберігати найактуальніші орієнтири, що підвищує точність і швидкість роботи алгоритму SLAM.

Аналіз результатів показав, що оптимальний діапазон значень співвідношення шумів Q/R для використаних наборів даних лежить у межах від 0,001 до 0,00001. З погляду балансу між швидкістю та якістю побудови траєкторії (порівняно з GPS-даними, які використовувалися як ground truth), оптимальна кількість релевантних орієнтирів для досліджуваних наборів даних становить від 50 до 80. Зокрема, як за співвідношення $Q/R = 0,001$ так і за співвідношення $Q/R = 1e-05$ оптимальною є кількість релевантних орієнтирів 60.

Подальші напрями вдосконалення візуально-інерційного SLAM полягають у: оптимізації вибору візуальних орієнтирів, оскільки це може дозволити підвищити точність і швидкість обробки, водночас знижуючи обчислювальні витрати; дослідженні використання монокулярних камер, оскільки вони є доступнішими, компактнішими та енергоефективними; а також у дослідженні інваріантного розширеного фільтра Калмана, який може підвищити точність картографування та стійкість системи у складних і динамічних умовах.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] А. Жарков, Р. Маслій, і В. Гармаш, «Аналіз підходів Visual SLAM для задачі навігації автономного робота», *Herald Khmelnytskyi Nat. Univ. Tech. sci.*, т. 335, № 3(1), с. 67-77, 2024. <https://doi.org/10.31891/2307-5732-2024-335-3-10>.
- [2] M. Quan, S. Piao, M. Tan, and S.-S. Huang, "Accurate Monocular Visual-Inertial SLAM Using a Map-Assisted EKF Approach", *IEEE Access*, vol. 7, pp. 34289-34300, 2019. <https://doi.org/10.1109/access.2019.2904512>.
- [3] Y. Ning, "A Comprehensive Introduction of Visual-Inertial Navigation," *Arxiv.org*. 2023. <https://arxiv.org/abs/2307.11758>.
- [4] J. A. Castellanos, J. Neira, and J. D. Tardós, "Limits to the consistency of EKF-based SLAM," *IFAC Proc.*, vol. 37, no. 8, pp. 716-721, 2004. [https://doi.org/10.1016/s1474-6670\(17\)32063-3](https://doi.org/10.1016/s1474-6670(17)32063-3).
- [5] S. Konatowski, P. Kaniewski, and J. Matuszewski, "Comparison of estimation accuracy of EKF, UKF and PF filters," *Annu. Navig.*, vol. 23, no. 1, pp. 69-87, 2016. <https://doi.org/10.1515/aon-2016-0005>.
- [6] C. Urrea, and R. Agramonte, "Kalman filter: Historical overview and review of its use in robotics 60 years after its creation," *J. Sensors*, pp. 1-21, 2021. <https://doi.org/10.1155/2021/9674015>.
- [7] G. P. Huang, A. I. Mourikis, and S. I. Roumeliotis, "Observability-based rules for designing consistent EKF SLAM estimators," *Int. J. Robot. Res.*, vol. 29, no. 5, pp. 502-528, 2009. <https://doi.org/10.1177/0278364909353640>.
- [8] A. Barrau, and S. Bonnabel, "Invariant kalman filtering," *Annu. Rev. Control, Robot., Auton. Syst.*, vol. 1, no. 1, pp. 237-257, 2018. <https://doi.org/10.1146/annurev-control-060117-105010>.
- [9] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231-1237, 2013. <https://doi.org/10.1177/0278364913491297>.
- [10] Andreas Geiger, *The KITTI vision benchmark suite*. [Electronic resource]. Available: <https://www.cvlibs.net/datasets/kitti/index.php>.
- [11] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF", in *2011 IEEE Int. Conf. Comput. Vis. (ICCV)*, Barcelona, Spain, 6–13 Nov. 2011. IEEE, 2011. <https://doi.org/10.1109/iccv.2011.6126544>.

Рекомендована кафедрою автоматизації та інтелектуальних інформаційних технологій ВНТУ

Стаття надійшла до редакції 20.02.0225

Жарков Анатолій Володимирович — аспірант кафедри автоматизації та інтелектуальних інформаційних технологій, e-mail: fkca.1akitjav@gmail.com ;

Маслій Роман Васильович — канд. техн. наук, доцент кафедри автоматизації та інтелектуальних інформаційних технологій, e-mail: maslij.r.v@vntu.edu.ua

Вінницький національний технічний університет, Вінниця

Visual-Inertial Slam Using the Extended Kalman Filter for Autonomous Navigation

¹Vinnitsia National Technical University

The study is devoted to the use of visual-inertial SLAM based on the extended Kalman filter (EKF) for autonomous navigation tasks. The study used data from the KITTI dataset, which include images from a stereo camera, as well as linear and angular velocities obtained from an inertial measurement device. To determine visual landmarks, it is proposed to use ORB features, which are characterized by their speed of calculation and resistance to changes in lighting, rotation and scaling. In addition, an algorithm for selecting relevant landmarks has been developed, which allows to increase the accuracy and speed of SLAM.

The results of a series of experiments have shown that the efficiency of the system largely depends on the adjustment of the noise ratio of the motion model (Q) and sensor measurements (R). It was determined that the optimal range of Q/R for the studied datasets is 0.001...0.00001.

To achieve a balance between speed and system performance, the influence of the number of relevant visual features in the range from 10 to 120 was investigated. The results of the experiments showed that the optimal number of landmarks is 50...80, and the best results were achieved at $Q/R = 0.001$ or $Q/R = 0.00001$, depending on the dataset used. The trajectory accuracy was assessed using the ATE metric (Absolute Trajectory Error). GPS data was used as ground truth for verification.

The results confirmed that visual-inertial SLAM is an effective tool for autonomous navigation, especially in the absence of GPS. The use of EKF in visual-inertial SLAM reduces computational costs compared to other methods, which makes it optimal for devices with limited resources, in particular mobile robots.

Further research will focus on optimizing landmark selection, improving noise modeling, and adapting the algorithm for monocular cameras to improve SLAM accuracy and efficiency.

Further research will focus on optimizing landmark selection, improving noise modeling, and adapting the algorithm for monocular cameras to improve SLAM accuracy and efficiency. In addition, research on the invariant extended Kalman filter (IEKF) is planned, which has the potential to improve mapping accuracy and enhance system robustness in complex and dynamic environments.

Keywords: visual-inertial SLAM, extended Kalman filter, inertial measurement unit, KITTI dataset.

Zharkov Anatolii — Post-Graduate Student of the Chair of Automatization and Intellectual Informational Technologies, e-mail: fkca.1akitjav@gmail.com ;

Maslii Roman V. — Cand. Sc. (Eng.), Associate Professor of the Chair of Automatization and Intellectual Informational Technologies, e-mail: maslij.r.v@vntu.edu.ua