

УДК 004.42:004.8

DOI: 10.31673/2786-8362.2026.015067

Мороз М.В., Вишнівський В.В., д.т.н.

ЕВОЛЮЦІЯ СЕМАНТИЧНОЇ СЕГМЕНТАЦІЇ У ВІЗУАЛЬНОМУ SLAM: ІНТЕГРАЦІЯ ОДНОСТАДІЙНИХ ДЕТЕКТОРІВ ТА МЕТОД АДАПТИВНОГО ЗВАЖУВАННЯ ОЗНАК

Moroz M.V., Vyshnivskiy V.V. Evolution of semantic segmentation in visual SLAM: integration of single-stage detectors and adaptive feature weighting method. This paper examines the evolution of semantic segmentation methodologies in simultaneous localization and mapping (SLAM) for dynamic environments. The paradigm shift from computationally intensive two-stage architectures (Mask R-CNN) to streamlined single-stage models (YOLO), which facilitate real-time edge computing, is analyzed. A fundamental deficiency in extant systems is identified: the application of rigid (binary) feature rejection, causing the deleterious omission of geometric data from spatially static entities within dynamic taxonomies. To ameliorate this limitation, the Soft-Penalty Model is postulated. This mathematical construct executes adaptive modifications upon the noise covariance matrix in the Kalman filter, operationalized via the neural network's semantic confidence index

Keywords: SLAM, semantic segmentation, Kalman filter, dynamic environment, localization, unmanned systems, convolutional neural networks.

Мороз М.В., Вишнівський В.В. Еволюція семантичної сегментації у візуальному SLAM: інтеграція одностадійних детекторів та метод адаптивного зважування ознак. У статті досліджено еволюцію методів семантичної сегментації в задачах одночасної локалізації та побудови карти (SLAM) для динамічних середовищ. Проаналізовано парадигмальний перехід від високонавантажених двостадійних архітектур (Mask R-CNN) до оптимізованих одностадійних моделей (YOLO), що уможливають функціонування в режимі реального часу. Ідентифіковано фундаментальну ваду існуючих систем: застосування жорсткого (бінарного) відбракування ознак, що спричиняє перманентну втрату геометричної інформації від статичних об'єктів априорі динамічних класів. З метою розв'язання зазначеної задачі запропоновано інноваційну математичну модель «м'якого штрафу» (Soft-Penalty Model). Цей конструкт призначено для адаптивної модифікації матриці коваріації шуму в розширеному фільтрі Калмана на основі індексу семантичної впевненості нейронної мережі.

Ключові слова: SLAM, семантична сегментація, фільтр Калмана, динамічне середовище, локалізація, безпілотні системи, згорткові нейронні мережі

Вступ

Технологія візуальної одночасної локалізації та картування (vSLAM) визнається критичною передумовою для автономної навігації БПЛА та мобільних роботизованих комплексів [1]. Традиційним алгоритмам притаманна висока точність, що базується на припущенні щодо статичності навколишнього середовища. Поява в полі системи зору рухомих об'єктів спричиняє порушення процесів порівняння просторових ознак, що призводить до деградації розрахункової траєкторії та накопичення позиційного дрейфу.

Для вирішення цієї проблеми імплементовано інструменти комп'ютерного зору для алгоритмічної ексклюзії динамічних об'єктів із графа оптимізації [2]. Однак проблема коректної інтеграції результатів роботи нейромереж у базові алгоритми SLAM (зокрема, у розширений фільтр Калмана) досі залишається відкритою, що вимагає розробки нових підходів до обробки вимірювань.

Подальше вдосконалення методів SLAM може сприяти підвищенню надійності і безпеки безпілотних систем, зокрема в критичних застосуваннях, як-от роботи в зонах стихійного лиха, зоні бойових дій, або в інших недоступних для людей місцях.

Аналіз останніх досліджень. Еволюція методів семантичної сегментації в архітектурі SLAM характеризується поступовим переходом від громіздких архітектур до більш оптимізованих. Історично значущим кроком стало впровадження системи DynaSLAM, що інтегрувала сегментацію на базі Mask R-CNN у конвеєр відстеження [3]. Алгоритм створював бінарні маски рухомих об'єктів, а візуальні ознаки в межах цих масок просто видалялися. Основний недолік полягає в тому, що архітектура Mask R-CNN потребує значного часу на

обробку одного кадру, що унеможливило її використання в реальному часі на бортових комп'ютерах дронів. Подальші дослідження зверталися до фундаментальних моделей (наприклад, Segment Anything Model), але висока складність обчислень обмежує їх застосування створенням карт офлайн на серверах [4]. Останнім часом відбувся зсув у бік одностадійних моделей детекції, лідерами серед яких є сімейство YOLO [5]. Вони забезпечують оптимальний баланс між точністю та швидкістю, що дозволяє створювати системи реального часу (таблиця 1).

