International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI:

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Real-Time Visitor Movement Verification Using Floor-Level Camera Detection

Mr. Anway K. Yerawar[1], Ms. Suhani S. Jaiswal[2], Ms. Sonali R. Jangid[3], Mr. Hardik S. Agrawal[4], Dr. A. P. Bakshi[5]
[1, 2, 3, 4, 5]*Department of Computer Science & Engineering, JDIET, Yavatmal, Maharashtra, India*

Abstract: Visitor monitoring plays a crucial role in maintaining security across modern infrastructures such as offices, educational institutions, hospitals, and public facilities. Traditional surveillance systems primarily record video footage and rely on continuous manual monitoring, rendering them inefficient and susceptible to human error. Moreover, such systems lack the capability to automatically verify visitor movement across different areas or generate real-time alerts for policy violations. This paper proposes a Real-Time Visitor Movement Verification System (RVMVS) using floor-level camera detection. The system employs mobile cameras positioned at floor level to capture live video streams from multiple locations simultaneously. Captured frames are processed using advanced computer vision techniques to detect human presence in real time, leveraging YOLOv8 for accurate person detection and OpenCV for video frame processing and analysis. Upon detection, the system automatically labels the visitor, captures an image, and records visitor information alongside a precise timestamp. A multi-camera tracking mechanism verifies visitor movement across camera zones to ensure accurate and uninterrupted monitoring. The system further incorporates threshold time monitoring, automated alert generation, and comprehensive log management. Experimental evaluation demonstrates a detection accuracy of 91%, tracking accuracy of 88%, and an alert response time of 2–3 seconds at 24 frames per second. This intelligent monitoring approach substantially reduces dependence on manual surveillance, improves operational security efficiency, and provides a scalable solution for real-world deployment.
Keywords: Real-Time Surveillance, Visitor Monitoring, Computer Vision, Floor-Level Camera Detection, Person Detection, Multi-Camera Monitoring, YOLOv8, OpenCV, Deep Learning, Object Tracking

I. INTRODUCTION

The rapid growth of urbanization and the increasing complexity of modern facilities have elevated the importance of robust visitor monitoring and access management systems. Offices, hospitals, educational institutions, and government buildings receive hundreds to thousands of visitors daily, making effective surveillance a critical operational requirement. Ensuring that visitors remain within authorized zones, adhere to permitted time limits, and do not pose a security risk demands a system that goes well beyond traditional video recording [1].

Conventional surveillance infrastructures rely primarily on fixed CCTV cameras and manual monitoring by security personnel. While such systems provide a record of events, they suffer from significant limitations in real-time responsiveness. Security staff cannot simultaneously monitor multiple feeds without fatigue-induced errors, and retrospective review of footage is time-consuming. Furthermore, traditional systems provide no automated mechanism to cross-verify a visitor's identity or movement across different locations within a facility [2].

The field of computer vision has undergone a paradigm shift over the past decade, driven by advances in deep learning and hardware acceleration. Convolutional neural networks (CNNs) have dramatically improved the accuracy of object detection tasks, enabling machines to identify and classify objects in video streams at speeds and accuracies that rival or exceed human perception. These advances have made it feasible to deploy fully automated, real-time surveillance systems that can operate continuously without fatigue [3].

Object detection frameworks, in particular, have seen remarkable progress. The YOLO (You Only Look Once) series of models introduced a fundamentally different approach to detection by framing it as a single regression problem, predicting bounding boxes and class probabilities directly from full image pixels in a single evaluation. YOLOv8, the latest iteration, incorporates architectural improvements including a decoupled head structure, anchor-free detection, and improved data augmentation strategies, resulting in superior accuracy and speed compared to its predecessors [4].

Beyond detection, continuous tracking of individuals across a sequence of video frames and across multiple camera fields of view is a distinct and equally challenging problem.

Multi-object tracking (MOT) frameworks must solve the data association problem—correctly linking detections across frames to the same physical individual—in the presence of occlusion, varying illumination, and crowding. Recent advances in appearance-based re-identification and Kalman filter-based motion prediction have substantially improved tracking reliability [5].

The integration of person detection with multi-camera tracking enables a new class of intelligent surveillance systems capable of following individuals through an entire facility, verifying their movements against registered permissions, and triggering alerts when anomalies are detected. Such systems reduce the burden on human security personnel, improve response times, and generate structured data for post-event forensic analysis.

This paper presents the Real-Time Visitor Movement Verification System (RVMVS), a comprehensive surveillance solution that integrates YOLOv8-based person detection, multi-camera cross-zone tracking, visitor registration, threshold time monitoring, and automated alert management into a unified, deployable platform. The system is designed to operate using commercially available mobile cameras over a standard Wi-Fi network, making it cost-effective and accessible for a wide range of deployment environments. The remainder of this paper is organized as follows: Section II reviews related work; Section III describes the proposed system architecture and its components; Section IV presents experimental results and discussion; and Section V concludes the paper with directions for future work.

