Advanced International Journal for Research (AIJFR)

E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

CCTV with Intruder Detection System

Sneha Shirur[1], Priya Navi[2], Surabhi Hasarangi[3], Vaishnavi Kathare[4], Shubha B. Baravani[5]

[1,2,3,4] UG Student, Department of Electronics and Communication, Maratha Mandal's Engineering College, Belagavi

[5]Associate Professor, Department of Robotics & Artificial Intelligence, Maratha Mandal's Engineering College, Belagavi

Abstract

With the advancement of technology, the detection of intruders has been introduced to reduce theft and criminal activities. Instead of merely recording theft incidents, the system aims to efficiently and accurately detect them in real time. The primary goal of this project is to introduce an innovative approach to closed-circuit television (CCTV) video image processing using computer vision. This system monitors and notifies users of potential threats, while also predicting a child's movements at home to prevent accidents. Real-time operation minimizes human effort and ensures accurate recognition through computer vision. This document outlines the design of an intrusion detection and danger prevention system for both indoor and outdoor home environments. The system employs image processing, face recognition, and machine learning for accurate prediction and alert generation. For instance, if a child approaches a dangerous area, such as a well in the garden, the system can predict the proximity based on pre-calculated distance parameters and send alerts to the user's mobile device or connected system. Similarly, the same predictive mechanism can be applied to detect and alert users about any intruders entering restricted areas within the home.

Keywords: Intrusion, Image Processing, Face Detection, Machine Learning, Rule-based Notification

1. Introduction

Surveillance involves monitoring human activities to ensure safety and detect suspicious behavior. Earlier, this was done manually, but with technological advancements, intelligent CCTV systems now automate the process. Due to rising crime and negligence, the need for effective surveillance has increased. Common techniques used in CCTV analysis include object tracking, background subtraction, suspicious activity detection, intruder detection, SVM, and Bayes decision theorem, sequence models, Multi-frame differencing method.

CCTV cameras record large amounts of footage, making manual monitoring difficult. Intelligent CCTV systems help by automatically identifying people and detecting intruders. This research aims to create a fast and economical home-security solution that can accurately detect intruders, differentiate familiar people, and send alerts based on movement.

1. With growing modernization, family safety has become a key priority.
2. Intelligent CCTV systems are now widely used because human actions and abnormal behaviors can be detected through computer vision.
3. Physical security alone is not enough to ensure a person's safety in all situations.
4. A proper detection system is necessary to monitor the perimeter, control access, and secure the area effectively.
5. Such systems can help track people's movements, predict a child's behavior, and identify suspicious activities before anything goes wrong.

2. Literature Survey

A. Face Detection

Face detection is the initial and most crucial step in any face recognition system. It focuses on identifying and isolating face regions from an image or video stream. Several algorithms are available for this task, and surveys have shown that Convolutional Neural Networks (CNNs) achieve the highest accuracy among them. CNN-based methods are significantly more reliable compared to techniques such as PCA with Decision Trees or KNN, often reaching accuracies of around 98% even with complex architectures.

YOLO (You Only Look Once) is a widely used end-to-end object detection framework that employs a single neural network to predict bounding boxes and class probabilities simultaneously. Its high speed and accuracy make it well-suited for real-time face detection and recognition systems. YOLO can be trained to detect faces using annotated face datasets and an appropriate network configuration.

The YOLO architecture is based on a CNN and includes two primary components:

- **Backbone network:** Extracts meaningful features from the input image.
- **Detection head:** Predicts bounding boxes and class probabilities.

During training, the model learns distinctive facial patterns and how to localize faces within an image. For prediction, YOLO divides the image into a grid and generates bounding box coordinates for each cell. One of the object classes is defined as "face," and regions with higher confidence scores indicate a detected face.

Since multiple bounding boxes may be generated for the same face, YOLO uses **Non-Maximum Suppression (NMS)** to remove duplicates. NMS compares overlapping boxes and retains only the one with the highest confidence score, ensuring accurate and clean detection results.

B. Face Recognition

Once the face regions are detected, the next step is **feature extraction and face recognition**, where the extracted face regions are compared with known identities. Face recognition is a biometric technique that identifies individuals based on their facial characteristics. Several methods have been introduced over the years, ranging from traditional approaches like PCA and LDA to modern deep learning–based models.

FaceNet is a deep convolutional neural network widely used for face recognition. It maps face images into a compact Euclidean embedding space, where the distance between two points directly reflects the similarity between the corresponding faces. FaceNet has achieved high accuracy on various benchmark datasets due to its ability to learn highly discriminative features. The model is trained on a large set of labeled face images and learns to extract unique facial features that distinguish one person from another. It generates **face embeddings**, which are numerical vectors representing the identity-specific characteristics of a face.

FaceNet uses a **siamese network architecture**, consisting of two identical subnetworks that share the same weights. This structure helps the model compare facial features effectively. Training is performed using a **triplet loss function**, which relies on triplets of images:

- an **anchor** image,
- a **positive** image of the same person, and
- a **negative** image of a different person.

The model is optimized to reduce the distance between the anchor and positive embeddings while increasing the distance between the anchor and negative embeddings.

