

ЖАРКОВ АНАТОЛІЙ

Вінницький національний технічний університет

<https://orcid.org/0009-0005-9857-644X>e-mail: fkca.lakitjav@gmail.com**МАСЛІЙ РОМАН**

Вінницький національний технічний університет

<https://orcid.org/0000-0003-3021-4328>e-mail: maslij.r.v@vntu.edu.ua**ГАРМАШ ВОЛОДИМИР**

Вінницький національний технічний університет

<https://orcid.org/0009-0007-1861-8772>e-mail: garmash.v.v@vntu.edu.ua

АНАЛІЗ ПІДХОДІВ VISUAL SLAM ДЛЯ ЗАДАЧІ НАВІГАЦІЇ АВТОНОМНОГО РОБОТА

Робота присвячена дослідженню задачі навігації автономного робота за допомогою методів Visual SLAM. У роботі представлено формулювання класичної задачі SLAM у вигляді факторного графа і розглянуто її рішення за допомогою оптимальної оцінки максимуму апостеріорної імовірності (MAI). Проведено аналіз доцільності та практичної ефективності основних підходів Visual SLAM (Simultaneous Localization and Mapping) для навігації автономного робота в закритому та відкритому середовищі. Розглянуто такі підходи Visual SLAM, як: тільки візуальний SLAM (Visual-Only SLAM), який в свою чергу поділяють на методи, на основі ознак (Feature-based methods) та прямі методи (Direct methods); візуально-інерційний SLAM (Visual-inertial SLAM) та RGB-D SLAM. Методи Visual SLAM вирішують головну проблему SLAM, яка полягає в тому, що автономну роботу необхідно постійно будувати карту, знаходячись у незнайомому середовищі, при цьому на кожній ітерації оцінювати як своє власне місцеположення, так і місцеположення орієнтирів, які він використовує для навігації. Запропоновано десять критеріїв оцінки ефективності кожного з підходів, які дозволяють в повній мірі зрозуміти їх сильні та слабкі сторони, та умови, при яких їх використання буде доцільним. Здійснено оцінку кожного підходу по кожному з критеріїв та в результаті отримано середній показник ефективності кожного підходу. Підходом, з найкращим середнім показником виявився візуально-інерційний SLAM, який має найкращі показники точності та стабільності роботи, проте його сенсорна система дорожча у порівнянні з іншими підходами. З іншого боку, методи, на основі ознак та прямі методи хоч показали дещо гіршу середню оцінку, проте мають досить збалансовані показники за кожним критерієм та є більш простими та доступними для реалізації. Загалом, отримані результати дозволяють чітко зрозуміти основні переваги та недоліки кожного з підходів, а також надають можливість зекономити час на проведенні аналізу дослідникам при виборі оптимального підходу.

Ключові слова: автономні роботи, одночасна локалізація і картографування, максимум апостеріорної імовірності.

ZHARKOV ANATOLIY, MASLIJ ROMAN, HARMASH VOLODYMYR

Vinnytsia National Technical University

ANALYSIS OF VISUAL SLAM APPROACHES FOR THE PROBLEM OF NAVIGATION OF AN AUTONOMOUS ROBOT

The work is devoted to the study of the problem of autonomous robot navigation using Visual SLAM methods. The paper presents the formulation of the classic SLAM problem in the form of a factorial graph and considers its solution using the optimal estimate of the maximum posterior probability. An analysis of the feasibility and practical effectiveness of the main Visual SLAM approaches for autonomous robot navigation in closed and open environments has been carried out. Visual SLAM approaches are considered, such as: Visual-Only SLAM, which in turn is divided into Feature-based methods and Direct methods; Visual-inertial SLAM and RGB-D SLAM. Visual SLAM methods solve the main problem of SLAM, which is that an autonomous robot must continuously build a map while in an unknown environment, while at each iteration estimating both its own location and the locations of the landmarks it uses for navigation. The authors proposed ten criteria for evaluating the effectiveness of each of the approaches, which allow to fully understand their strengths and weaknesses, and the conditions under which their use would be appropriate. Each approach was evaluated according to each of the criteria, and as a result, the average indicator of the effectiveness of each approach was obtained. The approach with the best average score turned out to be visual-inertial SLAM, which has the best performance accuracy and stability, but its sensor system is more expensive compared to other approaches. On the other hand, methods based on features and direct methods, although they showed a slightly worse average score, have fairly balanced indicators for each criterion and are simpler and more accessible to implement. In general, the obtained results allow a clear understanding of the main advantages and disadvantages of each of the approaches, as well as provide an opportunity to save time for the analysis of researchers when choosing the optimal approach.

Keywords: autonomous robots, simultaneous localization and mapping, maximum a posteriori estimation.

Вступ

Автономна навігація мобільних роботів стала активною дослідницькою сферою в останні роки завдяки прогресу в матеріалознавстві для конструювання роботів, компактному розміру батареї для більш тривалої віддаленої роботи та збільшенню потужності обчислювального обладнання для функціонування алгоритмічних методів і методів штучного інтелекту. Автономні роботи повинні пересуватися в навколишньому середовищі для виконання своїх завдань, вивчаючи його. Роботи «бачать і розуміють» світ, збираючи інформацію з приєднаних датчиків. Для того, щоб роботи могли взаємодіяти з реальним світом, їм

потрібні певні можливості розуміння сцени геометрично та інтерпретації її семантично. Точна локалізація робота має важливе значення для того, щоб робот виконував відповідні дії, а добре сприйняття та розуміння навколишнього середовища дозволяє роботу реагувати на навколишнє середовище [1].

Аналіз останніх досліджень

В загальному проблематика методу SLAM базується на тому, що робот повинен працювати в невідомому середовищі, щоб побудувати карту, оцінюючи своє невизначене розташування. SLAM є фундаментальною проблемою в численних додатках робототехніки, яка вимагає від робота автономної навігації в певному середовищі та взаємодії з реальним світом [2].

Автономні роботи, як правило, отримують інформацію від датчиків, приєднаних до них, наприклад, у формі зображення або 3D-просторових точок від високочастотного сканування LiDAR [3]. Однак цієї сприйнятої інформації часто недостатньо для робота, щоб орієнтуватися, оскільки йому бракує геометричного розуміння для реконструкції сцени. Геометричне моделювання особливо важливо в складних завданнях, де робот повинен локалізувати себе відносно змодельованої карти сцени, щоб орієнтуватися та досягати своїх цілей. Наприклад, навігація мобільного робота часто вимагає, щоб робот підтримував представлення карти для виконання планування в таких завданнях, як планування руху, навігація, переміщення роботизованих рук або навіть автономних автомобілів. Для робота неможливо виконувати безпечні автономні операції в динамічному середовищі, якщо йому не вистачає здатності сприймати й осмислювати потенційні перешкоди. Це особливо важливо для роботи робота в новому місці без попередньої інформації про навколишнє середовище, наприклад, під час дослідження планет або пошуково-рятувальних операцій [4, 5].

Труднощі геометричної реконструкції навколишнього середовища під час навігації в невідомому середовищі часто виникають через такі проблеми, як узгодження даних від різних датчиків, неузгодженість зображень карти з різних ракурсів спостереження або шуми спостереження. SLAM дозволяє автономному роботу уточнювати реконструйовану карту та своє поточне місцезнаходження ітераційно, досліджуючи невідоме середовище за допомогою різних датчиків [6]. У Visual-SLAM найпростіше представлення середовища часто складається з колекції розріджених 3D-точок, які постійно обробляються під час навігації. Створену карту можна використовувати для створення карти нового середовища або для автономного планування роботом своєї місії. Така система може підтримувати стабільність, планувати власні рухи та реагувати на динамічні зміни навколишнього середовища без втручання людини [7].

