Борсук Василь Юліанович

Національний університет “Львівська політехніка”, Львів

ORCID 0009-0009-1063-1010

**Відстеження об'єктів у відеозображеннях в реальному часі: сучасні підходи, виклики та перспективи розвитку**

Відстеження об'єктів у реальному часі є критично важливою технологією, яка знаходить застосування у багатьох сферах, сприяючи підвищенню безпеки, ефективності та інноваційності. У автономних транспортних засобах ця технологія забезпечує розпізнавання та прогнозування руху пішоходів і перешкод, що критично для безпечної навігації. У системах відеоспостереження вона дозволяє автоматично виявляти підозрілі дії, підвищуючи рівень громадської безпеки. У доповненій реальності відстеження об'єктів забезпечує інтерактивність, необхідну для навчання, розваг і дизайну. У медицині технології реального часу допомагають підвищити точність діагностики та хірургічних втручань, тоді як у робототехніці, спорті та екологічному моніторингу вони оптимізують процеси і сприяють прогресу в дослідженнях. Ця універсальність робить відстеження об'єктів ключовим елементом сучасних інтелектуальних систем.

Відстеження об'єктів на відео супроводжується низкою складнощів, зумовлених динамікою середовища та технічними обмеженнями. Однією з основних проблем є часткове або повне перекриття об'єктів, коли один об'єкт закриває інший, що може призводити до втрати або плутанини. Інша важлива складність – швидкий рух об'єктів, який спричиняє розмиття зображення та ускладнює точне визначення їхнього положення. Зміни зовнішнього вигляду об'єктів через варіації ракурсу, освітлення чи деформацію також створюють труднощі для алгоритмів, які залежать від візуальних ознак. Додаткові виклики включають складні умови освітлення, динамічний фон, зміну масштабу та перспективи об'єктів.

Водночас реальноважливою є оптимізація моделей для роботи на пристроях із обмеженими обчислювальними ресурсами, як-от дрони чи мобільні пристрої. Такі проблеми стимулюють розробку адаптивних алгоритмів, інтеграцію мультисенсорних даних та використання глибокого навчання для підвищення точності й надійності трекінгу в реальному часі.

Класичні методи відстеження об'єктів, такі як KLT-трекер [1] та алгоритми оптичного потоку, базуються на аналізі візуальних ознак і математичних моделях. KLT-трекер використовує ключові точки зображення для їх відстеження між кадрами, але є чутливим до змін масштабу, освітлення та перекриття. Методи оптичного потоку аналізують рух пікселів між кадрами, забезпечуючи точність у локальних рухах, але схильні до помилок при швидкому русі або шумі. Гібридні методи поєднують різні техніки, наприклад, KLT-трекер з оптичним потоком або використання фільтрів Калмана [2] для прогнозування положення об'єкта. Такі підходи покращують стійкість до втрати треку, але часто ускладнюють реалізацію та вимагають ретельної оптимізації параметрів.

Основними проблемами класичних методів є чутливість до змін середовища, нестабільність при складних рухах, низька ефективність для багатьох об'єктів і відсутність адаптивності до нових умов. Це обмежує їхню універсальність, особливо в складних сценаріях реального часу.

Хоча класичні методи є швидкими та легкими у впровадженні, їхні обмеження стимулюють перехід до сучасних підходів, таких як глибоке навчання, які краще адаптуються до складних умов і масштабних завдань.

Сучасні підходи до відстеження об'єктів у реальному часі активно використовують нейронні мережі, що дозволяє значно підвищити точність і адаптивність алгоритмів. Одним із популярних методів є застосування сіамських нейронних мереж (Siamese Networks) — архітектури, яка використовує дві ідентичні нейронні мережі з однаковими вагами для порівняння зразка (об'єкта) з поточним кадром. Цей підхід дозволяє моделі визначати, чи відповідає об'єкт у кадрі заданому шаблону, ґрунтуючись на подібності їхніх ознак.

Сіамські нейронні мережі добре працюють у задачах однооб'єктного трекінгу, таких як відстеження рухомих об'єктів у відео, де об'єкт може змінювати свою форму, масштаб чи орієнтацію. Наприклад, алгоритми на основі таких мереж, як SiamRPN [4], використовують шаблон об'єкта для визначення його місця розташування в наступних кадрах, забезпечуючи високу швидкість і точність.

Основною перевагою підходу є його стійкість до змін середовища, таких як варіації освітлення, перекриття та деформації об'єкта. Однак ці моделі можуть вимагати значних обчислювальних ресурсів, що стимулює розробку оптимізованих версій для роботи в реальному часі на пристроях із обмеженими ресурсами, наприклад дронах або мобільних телефонах.

LightTrack [5] — це сучасний підхід до відстеження об'єктів, що використовує сіамську нейронну мережу, оптимізовану для швидкості та ефективності. Основна ідея полягає у застосуванні алгоритмів пошуку нейронної архітектури (Neural Architecture Search) для автоматичного проєктування легкої та продуктивної архітектури трекера. На відміну від традиційних алгоритмів, LightTrack забезпечує високу продуктивність із суттєвим зменшенням обчислювальних ресурсів, що робить його придатним для використання на пристроях із обмеженими ресурсами, таких як мобільні чипсети.

Експерименти показують, що LightTrack перевершує найсучасніші трекери, такі як SiamRPN, за точністю, використовуючи значно менше параметрів і обчислювальних ресурсів. Наприклад, на платформі Snapdragon 845 Adreno GPU LightTrack працює в 12 разів швидше за SiamRPN, використовуючи в 13 разів менше параметрів і в 38 разів менше обчислень. Цей підхід дозволяє значно зменшити розрив між академічними моделями та їхньою практичною реалізацією, відкриваючи нові можливості для застосування відстеження об'єктів у реальному часі в промислових умовах.