As a result, the generated face embeddings lie in a high-dimensional space where:

- embeddings of the **same person** are grouped close together, and
- embeddings of **different individuals** are positioned far apart.

For recognition, a new face image is passed through the network to generate its embedding. This embedding is then compared with the embeddings stored in the database. If the measured similarity (or distance) meets a predefined threshold, the system classifies the face as a match.

C. Incremental Learning

The primary goal of incremental (or lifelong/continuous) learning is to enable models to learn from real, often noisy, data that arrives over time and may follow a non-stationary distribution. This is crucial for real-world applications where data patterns constantly change, allowing the model to adapt without the high computational cost of retraining on the entire dataset. Several incremental learning strategies exist, such as online learning, active learning, and transfer learning. An incremental method using OSD-eSVM is introduced, where identity labels are assigned based on current classifier scores. When an individual of interest (IoI) is detected, OSD-eSVM adds new classifiers to the association set. A Constraints module ensures the classifier set does not exceed a predefined size, while a Self-Healing module corrects potential misclassifications.

Classifiers are trained using a small number of positive samples and a limited pool of negatives, allowing efficient updates. The system can also remove classifiers when limits are reached or errors are detected. Although various incremental face recognition systems have been proposed using PCA, LDA, and deep learning, many still require retraining the entire dataset to add new faces—a process that is computationally expensive and impractical for real-time applications.

D. Contribution

We propose a novel incremental face recognition system that integrates YOLO for face detection and FaceNet for recognition. The system enables users to add new faces to the database dynamically, eliminating the need to retrain the entire model. When a face is detected, the system checks for a match; if the person is not recognized, the user can choose either to ignore the face or add it to the database. Upon adding, the system incrementally updates itself to learn the new identity. The main contribution of this work is an efficient and user-friendly incremental face recognition framework capable of adapting to changing data distributions without full retraining.

E. Limitation

CCTV systems still face several limitations and research challenges. Motion detection in dynamic scenes is particularly difficult due to variations in lighting, weather conditions, and shadows. There remains a strong need for faster and more accurate segmentation methods to enhance system performance.

3. Proposed Methodology

Our system offers several advantages over traditional face recognition methods. It is efficient, as it avoids retraining the entire dataset—an important benefit when dealing with large numbers of faces. It is also user-friendly, enabling users to add new faces through a mobile app, making database updates quick and convenient. Moreover, the system is adaptive, capable of recognizing newly added faces without full retraining, allowing it to respond effectively to changes in the data distribution. This section outlines the components and implementation of the proposed incremental face recognition system, which comprises three key modules: face detection, face recognition, and incremental learning.

A. Face Detection

The first stage of our system is face detection, performed using the YOLO object detection algorithm. YOLO analyses each image or video frame and generates bounding boxes along with class probabilities. In our case, the relevant class is the face. The detected face regions are then passed on to the recognition stage for further processing.

B. Face Recognition

The second stage of our system is face recognition, which is carried out using the FaceNet deep learning model. FaceNet processes each detected face and generates a compact feature representation, known as an embedding. These embeddings are then compared with those stored in the database to identify the person. In our system, the database consists of embeddings linked to their corresponding labels. To recognize a face, we compute the Euclidean distance between the input embedding and all stored embeddings. The label of the embedding with the smallest distance is taken as the recognition result.

C. Incremental Learning

The third stage of our system is incremental learning, which is activated whenever a new face is added to the database. Users can introduce new faces through a mobile app that communicates directly with the recognition system and sends the required face data. When a new face is captured, the system first applies face detection and recognition as previously described. If the face is not recognized, the user is given the choice to ignore it or add it to the database. If the user decides to add the face, the system updates the recognition model to learn this new identity. In our approach, incremental learning is carried out by fine-tuning the pre-trained FaceNet model using the newly added data. This allows the system to adapt to new faces without retraining the entire dataset, thereby reducing computational cost and improving efficiency.

D. Implementation Details

Our incremental face recognition system is implemented in Python using TensorFlow. YOLO, trained on the WIDER FACE dataset, is used for face detection, while the FaceNet model, pre-trained on VGGFace2, extracts face embeddings. A mobile app built with Flutter enables users to interact with the system easily.

YOLO first detects and crops face regions from the camera feed. These regions are then processed by FaceNet to generate compact, descriptive embeddings. The embeddings are classified using an SVM trained on labeled data of known individuals. If a face does not match any stored identity, the system raises an alert. Users can add the new face through the mobile app, triggering an incremental update. The new embeddings are added to the SVM model, allowing the system to recognize the individual in the future and reducing false alerts. By integrating YOLO, FaceNet, and SVM, the system provides accurate, real-time face recognition with the ability to adapt to new individuals over time.

E. Experimental Setup

The experimental results demonstrate that our incremental face recognition system is both **fast** and **highly accurate**. One of its key strengths is the ability to seamlessly recognize newly added faces without requiring a complete retraining of the model. This adaptability allows the system to adjust to changes in the data distribution over time, making it suitable for real-world, dynamic environments.