Методи Visual SLAM можна вважати особливо привабливими через простоту конфігурації, мініатюрний розмір і низьку вартість сенсорів, що вони використовують. Методи Visual SLAM поділяють на три основні групи: тільки візуальний SLAM, візуально-інерційний SLAM і RGB-D SLAM. Першу групу поділяють на дві окремих, а саме: SLAM на основі ознак та прямий SLAM. Ці дві підгрупи методів мають вищу технічну складність через обмежені вхідні дані в порівнянні з іншими двома групами методів [8].

Формулювання цілей

Аналіз основних підходів Visual SLAM може мати велике значення для дослідників, які займаються розробкою автономних роботизованих систем, адже дозволяє зекономити час на власний аналіз та вибрати відповідний підхід відносно задач проекту. Для аналізу важливо оцінити підходи Visual SLAM базуючись на критеріях ефективності. Тому важливим є розробка загальних для всіх підходів критеріїв ефективності та оцінювання кожного підходу за кожним критерієм з можливістю отримання інтегральної оцінки кожного підходу, що дасть можливість обрати оптимальний підхід.

Отже **метою дослідження** є загальне визначення задачі SLAM, аналіз існуючих підходів Visual SLAM та визначення найбільш оптимального підходу відповідного до запропонованих критеріїв якості.

SLAM: постановка проблеми

Складність проблеми пов'язана з рекурсивною залежністю: побудова карти часто залежить від того, що робот спостерігає навколишнє середовище з певного відомого місця, тоді як оцінка стану також часто вимагає від робота зробити висновок про своє місцезнаходження, спираючись на деякі відомі орієнтири. Алгоритм SLAM оцінює рух датчика і одночасно реконструює геометричну структуру відвіданої області. В роботі [9] вперше сформулювали постановку задачі в 1985 році для навігації мобільних роботів. Проблема полягає в необхідності моделювати навколишнє середовище і правильно визначати своє положення незважаючи на неточності, внесені датчиками. Запропонована методологія визначає загальний принцип роботи з невизначеністю в зібраних даних і знаходження мобільним роботом своїх орієнтирів під час дослідження навколишнього середовища.

Проблему Visual SLAM іноді формулюють за допомогою факторного графа (рисунок 1). Факторний граф — це дводольний граф, який визначає, як глобальну функцію багатьох змінних можна розкласти на множники та добуток локальних функцій [10].

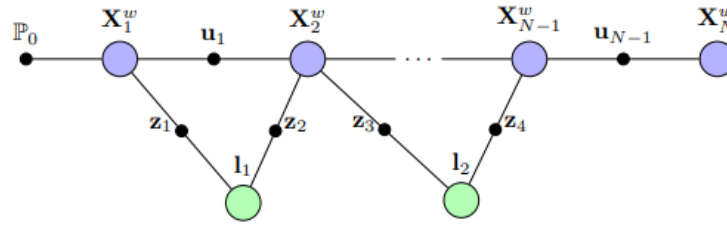


Рис. 1. Формулювання проблеми Visual SLAM за допомогою факторного графа

Кожен синій та зелений вузол на рисунку 1, відомі як змінні, представляє собою набір прихованих змінних, які потрібно оцінити, якими у випадку SLAM є стан робота та орієнтири. Вузли між змінними, також відомі як фактори, є набором обмежень та інформації між змінними. На рисунку 1 позиції камери позначаються як X_i^w , а орієнтири як l_j . Спостереження орієнтирів і одометрії в різних положеннях камери позначаються як z_k і u_i відповідно. Априорна імовірність щодо початкової позиції позначається як P_0 , і спільний розподіл ймовірностей проблеми MAI може бути обчислений як добуток зображених факторів.

Загалом, факторний граф можна використовувати, щоб розкласти на множники спільний розподіл ймовірностей за деякими випадковими змінними, кодуючи притаманну умовну незалежність деяких локальних змінних у спільний розподіл ймовірностей.

Формулювання SLAM

SLAM — це мультидисциплінарна проблема, яка охоплює як комп'ютерний зір, так і робототехніку, і традиційно формулюється як максимум апостеріорної імовірності [1]. У Visual SLAM можна визначити $X = \{X_i^w\}_{i=1}^N$ як траєкторію робота в часі, де X_i^w позначає позицію робота, параметризовану в наборі жорстких евклідових перетворень $\mathbb{SE}(3)$. Нехай $L = \{l_j\}_{j=1}^M$ позначає набір орієнтирів, параметризованих їх відповідним простором представлення, $Z = \{z_k\}_{k=1}^K$ буде набором спостережень виявлених орієнтирів і $U = \{u_i\}_{i=1}^{N-1}$ буде набором одометричних вимірювань між позиціями робота. Дані спостереження за Z орієнтирами збираються за деякою моделлю спостереження $h_k(\cdot)$, заданою як:

$$z = h_k(X_{ik}, l_{jk}) + \epsilon_k \quad (1)$$

де X_{ik} і l_{jk} позначають фактичний стан робота та позицію орієнтира, а ϵ_k — випадковий шум вимірювання. Рішенням проблеми SLAM є оптимальна оцінка MAI

$$X^*, L^* = \underset{X, L}{\operatorname{argarg\,max}} P(X, L | Z, U) \quad (2)$$

де $P(X, L | Z, U)$ є спільною ймовірністю всіх прихованих оціночних змінних, враховуючи всі попередні спостереження та вимірювання. Для класичної задачі SLAM без одометричних вимірювань [10] можна змінити (2) як

$$X^*, L^* = \underset{X, L}{\operatorname{argarg\,max}} P(Z) \quad (3)$$

$$X^*, L^* = \underset{X, L}{\operatorname{argarg\,max}} P(Z | X, L) P(X, L) \quad (4)$$

де $P(Z | X, L)$ це ймовірність отримання вимірювання Z за X і L , а $P(X, L)$ є попередньою інформацією про X та L . Припускаючи, що кожне спостереження z_k є незалежним, можна обрахувати (4) як

$$X^*, L^* = \underset{X, L}{\operatorname{argarg\,max}} \prod_{k=1}^K P(z_k | X_k, l_k) P(X, L) \quad (5)$$

Підходи Visual SLAM та критерії їх оцінювання

Методи Visual SLAM використовують одну або кілька камер у сенсорній системі, отримуючи 2D зображення як джерело інформації. Загалом методи SLAM на основі візуальної інформації поділяються на три основні частини: ініціалізація, відстеження та картографування [11]. В залежності від використання додаткових вимірювальних пристроїв методи Visual SLAM класифікують на три основних групи:

- тільки візуальний SLAM;
- візуально-інерційний SLAM;
- RGB-D SLAM.

Глибина та інерційні дані можуть бути додані до двовимірного вхідного 2D зображення для створення реконструкції розрідженої, напівщільної чи щільної карти.

На рисунку 2 показано основні компоненти Visual SLAM розглянуті вище.

