

ЩЕРБАТЮК Мирослав

Вінницький національний технічний університет

<https://orcid.org/0009-0008-2170-7896>

e-mail: [RickLestatDT@gmail.com](mailto:RickLestatDT@gmail.com)

МАСЛІЙ Роман

Вінницький національний технічний університет

<https://orcid.org/0000-0003-3021-4328>

e-mail: [maslij.r.v@vntu.edu.ua](mailto:maslij.r.v@vntu.edu.ua)

## АНАЛІЗ МЕТОДІВ ОДНООБ'ЄКТНОГО ТРЕКІНГУ, ЩО НЕЗАЛЕЖНІ ВІД КЛАСУ

У роботі досліджено чотири алгоритми однооб'єктного трекінгу незалежних від класу: KCF, CSRT, SAMURAI та MMTrack. Проаналізовано їх ефективність за критеріями швидкості обробки, якості відстеження (LaSOT AUC), стійкості до оклюзій, змін освітлення та масштабування. Експериментальні результати показали, що KCF забезпечує найвищу швидкодію (201 fps) при низькій точності (22% AUC), CSRT демонструє баланс швидкості (80 fps) та точності (28% AUC), SAMURAI досягає високої точності (70-74% AUC) за рахунок значних обчислювальних витрат, а MMTrack поєднує точність (70% AUC) з vision-language можливостями. Розглянуті сфери застосування кожного методу для систем відеоспостереження, безпілотних платформ та робототехніки. Результати створюють методологічну основу для обґрунтованого вибору алгоритмів трекінгу відповідно до специфічних вимог застосування.

Ключові слова: однооб'єктний трекінг; KCF; CSRT; samurai; MMTrack; OpenCV; алгоритми відстеження; комп'ютерний зір; відеоспостереження; безпілотні системи.

SHCHERBATIUK Myroslav, MASLII Roman

Vinnitsia National Technical University

## ANALYSIS OF CLASS-AGNOSTIC SINGLE-OBJECT TRACKING METHODS

This study presents a comprehensive analysis of four class-agnostic single-object tracking algorithms: KCF (Kernelized Correlation Filter), CSRT (Channel and Spatial Reliability Tracking), SAMURAI, and MMTrack. The research evaluates their performance across multiple criteria including processing speed, localization accuracy (measured by LaSOT AUC), robustness to occlusions, illumination changes, and scale variations. The experimental results demonstrate distinct performance profiles for each method: KCF achieves the highest processing speed (201 fps on CPU) but shows limited accuracy (22% LaSOT AUC) and poor resilience to occlusions and scale changes; CSRT provides a balanced trade-off between speed (80 fps) and accuracy (28% AUC) with improved robustness to partial occlusions and lighting variations; SAMURAI, built upon SAM2 with motion-aware memory mechanisms, delivers exceptional accuracy (70-74% AUC) and excellent robustness to various challenging conditions, but requires substantial computational resources (0.4 fps on CPU, 13 fps on GPU); MMTrack implements a unified token-based approach for vision-language tracking, achieving comparable accuracy (70% AUC) with moderate processing speed (4 fps CPU, 54 fps GPU) and superior adaptability to scale changes. The analysis confirms that no universal solution dominates across all scenarios, and the optimal choice depends on specific application requirements, available computational resources, and performance priorities. The study establishes a methodological framework for informed algorithm selection in video surveillance, autonomous systems, and robotics applications.

Keywords: single object tracking; KCF; CSRT; SAMURAI; MMTrack; OpenCV; tracking algorithms; computer vision; video surveillance; unmanned systems

Стаття надійшла до редакції / Received 30.07.2025

Прийнята до друку / Accepted 22.08.2025

## ПОСТАНОВКА ПРОБЛЕМИ У ЗАГАЛЬНОМУ ВИГЛЯДІ ТА ЇЇ ЗВ'ЯЗОК ІЗ ВАЖЛИВИМИ НАУКОВИМИ ЧИ ПРАКТИЧНИМИ ЗАВДАННЯМИ

Зростання складності завдань комп'ютерного зору в динамічних умовах підвищує інтерес до методів однооб'єктного трекінгу, які мають забезпечити високу точність, стійкість та роботу в режимі реального часу [1, 2]. Актуальність проблеми визначається потребою в надійних і водночас ефективних методах, котрі здатні стабільно відстежувати об'єкти при зміні освітлення, ракурсів, масштабів та умов часткової чи повної оклюзії. Особливо високі вимоги до таких методів висувають безпілотні автономні системи, системи спостереження або інтелектуального транспорту, де затримка в обробці або втрата об'єкта спостереження може призвести до аварійних ситуацій, порушень безпеки або неефективності роботи системи. Таким чином, аналіз сучасних методів трекінгу є необхідним для розвитку автономних систем, що функціонують за складних умов у різноманітних сценаріях із високим рівнем невизначеності [1, 2].

