DOI: https://doi.org/10.15407/kvt213.03.004

КИЙКО В.М., канд.техн.наук, старш.наук.співроб.,
ст. наук. співроб.**,** відд. розпізнавання образів,
https://orcid.org/0009-0005-6605-0339, e-mail: vkiiko@gmail.com
МАЦЕЛЛО В.В., канд.техн.наук, старш.наук.співроб.,
зав. відд. розпізнавання образів,
https://orcid.org/0000-0002-6969-1554, e–mail: matsello@gmail.com
Міжнародний науково-навчальний центр інформаційних технологій
та систем НАН України та МОН України,
пр. Акад. Глушкова, 40, м. Київ, 03187, Україна

ВІДСТЕЖЕННЯ В РЕАЛЬНОМУ ЧАСІ ОБ'ЄКТІВ У ВІДЕО НА ОСНОВІ АДАПТИВНИХ ГІСТОГРАМНИХ ОЗНАК

Вступ. *Відстеження об'єктів у відео є одною з відкритих проблем комп'ютерного зору та має широкий спектр практичних застосувань. Головні складнощі цього завдання полягають у тому, що об'єкт у процесі відстеження може значно змінюватись внаслідок змін умов освітлення, розмірів та орінтації у просторі, а також зникати з поля зору. Аналіз відомих алгоритмів показує, що кожний з них не повною мірою забезпечує надійне відстеження об'єктів за вказаних вище умов. Один з підходів до підвищення надійності відстеження полягає у розробленні засобу сумісного використання кількох алгоритмів, які є взаємодоповнювальними один одного за своїми можливостями.*

Метою *статті є розроблення алгоритму тривалого відстеження об'єктів у відео в реальному часі на основі використання взаємодоповнювальних ознак та алгоритмів для отримання надійніших результатів відстеження у складних умовах.*

Результати. *Розроблено алгоритм тривалого відстеження об'єктів у відео в реальному часі на основі сумісного використання двох алгоритмів з взаємодоповнювальними ознаками та можливостями – відомого алгоритму KCF (кернелізований кореляційний фільтр) з HOG (гістограма кутів напрямку градієнтів) ознаками градієнтів яскравості і розробленого алгоритму CH (гістограма кольорових ознак) із застосуванням HSV (колір, насиченість, яскравість) гістограмних ознак кольорових подань об'єкта і фону. Показано, що алгоритм має ширші можливості для свого використання порівняно з фільтрами KCF і CH. Проведено тестування розробленого алгоритму на відео з бази даних VOT (Visual Object Tracking).*

Висновки. *Розроблений алгоритм забезпечує відновлення локалізації об'єкта після його зникнення з поля зору, а також підвищення точності і надійності порівняно з KCF і CH алгоритмами. Відновлення локалізації виконується шляхом пошуку об'єкта*

на збільшеній за розмірами ділянці зображення за допомогою KCF або іншого алгоритму. Швидкісний алгоритм CH застосовується для попереднього зменшення кількості клітин на ділянці пошуку, що можуть відповідати об'єкту, і зменшення часу його пошуку. Підвищення точності і надійності локалізації досягається шляхом використання більш інформативного критерію у вигляді зваженої суми відгуків двох фільтрів, а також точнішого визначення прямокутника, який обмежує об'єкт, на основі сегментації кольорового представлення зображення.

Ключові слова: *розпізнавання об'єктів, відстеження об'єктів у відео, KCF алгоритм відстеження, HOG ознаки, гістограмні ознаки кольорів на зображенні, локалізація об'єкта.*

ВСТУП

Візуальне відстеження об'єкта полягає у визначенні положення (локалізації) цільового об'єкта у відео після його виділення оператором або програмою на початковому кадрі. Відстеження відеооб'єктів є одною з відкритих фундаментальних проблем комп'ютерного зору та має широкий спектр практичних застосувань, таких як відеоспостереження, розуміння поведінки, автономна навігація та розпізнавання відео. Головні складнощі цієї задачі полягають у тому, що об'єкт у процесі відстеження може значно змінювати своє подання внаслідок змін умов освітлення, поворотів, розмірів, орієнтації, швидкості та напрямків руху відносно камери, а також перекриття або короткочасного зникнення з поля зору.

Є два види методів відстеження: генеративний та дискримінаційний. Генеративні методи [1, 2] розв'язують завдання локалізації шляхом пошуку областей на зображенні, найбільш схожих на відстежуваний об'єкт. Дискримінаційні методи [3–5], на відміну від генеративних, спрямовано на диференціацію цілі від фону, подаючи відстеження як проблему двійкової класифікації. Водночас дискримінаційний класифікатор здебільшого навчається або адаптується онлайн, використовуючи зразки зображень цілі та фону. Відомий алгоритм KCF [3] такого типу базується на кореляційних фільтрах дискримінантів з ознаками у частотній області і використанні алгоритму швидкого дискретного перетворення Фур'є (ШПФ) для розв'язку завдань локалізації та навчання. Трекери на основі глибинного навчання на нейронній мережі (глибинні трекери) інтенсивно досліджуються і стали популярними протягом останніх років. Загальна картина є така, що так звані «керовані глибинні трекери», які попередньо навчаються на великій кількості офлайн-маркованих даних, демонструють порівняно високу надійність відстеження відомих об'єктів, і значну меншу надійність показують неконтрольовані глибинні трекери без попереднього навчання. Крім того, глибинні трекери мають високу обчислювальну складність та вимагають великого об'єму пам'яті, що обмежує можливості їхнього практичного використання.

Ефективність трекера вимірюється точністю локалізації об'єкта, надійністю (автоматичним відновленням після втрати відстеження), обчислювальною складністю та швидкістю (кількістю кадрів за секунду). Трекери розділяють на такі, що забезпечують короткочасне або тривале відстеження. Розроблено та оновлюються вісім основних баз даних (БД) з відео для оцінювання трекерів [6], серед яких VOT БД щорічно оновлюють та використовують для визначення найефективніших трекерів.

