

AI-Based Real-Time Theft Detection Using Surveillance Videos

Aditya Yogesh Rashinkar¹, Ayush Pradip Sonawane², Bhavya Deepak Pawar³,
Sahil Dnyaneshwar Mahale⁴, Prof. Mr. V. S. Walke⁵

Department of Computer Engineering¹⁻⁵

Guru Gobind Singh Polytechnic, Nashik, Maharashtra, India

adityarashinkar38@gmail.com¹, ayushsonawane74@gmail.com²

bhavya.pawar2008@gmail.com³, sahilmahale2801@gmail.com⁴

Abstract: CCTV cameras are now used in many places for security, but watching several camera feeds continuously is not an easy task for a person. When monitoring is done manually for a long time, important events can be missed. Because of this, researchers have started using computer vision and deep learning to automatically analyse surveillance videos and identify suspicious or criminal activities.

This literature survey looks at selected research on crime detection, video anomaly detection and suspicious activity recognition, with a specific focus on ideas that can be useful for theft detection. The reviewed studies show that YOLO-based object detection can help identify people and objects, while CNNs and temporal models can help understand what is happening across a sequence of video frames. Transfer learning is also useful when a project does not have a very large dataset.

At the same time, the reviewed work shows that real-world theft detection is still challenging. Problems such as limited and unbalanced datasets, different lighting conditions, camera angles, background changes, processing requirements and false alarms can affect the performance of a system. From these studies, a clear need can be seen for a practical real-time theft detection system that combines object detection with activity analysis and can generate a useful alert together with evidence of the detected event.

Keywords: Surveillance video, theft detection, crime detection, deep learning, computer vision, YOLO, anomaly detection, activity recognition, real-time monitoring.

I. INTRODUCTION

Video surveillance is commonly used in places such as shops, public areas, parking spaces, educational institutions, banks, streets and smart-city environments. CCTV is mainly installed to improve security and to help identify unusual events. However, when many cameras are running continuously, it becomes difficult for a person to watch every screen carefully. This can result in an important event being overlooked. For this reason, many researchers have worked on methods that can automatically identify abnormal activities from surveillance videos [2].

Crime detection is closely related to this problem. A surveillance system may be designed to identify activities such as burglary, theft, vandalism, violence or other suspicious behaviour. Different research papers use different approaches. Some focus mainly on detecting objects in individual frames, while others study a sequence of frames so that the system can understand how an activity develops over time.

This difference is especially important for theft detection. Seeing a person near an object does not automatically mean that theft is taking place. The system may need to understand what the person is doing before deciding whether an event is suspicious.



In this survey, selected studies are reviewed and compared based on their methods, findings and limitations. The main purpose is to understand what has already been done and then identify ideas that can be useful for developing a practical real-time theft detection system.

1.1 Gao et al. – Camera-Based Crime Behavior Detection and Classification

Gao et al. developed a surveillance-video system for detecting and classifying selected criminal activities. Their work considered arson, burglary and vandalism and compared several object detection models, including improved YOLOv5, Faster R-CNN, YOLOv7, SSD MobileNet and YOLOv6 [1]. This study is useful for the present project because it shows how different detection models can be used as part of a larger crime-monitoring system.

For their dataset, the researchers collected videos from sources such as UCF, YouTube and Storyblocks. They selected the relevant portions of the videos, converted them into frames, annotated the frames and divided the data into training, validation and testing sets. They also used preprocessing steps such as resizing and data augmentation before training the models. The paper reports that the prepared datasets contained thousands of frames for the different crime categories [1].

In the reported experiments, improved YOLOv5 achieved an mAP@50 of about 0.80 for arson, YOLOv7 achieved about 0.87 for burglary, and YOLOv6 achieved about 0.86 for vandalism [1]. These numbers should not be interpreted as a direct comparison of the models because the models were used for different crime categories and different datasets. The authors also combined the individual models so that video frames could be processed and the detected information could be used for final crime classification.

The study also points out some limitations. It covered only a limited number of crime categories, and adding more categories would require additional labelled data. Computational and deployment requirements were also important issues. The authors suggested expanding the dataset, adding person tracking and exploring temporal models such as LSTM in future work [1].

1.2 Nayak, Pati and Das – Deep Learning for Video Anomaly Detection

Nayak, Pati and Das presented a detailed review of deep-learning methods used for video anomaly detection [2]. Their work explains that an anomaly can be either an unusual activity or an unusual object/entity in a scene. Examples discussed in the review include fighting, riots, traffic violations, stampedes, weapons and abandoned objects.

One important point from this study is that abnormal behaviour is not always easy to define using one fixed visual pattern. Human behaviour can be complex, and changes in the environment can affect what the model sees. The availability and quality of datasets also have a strong effect on the final performance [2]. This is directly relevant to theft detection because the same type of movement may be normal in one situation and suspicious in another.

The review also shows that a useful surveillance system has to balance detection accuracy with the ability to work in practical or real-time conditions. This is important for a theft detection project because a highly accurate model is less useful if it takes too long to process a live video stream.