## АНАЛІЗ ДОСЛІДЖЕНЬ ТА ПУБЛІКАЦІЙ

Останні роки відзначаються значним прогресом у розробці методів однооб'єктного трекінгу. Найбільш успішні підходи, історично, можна поділити на три категорії: кореляційні фільтри, Siamese-базовані методи та підходи на основі трансформерів [3].

Одним із перших і водночас найефективніших методів на основі кореляційних фільтрів був KCF (Kernelized Correlation Filter), який забезпечував високу швидкість та придатність до задач реального часу, проте виявляв обмежену стійкість до масштабних змін та оклюзій [4]. Удосконаленням цього підходу став CSRT (Channel and Spatial Reliability Tracking), який використовує багатоканальні ознаки та просторові обмеження, що дозволило підвищити точність ціною більшої обчислювальної складності [5]. Вітчизняні статті, що досліджують кореляційні фільтри були представлені у роботах [6, 7].

Розвиток глибинного навчання сприяв появі Siamese-базованих методів. Одним із перших був SiamFC [8], де трекінг формулюється як задача порівняння ознак об'єкта між послідовними фреймами. Подальші роботи – Siamese-PRN [9], SiamMask [10] та Siam R-CNN [11] – значно підвищили точність за рахунок використання регресії пропозицій, інтеграції сегментації та застосування регіональних ознак відповідно. Проте такі методи потребують навчання на великих наборах даних і недостатньо гнучкі для застосування без донавчання.

Сучасний етап розвитку трекінгу характеризується широким використанням трансформерів, зокрема завдяки механізму самоуваги. Прикладом є OSTrack [12], що інтегрує глибокі ознаки та просторово-часові залежності, досягаючи високих результатів на наборах LaSOT [13] та GOT-10k [14]. Водночас такі моделі є надзвичайно ресурсомісткими та вимагають навчання на масштабних датасетах.

Особливий інтерес становлять трансформерні підходи, здатні працювати без донавчання на нових класах. До них належить SAMURAI [15] – метод, що об'єднує фундаментальну сегментаційну модель SAM 2 [16] разом з Калман-фільтром та motion-aware механізмом пам'яті, та MMTrack [17], що об'єднує візуальну та текстову інформацію у єдине представлення для гнучкого багатомодального трекінгу.

Таким чином, хоча Siamese-базовані та трансформерні методи демонструють найвищі результати за точністю, саме KCF, CSRT, SAMURAI та MMTrack мають високу практичну цінність завдяки здатності функціонувати без навчання на нових класах, що є важливим для систем реального часу та вбудованих застосувань.

## ФОРМУЛЮВАННЯ ЦІЛЕЙ СТАТТІ

Метою роботи є проведення аналізу та порівняння сучасних алгоритмів однооб'єктного трекінгу, що не потребують навчання на нових класах, тобто незалежних від класу (class-agnostic) – KCF, CSRT, SAMURAI та MMTrack з урахуванням таких критеріїв як, стійкість до оклюзій; чутливість до зміни масштабу та освітлення; швидкість; якість трекінгу відповідно до результатів на датасеті LaSOT.

## ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

**Метод KCF** був запропонований у 2014 році у роботі [4]. Це один із відомих алгоритмів кореляційних фільтрів, який застосовує “kernel trick” – математичного прийому, що дозволяє підвищити дискримінаційну здатність фільтра без пропорційного збільшення обчислювальної складності, зберігаючи при цьому винятково високу швидкість обробки. Алгоритмічна основа KCF ґрунтується на створенні навчальної множини через циклічні зсуви початкового зразка об'єкта, формуючи циркулянтну матрицю всіх можливих позицій. Для кожного нового кадру навчений кореляційний фільтр з ядром обчислює відгук на всі циклічно зсунуті варіанти, визначаючи нову позицію об'єкта як координату максимального відгуку.

Класифікатор даного алгоритму задається як:

$$w = \operatorname{argmin}_w i \sum ((w, \phi(x)) - y_i)^2 + \lambda ||w||^2$$

де  $\phi(x)$  – відображення простору Фур'є,

$\lambda \geq 0$  – параметр регуляризації,

$y_i$  – бал, згенерований для кожного  $i$ -того циклічного зсуву ( $y_i \in [0, 1]$ ),

$w$  – фільтр (ваги).

