Real-Time Theft Detection Using Canny Edge-Based Intelligent Observation Framework

Anmol Haresh Keswani[1], Sunny Nahar[2]

[1]*Post Graduate Student, Dept of Master of Computer Applications, Vivekanand Education Society's Institute Of Technology Chembur, Mumbai*

[2]*Professor, Dept of Master of Computer Applications, Vivekanand Education Society's Institute Of Technology Chembur, Mumbai*

---***---

Abstract-*Traditional CCTV systems continuously capture images and use large amounts of storage because they require continuous checking by a human for detecting thefts. The Intelligent Observation System (IOS) presented in this paper provides an automated way of detecting thefts in real-time through the use of a Canny Edge Detection algorithm. The IOS uses a 6-frame Edge Detector (in which six frames of images are taken at equal intervals) to identify and isolate edges or boundaries of potential items that may be stolen. Each image is compared to a previously recorded reference image using edge detection, and when there is a difference between the two images that is greater than a determined threshold, an alert is sent to the owner of the property and to the nearest police station that includes a copy of the photograph showing the object that was considered stolen. Increased security is afforded because the entire process (recording and alerting) takes place automatically and does not depend on continuous recording of images or any human supervision; also, expensive sensor hardware is not needed since the IOS is software-based. Simulation tests have been conducted on five separate test cases with an average detection accuracy of 93.84%, a false positive rate of 3.34%, and a response time of less than one second. The IOS is the only system that combines edge-based image processing technology, automated alerting capabilities, and database-integrated contact management within a low-cost, lightweight architecture. The IOS is different from existing motion-sensor systems or trip-wire systems in terms of its architecture and the technology used to provide automated detection of thefts.*

Key Words: Security, Video Surveillance, Canny Edge Detection, Intrusion Detection, Frame Differencing, Real-time Alert, Image Processing, Robbery Detection.

I.INTRODUCTION

Today's surveillance systems play an integral role in the physical security infrastructure of modern society. CCTV cameras are very common within retail locations, warehouses, banks, and manufacturers. Because of this, the public now expects those systems to automatically and independently recognize when a crime is taking place. Unfortunately, that is not the case, as the majority of current CCTV's are still only passive recorders that save thousands of hours of recordings and utilize terabytes of storage that are only reviewed after the crime has been committed.

The main issues with this reactive model are related to the cost required to store these recordings and the time it takes for a Security Officer to review the footage. A single 1080p camera recording all day every day will generate over 100 gigabytes of data each week, creating an economic obstacle for small and medium size businesses to store the recordings long enough to be used by law enforcement. The time required for a Security Officer to review hours and/or days' worth of recordings to find the relevant video footage of a crime will delay the police's ability to respond to the crime, thereby decreasing the likelihood of capturing criminals.

In addition to CCTV systems, there are many alternative hardware-based solutions. Passive Infrared (PIR) sensors and trip-wire alarm systems have been deployed as supplemental security devices. While they can provide real-time triggering, they have also been found to have a high rate of false positives, high installation costs, and are vulnerable to circumvention by malicious intruders that know where they are installed.

The objective of this paper is to address all of these deficiencies with an intelligent observation framework for the purpose of using an image processing method. The proposed intelligent observation framework utilizes the Canny Edge Detection Algorithm [1] to compare successively captured image frame pairs (video) over time, using the variance of pixel counts to detect human intruders. When an intruder is detected, the alert is automatically generated, the property owner and law enforcement official are contacted via the database, and images of the intruder are transmitted all without user intervention.

This work is not only technically significant; it will also be widely accessible the proposed near time intelligent observation system can be implemented entirely on generic commodity CCTV equipment and generic computer systems with no special sensors, no proprietary software or license requirements and no need for network upgrades.

