AI-POWERED SUPERMARKET ARMED ROBBERY DETECTION SYSTEM USING COMPUTER VISION

i. Basic Information

Name of Student: **AYELLA PATRICK OJOK**

Program: **Bachelor of Science in Computer Engineering**

Supervisor: **Dr. Godfrey Mirondo K (PhD)**

Date: **29th Sept, 2025**

Dissertation Topic: **Development of a Real-Time Computer Vision System for Detecting Armed Robbery in Supermarkets**

ii. Background

The retail sector, particularly supermarkets, is increasingly vulnerable to armed robberies, leading to significant financial losses and posing grave threats to the safety of employees and customers.

According to a 2022 report by the Uganda Police Force, armed robbery incidents increased by 30% over the past three years, with supermarkets being primary targets. These include A$B Supermarket near Namugongo catholic Shrine which was attacked by two armed robbers whereby unspecified amount of money was taken after shooting a security guard dead as reported by Kira Division Police Commander. More so on 24th June 2024, report by Mubiru Ivan in National news indicated that three robbers armed with AK47 attacked Spice Supermarket located in Mukono Municipality and made away with unspecified amount of money after shooting Sekandi Irene the cashier in the back. Traditional security measures, such as CCTV systems, are predominantly passive, serving as forensic tools for post-incident investigation rather than proactive mechanisms for crime prevention. The critical gap lies in the inability of these systems to provide real-time alerts, allowing perpetrators to escape before law enforcement can respond. Recent advancements in computer vision offer a transformative solution. Technologies like object detection and human pose estimation can automate the analysis of live surveillance feeds to identify the visual signatures of an armed robbery in progress. By leveraging deep learning models such as YOLO (You Only Look Once) for real-time object detection and OpenPose for pose estimation, it is feasible to create an intelligent system that transitions security from a reactive to a proactive stance, enabling immediate intervention.

iii. Problem Statement

Supermarkets are soft targets for armed robbers, resulting in trauma, loss of life, and financial damage. The existing CCTV infrastructure fails to provide timely alerts during such incidents, as security personnel cannot monitor all feeds simultaneously, and the footage is typically reviewed only after the crime has occurred. This delay in response allows criminals to execute their plans and escape, highlighting an urgent need for an automated, real-time detection system that can instantly identify robbery events and alert authorities.

iv. Purpose

To design and develop an integrated computer vision system that automatically detects active armed robbery in supermarkets by identifying weapons, suspicious attire, and coercive human postures in real-time, and triggers immediate alerts to security personnel and law enforcement.

v. Objectives

To develop a robust object detection module using a YOLO-based model to identify firearms and suspicious attire (e.g., ski masks, helmets) in real-time CCTV footage.

To implement a human pose estimation module using an OpenPose-based model to recognize coercive postures indicative of a robbery, such as individuals raising their hands or kneeling down.

To create a weighted decision algorithm that synthesizes detections from both modules to confirm a robbery event with high accuracy while minimizing false positives.

To integrate the detection system with a notification module that automatically alerts security officials via a reliable communication channel (phone call &SMS) when a threat is detected.

vi. Research Questions

What is the optimal combination of object and posture detection features to achieve high accuracy in identifying supermarket armed robberies?

How can a weighted scoring algorithm be designed to effectively balance the contributions of various robbery indicators (weapons, masks, postures) to minimize false alarms?

What level of precision and recall can be achieved by the integrated YOLO and OpenPose models in the dynamic and occluded environments of a supermarket?

What are the practical challenges in deploying this real-time computer vision system using existing supermarket CCTV infrastructure?

vii. Scope of the Study

This research will focus on developing a detection system for indoor supermarket environments. The scope is limited to specific, high-fidelity visual indicators:

Objects: Firearms (handguns, rifles) and face coverings (ski masks, helmets).

Poses: Raised hands and kneeling postures.

