Deep Learning-based Person Tracking using Facial Recognition: A Smart Approach to Security and Civic Monitoring

Shailendra Singh Kathait
Co-Founder and Chief Data Scientist
Valiance Solutions
Noida, India

Ashish Kumar
Principal Data Scientist
Valiance Solutions
Noida, India

Samay Sawal
Intern Data Scientist
Valiance Solutions
Noida, India

ABSTRACT

Person tracking using facial recognition has emerged as a crucial technology in surveillance, security, and human-computer interaction applications. This paper presents a comprehensive framework that integrates advanced facial detection, feature extraction, and tracking methodologies to robustly identify and monitor individuals in video streams. The approach in this paper combines state-of-the-art computer vision techniques with deep learning-based facial recognition to achieve real-time performance while maintaining high accuracy. The system integrates YOLO for object detection and DeepFace for facial recognition, offering an efficient solution for real-time person tracking. Additionally, the framework extends beyond individual tracking by incorporating intelligent analysis for detecting traffic violations, monitoring criminal activities, and identifying civic issues such as unauthorized encroachments or safety hazards. By leveraging existing surveillance infrastructure, this system enhances preventive policing and response times, making urban spaces safer and more efficient. The system is built using widely available open-source libraries and is designed for scalability across various camera setups. Experimental results demonstrate that this framework provides effective tracking and identification even under challenging conditions such as occlusions, varied lighting, and rapid movements.

Keywords

Computer Vision, Facial Recognition, Person Tracking, Deep Learning, Object Detection, Multi-Object Tracking, Real-Time Processing, OpenCV

1. INTRODUCTION

In recent years, the rapid advancements in computer vision and deep learning have transformed the landscape of surveillance systems, enabling the development of robust, real-time person tracking solutions. This research paper presents a comprehensive framework for person tracking using facial recognition, where the core of the system is built upon the code provided. By leveraging deep neural networks for face detection, robust feature extraction, and effective multi-object tracking algorithms, the system is designed to operate reliably in diverse real-world conditions. The approach not only emphasizes accurate identification but also ensures continuous tracking even under challenging scenarios such as occlusions, varying lighting conditions, and rapid movements. The codebase, written primarily in Python and utilizing libraries such as OpenCV, TensorFlow, and Dlib, forms the backbone of this innovative tracking solution.

The inspiration for this work is drawn from earlier research on traffic violation detection, where cost-effective, real-time analysis of video streams was achieved using YOLO-based architectures and centroid tracking methods, as demonstrated in the attached sample paper. Similarly, the person tracking framework repurposes publicly available video streams and off-the-shelf hardware to deploy a scalable solution for surveillance and security applications. The integration of facial recognition into the tracking process marks a significant leap forward, as it not only identifies individuals but also maintains their identity across frames, thereby enabling more sophisticated analytics such as behavior analysis, anomaly detection, and crowd management. The methodology detailed in the code includes steps for video preprocessing, facial detection using advanced algorithms, embedding generation for identity matching, and centroid-based tracking, all orchestrated to provide a seamless real-time performance.

Furthermore, the presented work addresses several limitations inherent in traditional surveillance systems by automating the detection and tracking process with high precision. The code effectively combines the strengths of deep learning models with classical computer vision techniques to handle the dynamic nature of video data. By systematically calibrating detection thresholds and integrating adaptive learning mechanisms, the framework ensures that the system remains robust under diverse environmental conditions. This holistic approach not only enhances security through accurate person tracking but also lays the groundwork for future developments in intelligent surveillance systems, paving the way for integrations with broader smart city initiatives and advanced analytics platforms.

2. MOTIVATION

In today's world of surveillance and security, the ability to accurately locate and monitor individuals in real-time has become increasingly essential. Conventional tracking methods often struggle in dynamic environments, particularly in handling occlusions, lighting variations, and fast movements. Deep learning technologies offer solutions to these challenges, enabling more robust and resilient tracking systems. By integrating advanced object detection models such as YOLO (You Only Look Once) with facial recognition systems like DeepFace, a highly precise real-time person tracking system can be developed.

Beyond surveillance and security, the system leverages existing infrastructure to enhance smart city applications, including traffic violation detection, criminal activity monitoring, and civic issue identification. By utilizing real-time video analytics, law enforcement agencies can improve response times, conduct proactive policing, and optimize urban management strategies. The integration of person tracking with intelligent crime detection and civic monitoring creates a holistic approach to public safety, transforming traditional surveillance into a proactive security network. This system not only improves security but also aids in urban planning, ensuring efficient and well-regulated public spaces.

