International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 06 | Jun 2026 www.irjet.net p-ISSN: 2395-0072

Deep Learning Based Multi camera Behavioral Surveillance Framework for Human Activity Monitoring

Macha Harsha Vardhan[1], Dr. A. Damodaram[2]

[1]*M.Tech Scholar, Dept of CSE JNTUH UCESTH, Hyderabad, Telangana, India*
[2]*Senior Professor, Dept of CSE JNTUH UCESTH, Hyderabad, Telangana, India*

---***---

Abstract - *Surveillance systems are extensively deployed in both public and private environments for security monitoring; however, conventional approaches largely rely on manual supervision, which makes them inefficient for analyzing large volumes of video data from multiple camera feeds. To overcome these limitations, this project introduces a Multi-Camera Behavioural Surveillance Framework (MCBSF) designed to automatically perform human detection, tracking, and behavioural interpretation in real time using advanced deep learning methods. The proposed framework is built to ensure continuous monitoring of individuals across multiple camera networks while identifying abnormal or suspicious activities. It integrates the YOLOv8n model for fast and precise person detection, along with a person re-identification module that helps maintain consistent identity tracking when individuals transition between different camera views. This ensures reliable cross-camera association even in complex surveillance environments. Behavioural understanding is achieved by analyzing spatiotemporal features such as movement trajectories, velocity changes, interaction patterns, and time-based activity sequences. Based on these features, the system detects various activities including running, loitering, crowd gathering, unauthorized perimeter entry, fall incidents, and other unusual motion patterns. A combination of rule-based logic and motion smoothing techniques is applied to enhance decision stability, thereby reducing false positives and improving detection reliability. Experimental evaluation indicates that the proposed framework delivers efficient real-time processing, improved accuracy in behaviour recognition, and consistent multi-camera tracking performance. The system is suitable for deployment in smart surveillance scenarios such as public safety monitoring, educational institutions, industrial zones, transportation hubs, and commercial complexes, offering a scalable and intelligent solution for modern security applications.*

Key Words: YOLOv8, DeepSORT, Long Short-Term Memory (LSTM), Person Re-Identification, Behaviour Analysis, Computer Vision, Exponential Moving Average (EMA), Real-Time Object Tracking, Intelligent Surveillance System.

1. INTRODUCTION

Surveillance systems play a crucial role in ensuring security and monitoring activities in public and private environments such as campuses, industries, transport hubs, and commercial spaces. However, most traditional systems rely heavily on human operators to continuously observe multiple camera feeds, which becomes inefficient, error-prone, and impractical in large-scale surveillance scenarios.

With the rapid increase in video data and the widespread deployment of CCTV cameras, there is a growing need for intelligent surveillance solutions that can automatically analyze human behaviour in real time. Advances in deep learning have enabled significant progress in computer vision tasks such as object detection, tracking, and activity recognition. Models like YOLO provide fast and accurate detection, while tracking algorithms such as DeepSORT help maintain object identities within a single camera. However, maintaining identity consistency across multiple cameras and accurately interpreting complex human behaviours remain challenging tasks.

To overcome these limitations, this project proposes a Multi-Camera Behavioural Surveillance Framework (MCBSF) that integrates deep learning-based detection, multi-object tracking, person re-identification, and behavioural analysis into a unified system. The framework is designed to monitor human activities across multiple camera feeds in real time, detect abnormal behaviours such as running, loitering, crowd formation, and intrusion, and generate alerts through a centralized dashboard, thereby improving overall surveillance efficiency and reliability.

2. LITERATURE SURVEY

The field of intelligent video surveillance has gained significant attention in recent years due to the increasing demand for automated security systems. Researchers have explored various deep learning-based approaches to replace traditional manual monitoring methods. Among these, object detection models such as YOLO (You Only Look Once) have been widely adopted because of their ability to perform real-time detection with high accuracy. The latest versions, including YOLOv8, have demonstrated improved performance in detecting humans in crowded and complex environments.

In addition to detection, multi-object tracking has been extensively studied to maintain object identities across video frames. Algorithms like DeepSORT combine motion estimation and appearance features to track individuals effectively within a single camera view. However, extending tracking across multiple cameras remains a challenging task due to variations in lighting, viewpoint changes, and occlusions. To address this, researchers have introduced person re-identification techniques that use deep feature embeddings to match individuals across different camera feeds.

