

Меркулова Катерина Володимирівна

к.т.н., доцент, доцент кафедри програмних систем і технологій
Київський Національний Університет імені Тараса Шевченка, Київ,
Україна ORCID: 0000-0001-6347-5191 k.merkulova@knu.ua

Ярошенко Андрій Олексійович аспірант

кафедри програмних систем і технологій
Київський Національний Університет імені Тараса Шевченка, Київ,
Україна
ORCID: 0009-0002-1598-8261
andrii.yaroshenko@knu.ua

**МЕТОД ВИЯВЛЕННЯ ТА СУПРОВОДЖЕННЯ МАЛИХ ПОВІТРЯНИХ ЦІЛЕЙ НА ОСНОВІ
СТАБІЛІЗОВАНОЇ МІЖКАДРОВОЇ РІЗНИЦІ З АДАПТИВНОЮ ЛОКАЛЬНОЮ ПОРОГОВОЮ
СЕГМЕНТАЦІЄЮ**

Анотація. У статті розв'язується актуальна науково-практична задача виявлення та муьлтитрекового супроводження малих повітряних цілей (безпілотних літальних апаратів, БПЛА) у відеопотоці, що отримується з бортової камери іншого спостережного апарата. Впровадження технологій комп'ютерного зору на борту безпілотників суттєво ускладнюється жорсткими обмеженнями щодо масо-габаритних та енергетичних характеристик (SWaP), що робить застосування сучасних ресурсномістких глибоких нейромереж (таких як YOLO або Faster R-CNN) вкрай складним без спеціалізованих графічних прискорювачів. Додатковими ускладнювальними факторами під час польоту є екстремально малі розміри цілей (часто менше ніж 15 пікселів), низький контраст об'єктів на тлі неба чи землі, а також постійний рух самої камери (ego-motion), що створює глобальне зміщення фону.

Метою дослідження є розробка ефективного та обчислювально легкого алгоритмічного конвеєра виявлення малих рухомих цілей, придатного для розгортання на малопотужних бортових Edge-системах у режимі реального часу на базових процесорах (CPU). Для досягнення цієї мети запропоновано модифікований підхід на основі класичних методів аналізу міжкадрової різниці. Запропонований метод послідовно інтегрує в єдину систему афінну стабілізацію зображення за алгоритмом Лукаса-Канаде з використанням RANSAC, адаптивну локальну порогову сегментацію на основі локального середнього та стандартного відхилення, фонову модель експоненційного ковзного середнього (EMA) для згладжування повільних змін освітленості, а також інноваційну концепцію безперервного темпорального голосування з плавним згасанням для фільтрації випадкового шуму. Використання локальної статистики дозволяє відмовитися від неефективних глобальних порогів, надійно зберігаючи чутливість системи до дрібних цілей і водночас пригнічуючи хибні спрацювання у висококонтрастних зонах.

Експериментальну перевірку розробленого методу здійснено на складному датасеті, що налічує 15 різноманітних відеокліпів загальною тривалістю понад 14 000 кадрів, анотованих у стандартному форматі MOT 1.1. Результати системного тестування продемонстрували, що метод забезпечує високу середню швидкість обробки на рівні 38,5 кадрів за секунду, що із запасом задовольняє вимоги реального часу (понад 25 кадрів/с). Середнє значення метрики F1 для базової конфігурації алгоритму становить 0,304. Проведений детальний пооб'єктний аналіз виявив, що якість виявлення критично залежить від розміру мішені та швидкості її руху: повнота виявлення (Recall) стрімко знижується для об'єктів розміром менше 15 пікселів та при швидкостях, що перевищують 10 пікселів за кадр. Дослідження переконливо доводить високий потенціал запропонованого алгоритму для автономних бортових обчислювачів, оскільки він гарантує належну ефективність виявлення без залучення енергозатратних графічних процесорів (GPU), залишаючи обчислювальні ресурси для виконання критичних польотних завдань.

Ключові слова: безпілотний літальний апарат, виявлення малих рухомих цілей, стабілізація зображення, міжкадрова різниця, адаптивна порогова сегментація, темпоральне голосування, муьлтитрекове супроводження.

Kateryna Merkulova

PhD in Engineering Sciences, Associate Professor, Lecturer
Taras Shevchenko National University of Kyiv, Kyiv,
Ukraine ORCID: 0000-0001-6347-5191
k.merkulova@knu.ua

Andrii Yaroshenko

1st-year PhD student
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
ORCID: 0009-0002-1598-8261
andrii.yaroshenko@knu.ua

METHOD FOR DETECTION AND TRACKING OF SMALL AERIAL TARGETS BASED ON STABILIZED INTER-FRAME DIFFERENCE WITH ADAPTIVE LOCAL THRESHOLD SEGMENTATION

Abstract. The article addresses the highly relevant scientific and practical problem of detecting and multi-track tracking of small aerial targets (Unmanned Aerial Vehicles, UAVs) in a video stream captured continuously from the onboard camera of another observing UAV. The broad integration of computer vision technologies onboard drones is severely hindered by strict Size, Weight, and Power (SWaP) constraints. This structural limitation makes the deployment of modern, resource-intensive deep neural networks (such as YOLO or Faster R-CNN) virtually impossible without specialized graphics processing units (GPUs). Additional complicating factors in real-world environments include the extremely small sizes of the targets (often less than 15 pixels), the inherently low contrast of objects against the sky or ground background, and the continuous movement of the camera itself (ego-motion), which inevitably creates a global background shift and motion artifacts.

The primary aim of this research is to develop an efficient, computationally lightweight detection pipeline for small moving targets, strictly suitable for real-time deployment on low-power onboard Edge systems running on standard central processing units (CPUs). To achieve this ambitious goal, a modified computer vision methodology based on inter-frame difference is proposed. The newly developed method seamlessly integrates Lucas-Kanade affine image stabilization utilizing RANSAC, adaptive local threshold segmentation based on local mean and standard deviation matrices, an Exponential Moving Average (EMA) background model to gracefully smooth out slow lighting changes, and a continuous temporal voting concept featuring smooth exponential decay to aggressively filter out random noise and artifacts. The strategic use of local spatial statistics allows the complete abandonment of fixed global thresholds, thereby maintaining a high sensitivity to extremely small targets while effectively suppressing false positives (FP) in high-contrast or noisy areas.

