E-ISSN: 2583-9667
Indexed Journal
Peer Reviewed Journal
https://multiresearchjournal.theviews.in

Received: 18-01-2025
Accepted: 26-03-2025

INTERNATIONAL JOURNAL OF ADVANCE RESEARCH IN MULTIDISCIPLINARY

Volume 3; Issue 2; 2025; Page No. 297-301

Shoplifting Detection System Using YOLOv5, MediaPipe Pose Estimation, and Machine Learning-Based Behavior Analysis

[1]Dr. Perumal S and [2]Maha Ganesh K

[1]Professor, Department of Computer Science and Information Technology, Vels Institute of Science, Technology and Advanced Studies, Pallavaram, Chennai, Tamil Nadu, India
[2]Student, Department of Computer Science and Information Technology, Vels Institute of Science, Technology and Advanced Studies, Pallavaram, Chennai, Tamil Nadu, India

DOI: https://doi.org/10.5281/zenodo.15590304

Corresponding Author: Dr. Perumal S

Abstract

Shoplifting represents a persistent and costly issue within the retail industry, contributing to considerable financial losses annually. To address this problem, this research introduces an intelligent surveillance system that integrates real-time object detection, human pose estimation, and behavioral analysis to proactively identify potential shoplifting incidents. The proposed system utilizes the YOLOv5 deep learning model for detecting individuals and store items in video feeds, MediaPipe for precise body pose landmark tracking, and a Random Forest classifier to categorize human behaviors as either suspicious or normal. The experimental results indicate that the system achieves a high detection accuracy with a minimal false positive rate, demonstrating its practical effectiveness in real-world retail scenarios.

Keywords: Shoplifting, YOLOv5, Pipe Pose, Estimation, Machine, Learning-Based, Behavior

1. Introduction
1.1 Background
Retail theft, particularly in the form of shoplifting, remains a persistent and costly challenge for businesses worldwide. According to global retail reports, shoplifting accounts for a significant portion of revenue losses, affecting both small businesses and large retailers alike. The increasing prevalence of such incidents not only undermines profitability but also compromises customer and employee safety.
To mitigate these losses, many retail establishments have adopted surveillance systems, most commonly through the deployment of Closed-Circuit Television (CCTV) cameras. While CCTV systems provide continuous monitoring of store premises, their effectiveness is often limited by reliance on manual observation by human security personnel. This traditional approach poses several challenges:

- Human fatigue due to prolonged monitoring, which can lead to oversight of critical events.
- Inefficiency in analyzing large volumes of video footage in real time.
- Subjectivity and inconsistency in detecting suspicious behavior.

As a result, many shoplifting incidents either go unnoticed or are identified too late for effective intervention.

1.2 Problem Statement
Despite advancements in surveillance technologies, retail stores continue to face challenges in effectively preventing shoplifting due to:

- The limitations of manual monitoring using conventional CCTV systems.
- The lack of automated behavior analysis, which prevents timely detection of suspicious actions.
- The difficulty in differentiating normal customer behavior from potentially malicious intent without real-time intelligence.

These gaps underscore the need for a more robust, automated solution capable of continuously analyzing customer behavior, identifying potential theft attempts, and promptly notifying security personnel.

1.3 Research Motivation
The motivation behind this research stems from the growing need to:
- Enhance store security by reducing the burden on human personnel.
- Leverage recent advancements in computer vision and machine learning to develop a proactive solution.
- Minimize financial losses associated with shoplifting through timely detection and intervention.
- Improve operational efficiency in monitoring surveillance footage without compromising accuracy.

By addressing these objectives, the proposed system aims to significantly reduce the number of undetected shoplifting incidents while maintaining a seamless shopping experience for customers.

1.4 Proposed Solution
To overcome the limitations of traditional surveillance systems, this research proposes the development of an automated shoplifting detection system that integrates:
- **Object Detection:** Identifying individuals and store products using state-of-the-art models such as YOLOv5.
- **Pose Estimation:** Monitoring human body movements with MediaPipe to detect unusual actions related to theft.
- **Behavioral Classification:** Analyzing motion patterns using a machine learning model (Random Forest) to classify behaviors as either "Normal" or "Suspicious (Shoplifting)."

The system processes real-time video feeds from CCTV cameras, continuously analyzes human-object interactions, and alerts security personnel when suspicious behavior is detected. This multi-stage approach enables early detection of potential shoplifting activities while reducing dependency on manual surveillance.