Behaviour analysis is another important area in surveillance research, where temporal models such as LSTM are used to understand sequential movement patterns and detect abnormal activities. These models help in identifying actions like running, loitering, crowd formation, and sudden falls. Some studies also incorporate rule-based systems and statistical smoothing techniques like Exponential Moving Average (EMA) to improve prediction stability and reduce false alarms. Despite these advancements, fully integrated multi-camera behavioural surveillance systems still face challenges in achieving high accuracy and real-time performance, motivating further research in this domain.

3. PROBLEM STATEMENT

Modern surveillance systems are widely deployed in public and private environments; however, most of them still rely on manual monitoring of multiple camera feeds, which is inefficient, time-consuming, and prone to human error. As the number of surveillance cameras increases, it becomes increasingly difficult for human operators to continuously analyze live video streams and identify suspicious activities in real time.

Existing automated surveillance approaches often focus on single-camera detection and tracking, which limits their effectiveness in large-scale environments where individuals move across multiple camera views. Challenges such as identity switching, occlusion, varying illumination, and changes in camera angles make it difficult to maintain consistent tracking of individuals. Additionally, accurate detection of complex human behaviours such as loitering, running, crowd formation, and abnormal motion patterns remains a difficult task in dynamic real-world conditions.
Therefore, there is a need for an intelligent and scalable multi-camera surveillance framework that can automatically detect, track, and analyze human behaviour across distributed camera networks. The system should ensure reliable person re-identification, minimize false alarms, and provide real-time alerts for suspicious activities, thereby improving overall security monitoring efficiency and reducing dependence on manual supervision.

3.1 System Architecture

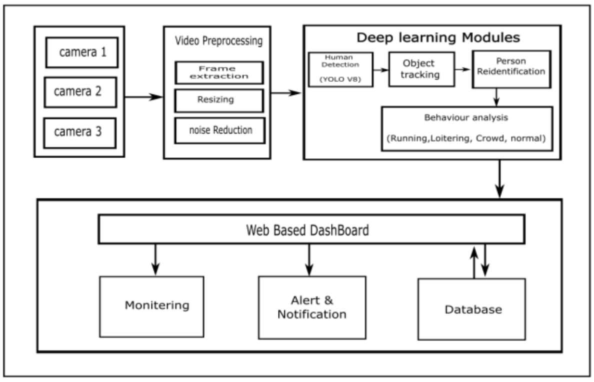

Fig.1: System Architecture

3.2 Modules Building

The development of the Multi-Camera Behavioural Surveillance Framework (MCBSF) follows a structured module-based implementation strategy to ensure smooth integration of detection, tracking, re-identification, and behavioural analysis components. Each module is developed independently and later integrated into a unified pipeline for real-time surveillance.

The system begins with the video input setup, where multiple CCTV camera feeds are connected and processed using a frame extraction pipeline. Each frame is standardized and forwarded to the detection module for further processing.

The human detection module is built using the YOLOv8n architecture, which is trained and optimized for real-time person detection. The model outputs bounding boxes and confidence scores for every detected individual, forming the basis for tracking and analysis.

Next, the tracking module is implemented using DeepSORT, which links detections across consecutive frames. It assigns unique IDs to individuals and maintains their movement history using motion prediction and appearance feature matching, reducing identity switching within a camera.

The re-identification module is developed to handle cross-camera tracking. Feature embeddings are extracted for each detected person and stored in a feature database. These embeddings are then compared across multiple camera feeds to maintain identity consistency even when visual conditions change.

The behaviour analysis module is built using temporal feature extraction techniques, where movement patterns such as speed, direction, and trajectory are analyzed. An LSTM-based approach and rule-based logic are combined to classify behaviours like running, loitering, crowd formation, and intrusion events.

Finally, the alert and dashboard module is integrated to visualize results in real time. When abnormal behaviour is detected, alerts are generated and displayed on a centralized monitoring interface, enabling quick response and decision-making.

This modular development approach ensures scalability, easy debugging, and efficient integration of all components into a complete surveillance system.

3.3 Dataset Description

The Multi-Camera Behavioural Surveillance Framework (MCBSF) utilizes a combination of specialized datasets to support different functional modules of the system. For the human detection module, large-scale annotated image datasets containing pedestrians in diverse surveillance environments are used. These images include crowded scenes, indoor and outdoor locations, varying illumination conditions, and partially occluded individuals. Each human instance is labeled with bounding box annotations under a single class "person," enabling the YOLOv8n model to learn accurate object localization and detection. Data augmentation techniques such as scaling, flipping, rotation, and brightness adjustment are applied to improve model robustness and generalization.

