Advanced Security Camera System Software using Pose Estimation, Facial Recognition and Weapon Detection

Dhananjay Sunil Menon[1], Shreya Chaudhary[2], Rakesh B Krishnan[3] and Dr. Shashank Mouli Satapathy[4]

dhananjaymenon@hotmail.com, shreya.shreya@gmail.com, rakeshbkrishnan2000@gmail.com

[1,2,3,4] School of Computer Science and Engineering, VIT University, Vellore, Tamil Nadu, India

----------***----------

Abstract - *The advanced security system recognizes faces of employees and estimates their pose. In case on a robbery, an alarm alert can be sent to the police station by registering a body gesture from the employee. Face recognition is a technique that identifies or locates human faces in digital images. A typical example of face detection occurs when we take photographs through our smartphones, and it instantly detects faces in the picture. Face detection is different from Face recognition. Face detection detects merely the presence of faces in an image while facial recognition involves identifying whose face it is. Face Recognition has many applications in biometrics, surveillance systems, information security and law enforcement.*

Key Words **security system, facial recognition, pose estimation, surveillance system, law enforcement.**

1. Introduction

Cameras are installed in many public places to improve security, and this works well with motion detectors if it is not necessary to automatically identify the person in the picture. This project aims to improve real-time alerts and overall bank safety by embedding advanced object, face, and gesture recognition features in security footage. Previously, security cameras were used to record video of everything that happened in the bank. Security cameras equipped with advanced security software can send real-time alerts to emergency services in the event of a robbery. When a face is detected in an image, it can be identified or matched against a face database.

The advanced camera system has mainly 3 modules. First being a pose detection system to detect if a bank robbery is taking place. The approach used for detecting pose and hand gestures would be through MediaPipe (Cao, Z., et all, 2019) Face detection would also be used to identify the employee when they are holding the pose. The approach used for face detection is by using a single picture for training and identification (Ji, H. K., et all, 2017). The second module is a quick approach to identify and retrieve information of people who enter the frame of the camera. The images of people are saved and are compared to a database containing pictures of many individuals and their details by using facial detection. Deepsort is the object tracking algorithm used to detect and track people visible in the camera footage. The third module is on weapon detection. Yolo algorithm is used for the detection as it is the fastest object recognition algorithm to date. An alert is activated in case a weapon is detected by the system.

Face recognition function is related to computer vision and image processing that is used to detect a human face and identify the individual by matching the detected face against a pre-defined database. MediaPipe is the proposed library to be used for face detection. Face_recognition in a prebuilt library which would be tested against the MediaPipe algorithm. MediaPipe draws points on the face, which is known as a face mesh. This can be used to create a distinct signature for the face and recognize the face based on the signature.

The Identity Profile Retrieval is used to retrieve a target individual's identity profile from a database. The identity profile retrieval is used to search for a particular person who walks into the screen even for a moment. Their picture and timestamps are saved in a folder. With this functionality, it would be easier to search and find a particular person without having to go through the entire video. Object Tracking is mainly used in this module. For this, the python library deep_sort is used, along with Yolo (for object recognition). Future Improvements of this feature would include finding the person from a database of photos.

Target object detection is related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class in digital images by use of a pre-defined dataset. The project will use functionalities such as face recognition, object detection, hand gesture recognition, identity profile reveals and

pose estimation. Yolo algorithm and OpenCV is used to search for weapons from a video. Future Improvements of this feature would include being able to identify partially concealed weapons from a person. For this, a better trained dataset would be required.

2. Literature Review

Multimodal dynamic sign language recognition method based on a deep 3-dimensional residual ConvNet and bidirectional LSTM networks, which is named as BLSTM-3D residual network (B3D ResNet). Despite recognising hands, it does not estimate the sign held up(Y. Liao, et all, 2019) Pose Estimation with mediapipe allows us to use 32 joint points on the body and 21 joint points on each hand. This helps for better evaluation of pose. Face Recognition is also implemented along with pose estimation(J.Hwang, et all, 2020)

Facial recognition was the motivation for the creation of eigenfaces. For this use, eigenfaces have advantages over other techniques available, such as the system's speed and efficiency. As eigenface is primarily a dimension reduction method, a system can represent many subjects with a relatively small set of data. Fisherface is the algorithm that is used for face detection. Fisherface is more complex than Eigenface in finding the projection of face space. Accessories, cosmetics and lighting can affect accuracy(Mustamin Anggo La Arapu, 2018).

