

УДК 004.932.2

Система відслідковування об'єктів на відеопослідовностях у інфрачервоному та видимому діапазонах на основі канално-незалежного просторово-регуляризованого дискримінантного кореляційного фільтру

Варфоломєєв А. Ю., Короткий Є. В.

Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

E-mail: a.varfolomeiev@kpi.ua

Представлено метод відслідковування для роботи на відеопослідовностях, що містять мультиспектральну інформацію. Розглянуто можливі способи об'єднання мультиспектральної інформації при вирішенні задачі відслідковування та обґрунтовано використання підходу на основі об'єднання на рівні обчислення ознак. З огляду на компромісні можливості по відношенню до якості та швидкодії, задачу відслідковування запропоновано вирішувати за допомогою дискримінантних кореляційних фільтрів (DCF). Розроблено метод обчислення канално-незалежного дискримінантного кореляційного фільтру із просторовою регуляризацією, що оснований на використанні метода множників зі зміною напрямків (ADMM). Обчислення DCF фільтру та локалізацію об'єкта при цьому запропоновано виконувати у спеціальному просторі ознак, що використовує багатоканальні ознаки FHOГ та ознаки на основі зворотного проектування зваженої гістограми об'єкта. Зазначені ознаки пропонується обчислювати для кожного каналу відповідного кадру мультиспектральної відеопослідовності та поєднувати отримані ознаки в єдиний тензор об'єднаного простору ознак. На тесті VOT Challenge RGBT2019 показано, що реалізація запропонованого методу за якістю відслідковування може конкурувати із більш складними рішеннями, у тому числі основаними на технологіях нейронних мереж. В ході експериментів також встановлено, що збільшення обсягу контекстно-фонові інформації дозволяє дещо підвищити якість відслідковування в порівнянні з базовою реалізацією запропонованого методу навіть у випадку застосування тільки ознак FHOГ.

Ключові слова: візуальне відслідковування об'єктів; мультиспектральні зображення; дискримінантні кореляційні фільтри (DCF); метод множників зі зміною напрямків (ADMM)

DOI: [10.20535/RADAP.2020.83.5-16](https://doi.org/10.20535/RADAP.2020.83.5-16)

Вступ

Задача візуального відслідковування об'єктів, що розглядається в даній роботі, полягає у пошуку заданого об'єкта на кожному кадрі відеопослідовності (об'єкт, при цьому, вважається апіорі невідомим і задається на першому кадрі відеопослідовності, зазвичай, у вигляді прямокутної області). Актуальність зазначеної задачі обумовлена відносно широким спектром її застосування, який включає робототехніку, системи допомоги водію, людино-машинні інтерфейси, медицину, обробку відео, охоронні та оборонні системи тощо.

Характерною проблемою під час відслідковування є виникнення ситуацій, коли об'єкт потрапляє

за заваду або до області, на фоні якої він має низьку контрастність, через що його стає важко розрізнити, а отже і надійно відслідковувати. Одне з рішень, яке частково дозволяє вирішити зазначену проблему, може полягати у сумісному використанні декількох оптичних діапазонів з різними властивостями – завдяки неоднаковим властивостям, залежно від умов, об'єкт все ще може добре розрізнятися в тому чи іншому діапазоні. Гарним прикладом такого суміщення є видимий (телевізійний) та дальній інфрачервоний діапазони (див. рис. 1).

Водночас, при застосуванні декількох оптичних діапазонів постає нове питання: як система відслідковування в кожен момент часу повинна обробляти не одне, як зазвичай, а вже декілька зображень,

отриманих у цих діапазонах? Одне з рішень у даному випадку може полягати у попередньому суміщенні зображень, представлених у різних діапазонах, в одне єдине композитне зображення. Для цього, зокрема, можуть застосовуватись методи мультимодального чи мультиспектрального об'єднання зображень, описані у роботах [1–3]. Інший спосіб суміщення може передбачати об'єднання результатів, отриманих від окремих систем відслідковування [4]. При цьому зображення кожного окремого оптичного діапазону обробляється своєю системою відслідковування, після чого знайдені ними координати об'єкта поєднуються, даючи остаточний результат.

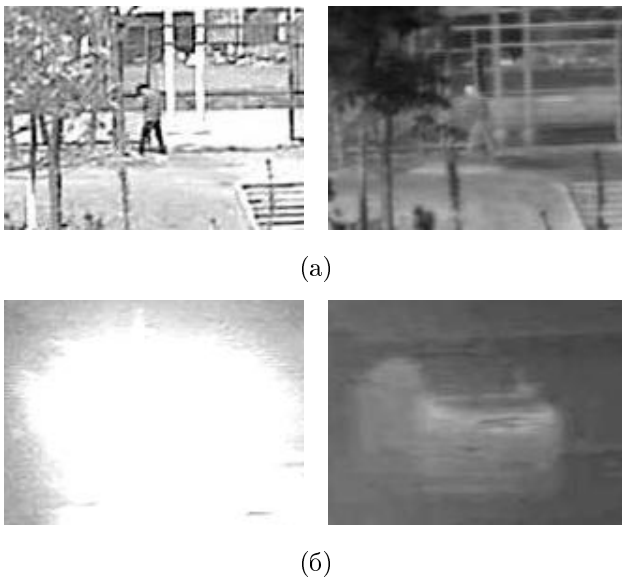


Рис. 1. Приклад сцен, на яких об'єкти важко розрізнити в одному оптичному діапазоні, проте легше в іншому: (а) людину добре видно у видимому діапазоні – лівий кадр, тоді як на інфрачервоному зображенні вона зливається з фоном – правий кадр; (б) через засвічування від фар, автомобіль неможливо побачити у видимому діапазоні – лівий кадр, проте його силует добре розрізняється в інфрачервоному діапазоні

Ще один варіант одночасної обробки декількох різномірних зображень може полягати у об'єднанні на рівні ознак [4]. Таке рішення передбачає обчислення ознак із зображень кожного з оптичних діапазонів та використання у методі відслідковування всіх отриманих ознак як єдиної множини. На думку авторів, на сьогодні саме цей варіант є найбільш доцільним, оскільки не передбачає зайвих операцій, які можуть призвести до втрати чи спотворення інформації при отриманні єдиного зображення під час об'єднання мультимодальних зображень, є простішим, ніж застосування декількох окремих підсистем відслідковування та може в явному вигляді використовуватись деякими сучасними методами відслідковування, основаними на згорткових нейронних мережах (CNN) та дискримінантних кореляційних фільтрах (DCF).

Згідно тестів [5–8] системи, що використовують згорткові нейронні мережі наразі забезпечують найвищу точність та надійність відслідковування. Водночас, вони потребують значних обчислювальних ресурсів, що ускладнює їх застосування, особливо на вбудованій апаратурі. Крім того, в таких системах частина згорткових шарів нейронних мереж є попередньо навченими [9–11], причому зазвичай на навчальних базах, що містять інформацію у видимому діапазоні. При безпосередньому застосуванні цих систем на інфрачервоних відеопослідовностях спостерігається певна деградація якості роботи, що обумовлено відмінностями представлення зовнішнього вигляду об'єктів у ГЧ діапазоні [7]. Таким чином, для забезпечення високої якості відслідковування, системи, що базуються на нейронних мережах мають навчатися на відповідних навчальних вибірках – вибірках з ГЧ зображеннями чи відеопослідовностями.