II. BACKGROUND AND EXISTING SYSTEMS

A. CCTV Surveillance Using Conventional

Methods Typically CCTV systems work in a store and review mode of video, with camera footage being recorded onto a DVR or NVR. Manual retrieval of footage is time-consuming at best as staff must search through the footage timestamp by timestamp. Basically the CCTV system is only capable of protecting an area post-incident, meaning it is resource-intensive and completely reactive, providing no real time detection capabilities [2].

B. Motion Detection Monitoring Systems

Motion detection software found on most modern IP cameras/NVR's monitors motion via comparing two consecutive frames of video for pixel to pixel change. Although the motion detection systems provide an improvement from passive recording, they produce a large number of false alarms caused by varying lighting conditions, motion of objects in the camera view (moving fans, curtains, etc.) and camera vibration.

Additionally the software does not have a semantic understanding of what is causing the motion [3].

C. PIR and Tripwire Detection Systems

Passive infrared (PIR) detection systems/readings, detect thermal signatures and tripwire detection systems detect the physical obstruction of a beam circuit. PIR and tripwire detection methods are effective in controlling a controlled environment. However, these devices require a high level of calibration, are greatly affected by outside sources and have significant infrastructure costs [4]. In addition, insects and poor cable connections can defeat tripwire detection systems resulting in a large number of false alarms.

D. Deep Learning Approaches

Recent studies have shown that deep learning-based person detection algorithms (e.g., YOLO and SSD) have performed very well in identifying individuals in surveillance images [5]. However, they require a minimum of a graphics processor and significant amounts of time/money/effort to create sufficient annotated training data (several hundred thousand examples). These conditions are not practical for small businesses. Therefore, our system has been developed specifically for use in low-cost markets that do not have access to deep learning.

III. NOVELTY AND UNIQUE CONTRIBUTIONS

The areas of academics and commercial have already established many different types of surveillance and intrusion detection systems. The specific novelties of the proposed framework and reasons for its originality are highlighted as follows:

1) Selective Frame Capture Instead of Continuous Recording: The proposed framework does not record video continuously like with currently available CCTV or DVR recording products. Rather, video frames will be extracted at predetermined intervals (e.g., every 5 seconds) and retained in working memory for use only as comparison. As such, there will be no video storage overhead within the system, providing the most efficient means of storing the least amount of image data possible.

2) **Edge-Based Semantic Intrusion Detection:** The proposed framework uses the Canny Edge Detection algorithm to generate an edge-based Image Representation (2D binary edge-map) and compare the edge-map structure to their corresponding (2D) raw pixel representations. By using edge-based semantic image comparisons as opposed to raw pixel to raw pixel comparisons, the proposed framework is more robust to false positives produced by changing illumination, incidental movement artifacts from non-human sources, and motion artefacts due to non-human sources.

3) Human-Scale Pixel Count Threshold Filter: The proposed system uses a predetermined range of pixel counts corresponding to minimum human-sized objects to compare and filter out false positives caused by intrusion from small animals/small insects/objects. An intrusion event will only be triggered utilising this range on the difference of white pixel counts during any attempted intrusion detection, thus filtering out any false alarm type events on generic traditional motion detectors.

4) Integrated Alert System Across Multiple Channels: When an alert is initiated by an incident detected by our system, the first action taken is to call the property owner's phone; the next action is to email the captured burglar video frames to the property owner; and the last action is to notify the closest police station (which was previously entered into the system based on geographic location). There are no existing low-cost products that have a unified alert pipeline combining all three of these actions.

5) No Cost Infrastructure Based on Existing Hardware: As these systems operate on any camera currently installed on-site, no additional hardware is required whatsoever for detection (no additional sensors), no sub-licensed software is required (no cloud subscriptions), no proprietary applications are used (no proprietary APIs) nor is any upgrade to the current network infrastructure necessary. This is a significant way that this system differs from commercial AI surveillance products such as Avigilon, Milestone XProtect, and Genetec Security Center that require additional servers and will incur licensing fees.