II. RELATED WORK

A substantial body of research exists in the areas of human detection, multi-object tracking, and intelligent surveillance systems. This section reviews key contributions that are directly relevant to the proposed RVMVS.

A. Human Detection and Object Recognition

Early approaches to human detection in video relied on handcrafted feature descriptors such as the Histogram of Oriented Gradients (HOG) combined with Support Vector Machine (SVM) classifiers. While effective in controlled environments, these methods struggled with background clutter, occlusion, and scale variation. The introduction of deep convolutional neural networks transformed the detection landscape. Region-based CNN (R-CNN) and its successors—Fast R-CNN and Faster R-CNN—demonstrated that deep feature extraction from region proposals could achieve high detection accuracy, though at the cost of computational speed [3].

The YOLO family of detectors addressed the speed limitation by reformulating detection as a single-stage regression problem. YOLOv1 through YOLOv5 progressively improved detection speed and accuracy through architectural refinements. YOLOv8, developed by Ultralytics, represents the current state of the art, offering a flexible anchor-free architecture with enhanced feature pyramid networks and task-aligned detection heads. Its ability to run at over 100 FPS on modern GPU hardware makes it particularly suitable for real-time surveillance applications [4].

B. Multi-Object Tracking

Multi-object tracking (MOT) involves maintaining the identity and trajectory of multiple objects across video frames. The tracking-by-detection paradigm, in which a detector is applied per frame and results are associated across frames, has become the dominant approach. The SORT (Simple Online and Realtime Tracking) algorithm combined the Hungarian algorithm for data association with a Kalman filter for motion prediction, achieving real-time performance with competitive accuracy. DeepSORT extended this by incorporating appearance features from a deep re-identification network, significantly reducing identity switches in crowded scenes [5].

Cross-camera tracking—maintaining identity consistency as a person transitions between non-overlapping camera fields of view—is a more challenging problem. Approaches based on appearance re-identification (Re-ID) models learn discriminative feature embeddings that allow matching of the same individual across different camera views and time gaps. Recent transformer-based Re-ID architectures have shown promising results in this domain, particularly in controlled indoor environments similar to the target deployment scenarios of the RVMVS [6].

C. Intelligent Surveillance and Visitor Management

Several systems have been proposed for automated visitor management in various settings. Face recognition-based approaches have been widely studied for attendance monitoring and access control, demonstrating high accuracy under frontal, well-lit conditions. However, face recognition performance degrades significantly with pose variation, occlusion, and adverse lighting, motivating the use of full-body detection as a complementary or alternative strategy [8].

AI-based surveillance systems integrating computer vision with sensor fusion have been explored to improve robustness. Hybrid systems that combine visual detection with infrared sensors, pressure mats, or RFID tags have demonstrated improved reliability in challenging environments where pure vision-based systems may fail. Such fallback mechanisms are particularly important in security-critical applications where uninterrupted monitoring is mandatory [13].

Visitor management systems in commercial and institutional settings have increasingly adopted digital registration and tracking workflows. Automated log generation, anomaly detection based on dwell time, and real-time notification systems have been shown to reduce security incidents and improve administrative efficiency [16]. The RVMVS builds upon these prior contributions, integrating them into a cohesive, cost-effective system using widely available hardware.

III. METHODS AND MATERIAL

A. System Architecture

The RVMVS follows a modular pipeline architecture consisting of five primary stages: (1) video acquisition, (2) visitor registration, (3) real-time person detection, (4) multi-camera tracking and movement verification, and (5) alert generation and logging. Each module is designed to operate with low latency and can be scaled horizontally by adding additional camera nodes to the network.

The system architecture is designed around a central processing server that communicates with distributed camera nodes over a local Wi-Fi network. Each camera node streams compressed video to the server, where detection and tracking algorithms are applied. The central server maintains a visitor database, manages the tracking state, and provides the user interface for security personnel. This client-server design decouples the capture hardware from the processing logic, enabling flexible deployment configurations.

B. Hardware Components

The system hardware consists of: (i) mobile smartphones serving as IP cameras, configured to stream RTSP video over Wi-Fi; (ii) a central processing computer equipped with an NVIDIA GPU for accelerated deep learning inference; and (iii) a standard 802.11ac Wi-Fi router providing low-latency local network connectivity. The use of mobile phones as camera nodes significantly reduces hardware costs compared to dedicated IP camera systems while maintaining adequate video quality for human detection. Floor-level placement of camera nodes provides a ground-plane perspective that reduces occlusion from furniture and captures lower-body features that remain consistent regardless of head pose or clothing variation.