Системи, в основі яких лежать дискримінантні кореляційні фільтри (DCF) та які використовують «створені вручну» (не згорткові) ознаки, згідно результатів [7] не поступаються за якістю відслідковування деяким нейромережним рішенням. При цьому вони мають значну швидкодію, достатньою для роботи навіть на вбудованій обчислювальній апаратурі. Як приклад можна навести систему [12], яка на звичайному комп'ютері забезпечує середню частоту обробки кадрів на рівні 170-290 кадрів/с. З огляду на це, дану роботу вирішено присвятити дослідженню саме системи на основі DCF.

Загалом система відслідковування, яку розглянуто нижче вперше була випробувана авторами в рамках тесту [13], де також наведено її короткий первинний опис під назвою CISRDCF (див. VOT-RGBT2019 тест [13]). Метою ж даної публікації є повноцінний опис системи CISRDCF для інфрачервоного і видимого діапазонів, виправлення деяких неточностей, що були виявлені в первинній реалізації системи [13], а також дослідження додаткових можливостей підвищення якості роботи системи CISRDCF за рахунок використання контекстної інформації.

1 Канально-незалежний дискримінантний кореляційний фільтр

Перш ніж описувати процедуру отримання канално-незалежного дискримінантного фільтру, для повноти викладення наведемо короткі відомості щодо особливостей застосування дискримінантних кореляційних фільтрів до задачі відслідковування об'єктів. Дискримінантний кореляційний фільтр (DCF) є певним фільтром, при обчисленні кореляції (або згортки) з яким має формуватись відгук у вигляді чіткого яскравого піку в області зображення,

де розташовано об'єкт, зовнішній вигляд якого закодовано у даному DCF фільтрі. За цим піком можна визначити положення об'єкта на кадрі, що й дозволяє здійснювати відслідковування. Отримання DCF здійснюється шляхом вирішення мінімізаційної задачі на основі метода найменших квадратів з регуляризацією за Тихоновим [12, 14]. Важливою особливістю при цьому є те, що рішення мінімізаційної задачі може бути знайдене у частотній області, завдяки чому асимптотична складність обчислень зменшується до $\mathcal{O}(N \cdot \log(N))$ операцій, проти $\mathcal{O}(N^3)$ операцій для класичного підходу, де N є кількістю елементів у зображенні, з якого обчислюється фільтр. Зауважимо, що у роботі [12] також було запропоноване розширення відоме як KCF, яке для підвищення дискримінантних властивостей фільтра застосовує спрямлення простору та вперше – багатоканальні ознаки, що дозволило значно збільшити надійність відслідковування.

Водночас, пізніше було встановлено, що обчислення фільтру в частотній області має один важливий недолік: через періодичність функцій, що передбачає перетворення Фур'є, неможливо отримати фільтр, який у просторовій області був би ненульовим лише в тій частині, що кодує об'єкт [15, 16]. Це призводить до того, що до фільтру потрапляє фонова інформація, яка може призводити до неточного кодування об'єкта і як наслідок до гіршої якості відслідковування. Щоб подолати цю проблему в ряді робіт запропоновано модифікації оригінальної оптимізаційної задачі. Дані модифікації можна розділити на дві групи. В першій групі для забезпечення пошуку фільтру потрібного вигляду у задачу мінімізації вводиться додаткове обмеження. До цієї групи належать підходи [15, 17, 18] тощо, які базуються на методі оптимізації ADMM. У другій групі фільтр потрібного вигляду отримують завдяки регуляризації. Цю групу складають підходи [10, 16, 19] та подібні, які для мінімізації використовують методи Гауса-Зейделя або спряжених градієнтів. В підходах обох груп загальна складність обчислення DCF дещо зростає, але асимптотично лишається на рівні методів [12, 14]. Зауважимо, що при належному налаштуванні параметрів методу ADMM, він не поступається за якістю рішення задачі мінімізації, а його швидкодія є трохи вищою, тому процедура отримання DCF фільтру на основі ADMM є більш виправданою.

Таким чином, в даній роботі пошук DCF фільтру пропонується виконувати за допомогою оптимізаційного методу ADMM. При цьому, оскільки положення об'єкта для відслідковування зазвичай задається прямокутником, а форма реальних об'єктів далеко не завжди є ідеально прямокутною, область, що кодує об'єкт у фільтрі має сенс зробити з нечіткою границею. Це досить зручно реалізувати шляхом придушення фонових складових фільтру за допомогою регуляризації, подібно до того як це

виконується в роботах [16, 19, 20]. Як і в більшості DCF фільтрів тут легко передбачити можливість роботи з багатоканальними зображеннями. Водночас, для спрощення обчислень та більш гнучкого керування об'єднанням відгуків по кожному з каналів ознак при розрахунку кореляції, як і в роботі [18] канали ознак пропонується вважати незалежними між собою. Оптимізаційна задача в такому випадку матиме вигляд [21]:

$$\min_h \sum_{i=1}^d (\|t_i * h_i - r\|^2 + \|w \cdot g_i\|^2), \quad (1)$$

так, що $h - g = \mathbf{0}$,

де $\|\cdot\|^2$ – позначає суму квадратів елементів матриці (квадрат норми Фробеніуса); $*$ – позначає згортку між фільтром h та шаблоном t , з якого формується даний фільтр; r – бажаний відгук, у ролі якого зазвичай виступає гаусіан з малим параметром дисперсії [12, 14, 16]; w – матриця регуляризації, за допомогою якої придушуються непотрібні складові фільтру; g – так само позначає фільтр, причому обмеження при оптимізації гарантує, що $g = h$; \cdot – позначає поелементне множення між w та g . Шаблон t в задачі (1) вважається багатоканальним зображенням, тому d – позначає кількість каналів, а індекс i – номер каналу.

Наведену оптимізаційну задачу (1) можна інтерпретувати наступним чином: необхідно знайти такий фільтр h , щоб сума різниць між бажаним відгуком r та згортками по каналам цього фільтру і шаблону t були мінімальними, а також придушувались складові фільтру визначені великими значеннями у матриці w . Тут явно враховано, що зображення може мати велику кількість каналів, тобто бути, наприклад, двоканальним і містити інфрачервоний та оптичний канал або взагалі представлятись у просторі деяких багатоканальних ознак.

Щоб вирішити задачу (1), скористаємось методом ADMM у масштабованій формі [22]. Даний варіант методу передбачає ітеративне вирішення наступних оптимізаційних підзадач:

$$\begin{aligned} h^{(k+1)} &= \arg \min_h \sum_{i=1}^d (\|t_i * h_i - r\|^2 + \frac{\rho}{2} \|h_i - g_i^{(k)} + u_i^{(k)}\|^2), \\ g^{(k+1)} &= \arg \min_g \sum_{i=1}^d (\|w \cdot g_i\|^2 + \frac{\rho}{2} \|h_i^{(k+1)} - g_i + u_i^{(k)}\|^2), \\ u^{(k+1)} &= u^{(k)} + (h^{(k+1)} - g^{(k+1)}), \end{aligned} \quad (2)$$

де ρ – штрафний параметр; u – масштабована двоїста змінна, що пропорційна множникам Лагранжа [22]; показник $(\cdot)^{(k)}$ визначає значення відповідних змінних на k -ій ітерації методу ADMM.

Стисло розглянемо розв'язок кожної з підзадач визначених у (2).