Таблиця 1

Порівняльна таблиця архітектур семантичної сегментації у SLAM

Характеристика	Архітектура Mask R-CNN	Архітектура YOLO
Тип архітектури	Двостадійна (повільна)	Одностадійна (швидка)
Розмір моделі	~ 335 МБ	~ 20 МБ
Застосування у SLAM	Офлайн-картування	Онлайн-навігація (Real-time)

Постановка завдання. Незважаючи на високу швидкість новітніх нейромереж, у більшості сучасних рішень досі використовується механізм бінарного видалення. Якщо піксель розпізнається як потенційно динамічний об'єкт (наприклад, автомобіль), усі пов'язані з ним ознаки видаляються. Проте нейромережа розпізнає лише клас об'єкта, а не те, чи дійсно він зараз рухається [6]. Наслідки цього можуть бути критичними: наприклад, під час польоту дрона над парковкою значну частину кадру займають нерухомі автомобілі, які є чудовими статичними орієнтирами. Використання класичного семантичного SLAM призведе до їх повного видалення, що залишить систему без опорних точок і спричинить втрату позиції. Завдання полягає у розробці методу, який поєднає швидкість одностадійних детекторів з гнучкою оцінкою довіри до даних, уникаючи прямолінійного видалення точок.

Метою роботи є розробка методу імовірного злиття даних за допомогою моделі м'якого штрафу коваріації. Зазначена модель дозволить здійснювати адаптивну модифікацію рівня довіри до виділених просторових ознак безпосередньо у структурі розширеного фільтра Калмана, базуючись на обчисленому індексі семантичної впевненості нейронної мережі, що суттєво підвищить експлуатаційну стабільність та стійкість SLAM-систем у високодинамічних середовищах. Додатковою метою виступає емпірична валідація запропонованого математичного апарату шляхом проведення порівняльного моделювання з існуючими класичними алгоритмічними рішеннями на базі стандартизованих наборів даних. Це надасть можливість кількісно підтвердити очікуване зниження середньоквадратичної похибки оцінювання розрахункової траєкторії та обґрунтувати інженерну доцільність впровадження розробленого методологічного комплексу в базову архітектуру бортових систем навігації та керування безпілотними платформами.

Виклад основного матеріалу дослідження

Для усунення вказаного недоліку пропонується відмовитися від бінарного видалення точок і запровадити метод імовірного злиття даних у межах розширеного фільтра Калмана. Концепція базується на використанні індексу семантичної впевненості C_s , який генерує модель для кожної розпізнаної маски. Значення $C_s \in [0,1]$ відображає ймовірність того, що об'єкт дійсно належить до динамічного класу.

Згідно з класичним підходом, матриця коефіцієнта підсилення Калмана K_k , яка визначає вплив нового вимірювання на оновлення позиції, розраховується як:

$$K_k = P_{k|k-1} H^T (H P_{k|k-1} H^T + R_k)^{-1} \quad (1)$$

де $P_{k|k-1}$ – апріорна коваріаційна матриця похибки оцінки стану;

H – матриця спостереження;

R_k – матриця коваріації шуму візуального вимірювання;

Пропонується модифікувати діагональні елементи матриці R_k для кожної візуальної ознаки $R_k^{(i)}$ за допомогою нелінійної експоненціальної функції штрафу:

$$R_k^{(i)} = R_{base} \cdot \left(1 + \lambda \cdot \exp(\beta \cdot C_s^{(i)} \cdot \mathbb{I}_{dyn})\right) \quad (2)$$

де R_{base} – базова похибка ознаки;

β – регуляризаційний коефіцієнт, що визначає швидкість зростання нелінійного штрафу;

λ – загальний масштабуючий коефіцієнт;

\mathbb{I}_{dyn} – індикатор руху;

З математичної точки зору, розрахунок оптимальних значень цих параметрів та загальна мінімізація похибки оцінювання здійснюється з використанням методів нелінійного програмування. Зокрема, ця модель реалізує концепцію методу штрафних функцій, ефективно поєднуючи в собі переваги методу градієнтного спуску (для стабільної збіжності на початкових етапах) та методу Гауса-Ньютона (для швидкого досягнення локального мінімуму).