Facial Recognition requires only 5 images of an employee's face to save into the system. Another program is used for taking these 5 images clearly (with faces detected in each shot). This improves the usability of the software (Kai Xiao,et all. 2020)

Object recognition is one of the research areas in the field of computer vision and image processing because of its varied applications in surveillance and security systems, biometrics, intelligent vehicle system, and content based image retrieval, etc. The main goal of this survey is to present a comprehensive study in the field of 2D object recognition (Bansal, Monika, et al, 2021). Two images with similar color histograms can possess different contents. Models often misclassify images even in the presence of correctly classified foregrounds. DNNs are extremely fragile against adversarial attacks (Robert Geirhos, et all, 2017). Yolo algorithm and OpenCV is used to search for weapons from a video. YOLO is a clever convolutional neural network (CNN) for doing object detection in real-time. Our approach will try to recognize partially visible weapons. This would require extensive training. However, it would be important to prevent false positive predictions as well.

OpenPose is used for pose estimation. MediaPipe is used for hand gesture recognition (Robert Geirhos, et al). The accuracy rate is high but only the right hand sign language is recognized.

Face Feature Extraction is done by using algorithms such as LDA (linear discriminant analysis), geometric feature, subspace analysis and neural network method (Hao Yang and Xiaofeng Han). This has higher accuracy than traditional systems but accessories, cosmetics and lighting can affect accuracy.

3. Proposed Methodology

The project will use functionalities such as face recognition, object detection, hand gesture recognition, identity profile reveals and pose estimation.

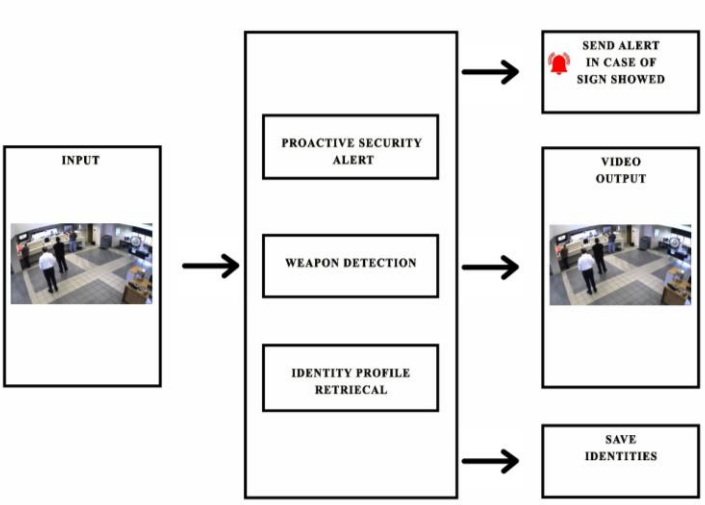


Figure 1: Proposed Approach

Here is a list of methodologies for each feature in figure 1.

- Proactive Security Alert
  - Pose Estimation
  - Hand Gesture Recognition
  - Face Detection
- Weapon Detection
  - Real-time Object Detection
- Identity Profile Retrieval
  - Object Tracking
  - Saving Timestamps
  - Identifying and saving information on object tracked

3.1 Proactive Security Alert

Pose Estimation would be used to detect the pose of an employee and detect a threatening gesture (like raising

the hands). **MediaPipe** is the python library that would be used for detecting body language.

When an employee is threatened, they may not be able to take any action to report a crime. Criminals may make it a point to stop employees from reaching under the desk. Hence cameras can report a crime if the employee makes a certain body and hand gesture.