The comprehensive experimental validation of the developed method was conducted on a challenging dataset consisting of 15 video clips with a total duration of over 14,000 frames, thoroughly annotated in the standardized MOT 1.1 format. The extensive testing results demonstrated that the proposed method provides an excellent average processing speed of 38.5 frames per second, which comfortably satisfies real-time operational requirements (exceeding the baseline of 25 FPS). The overall average F1-score metric reached 0.304 across the dataset. Furthermore, a detailed per-object analysis conclusively revealed that detection quality heavily correlates with the target's pixel size and motion characteristics: detection completeness (Recall) critically plummets for objects smaller than 15 pixels and at relative speeds exceeding 10 pixels per frame. The study solidifies the high potential of the proposed algorithmic pipeline for autonomous onboard drone computers, as it operates efficiently and robustly without the need for hardware graphics acceleration, effectively reserving critical power and computational resources for core flight management tasks.

Keywords: unmanned aerial vehicle, small moving target detection, image stabilization, inter-frame difference, adaptive threshold segmentation, temporal voting, multi-track tracking.

1. Вступ

Технологічний розвиток БПЛА зумовив їх широке впровадження в інженерному менеджменті, моніторингу інфраструктури та обороні. Однією з ключових задач є взаємне спостереження: виявлення одного БПЛА з борту іншого. У реальних умовах рух спостерігача створює глобальне відхилення фону, а ціль має малі розміри та низький контраст відносно неба чи землі.

Сучасні дослідження фокусуються на двох напрямках: глибинні нейромережеві методи (YOLO, Faster R-CNN, SSD) та класичні методи комп'ютерного зору на основі руху та оптичного потоку. Глибинні методи потребують значних обчислювальних ресурсів і розмічених датасетів, що робить їх розгортання на малогабаритних бортових обчислювачах (Edge-системах) вкрай складним. Сучасні БПЛА мають жорсткі обмеження щодо масо-габаритних та енергетичних характеристик (SWaP – Size, Weight, and Power). Концепція Edge AI вимагає перенесення обчислень безпосередньо на борт (наприклад, на одноплатні комп'ютери архітектури ARM) без залучення хмарних ресурсів, що виключає використання енергозатратних графічних прискорювачів (GPU). У цьому контексті класичні методи (міжкадрова різниця, оптичний потік) залишаються безальтернативними завдяки своїй легкості та здатності працювати на базових процесорах (CPU), проте вони чутливі до тремтіння камери та артефактів компенсації руху. Традиційні схеми з глобальним фіксованим порогом пригнічують хибні спрацювання в одних умовах, але пропускають цілі в інших. Темпоральне голосування «N з M кадрів» допомагає відсікти випадковий шум, проте має інерційний характер.

У роботі пропонується модифікований конвеєр виявлення, який зберігає обчислювальну простоту класичних методів і підвищує якість сегментації. Нові компоненти: (1) адаптивна локальна порогова сегментація; (2) фонові модель ЕМА; (3) концепція безперервного темпорального голосування з плавним згасанням.

Подальша структура статті організована таким чином: у Розділі 2 описана постановка проблеми; у Розділі 3 проаналізовано сучасний стан досліджень; у Розділі 4 детально описано мету та задачі дослідження; Розділ 5 детально описує архітектуру запропонованого методу; Розділи 6 і 7 присвячено експериментальній оцінці та аналізу результатів; у Розділах 8 та 9 наведено обговорення результатів і загальні висновки.

2. Постановка проблеми

Широке впровадження безпілотних літальних апаратів (БПЛА) зумовлює необхідність розробки ефективних систем взаємного спостереження, зокрема виявлення одного БПЛА з борту іншого в повітрі. Сучасні рішення для цієї задачі стикаються з фундаментальним протиріччям. З одного боку, глибинні нейромережеві методи (наприклад, YOLO чи Faster R-CNN) забезпечують високу точність, але потребують значних обчислювальних ресурсів. Це робить їхнє розгортання на малогабаритних бортових платформах із жорсткими обмеженнями SWaP (розмір, вага, енергоспоживання) вкрай складним без використання важких та енергозатратних графічних прискорювачів (GPU).

З іншого боку, класичні алгоритми комп'ютерного зору (на основі міжкадрової різниці чи оптичного потоку) є обчислювально легкими і здатними працювати на базових процесорах (CPU). Проте в базовому вигляді вони вкрай чутливі до руху самої камери (ego-motion), динамічних змін фону та артефактів стиснення відео. Крім того, традиційне застосування глобальних порогів бінаризації призводить до масових хибних спрацювань (False Positives) у контрастних зонах або, навпаки, до пропуску дрібних цілей в умовах низького контрасту.

Таким чином, існує невирішена проблема створення такого алгоритмічного конвеєра, який би поєднував обчислювальну ефективність (сумісність з Edge-системами) із високою стійкістю та точністю виявлення екстремально малих рухомих об'єктів у нестабільних умовах реальної аерозйомки.

Загалом, виявлення екстремально малих об'єктів залишається одним із найскладніших викликів комп'ютерного зору навіть для сучасних систем дистанційного зондування [16].

3. Аналіз останніх досліджень і публікацій

Розвиток БПЛА призвів до активних досліджень у галузі їх виявлення та супроводження [12]. Глибинні методи (YOLOv8, Faster R-CNN, спеціалізовані архітектури для малих цілей) досягають високої точності на великих об'єктах за умови GPU [6, 10, 11]. Проте їх важко розгорнути на бортових платформах з обмеженими ресурсами через енергоспоживання та затримки інференсу [5, 18].

Класичні методи виявлення руху ґрунтуються на міжкадровій різниці [9], оптичному потоці [7] та адаптивному відніманні фону [1, 14]. Алгоритм Лукаса-Канаде з афінною компенсацією широко застосовується для стабілізації відео [4]. Для подальшого мультитрекового супроводження традиційно застосовуються фільтри Калмана [17] або дискримінативні кореляційні фільтри [8], проте їхня робота сильно залежить від якості первинного виявлення. Традиційні підходи на основі гістограм направлений градієнтів (HOG) [3] або адаптивної оцінки щільності пікселів [19] не завжди забезпечують необхідну швидкість на бортових процесорах. Основні проблеми – залишкові артефакти на межах кадру, чутливість до динамічного фону та локальних змін контрасту. Темпоральне голосування для подолання FP: підтвердження мішені лише після стабільної реєстрації протягом M кадрів [2]. Дискретні схеми ефективні проти шуму, проте створюють інерційну затримку.

