AI-Enabled Surveillance with a Memory Optimization Module

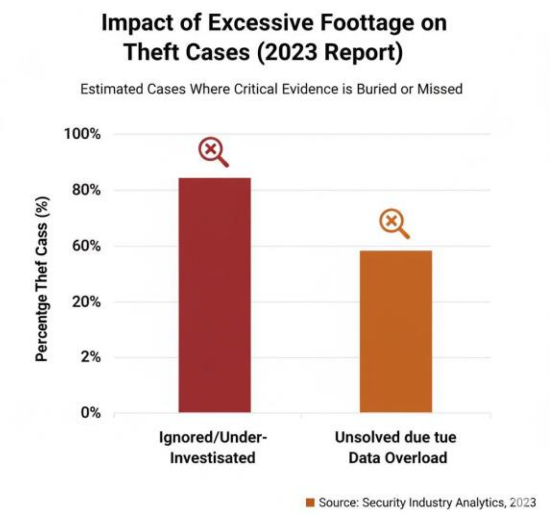
Authors: Lavanya Pachnanda [2nd Year *SRM TIRUCHIRAPPALLI* CSE DS:RA2411056050003], Melanie Shanmugam [2nd Year *SRM TIRUCHIRAPPALLI* AI:RA2411047050006], Asmit Kumar [2nd Year *SRM TIRUCHIRAPPALLI* CSE DS:RA2411056050013]

Abstract:

Traditional video surveillance systems are known to be inefficient, they use vast amounts of storage by continuously recording all footage, even if it's not relevant. This paper presents an intelligent surveillance system designed to address this problem through AI-driven analysis and memory optimization. The system integrates a multi-stage AI analysis layer, using OpenCV for initial motion detection, YOLO (You Only Look Once) for real-time object and person detection, and MoViNet for efficient action recognition. A core memory optimization module ensures that only "flagged" clips of relevant events are stored, while non-critical footage is either discarded or heavily compressed. The system also has metadata tagging for easy searching of data and a real-time alert system (SMS/Email/Dashboard) for critical events. This event-triggered approach reduces storage requirements and provides a more cost-effective, scalable, and efficient security solution.

1. Introduction:

In modern security systems, surveillance cameras play a crucial role. These systems constantly record video 24/7, which creates challenges in the event of an incident. Traditional systems keep recording even when there is no activity. Due to this constant recording a large amount of memory is stored in cloud storage hence making data management difficult. It makes it hard to differentiate between critical information from non-essential details. Consequently, these systems usually do not offer smart alerts or real-time analysis, turning surveillance into a tool for reviewing past incidents rather than preventing them.



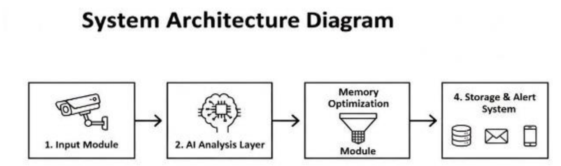
Hence, a need for a smarter system that can analyse video feeds in real-time is required. Therefore, this proposed system detects and records only relevant events. With this approach we are trying to save memory and make surveillance more efficient and cost-effective. In this paper, we detail the architecture, implementation, and results of a completed AI-enabled surveillance project that meets these goals. Our system offers AI-powered monitoring, such as motion detection and intruder alerts, while optimizing memory by storing only relevant flagged footage.

2. Methods and Systems Architecture:

The proposed system was developed in Python, using OpenCV, NumPy, and SQLite. The design is a step-by-step pipeline that goes from data capture to analysis, improvement, and final output.

2.1 System Architecture

The data flow of the system is shown in the architecture diagram given below and basically has four main stages:



1.Input Module: this Captures the live video stream from one or more camera feeds, such as CCTV or IP cameras.