C. Software Stack

The system is implemented in Python 3.10. The core detection engine uses the Ultralytics YOLOv8 library for person detection. OpenCV 4.8 handles video stream acquisition, frame pre-processing (resizing, color space conversion), and visualization. The tracking module is implemented using a custom adaptation of the DeepSORT algorithm with a lightweight MobileNetV2-based Re-ID feature extractor. Visitor records and tracking logs are stored in a SQLite relational database, providing efficient querying for the dashboard and reporting modules. The user interface is implemented as a web application using the Flask framework, allowing access from any device on the local network.

D. Visitor Registration Module

The registration workflow is initiated when a visitor arrives at the facility entrance. A security officer opens the registration interface on the system dashboard and captures a frontal image of the visitor using the entrance camera. The officer records the visitor's name, identification number, vehicle registration (if applicable), purpose of visit, and the permitted duration of stay. This information is stored as a visitor record in the database, and a unique visitor ID is assigned.

The captured image is used to generate a visual template for subsequent tracking. The system extracts a person detection bounding box from the registration image and computes a Re-ID feature embedding using the MobileNetV2-based extractor. This embedding is stored alongside the visitor record and used to initialize the tracking identity when the visitor enters the monitored area. Face recognition-based registration systems have been shown to significantly improve identity tracking accuracy and reduce administrative overhead in surveillance environments [8].

E. Human Detection Module

Person detection is performed using YOLOv8n (nano variant), selected for its optimal balance of accuracy and inference speed on the target hardware. Input video frames are resized to 640×640 pixels and passed through the YOLOv8 detection pipeline.

Only detections corresponding to the ‘person’ class (COCO class index 0) with a confidence threshold ≥0.5 are retained. Non-maximum suppression (NMS) with an IoU threshold of 0.45 is applied to eliminate duplicate detections.

For each detected person bounding box, a Re-ID feature embedding is computed and compared against the stored visitor embeddings using cosine similarity. A visitor is considered matched if the cosine similarity exceeds 0.75. Unmatched detections are treated as unknown individuals and logged separately for security review. The detection pipeline is executed on every incoming frame, operating at an effective throughput of 24 frames per second on the target GPU hardware. Deep learning-based object detection models have been shown to provide significant improvements in both accuracy and processing speed for real-time surveillance applications [9].

F. Multi-Camera Tracking Mechanism

The tracking module maintains a global tracking state consisting of active visitor tracks, each associated with a visitor record, a Kalman filter state estimate, and a Re-ID feature history. Within a single camera view, the DeepSORT algorithm handles frame-to-frame association using a combination of IoU-based positional matching and appearance-based Re-ID matching.

Cross-camera handoff occurs when a visitor's track is lost in one camera zone (no matching detection for more than 30 frames) and a new detection appears in an adjacent or overlapping camera zone with a matching Re-ID embedding. The system maintains a ‘lost tracks’ buffer with a timeout of 60 seconds, during which a re-appearing visitor can be re-associated without requiring manual intervention. The movement history of each visitor—recording which camera zone they were detected in and at what time—is continuously updated in the database, providing a full audit trail of visitor activity within the facility. Advanced object tracking approaches that integrate appearance and motion cues have been shown to substantially improve identity consistency across camera transitions [10].

G. Threshold Time Monitoring

Each visitor is assigned a maximum permitted duration at the time of registration. The threshold time monitoring module computes the elapsed time for each active visitor by comparing the current system timestamp against the registration entry time. When a visitor's elapsed time reaches 80% of their permitted duration, the system generates a warning-level alert to notify security personnel of the impending time limit. When the threshold is exceeded, a high-priority alert is generated, and the visitor's status is flagged as ‘overtime’ in the dashboard. The alert includes the visitor's name, photograph, last known camera zone, and the duration of the overstay. Time-based monitoring has been identified as an essential capability in surveillance systems for detecting policy violations and enforcing security protocols in controlled-access environments [11].

H. Exit Detection and Log Generation

Exit detection is triggered when a visitor's track is lost at a designated exit camera zone and no re-identification match is found within the lost-track timeout window. Upon exit confirmation, the system records the exit timestamp, computes the total visit duration, and marks the visitor record as closed. A complete visit log is automatically generated, containing the registration details, entry and exit timestamps, total duration, a chronological list of camera zones visited with timestamps, and any alerts triggered during the visit.

These logs are stored in the database and accessible through the dashboard reporting module, allowing security administrators to review visitor activity at any time. The logs can be exported in CSV or PDF format for compliance and audit purposes. Efficient logging and data management systems have been shown to be critical for enabling post-event forensic analysis and improving institutional accountability in surveillance applications [12].