Важливою особливістю методу є використання властивостей циркулянтних матриць для ефективного обчислення у частотній області. Застосування дискретного перетворення Фур'є та його інверсної версії дозволяє значно скоротити обчислювальну складність з  $O(n^3)$  до  $O(n \log n)$ , роблячи алгоритм придатним для застосувань реального часу. Kernel-версія додатково забезпечує збільшення кількості дискримінаційних ознак без суттєвого ускладнення алгоритму.

Алгоритм KCF поєднує в собі переваги високої швидкодії, відносну простоту реалізації та достатню точність для багатьох практичних застосувань. При цьому він не потребує попереднього навчання на великій кількості навчальних даних, як наприклад алгоритми, що базуються на глибинному навчанні, а отже сильно знижує вхідний поріг для початку застосування.

Обмеженням методу KCF є вразливість до змін масштабу об'єкта, а також трек може втрачатися при оклюзії, чи коли об'єкт, що відстежується, зливається з фоном [18]. Також присутній так званий ефект країв

(boundary effect), що зумовлено особливостями роботи алгоритму, а саме – циклічним зсувом. В результаті цього ефекту інколи трек об'єкту на відео втрачається ще до того, як об'єкт повністю покинув екран [18].

**CSRT-трекер** з бібліотеки OpenCV – це C++ реалізація методу CSR-DCF (Discriminative Correlation Filter with Channel and Spatial Reliability), запропонованого у 2016 році у роботі [5].

Метод CSR-DCF базується на кореляційних фільтрах, але вводить два основних вдосконалення, а саме: надійність каналів та просторову надійність. Це дозволяє фільтру адаптуватися лише до тієї частини об'єкта, яка ймовірно корисна для відстеження, а також оцінювати вагу різних каналів ознак залежно від їхньої якості. Зокрема, використання просторової карти надійності дає змогу ігнорувати або зменшувати вплив областей з фоном або низькою інформаційністю, а каналів з меншим внеском – зменшувати їх вагу.

Алгоритм CSR-DCF розпочинає роботу з ініціалізації, під час якої визначається обмежувальний прямокутник об'єкта. На цьому етапі витягуються характерні ознаки, зокрема гістограми орієнтованих градієнтів (HOG) та кольорові канали ColorNames. На основі цих ознак формується модель об'єкта у вигляді кореляційного фільтра, оптимізованого з урахуванням регуляризації, просторової маски та вагової надійності окремих каналів. Під час обробки наступних кадрів фільтр застосовується до зображення, утворюючи карту відгуку, де максимальне значення відповідає новому положенню об'єкта. Після цього модель оновлюється з урахуванням змін зовнішнього вигляду або часткових оклюзій, що дозволяє адаптуватися до нових умов. Використання просторових і каналних компонентів надійності забезпечує стійкість алгоритму та зменшує ймовірність появи шуму чи хибних детекцій [5].

Формула, яка використовується у CSR-DCF, для навчання кореляційного фільтра з просторовою та каналною надійністю, виглядає як мінімізація регуляризованої функції втрат:

$$\min_w \sum_c \mu_c ((w_c \times x_c) \odot m - y)^2 + \lambda ||w||^2$$

де  $w_c$  – фільтр ознак для каналу  $C$ ,

$x_c$  – ознаки (feature map) для каналу  $C$ ,

$m$  – просторова маска (spatial reliability map), котра задає, які пікселі або зони об'єкта більш “надійні”,

$\mu_c$  – коефіцієнт надійності каналу (channel reliability weight) для каналу  $C$ ,

$y$  – бажана відповідь,

$\lambda$  – параметр регуляризації.

Перевагою CSRT-трекера є досягнення балансу між точністю і надійністю, особливо в умовах, коли класичні кореляційні методи (наприклад KCF) можуть показувати великі помилки при зміні масштабу, фону або часткових чи повних оклюзіях. Даний трекер обирається, коли важлива якість локалізації об'єкта спостереження, а не максимальна швидкість обробки кадрів. Водночас, CSRT є корисним як базовий метод для порівняння в сучасних дослідженнях, оскільки він реалізує вдосконалення кореляційного підходу без надмірного ускладнення.