Як видно з рисунку 2, у системах Visual SLAM вхідними даними можуть бути як 2D-зображення, так і 2D-зображення разом з даними інерційного вимірювального пристрою, або 2D-зображення разом з даними глибини, залежно від використовуваного підходу, тобто тільки візуальний, візуально-інерційний або на основі RGB-D, відповідно [8]. Ініціалізація визначає глобальні координати та створює початкову карту, яка використовується для виконання двох основних кроків: відстеження та картографування. Процес відстеження відповідає за постійну оцінку положення датчика. Загалом, алгоритм встановлює 2D–3D відповідності між поточним кадром і картою, що становить проблему відому в літературі під назвою

Perspective-n-Point. Існує кілька способів вирішення цієї проблеми, зокрема одним із можливих рішень є метод EPnP [12]. Процес картографування відповідає за обчислення та розширення 3D-реконструкції під час руху камери. Обчислення даних глибини відрізняється залежно від використовуваного підходу. В результаті, процеси картографування призводять до створення розрідженої, напівщільної або щільної 3D-реконструкції, відповідно до реалізованої техніки.

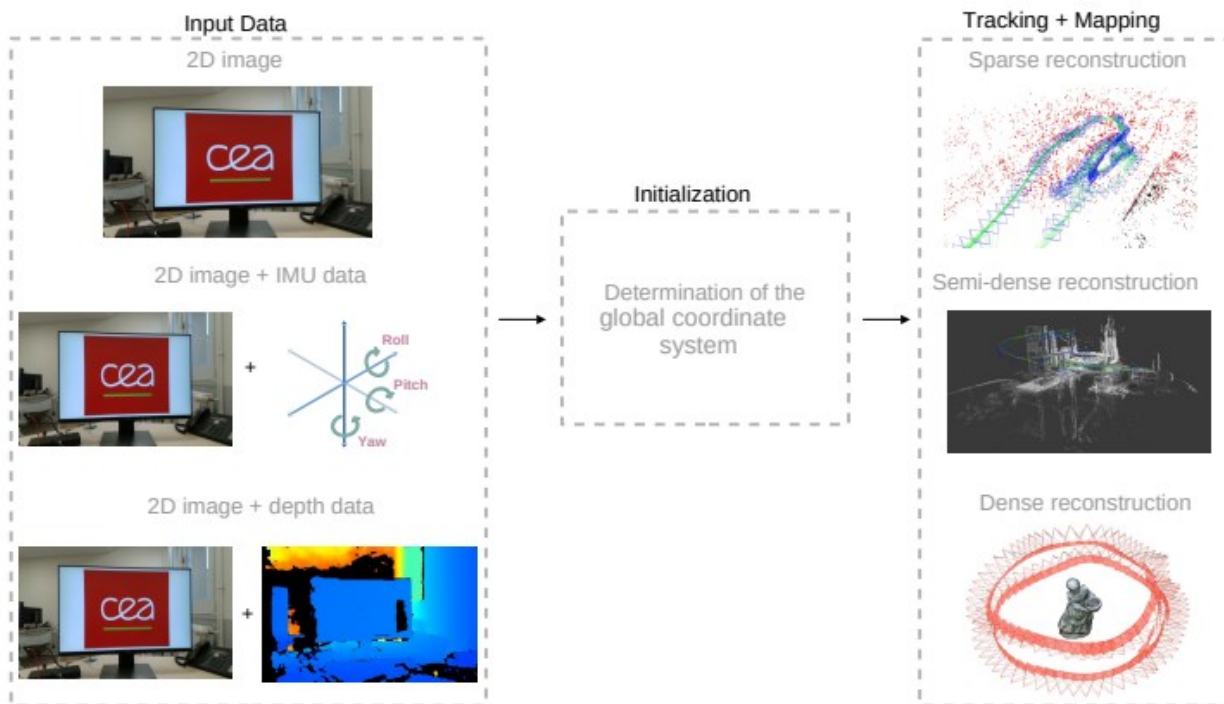


Рис. 2. Основні компоненти Visual SLAM

На основі аналізу літературних джерел та огляду підходів Visual SLAM визначені такі критерії порівняння підходів:

- **Точність:** якість визначення положення та створення карти; вимірюється відповідністю результатів роботи алгоритму до реальної геометрії середовища;
- **Стабільність:** здатність до підтримання стійкості та точності в різних умовах, таких як різка зміна освітлення, рух об'єктів тощо;
- **Обчислювальна складність:** кількість обчислень, яку вимагає алгоритм для обробки вхідних даних та видачі результатів. Важливо для реального часу та використання у вбудованих системах;
- **Вартість обладнання:** загальна вартість сенсорів та обладнання, яке використовується для виконання Visual SLAM. Включає вартість камер, датчиків глибини, інерційних сенсорів тощо;
- **Ефективність в реальному часі:** здатність алгоритму працювати в реальному часі без затримок у визначенні положення та створенні карти;
- **Робочий діапазон:** діапазон умов, в яких метод може ефективно функціонувати. Деякі методи можуть бути більш адаптованими для використання в закритому середовищі, інші у відкритому середовищі;
- **Відмовостійкість:** здатність алгоритму виявляти та виправляти помилки, такі як втрата відомостей про відстань, зміни освітлення, перешкоди тощо;
- **Необхідність калібрування:** спроможність методу працювати без необхідності постійного калібрування обладнання;
- **Сумісність з різними сенсорами:** можливість використання різних типів сенсорів (камери, датчики глибини, інерційні датчики) та їх комбінацій для отримання кращої точності та стабільності;
- **Енергоспоживання:** загальна кількість енергії, необхідна для роботи алгоритму.

Тільки візуальний SLAM

Візуальні системи SLAM засновані на обробці 2D-зображень. Після отримання зображень з кількох точок зору система виконує процес ініціалізації, щоб визначити глобальну систему координат і реконструювати початкову карту. В методах на основі ознак на першому кроці здійснюється ініціалізація точок карти з високою невизначеністю, які пізніше можуть сходитися до своїх фактичних позицій. Ця процедура супроводжується відстеженням, яке намагається оцінити позицію камери. Одночасно процес картографування включає нові точки в 3D-реконструкції, оскільки спостерігається більше невідомих сцен.

Тільки візуальний SLAM може використовувати монокулярну або стереокамеру. SLAM на основі монокулярної камери є добре дослідженою областю, враховуючи малий розмір датчика (найменший з усіх представлених підходів), його низьку ціну, легке калібрування та знижене енергоспоживання [13]. Незважаючи на ці переваги, системи на базі монокулярів пропонують більш високу складність ініціалізації системи, оскільки для визначення початкової глибини необхідні принаймні два різні точки огляду, а також проблеми, пов'язані з дрейфом і оцінкою масштабу. Останню проблему можна компенсувати за допомогою стереокамер, головною перевагою яких є стереозображення лише в одному кадрі. Однак розмір датчика значно більший, ніж проста монокулярна камера. Крім того, для кожного кадру потрібна додаткова обробка, в основному через необхідність процесу виправлення зображення на етапі стереозбігу.

Методи підходу тільки візуального SLAM ділять на дві категорії: методи, на основі ознак (непрямі) та прямі методи. На рисунку 3 наведені загальні відмінності між цими методами відповідно до частини алгоритму, відповідальної за абстракцію даних датчика (front end), і частини, відповідальної за інтерпретацію абстрагованих даних (back end) [16].

Як видно з рисунку 3 для методів основаних на ознаках використовується розріджена реконструкція карти, в свою чергу для прямих методів використовується або розріджена або напівщільна реконструкція карти.

