UDC 004.932.2

A.O. Shevchuk
O.N. Romanyuk

METHODS OF ANOMALOUS BEHAVIOR RECOGNITION IN INTELLIGENT VIDEO ANALYTICS SYSTEMS

Vinnytsia National Technical University

Анотація

У роботі розглянаються методи розпізнавання аномальної поведінки в системах інтелектуального відеоаналізу. Проаналізовано основні підходи до виявлення нестандартних ситуацій на відеозаписах, включаючи класичні алгоритмічні методи та сучасні підходи на основі глибокого навчання. Особлива увага приділяється конволюційним нейронним мережам, моделям LSTM та архітектурам-трансформерам, які дозволяють автоматизувати виявлення підозрілих подій в режимі реального часу.

Ключові слова: інтелектуальний відеоаналіз, аномальна поведінка, конволюційні нейронні мережі, LSTM, оцінка потоку оптичного довкілля, відеоспостереження.

Abstract

This paper examines methods for recognizing anomalous behavior in intelligent video analytics systems. The main approaches to detecting non-standard situations in video recordings are analyzed, including classical algorithmic methods and modern deep learning-based approaches. Special attention is given to convolutional neural networks, LSTM models, and transformer architectures, which enable automated detection of suspicious events in real time.

Keywords: intelligent video analytics, anomalous behavior, convolutional neural networks, LSTM, optical flow estimation, video surveillance.

Introduction

The rapid expansion of urban surveillance infrastructure has created an urgent demand for automated tools capable of detecting threatening or unusual events without continuous human oversight. Traditional rule-based video monitoring systems suffer from high false-positive rates and cannot generalize to novel scenarios. Intelligent video analytics (IVA) systems that incorporate machine learning provide a scalable solution, yet the diversity of human behaviors and environments makes the problem of anomaly detection fundamentally challenging [1]. This paper reviews the principal algorithmic approaches to anomalous behavior recognition and discusses their strengths and limitations in the context of territorial security.

Classical Algorithmic Approaches

Early IVA systems relied on background subtraction and frame differencing to isolate moving objects from a static scene. Gaussian Mixture Models (GMM) remain a widely used baseline: each pixel's intensity distribution is modeled by a mixture of Gaussians, and pixels that deviate significantly from the learned model are classified as foreground [2]. While computationally efficient, GMM-based detectors are sensitive to illumination changes, camera jitter, and dynamic backgrounds such as waving foliage.

Optical flow estimation, as formalized by Horn and Schunck, computes per-pixel motion vectors between consecutive frames and allows aggregation of motion statistics at the region or scene level. Histogram of Oriented Gradients (HOG) descriptors combined with Support Vector Machines (SVM) provide robust person detection but require hand-crafted feature engineering and scale poorly to complex temporal sequences [3].

Deep Learning-Based Methods

Convolutional Neural Networks (CNNs) have transformed the field by learning hierarchical spatial features directly from raw pixel data. Two-stream architectures process appearance (RGB frames) and motion (optical flow) in parallel branches, merging their outputs for action classification [4]. 3D convolutional networks (C3D, I3D) extend convolution into the temporal dimension, enabling joint spatio-temporal feature extraction from short video clips.

For modeling long-range temporal dependencies, recurrent architectures such as Long Short-Term Memory (LSTM) networks are employed. A common pipeline encodes each frame with a CNN backbone and feeds the

resulting feature vector sequence into an LSTM to capture behavioral dynamics over time. Anomalies are flagged when the LSTM prediction error for the next frame exceeds a learned threshold, effectively implementing a one-class classifier trained on normal behavior [5].

Transformer-based models, originally designed for natural language processing, have recently been adapted for video understanding. Vision Transformers (ViT) and their video extensions (TimeSformer, Video Swin Transformer) use self-attention mechanisms to model global context across spatial patches and temporal frames simultaneously. This global receptive field is particularly advantageous for detecting crowd anomalies or coordinated suspicious activities that span large areas of the scene [6].

Evaluation Metrics and Benchmark Datasets

Performance is commonly measured by the Area Under the ROC Curve (AUC) at the frame level, as anomaly detection is inherently a binary classification problem. Standard benchmarks include UCSD Ped1/Ped2, CUHK Avenue, and ShanghaiTech Campus datasets, with ShanghaiTech being the largest and most diverse. Methods achieving AUC above 0.90 on ShanghaiTech are considered state-of-the-art [7]. A critical challenge is the absence of anomaly samples during training: most competitive approaches adopt a reconstruction-based or prediction-based formulation, defining anomalies as events with high reconstruction error under a model trained exclusively on normal video.

Conclusions

The analysis demonstrates a clear evolutionary trajectory from handcrafted feature descriptors toward end-to-end deep learning solutions. Classical GMM and optical-flow approaches offer interpretability and low computational cost but lack generalization ability. CNN-LSTM hybrids provide a strong baseline for temporal anomaly detection, while transformer architectures currently represent the frontier, offering superior global context modeling at the cost of higher data and compute requirements. Future research directions include unsupervised domain adaptation to reduce dependency on labeled data, lightweight model designs for edge-device deployment, and multimodal fusion of video with audio or infrared signals to improve detection reliability under adverse conditions.

REFERENCES

1. Kiran B. R. et al. Deep Learning Based Anomaly Detection in Video Surveillance: A Survey // IEEE Access. 2022. Vol. 10. P. 28509–28524.
2. Stauffer C., Grimson W. E. L. Adaptive Background Mixture Models for Real-Time Tracking // Proc. IEEE CVPR. Fort Collins, 1999. P. 246–252.
3. Dalal N., Triggs B. Histograms of Oriented Gradients for Human Detection // Proc. IEEE CVPR. San Diego, 2005. Vol. 1. P. 886–893.
4. Simonyan K., Zisserman A. Two-Stream Convolutional Networks for Action Recognition in Videos // Advances in Neural Information Processing Systems. 2014. Vol. 27. P. 568–576.
5. Luo W. et al. Video Anomaly Detection with Sparse Coding Inspired Deep Neural Networks // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2021. Vol. 43. No. 3. P. 1070–1084.
6. Liu Z. et al. Video Swin Transformer // Proc. IEEE CVPR. 2022. P. 3202–3211.
7. Liu W. et al. Future Frame Prediction for Anomaly Detection – A New Baseline // Proc. IEEE CVPR. Salt Lake City, 2018. P. 6536–6545.

Шевчук Артем Олександрович – студент групи 1ПІ-22б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця, e–mail: artemshewchukvntu@gmail.com.

Романюк Олександр Никифорович – доктор технічних наук, професор, професор кафедри програмного забезпечення, завідувач кафедри програмного забезпечення, Вінницький національний технічний університет, м. Вінниця, e–mail: rom8591@gmail.com.

Shevchuk Artem O. – student of group 1PI-22b, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: artemshewchukvntu@gmail.com.

Romanyuk Oleksandr N. – Doctor of Technical Sciences, Professor, Professor of the Software Engineering Department, Head of the Software Engineering Department, Vinnytsia National Technical University, Vinnytsia, e–mail: rom8591@gmail.com.