Отже, аналіз літератури показує, що існує прогалина між високоточними, але ресурсоємними нейромережевими підходами, та легкими, але чутливими до завад класичними методами. Жоден з існуючих підходів у чистому вигляді не задовольняє повністю жорсткі вимоги (SWaP) для бортових Edge-систем в умовах динамічного фону. Для вирішення цієї проблеми у роботі запропоновано метод, що інтегрує три взаємодоповнюючі компоненти: адаптивну локальну порогову сегментацію стабілізованої міжкадрової різниці, фонову модель ЕМА та безперервний темпоральний вибірець.

4. Мета і задачі дослідження

Мета дослідження – підвищення точності та надійності виявлення й супроводження малих рухомих цілей у відеопотоці з бортової камери БПЛА за умови дотримання суворих апаратних обмежень бортових обчислювачів.

Практична постановка задачі продиктована необхідністю інтеграції алгоритму в Edge-систему БПЛА-перехоплювача або спостерігача. Алгоритм має функціонувати в умовах суворих апаратних обмежень (обмежена оперативна пам'ять, відсутність тензорних ядер або GPU, живлення від бортової батареї), обробляючи відеопотік безпосередньо в польоті.

Нехай відеопослідовність задана як $\{I_t\}_{t=1}^T$, де $I_t \in R^{H \times W \times 3}$ – кольоровий кадр у момент часу t . Кожна ціль

має малі розміри і займає площу від 3 до 500 пікселів. Під час зйомки камера зазнає впливу наступних дестабілізуючих факторів:

- ego-motion – глобального афінного перетворення (трансляція, обертання, масштаб); -
- апаратних недоліків: шуму сенсора та компресійних артефактів (H.264/AVC); -
- впливу середовища: різких змін освітленості та атмосферної турбулентності.

Математично завдання зводиться до побудови відображення $A: I_t \mapsto D_t$, де $D_t = \{(x_i, y_i, s_i, c_i)\}$ – множина

виявлень із координатами центра об'єктів. Запропоноване відображення повинно задовольняти такі критерії оптимізації:

1. Recall (повнота): більшість наявних у кадрі цілей має бути успішно виявлена.
2. Precision (точність): рівень хибних спрацювань (FP) не повинен перевищувати практично допустимого порога.
3. Realtime–продуктивність: метод має працювати зі швидкістю не менше 25 кадрів/с для забезпечення своєчасного реагування.
4. Edge–сумісність: метод повинен мати низьку обчислювальну складність, що лінійно залежить від тензорних процесорів), що Одозволить $(H * W)$ його безперешкодне виконання на типових малопотужних роздільної здатності кадру , і не потребувати апаратного прискорення (графічних або мікропроцесорах.

5. Запропонований метод

На N кожному $max = 500$ сірому $q_i = \text{кадри}$ $0, 01$ G_t шукаються $d_{min} M = 15$ характерні $P_t = \text{goodFeaturesToTrack}$ точки за методом $w_s = 21p$ Ші–Томасі $(t_i - G_{1t}, N_{max}, q [13]; l, d_{min})$ pit

(1)

де , , пікселів. Для кожної точки визначається нове положення методом пірамідального оптичного потоку Лукаса–Канаде з вікном .

Матриця афінного перетворення RANSAC, обмежуючись інлайєрами:

M

$$* = \arg \min_M \sum_{i \in I} \rho(\| M \tilde{p}_{t-1}^i - \tilde{p}_t^i \|) \quad \text{оцінюється методом}$$

(2) викривлення попереднього кадру до різниці інтенсивностей:

$$D_t \quad (x, y) = |G_t(x, y) - \hat{G}_{t-\Delta}(x, y)|$$

Отримана матриця використовується для поточного виду. згладжування $\sigma = 1, 5$. з ядром 5×5

Після компенсації руху обчислюється абсолютна $\tau_H = 70$, з фіксованим глобальним

(3)

Для придушення шуму застосовується гаусівське та У традиційних методах наступним кроком є порогова бінаризація порогом що призводить до масових FP у зонах високого контрасту або пропуску цілей при низькому контрасті.

$$\mu_D \text{ та } \sigma_D, \tau_{min} \text{ стандартне } \tau(x, y) = \max \mu + k \left(\sigma_D \text{ відхилення } \sigma_D \text{ за локальної статистики зберігати} \right)$$

У кожній $k = 3, 5$ точці $\tau_{min} (=x, y15)$ $w = 25$: обчислюються $w \times w$ локальне середнє

допомогою ковзного вікна розміром

(типовий розмір

(4)

де , . Використання замість глобального порога дозволяє: (1) пропускати хибні спрацювання в «тихих» зонах; (2) чутливість до дрібних цілей у зонах із високим шумом.

Для згладжування повільних фонових змін використовується фонові модель ЕМА:

(5)

де $\alpha = 0, 05$. Модель працює як низькочастотний $B^{t+1} = (1 - \text{фільтр:} \alpha) \cdot B^t + \text{повільні} \alpha \cdot I^t$ зміни інтегруються, а різкі

переміщення цілей залишаються на різнищевій карті.

Традиційні дискретні схеми «N з M кадрів» мають інерційну затримку. Запропонований безперервний виборець використовує експоненційну модель накопичення впевненості.

Для кожного виявлення рахунок S_t збільшується $(matched) = \beta$ при $S_t(tlost - \text{появі}) = G$ (нового βd) $+ S_t - (1 - \beta) \cdot score_{tadet}$ згасає при S_t відсутності: $\geq \theta_c = 0,5$

(6)

t (7)

де гаусівська зважена відстань. Об'єкт вважається підтвердженням при i
7, $G(d) —$

видаляється β $S \leq \theta =$ $= \text{при } 0, t \quad d \quad 0, 1.$

Переваги: немає «холодного старту», плавна втрата при короткочасних пропусках, просторова зваженість.