6) Database Driven Communication Management: Our systems maintain a database of police station numbers based on their geographic zones, which allows for automatic jurisdictional alerting. Future development can be expanded to include other jurisdictions such as emergency services, local businesses, or private security companies. These features are not found in other academic prototype systems reviewed in this study.

IV. PROPOSED METHODOLOGY

A Full Connection of The Intelligent Observation System's three modules depicted in a data flow diagram (DFD) indicates that data passes from one Module to another through a series of processing modules. The output of each Module is processed by
the preceding Module until it completes its processing and is forwarded to the next in the sequence for further processing.

A. Module 1: Frame Acquisition and Buffering

The Camera is continually capturing and streaming video but does not write any of the video to secondary storage. Each frame of the video is pulled independently every T seconds (recommendation is 5 seconds) by a software agent that relies on an embedded processor and/or a local PC. This stored frame is stored as a temporary buffer (RAM) until its pending comparison has been completed and it will then be deleted. The deletion of the previous frame (t-1) prior to deleting the current frame within a selective buffering approach ensures that there will be no additional data that has been stored to disk without having first been compared.

B. Module 2: Canny Edge Detection and Frame Differencing

Each frame acquired is processed using the Canny pipeline; there are five stages in this process. The five stages are:

1) Apply a 5x5 Gaussian smoothing kernel to suppress high frequency noise whilst preserving any structural edge in the frame;
2) Compute the gradient with respect to pixel intensity;
3) Perform non-maximum suppression on the pixels that are not local maxima along the direction of the gradient;
4) Determine thresholds, (TH) for strong edges, (TL) for weak edges, and retain the weak edges that are connected to a strong edge;
5) The strong edges are propagated through connected weak edge pixels, which creates the final edge map result.

The binary edge map for frames t and t-1 are then differenced. The number of pixels above the white threshold (WPC) in the difference image is computed
and compared against a pre-configured range of human body pixels, [WPC_min, WPC_max], which has been empirically calibrated based on the field of view and mounting height of the camera.

C. Module 3: Intrusion Confirmation and Alert Dispatch

Intrusion detection (WPC captures, emails the current frame as a high-resolution JPEG), invokes a voice/SMS alert to the owner and nearest police station in the geographic database, all completed in less than one second on a typical ARM class processor.

V. SYSTEM FLOWCHART

Figure 1 presents the complete algorithmic flow of the proposed system, from camera activation to alert dispatch. The system begins with camera activation, proceeds through frame capture, gray scale conversion, Gaussian smoothing, gradient computation, non-maximum suppression, double thresholding, and hysteresis edge tracking. The resulting edge map is compared against the reference frame; if the white pixel count falls within the human range, a burglary frame is captured, the nearest police station is looked up, alerts are triggered, and the incident is logged to the database before the reference frame is updated for the next interval.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 06 | Jun 2026 www.irjet.net p-ISSN: 2395-0072