1.5 Objectives
The primary objectives of this research are:
- To design and implement an automated system for real-time shoplifting detection using computer vision techniques.
- To evaluate the system's accuracy and effectiveness in differentiating between normal and suspicious behavior.
- To develop an intuitive user interface that allows security personnel to monitor live video streams and receive alerts for potential theft incidents.

1.6 Scope of the Study
The scope of this study includes:
- Implementation of object detection, pose estimation, and behavior classification models.
- Real-time processing of live CCTV feeds or pre-recorded video footage.
- Development of a user-friendly front-end interface for visualization and monitoring.
- Evaluation of system performance using publicly available datasets and controlled experiments.

The study does not focus on other forms of retail theft (e.g., employee theft or fraudulent returns) but instead centers specifically on shoplifting detection through human-object interaction analysis.

2. Methodology
The proposed shoplifting detection system is architected around three major functional modules, each responsible for a critical component of the detection pipeline. These modules are: Object Detection, Pose Estimation, and Behavioral Classification.

2.1 Object Detection with YOLOv5
The first stage of the system leverages YOLOv5 (You Only Look Once, version 5) for object detection within each frame of the input video stream. YOLOv5 is selected for its exceptional balance between speed and accuracy, making it highly suitable for real-time surveillance applications.
YOLOv5 is capable of detecting and classifying multiple object types simultaneously. In this system, it primarily identifies two categories:
- Persons (potential shoplifters)
- Store products (items at risk of theft)

For every detected entity, YOLOv5 assigns a bounding box and a corresponding class label (e.g., person, bottle, bag). This information serves as a foundational layer for subsequent modules that analyze interactions between detected persons and store products.

2.2 Pose Estimation using MediaPipe
The second component integrates MediaPipe Pose, a state-of-the-art real-time pose estimation framework developed by Google. MediaPipe detects 33 anatomical landmarks per person, including but not limited to:
- Wrists
- Elbows
- Shoulders
- Hips
- Knees

The system places particular emphasis on hand and hip movements, which are common indicators of suspicious behavior. For example, shoplifters often reach for products and attempt to conceal them in hidden areas such as:
- Behind a backpack
- Near the waistline
- Inside clothing pockets

Tracking these movements in real time allows the system to extract meaningful behavioral cues from the video stream.
2.3 Behavioral Classification using Random Forest
The final module employs a Random Forest classifier to categorize human behavior based on spatiotemporal features extracted from the previous stages. These features include:
- Velocity and trajectory of hand movements
- Distance between hands and nearby detected objects
- Proximity of objects to predefined concealment zones (e.g., backpack, pockets)

Using this feature set, the Random Forest model classifies each action into one of two categories:

- Normal – Regular shopping activity
- Suspicious (Shoplifting) – Behavior indicative of potential theft

The classifier is trained on a labeled dataset of human activities and is continuously updated to improve its predictive performance. This ensures that the system becomes more robust and accurate over time as more behavioral data is collected.

3. Implementation

The implementation phase of the shoplifting detection system translates the proposed methodology into a fully functional pipeline. This section describes, in detail, each step of the process — from video input acquisition to real-time visualization of detection results.

3.1 Input Source Acquisition

The system is designed to be compatible with two primary types of input sources:

- **Live CCTV Camera Feeds**: Enables real-time surveillance in retail environments. This mode requires access to a connected camera or IP stream.
- **Pre-recorded or Uploaded Video Footage**: Allows testing and analysis of previously captured video data. Useful for training, evaluation, and retrospective analysis of shoplifting incidents.

In both cases, the input is streamed frame-by-frame for processing, ensuring that real-time and recorded footage are handled with the same computational pipeline.

3.2 Frame-by-Frame Processing

To handle video streams effectively, the system utilizes OpenCV, a powerful open-source computer vision library. Each incoming frame undergoes the following sequential operations:

- **Frame Capture:** The frame is read from the video stream and resized if necessary to maintain consistent resolution.
- **Object Detection with YOLOv5:** Objects such as individuals and store items are detected and annotated with bounding boxes and class labels.
- **Pose Estimation with MediaPipe:** Human pose landmarks are extracted from each detected person, providing fine-grained spatial information about their posture and movements.