Методи, на основі ознак

Методи SLAM на основі ознак розглядають певну кількість точок інтересу, які називаються ключовими точками. Їх можна виявити на кількох зображеннях і порівняти їх дескриптори; цей процес надає інформацію про оцінку позиції камери. Дані дескриптора та розташування ключової точки складають ознаку, тобто дані, які використовуються алгоритмом для обробки відстеження та картографування.

Оскільки методи, засновані на ознаках, не використовують всю інформацію кадру, вони підходять для застосування у вбудованих системах. Однак вилучення ознак може не спрацювати в середовищі, зі слабко вираженою текстурою, до того ж результатом генерується розріджена карта, яка надає менше інформації, ніж щільна [14]. На рисунку 4 наведено приклад роботи методу ORB-SLAM [15], що належить до методів, на основі ознак.

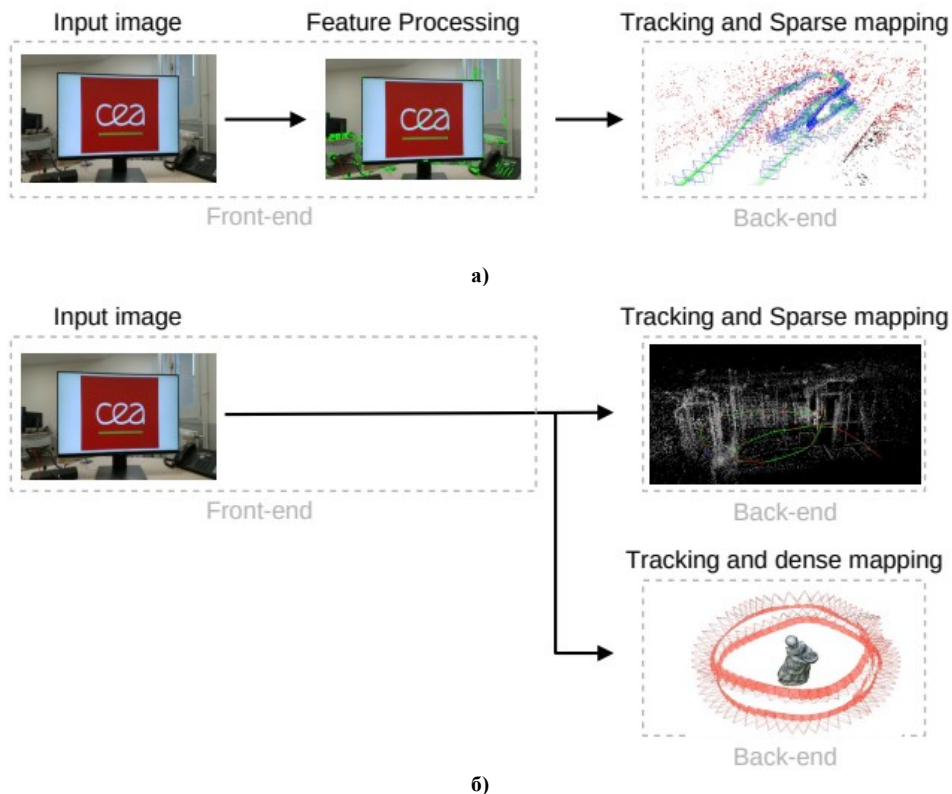


Рис. 3. Загальні відмінності між методами, на основі ознак та прямими методами Visual SLAM: а) методи на основі ознак; б) прямі методи

Проведемо оцінку даного підходу, відповідно до критеріїв, зазначених раніше:

- **Точність:** підхід має досить високу точність в умовах, коли на зображеннях присутні чітко виражені ознаки. Однак в умовах низької освітленості або на поверхнях без характерних ознак можуть виникати проблеми з точністю;
- **Стабільність:** методи SLAM на основі ознак мають проблеми зі стабільністю в умовах низької освітленості або на поверхнях без характерних ознак. Переміщення камери або зміна освітлення може також призвести до втрати даних про особливі точки;
- **Обчислювальна складність:** підхід є менш обчислювально витратним порівняно з іншими

методами, оскільки вимагає виявлення, відстежування та зберігання лише особливих точок на зображеннях;

- *Вартість обладнання:* методи SLAM на основі ознак вимагають використання високоякісних камер з хорошою роздільною здатністю та швидкістю зчитування. Вартість камер може бути помірною, але може залежати від точності та продуктивності;
- *Ефективність в реальному часі:* підхід є більш ефективним у реальному часі порівняно з іншими методами, оскільки вимагає виявлення та відстежування лише особливих точок на зображеннях, що є менш обчислювально витратним процесом;
- *Робочий діапазон:* підхід може працювати добре як у відкритих, так і закритих середовищах, але його ефективність залежить від наявності чітко виражених ознак на поверхнях для відстежування;
- *Відмовостійкість:* підхід вразливий до значних змін у середовищі або освітленні, оскільки він зазвичай використовує особливі точки для визначення положення. Потрібна додаткова обробка для виявлення та адаптації до змін;
- *Необхідність калібрування:* підхід є відносно легким для калібрування, хоча й потрібно правильно налаштувати параметри виявлення особливих точок та їх відстеження на зображеннях;
- *Сумісність з різними сенсорами:* підхід менш сумісний з сенсорами порівняно з іншими методами, оскільки він зазвичай спеціалізується на використанні візуальної інформації;
- *Енергоспоживання:* підхід як правило не є енерговитратним, проте вимагає певних обчислювальних ресурсів для виявлення та відстежування особливих точок на зображеннях.

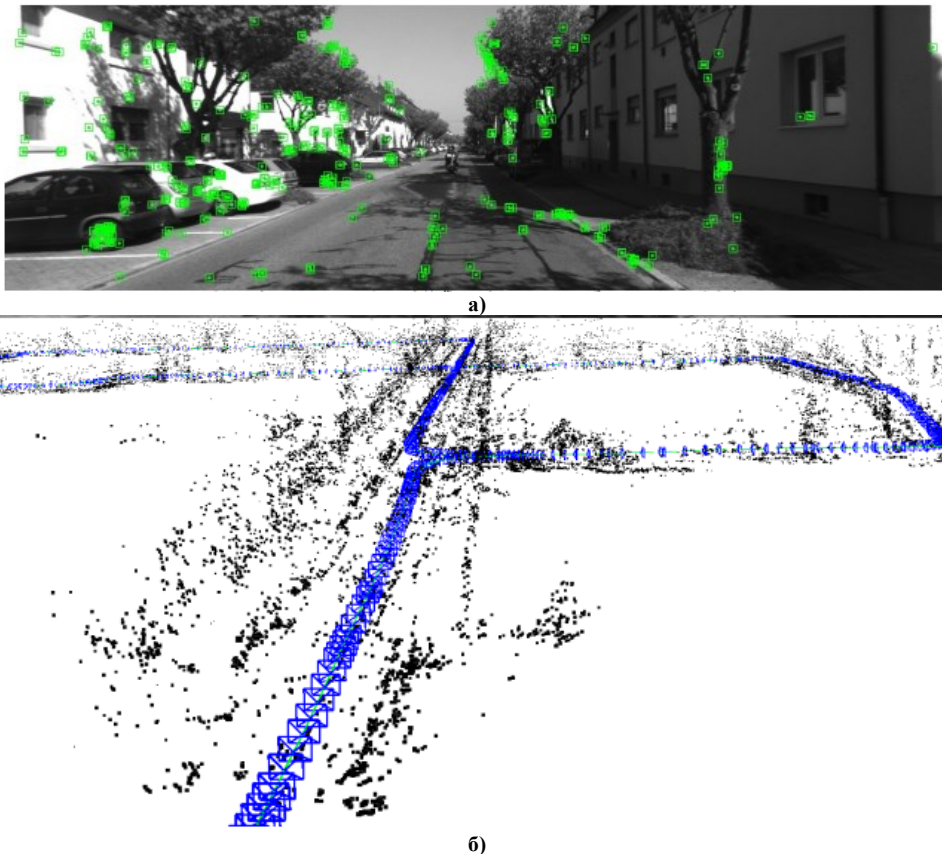


Рис. 4. Розріджена тривимірна реконструкція послідовності seq 00 набору зображень KITTI методом на основі ознак ORB-SLAM: а) кадр з послідовності, з ознаками ORB, що відстежуються; б) розріджена карта, отримана шляхом відстеження ознак ORB