To trigger the alarm three conditions should be met:

- Both arms of employee should be raised as shown in figure 3
- Hands should show a gesture as shown in figure 4.
- Arms should be up for at least 5 seconds

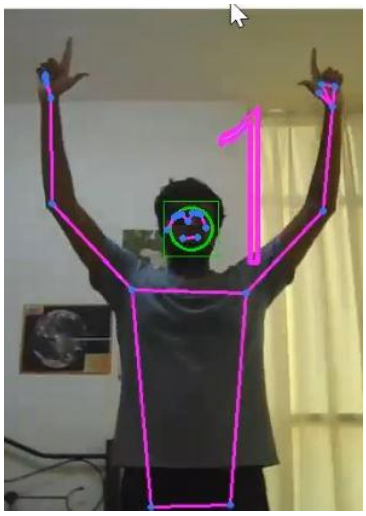
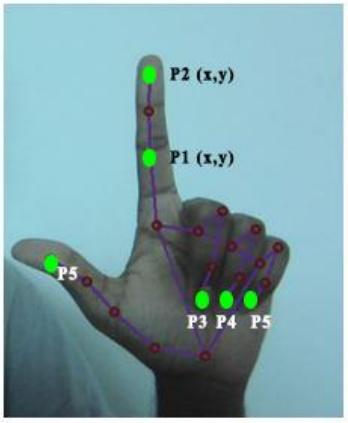


Figure 2: Pose detection

MediaPipe provides points for each body joint. This can help detect poses as sown in figure 3.

Hand Gesture Detection is an added security measure to prevent false positives (false alarms). The employee has to show a hand gesture for activating the alarm. An added timer of 5 seconds would be added to make sure that the signaling of alarm is legitimate.



Landmark	Point
Index_finger_pip	P1
Index_finger_tip	P2
Middle_finger_tip	P3
Ring_finger_tip	P4
Pinky_tip	P5
Thumb_tip	P6

	Conditions
1	P1.y > P2.y
2	P3.y > P1.y and P4.y > P1.y and P5.y > P1.y
3	Absolute Distance of nose from P2 is greater than absolute Distance of nose from P6

Figure 3: Hand gesture detection

Hand Gesture Detection is an added security measure to prevent false positives (false alarms). The employee has to show a hand gesture for activating the alarm that can be detected as shown in figure 3. An added timer of 5 seconds would be added to make sure that the signaling of alarm is legitimate.

MediaPipe creates 21 landmarks for one hand alone. With the help of conditions, we can create a function that can identify a specific hand gesture, For the identification of the hand gesture in our case, we can create the conditions listed in figure 3.

Face Detection is a feature for preventing anyone, apart from the employees, from setting off the alarm.

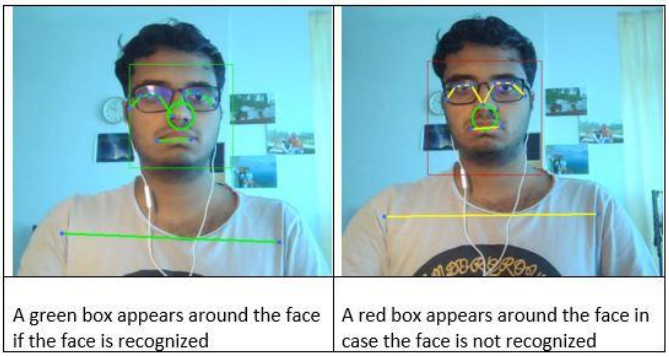


Figure 4: Facial recognition

The facial recognition requires only 5 photos of an employee to set up. This includes photos from different views (front, left, right, up, down). A program has been developed to capture and set up these photos in less than a minute. A python library, MediaPipe, was used to track a person's face and take it from different angles. These faces are compared to real time footage and one of two outcomes are displayed, as shown in figure 4.

3.2 Weapon Detection

This feature alerts employees if someone is holding a weapon that is visible to the camera. It is possible that a criminal wielding a weapon would go unnoticed in a crowd of people. This system would be faster than anyone else in detecting a weapon.