Підзадачу для h у (2) через наявність згортки доцільно вирішувати в частотній області. В цьо-

му випадку в підзадачі h всі масиви заміняться на відповідні частотні образи, які надалі позначені великими літерами. Крім того застосування «теореми про згортку» дозволяє замінити згортку на поелементне множення, а використання теореми Парсеваля – квадрати норм на квадрати модулів (більш детальне формулювання розглянуто в роботі [21]). Далі, щоб вирішити перетворену задачу, необхідно врахувати, що при пошуку екстремуму дійсної функції комплексного аргументу, диференціювання можна здійснити по комплексно-спряженій змінній H^* , а отримане рівняння вирішувати відносно звичайної змінної H [14, 23]. Оскільки згортка замінена на поелементне множення, у перетвореній формулі всі операції виконуються поелементно, а отже пошук похідної можна здійснити незалежно по кожному елементу. Врешті знайшовши похідну і вирішивши отримане рівняння відносно змінної H , знайдемо рішення для підзадачі h у частотній області:

$$H^{(k+1)} = \frac{RT^* + \frac{\rho}{2}(G^{(k)} - U^{(k)})}{TT^* + \frac{\rho}{2}}, \quad (3)$$

де H, G, T, R та U – фільтри h та g , шаблон t , бажаний відгук r та масштабована двоїста змінна u в частотній області відповідно; T^* позначає комплексне-спряження шаблону t у частотній області; операції множення і ділення виконуються поелементно. Для спрощення, у формулі (3) індекси біля змінних пропущені.

Підзадачу для g у (2) можна вирішувати одразу в просторовій області. Для цього слід розписати норми та виконати диференціювання [21]. Знову таки враховуючи, що всі операції над масивами виконуються поелементно, диференціювання можна проводити незалежно по кожному елементу масиву. Результат мінімізації матиме вигляд (індекси для спрощення пропущено):

$$g^{(k+1)} = \frac{\rho(h^{(k+1)} + u^{(k)})}{2w^2 + \rho}. \quad (4)$$

Відмітимо, що рішення для підзадачі g (4) є еквівалентним відповідному рішенням, отриманому в роботі [20].

Оскільки пошук h здійснюється в частотній області, а значення g з міркувань подальшого швидкого обчислення згортки також бажано мати в частотній області, вираз (4) доцільно перетворити наступним чином:

$$G^{(k+1)} = \mathcal{F} \left[\mathcal{F}^{-1} [H^{(k+1)} + U^{(k)}] \cdot \frac{\frac{\rho}{2}}{w^2 + \frac{\rho}{2}} \right], \quad (5)$$

де $\mathcal{F}[\cdot]$ та $\mathcal{F}^{-1}[\cdot]$ – позначає пряме та зворотне перетворення Фур'є відповідно. Таким чином, вираз (5) є рішенням підзадачі для g у частотній області.

Оновлення масштабованої двоїстої змінної u може виконуватись безпосередньо в частотній області

наступним чином:

$$U^{(k+1)} = U^{(k)} + (H^{(k+1)} - G^{(k+1)}). \quad (6)$$

Для забезпечення більш високої швидкості збіжності, розмір кроку ρ також може оновлюватись на кожній ітерації алгоритму. Це зазвичай здійснюється за формулою [17, 18, 20]:

$$\rho^{(k+1)} = \min(\beta \cdot \rho^{(k)}, \rho_{\max}), \quad (7)$$

де ρ_{\max} – максимально допустимий штрафний параметр (розмір кроку); β – коефіцієнт зміни штрафного параметра. Варто відзначити, що коли ADMM використовується у масштабованій формі, при оновленні параметру ρ також повинна змінюватись і масштабована двоїста змінна U , зокрема при збільшенні ρ в β раз, U має бути зменшене у відповідну кількість раз, тобто $U = \frac{U}{\beta}$ [22].

Після виконання необхідної кількості ітерацій методу, остаточне наближення результату формується як рішення підзадачі g , тобто фільтр, пошук якого здійснюється, міститиметься у змінній $G^{(k+1)}$ на останній ітерації.

Асимптотична складність наведеного вище методу обчислення DCF фільтра визначається головним чином перетвореннями Фур'є, наявними у виразі (5) та складає $\mathcal{O}(k \cdot d \cdot mn \cdot \log(mn))$, де mn – роздільна здатність шаблону в просторі ознак, d – кількість каналів та k – кількість ітерацій методу ADMM.

Після того, як DCF фільтр знайдено, з його допомогою пошук об'єкта на зображенні F виконується шляхом обчислення згортки у частотній області з подальшим переходом у просторову область:

$$c = \sum_{i=1}^d \mathcal{F}^{-1} [F_i \cdot G_i] = \mathcal{F}^{-1} \left[\sum_{i=1}^d F_i \cdot G_i \right], \quad (8)$$

де $\mathcal{F}^{-1}[\cdot]$ – зворотне перетворення Фур'є; c – відгук фільтру (згідно постановки задачі (1)); F_i – i -й канал ознак області пошуку об'єкта в частотній області; G_i – i -й канал дискримінантного фільтру в частотній області; операція множення є поелементною. Остаточне положення об'єкту визначається шляхом пошуку піку відгуку c (за умови, що у якості r була обрана функція також з чітко вираженим піком).

Процедура виявлення об'єкта є обчислювально простішою за пошук фільтру та має складність $\mathcal{O}(mn \cdot \log(mn))$. Таким чином, загальна асимптотична складність пошуку об'єкта разом із обчисленням фільтру є такою ж як для обчислення фільтру і становить $\mathcal{O}(k \cdot d \cdot mn \cdot \log(mn))$.

2 Процедура обчислення ознак

У описаному вище підході на основі дискримінантних кореляційних фільтрів (DCF) зображення

із різних оптичних діапазонів можна було б використовувати безпосередньо, розглядаючи їх як окремі канали. Однак, майже у всіх сучасних системах відслідковування, починаючи з роботи [12], зображення не використовуються напряму – їх переводять у деякий простір ознак. Це дозволяє підвищити дискримінантні властивості DCF фільтрів, а отже і надійність відслідковування.

В даній роботі також пропонується використовувати ознаки, зокрема FHOГ [24] та ознаки на основі зворотного проектування гістограм, подібно до підходу, що описаний у роботі [25]. Коротко наведемо відомості про зазначені види ознак.

Ознаки FHOГ [24] для монохромного зображення розраховуються наступним чином: зображення розбивається на невеликі квадратні фрагменти, які називаються комірками. У комірках для кожного пікселя оцінюється амплітуда та напрямок градієнту. Далі, для кожної комірки будується гістограма орієнтації градієнтів. Дана гістограма містить задану кількість стовпців, кожен з яких відповідає певному діапазону напрямків градієнтів. У стовпці гістограми заносяться суми амплітуд градієнтів пікселів комірки, що мають відповідні цим стовпцям напрямки градієнтів. На відміну від класичної реалізації HOГ, гістограма FHOГ включає стовпці напрямків градієнтів, стовпці орієнтацій градієнтів, що містять суму амплітуд градієнтів протилежних напрямків, а також 4 стовпці, що несуть інформацію про енергію градієнтів (текстурні ознаки). Стовпців орієнтацій вдвічі менше, ніж стовпців напрямків, оскільки кожен стовпець орієнтації враховує два діапазони напрямків градієнтів – прямиий та розвернутий на 180° . Це робить стовпці орієнтацій контрастно-нечутливими, так як вони можуть кодувати на зображеннях границі утворені переходами як з темної області в світлу, так і навпаки – зі світлої в темну. Крім того, комірки у FHOГ беруться без перекриття, натомість тут застосовується так званий м'який розподіл амплітуд по стовпчикам (soft binning), в якому передбачається вплив на гістограму даної комірки також пікселів із сусідніх до неї комірок, для чого застосовується інтерполяція [24]. Загалом гістограми ознак FHOГ можна інтерпретувати як багатоканальні зображення, в яких комірки відповідають пікселям, а стовпчики гістограм – окремим каналам зображення. Більш детально про ознаки FHOГ та про їх застосування до задачі відслідковування можна дізнатись у роботах [12, 24, 26].