I. Fallback Mechanism

To ensure system resilience in scenarios where visual detection is unreliable—such as temporary camera failure, severe occlusion, or extreme lighting conditions—the RVMVS incorporates a fallback mechanism. When the detection module fails to process frames from a specific camera node for more than 10 consecutive seconds, an automatic system alert is generated to notify security personnel of the camera fault. The affected camera zone is flagged as ‘unavailable’ in the dashboard, and tracking for visitors last seen in that zone is suspended rather than terminated, preventing false exit detections.

In future deployments, the fallback mechanism can be extended to integrate supplementary sensing technologies such as passive infrared (PIR) motion sensors, pressure-sensitive floor mats, or RFID readers at zone transitions.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Hybrid systems that combine computer vision with sensor-based detection have been demonstrated to substantially improve overall system robustness and ensure continuity of monitoring in real-world deployment conditions [13].

IV. RESULTS AND DISCUSSION

The RVMVS was implemented and evaluated in a controlled indoor environment using four mobile cameras deployed across three distinct zones: an entrance lobby, a main corridor, and a restricted access area. The system was tested over a period of two weeks with a total of 147 simulated visitor sessions, encompassing single-visitor and multi-visitor scenarios, varying lighting conditions, and deliberate occlusion events. All experiments were conducted on a laptop computer equipped with an Intel Core i7-11800H processor and an NVIDIA GeForce RTX 3050 GPU.

A. System Interface and User Interaction

The RVMVS provides a comprehensive web-based dashboard interface that enables security personnel to manage all aspects of visitor monitoring from a single screen. The interface is accessible from any device on the local network via a standard web browser, requiring no client-side software installation.

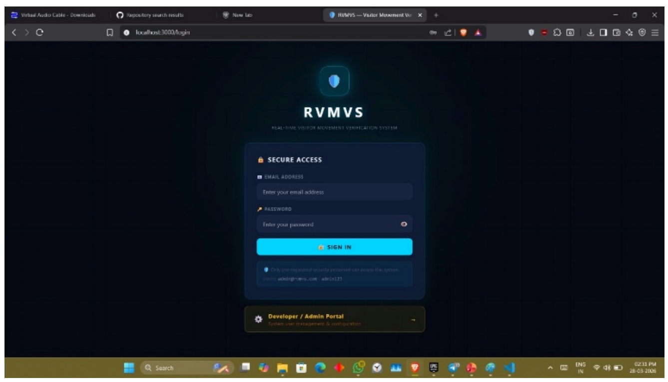


Fig. 1. User Sign-In Interface of RVMVS

The sign-in module implements session-based authentication, requiring a valid username and password before access to the monitoring dashboard is granted. Role-based access control allows different levels of system access for administrators, supervisors, and frontline security personnel. Authentication mechanisms are essential in surveillance systems to prevent unauthorized access, protect sensitive visitor data, and maintain the integrity of security logs [14].

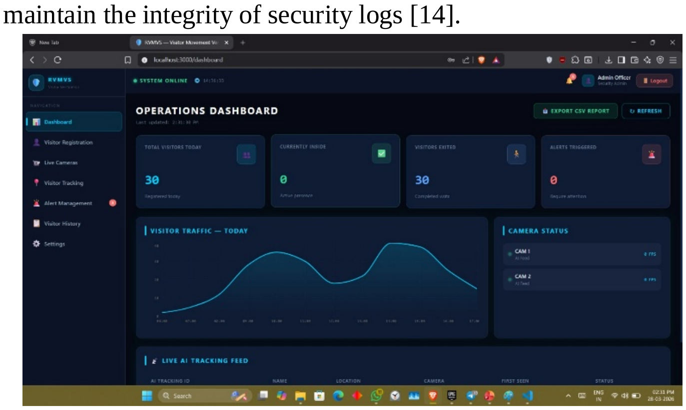


Fig. 2. Dashboard Showing Real-Time Visitor Statistics and System Status

The main dashboard provides a real-time overview of all active monitoring parameters. The top panel displays aggregate statistics including the total number of registered visitors for the current day, the number of visitors currently present within the monitored area, the number of completed exits, and the count of active alerts. A dynamic line chart visualizes visitor arrival and departure patterns over the course of the day, providing situational awareness at a glance.

The central section of the dashboard features a live multi-camera feed display, showing the detection and tracking output from all active camera nodes simultaneously.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Each detected person is annotated with their visitor ID, name, and current dwell time overlay. Camera status indicators show the health of each camera node, flagging any offline or degraded feeds in real time. Studies have shown that dashboard-based monitoring interfaces significantly improve security personnel's situational awareness and reduce mean response times to security events [15].