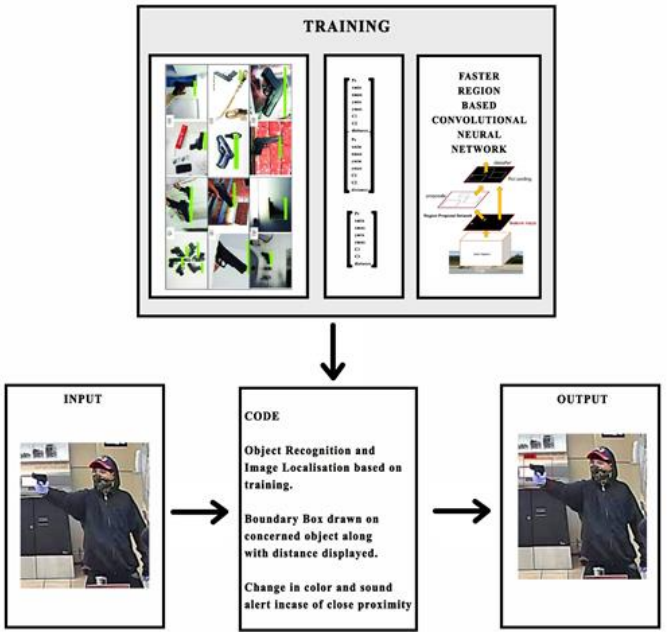


Figure 5: Working of the weapon dection module

For weapon detection, either one of two algorithms would be used for real-time object detection. The YOLO (You Only Look Once) algorithm and the Faster-RCNN algorithm are the two approaches shown in figure 5.

Both approaches would require using a labeled dataset and training code. The training code would be run on Google Colab (because of its free GPU resource).

3.3 Identity Profile Retrieval

A person can be recognized and tracked throughout a video segment using Object Tracking. The timestamps of the people on the screen can be saved and retrieved at a later time. Additional information about each individual can be noted and saved.

Figure 6: People detected

People who enter the frame have their screenshots saved along with the time stamps as shown in Figure 6 and each picture is saved with the time stamp first.

With so many people entering and exiting banks, saving and easily retrieving film of people involved in a crime might be beneficial to an investigation. Aside from the bank, this technology can be employed in other areas. This feature would be implemented using Object Tracking. Object tracking differs from object detection in that it requires an object tracker to follow a specific object (or person) throughout the movie.

4. Datasets used

For this study, two datasets are being used. The first dataset is a weapons dataset, which contains photos of various weapons and other dangerous things that will be utilized for weapon detection innovation. The second dataset is a database of various people's identity profiles, which will be utilized to perform the novelty function identity profile retrieval.

The weapon datasets, here the resources in this part are geared toward the creation of intelligent video surveillance automatic systems. An automatic weapon detection system can enable early identification of

potentially violent situations, which is critical for public safety. Detecting the presence of dangerous weapons such as pistols and knives in surveillance videos is one technique to avoid these situations. This type of object can be detected using Deep Learning algorithms based on Convolutional Neural Networks. Different techniques of combine a region proposal technique with a classifier, or incorporating both into one model, can be used to complete the weapon identification problem. Any deep learning model, on the other hand, must first learn a high-quality image dataset and an annotation that corresponds to the classification or detection tasks. Detection of weapons Open Data offers high-quality image datasets that can be used to train Deep Learning models for the creation of an automatic weapon identification system. Weapons datasets for image classification and object detection tasks are described and can be downloaded below. The public datasets are organized depending on the included objects in the dataset images and the target task.

The Pistol classification Dataset contains 9261 photos in total, divided into 102 item types. The handgun class contains 200 photos, while the other classes contain a variety of typical things such as an airplane, a chair, a cell phone, animals, and so on. This dataset is designed in the accompanying publication, which includes more details about the image dataset and experiment outcomes.