The system was evaluated using a diverse set of face images and video streams. During testing, it consistently delivered strong performance, with an **average processing time of just 0.1 seconds per image**, enabling real-time operation. The recognition model achieved an **overall accuracy of 98%**, validating the effectiveness of the feature extraction and matching techniques employed.

Additionally, the application interface enhances usability by allowing users to easily register new faces into the database. Once a face is added, the system can quickly incorporate it into future recognition tasks. The app also provides instant notifications whenever a recognized face appears in the camera feed, improving user convenience and ensuring timely alerts in practical scenarios such as surveillance or access control.

4. Experiments and Results

The experimental results indicate that our incremental face recognition system performs with both high speed and accuracy. It successfully recognizes newly added faces in the database and adapts effectively to changes in data distribution over time.

During testing on a collection of face images and video streams, the system demonstrated strong performance. The average processing time per image was **0.1 seconds**, and the recognition model achieved an average accuracy of **98%**. Additionally, the mobile app provides an easy and convenient interface for users to add new faces and receive instant notifications when a recognized face appears on camera.

5. Conclusions

In conclusion, this work presented a fast, efficient, and accurate incremental face recognition system built using deep learning–based face detection and recognition algorithms. A key advantage of the system is its ability to dynamically add new faces to the database without requiring complete retraining of the model. This incremental learning capability significantly reduces computational cost and makes the system adaptable to real-time updates. Because of its high accuracy, low processing time, and scalability, the proposed system is well-suited for real-world applications such as security monitoring, attendance systems, and personalized user authentication. Its integration with a mobile application further enhances usability by enabling users to easily add new faces and receive instant alerts, making the system practical and accessible for everyday deployment.

References

1. YOLO J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improve ment," arXiv preprint arXiv:1804.02767, 2018.
2. Facenet F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 815–823.
3. Rehmat Ullah , Hassan Hayat, Afsah Abid Siddiqui, Uzma Abid Siddiqui, Jebran Khan , Farman Ullah, Shoaib Hassan, Laiq Hasan, Waleed Albattah , Muhammad Islam, and Ghulam Mohammad Karami ,"A Real Time Framework for Human Face Detection and Recognition in CCTV Images " 3March 2022
4. P. Kamency, M. Benco, T. Mizdos, and R. Radil, "A new method for face recognition using convolutional neural network," Digital Image Processing And Computer Graphics, vol. 15, pp. 663–672, 2017
5. Guruh fajar shikik, Adhitya Nugraha,"A systematic Review of intelli gence video Surveillance: Trends, Techniques, Frameworks and Datasets "December 10, 2019
6. Lixiang Li , Xiaohui Mu, Siying Li and Haipeng Peng , "A Review of Face Recognition Technology" July 21, 2020.
7. Sammed Babaso Shirote, Vaidehi Vaibhav Mhaiskar ,"IOT BASED BUR GLAR NOTIFICATION SYSTEM USING IMAGE PROCESSING". PLoS One. 2016.
8. K. Dailey, "The rise of CCTV surveillance in the US," 29, April 2013.
9. M. Vidanapathirana, "Real-time Human Detection in Computer Vision — Part 1," 24 Mar 2018. [Online]. Available: https://medium.com/@madhawavidanapathirana/https- medium-com-madhawavidanapathirana-real-time human-detection-in- computer-vision-part-1 2acb851f4e55. [Accessed 07 Jun 2020].
10. D. Marikyan, S. Papagiannidis, and E. Alamanos, "A systematic review of the smart home literature: A user perspective," Technol. Forecast. Soc. Change, vol. 138, no. June 2018, pp. 139-154, 2019. https://doi.org/10.1016/j.techfore.2018.08.015

11. K. Maswadi, N. B. A. Ghani, and S. B. Hamid, "Systematic Literature Review of Smart Home Monitoring Technologies Based on IoT for the Elderly," IEEE Access, vol. 8, pp. 92244-92261, 2020. https://doi.org/10.1109/ACCESS.2020.2992727
12. R. Khalaf, A. Mohammed, E. Essa, and H. Ali, "Controlling Smart Home Activities Using IoT," ICCISTA 2019 - IEEE Int. Conf. Comput. Inf. Sci. Technol. their Appl. 2019, pp. 1-6, 2019. https://doi.org/10.1109/ICCISTA.2019.8830664
13. Taryudi Nico Surantha and Wingky R. Wicaksono. "An IoT based House Intruder Detection and Alert System using Histogram of Oriented Gradients". Journal of Computer Science 2019, 15 (8): 1108.1122.
14. N. S. Irjanto and N. Surantha, "Home Security System with Face Recognition based on Convolutional Neural Network," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 11, pp. 408-412, 2020.[5] B. Ingre and A. Yadav, "Performance analysis of NSL-KDD dataset using ANN," in Proc. Int. Conf. Signal Process. Commun. Eng. Syst., Andhra Pradesh, India, Jan. 2015.
15. Rahmatur Ramadhan, Ema Rachmawati, Muhammad Hari, "Analysis of SRGAN to upscale CCTV images", e-Proceeding of Engineering Volume 09 Issue 03, June 2022.