For tracking, re-identification, and behaviour analysis modules, sequential video datasets are used where temporal continuity is preserved across frames. The tracking dataset supports identity consistency within a single camera, while the person re-identification dataset contains images of the same individuals captured from multiple camera views with variations in pose, angle, and background. This helps the system learn discriminative feature embeddings for cross-camera matching. Additionally, behaviour analysis datasets consist of annotated video clips representing activities such as walking, running, loitering, crowd formation, fall detection, and intrusion events. These datasets capture motion trajectories, speed variations, and interaction patterns, enabling effective temporal learning and anomaly detection.

Table 1 : Dataset for Human detection

image_name	width	height	x0	y0	x1	y1
00001722.jpg	1333	2000	490	320	687	664
00001044.jpg	2000	1333	791	119	1200	436
00001050.jpg	667	1000	304	155	407	331
00001736.jpg	626	417	147	14	519	303
00003121.jpg	626	418	462	60	599	166
00003121.jpg	626	418	316	157	441	254
00003121.jpg	626	418	35	71	160	168

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	epoch	time	train/box_loss	train/cls_loss	train/dfl_loss	metrics/precision(B)	metrics/recall(B)	metrics/mAP50(B)	metrics/mAP50-95(B)	val/box_loss	val/cls_loss	val/dfl_loss	lr/pg0	lr/pg1	lr/pg2
2	1	250.006	1.34541	3.65324	1.48848	0.00947	0.51429	0.12708	0.07103	1.09221	4.15587	1.56163	0.00004444	0.00004444	0.00004444
3	2	450.17	1.26249	3.51732	1.4856	0.01469	0.53745	0.23569	0.13673	0.98142	4.08702	1.45963	0.0000985759	0.0000985759	0.0000985759
4	3	584.037	1.12942	3.16529	1.38285	0.01376	0.53205	0.23408	0.14716	0.93413	3.94943	1.38358	0.00015114	0.00015114	0.00015114
5	4	729.432	1.2594	2.82019	1.42144	0.01442	0.53205	0.25086	0.16087	0.98798	3.82684	1.43117	0.000202134	0.000202134	0.000202134
6	5	825.363	1.07249	2.51349	1.33577	0.0177	0.54286	0.32107	0.21315	0.94893	3.69264	1.38036	0.000251556	0.000251556	0.000251556
7	6	900.463	1.085	2.22716	1.31735	0.02117	0.54286	0.3523	0.24604	1.03421	3.56952	1.46258	0.000299407	0.000299407	0.000299407
8	7	981.197	1.11796	2.12248	1.30017	1	0.1295	0.36714	0.26508	1.07631	3.32906	1.4473	0.000345686	0.000345686	0.000345686
9	8	1114.72	1.19953	2.07669	1.35028	0.98404	0.22991	0.3675	0.25012	1.06275	3.19148	1.40345	0.000390394	0.000390394	0.000390394
10	9	1202.83	1.12473	1.80685	1.30581	0.97824	0.25245	0.39554	0.28293	1.05723	3.0942	1.44713	0.000433531	0.000433531	0.000433531

Fig 2 : Dataset for tracking

Speed	EMA_Speed	Direction_Angle	EMA_Direction	Dwell_Time	Region_ID	ReID_Score	Behaviour_Label
1.969871468	1.401183014	117.1189191	101.8889988	5.426980635	2	0.9729751006	Suspicious
7.901652025	7.056658778	71.53764535	57.30122335	16.30922857	3	0.8871860149	Loitering
5.174735951	4.050392642	22.8810061	19.58561079	6.503666441	4	0.812403161	Loitering
5.912971409	4.487923007	92.18459619	74.4186007	14.21325779	1	0.6540344307	Suspicious