Виявлення передаються багатоцільовому трекеру. Асоціація — жадібний найближчий-сусід із
максимальною відстанню пікселів. Для треків, що тимчасово втратили виявлення, активується

пошук у ROI радіусом $R = d100^{max}$ пікселів = 70 із низьким порогом $\tau_L = 30$.

Отже, розроблений конвеєр обробки інтегрує стабілізацію зображення, адаптивну локальну сегментацію, згладжування фону та безперервне темпоральне голосування. Така архітектура дозволяє відмовитися від ресурсомістких нейромереж і глобальних порогів, забезпечуючи потенційну можливість роботи в режимі реального часу на Edge-платформах.

6. Параметри методу

Для забезпечення ефективної роботи запропонованого алгоритму було визначено набір ключових параметрів, які розділено на дві функціональні групи. Перша група стосується етапів компенсації руху камери та виділення мішеней на фоні завад. У Таблиці 1 наведено діапазони тестування та підібрані базові значення для параметрів стабілізації та локальної порогової сегментації.

Друга група параметрів, зведена у Таблицю 2, відповідає за логіку підтвердження наявності об'єкта (темпоральне голосування) та його подальше мультитрекове супроводження.

Наведені базові значення параметрів були підібрані емпіричним шляхом під час попередніх експериментів. Вони забезпечують компроміс між чутливістю до дрібних цілей та стійкістю до хибних спрацювань, і саме ця конфігурація використовується як базова для подальшої експериментальної оцінки на тестовому датасеті.

Таблиця 1

Параметри стабілізації та сегментації

Параметр	Позначення	Діапазон	Значення
Глобальний резервний поріг	τ_H	50–90	70
Множник локального відхилення	k	1,5–6,0	3,5
Вікно локальної статистики	w	15–35	25

Мінімальний поріг	τ_{min}	5–25	15
Коефіцієнт адаптації фону	α	0,01–0,1 0	0,05
Вікно LK	w_s	15–31	21

Таблиця 2

Параметри виявлення та трекінгу

Параметр	Позначення	Діапазон	Значення
Мін. площа контуру	A_{min}	1–10	3
Макс. площа контуру	A_{max}	200–800	500
Мін. циркулярність	ψ_{min}	0,10–0,4 0	0,25
Відстань NMS	d_{NMS}	15–35	25
Коефіцієнт пам'яті	β	0,50–0,9 0	0,70
Поріг підтвердження	θ_c	0,20–0,8 0	0,50
Макс. втрата трека	T_{lost}	15–50	30
Макс. відстань асоціації	d_{max}	50–100	70
ROI радіус пошуку		50–150	100

7. Результати

Метод оцінено на загальній тривалістю 30

дослідження

датасеті з 15 відеокліпів понад 14 000 кадрів (,

кадрів/с). Анотація у форматі MOT 1.1.

1920×1080

Критерії вибору кліпів: різноманітність розмірів цілей (від 7 до 60 пікселів), кількості треків (1–14 на кліп), фону (небо, хмари, місцевість), швидкості камери та освітлення. Серед обраних: clip_1 (мала ціль, короткий), clip_10 (довгий, стабільний), clip_38 (великі цілі), clip_20 (дуже малі цілі, <11 пікселів), clip_30 (багато треків – 12), clip_40 (великі цілі на світлому фоні).

Метрики оцінювання – стандартні бінарної класифікації на рівні кадра з порогом відповідності пікселів:

$$\varepsilon = 25$$

(8)

Для кількісної оцінки ефективності запропонованого $P = TP_{TP+FP}$, R підходу $= TP_{TP+FN}$ базова, F^1 = конфігурація $P_2 + PR^R$ параметрів ($k = 3, 5$,

$\beta = 0, 7$, $\theta = 0, 5$) була протестована на всьому датасеті з 15 кліпів. У Таблиці 3 зведено детальні результати

тестування, що включають абсолютну кількість істинно-позитивних (TP), хибно-позитивних (FP) та хибно-негативних (FN) спрацювань, а також розраховані метрики Precision, Recall, F1-меру та швидкість обробки (FPS) для кожного відеофрагмента.

Таблиця 3

Базові метрики на вибірці з 15 кліпів

Кліп	Кадрів	TP	FP	FN	Precision	Recall	F1	FPS
clip_1	297	150	775	147	0,162	0,505	0,246	42,3
clip_8	1800	218	8175	3317	0,026	0,062	0,037	37,7
clip_10	1800	1348	4425	452	0,234	0,749	0,356	43,1
clip_16	798	473	1354	578	0,259	0,450	0,329	18,7
clip_17	362	59	289	360	0,170	0,141	0,154	75,1
clip_19	784	410	563	374	0,421	0,523	0,467	22,3
clip_20	707	403	328	622	0,551	0,393	0,459	24,4
clip_23	122	0	67	122	0,000	0,000	0,000	35,6
clip_28	1685	912	3546	2038	0,205	0,309	0,246	82,4
clip_30	939	889	3321	1046	0,211	0,459	0,289	19,0
clip_38	1750	931	1902	1625	0,329	0,364	0,346	71,0
clip_40	1122	67	973	1101	0,064	0,057	0,061	22,5
clip_42	1411	924	5054	1117	0,155	0,453	0,231	46,1
clip_43	730	683	205	671	0,769	0,504	0,609	72,4
clip_45	794	595	1684	660	0,261	0,474	0,337	20,5
Середнє	—	—	—	—	0,252	0,317	0,304	38,5