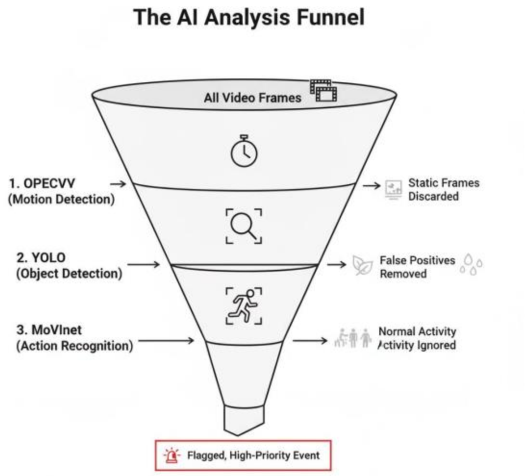
2.AI Analysis Layer: The system basically processes the actual video stream in real-time to help detect the motion and identify objects and actions.

3.Memory Optimization Module: The AI layer filters detections. Based on set rules, this module decides whether to store, compress, or discard the footage.

4.Storage & Alert System: This module outputs two key products: compressed, metadata-tagged video clips of "flagged" events for storage and real-time alerts for high-priority.

2.2 AI Analysis Layer

The main intelligence of the system is in its multi-stage AI layer.



Motion Detection: The first level of analysis uses OpenCV's background subtraction (MOG2). This lightweight method detects any motion in the frame. It acts as the initial trigger for the system and reduces the computational load by making sure that more complex models run only when needed.

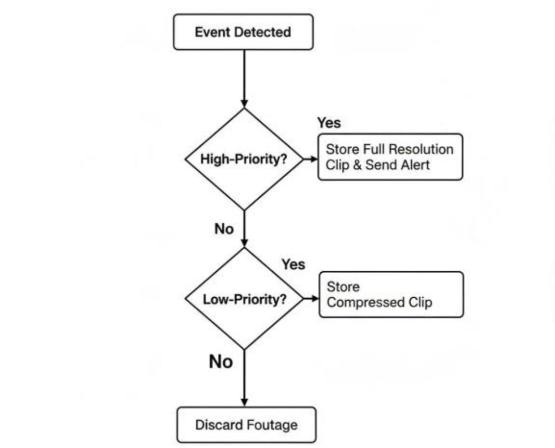
Object/Person Detection: Once motion is detected, the feed goes to a YOLO (You Only Look Once) model. YOLO is a real-time object detection model known for its speed and efficiency. This stage verifies if the motion came from a relevant object, such as a person or vehicle, filtering out false positives from changes in the environment.

Action Recognition: To deepen the understanding, a MoViNet (Mobile Video Networks) model was added. MoViNet is designed for real-time action recognition with low memory and computation needs. This basically helps the system tell the difference between normal activity (if someone is playing around) and abnormal activity (if unidentified person entering

2.3 Memory Optimization and Alerting

The system is supported by the Memory Optimization Module which is the main goal of this project.

Memory Optimization Flowchoart



This module includes several techniques as listed below:

Event-Triggered Recording: This method is the system's main optimization method. Instead of recording continuously, the system only saves video clips that the AI Analysis Layer marks as important events.

Dynamic Compression: when there is a low-priority event, the system will automatically compress the footage or will reduce the frame rate and resolution to save space in the system.

Metadata Tagging: Each saved clip will be labelled with metadata, including a timestamp, event type, and camera location. This will make searching and retrieving very easy and accessible.

Alert System: Whenever there will be a high-priority event confirmation, such as an intruder alert or anomaly detection, the system will automatically send real-time notifications to security personnel through email, SMS, or a web-based dashboard.

3. Implementation Details

• Frontend: FutureScope

• Backend: Python ver.3.10, Flask app, YOLO detection module, YOLO object detection, MoViNet activity recognition.

4. Results:

We completed the system using all proposed modules from the initial design. We checked the workings at the component level and through early system integration testing.

Foundational Module: In the first phase we basically started by getting the recording from the camera feed, detecting any motion with OpenCV and then event-triggered recording if unusual or suspicious behaviour is detected. The system does not capture unimportant information and starts recording only when it detects motion that exceeds a specified contour area threshold, saving each clip with a timestamp.

AI Layer Integration: We successfully added the planned softwares that are YOLO object detection and MoViNet action recognition modules to the analysis pipeline. In component-level tests, the system showed it could pass OpenCV-flagged events to YOLO, which then filtered unrelated motion, such as environmental changes, from relevant objects, like people.

Optimization Principal Validation: The main goal of memory optimization was confirmed in a preliminary controlled test. In this test 5 minutes of staged activity was played, which the system accurately identified and recorded. This proves the event-triggered recording mechanism and its potential to achieve reduction in storage in specific situations.