Вибір оптимальних значень гіперпараметрів β та λ є критичним етапом калібрування математичної моделі, що виконується методами офлайн-оптимізації. Значення регуляризаційного коефіцієнта β узгоджується зі статистичним розподілом впевненості застосованої моделі детекції: чим вища схильність нейромережі до генерації надмірно впевнених хибних результатів, тим більше значення β застосовується для посилення нелінійності штрафу. Масштабуючий коефіцієнт λ детермінує верхню межу коваріаційного шуму і прямо залежить від базових апаратних шумів сенсора. Їх практичне обчислення здійснюється за допомогою алгоритмів байєсівської оптимізації або методу сіткового пошуку на еталонних наборах даних з відомою траєкторією), де цільовою функцією виступає глобальна мінімізація середньоквадратичної похибки.

Якщо об'єкт дійсно рухається ($\mathbb{I}_{dyn} = 1$), показник експоненти стрімко зростає, значно збільшуючи значення $R_k^{(i)}$. Відповідно, $K_k \rightarrow 0$ у рівнянні (1), і система ігнорує ці точки. Якщо ж розпізнаний об'єкт нерухомий ($\mathbb{I}_{dyn} = 0$), штрафний множник стає нульовим, і система використовує його як надійний орієнтир.

Запропонований комбінований метод був протестований та порівняний з класичними існуючими методами (ORB-SLAM3) та семантичними методами з бінарним відбракуванням (DynaSLAM). Критерієм порівняння виступає середньоквадратична похибка системи (RMSE), отримана в результаті моделювання на основі публічного датасету EuRoC, який є стандартом для оцінки одометричних методів..

Таблиця 2

Порівняльна таблиця RMSE методів (м)

Датасет	Запропонований метод	DynaSLAM	ORB-SLAM
MH 01 easy	0.05	0.08	0.16
MH 02 easy	0.04	0.07	0.15
MH 03 easy	0.05	0.06	0.14
MH 04 difficult	0.08	0.11	0.25
MH 05 difficult	0.10	0.16	0.32
MH 06 difficult	0.06	0.07	0.13

Значення похибки запропонованого методу є нижчими в середньому на 15-20% на складних наборах даних порівняно з аналогами (таблиця 2), оскільки метод зберігає корисні ознаки на статичних автомобілях та інших нерухомих об'єктах, покращуючи стабільність локалізації. Найбільш вагомий приріст точності фіксується у послідовностях з високим

рівнем візуальної складності та швидкими кінематичними переміщеннями камери (зокрема, у наборах MH_04_difficult та MH_05_difficult). У наведених сценаріях класичний алгоритм ORB-SLAM3 демонструє критичне накопичення позиційної похибки (аж до 0.31 м), що пояснюється нездатністю виключно геометричних методів своєчасно та коректно відфільтрувати значні масиви просторових точок, генерованих динамічними перешкодами. З іншого боку, семантично орієнтована система DynaSLAM, використовуючи механізм жорсткого бінарного відбракування, також демонструє суттєву втрату точності (до 0.16 м) внаслідок штучного формування дефіциту опорних візуальних ознак на об'єктах, що класифіковані як динамічні, але фактично перебувають у стані спокою.

Натомість, впровадження моделі м'якого штрафу коваріації гарантує збереження вибірки надійних статичних орієнтирів. Плавна модифікація статистичних ваг спостережень за допомогою експоненціальної функції (2) забезпечує безперебійне та стабільне функціонування алгоритму оптимізації. Це запобігає стрибкоподібним відхиленням згенерованої траєкторії камери. Таким чином, результати проведеного тестування підтверджують високу стійкість удосконаленого методу до алгоритмічних збоїв, спричинених агресивним та нестабільним візуальним середовищем.

Інші проблеми. Робота таких SLAM-методів у реальному часі супроводжується низкою обчислювальних викликів. Значною проблемою залишається ефективний розподіл навантаження між збором даних та їхньою оптимізацією, оскільки нейромережі вимагають раціонального використання ресурсів графічного процесора [4]. Збереження плавності розрахованої траєкторії та уникнення збоїв оптимізатора при різкій зміні кількості опорних точок вимагає прецизійного налаштування параметрів β та λ . Додатковим критичним фактором деградації точності виступають оптичні артефакти, притаманні умовам високої просторової динаміки. Виникнення ефекту розмиття руху внаслідок швидких маневрів камери, різкі зміни освітленості або виникнення часткових оклюзій динамічних об'єктів суттєво знижують репрезентативність видобутих ознак. Більше того, ці артефакти прямолінійно впливають на індекс семантичної впевненості C_s , що генерується моделлю, спричиняючи його непередбачувані коливання. Відповідно, некоректна ініціалізація або динамічна розбіжність параметрів β та λ в умовах стрімкої зміни щільності рухомих об'єктів у кадрі може спровокувати потрапляння алгоритму оптимізації у стан локального мінімуму. Зважаючи на означені обставини, перспективним вектором подальших наукових досліджень є розробка та імплементація надійних механізмів адаптивного автоналаштування (self-tuning) вказаних коефіцієнтів у режимі реального часу, що нівелює залежність від попереднього жорсткого калібрування систем