ПОСТАНОВКА ПРОБЛЕМИ

Для відстеження необхідно після виявлення об’єкта на першому кадрі визначити його положення у вигляді мінімального обмежувального прямокутника на усіх наступних кадрах. Алгоритм повинен забезпечувати тривале відстеження об’єктів в реальному часі та умовах змін подання на зображеннях, короткочасного зникнення з поля зору або суттєвого зміщення на сусідніх кадрах унаслідок змін положення об’єкта або відео камери. Припустимо, що об’єкт на зображеннях може змінюватись головним чином внаслідок поворотів, змін масштабу та умов освітлення і, меншою мірою, внаслідок деформації і змін складових частин — переважна орієнтація на відстеження "жорстких" рухомих об’єктів (транспортні засоби тощо). Повороти можуть бути не тільки у площині зображення і призводити до зміни співвідношення горизонтального та вертикального розмірів об’єкта. Результати відстеження повинні мало залежати від похибки виявлення об’єкта на першому кадрі і отримуватись за відсутності попередньо одержаних даних про об’єкт. Адаптування класифікатора також повинно виконуватись в режимі онлайн на наступних кадрах без повторного оброблення попередніх кадрів на основі використання нових зображень об’єкта та фону.

Аналіз відомих алгоритмів показує, що кожний з них не повністю забезпечує надійне відстеження об’єктів за вказаних вище умов. Один з відомих підходів до підвищення надійності відстеження полягає у розробленні засобу сумісного використання кількох алгоритмів [7–10], які є порівняно ефективними, а також взаємодоповнювальними один одного як за типом ознак про об’єкт відстеження, так і за засобами прийняття рішень на основі цих ознак. Застосування не одного, а кількох алгоритмів зумовлено тим, що кожний з них недостатньо задовольняє практичним потребам і зазвичай має певні переваги, але також і недоліки порівняно з іншими алгоритмами. Тому використання алгоритмів зі спільною і доповнювальною дією створює передумови для отримання надійніших результатів.

Метою роботи є розроблення засобу використання взаємодоповнювальних ознак та алгоритмів, порівняно стійких до вказаних вище змін, для отримання надійніших результатів відслідковування об’єктів у складних умовах.

Важливим є також питання визначеності умов, за яких прийнятно виконувати корекцію моделі об’єкта, тому що накопичення помилок часто призводить до так званого «дріфту» — підлаштування моделі до оточення і втрати спостереження об’єкта. Пропонується спільно застосовувати такі два алгоритма. Перший — KCF, є ефективним за надійністю та швидкістю відстеження, тому його часто використовують як окремо, так і у сукупності з іншими алгоритмами. Алгоритм KCF використовує ознаки як сукупність гістограм кутів напрямку градієнтів яскравості значень цих кутів на ділянках 4x4 зображення об’єкта — так зване HOG подання [11, 12]. Другий алгоритм використовує кольорові ознаки і опис об’єкта у вигляді гістограми значень кольору клітин його зображення у дискретизованому HSV кольоровому просторі. Цей алгоритм має деякі спільні, але також значно відмітні риси порівняно з алгоритмом короткочасного відстеження кольорових об’єктів [7]. У подальших розділах ці алгоритми, а також засіб і експериментальні результати їхнього спільного використання розглянуто детальніше.

СКЛАДОВІ АЛГОРИТМИ

Алгоритм KCF використовує ШПФ у навчанні дискримінаційного кореляційного фільтра W та відстеженні цим фільтром об'єктів у відео. Навчання фільтра виконується на основі лінійної або нелінійної регресії шляхом розв'язання одної з двох відповідних задач пошуку мінімума цільової функції як суми квадратичних відхилень кореляцій від заданих значень:

$$W = \arg\min_{w} \sum_{m,n} \left|\langle x_{m,n}, W\rangle - y(m,n)\right|^2 + \lambda |w|^2, \tag{1}$$

$$\sum_{m,n} \alpha(m,n)\cdot \phi(x_{m,n}) = \arg\min_{w} \sum_{m,n} \left|\langle \phi(x_{m,n}), W\rangle - y(m,n)\right|^2 + \lambda |w|^2, \tag{2}$$

де $x{=}x_{0,0}$ — зображення ділянки $M \times N$ на початковому кадрі, який містить об'єкт в оточенні фону (Рис. 1); $x_{m,n}, (m,n) \in \{0,1,...,M-1\} \times \{0,1,...,N-1\}$ — сукупність прикладів об'єкта для навчання шляхом застосування циклічного оператора циркулянтної матриці для зсувів еталонного зображення x у горизонтальному та вертикальному напрямках; $y(m,n)$ — мітки, що відповідають $x_{m,n}$ і дорівнюють відстаням за функцією Гауса між x та $x_{m,n}$; ϕ — невідома функція відображення первинних ознак на зображенні у Гільбертовий простір, яка індукована ядром k і задовольняє умові $\langle \phi(f), \phi(g)\rangle = k(f,g)$; $\lambda \geq 0$ — параметр регуляризації, для протидії так званому «перенавчанню»; α — матриця коефіцієнтів $\alpha(m,n)$, що визначається внаслідок розв'язання задачі (2).

Задача навчання (1) розв'язується з використанням первинних ознак (яскравості клітин) об'єкта або їхніх лінійних перетворень, а (2) — ефективнішого ознакового подання, зокрема, HOG ознак, що є результатом нелінійного перетворення цих первинних ознак. Коефіцієнти α у результаті розв'язання (2) визначаються за формулою

$$A = F(\alpha) = \frac{F(y = \{y(m,n) \mid (m,n) \in \{0,1,...,M-1\} \times \{0,1,...,N-1\}\}}{F\{k(x_{m,n}, x)\} + \lambda}, \tag{3}$$

де F – функція ШПФ. Якщо $k(f_{m,n}, g_{m,n}) = k(f, g)$ для усіх m, n, f, g, що виконується, зокрема, для Гаусова RBF ядра k, який застосовується у *KCF*.

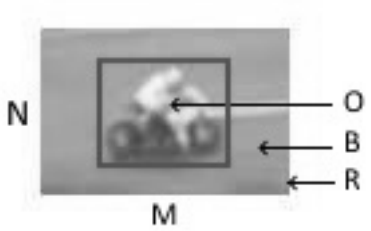

Рис. 1. Ділянки зображень об'єкта (*O*) та фону (*B*) у навчанні фільтрів.