1.3 Himeur et al. – Transfer Learning and Domain Adaptation

Himeur et al. reviewed the use of deep transfer learning and deep domain adaptation in automated video surveillance [3]. According to the study, deep-learning surveillance systems often require a large amount of labelled training data and considerable computing resources. Another problem is that a model trained in one environment may not work equally well in another environment.

Transfer learning provides a practical way to reduce the amount of training needed. Instead of training a model completely from the beginning, a model that has already learned general visual features can be used as a starting point. Domain-adaptation techniques can then help reduce the difference between the training environment and the environment in which the system will actually be used [3].



For a theft-detection project with a limited custom dataset, this approach can be helpful. However, it does not completely solve the problem. Camera position, lighting, image resolution, background and other differences in the deployment environment can still affect the results.

1.4 Gandapur – E2E-VSDL for Criminal Activity Detection

Gandapur proposed an end-to-end video-surveillance deep-learning model called E2E-VSDL for detecting and preventing criminal activities [4]. The model combines CNN-based spatial feature extraction with recurrent components, including GRUs and SRUs, so that information from a sequence of video frames can be used. The study also gives attention to selecting important frames instead of treating every frame in the same way, which can help reduce unnecessary computation.

This work is particularly relevant to theft detection because theft is normally not a single-frame event. It is a sequence of actions—for example, a person approaching an object, interacting with it and then moving away with it. A model that considers the relationship between successive frames may therefore be able to distinguish normal movement from a suspicious sequence more effectively than a model that only looks at one frame [4].

The study also evaluates the approach using continuous video streams from multiple surveillance cameras and discusses alarm generation when a criminal activity is detected. This makes the work useful as a reference for systems that are expected to operate continuously rather than only on stored videos.

1.5 Khanam et al. – Suspicious Action Recognition in Surveillance

Khanam et al. reviewed different approaches for recognising suspicious actions in surveillance videos [5]. Their survey includes both traditional handcrafted-feature methods and deep-learning approaches such as C3D, GAN-based methods, YOLO-based methods and ConvLSTM-related techniques. The authors also discuss commonly used datasets and the practical problems involved in suspicious-action recognition.

A useful observation from this survey is that a model can perform well on a benchmark dataset but still have difficulty when it is used with a different camera angle, lighting condition or real-world environment [5]. More complex deep-learning models can also require more computing power. These points are important for the proposed project because the goal is to work with live CCTV footage rather than only with carefully prepared laboratory datasets.

TABLE I: COMPARATIVE ANALYSIS OF REVIEWED STUDIES

Study	Problem / Focus	Main Technique	Key Finding	Main Limitation
Gao et al. [1]	Crime detection from surveillance video	YOLOv5, YOLOv7, YOLOv6, Faster R-CNN, SSD, MobileNet	Automated detection of selected crime categories with reported mAP values	Limited crime classes, different datasets, and computation/deployment constraints
Nayak et al. [2]	Video anomaly detection	Deep-learning approaches and evaluation methods	Shows the difficulty of detecting abnormal activities/entities in complex video	Dataset and environmental variability; ambiguity of anomalies
Himeur et al. [3]	Generalization of video surveillance models	Transfer learning and domain adaptation	Can reduce training limitations and help with cross-domain use	Differences between domains can still affect performance
Gandapur [4]	Crime detection using video sequences	CNN + GRU/SRU and focused frame selection	Uses spatial and temporal information while reducing unnecessary processing	Temporal models can make implementation more complex
Khanam et al. [5]	Suspicious action recognition	C3D, GANs, YOLO, ConvLSTM and	Reviews methods, datasets and real-world	Generalization and computational cost remain issues



		related methods	challenges	
--	--	-----------------	------------	--

1. Research Gap Identified from the Literature

The reviewed papers show that automated surveillance has improved considerably, but there are still several issues to address when the goal is a practical real-time theft detection system.

Theft-specific detection: Many studies focus on broad crime or anomaly categories. A system designed specifically around theft-related behaviour can define the problem more clearly.

Frame-level versus activity-level understanding: Object detectors are good at identifying people and objects, but theft usually depends on a sequence of actions. Temporal analysis can provide the extra context needed.

Real-time processing: A practical system has to process continuous video while keeping detection delay and computing requirements within reasonable limits.

Multiple-camera operation: Real surveillance locations may have several cameras. Each stream needs to be analysed while the results are brought together in one monitoring interface.

Environmental variation: Lighting, camera angle, resolution, background and the number of people in a scene can all affect model performance.

Limited datasets: Building a large and diverse theft dataset is difficult. Transfer learning and data augmentation can help, but the system still needs to be tested using realistic data.

Actionable alerts: Detection should lead to a useful alert and evidence record so that the monitoring person can understand which camera detected the event and why it was considered suspicious.