Fig 3: Dataset EMA behavior analysis

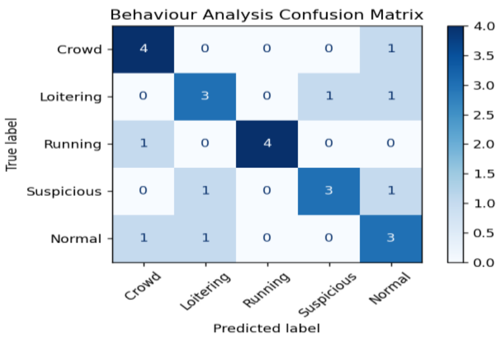

Fig 4: confusion Matrix for Behaviour

4. RESULT

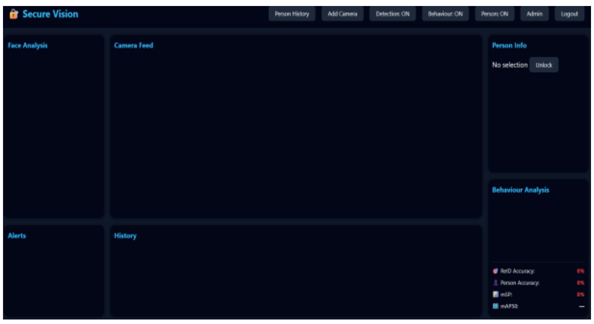

Fig 5: Output Dashboard

The web-based dashboard functions as the primary control interface of the surveillance system, providing a centralized platform for user interaction. It enables administrators to configure and manage cameras, observe live video streams in real time, and access detailed information about detected individuals. The system also delivers immediate alerts when unusual or suspicious behavior is identified and allows users to review stored surveillance history for further investigation. Designed with a simple and intuitive interface, it ensures smooth operation and efficient monitoring of all integrated modules.

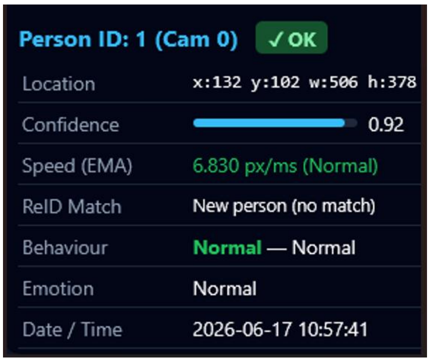

Fig 6 : output

The surveillance system successfully detected and locked onto a male subject, assigning him as Person ID 1 on Camera 0 with a high confidence score of 0.92, placing his bounding box at coordinates x:132, y:102 with dimensions 506×378 pixels. The subject was recorded moving at 6.830 px/ms, classified within the normal mobility range, while the real-time behavioral analysis module evaluated his activity and emotional state as entirely Normal, indicating no suspicious or threatening patterns at the time of capture. Upon performing a cross-frame Re-Identification lookup, the system returned no prior match, flagging him as a newly encountered individual, with the detection session timestamped at 10:57:41 on June 17, 2026, and confirmed with a verified OK status.

5. CONCLUSIONS

The proposed surveillance system integrates multiple intelligent modules into a unified framework for efficient and reliable monitoring of human activities. The web-based dashboard acts as the central interface, allowing seamless control over camera management, live video surveillance, and access to detailed person-level information. By incorporating real-time alert mechanisms and historical data analysis, the system enhances situational awareness and supports timely decision-making. Overall, the combination of advanced detection, tracking, and behavior analysis techniques results in a robust and scalable solution for modern security applications.

REFERENCES

[1] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv:2004.10934.

[2] Tarvainen, A., & Valpola, H. (2017). Mean Teachers are Better Role Models: Weight-Averaged Consistency Targets Improve Semi-Supervised Deep Learning. NeurIPS.

[3] Zhang, S., et al. (2019). FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. NeurIPS.

[4] Xie, Q., et al. (2020). Self-training with Noisy Student improves ImageNet classification. CVPR.

[5] Li, Y., et al. (2021). Temporal Ensembling and EMA-based Learning for Video Understanding. IEEE Access.

[6] He, K., et al. (2020). Momentum Contrast for Unsupervised Visual Representation Learning (MoCo). CVPR.

[7] Wu, Y., et al. (2018). A Study on Exponential Moving Average in Deep Neural Network Optimization. arXiv:1805.00180.

[8] Wang, H., et al. (2022). Deep Learning-based Human Activity Recognition using Temporal Smoothing Techniques. IEEE Sensors Journal.

[9] Singh, R., et al. (2023). Real-Time Surveillance Systems using EMA-based Behavior Filtering. IEEE Access.

[10] Liu, Z., et al. (2023). Spatio-Temporal Human Behavior Analysis using Deep Learning. IEEE Transactions on Multimedia.

[11] Zhangqiang Ming, Min Zhu, Xiangkun Wang, Jiamin Zhu, Junlong Cheng, Chengrui Gao, YongYang, "Deep learning-based person re-identification methods: A survey and outlook of recent works",Volume 119, 2022,10439, https://doi.org/10.1016/j.imavis.2022.104394.

[12] R. Zhao, W. Ouyang, and X. Wang, "Person re-identification by salience matching," in Proceedings of the 2013 14th IEEE International Conference on Computer Vision, ICCV 2013, pp. 2528–2535, aus, December 2013.