The Knife classification Dataset comprises 10039 images gathered from the internet and categorized into 100 item types. Apart from other handled objects like a pen, smartphone, or cigar, or common backdrop objects like a car, barrel, various plants, and some animals, the target class Knife contains 635 photos. This dataset is designed in the accompanying publication, which includes more details about the image dataset and experiment outcomes.

The Sohas weapon Dataset includes weapons and small things that are handled in a similar manner. It is divided into six categories: firearm, knife, bill, purse, smartphone, and card. The classification images are created by cutting out the object's bounding box from the detection photos. This dataset is designed in the accompanying publication, which includes more details about the image dataset and experiment outcomes.

Identity Profile Dataset is a custom dataset. It was built from the ground up by the members of this project. A binary file containing objects of the class "IDENTITY" makes up the dataset. "Name," "Age," "Born," and "Description" are the properties of the class. These characteristics will be integrated to form an individual's identity profile. To manipulate this dataset, a special application was constructed. The Python software can be used to add, alter, and delete records from the dataset. Each object of the class "IDENTITY" is a database record. The details known are then matched with the identity profile database in order to retrieve a record containing specific information once the face recognition function is done and the individual's face and name are identified. In real time implementation, if this software were to be implemented in a bank for example, the bank authorities may make use of legal systems to acquire access to high end profile databases such as the adhaar database or voters ID database. If the legal permissions can be attained, then even the NCRB (National Crime Records Bureau) database can be made use of for better and higher quality implementation. Since this is a college level project, the database for this particular project, as mentioned above, is custom made by the members working on this project.

5. Novelties

Identity Profile Retrieval: It's a feature that's been added to the face recognition feature. Identity profile retrieval is a function that takes an identifiable face as an input from a face recognition algorithm and utilizes it to cross-match the input against a big database in order to extract the individual's identity profile and other records.
For example: If an individual walks into a bank and attempts to use a fake identity at the bank, the system at the back may analyze the photograph of the individual and cross match it with a database such as adhaar database or the voters ID database in order to confirm the identity of the individual. The function will process the input photograph and return the real identity profile of the individual which will alert the bank officials that the individual is using a fake identity at the bank and necessary actions may be taken based of the given information.

Pose Estimation: Posture estimation is a computer vision and image processing approach that is used to determine human posture. It is a more advanced version of gesture

recognition. By analyzing digital photos and videos, it can recognize and identify particular human poses. For security needs, the function can be extremely valuable.

For example: A security camera armed with this function can be used in places such as banks or shops. The camera can analyze the footage it collects during surveillance and identify any individual who is suspicious. There are specific postures and actions that can be used to decide if a particular individual is acting suspicious or not. It can be used in shops to detect activities such as shoplifting.

Target Object Identification: Target object identification is a computer vision and image processing function that can be used to recognize and identify certain items in a digital image or video clip. It is an advanced form of the object detection technique. The function can be used to improve security.
For example: A security camera at a bank equipped with this ability can be used to identify any individual(s) that enter the bank with any malicious intent. The camera may analyze images or surveillance footage collected and uses this as input to detect any individual(s) who seem to be carrying any weaponry such as guns or knives. If such an individual is detected, the software may alert the concerned people and thus help to prevent a potential robbery or hostage situation at the bank.

6. Threats to validity

Bank robberies can be unpredictable, and various scenarios can take place. An employee may not be able to show the pose required to activate an alarm during the robbery. A robber may instruct a person to not to raise their hands. In such a case, stress levels on facial expressions could be used in the future to give better results.

Identity Retrieval will not work incase the robber is wearing a mask. However it is a common practice for robbers to visit a bank in plain clothes and study the scheduling and organization in banks. With the identity profile retrieval feature, identities can be narrowed down.

Weapon Detection works well when the weapon is openly visible. However, many robbers conceal their weapons when entering to rob a bank. However, provided they openly show the weapon to threaten the people present in the bank, the weapon detection system would work well.

The research has not been tested in real bank robbery footage. Other footages containing people and weapons have been used for testing the functionality of each component.