Після навчання пошук об'єкта виконується на ділянці z у наступному кадрі з розмірами $M \times N$ і центральною точкою, що збігається з положенням об'єкта у попередньому кадрі, шляхом обчислення матриці відгуків фільтра за формулою

$$\bar{y} = F^{-1}\left(A \circ F\left(\phi(z) \cdot \phi(\bar{x})\right)\right), \tag{4}$$

де $\bar{x}$ — ознакове подання об'єкта, що оновлюється під час відстеження; $\circ$ — поелементне множення двох матриць, F^{-1} — зворотне перетворення Фур'є. Після цього визначаються координати об'єкта, які відповідають положенню максимального значення відгуку на матриці $\bar{y}$.

Подальший контроль зміни масштабу виконується шляхом обчислення додаткових значень відгуків фільтра, які відповідають зміненим розмірам об'єкта по горизонталі та вертикалі (без зміни їх відношення та координат центру), і вибору масштабу, який відповідає максимальному з цих значень. Хай $P \times Q$ — розміри об'єкта, і S — сукупність *ns* оцінюваних масштабів, наприклад, у вигляді *S={1-dk, k=n, n-1, n-2,..., -n}*, де $n = ns \cdot 0{,}5$ і d — покрокова зміна масштабу. Якщо, наприклад, *ns*=5 і d =0,02, то *S={0,96, 0,98, 1, 1,02, 1,04}*. Під час оцінювання кожного з масштабів $s \in S$ на зображенні виділяється ділянка *xs* з розмірами $sP \times sQ$ і центром в точці положення об'єкта, після чого розміри шляхом масштабування змінюють на $P \times Q$ з подальшим визначенням оцінки масштабу за формулою

$$s^* = \arg\max_{s \in S} cor(\bar{x}, xs),$$

де $cor(\bar{x}, xs)$ — нормована кореляція двох ознакових подань об'єкта — адаптованого $\bar{x}$ і масштабованого *xs*.

Модель об'єкта складається з матриці коефіцієнтів A та ознакового подання об'єкта $\bar{x}$ і оновлюється під час відстеження за формулами

$$\begin{aligned} \bar{x}(t) &= (1-\beta)\,\bar{x}(t-1) + \beta\,\bar{x}(t), \\ \bar{A}(t) &= (1-\beta)\,\bar{A}(t-1) + \beta\,\bar{A}(t), \end{aligned},$$

де t — індекс поточного кадру і β — коефіцієнт оновлення моделі у навчанні.

Алгоритм KCF використовує HOG ознаки — 31 значення гістограми орієнтації градієнта яскравості на кожній ділянці 4 х 4 клітин зображення. Такі ознаки є стійкими до змін умов освітлення, але менш надійними у разі деформацій об'єкта та змін його орієнтації. Крім того, використання циркулянтної структури матриць цих ознак призводить до небажаних крайових ефектів, які лише частково можуть бути зменшені шляхом застосування віконного фільтру Ханна. З другого боку, ознаки на основі значень кольору клітин зображення менше залежать від останніх двух змін, але більше від умов освітлення, тобто є взаємодоповнювальними до HOG.

Одне з відомих подань кольорових зображень належить HSV моделі, яка складається з трьох каналів — колір (Hue, 0–360), насиченість (Saturation, 0–100) та яскравість (Value, 0–255). Основні переваги цієї моделі порівняно з RGB полягають у тому, що колір та насиченість мало змінюються у місцях тіні, а також є можливість збільшення надійності відстеження в умовах змін освітлення шляхом оброблення відокремленого каналу яскравості. Один із засобів використання кольорових ознак полягає у доданні середніх значень кольору кожного з трьох HSV каналів на ділянках 4 x 4 до HOG ознак і збільшення таким чином загальної кількості ознак до 34, що не приводить до суттєвого підвищення надійності відстеження. Порівняно ефективнішим є застосування окремого дискримінантного фільтра на основі значень гістограми дискретизованого (квантованого) HSV подання зображення об'єкта. Додаткові переваги такого фільтру полягають у тому, що, по-перше, використані ним ознаки є незалежними від HOG ознак, а по-друге, вільними від недоліків KCF фільтрів, пов'язаних з використанням циркулянтних матриць (гістограма не залежить від будь-яких зсувів зображення). Розроблено кілька варіантів такого фільтру, порівняно швидкісний з яких CH. Його суть така.

Позначимо як G гістограму дискретизованих значень HSV подання зображення. Наприклад, якщо ступені дискретизації каналів дорівнюють відповідно 36, 20 і 64, то гістограма складається з K = (360 / 36) * (100 / 20) * (256 / 64) = 200 комірок.

Дискримінаційний фільтр на основі цієї гістограми є вектором ваг $W_j, j = 1,...,K$, що визначається на прямокутній ділянці R зображення, внутрішній прямокутник якої є зображенням об'єкта O, а зовнішні до нього клітини $t \in R$ є оточенням (контекстом) B об'єкта (Рис. 1). Ознаками кольору $\psi[t]$ кожної клітини t зображення у загальному випадку є інший вектор $\psi_j[t], j = 1,...,K$, кожна компонента якого є мірою близькості кольору цієї клітини до відповідної комірки гістограми. Обидва вектори W і ψ є нормованими і сума їхніх компонент дорівнює одиниці. Відгук фільтра на клітині t дорівнює скалярному добутку $W^T\psi$, а відгук цього фільтра на ділянці зображення дорівнює середньому значенню цього добутку. З метою зменшення обчислень та прискорення роботи фільтру у якості ознаки кольору кожної клітини зображення беремо номер комірки гістограми G, яка відповідає кольору цієї клітини [7]. Це означає, що всі компоненти вектора ознак $\psi[t]$ мають нульові значення, крім одної компоненти, яка дорівнює одиниці, і номер цієї компоненти є ознакою клітини t.