В даній роботі ознаки FHOГ обчислюються для комірок розміром 4×4 пікселі та мають 9 орієнтацій, що дає гістограми з 9 стовпцями орієнтацій, $2 \times 9 = 18$ стовпцями напрямків та 4 стовпцями, що описують енергії градієнтів. Таким чином, зображення у просторі ознак FHOГ є в 4 рази меншими по ширині та висоті, однак мають $9 + 18 + 4 = 31$ канал.

Ознаки FHOГ головним чином кодують інформацію про границі об'єктів на зображенні. Водночас, щоб додатково враховувати яскравість або у випадку інфрачервоних зображень нагрів притаманний об'єкту, пропонується використовувати ознаки на основі зворотного проектування гістограм. Дані ознаки несуть інформацію про відносну повторюваність пікселів характерних для зображення об'єкта. Процедура їх обчислення ґрунтується на підході, описаному в роботі [25], розглянемо її більш детально.

Обчислення ознак на основі зворотного проектування гістограм передбачає отримання зваженої гістограми об'єкта. Дана гістограма визначає імовірність приналежності кожної градації яскравості до об'єкту та розраховується наступним чином [25]:

$$\beta_j = \frac{p_j(O)}{p_j(O) + p_j(B) + \lambda}, \quad (9)$$

де j – номер стовпця гістограми, асоційований з відповідною градацією яскравості; $p_j(O) = \frac{N_j(O)}{|O|}$ та $p_j(B) = \frac{N_j(B)}{|B|}$ – відносні частоти появи пікселя з яскравістю j в області об'єкта та фону відповідно; $N_j(\cdot)$ визначає кількість пікселів з яскравістю j у заданій області; $|\cdot|$ – загальна кількість пікселів у області; O та B – множини пікселів області об'єкта та фону відповідно; λ – параметр регуляризації, що запобігає діленню на мале значення або 0.

За знайденою зваженою гістограмою β виконують зворотне проектування. Для цього на поточному кадрі, в області центрованої на минулому положенні об'єкту обчислюється вага кожного з пікселів. При цьому кожен піксель замінюється на відповідне його яскравості значення β_j , формуючи мапу правдоподібності. Очевидно, що згідно з (9) пікселі, яскравості яких часто зустрічаються на об'єкті, але майже відсутні у фоні матимуть значення ваг β_j близькі до 1, водночас пікселі, що часто присутні у фоні, але майже не зустрічаються в області об'єкта навпаки матимуть ваги близькі до 0. Таким чином, в мапі правдоподібності більші значення матимуть точки, що відповідають положенням пікселів з розподілом кольорів притаманним саме об'єкту.

В даній роботі на відміну від [25] мапу правдоподібності пропонується вважати додатковим каналом ознак, що об'єднується з каналами ознак FHOГ. При цьому обчислення FHOГ та ознак на основі зворотного проектування гістограм (мап правдоподібностей) здійснюється як для оптичного, так і для інфрачервоного зображень, після чого всі отримані ознаки поєднуються (рис. 2). При поєднанні мапа правдоподібності масштабується до розміру матриць в каналах FHOГ.

При поєднанні ознак FHOГ із ознаками на основі зворотного проектування гістограм слід враховувати, що FHOГ вже мають певну кількість каналів, тоді як мапа правдоподібності є одним єдиним

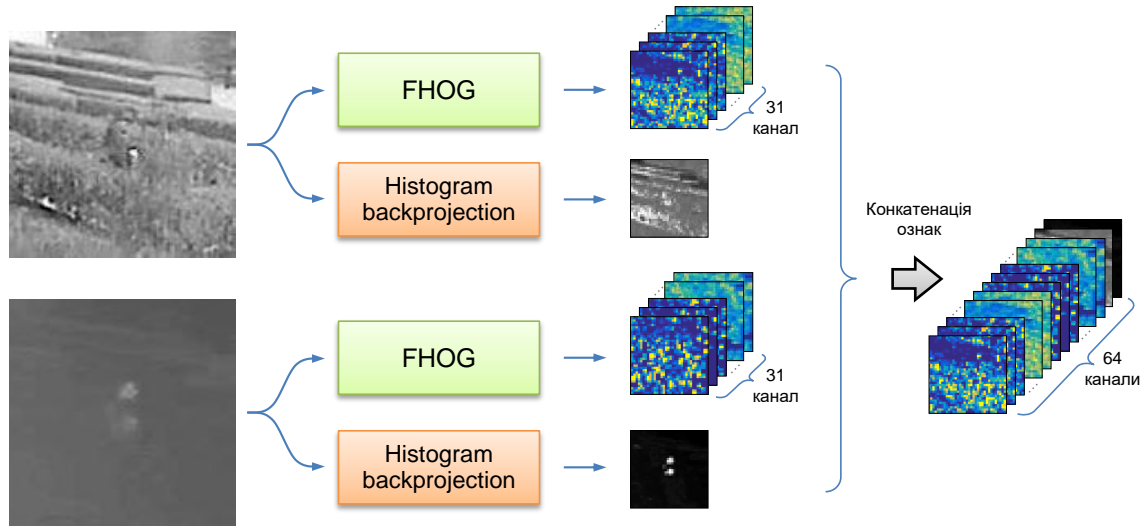


Рис. 2. Процедура обчислення ознак із оптичного та інфрачервоного зображень

каналом. Тому для забезпечення збалансованості, об'єднання відгуків на етапі обчислення кореляції при локалізації об'єкта виконується за принципом подібним, описаному в роботі [25]:

$$c = c_{HOG} + \gamma \cdot c_{HBP}, \quad (10)$$

де c_{HOG} – сумарний кореляційний відгук, отриманий по всім каналам FHOГ ознак; c_{HBP} – відгук отриманий із мапи правдоподібності знайденої шляхом зворотного проектування нормалізованої гістограми β ; γ – коефіцієнт об'єднання, що враховує ступінь важливості відгуків.

Під час відслідковування, для кращого пристосування до зміни зовнішнього вигляду об'єкта, його зважену гістограму β як і в роботі [25] має сенс оновлювати на кожному кадрі:

$$\beta^{(t+1)} = (1 - \eta_{HBP}) \cdot \beta^{(t)} + \eta_{HBP} \cdot \beta, \quad (11)$$

де $\beta^{(t+1)}$ та $\beta^{(t)}$ – зважені гістограми об'єкта для наступного та поточного кадрів відповідно; β – гістограма знайдена для області розташування об'єкта на поточному кадрі; η_{HBP} – коефіцієнт оновлення.

3 Експериментальні випробування

3.1 Особливості реалізації методу відслідковування CISRDCF

Щоб перевірити запропоновані вище вдосконалення, у середовищі MATLAB було створено програмну реалізацію методу відслідковування CISRDCF, яка має параметри, що описані в даному розділі.