Одним з основних обмежень CSRT можна назвати відносно меншу швидкість роботи, у порівнянні з класичним KCF або іншими “легкими” фільтрами, оскільки покадрове оновлення моделі та обчислення просторових карт ускладнює процес роботи та потребує більшої кількості ресурсів. При застосуванні у реальних потокових відео з великою роздільною здатністю або при жорстких апаратних обмеженнях, даний алгоритм може демонструвати зменшення кількості оброблюваних кадрів на секунду, що може негативно вплинути на виконання деяких специфічних задач [2].

**Метод MMTrack** був запропонований у 2023 році у роботі [17]. Він належить до класу vision-language tracking (VLT) методів [19-21], де об'єкт уточнюється за допомогою природномовного опису. Основною особливістю даного методу є формулювання візуально-мовної задачі трекінгу як єдиної задачі генерації токенів, що дозволяє уникнути складних “prior knowledge” модулів та залежних від задач функцій втрат, характерних для попередніх VLT методів [19-21].

MMTrack серіалізує опис об'єкта та координати його обмежувального прямокутника у послідовність дискретних токенів. Далі відбувається авторегресивна генерація цільової послідовності токенів, які безпосередньо кодують координати об'єкта в поточному кадрі. Такий підхід до представлення та обробки даних дозволяє уникнути використання окремих спеціалізованих модулів і натомість застосовувати єдину функцію втрат — перехресну ентропію. Це спрощує процес навчання та підвищує ефективність у VLT-сценаріях [17].

Навчання зводиться до мінімізації перехресної ентропії над послідовністю координатних токенів:

$$L_{CE} = - \sum_{t=1}^T \log p(y_t | y_{<t}, X_{vis}, X_{lang})$$

де  $y_t$  – поточний токен координат,

$y_{<t}$  – попередній токен координат,  
 $X_{vis}$  – ознаки отримані із зображення,  
 $X_{lang}$  – ознаки отримані із тексту.

Основними обмеженнями методу MMTrack є підвищені обчислювальні витрати, зумовлені токенизацією координат та авторегресивним декодуванням, а також залежність якості результатів від точності дискретизації координат та семантичної точності текстового опису. Додатковою проблемою є накопичення помилок автогенерації на довгих відеопослідовностях, що потребує застосування спеціалізованих механізмів згладжування та корекції траєкторії відстеження.

**Метод SAMURAI** був представлений у 2024 році у роботі [15]. В основі методу лежить модель для сегментації зображень та відео Segment Anything Model 2 (SAM 2) [16]. SAMURAI адаптує SAM 2 до задачі відстеження, додаючи механізм “motion-aware memory” (пам’ять із обізнаністю про рух), щоб поліпшити вибір серед збережених масок і враховувати рух об’єкта поміж кадрів.

Алгоритм SAMURAI працює таким чином: спочатку отримує початкове положення об’єкта (обмежувальний прямокутник або сегментаційну маску) у першому кадрі, після чого використовує накопичену історію масок та інформацію про рух для точного визначення положення об’єкта в подальших кадрах. Механізм motion-aware memory динамічно відбирає збережені представлення об’єкта на основі їх якості та актуальності, що дозволяє ефективно запобігати накопиченню помилок під час довготривалого трекінгу.

Метод SAMURAI розширює підхід відстеження на основі сегментації, використовуючи механізм збереження “пам’яті”:

$$B_t = \{ m_i \mid f(S_{mask}, S_{obj}, S_{kf}) = 1, t - N_{max} \leq i \leq t \}$$

де  $B_t$  – набір знімків пам’яті (memory bank) на кадрі  $t$ ,  
 $m_i$  – знімок пам’яті з кадру  $i$ -того кадру, в момент проходження перевірки за критеріями,  
 $S_{mask}$  – оцінка достовірності маски (mask affinity score),  
 $S_{obj}$  – оцінка появи об’єкта (object occurrence score),  
 $S_{kf}$  – оцінка від Калман-фільтра (motion score),  
 $N_{max}$  – максимальна дозволена кількість кадрів, серед яких шукатимуться кандидати для пам’яті.

Завдяки zero-shot властивостям базової моделі SAM 2, метод SAMURAI демонструє високу узагальненість, забезпечуючи ефективне відстеження об’єктів нових класів без необхідності додаткового навчання. Алгоритм досягає високих результатів на складних бенчмарках LaSOT [13] та GOT-10k [14], суттєво перевершуючи показники більшості супервізійних методів однооб’єктного трекінгу [15]. Водночас метод характеризується низкою технічних обмежень: значними обчислювальними вимогами через складність SAM 2, залежністю від якості початкової ініціалізації об’єкта, а також підвищеною чутливістю до сильних деформацій об’єкту відстеження або швидких змін масштабу [15].