Прямі методи

На відміну від методів, заснованих на ознаках, прямі методи використовують дані камери без попередньої обробки, враховуючи інтенсивність пікселів і мінімізуючи фотометричну похибку. Існує багато різних алгоритмів, заснованих на цьому підході, і залежно від обраної методики реконструкція може бути щільною, напівщільною або розрідженою. Щільність реконструкції є суттєвим обмеженням для роботи алгоритму в реальному часі, оскільки спільна оптимізація позицій карти та камери є більш дорогою з точки зору обчислень для щільної та напівщільної реконструкції, ніж для розрідженої. На рисунку 5 зображено результати роботи широкомасштабного прямого монокулярного SLAM [17], що належить до прямих методів. Цей метод створює узгоджену глобальну карту, використовуючи пряме вирівнювання зображення та ймовірнісні напівщільні карти глибини замість ключових точок.

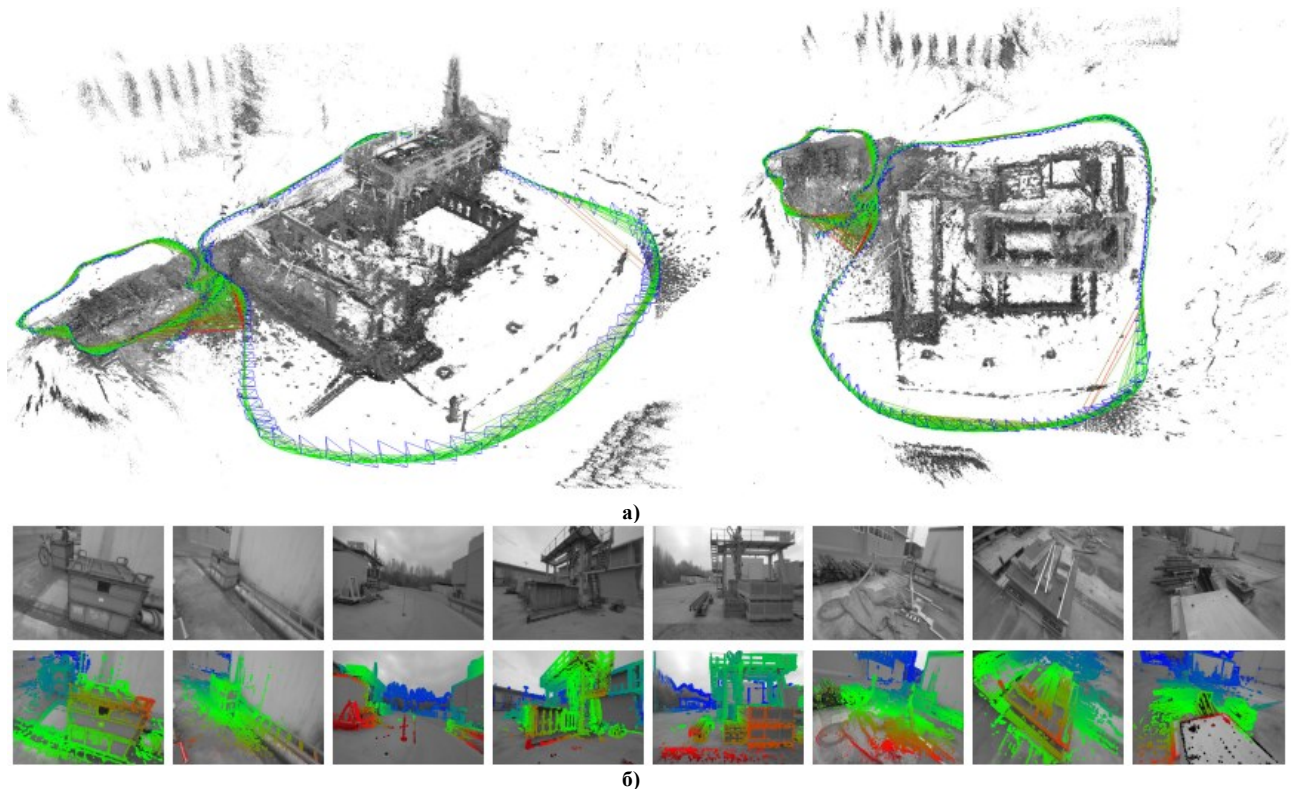


Рис. 5. Результат роботи широкомасштабного прямого монокулярного SLAM: а) накопичені хмари точок усіх ключових кадрів траєкторії середнього розміру (з ручної монокулярної камери), згенеровані в режимі реального часу; б) набір ключових кадрів із кольоровою напівщільною інверсною картою глибини

Проведемо оцінку даного підходу, відповідно до критеріїв, зазначених раніше:

- **Точність:** методи, які розглядають всю інформацію зображення як неперервні функції, мають високу точність за умови стабільного освітлення та достатньої кількості деталей для використання у процесі відстеження. Проте, вони можуть бути менш точними на поверхнях зі слабо вираженою текстурою чи в умовах низької освітленості;
- **Стабільність:** прямі методи менш стабільні в умовах низької освітленості або на поверхнях з слабо вираженою текстурою, оскільки вони розглядають всю інформацію зображення як неперервну функцію;
- **Обчислювальна складність:** прямі методи є більш обчислювально витратними порівняно з методами SLAM на основі ознак, оскільки вони працюють з усіма пікселями зображеннями, не вимагаючи відстеження особливих точок. Проте, обчислювальна складність залежить від розмірів зображень та потреби у великих об'ємах пам'яті;
- **Вартість обладнання:** прямі методи вимагають використання високоякісних камер, але вони можуть бути менш вимогливими до роздільної здатності і швидкості зчитування, оскільки працюють з усіма пікселями зображеннями і ціна таких камер досить помірною;
- **Ефективність в реальному часі:** прямі методи менш ефективні у реальному часі через необхідність обробки усіх пікселів зображення та велику обчислювальну складність;
- **Робочий діапазон:** прямі методи мають обмежену ефективність в умовах низької освітленості або на поверхнях без характерних ознак, оскільки вони працюють без використання особливих точок;
- **Відмовостійкість:** прямі методи менш відмовостійкі у порівнянні з іншими методами, оскільки вони частіше вимагають хорошого освітлення та добре видимих текстур на поверхні для ефективної роботи;
- **Необхідність калібрування:** прямі методи вимагають складного калібрування, оскільки потрібно налаштувати параметри для правильного врахування особливостей освітлення та оптичних спотворень, проте сучасні камери спрощують цей процес;
- **Сумісність з різними сенсорами:** прямі методи обмежені у сумісності з іншими сенсорами через їхню специфічність у використанні усіх пікселів зображення;
- **Енергоспоживання:** прямі методи енергоємні через необхідність обробки усіх пікселів зображення.