В програмній реалізації використовувались як ознаки FHOГ [24], так і описані вище ознаки на основі зворотного проектування гістограм. Обчислення ознак на кожному кадрі здійснювалось для

видимого та інфрачервоного каналу та формувалася об'єднаний тензор ознак з 64 каналами. Процедуру його формування ілюструє рис. 2. В даній процедурі передбачається, що розміри інфрачервоного та видимого зображень є однаковими.

Обчислення дискримінантного фільтру для виявлення об'єкта здійснювалось незалежно по кожному каналу ознак за формулами (3), (5)–(7). При цьому в якості бажаного відгуку r використовувався гаусіан із параметром стандартного відхилення $\sigma = 0,1$. У якості матриці регуляризації w використовувалася матриця, що має великі значення скрізь за межами прямокутника, який задає положення об'єкта (рис. 3).

Великі значення у матриці w були прийняті рівними 100, а значення в межах прямокутника – 0,01. Перехід між малими і великими значеннями у w додатково згладжувався фільтром Гауса розміром 7×7 пікселів зі стандартним відхиленням $\sigma = 0,25$. Згладжування застосовувалось для того, щоб зробити границю області придушення при обчисленні фільтру плавною на випадок, якщо положення об'єкта задано неточно. В деяких експериментах область об'єкта у матриці w бралася більшою, ніж актуальний розмір об'єкта, що робилося для додаткового включення у фільтр невеликої кількості контекстно-фонові інформації.

У формулах (5), (8)–(10) застосовувались наступні параметри: початковий крок алгоритму $\rho^{(0)} = 1$; максимально допустимий крок $\rho_{\max} = 1000$; коефіцієнт зміни кроку $\beta = 30$. Кількість ітерацій при кожному обчисленні фільтру складала 2. При цьому початкові значення $G^{(0)}$ та $U^{(0)}$ для першої ітерації ADMM бралися із результатів обчислення фільтру для попереднього кадру (у роботі [22] цей підхід названо «теплим стартом»). Відзначимо, що таке наближення можливе, оскільки між суміжними кадрами об'єкт зазвичай змінюється досить слабо.

Для самого першого кадру $G^{(0)}$ та $U^{(0)}$ приймалися рівними 0.

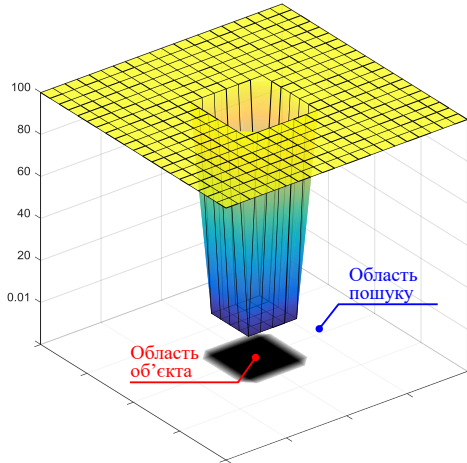


Рис. 3. Типовий вигляд матриці регуляризації w : темна область містить малі значення і відповідає ознакам об'єкта, що не мають придушуватись при обчисленні фільтра; світла область навпаки містить великі значення, що вказує на необхідність придушення саме цих складових у фільтрі, оскільки вони відповідають ознакам фону

Щоб ефективніше пристосовуватись до зміни зовнішнього виду об'єкта, як і в більшості подібних систем відслідковування [12, 14, 16, 25], на кожному кадрі після обчислення фільтра здійснювалось його експоненціальне усереднення:

$$G^{(t+1)} = (1 - \eta_f) \cdot G^{(t)} + \eta_f \cdot G, \quad (12)$$

де $G^{(t+1)}$ та $G^{(t)}$ – усереднені фільтри об'єкта для наступного та поточного кадрів відповідно; G – значення фільтра знайдене для розташування об'єкта на поточному кадрі; η_f – коефіцієнт експоненціальної фільтрації. В цій роботі використовувалось $\eta_f = 0,025$.

Також після кожного виявлення об'єкта, обчислювалась нормалізована гістограма нової області розташування об'єкта та виконувалось її оновлення згідно формули (11), при цьому застосовувався коефіцієнт $\eta_{HVP} = 0,125$.

Обчислення кореляційних відгуків при пошуку об'єкта здійснювалось для FHOГ ознак та ознак на основі зворотного проектування гістограм за формулою (8) незалежно, в результаті чого формувалось два відгуки $c_{HOГ}$ та c_{HVP} відповідно. Фінальний відгук обчислювався за формулою (10) із використанням коефіцієнту об'єднання $\gamma = 0,55 \times \frac{\max(c_{HOГ})}{\max(c_{HVP})}$. Параметр 0,55 було встановлено емпірично.

Для пристосування до зміни розміру об'єкта обчислення кореляції проводилось на $n_s = 3$ масштабах. Коефіцієнти масштабування при цьому визначались як: $s = 1,03^p$, де $p = \{-(n_s - 1)/2, \dots, +(n_s - 1)/2\}$.

В усіх експериментах область пошуку бралася квадратною та більшою за об'єкт приблизно в 4,25 рази і далі масштабувалася так, щоб в просторі ознак роздільна здатність каналів не перевищувала 38×38 пікселів.

Перед переходом до частотної області до зображення в просторі ознак поканально застосовувалось косинусне вікно фон Хана.

3.2 Оцінювання якості роботи методу CISRDCF

Оцінювання якості роботи створеної системи відслідковування виконувалось на тесті VOT Challenge VOT-RGBT2019 [13]. Цей тест включає в себе 60 анотованих відеопослідовностей з різними об'єктами, що синхронно зняті на камеру видимого діапазону та тепловізор. Роздільна здатність кадрів видимого та інфрачервоного діапазонів є однаковими для кожної відеопослідовності. Для оцінки якості відслідковування у тесті застосовується три показники: точність, надійність та очікуване усереднене перекриття [5, 8]. Всі вказані показники розраховуються на основі значення перекриття (overlap), під яким у тесті розуміється значення так званого «перетину-над-об'єднанням» (індекс Жакара), рівного $IoU = \frac{|r_t \cap r_{GT}|}{|r_t \cup r_{GT}|}$, де r_t – множина точок, яка задає положення об'єкту, знайдене системою відслідковування, r_{GT} – множина точок опорного (еталонного) положення об'єкту, тут $|\cdot|$ позначає кількість елементів множини [25]. Точність (A) – усереднене спочатку по всіх кадрах, а потім по всіх відеопослідовностях значення IoU на тих частинах відеопослідовностей, де відслідковування було визнано надійним. Надійність (R) – величина пропорційна середній кількості зривів відслідковування. При цьому вважається, що зрив відбувається тоді, коли перекриття між опорним та знайденим системою положеннями стає рівним нулю. Очікуване усереднене перекриття (EAO) – усереднене перекриття, очікуване для даної системи відслідковування на великій кількості фрагментів відеопослідовностей приведених до певної однакової довжини [5, 13]. Дана величина враховує збільшення дисперсії та зміщення значення усередненого перекриття для послідовностей різної довжини. Результати тесту впорядковуються саме за величиною EAO [13].