2. Relevance to the Proposed Real-Time Theft Detection System

Based on the reviewed literature, a practical architecture for a real-time theft detection system can combine object detection with activity analysis. The object-detection part can identify relevant people and objects in each frame, while tracking or temporal analysis can be used to understand how their behaviour changes over several frames.

A simple workflow based on the reviewed studies is:

LIVE CCTV → FRAME PROCESSING → OBJECT DETECTION → TRACKING / ACTIVITY ANALYSIS → THEFT DECISION → ALERT GENERATION → EVIDENCE / EVENT LOG

The reviewed literature does not show that one particular model is always the best choice. The right model depends on the target classes, available dataset, hardware, required speed and evaluation method. Therefore, for the proposed project, suitable models should be compared under the same or clearly defined test conditions rather than comparing numbers taken from unrelated datasets.

3. Recommended Evaluation Considerations

The reviewed studies use measures such as precision, recall, mean average precision (mAP), confusion matrices, processing speed and validation/test performance. For a theft detection system, two more practical issues are especially important: false alarms and detection delay. A system that raises too many incorrect alerts, or detects a theft too late, may not be useful in a real situation.

Precision – the proportion of predicted theft events that are actually related to theft.

Recall – the proportion of theft events in the test data that the system successfully detects.

F1-score – a combined measure of precision and recall.

mAP / IoU – useful for checking object-detection and bounding-box performance.

FPS / latency – shows whether the system can process video close to real time.

False-alarm rate – shows how often normal activity is incorrectly reported as theft.

Cross-condition testing – checks performance under different lighting, camera positions, backgrounds and scene densities.



4. Academic Writing and Plagiarism Considerations

This literature survey is written as an independent summary of the reviewed research. The wording has been paraphrased instead of copied from the original papers, while findings belonging to particular authors are identified using numbered citations. Technical terms such as YOLO, CNN, GRU, LSTM, object detection, surveillance video and mAP are standard terms in this field and may naturally appear in other research papers.

The citations should remain connected to the claims they support. For example, the numerical results reported in Section 2.1 are presented as findings reported by Gao et al. [1], not as results obtained from the proposed project. In the same way, the discussions of anomaly detection, transfer learning, temporal models and suspicious-action recognition are connected to the corresponding studies [2]–[5]. This helps maintain academic integrity and makes it clear which ideas come from previous research and which direction is being proposed for the current project.

II. CONCLUSION

The studies reviewed in this survey show that deep learning and computer vision can be effectively used to support automated surveillance and crime-related activity detection. YOLO-based object detectors are useful for locating objects in video, while temporal approaches such as GRU- and ConvLSTM-related models can provide additional information about activities that happen over a sequence of frames [1], [4], [5]. Transfer learning is also useful when only a limited amount of application-specific training data is available [3].

However, the literature also makes it clear that real-world deployment is not straightforward. Limited datasets, changing lighting and camera views, computing requirements, real-time processing and generalization to new environments can all affect performance. Because of this, a theft detection system should be tested under realistic conditions instead of depending only on a controlled dataset.

Overall, the literature supports the idea of combining object detection, activity understanding, continuous video processing and timely alert generation for theft detection. The reviewed studies provide a useful technical foundation for the proposed system, while also highlighting the limitations that need to be considered during implementation and evaluation.

III. ACKNOWLEDGMENT

The authors acknowledge the researchers and published studies reviewed in this paper, which provided the technical foundation for understanding real-time surveillance, crime detection, anomaly detection, transfer learning, temporal activity analysis, and suspicious-action recognition for the proposed theft detection system.

REFERENCES

- [1] J. Gao, J. Shi, P. Balla, A. Sheshgiri, B. Zhang, H. Yu, and Y. Yang, "Camera-Based Crime Behavior Detection and Classification," *Smart Cities*, vol. 7, pp. 1169–1198, 2024. doi: 10.3390/smartcities7030050.
- [2] R. Nayak, U. C. Pati, and S. K. Das, "A comprehensive review on deep learning-based methods for video anomaly detection," *Image and Vision Computing*, vol. 106, 104078, 2021. doi: 10.1016/j.imavis.2020.104078.
- [3] Y. Himeur, S. Al-Maadeed, H. Kheddar, N. Al-Maadeed, K. Abualsaud, A. Mohamed, and T. Khattab, "Video surveillance using deep transfer learning and deep domain adaptation: Towards better generalization," *Engineering Applications of Artificial Intelligence*, vol. 119, 105698, 2023. doi: 10.1016/j.engappai.2022.105698.
- [4] M. Q. Gandapur, "E2E-VSDL: End-to-end video surveillance-based deep learning model to detect and prevent criminal activities," *Image and Vision Computing*, vol. 123, 104467, 2022. doi: 10.1016/j.imavis.2022.104467.
- [5] S. Khanam, M. I. Sharif, X. Cheng, and S. N. Kadry, "Suspicious action recognition in surveillance based on handcrafted and deep learning methods: A survey of the state of the art," *Computers & Electrical Engineering*, vol. 120, 109811, 2024. doi: 10.1016/j.compeleceng.2024.109811.