Відповідно до поставленої мети дослідження, проведено кількісне та якісне порівняння розглянутих методів. Кількісне оцінювання здійснювалося за критеріями продуктивності (швидкодія у fps на CPU та GPU) та точності (показник AUC на датасеті LaSOT). Якісний аналіз охоплював оцінку стійкості алгоритмів до часткових і повних оклюзій, варіацій освітлення та динамічних змін масштабу об’єктів. Експериментальну верифікацію швидкодії проведено на тестовій платформі з процесором Intel Core I7-12700KF та відеокартою NVIDIA GeForce RTX 3070 TI. Узагальнені результати порівняльного аналізу занесено до таблиці 1.

Таблиця 1.

Порівняльний аналіз методів однооб’єктного трекінгу незалежних від класу

| Критерій\Метод                | KCF (2014) | CSRT (2017) | MMTrack (2023) | SAMURAI (2024) |
|-------------------------------|------------|-------------|----------------|----------------|
| Швидкодія CPU/GPU, fps        | 201.36 / - | 80.6 / -    | 4 / 54         | 0.4 / 13.02    |
| LaSOT, AUC                    | 22%        | 28%         | 70%            | 70-74%         |
| Стійкість до оклюзій          | Низька     | Висока      | Висока         | Дуже висока    |
| Стійкість до зміни освітлення | Середня    | Висока      | Висока         | Дуже висока    |
| Стійкість до зміни масштабу   | Низька     | Середня     | Дуже висока    | Висока         |

Порівняльний аналіз алгоритмів KCF, CSRT, Samurai та MMTrack демонструє істотні відмінності як у точності відстеження, так і в обчислювальній ефективності. KCF характеризується високою швидкістю роботи на CPU (201 fps) та простотою реалізації, проте показує низьку стійкість до оклюзій і змін масштабу

об'єкта, що підтверджується низьким значенням AUC – 22% на датасеті LaSOT. Алгоритм CSRT, маючи дещо меншу але досить високу швидкодію (80.6 fps на CPU), забезпечує кращу точність (AUC – 28%), що підвищує його стійкість до часткових оклюзій і варіацій освітлення, але лиш середню можливість адаптації до зміни масштабу. Метод SAMURAI досягає вищого показника точності (AUC – 74% для large моделі та 70% для small моделі), демонструючи дуже високу стійкість до оклюзій і змін освітлення, а також хорошу здатність щодо відстеження змін масштабу, проте його швидкість є суттєво нижчою (0.4 fps на CPU та 13 fps на GPU), що обмежує застосування в системах реального часу без потужних обчислювальних ресурсів. VLT метод MMTrack, демонструє співставну з SAMURAI точність (AUC  $\approx$  70%) дещо вищу швидкодію (4 fps на CPU та 54 fps на GPU) що дозволяє його використовувати в системах реального часу при наявності потужних обчислювальних ресурсів. А завдяки використанню глибинних ознак та мультимодальної інтеграції метод забезпечує високий показник стійкості до оклюзій та змін освітлення, а також дуже високу стійкість до зміни масштабу.

Результати аналізу можливих застосувань методів, що розглядаються, у таких напрямках як відеоспостереження, БПЛА та робототехніка наведені у таблиці 2.

Таблиця 2.

**Огляд можливих застосувань методів однооб'єктного трекінгу незалежних від класу**

| Метод/Застосування | Відеоспостереження                                                                                              | БПЛА                                                                                                                | Робототехніка                                                                                 |
|--------------------|-----------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|
| KFC                | - системи реального часу;<br>- масштабні мережі з великою кількістю камер.                                      | - embedded-системи з обмеженими ресурсами;<br>- енергоефективні рішення;<br>- патрулювання в режимі реального часу. | - роботи з жорсткими вимогами до швидкодії;<br>- мобільні платформи з мінімальними ресурсами. |
| CSRT               | - високі вимоги до точності локалізації;<br>- робота в умовах зміни освітлення;<br>- часткові оклюзії об'єктів. | - моніторинг у різних погодних умовах;<br>- складні умови польоту;<br>- відстеження змінної текстури/кольору.       | - допустима втрата швидкості заради точності;<br>- робота у середовищах зі змінними умовами.  |
| MMTrack            | - відстеження неоднозначних об'єктів;<br>- пошук за текстовим описом;                                           | - відстеження нетипових об'єктів;<br>- керування через мовні команди.                                               | - доповнена реальність;<br>- мовна взаємодія з роботами;<br>- складні/неоднозначні об'єкти    |
| SAMURAI            | - адаптивні системи без перенавчання<br>- моніторинг різнотипних об'єктів                                       | - пошуково-рятувальні операції;<br>- довготривалий моніторинг;                                                      | - високоточні роботи-маніпулятори;<br>- адаптивні маніпулятори для різнотипних об'єктів       |