The system will be designed to process video from standard CCTV cameras. It will function as a detection and alert system, and the study will not extend to automated physical intervention or facial recognition due to privacy considerations. Testing will be conducted using a curated dataset and simulated scenarios.

viii. Justification

This project is critically justified by the escalating threat of armed robberies in retail spaces and the demonstrable inadequacy of passive surveillance systems. The tragic consequences of these crimes, including loss of life and economic damage, necessitate a technological leap in security protocols. Investing in an AI-powered, proactive detection system is a imperative to protect human life, safeguard business assets, and foster a secure commercial environment.

ix. Significance

The successful implementation of this system will represent a significant advancement in retail security. Its primary significance lies in its potential to save lives and prevent injuries by enabling a near-instantaneous police response. Furthermore, it will protect businesses from financial ruin, thereby contributing to economic stability. Academically, it provides a practical framework for applying multi-modal computer vision to a critical real-world problem, with potential for adaptation in other vulnerable retail environments.

x. Theoretical/Conceptual Framework

The system will be grounded in an integrated application of deep learning for computer vision. The framework follows a multi-stage pipeline:

Input: Live video feed from supermarket CCTV cameras.

Processing: Each frame will be processed by two parallel deep learning models:

YOLO-based Object Detector: A Convolutional Neural Network (CNN) that localizes and classifies objects (guns, masks) in real-time.

OpenPose-based Pose Estimator: A CNN that detects 2D human body keypoints (e.g., wrists, elbows, knees, ankles) to reconstruct and classify postures.

Fusion and Decision Making: A rule-based algorithm synthesizes the outputs. Detections shall be assigned weighted scores based on their reliability and importance (e.g., a firearm scores higher than a raised hand). The scores are will be summed, and if a predefined threshold is exceeded, a robbery is confirmed.

Output: The confirmed event triggers an immediate alert to security personnel and police.

Literature review

The application of computer vision in security is a well-established research domain. Previous work has largely focused on singular approaches. Studies like Michał Grega et al. (2016) demonstrated automated firearm and knife detection in CCTV images, highlighting the efficacy of object detection for security. Meanwhile, Zhe Cao et al. (2017) made significant strides in real-time multi-person 2D pose estimation with OpenPose, enabling complex human behavior analysis.

The paper "Bank Robbery Detection System Using Computer Vision" by Sharma et al. is a direct predecessor to this work, successfully integrating YOLO for object detection and OpenPose for pose recognition in a financial context. Their use of a weighted scoring formula to fuse detections provides a valuable methodological template. However, a gap exists in tailoring this integrated approach to the unique challenges of a supermarket environment, which has different layouts, lighting, and customer densities compared to banks, more so there was no real-time alert to the Authority. This project will adapt and refine this integrated model, focusing on the specific threat indicators and operational realities of supermarkets and trigger an alert to the nearest police station to fill this research gap hence enhancing the state of the art in retail security.

xii. Methodology

The research will adopt the Design Science Research Methodology (DSRM), involving the creation and evaluation of an artifact—the detection system.

Data Collection & Preparation: A dataset will be curated from open-source videos and simulated scenarios, annotated for weapons, masks, and relevant poses. The dataset will be split into training and validation sets.

Model Development & Training:

Object Detection: A YOLOv8 (or newer) model will be fine-tuned on the custom dataset using transfer learning.

Pose Estimation: A pre-trained OpenPose model will be used and adapted to recognize "raised hands" and "kneeling" based on the coordinates of keypoints (wrists, elbows, knees, ankles).

System Integration & Algorithm Design: The two models will be integrated into a single application using Python. A decision algorithm will be implemented, using a refined version of the scoring formula from the source paper:

score = (hndup/num * weight_pose) + (knldn/num * weight_pose) + (mask_count * weight_mask) + (gun_count * weight_gun).

Thresholds and weights will be tuned empirically.

Testing & Evaluation: The system will be evaluated on its:

Accuracy: Precision, Recall, and F1-Score.

Speed: Frame processing rate (FPS) to ensure real-time performance.

Robustness: Low false positive rate in non-threatening scenarios.

xiii. References

Ultralytics. (2023). YOLOv8 Documentation. Retrieved from https://docs.ultralytics.com/

Jocher, G., et al. (2023). ultralytics/ultralytics: YOLOv8. GitHub. https://github.com/ultralytics/ultralytics

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields.

Sharma, A. K., Mittal, P., Ranjan, R., & Chaturvedi, R. (Bank Robbery Detection System Using Computer Vision). [Adapted for Supermarket Context].