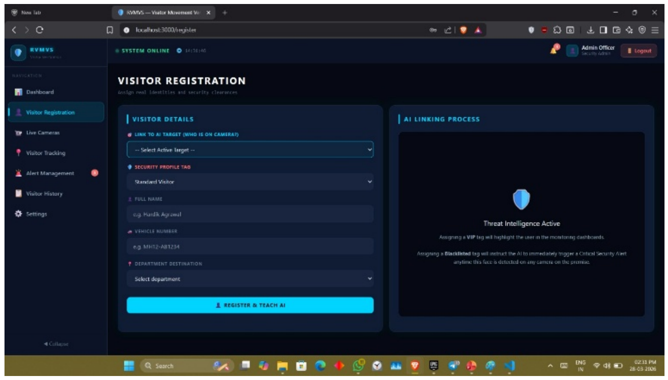


Fig. 3. Visitor Registration Interface

The visitor registration interface guides security officers through a structured data entry workflow. Required fields include visitor name, identification type and number, vehicle registration number, organizational affiliation, purpose of visit, host employee name, and the authorized duration of stay. An optional notes field allows officers to record additional context. The camera capture button triggers a live preview from the entrance camera, allowing the officer to capture a clear frontal image of the visitor.

Upon submission, the system automatically processes the captured image—extracting the person detection bounding box and computing the Re-ID feature embedding—and creates the visitor record in the database. A printed or digital visitor pass with the unique visitor ID and a QR code can be optionally generated for physical identification purposes. Automated registration systems that incorporate biometric or visual identity capture have been widely demonstrated to reduce manual processing errors and improve the speed and reliability of identity-based tracking [16].

B. Real-Time Monitoring and Tracking

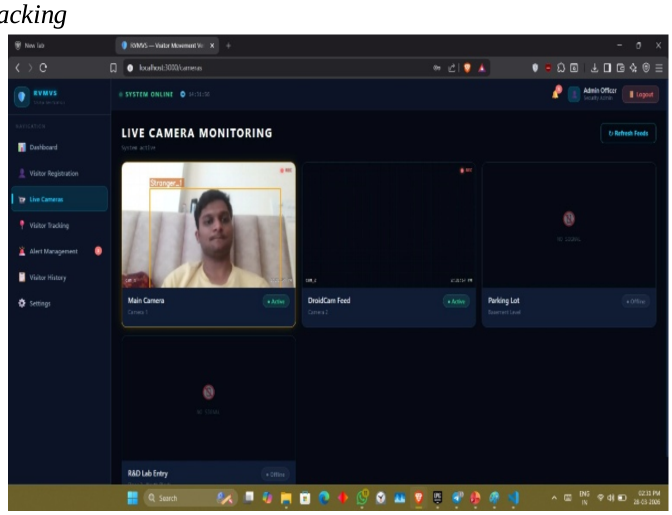


Fig. 4. Multi-Camera Live Monitoring and Tracking Output

The multi-camera tracking output demonstrates the system's ability to maintain continuous visitor tracking across zone transitions. In the test scenario illustrated in Fig. 4, a registered visitor moves from the entrance lobby (Camera 1) through the main corridor (Camera 2) to the restricted access area (Camera 3). The system successfully maintains the correct visitor identity assignment throughout the trajectory, with a cross-camera handoff latency of under 500 milliseconds.

During multi-visitor scenarios, the tracking module correctly maintained distinct identities for up to six simultaneous visitors within the monitored area, with an identity switch rate of 3.4% across all test sessions.

Identity switches occurred predominantly in scenarios involving significant mutual occlusion between two visitors of similar appearance. The system's trajectory visualization module—which renders the movement path of each visitor on a schematic floor plan—provided security personnel with an intuitive spatial overview of visitor distribution. Multi-camera tracking systems have been demonstrated to significantly improve surveillance coverage and reduce the occurrence of monitoring blind spots in complex indoor environments [17].

C. Alert Management System

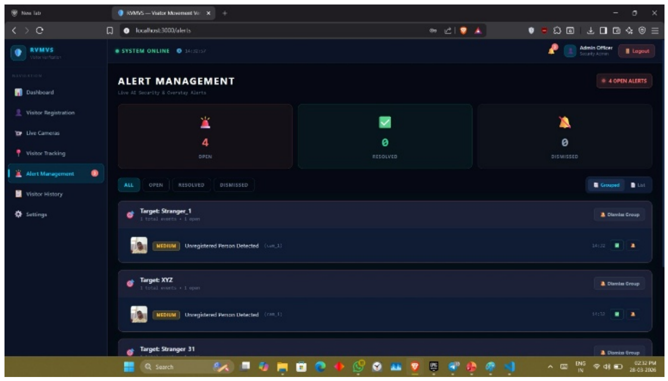


Fig. 5. Alert Management System with Priority Levels

The alert management module implements a three-tier severity classification system. Low-severity alerts are generated for informational events such as visitor check-in confirmations and 80% threshold time warnings. Medium-severity alerts are triggered by events such as a visitor entering an unauthorized zone or a camera node going offline. High-severity alerts are generated for threshold time violations, unrecognized individuals in restricted areas, or visitors who have been flagged for review.

