

Д-р. ф., асистент Сергієнко П.А., магістрант Філіпенко Д.О.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

СПОСІБ РОЗПІЗНАВАННЯ ПОВЕДІНКОВИХ ПАТЕРНІВ НА ОСНОВІ ПОПЕРЕДНЬО ВІДОМИХ ОБРАЗІВ

Abstract

Serhiienko Pavlo, assistant, PhD; Filipenko Danylo, student

A Method for Recognizing Behavioral Patterns Based on Predefined Templates

This paper proposes a method for recognizing human behavioral patterns in video using predefined action templates. The approach employs YOLOv8n for detection, DeepSORT for tracking, and bidirectional LSTM to match temporal sequences against 10 known behavioral templates. The dataset includes 700 videos (320x240). Key optimizations: feature caching and frame skipping. Achieved 89.4% pattern recognition accuracy with 2.5x speedup. Suitable for real-time surveillance.

Вступ

Автоматизоване розпізнавання поведінкових патернів є важливим завданням сучасних систем відеоспостереження, зокрема у сфері безпеки, моніторингу громадських просторів та промислових об'єктів. Складність проблеми зумовлена необхідністю аналізу просторово-часових залежностей руху об'єктів у реальному часі. Традиційні підходи, що використовують 3D-згорткові мережі, забезпечують високу точність, проте потребують значних обчислювальних ресурсів, що ускладнює їх застосування на пристроях із обмеженою продуктивністю [1, 2].

У даній роботі запропоновано спосіб розпізнавання десяти поведінкових патернів на основі еталонних послідовностей ознак, сформованих під час навчання, що дозволяє суттєво зменшити обчислювальні витрати та забезпечити роботу в умовах, наближених до реального часу.

У сучасних умовах швидкого зростання кількості камер відеоспостереження особливої актуальності набуває розроблення методів і способів, здатних працювати на відео малої роздільності та на обмежених обчислювальних ресурсах. Значна частина інфраструктурних систем використовує компактні камери, вбудовані процесори або хмарні обчислення з обмеженою пропускною здатністю каналів зв'язку. Тому

постає потреба у способах, що забезпечують прийнятний компроміс між точністю та швидкістю. Саме цій проблемі присвячена дана робота.

Постановка задачі

Метою роботи є підвищення ефективності розпізнавання поведінкових патернів шляхом поєднання детекції, трекінгу та шаблонного зіставлення ознак об'єктів руху. Завданням є розроблення способу, який забезпечить точність класифікації не нижче 85%, зменшення часу обробки відеопослідовності у 2–3 рази порівняно з базовими підходами, а також можливість роботи з відео малої роздільності. Додатковою вимогою є забезпечення стійкості роботи способу при наявності шумів у вхідному відео, частковому перекритті об'єктів та варіаціях тривалості поведінкових фрагментів. Також важливою є можливість масштабування способу на більшу кількість класів без суттєвого зростання обчислювальних витрат. У межах роботи здійснено вибір архітектурних рішень, які дозволяють досягти зазначених цілей.

Термінологія

Поведінковий патерн – узагальнена послідовність дій чи рухів об'єкта, що характеризує певний тип поведінки.

Еталонна послідовність – записана під час навчання послідовність ознак, яка використовується як шаблон для зіставлення.

Трек — послідовність детекцій одного об'єкта, об'єднаних трекером у єдину траєкторію.

Ознаковий вектор — набір числових параметрів, що описують стан або рух об'єкта в конкретний момент часу.

Аналіз існуючих способів

Для аналізу відеоповедінки використовують два основні підходи: генеративні моделі, такі як SlowFast, та шаблонні способи, що поєднують детектори об'єктів із рекурентними мережами [3]. У роботі [4] продемонстровано зіставлення траєкторій із шаблонами, проте відсутність механізму кешування призводить до повторної обробки кадрів детектором YOLO. У запропонованому підході цю проблему усунуто через попереднє формування та збереження еталонних ознак, що дозволяє зменшити затрати часу на інференс.

Значна частина сучасних робіт у сфері розпізнавання дій використовує глибинні архітектури, такі як I3D, TSM або конструкції на основі Transformer-моделей. Попри високу точність, вони мають суттєву обчислювальну складність, потребують великих датасетів та оптимально працюють на відео високої роздільності [5]. Це створює бар'єри для їхнього

практичного застосування у задачах моніторингу, де важливою є оперативність обробки та обмежена якість відеопотоку. Таким чином, постає потреба у способах, здатних працювати на компактних даних і водночас забезпечувати прийнятний рівень точності.

Пропонований спосіб

Запропоновано спосіб розпізнавання поведінкових патернів, який полягає у зіставленні траєкторій об'єктів із попередньо сформованими еталонними послідовностями ознак, отриманими на основі детекції YOLOv8n та трекінгу DeepSORT. Від існуючих він відрізняється використанням попереднього кешування ознак та фіксованої структури ознакових послідовностей, що усуває повторну детекцію кадрів і зменшує обчислювальні витрати. Це дозволяє забезпечити прискорення обробки у 2–3 рази та підвищити стабільність класифікації поведінкових патернів на відео малої роздільності.