Сімейство сіамських нейронних мереж FEAR [6] поєднує компактну нейронну мережу та вводить нову двошаблонну репрезентацію обʼєкта. Основною інновацією є використання двох шаблонів: статичного, який запобігає дрейфу і зберігає початковий вигляд об'єкта, і динамічного, що адаптується до поточних умов. FEAR опрацьовує 205 кадрів в секунду на iPhone 11, що в 4.2 рази швидше за LightTrack, з високою точністю на бенчмарках. Цей трекер є одним з найбільш швидких і точних у своєму класі, а також енергоефективним.

LightTrack та FEAR використовують згорткові нейронні мережі, які поступово починають витіснятися трансформерами. Трансформери значно покращують моделювання візуальних даних завдяки здатності ефективно обробляти довгострокові залежності в зображеннях. Вони можуть захоплювати глобальні контексти та взаємодії між різними частинами зображення, що є складним завданням для згорткових мереж, які зазвичай обробляють зображення локально. Трансформери дозволяють більш точно моделювати складні сцени з різноманітними об'єктами та взаємодіями, що підвищує їхню ефективність у відстеженні. Однак їхні моделі часто мають високу обчислювальну складність, що обмежує їх використання на пристроях з обмеженими ресурсами.

HiT [7] — це нова сім'я ефективних трекерів, заснованих на трансформерах, яка вирішує проблему низької швидкості трансформерів, зберігаючи високу продуктивність. Ключовою інновацією є Bridge Module, який поєднує глибокі особливості з поверхневими велико-роздільними ознаками. Також, використовується нова техніка дво-образного кодування позицій, яка одночасно кодує інформацію про позицію пошукової області та шаблонів зображень. HiT досягає високої швидкості і конкурентоспроможних результатів на бенчмарках, перевершуючи всі попередні ефективні трекери. Цей підхід поєднує швидкість і точність, роблячи HiT актуальним для застосувань на пристроях з обмеженими ресурсами.

Також, недавній метод ABTrack [8] пропонує ще додаткові методи оптимізації трансформерів. Основна ідея ABTrack полягає в адаптивному пропусканні трансформерних блоків, оскільки не всі семантичні ознаки та відносини мають однаковий вплив на точність відстеження на різних рівнях абстракції. Залежно від характеристик цілі та сцени, де вона знаходиться, деякі ознаки можуть бути менш важливими, і їх пропуск не вплине на результат. Для цього введено Bypass Decision Module (BDM), який визначає, чи має бути пропущений певний трансформерний блок, що дозволяє адаптивно спрощувати архітектуру і прискорювати процес відстеження.

Відстеження об'єктів у відеозображеннях у реальному часі є важливою технологією для численних застосувань, таких як автономні транспортні засоби, системи відеоспостереження, доповнена реальність та медична діагностика. Хоча класичні методи відстеження, як KLT-трекер та алгоритми оптичного потоку, мають обмеження в контексті швидкості та адаптивності, сучасні підходи на основі нейронних мереж, зокрема трансформерів, значно підвищують точність і стійкість до змін середовища. Проте, проблема високих обчислювальних вимог залишається актуальною. Останні досягнення, такі як LightTrack, FEAR, HiT та ABTrack, сприяють значному покращенню швидкості та ефективності моделей, що дозволяє їх успішно застосовувати на пристроях з обмеженими ресурсами. Також, важливим напрямком досліджень є розробка ефективніших методів репрезентації обʼєктів, по прикладу двошаблонної репрезентації FEAR. Подальші покращення в цих напрямках обіцяють подолати існуючі бар'єри, наближаючи наукові розробки до реальних промислових застосувань у відстеженні об'єктів у реальному часі.

**Література**

1. Lucas, B., & Kanade, T. (1981). An Iterative Image Registration Technique with an Application to Stereo Vision. In Proceedings of the 7th International Joint Conference on Artificial Intelligence-Volume 2 (pp. 674-679). San Francisco, CA: Morgan Kaufmann Publishers Inc.
2. Kalman, Rudolph Emil. "A new approach to linear filtering and prediction problems." (1960): 35-45.
3. Koch, Gregory, Richard Zemel, and Ruslan Salakhutdinov. "Siamese neural networks for one-shot image recognition." In ICML deep learning workshop, vol. 2, no. 1, pp. 1-30. 2015.
4. Li, Bo, Junjie Yan, Wei Wu, Zheng Zhu, and Xiaolin Hu. "High performance visual tracking with siamese region proposal network." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8971-8980. 2018.
5. Yan, Bin, Houwen Peng, Kan Wu, Dong Wang, Jianlong Fu, and Huchuan Lu. "Lighttrack: Finding lightweight neural networks for object tracking via one-shot architecture search." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15180-15189. 2021.
6. Borsuk, Vasyl, Roman Vei, Orest Kupyn, Tetiana Martyniuk, Igor Krashenyi, and Jiři Matas. "FEAR: Fast, efficient, accurate and robust visual tracker." In European Conference on Computer Vision, pp. 644-663. Cham: Springer Nature Switzerland, 2022.
7. Kang, Ben, Xin Chen, Dong Wang, Houwen Peng, and Huchuan Lu. "Exploring lightweight hierarchical vision transformers for efficient visual tracking." In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9612-9621. 2023.
8. Yang, Xiangyang, Dan Zeng, Xucheng Wang, You Wu, Hengzhou Ye, Qijun Zhao, and Shuiwang Li. "Adaptively bypassing vision transformer blocks for efficient visual tracking." Pattern Recognition (2024): 111278.