**ВИСНОВКИ З ДАНОГО ДОСЛІДЖЕННЯ**

**І ПЕРСПЕКТИВИ ПОДАЛЬШИХ РОЗВІДОК У ДАНОМУ НАПРЯМІ**

У ході дослідження було здійснено аналіз сучасних методів однооб'єктного трекінгу незалежних від класу, серед яких KCF, CSRT, SAMURAI та MMTrack. Отримані результати показали, що кожен з алгоритмів має свої переваги й обмеження, які визначають сфери їхнього практичного застосування. Зокрема, традиційні кореляційні методи забезпечують високу швидкодію і можуть ефективно використовуватися у системах реального часу з обмеженими ресурсами, проте суттєво поступаються сучасним глибинним моделям у точності та стійкості до складних сценаріїв. Методи нового покоління, зокрема SAMURAI та MMTrack, продемонстрували значно вищу якість відстеження та стійкість до оклюзій, змін масштабу й освітлення, однак їхнє впровадження потребує потужних обчислювальних засобів, що звужує коло можливих застосувань.

Зроблений порівняльний аналіз підтвердив відсутність єдиного універсального рішення для всіх сценаріїв трекінгу, оскільки вибір алгоритму має визначатися специфікою завдань, вимогами до точності та швидкодії, а також доступними апаратними ресурсами. Практична значущість проведеної роботи полягає у створенні методологічної основи для обґрунтованого вибору трекінг-алгоритмів у таких сферах, як відеоспостереження, робототехніка та безпілотні системи.

Подальші дослідження доцільно спрямувати на розробку гібридних рішень, що поєднують швидкість класичних кореляційних методів із точністю сучасних глибинних моделей, а також на інтеграцію алгоритмів трекінгу з інтелектуальними системами, здатними забезпечувати високу автономність і надійність у реальних умовах. Не менш актуальним завданням залишається стандартизація критеріїв оцінювання ефективності трекінгу, що сприятиме підвищенню відтворюваності результатів та практичній цінності отриманих розробок.

**Література**

1. Wu, Y., Lim, J., & Yang, M. H. (2015). Object tracking benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9), 1834–1848. <https://doi.org/10.1109/TPAMI.2014.2388226>.