(а) Метрики базової конфігурації

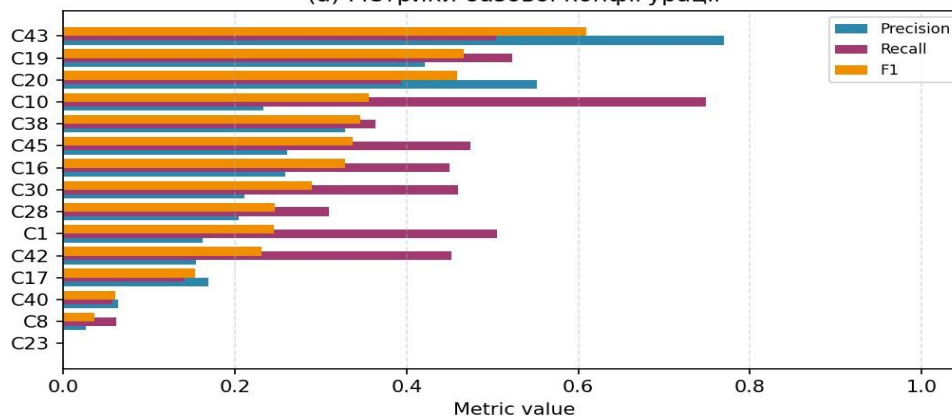


Рис. 1. Метрики базової конфігурації

Як свідчать дані з таблиці, алгоритм демонструє суттєву варіативність результатів залежно від умов сцени. Найвищі значення F1-міри досягнуто на кліпах clip_43 (0,609), clip_19 (0,467) та clip_20 (0,459). Водночас система не дала прийнятних результатів на clip_23 та clip_40. У випадку clip_23 це пояснюється надзвичайно швидким рухом камери (22,7 рх/кадр), тоді як для clip_40 проблемою стали великі розміри цілей на тлі, що зливається з фоном.

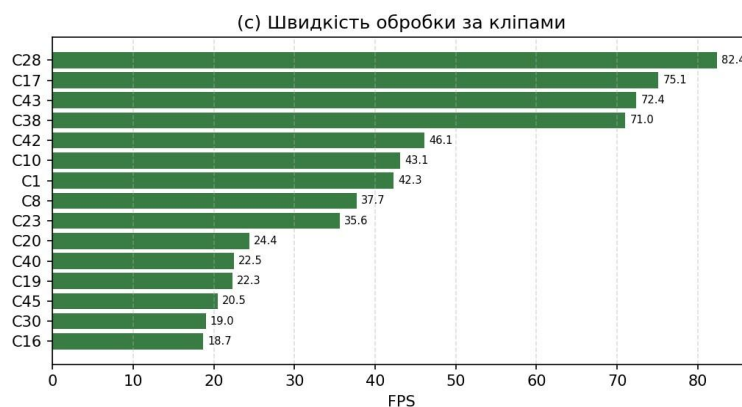


Рис. 2. Швидкість обробки за кліпами

Візуалізація метрик Precision, Recall та F1 (Рис. 1), впорядкованих за зростанням загальної якості, підтверджує цю тенденцію. Найбільш успішні кліпи (clip_43, clip_19, clip_20) характеризуються високою точністю (Precision > 0,42) при помірній повноті (Recall). На противагу їм, проблемні сцени (clip_8, clip_23, clip_40) демонструють критичне падіння всіх трьох метрик (нижче 0,1), що безпосередньо пов'язано зі специфікою фону або екстремальним рухом камери.

Окрім точності виявлення, критичним параметром для бортових систем є швидкість роботи алгоритму (Рис. 2). Аналіз швидкодії демонструє, що на більшості коротких кліпів обробка перевищує 20 кадрів/с. Бажана обчислювальна легкість методу яскраво проявляється на кліпах зі стабільним фоном та низькою щільністю треків (clip_17, clip_28, clip_38, clip_43), де швидкість сягає 70–82 кадрів/с. Найнижча ж швидкість зафіксована на clip_16 (18,7 кадрів/с) та clip_30 (19,0 кадрів/с). Це є прямим наслідком необхідності обробки 6 та 12 треків відповідно, що прогнозовано збільшує навантаження на етапи трекінгу та темпорального голосування.

Більш глибоке розуміння причин деградації метрик дає аналіз розподілу абсолютних значень TP, FP та FN (Рис. 3). Найбільша кількість хибнопозитивних спрацювань (FP) спостерігається на clip_8 (8175), clip_28 (3546) та clip_30 (3321). Це очікувано пояснюється наявністю високодинамічного фону, такого як хмари чи будівлі. При цьому лідерами за кількістю правильних виявлень (TP) є clip_10 (1348), clip_38 (931) та clip_28 (912), де мішень знаходилася на відносно стабільному та контрастному фоні.

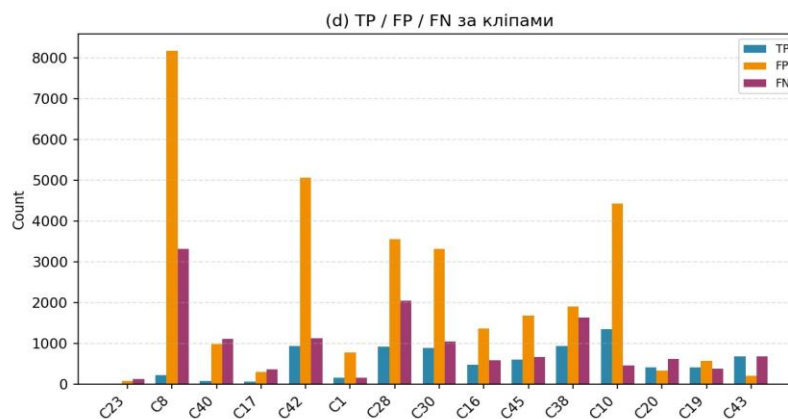


Рис. 3. TP / FP / FN за кліпами

Загалом, результати на тестовій вибірці показують, що розроблений конвеєр забезпечує високу швидкість обробки та гідну якість виявлення у сприятливих умовах. Однак наявність динамічного фону та різких рухів камери провокує стрімке зростання хибних спрацювань. Це підтверджує, що для точного розуміння меж застосовності алгоритму необхідний детальний розгляд результатів на рівні окремих об'єктів.

На рис. 4–5 наведено приклади кадрів із відео із виділеними ground-truth bounding box'ами (червоні прямокутники), що демонструють різноманітність умов зйомки та розмірів цілей.

На рис. 4 показано дрібну ціль (clip_19, середній розмір ~12 пікселів) на фоні зеленої місцевості. Повний кадр (зліва) демонструє, що дрон займає лічені пікселі. Збільшена зона (справа) показує розмитий контур білого об'єкта на однорідному тлі; червоний прямокутник позначає ground-truth bounding box. Навіть при 4-кратному збільшенні ціль залишається розпливчатою плямою, що пояснює низький Recall (0,508) і високий рівень FP через шум фону.