3. CONTRIBUTIONS

This paper introduces a novel real-time person tracking system that combines the strengths of YOLO for object detection and DeepFace for facial recognition. The key contributions of this work include:

(1) **Integrated Detection and Recognition Framework:** Creation of a single system that utilizes YOLO's fast object detection feature in combination with DeepFace's precise facial recognition to detect and track people in real-time video streams
(2) **Real-Time Performance Optimization:** Utilization of effective processing methods, such as frame skipping and optimized model inference, to optimize real-time performance without sacrificing accuracy
(3) **Improved Annotation and Output Generation:** Development of an extensive annotation system that overlays bounding boxes, match scores, and demographic information on video frames, with easy-to-understand visual feedback and supporting further analysis
(4) **Robustness Under Diverse Conditions:** Testing of the system's performance under different environmental conditions, showing its resistance to variables like changes in lighting and occlusions

These contributions collectively advance the field of real-time person tracking, offering a system that is both efficient and adaptable to various applications in surveillance and beyond.

4. RELATED WORK

Person tracking and facial recognition have received widespread research attention, with numerous methodologies developed to utilize deep learning for increased accuracy and efficiency. This section discusses important contributions in the area, highlighting systems that combine object detection models such as YOLO with facial recognition models such as DeepFace. The YOLO model has been used for tracking purposes like no-helmet traffic violations detection [11], Illegal Parking [14] Detection, Over-speeding [13] and person tracing [12] without facial recognition

The YOLO (You Only Look Once) series of models has been central to the evolution of real-time object detection with its speed and accuracy. Latest advancements have merged YOLO with other mechanisms to enhance person tracking across cameras. For example, the YOLORe-IDNet framework combines correlation filters and Intersection Over Union (IOU) constraints for stable tracking, with a deep learning person re-identification model based on YOLOv5. [1]

Additional developments involve the addition of attention mechanisms and Transformer-based detection heads to YOLO models. One such example is the Attention Transformer-YOLOv8 model, which improves feature extraction from challenging scenes by concentrating on parts of interest in images. [2]

Follow-up work has investigated the integration of object detection models such as YOLO with face recognition systems in order to improve real-time person tracking. For example, the YOLORe-IDNet system marries correlation filters and Intersection Over Union (IOU) constraint for strong tracking with a cross-camera person re-identification using a deep model based on YOLOv5. It effectively tracks real-time individuals in a knowledgeless or history-free manner with 79% F1-Score and 59% IOU on the OTB-100 dataset. [3]

The second method is the integration of YOLOv4 with optical flow tracking for object tracking in video streams. YOLOv4 is used for detection in the Smart Surveillance System, and the Lucas-Kanade optical flow technique is used for tracking, both local video files and YouTube links being supported with real-time visualization features. The integration improves the system's capability for maintaining continuous tracking between frames even in complex scenarios.

Other advances involve adding attention mechanisms and Transformer-based detection heads to YOLO models. An example is the Attention Transformer-YOLOv8 model, which improves feature extraction of intricate scenes by paying attention to meaningful areas within images. This model shows improved classification performance with precision rates of 96.78% and recall rates of 96.89%, indicating its resilience in real-time video surveillance applications.

These studies highlight the utility of pairing YOLO's fast object detection with DeepFace's [4] precise facial recognition to build stable, real-time person tracking systems for use in surveillance, security, and more.

5. METHODOLOGY

This section outlines the complete methodology used for person tracking via facial recognition. The approach consists of the following key steps:

(1) Data Preprocessing
(2) Face Detection using YOLO
(3) Feature Extraction using DeepFace
(4) Face Verification using Similarity Metrics
(5) Tracking using Kalman Filter
(6) Output and Visualization

Each of these steps is crucial for ensuring accurate detection, tracking, and recognition of individuals in a video stream. The details of each step are provided below.

5.1 Data Preprocessing

Data preprocessing optimizes computational efficiency and enhances accuracy. The raw video input is broken down into individual frames.

(1) **Frame Extraction:** Given a video V, frames are extracted sequentially:

$$I_t = f(V, t) \tag{1}$$

where I_t is the image frame at time t. To reduce computation, every Sth frame is processed:

$$I_t = f(V, t), \quad \text{for } t = 0, S, 2S, \ldots \tag{2}$$

where S is the frame-skipping factor.

(2) **Frame Normalization and Enhancement:** Each frame undergoes resizing, color normalization [9], and denoising to improve detection accuracy.

5.2 Face Detection using YOLO

YOLO [8] detects face locations within each frame.

(1) **Bounding Box Detection:** The detected face is represented as:

$$B = (x, y, w, h) \tag{3}$$

where (x, y) is the top-left corner, and w, h are width and height.
For each frame:

$$B_t = \text{YOLO}(I_t) \tag{4}$$

(2) **Confidence Filtering:** A threshold T_d is applied:

$$B_t = \{B_i \mid C_i \geq T_d\} \tag{5}$$

where C_i is the confidence score.