2. Kristan, M., Leonardis, A., Matas, J., et al. (2019). The sixth Visual Object Tracking (VOT2018) challenge results. Proceedings of the European Conference on Computer Vision Workshops (ECCVW). [https://doi.org/10.1007/978-3-030-11009-3\\_1](https://doi.org/10.1007/978-3-030-11009-3_1).
3. Soleimanitaleb, Z., & Keyvanrad, M. A. (2022). Single object tracking: A survey of methods, datasets, and evaluation metrics. arXiv preprint arXiv:2201.13066. <https://arxiv.org/pdf/2201.13066>.
4. Henriques, J. F., Caseiro, R., Martins, P., & Batista, J. (2015). High-speed tracking with kernelized correlation filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(3), 583-596. <https://doi.org/10.1109/TPAMI.2014.2345390>.
5. Lukežič, A., Vojříř, T., Čehovin, L., Matas, J., & Kristan, M. (2017). Discriminative correlation filter with channel and spatial reliability. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 6309-6318. <https://doi.org/10.1109/CVPR.2017.515>.
6. Кийко, В. М., & Мацелло, В. В. (2020). Простеження об'єктів при відеоспостереженні. Control Systems & Computers, 2, 12-22. <https://nasplib.isoftware.kiev.ua/server/api/core/bitstreams/c139db73-0751-4f9e-ba50-f4a579874f77/content>.
7. Крепич, Р. (2025). Дослідження ефективності кореляційних методів візуального відстеження рухомих об'єктів безпілотними системами. Herald of Khmelnytskyi National University. Technical Sciences, 355(4), 264-268. <https://doi.org/10.31891/2307-5732-2025-355-38>.
8. Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A., & Torr, P. H. S. (2016). Fully-convolutional siamese networks for object tracking. European Conference on Computer Vision Workshops (ECCVW), 850-865. [https://doi.org/10.1007/978-3-319-48881-3\\_56](https://doi.org/10.1007/978-3-319-48881-3_56).
9. Li, B., Yan, J., Wu, W., Zhu, Z., & Hu, X. (2018). High performance visual tracking with Siamese region proposal network. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 8971-8980. <https://doi.org/10.1109/CVPR.2018.00935>.
10. Wang, Q., Zhang, L., Bertinetto, L., Hu, W., & Torr, P. H. S. (2019). Fast online object tracking and segmentation: A unifying approach. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 1328-1338. <https://doi.org/10.1109/CVPR.2019.00142>.
11. Voigtlaender, P., Luiten, J., Torr, P. H. S., & Leibe, B. (2020). Siam R-CNN: Visual tracking by re-detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 6578-6588. <https://doi.org/10.1109/CVPR42600.2020.00660>.
12. Ye, B., Chang, H., Ma, B., Shan, S., & Chen, X. (2022, October). Joint feature learning and relation modeling for tracking: A one-stream framework. In *European conference on computer vision* (pp. 341-357). Cham: Springer Nature Switzerland. [https://www.ecva.net/papers/eccv\\_2022/papers\\_ECCV/papers/136820332.pdf](https://www.ecva.net/papers/eccv_2022/papers_ECCV/papers/136820332.pdf).
13. Fan, H., Lin, L., Yang, F., Chu, P., Deng, G., Yu, S., Bai, H., Xu, Y., Liao, C., & Ling, H. (2019). LaSOT: A high-quality large-scale single object tracking benchmark. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 5374-5383. <https://doi.org/10.1109/CVPR.2019.00552>.
14. Huang, L., Zhao, X., & Huang, K. (2019). Got-10k: A large high-diversity benchmark for generic object tracking in the wild. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(5), 1562-1577. <https://arxiv.org/pdf/1810.11981>.
15. Wang, H., Zhou, D., Zhang, L., & Zhu, S.-C. (2024). SAMURAI: Adapting segment anything model for zero-shot visual tracking with motion-aware memory. arXiv preprint arXiv:2411.11922. <https://arxiv.org/abs/2411.11922>.
16. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W. Y., Dollar, P., & Girshick, R. (2024). Segment Anything Model 2. arXiv preprint arXiv:2408.00714. <https://arxiv.org/abs/2408.00714>.
17. Zheng, Y., Zhong, B., Liang, Q., Li, G., Ji, R., & Li, X. (2023). Toward unified token learning for vision-language tracking. IEEE Transactions on Circuits and Systems for Video Technology, 34(4), 2125-2135. <https://arxiv.org/pdf/2308.14103>.
18. Yadav, S., & Payandeh, S. (2021). Critical overview of visual tracking with kernel correlation filter. Technologies, 9(4), 93. <https://www.mdpi.com/2227-7080/9/4/93>.
19. Feng, Q., Ablavsky, V., Bai, Q., & Sclaroff, S. (2019). Robust visual object tracking with natural language region proposal network. arXiv preprint arXiv:1912.02048. <https://arxiv.org/pdf/1912.02048>.
20. Wang, X., Shu, X., Zhang, Z., Jiang, B., Wang, Y., Tian, Y., & Wu, F. (2021). Towards more flexible and accurate object tracking with natural language: Algorithms and benchmark. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 13763-13773). <https://arxiv.org/pdf/2103.16746>.
21. Yang, Z., Kumar, T., Chen, T., Su, J., & Luo, J. (2021). Grounding-tracking integration. IEEE Transactions on Circuits and Systems for Video Technology, 31(9), 3433-3443. <https://arxiv.org/pdf/1912.06316>.