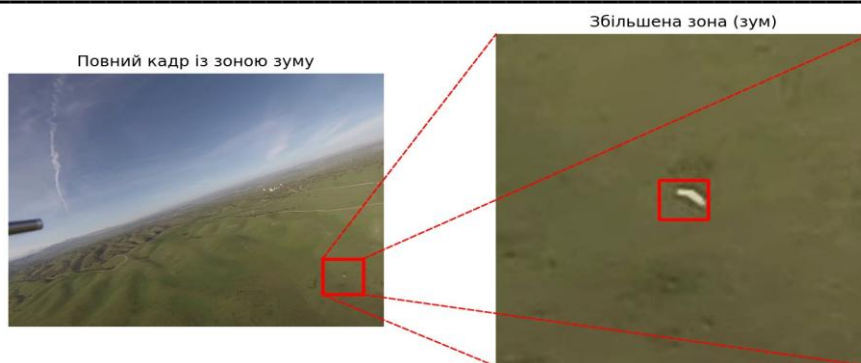


Рис. 4. Дрібна ціль із ground-truth bounding box та (clip_19)

На рис. 5 представлено сцену із двома цілями (clip_30, середній розмір ~ 16 пікселів). На повному кадрі (зліва) обидва об'єкти виглядають як мікроскопічні білі точки на зеленій поверхні. Права панель — збільшена зона із зазначеними ground-truth bounding box'ами (червоні прямокутники). Цілі розташовані на однорідному тлі, що спрощує їх виявлення, проте невеликий розмір і швидкість руху (до 10 px/кадр) ускладнюють стабільне супроводження.

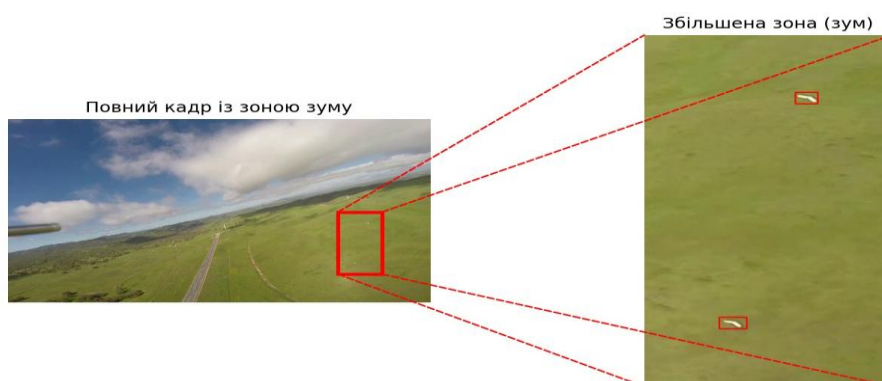


Рис. 5. Дві середні цілі із ground-truth bounding box'ами (clip_30)

Наведені якісні результати (візуалізація) підтверджують кількісні метрики: запропонований алгоритм здатен успішно локалізувати об'єкти, що займають менше 20 пікселів, однак головним викликом залишається семантична схожість цілі з фоновими структурами, що провокує хибні спрацювання.

Щоб зрозуміти причини пропусків цілей (FN), було проведено аналіз метрик окремо для кожного з 89 треків на всіх 15 кліпах. Розглянемо детальніше два полярні випадки. У Таблиці 4 наведено результати виявлення екстремально малих цілей на прикладі clip_19, де середній розмір об'єктів становить близько 12–17 пікселів.

Таблиця 4

Пооб'єктний аналіз для clip_19 (малі цілі)

ID	Кадрів	Середній розмір (px)	Швидкість (px/кадр)	TP	FN	Recall
1	461	17,0	3,03	234	227	0,508
2	291	12,0	3,26	174	117	0,598
3	31	8,5	4,08	0	31	0,000
4	1	17,0	0,00	0	1	0,000

Як видно з Таблиці 4, для надмалих об'єктів повнота виявлення (Recall) коливається в межах 0,5, а дуже короткі треки (ID 3, 4) губляться повністю. Для порівняння, у Таблиці 5 проаналізовано ситуацію з великими цілями на прикладі clip_40, де розмір об'єктів досягає 39 пікселів.

Таблиця 5

Пооб'єктний аналіз для clip_40 (великі цілі)

ID	Кадрів	Середній розмір (px)	Швидкість (px/кадр)	TP	FN	Recall
1	313	39,2	5,05	13	300	0,042

2	424	37,8	4,67	4	420	0,009
3	46	28,6	7,79	0	46	0,000
4	384	36,6	4,78	45	339	0,117
5	1	39,2	0,00	0	1	0,000

Аналіз показує, що Recall корелює з розміром цілі та тривалістю її присутності. Треки, що мають менше 30 кадрів або швидкість руху понад 10 пікселів/кадр, майже завжди пропускаються. На clip_40 при великих розмірах цілей (28–39 px) спостерігається стабільно низький Recall через невдалий вибір фону (білий дрон на тлі хмар).

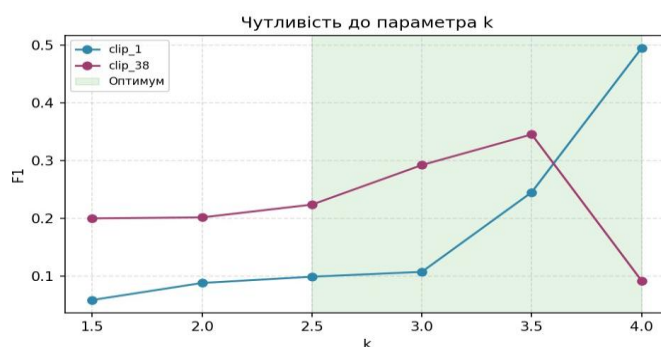
Пооб'єктний аналіз довів, що основними факторами деградації якості відстеження є недостатній розмір цілі (менше 15 пікселів), надмірна швидкість руху (понад 10 пікселів/кадр) та злиття кольору БПЛА з фоном. Тривалість присутності об'єкта в кадрі також є критичною: алгоритму потрібно понад 30 кадрів для впевненої ініціалізації треку за рахунок темпорального голосування.