This frame-by-frame pipeline ensures that all subsequent analysis is grounded in accurate, per-frame data.

3.3 Multi-Frame Object Tracking

A crucial aspect of behavior recognition is tracking detected entities (e.g., people, products) across multiple frames. For this purpose, a lightweight tracking mechanism is implemented to:

- **Assign Persistent IDs:** Unique identifiers are given to each tracked object or person, enabling temporal continuity.
- **Handle Occlusions and Reappearance:** Tracking logic ensures robustness even if an object disappears momentarily and reappears later in the scene.
- **Monitor Trajectories:** By comparing positions of tracked entities across frames, their speed and direction of movement are estimated.

This tracking module plays a central role in understanding the context of actions over time, rather than analyzing them in isolation.

3.4 Feature Extraction

Once entities are tracked, spatiotemporal features are extracted from the data. These features form the basis for behavior classification and include:

- **Movement Velocity and Direction:** Helps distinguish between casual browsing and abrupt or evasive movements.
- **Hand-to-Object Distance:** Determines whether a person's hand is interacting with or reaching for a store item.
- **Proximity to Concealment Areas:** Analyzes the position of store items relative to suspicious regions like the hips, backpack zones, or coat pockets.

Each of these features is computed on a per-individual basis across time, producing a rich dataset for behavioral analysis.

3.5 Behavior Classification using Random Forest

The extracted features are passed to a pre-trained Random Forest classifier, which makes a decision for each observed action. The classification model is trained on a labeled dataset and outputs:

- **"Normal":** Indicating no suspicious activity.
- **"Suspicious (Shoplifting)":** Indicating potential theft based on learned behavioral patterns.

The use of Random Forest ensures robustness and interpretability, as the ensemble of decision trees captures complex decision boundaries while mitigating overfitting.

3.6 Real-Time Monitoring Interface with Streamlit

To ensure the system is user-friendly and deployable in real-world retail environments, a web-based user interface is developed using Streamlit. Key features of the interface include:

- **Live Video Display:** Annotated video feed with bounding boxes, pose skeletons, and labels.
- **Status Notifications:** Real-time alerts when suspicious behavior is detected.
- **Logging and Reports:** History of events detected, including timestamps and visual evidence.
- **Upload and Analysis Tools:** For non-real-time use, users can upload videos for analysis through the same interface.

The Streamlit front-end makes the system easily accessible to non-technical users and allows seamless integration with existing retail surveillance setups.

4. Results

The system's performance is evaluated using various visual and analytical metrics:

Fig 1: YOLOv5 and Media pipe Detection Output
Displays bounding boxes around detected people and products in each video frame.

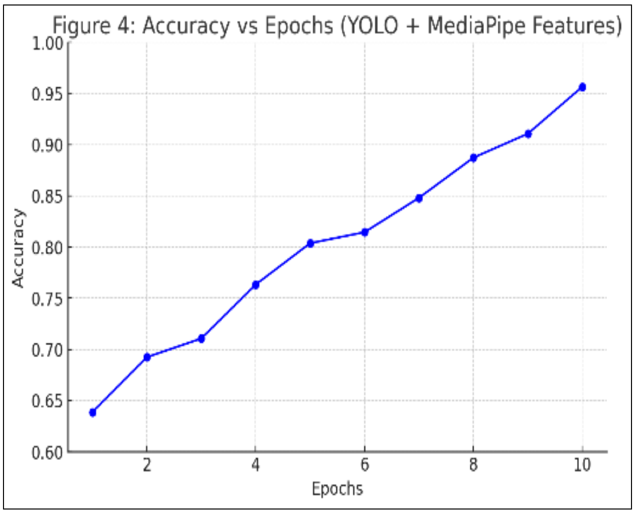

Fig 2: Accuracy vs Epochs Graph
Visualizes the learning curve of the behavioral classification model during training.