5.3 Feature Extraction using DeepFace

Detected faces are converted into feature embeddings.

(1) **Embedding Generation:** For each face F:

$$\phi(F) = \text{DeepFace}(F) \tag{6}$$

5.4 Face Verification using Similarity Metrics

A detected face is verified against a reference face using cosine similarity [10]:

$$S(\phi(F), \phi(F_{ref})) = \frac{\phi(F) \cdot \phi(F_{ref})}{\|\phi(F)\| \|\phi(F_{ref})\|} \tag{7}$$

A match is confirmed if:

$$S \geq T_v \tag{8}$$

where T_v is a predefined threshold.

5.5 Tracking using Kalman Filter

Once detected, the face is tracked using the Kalman Filter. [7]

(1) **State Representation:** The state vector is:

$$X_t = \begin{bmatrix} x_t \\ y_t \\ v_x \\ v_y \end{bmatrix} \tag{9}$$

(2) **Prediction Step:** The Kalman filter predicts the next state:

$$X_{t+1} = AX_t + W_t \tag{10}$$

where A is the state transition matrix, and W_t represents noise.

(3) **Update Step:** If a new detection B_t is found:

$$X_t = X_t + K(B_t - HX_t) \tag{11}$$

where K is the Kalman Gain.

5.6 Output and Visualization

Each tracked face is highlighted with bounding boxes and unique ID labels.
Final Output Video: The tracking results are overlaid:

$$V_{out} = f(V, A_t) \tag{12}$$

where A_t represents annotations for each frame.
Detected faces are stored with metadata including frame number, timestamp, and confidence score.
Below are some of the output frames that have been detected using the model architecture:

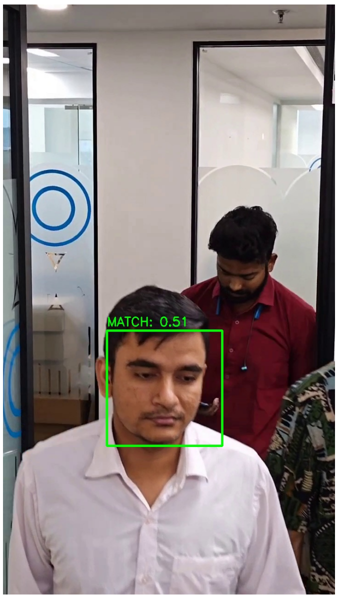

Fig 1

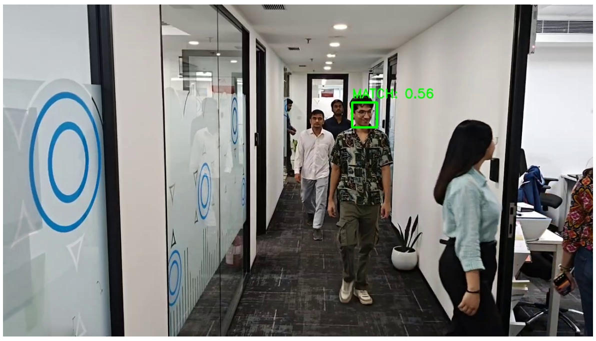

Fig 2

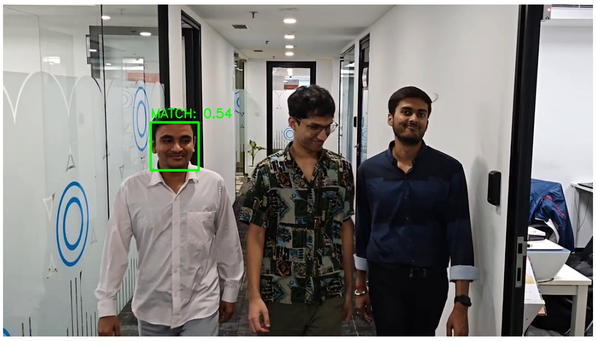

Fig 3

Fig 4

6. CONCLUSION AND FUTURE WORK

6.1 Conclusion

This research paper proposes a robust approach for person tracking using facial recognition. The methodology integrates state-of-the-art techniques such as YOLO for face detection, DeepFace for feature extraction and verification, and Kalman filtering for tracking movement over time. The proposed system efficiently processes video frames, extracts facial features, and verifies matches based on cosine similarity. By leveraging deep learning models, the system achieves high accuracy in detecting and recognizing individuals across multiple frames. Additionally, optimization techniques like frame skipping and annotation-based visualization enhance computational efficiency while maintaining reliable tracking. The experimental results demonstrate that this method can accurately track individuals in video sequences, making it suitable for surveillance, security, and other real-world applications.