7. Conclusion

This project is an advanced security system camera software with functionalities such as proactive security system alert, weapon detection and identity profile retrieval. The proactive system alert detects if a robbery is taking place by scanning the pose of the employees and sends a notification to emergency services. Weapon detection uses object detection to find a weapon. Identity profile retrieval is used to get information on people if an incident has taken place.

8. Future Work

With increasing technology in various aspects of security in banks, security cameras have always been the same. Extensive investigation must be done after a crime to gather evidence and leads in finding criminals. With smart vision technology, a lot of time can be saved with proactive alarm systems and criminal information can easily be extracted after the crime. Such security features can make robberies more difficult and can dissuade such crimes in the future.

With high-tech security and low crime rates, people of a community can be safe and have peace of mind. Banks can also be saved from any loss of money. In some countries, security cameras are present extensively. This makes it easier to nab criminals and dissuade people from doing anything immoral and wrong.

This application can be extended to banks for better camera security. The required cameras can be installed at specific angles and locations in order to get the best view needed for the implementation of this application.

REFERENCES

[1] Dynamic Sign Language Recognition Based on Video Sequence with BLSTM-3D Residual Networks (Y. Liao, P. Xiong, W. Min, and J. Lu), 2019
[2] Exploring Rare Pose in Human Pose Estimation (J.Hwang, J. Yang and N. Kwak), 2020
[3] The Progress of Human Pose Estimation: A survey and taxonomy of Models Applied in 2D Human Pose Extimation (T. L. Munea, Y. Z. Jembre,

H. T. Weldegebriel, L. Chen, C. Huang and C. Yang), 2020
[4] Face Recognition Using Fisherface Method (Mustamin Anggo La Arapu), 2018
[5] Face Recognition Using Eigenface Approach (Marijeta Slavković Dubravka R. Jevtić),2012
[6] Noise or Signal: The Role of Image Backgrounds in Object Recognition (Kai Xiao, Logan Engstrom, Andrew Ilyas, Aleksander Madry),2020
[7] 2D Object Recognition Techniques: State-of-the Art Work (Bansal, Monika; Kumar, Munish; Kumar, Manish), 2021
[8] Comparing deep neural networks against humans: object recognition when the signal gets weaker (Robert Geirhos, David H. J. Janssen, Heiko H. Schütt, Jonas Rauber, Matthias Bethge, Felix A. Wichmann), 2017
[9] A New Approach for Video Action Recognition: CSP-Based Filtering for Video to Image Transformation (Robert Geirhos, David H. J. Janssen, Heiko H. Schutt, Jonas Rauber, Matthias Bethge, Felix A. Wichmann)
[10] Face Recognition Attendance System Based on Real-Time Video Processing (Hao Yang and Xiaofeng Han)
[11] Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2019). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *IEEE transactions on pattern analysis and machine intelligence*, *43*(1), 172-186.
[12] Choutas, V., Weinzaepfel, P., Revaud, J., & Schmid, C. (2018). Potion: Pose motion representation for action recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7024-7033).
[13] Parelli, M., Papadimitriou, K., Potamianos, G., Pavlakos, G., & Maragos, P. (2020, August). Exploiting 3d hand pose estimation in deep learning-based sign language recognition from rgb videos. In *European Conference on Computer Vision* (pp. 249-263). Springer, Cham.
[14] Gao, Q. X., Zhang, L., & Zhang, D. (2008). Face recognition using FLDA with single training image per person. *Applied Mathematics and Computation*, *205*(2), 726-734.
[15] Ji, H. K., Sun, Q. S., Ji, Z. X., Yuan, Y. H., & Zhang, G. Q. (2017). Collaborative probabilistic labels for face recognition from single sample per person. *Pattern Recognition*, *62*, 125-134.
[16] Gupta, S., Sahoo, O. P., Goel, A., & Gupta, R. (2010). A new optimized approach to face recognition using eigenfaces. *Global Journal of Computer Science and Technology*.