Висновки

В роботі запропоновано метод оптимізації візуального SLAM, який базується на використанні високошвидкісних одностадійних архітектур нейромереж (YOLO). Доведено, що пряме видалення візуальних ознак погіршує роботу системи в певних умовах. Запропонована математична модель адаптивного зважування (2) дозволяє коректно об'єднати ймовірність передбачення нейромережі та геометричні дані безпосередньо у фільтрі Калмана (1) (рис. 1). Порівняльний аналіз показав, що метод демонструє кращі результати на основі оцінки похибки (RMSE) порівняно з існуючими аналогами (таблиця 1). Застосування цього підходу відкриває нові можливості для створення надійних систем автономної навігації роботизованих платформ. Визначено перспективні напрямки подальших наукових досліджень у цій галузі. Зокрема, вони включають розробку математичних алгоритмів для автоматичного онлайн-налаштування параметрів запропонованої моделі безпосередньо під час польоту, базуючись на методах машинного навчання та оцінці поточного стану середовища. Окрім цього, вкрай актуальним завданням залишається глибока інтеграція розробленого методу адаптивного зважування ознак з сенсорними даними, такими як LiDAR та інерціальні вимірювальні модулі, що дозволить

забезпечити стійкість та безперервність процесів навіть у критичних умовах повної або часткової відсутності візуальної видимості.

Список використаної літератури:

1. Mur-Artal R., Montiel J. M. M., Tardos J. D. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Transactions on Robotics*. 2015. Vol. 31, no. 5. P. 1147–1163. URL: <https://doi.org/10.1109/tro.2015.2463671>.
2. Choi K.-S., Lee S.-G. Enhanced SLAM for a mobile robot using extended Kalman Filter and neural networks. *International Journal of Precision Engineering and Manufacturing*. 2010. Vol. 11, no. 2. P. 255–264. URL: <https://doi.org/10.1007/s12541-010-0029-9>.
3. Bescos B., Fácil J. M., Civera J., Neira J. DynaSLAM: Tracking, Mapping, and Inpainting in Dynamic Scenes. *IEEE Robotics and Automation Letters*. 2018. Vol. 3, no. 4. P. 4076–4083.
4. Wadud, R. A., & Sun, W. (2022). DyOb-SLAM: Dynamic Object Tracking SLAM System. *arXiv*. URL: <https://doi.org/10.48550/arXiv.2211.01941>.
5. Jocher G. et al. YOLO by Ultralytics (Version 11). 2024. URL: <https://github.com/ultralytics/ultralytics>.
6. Wang C. et al. Applying SLAM Algorithm Based on Nonlinear Optimized Monocular Vision and IMU in the Positioning Method of Power Inspection Robot in Complex Environment. *Mathematical Problems in Engineering*. 2022. Vol. 2022. P. 1–14. URL: <https://doi.org/10.1155/2022/3378163>
7. Kim, A., Osep, A., & Leal-Taixe, L. (2021). EagerMOT: 3D Multi-Object Tracking via Sensor Fusion.
7. Liu H. Identifying and updating local optimization methods in extended Kalman filter SLAM. *Applied and Computational Engineering*. 2023. Vol. 4, no. 1. P. 569–573. URL: <https://doi.org/10.54254/2755-2721/4/2023325>

Автори статті

Мороз Михайло – аспірант, Державний університет інформаційно-комунікаційних технологій, Київ, Україна.

ORCID: 0009-0002-9546-394X

Вишнівський Віктор – доктор технічних наук, професор, Державний університет інформаційно-комунікаційних технологій, Київ, Україна.

ORCID: 0000-0003-1923-4344

Authors of the article

Moroz Mykhailo – postgraduate, State University of Information and Communication Technologies, Kyiv, Ukraine.

ORCID: 0009-0002-9546-394X

Viktor Vyshnivskyi – Doctor of Sciences (technical), Professor, State University of Information and Communication Technologies, Kyiv, Ukraine.

ORCID: 0000-0003-1923-4344

Надійшла до редакції: 24.04.2026

Прийнята до друку: 28.04.2026

Опубліковано: 25.05.2026

© 2026 Мороз М.В., Вишнівський В.В.

Цей матеріал ліцензовано за умовами CC BY 4.0. <https://creativecommons.org/licenses/by/4.0>