Візуально-інерційний SLAM

Візуально-інерційний SLAM окрім візуальних даних отриманих з камери включає також інерційні дані для оцінки положення датчика та реконструкції карти. Інерційні дані отримуються за допомогою використанням інерційного вимірювального пристрою (ІВП), який складається з набору приладів, а саме: гіроскопа, акселерометра та, додатково, магнітометра. Таким чином ІВП здатний надавати інформацію щодо кутової швидкості (гіроскоп) і прискорення (акселерометр) уздовж осей x , y та z , а також, крім того, щодо

магнітного поля навколо пристрою (магнітометр). Хоча додавання ІВП може збільшити інформаційну насиченість середовища та забезпечити більш високу точність, це також збільшує складність алгоритму, особливо на етапі ініціалізації. Це пов'язано з тим що, крім початкової оцінки позиції камери, алгоритм також має оцінити позицію ІВП. Алгоритми візуально-інерційного SLAM можна розділити відповідно до типу об'єднання даних камери та ІВП, які можуть бути слабо або тісно пов'язані. Методи слабого зв'язку не об'єднують дані ІВП для оцінки позиції: замість цього дані ІВП використовуються для оцінки орієнтації та змін у положенні датчика [18]. В свою чергу, тісно пов'язані методи базуються на об'єднанні даних камери та ІВП в рівняння руху, що призводить до оцінки стану, яка враховує дані обох пристроїв.

Крім того, алгоритми візуально-інерційного SLAM представляють різні реалізації відповідно до їх внутрішнього підходу, який може бути заснований на фільтрації або на основі оптимізації. Методи на основі фільтрації для візуально-інерційного SLAM покладаються на вилучення ознак, тоді як методи на основі оптимізації (також відомі як підходи на основі ключових кадрів) покладаються на глобальну оптимізацію, яка підвищує точність системи, а також обчислювальну вартість алгоритму. Результат роботи візуально-інерційного прямого методу SLAM [19] наведено на рисунку 6. Представлені зразки зображень і реконструйовані сцени в результаті руху камери по кімнаті. Орієнтовна траєкторія руху камери показана червоним кольором.

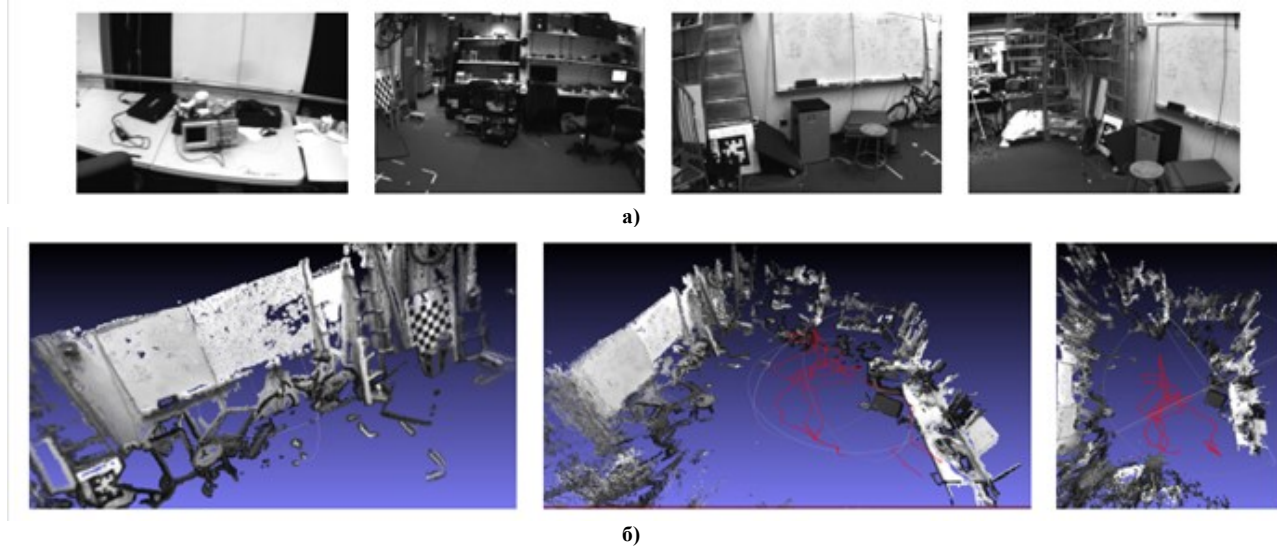


Рис. 6 Результат роботи алгоритму візуально-інерційного прямого методу SLAM: а) зразки зображень кімнати, по якій рухається камера б) реконструйовані сцени

Проведемо оцінку даного підходу, відповідно до критеріїв, зазначених раніше:

- **Точність:** комбінація візуальної інформації з даними інерційних сенсорів допомагає забезпечити високу точність, особливо під час руху і за умов низького рівня освітлення;
- **Стабільність:** комбінація візуальної інформації з даними інерційних сенсорів забезпечує високу стабільність, оскільки інерційні сенсори допомагають компенсувати рух камери та забезпечують стійке визначення місцеположення;
- **Обчислювальна складність:** комбінація візуальної інформації з даними інерційних сенсорів збільшує обчислювальну складність методу, оскільки потрібно обробляти дані з різних джерел та забезпечувати їх інтеграцію. Це вимагає додаткових обчислювальних ресурсів;
- **Вартість обладнання:** використання інерційних сенсорів разом з візуальними даними збільшує загальну вартість обладнання, але інерційні сенсори зазвичай є не дуже вартісними, і можуть бути доданими до системи з камерами за більш помірною ціною;
- **Ефективність в реальному часі:** комбінація візуальних та інерційних даних надає можливість забезпечити ефективну роботу в реальному часі, особливо під час руху;
- **Робочий діапазон:** комбінація візуальних та інерційних даних покращує ефективність системи як у відкритих, так і у закритих середовищах, але точність залежить від якості інерційних сенсорів;
- **Відмовостійкість:** комбінація візуальних та інерційних даних покращує відмовостійкість у випадку відмови візуального відстежування, оскільки інерційні сенсори допомагають компенсувати неякісні візуальні дані;
- **Необхідність калібрування:** калібрування систем, які поєднують візуальні та інерційні датчики, є складнішим, але правильне налаштування може покращити точність та стабільність системи;
- **Сумісність з різними сенсорами:** використання комбінації візуальних та інерційних сенсорів надає можливість для широкого спектру сумісності з іншими типами сенсорів;
- **Енергоспоживання:** комбінація візуальних та інерційних даних збільшує енерговитрати в порівнянні з іншими підходами, але ефективне управління енергією може зменшити це навантаження.

RGB-D SLAM

Системи SLAM, засновані на даних RGB-D, почали привертати більше уваги з появою Kinect від Microsoft у 2010 році. Датчики RGB-D складаються з монокулярної RGB-камери та датчика глибини, що дозволяє системам SLAM безпосередньо отримувати інформацію про глибину з достатньо високою точністю, отриманою в режимі реального часу за допомогою недорогого обладнання.