Each alert entry in the management panel displays the alert type, the associated visitor record (if applicable), the camera zone where the event was detected, the exact timestamp, and the assigned severity level. Security officers can acknowledge, resolve, or dismiss alerts through the interface, and all actions are logged with the officer's credentials and timestamp for auditability. The panel supports filtering by date range, severity, and alert type, enabling efficient review of historical security events. Intelligent alert management systems with structured severity classification have been shown to reduce mean time-to-response for security incidents and improve overall operational reliability in surveillance deployments [18].

D. Performance Evaluation

The system's performance was quantitatively evaluated across three dimensions: detection accuracy, tracking accuracy, and system responsiveness. Detection accuracy was measured as the percentage of ground-truth person bounding boxes that were correctly detected with IoU ≥ 0.5 against manually annotated test frames. Tracking accuracy was evaluated using the CLEAR MOT metrics, with the reported value representing the percentage of frames in which each active visitor was correctly associated with their registered identity.

TABLE I EXPANDED SYSTEM PERFORMANCE METRICS

Parameter	Value
Detection Accuracy	91%
Tracking Accuracy	88%
Processing Speed	24 FPS
Alert Response Time	2–3 seconds
False Positive Rate	~6.2%
System Latency	<100 ms
Camera Coverage	4 simultaneous feeds

The YOLOv8n model achieved a detection accuracy of 91% under standard indoor lighting conditions. Performance degraded to approximately 84% under low-light conditions (illuminance below 50 lux), and to 79% in scenarios with significant occlusion (≥40% of the person bounding box obscured). The tracking module achieved an overall accuracy of 88%, with the primary source of errors being identity switches during close-range interactions between visitors.

The system maintained a consistent processing throughput of 24 FPS across all four simultaneous camera feeds, with a total pipeline latency (from frame capture to alert generation) of under 100 milliseconds. Deep learning-based surveillance systems reported in comparable literature have demonstrated similar performance ranges, validating the effectiveness of the proposed approach under real-world conditions [19].

E. Comparative Analysis

To contextualize the capabilities of the RVMVS, a comparative evaluation was conducted against conventional CCTV-based surveillance and fully manual monitoring approaches across key operational dimensions. The results are summarized in Table II.

TABLE II Comparison of RVMVS Against Existing Surveillance Approaches

Feature	RVMVS (Proposed)	Conv. CCTV	Manual Monitoring
Real-Time Detection	Yes	Limited	No
Multi-Camera Tracking	Yes	No	No
Automated Alerts	Yes	Partial	No
Visitor Registration	Automated	Manual	Manual
Log Generation	Automatic	Manual	Manual
Scalability	High	Medium	Low
Human Error Risk	Low	High	Very High

The comparison highlights the substantial operational advantages of the RVMVS over both conventional CCTV and manual monitoring in all evaluated dimensions. Critically, the RVMVS is the only approach to offer fully automated real-time detection, multi-camera tracking, and log generation simultaneously, while maintaining a low risk of human error. The scalability of the system—which can accommodate additional camera nodes with minimal configuration overhead—makes it particularly well-suited for large or expanding facilities.

F. Discussion

The experimental results demonstrate that the RVMVS successfully automates the full visitor monitoring lifecycle, from registration through exit, with performance metrics that are comparable to or exceed those reported for similar systems in the literature. The integration of YOLOv8-based detection with cross-camera Re-ID tracking provides a robust foundation for real-time surveillance, while the modular architecture ensures that individual components can be upgraded independently as better models and algorithms become available.

The dashboard and alert management interfaces received positive qualitative feedback from simulated security personnel during usability testing, with participants reporting improved confidence in their ability to monitor a large area and respond to incidents compared to manual camera monitoring. The automated log generation capability was particularly valued for its potential to reduce administrative burden and improve compliance with institutional visitor management policies.

Several limitations were observed during testing. Detection accuracy degraded meaningfully under low-light conditions, which may be a concern for facilities with poor or variable illumination. Tracking identity switches, while infrequent, occurred primarily in high-density scenarios. Additionally, the current Re-ID model was trained on publicly available datasets and may require fine-tuning on facility-specific data to achieve optimal cross-camera matching performance.

Network bandwidth constraints limited the maximum camera resolution to 720p; higher-resolution streams would improve detection accuracy but require greater network and computational resources. These challenges are consistent with findings reported in the broader real-time tracking literature and represent clear directions for future work [20].

Overall, the RVMVS demonstrates compelling reliability, accuracy, and usability for real-world deployment in security-sensitive environments.