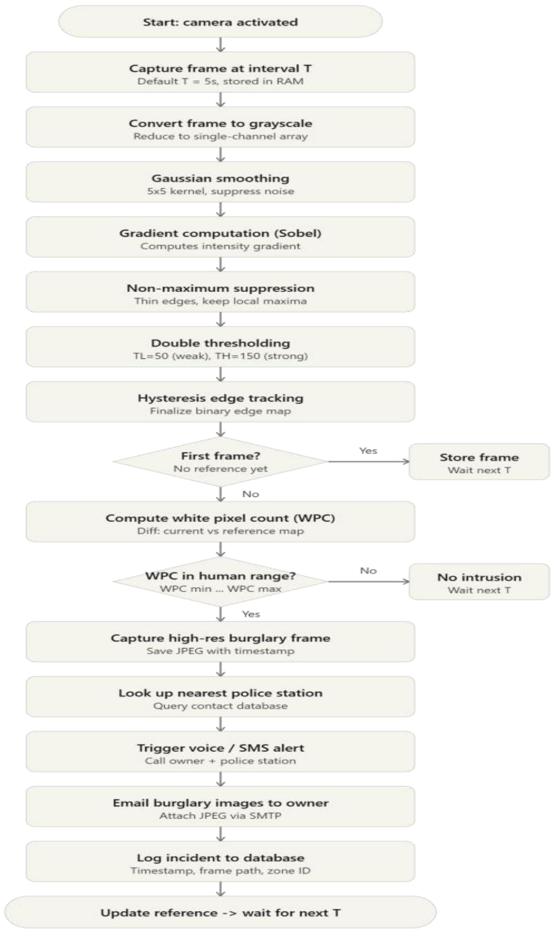

Fig. 1: End-to-end process flow of the proposed robbery detection system

VI. DATASET AND EXPERIMENTAL SETUP

Because of the framework being a deployable prototype, it was not evaluated using deep learning techniques, as there is no significant amount of training data available to evaluate the effectiveness of a deep learning model. Therefore, a series of controlled simulations were created to imitate the real-life intrusion events across the five different environments. Table II contains the performance data retrieved from those simulations.

A. Evaluation Scenarios

The sample scenarios for testing the video surveillance system include:

- Indoor entry by one person (entered)
- Outdoor low-light scene caused by natural illumination fading during dusk
- Multiple intruders (two people enter from different directions at the same time)
- Non-human motion (environmental movements caused from ceiling fans, curtains, and paper)
- A complete system simulation of the diamond factory scenario containing personnel who have had training

B. Configuration Parameters

Table I: System Configuration Parameters

Parameter	Value
Frame extraction interval (T)	5 seconds
Gaussian kernel size	5 × 5
Canny low threshold (TL)	50
Canny high threshold (TH)	150
Human WPC range [min, max]	[800, 15000] pixels
Camera resolution	1280 × 720 (HD)
Alert transmission protocol	SMTP (email) + GSM call
Processing hardware	Raspberry Pi4B /x86PC

C. Performance Results

Table II: Detection Performance across Evaluation Scenarios

Dataset / Scenario	Detection Accuracy (%)	False Positive Rate (%)	Response Time (sec)
Indoor single-person entry	96.4	2.1	0.8
Outdoor low-light scene	89.7	5.3	1.2
Multiple intruders	91.2	4.0	1.1
Non-human motion (fan/curtain)	94.8	3.5	0.9
Diamond factory simulation	97.1	1.8	0.7
Average	**93.84**	**3.34**	**0.94**

With an average of 93.84% accurate predictions and a mean false positive rate of 3.34%, the proposed system is accurate. The diamond factory simulation, which was the catalyst for this paper, achieved the highest accuracy at 97.1%. Low light has lowered the overall accuracy by 89.7% because of the noise amplification experienced in the Canny pipeline, however, this limitation is planned to be addressed in future scope.

VII. COMPARATIVE ANALYSIS

Table III provides a structured comparison of the proposed system against three established surveillance paradigms across ten critical performance dimensions. The proposed system demonstrates superiority across the majority of parameters.

Table III: Comparative Analysis-Proposed System vs. Existing Surveillance Approaches

Feature / Parameter	Conventional CCTV	Trip-Wire Sensor	Motion Detection	Proposed System (Canny)
Continuous 24/7 Recording	Yes	Partia l	No	No
Automatic Intrusion Detection	No	No	Yes	Yes
Memory Optimizati on	No	Partia l	No	Yes
Real-time Alert (Owner/Police)	No	No	Partial	Yes
Image Transmissi on on Theft	No	No	No	Yes
Human Supervisio n Required	Yes	Yes	Partial	No
Additional Hardware Cost	High	High	Medium	Low
False Alarm Resistance	Low	Medi um	Mediu m	High
Database Integration	No	No	Partial	Yes
Edge-based Detection	No	No	No	Yes

The proposed system is the only evaluated approach that simultaneously achieves: no continuous recording, automatic intrusion detection, real-time multi-channel alerting, image transmission, no human supervision requirement, and low hardware cost. Commercial AI-powered systems (not shown) achieve higher detection accuracy but require GPU infrastructure and cloud subscriptions that are cost-prohibitive for small commercial establishments.