Більшість систем на основі RGB-D використовують ітераційний алгоритм найближчих точок для визначення місцезнаходження датчика, об'єднуючи карти глибини для отримання реконструкції загальної карти. Системи RGB-D мають такі переваги, як надання даних кольорового зображення та щільної карти глибини без будь-якого етапу попередньої обробки, що зменшує складність ініціалізації SLAM. Недоліками підходу є те, що він більш придатний для закритого середовища, і вимагає високого споживання пам'яті та енергії [20]. Приклад роботи алгоритму RGB-D SLAM [21] наведено на рисунку 7.

Проведемо оцінку даного підходу, відповідно до критеріїв, зазначених раніше:

- **Точність:** використання даних глибини допомагає збільшити точність визначення положення та створення карти, зокрема за рахунок розширення робочого діапазону;
- **Стабільність:** використання даних глибини допомагає підвищити стабільність методу, зокрема за рахунок розширення робочого діапазону;
- **Обчислювальна складність:** обчислювальна складність методу збільшується, оскільки потрібно обробляти не лише RGB-зображення, а й дані глибини. Це вимагає додаткового часу та ресурсів для обробки даних;
- **Вартість обладнання:** додавання камер глибини до системи збільшує загальну вартість обладнання, але надає додаткову інформацію, що дозволяє покращити точність і стабільність методу;
- **Ефективність в реальному часі:** використання даних глибини може полегшити і покращити процес відстеження ознак та оцінки відстані до об'єктів, що може допомогти забезпечити ефективну роботу в реальному часі;
- **Робочий діапазон:** використання даних глибини розширює робочий діапазон за рахунок підвищення ефективності роботи в умовах низької освітленості або на поверхнях без характерних ознак, але система все ще обмежена наявними даними глибини, зокрема у відкритих середовищах;
- **Відмовостійкість:** використання даних глибини покращує відмовостійкість, за рахунок виявлення та уникнення проблем, таких як неправильна оцінка глибини або виявлення фальшивих ознак;
- **Необхідність калібрування:** використання даних глибини спрощує калібрування, оскільки допомагає визначити відстані до об'єктів та уникнути проблем, пов'язаних з освітленням та спотвореннями;
- **Сумісність з різними сенсорами:** використання даних глибини забезпечує додаткові можливості для інтеграції з іншими сенсорами, такими як RGB камери або інерційні сенсори;
- **Електроспоживання:** використання даних глибини збільшує енергозатрати, але деякі реалізації можуть оптимізувати цей процес.

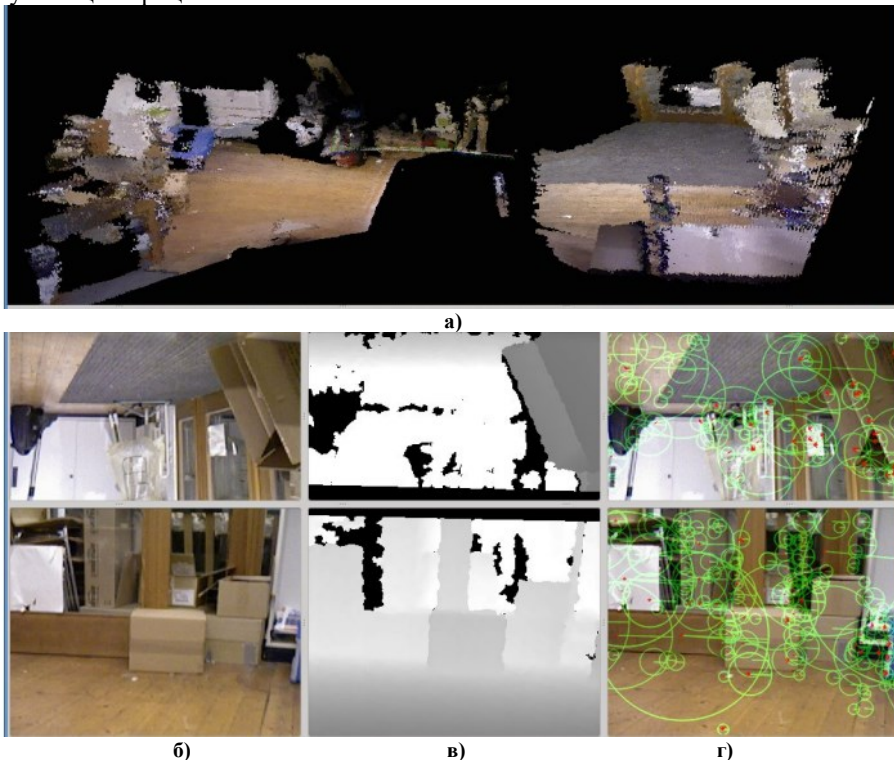


Рис. 7. Приклад роботи алгоритму RGB-D SLAM, що оброблює вхідні дані від катадіоптричного датчика: а) 3D-модель, створена з двох зображень за допомогою оціненого калібрування; б) вхідні кольорові зображення; в) вхідні зображення глибини; г) визначені візуальні ознаки

В якості оцінювання буде використовуватись 5-ти бальна система, де 5 – це найкращий результат, та відповідно 1 – найгірший.

Таблиця 1

Результати аналізу підходів Visual SLAM за набором критеріїв

Критерії	Методи SLAM, на основі ознак	Прямі методи SLAM	Візуально-інерційний SLAM	RGB-D SLAM
Точність	3	3	5	4
Стабільність	3	3	5	4
Обчислювальна складність	3	2	4	3
Вартість обладнання	4	5	3	3
Ефективність в реальному часі	5	3	5	4
Робочий діапазон	3	3	5	4
Відмовостійкість	2	3	5	4
Необхідність калібрування	5	2	3	4
Сумісність з різними сенсорами	3	3	5	5
Електроспоживання	4	2	3	3
Середнє значення ефективності	3,5	3,2	4,3	3,8

За результатами, найкращий середній показник показав підхід з використанням візуально-інерційного SLAM. Проте важливо враховувати, що вибір методу Visual SLAM повинен залежати від конкретних потреб та обмежень задачі, таких як доступність обладнання, середовище роботи, швидкість роботи та точність, які необхідно досягти. Так наприклад, візуально-інерційний SLAM та RGB-D SLAM показують високі показники точності та надійності, проте мають вищу ціну обладнання, є більш енергозатратними та вимагають вищі обчислювальні потужності. Тоді як методи на основі ознак та прямі методи мають хоч і дещо програють у точності та стабільності, проте є більш бюджетними та доступними, що робить реалізацію методів на основі цих підходів значно доступнішою та простішою.

Висновки

У даній статті представлено формулювання задачі SLAM у вигляді факторного графа та розглянуто її вирішення за допомогою оптимальної оцінки максимуму апостеріорної імовірності. Проведено аналіз доцільності та практичної ефективності основних підходів Visual SLAM для навігації автономного робота як в закритому так і у відкритому середовищі. Розглянуто такі підходи Visual SLAM як: тільки візуальний SLAM, у якому у якості сенсора використовується лише камера; візуально-інерційний SLAM, в якому крім камери додатково використовуються інерційні вимірювальні пристрої (гіроскоп, акселерометр та магнітометр); та RGB-D SLAM, який використовує RGB-D камери, що окрім візуальної інформації надають інформацію про глибину.