Для кожного навчального відео спочатку виконується детекція об'єктів за допомогою YOLOv8n та побудова треків DeepSORT. На основі координат, розмірів та динамічних характеристик руху об'єктів формуються ознакові послідовності, які описують траєкторію впродовж фрагмента з 20 кадрів. У формуванні ознак використовуються значення (x_1 , y_1 , x_2 , y_2), координати центру об'єкта, його ширина, висота та елементи оптичного потоку. Після нормалізації отримані дані агрегуються у фіксовані послідовності та кешуються у файлах формату *.pkl*, що дозволяє уникнути повторної обробки відео під час інференсу і значно скорочує обчислювальні витрати.

Під час застосування запропонованого способу кожне нове відео обробляється один раз: на вході формуються детекції та треки, після чого відповідна послідовність ознак зіставляється з еталонними шаблонами. Для цього використовується архітектура LSTM, здатна виявляти просторово-часові залежності у русі об'єкта. Спосіб приймає на вхід матрицю розміру 20×20 та генерує ймовірність відповідності траєкторії одному з десяти попередньо сформованих патернів.

Реалізація запропонованого способу передбачає використання LSTM-архітектури, що включає два рекурентні шари розмірністю 64 та 128 одиниць з подальшим повнозв'язним шаром, який виконує класифікацію за допомогою функції softmax. Для забезпечення можливості роботи в режимі, близькому до реального часу, реалізовано кілька оптимізацій: кешування еталонних ознак, пропуск окремих кадрів без втрати якості розпізнавання та використання багатопоточності під час попередньої обробки. Завдяки цим рішенням запропонований спосіб забезпечує точність розпізнавання 89,4%

на відео роздільності 320×240 та демонструє прискорення обробки приблизно у три рази порівняно з базовим підходом.

Під час реалізації запропонованого способу було приділено особливу увагу нормалізації ознак, що дозволило зменшити варіативність між різними відеофрагментами. Для траєкторій різної довжини застосовано операції доповнення або скорочення послідовностей до фіксованого розміру, що забезпечило коректну роботу рекурентної мережі. Важливим етапом є також фільтрація шумових детекцій та видалення коротких треків, які не містять достатнього обсягу інформації для подальшого аналізу. Такий підхід підвищує стабільність способу та позитивно впливає на точність класифікації.

Результати експериментальних досліджень

Запропонований спосіб було реалізовано на мові Python. Для оцінювання точності застосовано датасет із 700 відеофрагментів десяти поведінкових патернів роздільністю 320×240 пікселів. Модель показала середню точність класифікації 89,4%, що перевищує базовий шаблонний підхід без кешування ознак на 8,2%. Завдяки попередньому формуванню та збереженню еталонних послідовностей, час обробки одного відео скоротився з 48,2 с до 15,7 с, що забезпечило прискорення у 3,07 рази. Додаткові оптимізації, такі як пропуск кадрів, також позитивно вплинули на швидкодію без суттєвого погіршення точності. Отримані результати підтверджують ефективність запропонованого способу та перспективність його використання у системах відеоаналітики реального часу.

Висновки

Розроблений спосіб розпізнавання поведінкових патернів полягає у зіставленні траєкторій об'єктів із попередньо сформованими еталонними послідовностями ознак, сформованими за допомогою детекції YOLOv8n та трекінгу DeepSORT. Від існуючих рішень він відрізняється застосуванням кешування ознак та уніфікованої структури ознакових векторів, що дозволяє уникнути повторної обробки кадрів і суттєво знизити обчислювальні витрати. Це дало можливість досягти трьохразового прискорення обробки відео та підвищити стабільність класифікації на відео малої роздільності, що підтверджено експериментальними дослідженнями. Запропонований підхід формує основу для подальшого розвитку способу, зокрема для інтеграції компактних згорткових архітектур на кшталт ResNet18, адаптації до вбудованих обчислювальних платформ та розширення на задачі розпізнавання взаємодій між кількома об'єктами.

Література

1. Carreira J., Zisserman A. Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset // CVPR. – 2017. – P. 4724-4733. DOI: 10.1109/CVPR.2017.502.
2. Feichtenhofer C., Fan H., Malik J., He K. SlowFast Networks for Video Recognition // ICCV. – 2019. – P. 6202-6211. DOI: 10.1109/ICCV.2019.00630.
3. Wang L., Xiong Y., Wang Z., Qiao Y., Lin D., Tang X., Van Gool L. Temporal Segment Networks for Action Recognition in Videos // TPAMI. – 2018. – P. 2740-2755. DOI: 10.1109/TPAMI.2018.2868668.
4. Donahue J., Hendricks L. A., Rohrbach M., Venugopalan S., Guadarrama S., Saenko K., Darrell T. Long-term Recurrent Convolutional Networks for Visual Recognition and Description // TPAMI. – 2017. – P. 677–691. DOI: 10.1109/TPAMI.2016.2599174.
5. Wojke N., Bewley A. Deep Cosine Metric Learning for Person Re-Identification // WACV. – 2018. – P. 748–756. DOI: 10.1109/WACV.2018.00087.