Для аналізу стійкості проведено серію запусків із варіюванням коефіцієнта адаптивності на кліпах clip_1 та clip_38.

При поріг надто низький, що призводить до масових FP (Precision 0,03–0,13) і низького F1. Оптимальний $k = 1$, діапазон $5 - 2, 0$ $k = 2, 5 - 4, 0$: при $k = k3, >5$ досягнуто $4, 5$ найкращого балансу ($F1 = 0,246$ для clip_1

та $F1 = 0,346$ для clip_38). Подальше збільшення робить поріг занадто консервативним: Precision наближається до 1,00, але Recall падає до 0.

clip_1 крива має максимум при $k = 3, 5$ ($F1 = 0,246$), тоді як для clip_38 оптимум також припадає на $k = 3, 5$. На рис. 6 показано залежність F1 від коефіцієнта адаптивності для двох репрезентативних кліпів. Для Precision зростає, але $k < \text{Recall}$ 2, 5 падає, що зумовлює деградацію F1. Зеленою зоною відмічено ($F1 = 0,346$). При зменшенні спостерігається різке падіння F1 через зростання FP. При збільшенні рекомендований $k > 4, 0$ діапазон $k = 2, 5 - 4, 0$.



Аналіз чутливості до β (0,5–0,9) не виявив змін метрик k через k те, що в поточній реалізації Рис. 6. Чутливість до параметра

використовується дискретна схема N-of-M. Таким чином, хоча параметр критично впливає на баланс між повнотою та точністю, параметри темпорального згладжування демонструють високу стійкість до змін у поточній версії алгоритму. Їх тонке налаштування стане актуальним на етапі впровадження повноцінного безперервного виборця.

Метод забезпечує обробку зі швидкістю 18–82 кадрів/с залежно від складності сцени (середнє 38,5 кадрів/с). Найнижча швидкість на clip_16 (18,7 кадрів/с) та clip_30 (19,0 кадрів/с), пов'язана з великою кількістю треків (6 та 12 відповідно).

Такі показники свідчать про успішне виконання вимоги щодо роботи в режимі реального часу (Realtime), заявленої у постановці задачі (не менше 25 кадрів/с). Падіння швидкості нижче 20 кадрів/с спостерігається виключно в аномально навантажених сценах з великою кількістю треків. Загалом, отримана обчислювальна легкість підтверджує придатність методу для розгортання на малопотужних процесорах (CPU) без необхідності використання графічних прискорювачів, що є критичним для бортових систем.

8. Обговорення

Запропонований конвеєр продемонстрував працездатність на різноманітному датасеті. Основним джерелом приросту якості є адаптивна локальна сегментація, що дозволяє працювати в умовах нерівномірного фону.

Пооб'єктний аналіз виявив ключові закономірності:

1. Розмір цілі: при $\text{avg_size} < 15 \text{ px}$ Recall майже завжди $< 0,3$.
2. Тривалість: треки коротші за 50 кадрів пропускаються через темпоральне голосування.
3. Швидкість руху: $> 10 \text{ px/кадр}$ призводить до артефактів стабілізації (clip_23).
4. Контраст фону: білий БПЛА на тлі яскравої хмари (clip_40) створює систематичні FN.

ЕМА–модель теоретично згладжує коливання освітлення, але на коротких кліпах потенціал не повністю реалізований. Безперервний виборець на поточній версії не демонструє відхилення від дискретного N-of-M через обмеження реалізації.

9. Висновки та перспективи подальших досліджень

У роботі вирішується актуальна задача виявлення та супроводження малих повітряних цілей з борту БПЛА. Для подолання проблем зміщення фону та низького контрасту запропоновано та експериментально верифіковано модифікований конвеєр, що включає три нові компоненти: адаптивну локальну порогову сегментацію, фонову модель ЕМА та концепцію безперервного темпорального голосування.

Експериментальна оцінка на датасеті з 15 відеокліпів, що охоплюють різноманітні фонові умови та сценарії руху, підтвердила ефективність методу. Досягнуто середній показник $F1 = 0,304$ при середній швидкості обробки 38,5 кадрів/с. Аналіз чутливості до параметрів дозволив визначити оптимальний діапазон коефіцієнта адаптивності $k=2,5-4,0$, що дає змогу гнучко налаштовувати конвеєр під специфіку різних умов спостереження.

Практична цінність отриманих результатів полягає у можливості створення автономних бортових систем моніторингу та захисту для малого безпілотного транспорту без збільшення його ваги та енергоспоживання. Запропонований конвеєр обробки може бути інтегрований у програмне забезпечення існуючих польотних контролерів або одноплатних обчислювачів (наприклад, архітектури ARM) як фоновий модуль первинної обробки відеопотоку. Це дозволяє реалізувати сценарії автоматичного виявлення повітряних завад чи цілей-порушників у режимі реального часу без постійного залучення оператора та без використання високобюджетних апаратних прискорювачів.

Отримані показники швидкодії переконливо свідчать про те, що алгоритм не лише із запасом задовольняє вимоги реального часу (понад 25 кадрів/с), але й має високий потенціал для практичного розгортання в Edge–системах. Унікаючи ресурсомістких нейромережових обчислень, метод здатен працювати як незалежний програмний модуль, не перевантажуючи бортовий обчислювач апарата і залишаючи обчислювальні ресурси для виконання критичних польотних завдань.

Наукова новизна виконаної роботи полягає в комплексній інтеграції методів локальної адаптивної сегментації та темпорального згладжування, що дозволило компенсувати класичні недоліки методів міжкадрової різниці при роботі з рухомою платформою. Отримані залежності якості виявлення від геометричних і динамічних параметрів цілей (розміру, швидкості, контрасту) формують підґрунтя для створення гібридних двохрівневих систем: де класичний обчислювально-легкий алгоритм виступає в ролі первинного фільтра (ROI detector), а складніші нейромережові моделі залучаються точково лише для підтвердження складних або сумнівних кандидатів.