В роботі запропоновано десять критеріїв оцінки ефективності кожного з підходів, які дозволяють в повній мірі зрозуміти сильні та слабкі сторони підходів що розглядаються, та умови, при яких їх використання буде доцільним. Здійснено оцінку кожного підходу по кожному з критеріїв та в результаті отримано середній показник ефективності кожного підходу. За результатами експериментів було визначено підхід з найкращим середнім показником, а саме візуально-інерцій SLAM, який має найкращі показники точності та стабільності роботи, проте його сенсорна система дорожча у порівнянні з іншими підходами. З іншого боку, методи, на основі ознак та прямі методи хоч показали дещо гіршу середню оцінку, проте мають досить збалансовані показники за кожним критерієм та є більш простими та доступними для реалізації.

Література

1. Tin Lai “A Review on Visual-SLAM: Advancements from Geometric Modelling to Learning-Based Semantic Scene Understanding Using Multi-Modal Sensor Fusion” School of Computer Science, The University of Sydney, Camperdown, NSW 2006, Australia, Sensors 2022, 22(19), 7265; <https://doi.org/10.3390/s22197265>
2. H. J. S. Feder, J. J. Leonard, and C. M. Smith, “Adaptive mobile robot navigation and mapping,” The International Journal of Robotics Research, vol. 18, no. 7, [A Review on Visual-SLAM: Advancements from Geometric Modelling to Learning-Based Semantic Scene Understanding Using Multi-Modal Sensor Fusion](https://doi.org/10.1177/02783649922066484) pp. 650–668, 1999. <https://doi.org/10.1177/02783649922066484>
3. C. Kolhatkar and K. Wagle, “Review of SLAM algorithms for indoor mobile robot with LIDAR and RGB-D camera technology,” Innovations in electrical and electronic engineering, pp. 397–409, 2021.

https://www.researchgate.net/publication/343223545_Review_of_SLAM_Algorithms_for_Indoor_Mobile_Robot_with_LIDAR_and_RGB-D_Camera_Technology

4. T. Lai and F. Ramos, "Adaptively Exploits Local Structure With Generalised Multi-Trees Motion Planning," IEEE Robotics and Automation Letters, vol. 7, no. 2, pp. 1111–1117, 2021. <https://arxiv.org/pdf/2207.08130.pdf>

5. C. Katrakazas, M. Quddus, W.-H. Chen, and L. Deka, "Real-time motion planning methods for autonomous on-road driving: State-of-the-art and future research directions," Transportation Research Part C: Emerging Technologies, vol. 60, pp. 416–442, 2015. https://www.researchgate.net/publication/283782140_Real-time_motion_planning_methods_for_autonomous_on-road_driving_State-of-the-art_and_future_research_directions

6. T. Chong, X. Tang, C. Leng, M. Yogeswaran, O. Ng, and Y. Chong, "Sensor Technologies and Simultaneous Localization and Mapping (SLAM)," Procedia Computer Science, vol. 76, pp. 174–179, 2015. <https://doi.org/10.3390/rs15041156>

7. H. Bavle, J. L. Sanchez-Lopez, E. F. Schmidt, and H. Voos, "From SLAM to Situational Awareness: Challenges and Survey," arXiv preprint arXiv:2110.00273, 2021 <https://www.mdpi.com/1424-8220/23/10/4849>

8. Macario Barros, A.; Michel, M.; Moline, Y.; Corre, G.; Carrel, F. A Comprehensive Survey of Visual SLAM Algorithms. Robotics 2022, 11, 24. https://www.researchgate.net/deref/https%3A%2F%2Fdoi.org%2F10.3390%2Frobotics11010024?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmV2Y2F0aW9uIn19

9. R. Chatila and J.-P. Laumond, "Position referencing and consistent world modeling for mobile robots," in Proceedings. 1985 IEEE International Conference on Robotics and Automation, vol. 2, IEEE, 1985, pp. 138–145. https://www.researchgate.net/publication/3979283_Position_referencing_and_consistent_world_modeling_for_mobile_robots

10. F. Dellaert and M. Kaess, "Factor graphs for robot perception," Foundations and Trends® in Robotics, vol. 6, no. 1-2, pp. 1–139, 2017. <https://www.cs.cmu.edu/~kaess/pub/Dellaert17fnt.pdf>

11. Taketomi, T.; Uchiyama, H.; Ikeda, S. Visual SLAM algorithms: A survey from 2010 to 2016. IPSJ Trans. Comput. Vis. Appl. 2017, 9, 1–11. <https://link.springer.com/article/10.1186/s41074-017-0027-2>

12. Lepetit, V.; Moreno-Noguer, F.; Fua, P. EPnP: An Accurate O(n) Solution to the PnP Problem. Int. J. Comput. Vis. 2008, 81, 155. <https://link.springer.com/article/10.1007/s11263-008-0152-6>

13. Davison, A.J.; Reid, I.D.; Molton, N.D.; Stasse, O. MonoSLAM: Real-Time Single Camera SLAM. IEEE Trans. Pattern Anal. Mach. Intell. 2007, 29, 1052–1067. <https://ieeexplore.ieee.org/document/4160954>

14. Azzam, R., Taha, T., Huang, S. *et al.* Feature-based visual simultaneous localization and mapping: a survey. *SN Appl. Sci.* 2, 224 (2020). <https://doi.org/10.1007/s42452-020-2001-3>

15. Nicola Krombach, David Droschel, Sebastian Houben, Sven Behnke "Feature-based visual odometry prior for real-time semi-dense stereo SLAM", August 2018 Robotics and Autonomous Systems 109, DOI: <https://arxiv.org/pdf/1810.07768v1.pdf#page=40>

16. Boikos, K.; Bouganis, C.S. Semi-dense SLAM on an FPGA SoC. In Proceedings of the 2016 26th International Conference on Field Programmable Logic and Applications (FPL), Lausanne, Switzerland, 29 August–2 September 2016; pp. 1–4. <https://ieeexplore.ieee.org/document/7577365>

17. Engel, J., Schöps, T., Cremers, D. (2014). LSD-SLAM: Large-Scale Direct Monocular SLAM. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8690. Springer, Cham. https://doi.org/10.1007/978-3-319-10605-2_54

18. Chen, C.; Zhu, H.; Li, M.; You, S. A Review of Visual-Inertial Simultaneous Localization and Mapping from Filtering-Based and Optimization-Based Perspectives. Robotics 2018, 7, 45 <https://www.mdpi.com/2218-6581/7/3/45>

19. A. Concha, G. Loianno, V. Kumar and J. Civera, "Visual-inertial direct SLAM," 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 2016, pp. 1331-1338, http://webdiis.unizar.es/~jcivera/papers/concha_etal_icra16.pdf

20. Canovas, B.; Rombaut, M.; Nègre, A.; Pellerin, D.; Olympieff, S. Speed and Memory Efficient Dense RGB-D SLAM in Dynamic Scenes. In Proceedings of the IROS 2020—IEEE/RSJ International Conference on Intelligent Robots and Systems, Las Vegas, NV, USA, 25–29 October 2020; pp. 4996–5001. <http://ras.papercept.net/images/temp/IROS/files/0404.pdf>

21. Felix Endres "Robot Perception for Indoor Navigation" Dissertation zur Erlangung des akademischen Grades Doktor der Naturwissenschaften Technische Fakultät, Albert-Ludwigs-Universität " at Freiburg, 4. August 2015, <http://www2.informatik.uni-freiburg.de/~endres/files/publications/felix-endres-phd-thesis.pdf>