Компоненти фільтра W визначаються (навчання фільтра) шляхом застосування лінійної регресії до клітин зображення на прямокутній ділянці $R = O \cup B$ (Рис. 1) та мінімізації такої цільової функції:

$$L_C = \frac{1}{|O|}\sum_{t\in O}\left(W^T\psi(t)-1\right)^2 + \frac{1}{|B|}\sum_{t\in B}\left(W^T\psi(t)\right)^2. \qquad (5)$$

З врахуванням особливостей введених ознак (5) можна надати у вигляді

$$L_C = \sum_{j=1}^{K} \left(\frac{n_j(O)}{|O|} \left(W_j - 1 \right)^2 + \frac{n_j(B)}{|B|} W_j^2 \right). \quad (6)$$

Значення компонент фільтру, які мінімізують (6), обчислюють за формулою:

$$W_j = Z_j(O) / (Z_j(O) + Z_j(B) + \lambda), j = 1,...,K, \quad (7)$$

де $Z_j(O) = n_j(O)/|O|$, $Z_j(B) = n_j(B)/|B|$ — відношення кількості клітин з ознакою j в O і B до загальної кількості клітин в O, B; λ — параметр регуляризації, який протидіє перенавчанню. Як показує практика, трапляються ситуації, коли $z_j(O)$ має мале значення, а W_j порівняно значне (близьке до 50) внаслідок відсутності відповідного кольору на локальному фоні B під час навчання. Це може призвести до значних відгуків фільтра під час подальшого детектування об'єкта на можливих ділянках фону з переважним кольором, що відповідає j-й комірці гістограми. Для запобігання цьому, додатково до (7) застосовується $W_j = 0$, якщо $z_j(O) < thrz, j = 1,...,K$, де $thrz$ є порівняно мале порогове значення. Кольорова модель фільтра W_a оновлюється під час відстеження за формулою

$$W_a = (1 - \mu) W_a(t-1)) + \mu W(t), \quad (8)$$

де t — індекс поточного кадру, μ — коефіцієнт оновлення моделі.

Пошук об'єкта з використанням W_a полягає у скануванні по зображенню кадру прямокутної ділянки з розмірами об'єкта і обчисленні для кожного з положень R кореляції $C(R) = W_a^T G_R$, де G_R — гістограма ознак кольору на ділянці R. Обчислення кореляції виконують з оцінкою складності $O(N)$ за таким алгоритмом. Спочатку під час перегляду області пошуку на зображенні у кожну клітину записують значення W_j, де j — ознака кольору цієї клітини. Після цього формують так зване «інтегральне» подання цього зображення [13]. На заключному етапі обчислюють кореляцію на кожному прямокутнику R шляхом виконання 4-х простих операцій над значеннями, які відповідають кутам цього прямокутника на інтегральному зображенні. Одна із застосованих модифікацій кольорового фільтра полягає у тому, що об'єкт розділяють на кілька частин і, як результат, гістограма є об'єднанням складових гістограм на цих частинах, що підвищує відмінні властивості фільтра від інших об'єктів та фону.

Звісно, що розглянутий вище фільтр на основі HSV гістограмних ознак — це тільки один з варіантів реалізації доповнювального до KCF фільтра. Порівняно ефективними можуть бути також фільтри на основі гістограмних ознак «назви кольорів» (color names — CN) [14], а також LBP (локальні бінарні образи) ознак [15]. Ознаки першого типу мають гарний баланс між фотометричною інваріантністю та дискримінаційною силою, формуються шляхом навчання і відповідають уявленню людини про 11 базових кольорів (чорний, синій, корич-

невий, сірий, зелений, оранжевий, рожевий, фіолетовий, червоний, білий і жовтий). Ознаки LBP контурів на півтонових зображеннях, на відміну від HOG ознак, відповідають не кутам напряму градієнтів яскравості, а іншим властивостям зображення в околі цих градієнтних точок. Всі ці ознаки є також незалежними від HOG ознак, а фільтри на їх основі є близькими до розглянутого вище на основі HSV ознак і вільними від недоліків KCF, пов'язаних з використанням циркулянтних матриць.

СУМІСНЕ ВИКОРИСТАННЯ ТА АДАПТИВНЕ ОНОВЛЕННЯ ДВОХ ФІЛЬТРІВ

Для забезпечення тривалого відстеження об'єкта необхідно вирішити два основні завдання: відновлення локалізації після короткочасного зникнення з поля зору і запобігання втрати відстеження внаслідок дріфту, тобто накопичення помилки відстеження, що в результаті призводить до помилкового визначення частини фону як об'єкта. Перше завдання зазвичай розв'язують шляхом застосування додаткового детектора, який навчається на порівняно надійно відстежених прикладах об'єкта упродовж його відстеження [16]. Цей детектор є більш швидкісним для пошуку об'єкта на ділянках більшого розміру, але менш надійним порівняно з базовим фільтром. Це призводить до зниження надійності відстеження, особливо коли об'єкт суттєво змінює своє подання за час відсутності. Для ефективнішого розв'язання цього завдання далі розглядається використання базового або додаткового фільтра, але на суттєво зменшеній ділянці шляхом попереднього застосування модифікованого швидкісного фільтра CH на основі кольорових ознак. Запобігання дріфту у роботі головним чином забезпечується за рахунок підвищення точності локалізації шляхом спільного використання даних про об'єкт відстеження, отриманих за допомогою KCF та CH фільтрів. Розглянемо алгоритм спільного використання двох фільтрів під час відстеження детальніше.

Після локалізації об'єкта з координатами центральної точки (x_t, y_t) на поточному кадрі t виконують його пошук на наступному $(t+1)$ кадрі на ділянці z_{t+1} з розмірами $M \times N$ і центром (x_t, y_t). Для цього визначають матрицю відгуків $\bar{y}(z_{t+1})$, найбільше значення відгуку $y(x_{t+1}, y_{t+1})$ у точці (x_{t+1}, y_{t+1}), а також значення відгуку $y_{col}(x_{t+1}, y_{t+1})$ фільтра CH. Далі приймають рішення про локалізацію об'єкта у точці (x_{t+1}, y_{t+1}), якщо виконується одна з таких умов:

$$\begin{aligned} &\langle y(x_{t+1}, y_{t+1}) + y_{col}(x_{t+1}, y_{t+1}) > T_{obj}, \\ &y_{col}(x_{t+1}, y_{t+1}) > T_{col}\rangle \end{aligned} \tag{9}$$

$$\begin{aligned} &\langle y(x^*_{t+1}, y^*_{t+1}) + C_n y_{col}(x^*_{t+1}, y^*_{t+1}) > T_{obj}, \\ &y(x^*_{t+1}, y^*_{t+1}) > T_{kcf}, \\ &y_{col}(x^*_{t+1}, y^*_{t+1}) > T_{col}\rangle \end{aligned} \tag{10}$$

$$\begin{aligned}
&\langle y(x^*_{t+1}, y^*_{t+1}) + C_n y_{col}(x^*_{t+1}, y^*_{t+1}) > T_{obj1} < T_{obj},\\
&y(x^*_{t+1}, y^*_{t+1}) > T_{kcf},\\
&y(x^*_{t+1}, y^*_{t+1}) < y(x^*_t, y^*_t) * 0{,}8,\\
&cor(\hat{W}, W(x^*_{t+1}, y^*_{t+1}) > 0{,}6,\\
&y_{col}(x^*_{t+1}, y^*_{t+1}) > T_{col}\rangle
\end{aligned} \tag{11}$$

де $(x^*_{t+1}, y^*_{t+1}) = \underset{(xc,yc)\in z_{t+1}}{\arg\max}\ (y(xc, yc) + C_n y_{col}(xc, yc))$, C_n нормувальний коефіцієнт двох відгуків (фактор злиття); $T_{obj}, T_{col}, T_{kcf}$ — порогові значення.

Для прискорення перевірка приведених вище умов (9)–(11) виконується послідовно від простіших до складніших. Умови (9) полягають у порівнянні з порогами відгуків KCF та CH фільтрів у точці максимального відгуку KCF, а умови (10), (11) — максимального значення зваженої суми цих фільтрів на ділянці пошуку, причому (11) відповідає ситуації більшого зниження відгуку KCF порівняно з CH внаслідок, наприклад, зміни орієнтації або повороту об'єкта у поточному кадрі. У разі виконання умов порівняно надійної локалізації об'єкта обчислюють оновлені значення параметрів обох фільтрів на поточному кадрі за формулами (3) і (7), а також оновлення (адаптація) моделі об'єкта за формулами (5) і (8).

Якщо умови (9)–(11) не виконуються, тоді об'єкт вважають зниклим з поля зору і на наступних кадрах ділянку пошуку поступово збільшують, доти розміри не перевищують задані максимальні значення. Після відновлення локалізації розміри ділянки пошуку повертаються до значень перед першим зникненням. Інакше виконують дії для якнайбільшого зменшення кількості застосувань KCF фільтра для пошуку об'єкта на решті ділянки z_{t+1}. Для кожної клітини $p \in z_{t+1}$ визначають значення двох відгуків фільтра CH – один $CH(R_0)$ на прямокутній ділянці R_o з розмірами об'єкта і центром у клітині p, а другий $CH(R)$ на навколишній ділянці $R = R_o \cup R_B$ (Рис. 1). Якщо $CH(R_0) > thr$, де thr — задане порогове значення, і $CH(R_B) = CH(R) - CH(R_o)$ є відносно малим порівняно з $CH(R_0)$, то клітина p відмічається як одна з кандидатів на центральну точку об'єкта.

За відсутністю відмічених точок приймається рішення про зникнення об'єкта з поля зору. Інакше формується бінарне зображення v зі зменшеними у два рази розмірами, ненульові значення якого відповідають відміченим клітинам на z_{t+1}, на якому виконується пошук однозв'язкових сукупностей (компонент) клітин зі значеннями $v(p) > 0$. Кожна з цих компонент, якщо не є малою за розмірами, покривається без перекриття ділянками $M \times N$, що відповідають подальшим застосуванням KCF фільтра і обчисленню на них відгуків цього фільтра. Для кожної відміченої клітини $p \in z_{t+1}$ обчислюють також значення відгуку CH фільтра. Після цього виконують пошук на z_{t+1} точки з максимальним сумарним відгуком двох фільтрів і приймають для неї рішення, чи є вона центром відстежуваного

об'єкта за умовами (10) або (11). У відео *road* з розмірами кадрів 1210x680 з БД VOT одне зі зникнень виникає внаслідок перекриття об'єкта на 40-ому – 44-ому кадрах. На рис. 2а–2в показано відповідно результати локалізації об'єкта на 39-ому та 45-ому кадрах, а також зображення ділянки пошуку 290x224 на 44-му кадрі. На кожному з 40–44 кадрів не було відмічених точок внаслідок того, що значення $CH(R_0)$ на ділянках пошуку цих кадрів не перевищували thr, тому швидкість їх оброблення практично була такою, як і для решти кадрів відео.

Надійність відстеження суттєво залежить від точності визначення прямокутника, який обмежує цільовий об'єкт у кожному кадрі відео — чим менше цей прямокутник містить клітин фону, тим вищими є дискримінаційні властивості створюваних фільтрів і меншою — вірогідність втрати об'єкта внаслідок дріфту. Фільтр KCF відстежує зміни масштабу, але у разі змін орієнтації часто визначає обмежувальний прямокутник з помилками і не має можливостей для зміни відношення довжин його сторін.

Фільтр CH має додаткові можливості визначення положення і розмірів прямокутника на основі сегментації, але зі зниженням точності у випадках, коли кольори об'єкта мало відрізняються від кольорів фону, тому спільне використання обох фільтрів є доцільним не тільки для розглянутого вище більш точного визначення координат центральної точки об'єкта, а також і обмежувального прямокутника.

Позначимо як v зображення в межах незначно збільшеного за розмірами обмежувального прямокутника, визначеного KCF фільтром. Розглянемо алгоритм сегментації v шляхом виконання таких дій. Спочатку колір кожної клітини замінюється на значення ваги W_j, де j — номер комірки гістограми, що відповідає цьому кольору. Далі виконується сегментація отриманого подання на два класи (об'єкт та фон) і позначення кожної клітини p однією з двох міток $l_p \in \{0,1\}$. Пошук оптимальної сегментації v_b^* виконують на основі моделі Марковських випадкових полів (Markov Random Fields) за формулою:

$$v_b^* = \arg\min_{v_b} \sum_p \rho(p, l_p) + \sum_{\{p,q\} \in \mathrm{N}} w_{p,q} |l_p - l_q|, \tag{12}$$

де v_b — бінарне подання зображення v; $\rho(p, l_p)$ — штраф за позначення p міткою l_p; N — сукупність 4-х сусідніх до p клітин; $w_{p,q}$ — штраф за пару сусідніх клітин, що мають мітки l_p і l_q. Наявність першої суми в (11) приводить до близькості за кольором клітин, які мають однакові мітки, а другої — до переважного збігу міток у сусідніх клітинах.

Рис. 2. Приклад зникнення об'єкта з поля зору.

Відомо алгоритми, які ефективно з практичної точки зору розв'язують задачу (12) [17–19]. Порівняно більш швидкісний з них [18] зводить (11) до задачі пошуку максимального потоку на певним чином побудованому графі. Цей алгоритм є ітераційним і потребує на кожній ітерації двох значень кольорів або яскравості, що відповідають об'єкту і фону, наприклад, у вигляді двох характерних клітин, одна з яких належить об'єкту, а друга — фону. Під час відстеження такі дані по факту не можуть бути надано за допомогою оператора, тому пропонується їх одержувати у вигляді середніх оцінок яскравості клітин об'єкта і фону на зображенні v. Для цього спочатку застосовують алгоритм Отсу для обчислення оптимального порогу бінаризації θ зображення v [20]. Після цього обчислюють два середні значення за клітинами зображення, яскравість яких менше (більше) порога θ, які є оцінками середньої яскравості клітин об'єкта (фону) і подаються на вхід алгоритму сегментації.

Результатом сегментації є півтонове зображення v_S, на якому чим більше яскравість, тим більше вірогідність належності цього зображення до об'єкта. На цьому зображенні клітини з нульовим значенням, як правило, належать фону, але решта належить об'єкту, тільки якщо кольори суттєво відрізняються від кольорів на фоні. Інакше необхідно визначати обмежувальний прямокутник в умовах наявності завад шляхом пошуку складових горизонтальних (вертикальних) сторін, що відповідають суттєвим різницям яскравості вздовж рядків (стовпців) зліва-справа (зверху-знизу) цих сторін на зображенні v_s. Для зменшення кількості завад попередньо відмічають всі клітини, які належать до одного з так званих «максимальних» квадратів $q_m \in v_s$, що за визначенням задовольняє таким умовам:

1) для кожної клітини $p \in q_m$ виконується $v_s(p) > 0$;

2) не існує іншого максимального квадрата, частиною якого є q_m;

3) розмір q_m є меншим за задане порогове значення.

Пошук таких клітин виконують за дворазовий перегляд клітин зображення v_s, після чого всі відмічені клітини вважають належними до фону. Якщо визначені сторони обмежувального прямокутника на зображенні v_s не відповідають суттєвим змінам яскравості, то у якості обмежувального береться прямокутник, який визначено KCF фільтром. Якщо розміри прямокутника на основі сегментації відрізняються від визначених KCF фільтром і не змінюються суттєво протягом кількох кадрів, то виконується ініціалізація KCF фільтра на обмежувальному прямокутнику з новими розмірами. Для визначення нових параметрів CH фільтра на поточному кадрі за формулою (7) виконують обчислення гістограми кольорів об'єкта з врахуванням результатів сегментації.

На рис. 3 надано приклади різних подань одного і того саме об'єкта у відео *road* з БД VOT внаслідок змін масштабу та орієнтації відносно мобільної камери.

Рис. 3. Приклади зміни подання об'єкта внаслідок змін масштабу та орієнтації відносно мобільної камери.

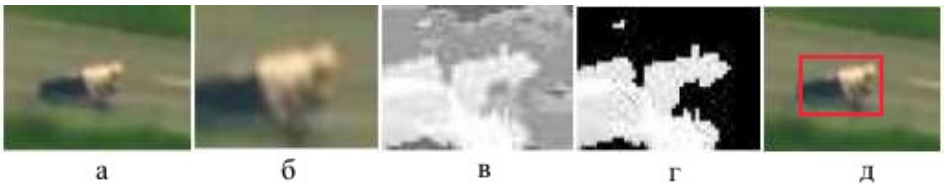

Рис. 4. Визначення обмежувального прямокутника на основі сегментації кольорового зображення.

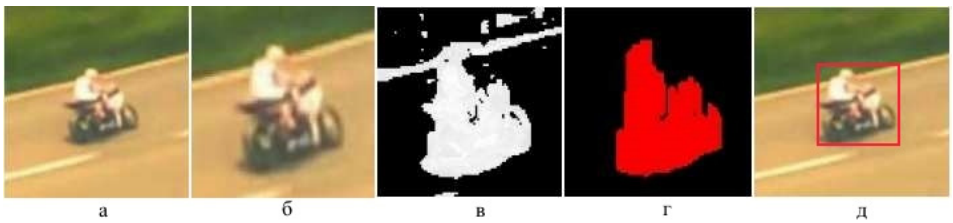

Рис. 5. Визначення обмежувального прямокутника в умовах навколишніх завад.

На рис. 4 показано приклад визначення точнішого (порівняно з KCF фільтром) прямокутника, який обмежує об'єкт на основі сегментації зображення: 4а — зображення ділянки пошуку об'єкта на кадрі; 4б — зображення v об'єкта внаслідок локалізації KCF фільтром; 4в — півтонове зображення v_s зі значеннями в клітинах, що є вагами комірок гістограми, відповідних кольорам клітин на Рис. 4б; 4г — результат сегментації зображення v_S, і 4д — результат визначення обмежувального прямокутника. На рис. 5 показано приклад визначення обмежувального прямокутника в умовах більшої кількості завад: 5а — зображення ділянки пошуку об'єкта на кадрі; 5б — зображення v об'єкта у результаті локалізації KCF фільтром; 5в — результат сегментації зображення v_S; 5г — результат усунення завад на зображенні v_S, і 5д — результат визначення обмежувального прямокутника.

РЕЗУЛЬТАТИ ЕКСПЕРИМЕНТАЛЬНОЇ ПЕРЕВІРКИ

У процесі тестування було використано дванадцять відео з БД VOT, а також чотири відео руху транспортних засобів, знятих з БПЛА. Розглядали та порівнювали між собою три алгоритми відстеження на основі навчання та застосування кореляційних фільтрів — KCF, CH та KCFCH з спільним використанням KCF і CH. Кожний з цих алгоритмів має свою сферу для переважного застосування.

Алгоритм KCF використовує HOG подання контурів на зображенні і його можна застосовувати для простеження об'єктів у відео, якщо зміна швидкості їх руху, масштабу (розмірів) та орієнтації відбувається порівняно повільно, а також якщо ці об'єкти не надто сильно і швидко змінюють свою форму і контури під час відстеження. Типовим прикладом, який задовольняє цим вимогам, є простеження транспортних засобів на дорогах під час відеоспостереження на значній відстані. Алгоритм CH використовує гістограмні ознаки кольору об'єкта, які на відміну від KCF не залежать від кругових зсувів його зображення. Цей алгоритм може бути використаний для простеження об'єктів у відео, якщо його кольори суттєво відрізняються від кольорів на фоні. Алгоритм KCFCH зі спільним використанням KCF і CH має ширші можливості для застосування порівняно з фільтрами KCF і CH.

Алгоритми відстеження зазвичай оцінюють за двома основними критеріями — точністю та надійністю. Точність визначається за усередненим по кадрах критерієм

$$P_t = \frac{|r_t \cap r_t^{gr}|}{|r_t \cup r_t^{gr}|} > thrp\,, \tag{13}$$

де r_t — визначений обмежувальний прямокутник простежуваного об'єкта у t-му кадрі, r_t^{gr} — еталонний обмежувальний прямокутник цього об'єкта, визначений за допомогою оператора. Якщо $thrp$ = 0,5, тоді умова (13) відповідає PASCAL критерію [21], за виконання якого вважається, що результат простеження є правильним і таким, що відповідає еталонним даним. Інакше у багатьох роботах використовують значення $thrp = 0$, і надійність трекера при цьому визначається як загальна кількість помилок відстеження у відео, кожна з яких відповідає умові $P_t = 0$. За умовами тестування на БД VOT на наступному 5-му кадрі після кожної такої помилки повинна автоматично виконуватись реініціалізація трекера згідно з відомими еталонними даними. Швидкість трекера є ще одним параметром, що оцінюється як середній час оброблення одного кадру у відео або кількість оброблених кадрів за секунду (FPS).

Алгоритм імплентовано на C++ з реалізацією на Intel Core(TM) i7-5500U 2.4 GHz CPU. Основні результати тестування на 12-ти відео з VOT для оцінювання результатів як короткочасного, так і тривалого відстеження об'єктів надано в таблиці 1, де надійності відповідає середнє значення кількості помилок по всіх відео. У тестуванні використано такі значення основних параметрів алгоритму KCFCH: коефіцієнти оновлення моделей об'єкта $\beta = 0{,}02$ і $\mu = 0{,}04$; кольорові ознаки – значення в комірках 10x5x4 HSV гістограми на зображенні; максимальний розмір об'єкта – 100 клітин по вертикалі та горизонталі; HOG ознаки – 31-е значення гістограми кутів градієнта яскравості на кожній ділянці 4x4 зображення об'єкта; фактор злиття відгуків двох фільтрів C_n=0,4; порогові значення T_{obj}=0,3, T_{col} =0,34, T_{kcf}=0,15.

Таблиця 1. Основні результати тестування.

Алгоритм	Точність (P_t)	Надійність	Швидкість (FPS)
KCF_CH	0,87	5,3	28
KCF	0,64	12,2	46

У роботі [7] приведено результати тестування 12-и алгоритмів короткочасного відстеження об'єктів на 28-и відео з БД VOT14 зі значеннями точності та надійності розробленого алгоритму Staple на основі HOG та RGB гістограмних кольорових ознак, що відповідно дорівнюють 0,644 і 9,38. Для коректнішого порівняння з цим та іншими відомими алгоритмами необхідно провести детальнішу перевірку на однакових вибірках даних.

ВИСНОВКИ

Розроблено алгоритм тривалого відстеження об'єктів у відео в реальному часі на основі спільного застосування двох алгоритмів з взаємодоповнювальними ознаками та можливостями — відомого алгоритму KCF з HOG ознаками градієнтів яскравості і розробленого CH алгоритму із застосуванням HSV гістограмних ознак кольорових подань об'єкта і фону. Розроблений алгоритм KCFCH забезпечує відновлення локалізації об'єкта після його зникнення з поля зору, а також підвищення точності і надійності локалізації порівняно з KCF і CH алгоритмами. Відновлення локалізації виконується шляхом пошуку об'єкта на збільшеній за розмірами ділянці зображення за допомогою KCF або іншого алгоритму. Порівняно більш швидкісний алгоритм CH застосовують для попереднього зменшення кількості клітин на ділянці пошуку, які можуть відповідати об'єкту, і зменшення часу його пошуку. Підвищення точності і надійності локалізації алгоритмом KCFCH досягається внаслідок використання більш інформативного критерію у вигляді зваженої суми відгуків двох фільтрів, а також точнішого визначення прямокутника, який обмежує об'єкт, на основі сегментації кольорового подання зображення.

У подальших дослідженнях необхідно приділити більшу увагу розробленню засобу відстеження на основі спільного застосування алгоритмів, які не є наперед заданими, а такими, що формуються або уточнюються залежно від властивостей об'єкта відстеження і зображень у відео.

REFERENCES

1. C. Bao, Y.Wu, H. Ling, and H. Ji. Real time robust l1 tracker using accelerated proximal gradient approach. In CVPR, pp.1830–1837, 2012.
2. B. Liu, J. Huang, L. Yang, and C. Kulikowski. Robust tracking using local sparse appearance model and k-selection. In CVPR, pp. 1313–1320, 2011.
3. J. F. Henriques, R. Caseiro, P. Martins, and J. Batista. High-speed tracking with kernelized correlation filters. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 37, no. 3, pp. 583–596, 2015.
4. S. Hare, A. Saffari, and P. Torr. Struck: Structured output tracking with kernels. In IEEE Trans. on PAMI, Vol. 2, No 7, pp. 1–14, 2015.

5. M. Danelljan, F. S. Khan, M. Felsberg, and J. van de Weijer. Adaptive Color Attributes for Real-Time Visual Tracking. In Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 1090–1097.
6. Weiwei Xing, Weibin Liu, Jun Wang, Shunli Zhang, Lihui Wang, Yuxiang Yang, Bowen Song. Visual Object Tracking from Correlation Filter to Deep Learning. Springer Nature, 2021, P. 193.
7. L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, P.H. Torr. Staple: Complementary learners for real-time tracking. In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016, pp. 1401–1409.
8. Zhou, H. Fu, S. You, and C. Kuo. Unsupervised lightweight single object tracking with UHP-SOT++. ArXiv: 2111.07548, 2021, P. 13.
9. M. Dunnhofer and C. Micheloni. CoCoLoT: Combining Complementary Trackers in Long-Term Visual Tracking, arXiv: 2205, 04261v1 [cs.CV] 9 May, 2022, P. 8.
10. Kyyko V.M., Matsello V.V. Object tracking by co-operative actions of background *Control Systems and Computers*, 2020. No 2, pp. 23–29. (In Ukrainian).
11. N. Dalal, B. Triggs. Histograms of oriented gradients for human detection. In Proc. of the IEEE Conf. on CVPR, San Diego, USA, 20–25 June 2005; pp. 886–893.
12. Felzenszwalb, P.F.; Girshick, R.B.; McAllester, D. Object detection discriminatively trained part-based models. IEEE Trans. PAMI. 2009, 32, pp. 1627–1645.
13. Schlesinger, M.I. Fast implementation of one class of linear convolution. Theoretical and applied questions of image recognition: Kiev: Institute of Cybernetics, 1991, pp. 61–69.
14. J. van de Weijer, C. Schmid, J. J. Verbeek, and D. Larlus. Learning color names for real-world applications. TIP, 18(7):1512–1524, 2009.
15. T. Ojala, M. Pietika.inen, and T. Ma.enpa. Multiresolution Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Trans. PAMI, vol. 24, no. 7, pp. 971–987, July 2002.
16. Chao Ma, Xiaokang Yang, Chongyang Zhang1, and Ming-Hsuan Yang. Long-term Correlation Tracking. IEEE Conf. On Computer Vision and Pattern Recognition (CVPR), 2015, pp. 5388–5396.
17. Schlesinger, M.I. Pattern recognition as an implementation of a certain subclass of thought processes. *Control Systems and Computers*, 2017. No 2, pp. 20–37.
18. Yuri Boykov, Gareth Funka-Lea. Graph Cuts and Efficient N-D Image Segmentation. Int. *Journal of Computer Vision*, Vol. 70, 2006, pp. 109–131.
19. Q. Chen and V. Koltun. Fast mrf optimization with application to depth reconstruction. in Proc. of IEEE Conf. on Computer Vision and Pattern Recognition. 2014, pp. 3914–3921.
20. Nobuyuki Otsu. A threshold selection method from gray-level histograms. IEEE Trans. Sys., Man., Cyber. 9 (1), 1979, pp. 62–66.
21. M. Everingham, L. J. V. Gool, C. K. I. Williams, J. M. Winn, and A. Zisserman. The pascal visual object classes (VOC) challenge. *IJCV*, vol. 88, no. 2, pp. 303–338, 2010.

Received 23.06.2023

Kyyko V.M., PhD (Engineering),
Senior Researcher of the Pattern Recognition Department,
https://orcid.org/ 0009-0005-6605-0339, e-mail: vkiiko@gmail.com,
Matsello V.V., PhD (Engineering),
Head of the Pattern Recognition Department,
https://orcid.org/0000-0002-6969-1554, e-mail: matsello@gmail.com

International Research and Training Center for Information
Technologies and Systems of the National Academy of Sciences
of Ukraine and the Ministry of Education and Science of Ukraine,
40, Acad. Glushkov av., Kyiv, 03187, Ukraine

REAL-TIME TRACKING OF OBJECTS IN VIDEO BASED
ON ADAPTIVE HISTOGRAM FEATURES

Introduction*. Object tracking in video is one of the open problems in computer vision and has a wide range of practical applications. The main difficulties of this task are that the object in the process of tracking can significantly change its appearance due to changes in lighting conditions, size and orientation in space, as well as disappear from the field of view. Analysis of known algorithms shows that each of them does not fully ensure reliable tracking of objects under the above conditions. One approach to improving tracking reliability is to develop a means of using multiple algorithms that complement each other in their capabilities.*

The purpose of the paper *is to develop an algorithm for long-term real-time tracking of objects in video based on the use of complementary features and algorithms to obtain more reliable tracking results in difficult conditions.*

The results. *An algorithm for long-term real-time tracking of objects in video has been developed based on the combined use of two algorithms with complementary features and capabilities - the well-known KCF algorithm with HOG features of brightness gradients and the developed CH algorithm using HSV histogram features of color representations of the object and background. It was shown that the algorithm has wider possibilities for its use compared to KCF and CH filters. The developed algorithm was tested on video from the VOT (Visual Object Tracking) database.*

Conclusions. *The developed algorithm ensures restoration of object localization after its disappearance from the field of view, as well as increasing the accuracy and reliability of localization in comparison with KCF and CH algorithms. Localization recovery was performed by searching for an object on an enlarged area of the image using KCF or another algorithm. The high-speed CH algorithm was used to preliminarily reduce the number of cells in the search area that can match the object and reduce its search time. Increasing the accuracy and reliability of localization was achieved by using a more informative criterion in the form of a weighted sum of the responses of two filters, as well as a more accurate definition of the rectangle bounding the object based on the segmentation of the color representation of the image.*

Keywords: *OBJECT RECOGNITION, object tracking in video, KCF tracking algorithm, HOG features, histogram features of colors in the image.*