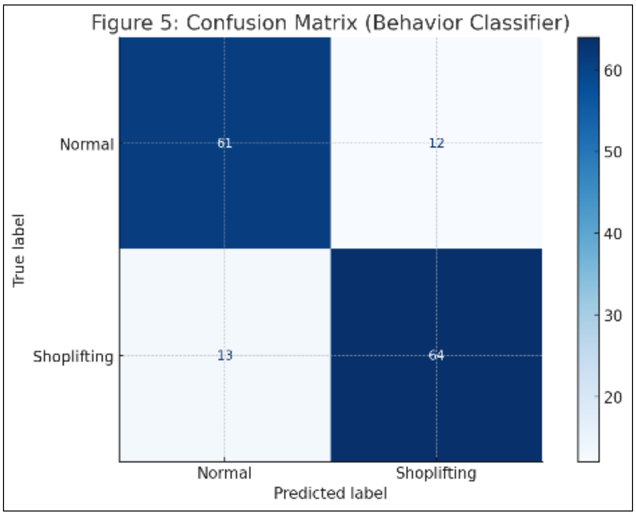

Fig 3: Confusion Matrix Breaks down the classifier's predictions into true positives, false positives, true negatives, and false negatives.

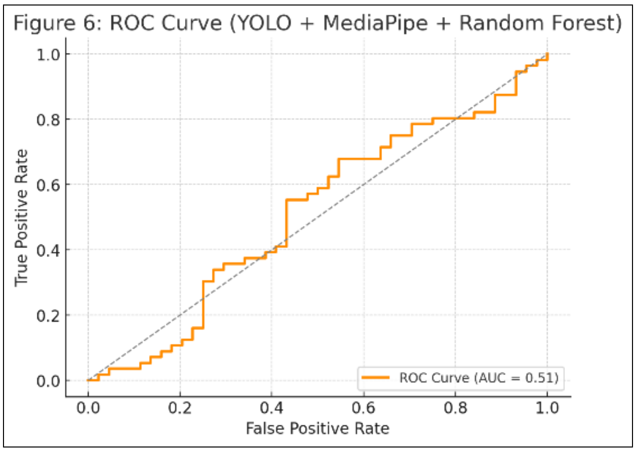

Fig 4: ROC Curve Illustrates the trade-off between sensitivity and specificity, highlighting the model's thresholding efficiency.

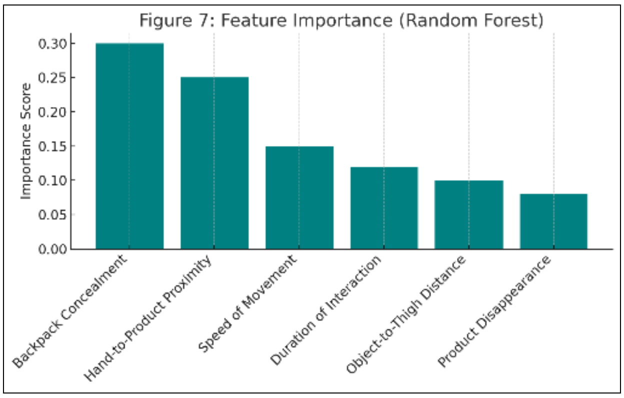

Fig 5: Feature Importance Plot Ranks the most influential features contributing to the model's decision-making process.

5. Conclusion and Future Work

This research presents a robust, automated system capable of detecting shoplifting behaviors in real time with high accuracy and a low false positive rate. The integration of YOLOv5, MediaPipe Pose, and Random Forest classification proves to be an effective combination for real-world surveillance applications.

6. Future enhancements will focus on

- Scaling the system to handle inputs from multiple camera angles and locations.
- Improving pose tracking under occlusions and varying lighting conditions.
- Integrating facial recognition or customer ID tracking to detect repeat offenders.
- Deploying the solution on edge devices for real-time inference on-site.

7. References

1. Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv. 2020; arXiv:2004.10934. Available from: https://arxiv.org/abs/2004.10934
2. Google Research. MediaPipe Pose: Real-time human pose estimation [Internet]. 2020 [cited 2025 Jun 3]. Available from: https://google.github.io/mediapipe/
3. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, *et al*. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research. 2011;12:2825–30.
4. Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv. 2018; arXiv:1804.02767. Available from: https://arxiv.org/abs/1804.02767
5. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv. 2015; arXiv:1412.6980. Available from: https://arxiv.org/abs/1412.6980
6. Streamlit Inc. Streamlit – The fastest way to build data apps [Internet]. 2019 [cited 2025 Jun 3]. Available from: https://streamlit.io
7. Yadav JS. Shoplifting dataset [Internet]. Kaggle; 2024 [cited 2025 Jun 3]. Available from: https://www.kaggle.com/datasets/jashwantsinghyadav/shoplifting-dataset