Водночас пооб'єктний аналіз виявив певні обмеження запропонованого підходу. Зокрема, зафіксовано зниження повноти виявлення (Recall) для екстремально малих цілей (менше 15 пікселів) та наявність систематичних хибних спрацювань через злиття кольору цілі з фоном (наприклад, білий БПЛА на тлі яскравих хмар). З огляду на ці виклики, подальші напрямки вдосконалення системи включають:

1. Інтеграція легковагового класифікатора (MobileNet) для зниження FP.
2. Багатомасштабний пошук для цілей < 15 пікселів.
3. Повноцінна реалізація безперервного виборця.
4. Розширення датасету нічними сценами.

Внесок авторів (обов'язково якщо більше 1 автора).

Ярошенко Андрій – аналіз джерел, створення математичного апарату, методика, збір і перевірка емпіричних даних, написання статті.

Меркулова Катерина – керівництво роботою, постановка задачі, створення математичного апарату, концептуалізація, консультування, редагування.

Декларація про штучний інтелект

Під час написання статті штучний інтелект не використовувався.

Конфлікт інтересів

Автор заявляє про відсутність конфлікту інтересів та підтверджує, що під час підготовки цієї роботи не існувало жодних комерційних, фінансових чи інших взаємовідносин, які могли б бути розцінені як такі, що здатні вплинути на результати дослідження або їх інтерпретацію. Робота виконана відповідно до принципів академічної доброчесності, етичних норм проведення наукових досліджень та вимог редакційної політики щодо запобігання конфлікту інтересів.

Список використаної літератури

1. Bouwmans, T. (2014). Traditional and recent approaches in background modeling for foreground detection. *Computer Science Review*, 11, 31–66.
2. Cucchiara, R., & Al., E. (2003). Detecting moving objects, ghosts, and shadows in video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10), 1337–1342.
3. Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *Y 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR)* (c. 886–893).
4. Evangelidis, G. D., & Psarakis, E. Z. (2008). Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 30(10), 1858–1865.
5. Hsieh, M., & Al., E. (2017). Drone-based object counting. *Y Proceedings of the IEEE international conference on computer vision (ICCV)* (c. 4145–4153).
6. Liu, W., & Al., E. (2016). SSD: Single shot multibox detector. *Y European conference on computer vision (ECCV)* (c. 21–37).
7. Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. *Y Proceedings of the international joint conference on artificial intelligence (IJCAI)* (c. 674–679).
8. Lukežič, A., & Al., E. (2017). Discriminative correlation filter with channel and spatial reliability. *Y 2017 IEEE conference on computer vision and pattern recognition (CVPR)* (c. 6309–6318).
9. Radke, R. J., & Al., E. (2005). Image change detection algorithms: A systematic survey. *IEEE Transactions on Image Processing (TIP)*, 14(3), 294–307.
10. Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. *Y Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (c. 7263–7271).
11. Ren, S., & Al., E. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 39(6), 1137–1149.
12. Rozantsev, A., Lepetit, V., & Fua, P. (2017). Flying objects detection from a single moving camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(5), 879–892.
13. Shi, J., & Tomasi, C. (1994). Good features to track. *Y Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (c. 593–600).
14. Sobral, A., & Vacavant, A. (2014). A comprehensive review of background subtraction algorithms. *Computer Vision and Image Understanding (CVIU)*, 122, 4–21.
15. Viola, P., & Jones, M. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57(2), 137–154.
16. Wang, K., & Al., E. (2021). Small-Object detection in remote sensing images. *IEEE Access*, 9, 78542–78554.
17. Welch, G., & Bishop, G. (2006). An introduction to the kalman filter. *UNC Technical Report*.
18. Zhang, Y., & Al., E. (2022). Small object detection via precise region proposal. *IEEE Access*, 10, 45678–45689.
19. Zivkovic, Z., & van der Heijden, F. (2006). Efficient adaptive density estimation per image pixel. *Pattern Recognition Letters*, 27(7), 773–780.

References

1. Bouwmans, T. (2014). Traditional and recent approaches in background modeling for foreground detection. *Computer Science Review*, 11, 31–66.
2. Cucchiara, R., & Al., E. (2003). Detecting moving objects, ghosts, and shadows in video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10), 1337–1342.
3. Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR)* (pp. 886–893).
4. Evangelidis, G. D., & Psarakis, E. Z. (2008). Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 30(10), 1858–1865.
5. Hsieh, M., & Al., E. (2017). Drone-based object counting. *In Proceedings of the IEEE international conference on computer vision (ICCV)* (pp. 4145–4153).
6. Liu, W., & Al., E. (2016). SSD: Single shot multibox detector. *In European conference on computer vision (ECCV)* (pp. 21–37).
7. Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. *In Proceedings of the international joint conference on artificial intelligence (IJCAI)* (pp. 674–679).
8. Lukežič, A., & Al., E. (2017). Discriminative correlation filter with channel and spatial reliability. *In 2017 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 6309–6318).
9. Radke, R. J., & Al., E. (2005). Image change detection algorithms: A systematic survey. *IEEE Transactions on Image Processing (TIP)*, 14(3), 294–307.
10. Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. *In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 7263–7271).
11. Ren, S., & Al., E. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 39(6), 1137–1149.

12. Rozantsev, A., Lepetit, V., & Fua, P. (2017). Flying objects detection from a single moving camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(5), 879–892.
13. Shi, J., & Tomasi, C. (1994). Good features to track. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 593–600).
14. Sobral, A., & Vacavant, A. (2014). A comprehensive review of background subtraction algorithms. *Computer Vision and Image Understanding (CVIU)*, 122, 4–21.
15. Viola, P., & Jones, M. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57(2), 137–154.
16. Wang, K., & Al., E. (2021). Small-Object detection in remote sensing images. *IEEE Access*, 9, 78542–78554.
17. Welch, G., & Bishop, G. (2006). An introduction to the kalman filter. *UNC Technical Report*.
18. Zhang, Y., & Al., E. (2022). Small object detection via precise region proposal. *IEEE Access*, 10, 45678–45689.
19. Zivkovic, Z., & van der Heijden, F. (2006). Efficient adaptive density estimation per image pixel. *Pattern Recognition Letters*, 27(7), 773–780.

Надійшла до редакції: 26.06.2026

Прийнята до друку: 15.09.2026

Опубліковано: 30.09.2026