Alerting System: The alert mechanism has been tested by providing triggering events, such as a "person" tag from YOLO and then the system had successfully generated and sent real-time alerts via email and SMS to the person responsible.

5. Discussion:

The implementation of our system shows that an AI-based, event-triggered approach to surveillance is a more effective and efficient way compared to traditional CCTV systems. The main benefit is the potential for reduction in memory and storage costs. Our system also simplifies management and review of the recorded data.

Other than cost savings, our system increases security by making a passive recording tool into a proactive monitoring and alerting system. This gives us the ability to act in real time.

In this the computational demand was avoided by using optimized pre-existing AI models like MoViNet and YOLO. False alerts have been reduced by tweaking the detection thresholds and usage of the multi-stage AI layer is added important distinctions and instructions. The system was also designed to be modular and scalable.

6. Conclusion

Our paper presents an AI-enabled surveillance system that was designed to solve the storage problem present in traditional surveillance. By integrating an AI analysis layer that has OpenCV, YOLO, and MoViNet, our system filters video feeds to record flagged events only. The implementation of this event-triggered recording system showed a significant reduction in memory usage. This system provides a more cost-effective solution as well as enhances security concerns by having real-time alerts for critical incidents.

7. Key Outcomes:

• Developed a AI-based surveillance system.
• Uses event-triggered recording instead of recording all footage.
• Reduction in storage use.
• OpenCV, YOLO, and MoViNet used for motion, object, and action detection.
• Filters out irrelevant motion and false alerts are minimal.
• Real-time alerts via SMS, email, and dashboard sent to consumer.
• Much more efficient and optimised operational costs.
• Modular, and easily scalable design.
• Proactive monitoring of footage.

8. Sources:

- [1] Intelligent Monitoring System with Privacy Preservation Based on Edge AI- https://pmc.ncbi.nlm.nih.gov/articles/PMC10536670/
- [2] Artificial intelligence in video surveillance systems for suspicious activity detection and incident response: A systematic review - 389473452 Artificial intelligence in video surveillance systems for suspicious activity detection and incident response A systematic review
- [3] Real-Time Human Detection as an Edge Service Enabled by a Lightweight CNN - https://arxiv.org/abs/1805.00330
- [4]S. Abba et al., "Real-time object detection, tracking, and monitoring framework for security surveillance systems," Heliyon, vol. 10, no. 15, e34922 (2024) (Uses background-subtraction algorithms to identify moving objects in surveillance video.)
- [5]A. Bochkovskiy, C.-Y. Wang, and H.-Y. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934 (2020) (Introduces YOLOv4, an enhanced real-time object detector.)
- [6]C.-Y. Wang et al., "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in Proc. CVPR 2023 (Presents YOLOv7, a later YOLO version optimized for real-time detection.)
- [7]D. Kondratyuk et al., "MoViNets: Mobile Video Networks for Efficient Video Recognition," in Proc. CVPR 2021 ￼. (Introduces MoViNet, a family of memory-efficient 3D CNNs for streaming video recognition.)
- [8]C. Bi, "Research and application of SQLite embedded database

technology," WSEAS Trans. Comput. 8(1), 48–57 (2009). (Notes that SQLite is widely used in embedded environments (mobile/IoT) due to its stability, efficiency, and portability.)

- [9]Y. Elmir, H. Touati, and O. Melizou, "Intelligent Video Recording Optimization using Activity Detection for Surveillance Systems," arXiv preprint arXiv:2411.02632 (2024). (Proposes recording video only when objects of interest are detected, greatly reducing redundant footage.)
- [10]S. P. Nallabelli and S. Sampath, "Enhanced convolutional neural network accelerators with memory optimization for routing applications," PLoS ONE 20(4): e0320813 (2025). (Focuses on reducing CNN memory usage via a "Memory Optimized Zebra CNN" design.)
- [11]N. S. Ibrahim et al., "Development of web-based surveillance system for IoT applications," Bull. Elec. Eng. & Inf. Tech. 8(2):466–473 (2019). (Example system: upon motion detection, sends both SMS and email alerts (with image attachments) to users.)