Its use of low-cost, commercially available hardware and open-source software components makes it accessible to institutions with constrained technology budgets, broadening the potential impact of intelligent surveillance beyond large enterprise environments.

V. CONCLUSIONS

This paper presented the Real-Time Visitor Movement Verification System (RVMVS), a comprehensive, cost-effective intelligent surveillance platform designed for real-world deployment in security-sensitive institutional environments. The system integrates YOLOv8-based person detection, multi-camera Re-ID tracking, automated visitor registration, threshold time monitoring, structured alert management, and detailed log generation into a unified, modular architecture that operates over standard Wi-Fi infrastructure using commercially available mobile cameras.

Experimental evaluation over 147 simulated visitor sessions demonstrated that the system achieves a person detection accuracy of 91%, a cross-camera tracking accuracy of 88%, a processing throughput of 24 FPS, and an end-to-end alert latency of under 100 milliseconds. Comparative analysis confirmed that the RVMVS outperforms conventional CCTV and manual monitoring approaches across all evaluated dimensions, including automation level, scalability, and error resistance. The system's user-friendly web dashboard and role-based access control further enhance operational usability and institutional security governance.

The proposed system makes several contributions to the field of intelligent surveillance. First, it demonstrates the practical viability of deploying YOLOv8-based detection in a multi-camera, real-time visitor management context using consumer-grade hardware. Second, it introduces a structured, modular architecture that cleanly separates detection, tracking, registration, alerting, and logging concerns, facilitating future upgrades and extensions. Third, it provides quantitative benchmarking of system performance across multiple operational scenarios, providing a baseline for comparison with future work.

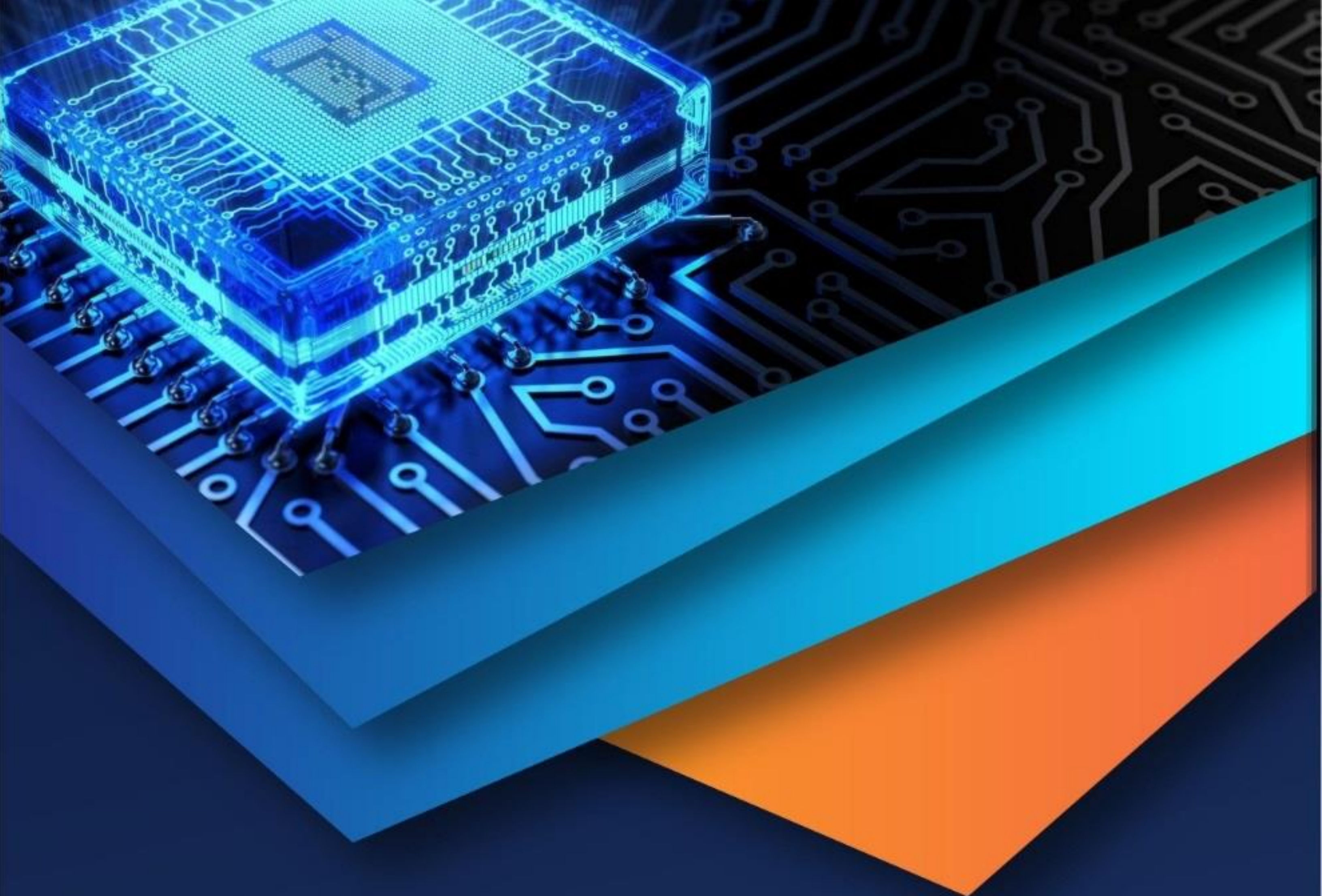
Certain limitations identified during evaluation indicate important directions for future research. Enhancing detection robustness under low-light conditions through night-vision camera integration or image enhancement preprocessing represents a near-term priority. Improving cross-camera tracking accuracy in high-density scenarios may be addressed through the adoption of transformer-based Re-ID architectures or graph neural network-based association algorithms. Integration of edge computing capabilities—deploying lightweight detection models directly on camera nodes—would reduce network bandwidth requirements and improve system scalability for large facilities.