Результати оцінювання запропонованої системи відслідковування (CISRDCF) за методикою VOT-RGBT2019 наведені у табл. 1 [13]. Зауважимо, що у коді поданому на тест була виявлена помилка¹. При локалізації об'єкта шляхом розрахунку кореляції з DCF фільтром, ознаки двічі обчислювались із зображення видимого діапазону, а не з видимого та інфрачервоного діапазонів як передбачалось. Через це оцінка системи CISRDCF виявилась заниженою. Щоб виправити ситуацію, у даній роботі наводи-

¹Вихідний код доступний за посиланням: [CISRDCF-code-2019-06-09T12_56_48.961403.zip](https://github.com/sergey-ivanov/cisrDCF-code-2019-06-09T12_56_48.961403.zip)

тсь результат для скорегованої версії системи, що у табл. 1 позначена як CISRDCF*.

Табл. 1 Результати тестування запропонованої системи відслідковування (CISRDCF) на тесті VOT-RGBT2019 у порівнянні з іншими системами відслідковування

Система відслідковування	ЕАО	А	R
JMMAC	0.4826	0.6649	0.8211
SiamDW-T	0.3925	0.6158	0.7839
mfDiMP	0.3879	0.6019	0.8036
FSRPN	0.3553	0.6362	0.7069
MANet	0.3463	0.5823	0.7010
MPAT	0.3180	0.5723	0.7242
CISRDCF	0.2923	0.5215	0.6904
gesbtt	0.2896	0.6163	0.6350
CISRDCF*	0.3305	0.5607	0.7483
CISRDCF (1,25)	0.3247	0.5720	0.7128
CISRDCF (1,5)	0.3373	0.5728	0.7514

* виправлений варіант системи

Як видно з табл. 1, скорегована версія системи CISRDCF* за надійністю відслідковування (показником R) є четвертою та поступається лише деяким нейромережним системам (SiamDW-T, mfDiMP) спеціально навченим для роботи з інфрачервоним зображенням, а також системі JMMAC, що також працює на основі дискримінантного фільтра [10], однак додатково використовує ключові точки для стабілізації кадру та фільтр Калмана [13]. При цьому CISRDCF* за надійністю (показником R) випереджає FSRPN та MANet. Зауважимо, що всі наведені системи є більш обчислювально-складними, ніж CISRDCF*. Водночас, основним недоліком CISRDCF* є невисока точність локалізації об'єкта, яка обумовлена головним чином неточним оцінюванням масштабу. Це видно з прикладів відслідковування, наведених на рис. 4. Тим не менш на практиці така точність все ж може вважатися прийнятною.

В рамках даної публікації також було проведено додаткове тестування впливу обсягу контекстно-фонові інформації на якість роботи розглянутої системи відслідковування. Для проведення зазначеного тесту використовувались матриці регуляризації w (рис. 3), в яких область об'єкта (область, що не придушувалася за допомогою регуляризації при обчисленні фільтра) бралася в 1,25 та 1,5 рази більшою за актуальний розмір об'єкта. Випробування показали, що в такому випадку краще обмежитись використанням тільки ознак FHOГ. Отримані результати тестування наведені у двох останніх рядках табл. 1. З них, зокрема, видно, що при використанні збільшеної в 1,5 рази області об'єкта і лише ознак FHOГ вдається перевершити якість роботи звичайної версії CISRDCF* та на-

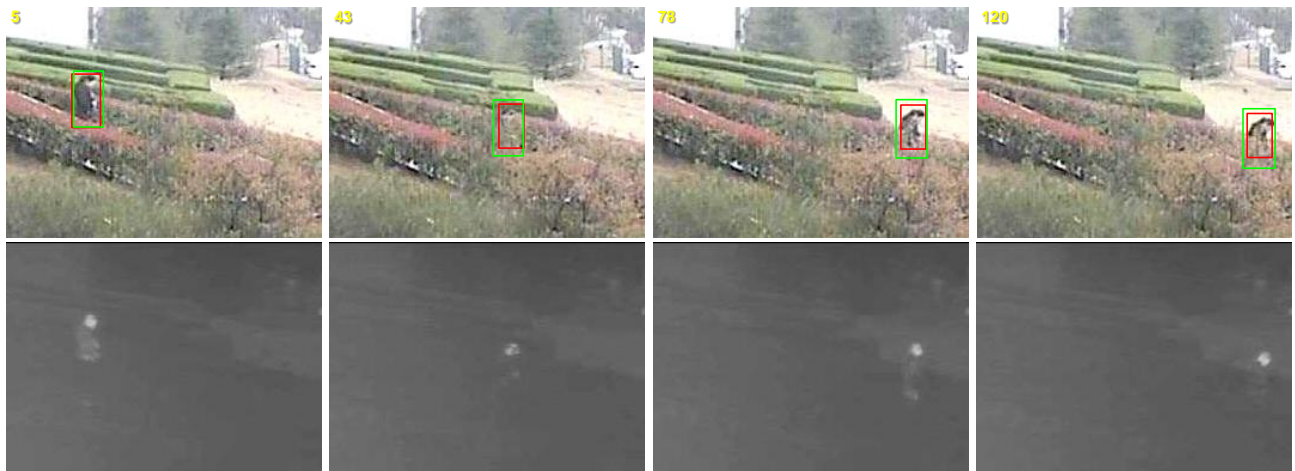
близитись за показником ЕАО до більш складної нейромережної системи MANet.

Висновки

В роботі запропоновано систему відслідковування об'єктів CISRDCF для мультиспектральних відео послідовностей. В основі даної системи лежить дискримінантний кореляційний фільтр (DCF), який обчислюється незалежно по каналам ознак за допомогою методу множників зі зміною напрямків (ADMM) та передбачає просторову регуляризацію для більш точного кодування об'єкта у фільтрі. Запропонована система передбачає можливість роботи із відеопослідовностями, кадри яких містять одночасно видимий та інфрачервоний канали. Вказана можливість досягається за рахунок використання одного і того ж набору ознак (ознак FHOГ та ознак на основі зворотного проектування гістограм) для каналів видимого та інфрачервоного діапазонів із подальшим об'єднанням отриманих ознак у єдиний тензор. На тесті VOT-RGBT2019 показано, що система CISRDCF за якістю відслідковування може конкурувати із більш складними рішеннями, в тому числі на основі нейронних мереж, а за показником надійності навіть перевершувати деякі з них. Крім того, встановлено, що при використанні лише ознак FHOГ та збільшенні обсягу контекстно-фонові інформації (збільшення області об'єкта у матриці регуляризації в 1,5 рази) вдається підвищити якість відслідковування в порівнянні з базовою реалізацією системи.

Перелік посилань

1. Multi-Sensor Image Fusion and Its Applications / Ed. by R. S. Blum, Z. Liu. – CRC Press, 2006. – 528 p. ISBN: 9780849334177.
2. Image Fusion and Its Applications / Ed. by Y. Zheng. – InTechOpen, 2011. – 242 p. ISBN: 9789533071824.
3. Гривачевський А. П. Підвищення інформативності мультиспектральних зображень шляхом мультимодального комплексування / А. П. Гривачевський, І. Н. Прудис // Вісник НТУУ «КПІ». Серія Радіотехніка, Радіоапаратобудування. – 2018. – № 73. – С. 40–49. DOI: 10.20535/RADAP.2018.73.40-49.
4. Maggio E., Cavallaro A. Video tracking. Theory and Practice. – John Wiley & Sons Ltd, 2011. – 294 p. ISBN: 9780470749647.
5. Kristan M. The Visual Object Tracking VOT2015 challenge results / M. Kristan, J. Matas, A. Leonardis, M. Felsberg та ін. // IEEE International Conference on Computer Vision Workshop (ICCVW). – 2015. – P. 564–586.
6. Kristan M. The Visual Object Tracking VOT2016 challenge results / M. Kristan, A. Leonardis, J. Matas та ін. // European Conference on Computer Vision – ECCV 2016 Workshops. ECCV 2016. Lecture Notes in Computer Science. – Vol. 9914. – 2016. – pp. 777–823.



(a)



(b)



(v)

Рис. 4. Приклад відслідковування різних об'єктів системою CISRDCF. На кожній відеопослідовності (a)–(v) показано видимий (зверху) та інфрачервоний (знизу) канали. На видимих каналах у верхньому лівому куті нанесено номери кадрів у відповідних відеопослідовностях, червоним прямокутником позначено еталонні положення об'єкту, а зеленим – визначені системою відслідковування

7. Kristan M. The visual object tracking VOT2017 challenge results / M. Kristan, A. Leonardis, J. Matas, M. Felsberg та ін. // ICCV2017 Workshops, Workshop on visual object tracking challenge. – 2017.
8. Kristan M. The sixth Visual Object Tracking VOT2018 challenge results / M. Kristan, A. Leonardis, J. Matas, M. Felsberg та ін. // Computer Vision – ECCV 2018 Workshops. ECCV 2018. Lecture Notes in Computer Science. – Vol. 11129. – 2018. – P. 3–53.
9. Nam H. Learning multi-domain convolutional neural networks for visual tracking / H. Nam, B. Han // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2016. – P. 4293–4302.
10. Danelljan M. ECO: Efficient convolution operators for tracking / M. Danelljan, G. Bhat, F. S. Khan, M. Felsberg // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – P. 6931–6939.
11. Bertinetto L. Fully-convolutional siamese networks for object tracking / L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, P. H. S. Torr // European Conference on Computer Vision – ECCV 2016 Workshops. ECCV 2016. Lecture Notes in Computer Science. – Vol. 9914. – 2016. – P. 850–865.
12. Henriques J. F. High-speed tracking with kernelized correlation filters / J. F. Henriques, R. Caseiro, P. Martins, J. Batista // IEEE Trans. Pattern Anal. Mach. Intell. (PAMI). – Vol. 37, Iss. 3. – 2015. – P. 583–596.
13. Kristan M. The Seventh Visual Object Tracking VOT2019 Challenge Results / M. Kristan, J. Matas, A. Leonardis, M. Felsberg та ін. // 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). – 2019.
14. Bolme D. S. Visual Object Tracking using Adaptive Correlation Filters / D. S. Bolme, R. J. Beveridge, B. A. Draper, Y. M. Lui // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2010. – P. 2544–2550.
15. Galoogahi H. K. Correlation Filters with Limited Boundaries / H. K. Galoogahi, T. Sim, S. Lucey // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2015. – P. 4630–4638.
16. Danelljan M. Learning spatially regularized correlation filters for visual tracking / M. Danelljan, G. Häger, F. Khan, M. Felsberg // IEEE International Conference on Computer Vision (ICCV). – 2015. – P. 4310–4318.
17. Galoogahi H. K. Learning background-aware correlation filters for visual tracking / H. K. Galoogahi, A. Fagg, S. Lucey // IEEE International Conference on Computer Vision (ICCV). – 2017. – P. 1144–1152.
18. Lukežič A. Discriminative correlation filter tracker with channel and spatial reliability / A. Lukežič, T. Vojtíř, L. Č. Zajc, J. Matas, M. Kristan // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – P. 6309–6318.
19. Danelljan M. Beyond Correlation Filters: Learning Continuous Convolution Operators for Visual Tracking / M. Danelljan, A. Robinson, F. S. Khan, M. Felsberg // 14th European Conference on Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science. – Vol. 9909. – 2016. – P. 472–488.
20. Feng L. Learning spatial-temporal regularized correlation filters for visual tracking / L. Feng, C. Tian, W. Zuo, L. Zhang, M.-H. Yang // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2018. – P. 4904–4913.
21. Varfolomeiev A. Channel-independent spatially regularized discriminative correlation filter for visual object tracking / A. Varfolomeiev // Journal of Real-Time Image Processing (RTIP). – 2020. – P. 1–11.
22. Boyd S. Distributed optimization and statistical learning via the Alternating Direction Method of Multipliers / S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein // Foundations and Trends in Machine Learning. – Vol. 3, Iss. 1. – 2010. – P. 1–122. DOI: 10.1561/22000000016.
23. Messerschmitt D. Stationary points of a real-valued function of a complex variable. Tech. Report, EECS, U.C. Berkeley. – 2006.
24. Felzenszwalb P. F. Object Detection with Discriminatively Trained Part Based Models / P. F. Felzenszwalb, R. B. Girshick, D. McAllester, D. Ramanan // IEEE Trans. Pattern Anal. Mach. Intell. (PAMI). – Vol. 32, Iss. 9. – 2010. – P. 1627–1645.
25. Bertinetto L. Staple: Complementary learners for real-time tracking / L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, P. H. S. Torr // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2016. – P. 1401–1409.
26. Varfolomeiev A. Modification of the KCF tracking method for implementation on embedded hardware platforms / A. Varfolomeiev, O. Lysenko // International Conference Radio Electronics & Info Communications (UkrMiCo), Kyiv. – 2016.

References

- [1] Blum R. S., Liu Z. (2006) *Multi-Sensor Image Fusion and Its Applications*. CRC Press, 528 p. ISBN: 9780849334177.
- [2] Zheng Y. (2011) *Image Fusion and Its Applications*. InTech Open, 242 p. ISBN: 9789533071824.
- [3] Hryvachevskiy A. P., Prudyus I. N. (2018) Enhancing the Informativeness of Multi-spectral Images by means of Multimodal Image Fusion. *Visnik NTUU KPI. Ser. Radiotekh. radioaparaturbuduv.*, Iss. 73, pp. 40–49. DOI: 10.20535/RADAP.2018.73.40-49.
- [4] Maggio E., Cavallaro A. (2011) *Video tracking. Theory and Practice*. John Wiley & Sons Ltd, 294 p. ISBN: 9780470749647.
- [5] Kristan M., Matas J., Leonardis A., Felsberg M. et al (2015) The Visual Object Tracking VOT2015 challenge results. *IEEE International Conference on Computer Vision Workshop (ICCVW)*, pp. 564–586. DOI: 10.1109/ICCVW.2015.79.
- [6] Kristan M., Leonardis A., Matas J. et al (2016) The Visual Object Tracking VOT2016 challenge results. *European Conference on Computer Vision – ECCV 2016 Workshops. ECCV 2016. Lecture Notes in Computer Science*, vol. 9914, pp. 777–823. DOI: 10.1007/978-3-319-48881-3_54.
- [7] Kristan M., Leonardis A., Matas J., Felsberg M. et al (2017) The visual object tracking VOT2017 challenge results. *ICCV2017 Workshops, Workshop on visual object tracking challenge*. DOI: 10.1109/ICCVW.2017.230.

- [8] Kristan M., Leonardis A., Matas J., Felsberg M. et al (2018) The sixth Visual Object Tracking VOT2018 challenge results. *Computer Vision – ECCV 2018 Workshops. ECCV 2018. Lecture Notes in Computer Science*, vol. 11129, pp. 3–53. DOI: 10.1007/978-3-030-11009-3_1.
- [9] Nam H., Han B. (2016) Learning multi-domain convolutional neural networks for visual tracking. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4293–4302. DOI: 10.1109/CVPR.2016.465.
- [10] Danelljan M., Bhat G., Khan F. S., Felsberg M. (2017) ECO: Efficient convolution operators for tracking. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6931–6939. DOI: 10.1109/CVPR.2017.733.
- [11] 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 – ECCV 2016 Workshops. ECCV 2016. Lecture Notes in Computer Science*, Vol. 9914, pp. 850–865. DOI: 10.1007/978-3-319-48881-3_56.
- [12] Henriques J. F., Caseiro R., Martins P., Batista J. (2015) High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)*, Vol. 37, Iss. 3, pp. 583–596. DOI: 10.1109/TPAMI.2014.2345390.
- [13] Kristan M., Matas J., Leonardis A., Felsberg M. et al (2019) The Seventh Visual Object Tracking VOT2019 Challenge Results. *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. DOI: 10.1109/ICCVW.2019.00276.
- [14] Bolme D. S., Beveridge R. J., Draper B. A., Lui Y. M. (2010) Visual Object Tracking using Adaptive Correlation Filters. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2544–2550. DOI: 10.1109/CVPR.2010.5539960.
- [15] Galoogahi H. K., Sim T., Lucey S. (2015) Correlation Filters with Limited Boundaries. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4630–4638. DOI: 10.1109/CVPR.2015.7299094.
- [16] Danelljan M., Häger G., Khan F., Felsberg M. (2015) Learning spatially regularized correlation filters for visual tracking. *IEEE International Conference on Computer Vision (ICCV)*, pp. 4310–4318. DOI: 10.1109/ICCV.2015.490.
- [17] Galoogahi H. K., Fagg A., Lucey S. (2017) Learning background-aware correlation filters for visual tracking. *IEEE International Conference on Computer Vision (ICCV)*, pp. 1144–1152. DOI: 10.1109/ICCV.2017.129.
- [18] Lukežič A., Vojří T., Zajc L. Č., Matas J., Kristan M. (2017) Discriminative correlation filter tracker with channel and spatial reliability. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6309–6318. DOI: 10.1007/s11263-017-1061-3.
- [19] Danelljan M., Robinson A., Khan F. S. and Felsberg M. (2016) Beyond Correlation Filters: Learning Continuous Convolution Operators for Visual Tracking. *14th European Conference on Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science*, Vol. 9909, pp. 472–488. DOI: 10.1007/978-3-319-46454-1_29.
- [20] Li F., Tian C., Zuo W., Zhang Lei, Yang M.-H. (2018) Learning spatial-temporal regularized correlation filters for visual tracking. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4904–4913. DOI: 10.1109/CVPR.2018.00515.
- [21] Varfolomeiev A. (2020) Channel-independent spatially regularized discriminative correlation filter for visual object tracking. *Journal of Real-Time Image Processing (RTIP)*, DOI: 10.1007/s11554-020-00967-y.
- [22] Boyd S., Parikh N., Chu E, Peleato B., Eckstein J. (2010) Distributed optimization and statistical learning via the Alternating Direction Method of Multipliers. *Foundations and Trends in Machine Learning*, Vol. 3, Iss. 1, pp. 1–122. DOI: 10.1561/22000000016.
- [23] Messerschmitt D. (2006) Stationary points of a real-valued function of a complex variable. *Tech. Report, EECS, U.C. Berkeley*.
- [24] Felzenszwalb P. F., Girshick R. B., McAllester D., Ramanan D. (2010) Object Detection with Discriminatively Trained Part Based Models. *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)*, Vol. 32, Iss. 9, pp. 1627–1645. DOI: 10.1109/TPAMI.2009.167.
- [25] Bertinetto L., Valmadre J., Golodetz S., Miksik O., Torr P. H. S. (2016) Staple: Complementary learners for real-time tracking. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1401–1409. DOI: 10.1109/CVPR.2016.156.
- [26] Varfolomeiev A., Lysenko O. (2016) Modification of the KCF tracking method for implementation on embedded hardware platforms. *International Conference Radio Electronics & Info Communications (UkrMiCo)* DOI: 10.1109/UkrMiCo.2016.7739644.

Система отслеживания объектов на видеопоследовательностях в инфракрасном и видимом диапазонах на основе канально-независимого пространственно-регуляризованного дискриминантного корреляционного фильтра

Варфоломеев А. Ю., Короткий Е. В.

Представлен метод отслеживания объектов для работы на видеопоследовательностях, содержащих мультиспектральную информацию. Рассмотрены возможные способы объединения мультиспектральной информации при решении задачи отслеживания и обосновано использование подхода на основе объединения на уровне вычисления признаков. С точки зрения на компромисс в отношении качества и быстродействия, задачу отслеживания предложено решать с помощью дискриминантных корреляционных фильтров (DCF).

Разработан метод вычисления канально-независимого дискриминантного корреляционного фильтра с пространственной регуляризацией, который основан на использовании метода множителей меняющихся направлений (ADMM). Вычисление DCF фильтра и локализацию объекта при этом предложено выполнять в специальном пространстве признаков, которое использует многоканальные признаки FHOГ и признаки, основанные на обратном проектировании взвешенной гистограммы объекта. Указанные признаки предлагается вычислять для каждого канала соответствующего кадра мультиспектральной видеопоследовательности и объединять получаемые признаки в единственный тензор объединенного пространства признаков.

На тесте VOT Challenge RGBT2019 показано, что реализация предложенного метода по качеству отслеживания может конкурировать с более сложными решениями, в том числе, основанными на нейросетевых технологиях. В ходе экспериментов также установлено, что увеличение объема контекстно-фоновой информации позволяет несколько увеличить качество отслеживания в сравнении с базовой реализацией предложенного метода, даже в случае использования только признаков FHOG.

Ключевые слова: визуальное отслеживание объектов; мультиспектральные изображения; дискриминантные корреляционные фильтры (DCF); метод множителей меняющих направление (ADMM)

The Object Tracker for Infrared and Visual Bands based on Channel-Independent Spatially-Regularized Discriminative Correlation Filter

Varfolomeiev A. Y., Korotkyi I. V.

The method of visual object tracking intended for the application on multispectral video sequences is considered.

Introduction. The possible techniques of multispectral information fusion for visual object tracking are considered and the use of feature based fusion approach is justified. The tracker is suggested to be implemented using the discriminative correlation filters (DCF), since this approach

is known to provide the compromise in terms of tracking quality and speed.

Theoretic results. The method for channel-independent discriminative correlation filter with spatial regularization calculation based on the use of alternating direction method of multipliers (ADMM) is proposed. The calculation of DCF filter and the object localization is suggested to be performed in special feature space, which employs the multichannel FHOG features and the features that are based on the backprojection of object weighted histogram. In particular, we propose to calculate the mentioned features for each channel of the respective frame of the multispectral video sequence with subsequent concatenation of obtained features into a single tensor, which forms the joint feature space.

Conclusions. Using the VOT Challenge RGBT2019 subchallenge, it was shown that the implementation of suggested method is competitive in terms of tracking robustness with more sophisticated approaches, including the ones that are based on the convolutional neural networks. During the experiments, it was additionally established that the increasing of context-background information gives slight tracking quality improvement compared to the basic method implementation, even when only FHOG features are used.

Key words: visual object tracking; multispectral images; discriminative correlation filters (DCF); alternating direction method of multipliers (ADMM)