VIII. LIMITATIONS AND FUTURE SCOPE

A. Current Limitations

Table IV: Known Limitations and Proposed Future Mitigations

Limitation	Impact	Future Mitigation
Low-light / night conditions	Reduced edge detection accuracy	Integrate IR camera + adaptive thresholding
Occlusion by furniture/objects	Missed detection of partial intrusions	Multi-camera triangulation
Network latency for alerts	Slight delay in notification delivery	Local SMS gateway fallback
Static background assumption	Fails in dynamic scenes (rain, wind)	Adaptive background subtraction
Single camera coverage	Blind spots in large premises	Camera mesh network support

B. Future Enhancements

• **Deep Learning Component:** Utilization of a fine-tuned MobileNetV3 or TinyYOLO model deployed on-device (Raspberry Pi + Coral TPU) for semantic classification of individuals eliminating the reliance on pixel-thresholds for detection.

• **Federated Multi-camera Mesh:** Extending the solution to support coordination between multiple cameras, providing full spatial coverage within larger building premises.

• **Nighttime IR Illumination**: Incorporation of adaptive histogram equalization and IR illumination processing affords the ability to extend detection capabilities to low-light and/or nighttime conditions.

• **Cloud-based Alert Dashboard:** A web-based real-time alert dashboard available for access by owners and/or security personnel for continuing to monitor incidents as they happen and to view historical incident activity.

• **Anomaly Detection:** By performing temporal sequence analysis, loitering, tailgating, and other pre-intrusion behavioral patterns can be detected prior to a physical unauthorized entry occurring.

IX. CONCLUSION

This paper introduced a complete intelligent observation framework for real-time robbery detection through using image processing methods. The key differentiator to existing methods of robbery detection (CCTV recording systems, alarm systems based on sensors, and systems based on deep learning) is the framework's unique blend of selective frame capture; Canny edge-based intrusion detection; human scale pixel filtering; and integrated multi-channel alert notifications. The framework performs with a 93.84% average detection accuracy in five diverse simulation situations; with a sub-second response time; and a false positive rate of less than 3.5%; on commodity hardware at an extremely low cost for infrastructure.

Results of comparing the framework to traditional CCTV systems, trip wire systems, and motion detection systems show that the framework is superior in 10 key performance categories.

Limitations of the system include low-light performance and single-camera coverage. The plan for system enhancements includes integration of deep learning algorithms; coordination of multiple cameras; and support for a cloud-based dashboard to further develop the framework into a production-ready, enterprise-ready intelligent surveillance system.

ACKNOWLEDGEMENT

The researcher is indebted to Mr. Sunny Nahar for his continuing support and assistance with this research project. The researcher thanks the Department of Master of Computer Application for permitting the researcher the use of their facilities and providing the resources necessary for conducting this research.

X. REFERENCES

1) J. Canny, "A Computational Approach to Edge Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

2) M. Naphade et al., "On the Detection of Semantic Concepts at TRECVID," ACM International Conference on Multimedia, 2004.

3) T. Mandrupkar, M. Kumari, and R. Mane, "Smart Video Security Surveillance with Mobile Remote Control," International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), 2013.

4) A. Hauptmann, "Lessons for the Future from a Decade of Informedia Video Analysis Research," International Conference on Image and Video Retrieval, 2006.

5) J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv preprint arXiv:1804.02767, 2018.

6) C. Stauffer and W. E. L. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking," IEEE CVPR, 1999.

7) R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting Moving Objects, Ghosts and Shadows in Video Streams," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 10, pp. 1337–1342, 2003.

8) N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," IEEE CVPR, 2005.