

МОДЕЛІ ТА АЛГОРИТМИ СЕМАНТИЧНОГО АНАЛІЗУ ВІДЕОДАНИХ ДЛЯ ІНТЕЛЕКТУАЛЬНИХ ІНФОРМАЦІЙНИХ СИСТЕМ

Макаров Д.С.

Кафедра Інформатики,
Харківський національний університет
радіоелектроніки,

Україна.

E-mail: dmytro.makarov@nure.ua

Abstract

Semantic analysis of video data is a key component in the development of intelligent information systems. It allows for the extraction of meaningful insights and structured knowledge from large volumes of dynamic video streams. With the widespread use of surveillance systems, autonomous vehicles, telemedicine platforms, and smart environments, the demand for accurate and real-time video understanding is rapidly growing. This paper presents a comprehensive overview of models and algorithms used for semantic video analysis. It focuses on object detection and tracking, action and activity recognition, multimodal data fusion, and contextual scene understanding. Special attention is paid to deep learning approaches, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformer architectures. Challenges such as scalability, ambiguity of actions, and limited annotated datasets are discussed. The proposed direction of development emphasizes multimodal processing, weakly supervised learning, and domain adaptation for complex real-world applications.

Зі зростанням обсягів відеоданих, що генеруються у сфері безпеки, транспорту, медицини, торгівлі та цифрових медіа, виникає потреба в автоматизованих методах семантичного аналізу, які забезпечують розпізнавання, класифікацію та інтерпретацію візуальної інформації у динамічному середовищі [1, 2].

Семантичний аналіз відео передбачає виявлення об'єктів, визначення дій та активностей, інтерпретацію сцени та її контексту, а також виявлення взаємозв'язків між різними об'єктами й подіями у часовій послідовності [3]. Такий підхід є ключовим для побудови інтелектуальних систем спостереження, автономного управління та медичних рішень, де точність і швидкість прийняття рішень критично важливі.

Ключовим компонентом аналізу є моделі обробки зображень на основі глибоких згорткових нейронних мереж (CNN), які застосовуються для виявлення та класифікації об'єктів у відеопотоці. Найпоширенішими є архітектури YOLO (You Only Look Once), SSD (Single Shot Detector) та Faster R-CNN, що демонструють високу точність при детекції об'єктів у реальному часі [2].

Для розпізнавання дій та активностей у відео використовуються рекурентні нейронні мережі, зокрема LSTM (Long Short-Term Memory) та GRU (Gated Recurrent Units), які дозволяють враховувати часову динаміку та взаємозалежності між кадрами. Гібридні моделі, що поєднують CNN і RNN, забезпечують ефективну обробку як просторових, так і часових ознак [3].

Останнім часом активно розвиваються трансформерні архітектури (наприклад, ViViT, TimeSformer), які здатні обробляти послідовності кадрів без втрати глобального контексту. Такі моделі демонструють високу ефективність при аналізі складних сцен із багатьма об'єктами та різними типами взаємодій [4].

Особливу увагу дослідники приділяють інтеграції мультимодальних даних — зображень, аудіо та тексту (наприклад, субтитрів або метаданих) — для більш глибокого розуміння подій у відео [4].

Об'єднання візуальних та невербальних сигналів дозволяє значно підвищити точність інтерпретації складних ситуацій, таких як діагностика на основі медичного відео чи аналіз поведінки у натовпі.

Серед сучасних викликів – необхідність масштабування алгоритмів для обробки великих обсягів відеоданих, забезпечення роботи в реальному часі, подолання неоднозначності інтерпретації дій, а також обмеженість розмічених навчальних вибірок [2, 4]. У відповідь на ці виклики дослідники спрямовують зусилля на створення моделей, здатних навчатися без жорсткої залежності від розмічених даних, адаптуватися до нових доменів, ефективно обробляти інформацію на периферії мережі, а також бути стійкими до спотворень і змін зовнішнього середовища [1, 3].

Семантичний аналіз відео вже сьогодні активно використовується у сфері відеоспостереження – для виявлення загроз, аналізу поведінки та прогнозування небезпечних ситуацій [3]; у медицині – для моніторингу стану пацієнтів, аналізу хірургічних втручань та оцінки функціонального стану [4]; у транспорті – для виявлення перешкод, розпізнавання дорожніх знаків і планування руху автономних транспортних засобів [4]; а також у розважальній сфері, зокрема в контексті персоналізованого відеоконтенту, доповненої та віртуальної реальності [1].

Крім технічних аспектів, важливо враховувати й етичні виклики, пов'язані з конфіденційністю, обробкою біометричних даних та забезпеченням прозорості алгоритмів. Ці питання особливо актуальні у сферах відеоспостереження та медицини, де точність не повинна суперечити правам користувачів.

Моделі та алгоритми семантичного аналізу відеоданих стають фундаментальним елементом побудови інтелектуальних інформаційних систем. Високий рівень складності оброблюваної інформації, потреба у швидкості реагування та багатовимірність контексту подій вимагають застосування складних і водночас адаптивних підходів. Подальший розвиток у цій сфері має ґрунтуватися на поєднанні глибокого навчання, мультимодального аналізу та методів навчання з обмеженою розміткою. Це дозволить підвищити точність і гнучкість систем у реальних умовах, особливо у критично важливих сферах, таких як медицина, безпека та автономний транспорт.

Крім того, перспективним є створення енергоефективних архітектур для роботи на пристроях з обмеженими ресурсами та розвиток алгоритмів, здатних до самонавчання в умовах змінного середовища. Окрему увагу варто приділити розробці відкритих датасетів та інструментів анотації, що сприятимуть пришвидшенню дослідницьких процесів і забезпечать репродуктивність експериментів. Також важливо впроваджувати етичні принципи в моделі відеоаналітики, зокрема щодо захисту приватності та запобігання упередженням. Варто активізувати співпрацю між академічною спільнотою та індустрією задля трансферу технологій та швидшої імплементації нових рішень. У майбутньому можлива поява автономних відеоаналітичних систем, здатних до контекстуального мислення та самопояснення прийнятих рішень. Інтеграція таких рішень у складні інформаційні системи відкриває нові можливості для глибшого розуміння візуального контенту, оперативного прийняття рішень і масштабованого застосування в практичних задачах різної складності.

Таким чином, міждисциплінарний підхід, що поєднує комп'ютерний зір, обробку природної мови, сенсори та глибоке навчання, стане визначальним у майбутньому розвитку відеоаналітики як складової цифрового інтелекту.

Література

1. Кобилін О. А., Гороховатський В. О., Творошенко І. С., Передрій О. О. Застосування методів непараметричної статистики в класифікаторах зображень на основі компонентів структурного опису // Телекомунікації та радіоінженерія. 2020. №10. С. 79.
2. Коваленко, О. І., & Мартиненко, В. В. (2021). Аналіз методів розпізнавання об'єктів в відеопотоці на основі глибоких нейронних мереж. Інформаційні технології і засоби навчання, 71(2), 46-58.
3. Мельник П. В., Шестопапов С. В. Методи семантичного аналізу в системах відеоспостереження // Вісник Київського національного університету технологій та дизайну. 2018. №2. С. 45-51.
4. Шаповалов А. С., Мелешко С. І. (2022). Застосування нейронних мереж для аналізу відеоданих у системах безпеки. Інформаційні технології в освіті, науці і техніці, №1(42), С. 102–109.