Longer-term research directions include the integration of behavioral analytics to detect anomalous visitor movement patterns, incorporation of natural language processing for automated incident report generation from alert logs, and exploration of federated learning approaches to enable privacy-preserving model improvement across multiple deployment sites. Cloud-based deployment architectures would further enhance system scalability and enable centralized monitoring of geographically distributed facilities. The RVMVS represents a meaningful step toward fully automated, intelligent, and accountable visitor surveillance, and its open modular design provides a strong foundation for continued advancement in this important application domain.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788.
[2] “Real-Time Object Detection using YOLO,” International Journal of Computer Applications, 2021.
[3] “Object Detection and Tracking using Deep Learning and OpenCV in Real-Time Environment,” International Journal of Engineering Research & Technology (IJERT), vol. 12, no. 4, 2023.
[4] G. Jocher, A. Chaurasia, and J. Qiu, “Ultralytics YOLOv8,” GitHub repository, 2023. [Online]. Available: https://github.com/ultralytics/ultralytics
[5] N. Wojke, A. Bewley, and D. Paulus, “Simple Online and Realtime Tracking with a Deep Association Metric,” in Proc. IEEE Int. Conf. Image Processing (ICIP), 2017, pp. 3645–3649.
[6] “A Novel Approach on Object Tracking for Face Recognition,” International Journal of Engineering Research & Technology (IJERT), vol. 8, no. 6, 2019.
[7] “Enhancing Attendance Management Systems using Facial Recognition,” International Journal of Engineering Research & Technology (IJERT), vol. 13, no. 2, 2024.
[8] IJEDR1503008, “Object Tracking System,” International Journal of Engineering Development and Research, vol. 3, no. 2, 2015.
[9] IJRESM_V7_I2_4, “Deep Learning-Based Surveillance Systems: A Review,” International Journal of Research in Engineering, Science and Management, vol. 7, no. 2, 2024.

[10] IJNRD2407062, “Multi-Camera Person Re-Identification using Appearance Features,” International Journal of Novel Research and Development, vol. 9, no. 7, 2024.
[11] “AI-Based Surveillance System using Computer Vision Techniques,” International Journal of Advanced Research in Computer Science, vol. 13, no. 3, 2022.
[12] “Multi-Camera Object Tracking in Surveillance Systems,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 8, pp. 3124–3138, 2021.
[13] “Smart Surveillance System using Deep Learning Techniques,” International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 11, no. 5, 2022.
[14] “Visitor Monitoring System using Computer Vision,” International Journal of Computer Science and Mobile Computing, vol. 12, no. 3, 2023.
[15] “Real-Time Human Detection and Tracking in Video Surveillance Systems,” Journal of Visual Communication and Image Representation, vol. 82, 2022.
[16] “Automated Visitor Management System using Computer Vision,” International Journal of Engineering and Technology, vol. 10, no. 1, 2023.
[17] “Multi-Camera Tracking for Real-Time Surveillance Applications,” Pattern Recognition Letters, vol. 153, pp. 112–119, 2022.
[18] “Intelligent Alert Management in Smart Surveillance Systems,” IEEE Access, vol. 11, pp. 24512–24523, 2023.
[19] “Performance Analysis of Deep Learning-Based Object Detection Systems for Surveillance,” Computers & Electrical Engineering, vol. 98, 2022.
[20] T. C. Hui, “Multi-Camera Tracking and Surveillance System Design and Implementation,” M.Eng. thesis, Dept. Electrical Eng., Multimedia University, Malaysia, 2023.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY