

PROJECT ABSTRACT

Shoplifting Detection using Pose Estimation

A MediaPipe pose pipeline that scores concealment-like gestures in store footage and queues advisory alerts for human review — not evidence, with an explicit privacy and ethics boundary.

01. Title

Shoplifting Detection using Pose Estimation

02. Introduction / Background

Retail shrinkage costs stores real money, and reviewing hours of CCTV to find one concealment gesture is exactly the kind of task humans do badly: attention fades, footage piles up, and incidents are found hours too late. Pose estimation offers a different angle from raw video classification — instead of asking whether a frame looks like theft, it tracks what the body is doing: where the hands go, how long they stay, whether a product-handling motion is followed by a pocket-zone motion. This project builds that pipeline: MediaPipe pose extracts real-time body landmarks, a tracking layer follows hand and wrist positions across frames, and a gesture-heuristics engine scores concealment-like patterns sustained over a time window. The output is deliberately an advisory alert log for human review, never an accusation — the honest framing for a student system with a high false-positive rate, with the privacy and ethics boundary documented explicitly.

03. Problem Statement

Stores cannot afford continuous human review of all CCTV footage, but any automated flagging must be honest about its limits: gestures are not intent, and false accusations harm real people. The need is a pose-based pipeline that tracks body landmarks in store footage, scores concealment-like gesture patterns, queues advisory alerts with frame crops for human review, and evaluates its alert rule on staged footage — while being explicit that alerts are not evidence and testing must use staged or consenting-subject footage only.

04. Objectives

The objectives are to extract real-time body landmarks from store footage with MediaPipe pose; to track hand and wrist positions across frames with dwell-time measurement; to design gesture heuristics for concealment-like behaviors; to implement a time-window alert rule over sustained patterns; to present alerts as a

human review queue with frame crops; and to evaluate the alert rule's precision and recall on staged footage via the included notebook.

- Extract real-time body landmarks with MediaPipe pose.
- Track hand and wrist positions across frames with dwell-time measurement.
- Design gesture heuristics for concealment-like behaviors.
- Implement a time-window alert rule over sustained patterns.
- Present alerts as a human review queue with frame crops.
- Evaluate alert-rule precision and recall on staged footage during the build.

05. Proposed System / Working

Store footage frames are sampled and MediaPipe pose extracts body landmarks for each visible person. Hand and wrist landmarks are tracked across frames; dwell time in torso and pocket zones is measured per person. The gesture-heuristics engine scores concealment-like behaviors: hand-to-torso contact with rapid retraction, repeated reaching motions, and product-handling followed by pocket-zone motion. Gesture patterns sustained over a time window become advisory alerts with a timestamped log entry and a frame crop. The demo app presents alerts as a review queue — a human watches each flagged segment and decides, because the system never accuses. The notebook runs the staged evaluation: labeled actors perform normal shopping and concealment scenarios, and the alert rule's precision and recall are computed.

- MediaPipe pose extraction of body landmarks per person.
- Hand and wrist tracking across frames with dwell-time measurement.
- Gesture heuristics: hand-to-torso contact, rapid retraction, repeated reaching, product-then-pocket patterns.
- Time-window alert rule over sustained gesture patterns.
- Advisory alert log with timestamps and frame crops.
- Demo app presenting alerts as a human review queue.
- Notebook evaluation: alert-rule precision and recall on staged footage.

06. Technologies / Components Used

The pipeline uses Python 3.10 with MediaPipe for real-time body landmarks and OpenCV for video capture, landmark overlays and frame crops. A gesture-heuristics engine in Python scores dwell times and motion patterns over time windows; NumPy and pandas manage gesture features and alert logs; Matplotlib plots gesture timelines and evaluation results; a Jupyter notebook runs the staged-footage evaluation procedure; and a Flask app serves the demo with a footage-upload and alert-review interface.

- Python 3.10, OpenCV (video capture, landmark overlays, frame crops).
- MediaPipe pose (real-time body landmarks, Lugaresi et al.).

- Gesture heuristics engine (hand-to-torso rules, dwell times, time windows).
- NumPy, pandas (gesture features and alert logs).
- Jupyter notebook (staged-footage evaluation procedure).
- Matplotlib (gesture timelines and evaluation plots).
- Flask demo app with footage upload and alert-review interface.

07. Key Features

Key features include MediaPipe pose landmark extraction, hand and wrist tracking with dwell-time measurement, gesture heuristics for concealment-like behaviors, a time-window alert rule, an advisory alert log with timestamps and frame crops, a review-oriented demo app, and a buyer-run staged-footage evaluation computing the alert rule's precision and recall — always framed as advisory, never as evidence.

- MediaPipe pose extracting real-time body landmarks.
- Hand and wrist tracking with dwell-time measurement.
- Gesture heuristics for concealment-like behaviors.
- Time-window alert rule for sustained patterns.
- Advisory alert log with timestamps and frame crops.
- Review-oriented demo app with alert queue.
- Staged-footage evaluation: alert-rule precision and recall.

08. Applications / Use Cases

Pose-based gesture alerting applies to retail footage-review prioritization, loss-prevention training, CCTV analytics research, and store-layout studies. It also demonstrates real-time pose estimation, landmark tracking, and heuristic rule design as viva-ready computer-vision concepts — always framed as advisory tooling with an explicit ethics boundary.

- Retail footage-review prioritization prototypes.
- Loss-prevention training material.
- CCTV analytics research.
- Store-layout and traffic studies.
- Teaching pose estimation and landmark tracking.

09. Results & Scope

The expected outcome is a working gesture-alerting pipeline whose quality is measured, not asserted: the included notebook replays staged footage — labeled actors performing normal shopping and concealment scenarios — and computes the alert rule's precision and recall on that labeled set. Scope covers upper-body-visible store footage from fixed camera angles; documented limits include the high false-positive rate (adjusting clothing, pocketing a phone, holding items against the torso all

resemble concealment), occlusion, crowded aisles, and landmark quality at off-angles. Alerts are advisory only, NOT evidence: the system must never be used to accuse, detain or report any real person. Testing is restricted to staged footage with actors or consenting subjects with clear notice; running it on real customers without consent is a privacy violation. Future scope includes product-interaction tracking, multi-person scene analysis, and zone-based dwell analytics.

- Measured outcome: alert-rule precision and recall computed by the notebook on staged footage.
- Staged scenarios with labeled actors performing normal and concealment behaviors.
- Scope: upper-body-visible store footage from fixed camera angles.
- Documented limits: high false-positive rate, occlusion, crowds, off-angle landmarks.
- Honest framing: advisory only, never evidence; staged/consenting footage for testing only.
- Future scope: product-interaction tracking, multi-person analysis, zone dwell analytics.

10. Conclusion

MediaPipe pose plus gesture heuristics plus a time-window alert rule makes suspicious-gesture flagging a tractable student-buildable system, and this project delivers it as a complete auditable experiment: pose pipeline, gesture engine, alert logic, review-queue demo, and a notebook that measures the alert rule's precision and recall on staged footage. Built to order, it ships with source code, the evaluation notebook, report, presentation, and viva preparation material.

11. References

The project cites the MediaPipe framework, a pose-based action-recognition method, and a gesture-analysis reference.

- V. Lugaresi et al., 'MediaPipe: A Framework for Building Perception Pipelines,' arXiv:1906.08172, 2019.
- S. Yan, Y. Xiong, and D. Lin, 'Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition,' Proc. AAAI, 2018.
- Z. Cao et al., 'OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,' IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 43, no. 1, 2021.