## References

1. Wu, Y., Lim, J., & Yang, M. H. (2015). Object tracking benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9), 1834–1848. <https://doi.org/10.1109/TPAMI.2014.2388226>.
2. Kristan, M., Leonardis, A., Matas, J., et al. (2019). The sixth Visual Object Tracking (VOT2018) challenge results. *Proceedings of the European Conference on Computer Vision Workshops (ECCVW)*. [https://doi.org/10.1007/978-3-030-11009-3\\_1](https://doi.org/10.1007/978-3-030-11009-3_1).
3. Soleimanitaleb, Z., & Keyvanrad, M. A. (2022). Single object tracking: A survey of methods, datasets, and evaluation metrics. *arXiv preprint arXiv:2201.13066*. <https://arxiv.org/pdf/2201.13066>.
4. Henriques, J. F., Caseiro, R., Martins, P., & Batista, J. (2015). High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3), 583–596. <https://doi.org/10.1109/TPAMI.2014.2345390>.
5. Lukežić, A., Vojř, T., Čehovin, L., Matas, J., & Kristan, M. (2017). Discriminative correlation filter with channel and spatial reliability. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6309–6318. <https://doi.org/10.1109/CVPR.2017.515>.
6. Kyiko, V. M., & Matsello, V. V. (2020). Object tracking in video surveillance. *Control Systems & Computers*, 2, 12–22. <https://nasplib.isofts.kiev.ua/server/api/core/bitstreams/c139db73-0751-4f9e-ba50-f4a579874f77/content>.
7. Krepych, R. (2025). Research on the effectiveness of correlation methods for visual tracking of moving objects by unmanned systems. *Herald of Khmelnytskyi National University. Technical Sciences*, 355(4), 264–268. <https://doi.org/10.31891/2307-5732-2025-355-38>.
8. Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A., & Torr, P. H. S. (2016). Fully-convolutional siamese networks for object tracking. *European Conference on Computer Vision Workshops (ECCVW)*, 850–865. [https://doi.org/10.1007/978-3-319-48881-3\\_56](https://doi.org/10.1007/978-3-319-48881-3_56).
9. Li, B., Yan, J., Wu, W., Zhu, Z., & Hu, X. (2018). High performance visual tracking with Siamese region proposal network. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8971–8980. <https://doi.org/10.1109/CVPR.2018.00935>.
10. Wang, Q., Zhang, L., Bertinetto, L., Hu, W., & Torr, P. H. S. (2019). Fast online object tracking and segmentation: A unifying approach. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1328–1338. <https://doi.org/10.1109/CVPR.2019.00142>.
11. Voigtlaender, P., Luiten, J., Torr, P. H. S., & Leibe, B. (2020). Siam R-CNN: Visual tracking by re-detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6578–6588. <https://doi.org/10.1109/CVPR42600.2020.00660>.
12. Ye, B., Chang, H., Ma, B., Shan, S., & Chen, X. (2022, October). Joint feature learning and relation modeling for tracking: A one-stream framework. In *European conference on computer vision* (pp. 341–357). Cham: Springer Nature Switzerland. [https://www.ecva.net/papers/eccv\\_2022/papers\\_ECCV/papers/136820332.pdf](https://www.ecva.net/papers/eccv_2022/papers_ECCV/papers/136820332.pdf).
13. Fan, H., Lin, L., Yang, F., Chu, P., Deng, G., Yu, S., Bai, H., Xu, Y., Liao, C., & Ling, H. (2019). LaSOT: A high-quality large-scale single object tracking benchmark. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5374–5383. <https://doi.org/10.1109/CVPR.2019.00552>.
14. Huang, L., Zhao, X., & Huang, K. (2019). Got-10k: A large high-diversity benchmark for generic object tracking in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(5), 1562–1577. <https://arxiv.org/pdf/1810.11981>.
15. Wang, H., Zhou, D., Zhang, L., & Zhu, S.-C. (2024). SAMURAI: Adapting segment anything model for zero-shot visual tracking with motion-aware memory. *arXiv preprint arXiv:2411.11922*. <https://arxiv.org/abs/2411.11922>.
16. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W. Y., Dollar, P., & Girshick, R. (2024). Segment Anything Model 2. *arXiv preprint arXiv:2408.00714*. <https://arxiv.org/abs/2408.00714>.
17. Zheng, Y., Zhong, B., Liang, Q., Li, G., Ji, R., & Li, X. (2023). Toward unified token learning for vision-language tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(4), 2125–2135. <https://arxiv.org/pdf/2308.14103>.
18. Yadav, S., & Payandeh, S. (2021). Critical overview of visual tracking with kernel correlation filter. *Technologies*, 9(4), 93. <https://www.mdpi.com/2227-7080/9/4/93>.
19. Feng, Q., Ablavsky, V., Bai, Q., & Sclaroff, S. (2019). Robust visual object tracking with natural language region proposal network. *arXiv preprint arXiv:1912.02048*. <https://arxiv.org/pdf/1912.02048>.
20. Wang, X., Shu, X., Zhang, Z., Jiang, B., Wang, Y., Tian, Y., & Wu, F. (2021). Towards more flexible and accurate object tracking with natural language: Algorithms and benchmark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13763–13773). <https://arxiv.org/pdf/2103.16746>.
21. Yang, Z., Kumar, T., Chen, T., Su, J., & Luo, J. (2021). Grounding-tracking integration. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(9), 3433–3443. <https://arxiv.org/pdf/1912.06316>.