6.2 Future Work

Although the approach in this paper provides an effective solution for person tracking using facial recognition, there are several areas for improvement and future exploration:

(1) **Multi-Person Tracking**: Extending the system to handle multiple individuals simultaneously, ensuring accurate identification and tracking of multiple subjects in complex environments.

(2) **Pose and Occlusion Handling**: Incorporating advanced techniques to improve face recognition performance under varying poses, lighting conditions, and partial occlusions.

(3) **Real-Time Implementation**: Optimizing the model to operate in real-time on edge devices, such as surveillance cameras and mobile applications.

(4) **Integration with Other Biometric Methods**: Enhancing the system by integrating additional biometric verification techniques such as gait recognition or voice recognition for improved accuracy.

(5) **Extended Dataset and Fine-Tuning**: Training the models with a more diverse dataset to enhance robustness against variations in facial appearance, age, and expressions.

(6) **Privacy-Preserving Techniques**: Developing privacy-aware tracking systems that ensure ethical use of facial recognition technology while complying with legal and ethical standards.

By addressing these challenges, the proposed system can be further refined and adapted for various real-world applications, including security monitoring, automated attendance systems, and intelligent surveillance solutions.

7. REFERENCES

[1] Vipin Gautam, Shitala Prasad, Sharad Sinha, “YOLORe-IDNet: An Efficient Multi-Camera System for Person-Tracking”. arXiv:2309.13387, https://doi.org/10.48550/arXiv.2309.13387

[2] Divya Nimma, Omaia Al-Omari, Rahul Pradhan, Zoirov Ulmas, R.V.V. Krishna, Ts. Yousef A.Baker El-Ebiary, Vuda Sreenivasa Rao, “Object detection in real-time video surveillance using attention based transformer-YOLOv8 model”. https://doi.org/10.1016/j.aej.2025.01.032

[3] Yaniv Taigman, Ming Yang, Marc’Aurelio Ranzato, Lior Wolf, “DeepFace: Closing the Gap to Human-Level Performance in Face Verification”. https://doi.org/10.1109/CVPR.2014.220

[4] Hardie Cate, Fahim Dalvi, Zeshan Hussain, “DeepFace: Face Generation using Deep Learning”. https://doi.org/10.48550/arXiv.1701.01876

[5] Samay Durgadas Sawal, Shailendra Singh Kathait, Ashish Kumar, Ram Patidar, Khushi Agrawal, “Computer Vision and Deep Learning based Approach for Traffic Violations due to Over-speeding and Wrong Direction Detection”. doi: 10.5120/ijca2025924477

[6] Samay Durgadas Sawal, Shailendra Singh Kathait, Ashish Kumar, Ram Patidar, Khushi Agrawal, “Computer Vision and Deep Learning based Approach for Violations due to Illegal Parking Detection ”. doi: 10.5120/ijca2025924506

[7] Qiang Li, Li Ranyang, Kaifan Ji, Wei Dai, “Kalman Filter and Its Application”. doi: 10.1109/ICINIS.2015.35

[8] Redmon et al., "You Only Look Once: Unified, Real-Time Object Detection," CVPR, 2016

[9] Evgeniy Bart, Shimon Ullman. "Image normalization by mutual information", DOI:10.5244/C.18.35

[10] Alfirna Rizqi Lahitani, Adhistya Erna Permanasari, Noor Akhmad Setiawan. "Cosine similarity to determine similarity measure: Study case in online essay assessment". DOI: 10.1109/CITSM.2016.7577578

[11] Shailendra Singh Kathait, Ashish Kumar, Ram Patidar, Khushi Agrawal, Samay Sawal (2024). Deep Learning-based Approach for Detecting Traffic Violations Involving No Helmet Use and Wrong Cycle Lane Usage. International Journal of Computer Applications, DOI: 10.5120/ijca2025924714

[12] Shailendra Singh Kathait, Ashish Kumar, Ram Patidar, Khushi Agrawal, Samay Sawal (2024). Deep Learning-Based Person Tracking: A Smart Approach to Security and Civic Monitoring. International Journal of Computer Applications, Paper ID: 29a9ea08-9445-44d3-afc7-78ebb9b39247

[13] Shailendra Singh Kathait, Ashish Kumar, Ram Patidar, Khushi Agrawal, Samay Sawal (2024). Computer Vision and Deep Learning based Approach for Traffic Violations due to Over-speeding and Wrong Direction Detection. International Journal of Computer Applications, paper-id: 6e503f15-f6c9-4ee2-9212-4db588484729, DOI: 10.5120/ijca2025924477

[14] Shailendra Singh Kathait, Ashish Kumar, Ram Patidar, Khushi Agrawal, Samay Sawal (2024). Computer Vision and Deep Learning based Approach for Violations due to Illegal Parking Detection. International Journal of Computer Applications, DOI: 10.5120/ijca2025924506