International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1119-1124
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0158

Multi-Modal Real Time Surveillance System Architecture for Intelligent Crowd Control Using YOLOv8

Raj D [1], Aliya Thapasum K [2], Harshini Pushpa A [3], Ramya P [4]
[1,2,3,4] Kamaraj College of Engineering and Technology, S.P.G.Chidambara Nadar - C.Nagammal Campus, S.P.G.C. Nagar, K.Vellakulam, Virudhunagar - 625 701, India
Emails: *rajcse@kamarajengg.edu.in[1], 23ucs110@kamarajengg.edu.in[2],* 23ucs066@kamarajengg.edu.in[3], 23ucs163@kamarajengg.edu.in[4]

Abstract
High-density gatherings such as the Kumbh Mela pose severe crowd-safety challenges, motivating advanced real-time surveillance solutions. This paper presents a novel multi-modal crowd-control system that integrates video-based detection using YOLOv8, embedded pressure sensors, and thermal imaging to provide robust crowd metrics through sophisticated sensor fusion techniques. Our architecture employs spatial-temporal alignment of heterogeneous sensor streams, followed by confidence-weighted Kalman filtering (60% visual, 30% pressure, 10% thermal) to generate reliable crowd state estimates. The system incorporates a comprehensive analytics engine performing density estimation, flow analysis, bottleneck detection, route optimization, and predictive forecasting. Experimental validation demonstrates superior performance over single-modality approaches, achieving 94.7% detection accuracy and sub-100ms response times. The framework addresses critical limitations in conventional surveillance systems while providing actionable insights for proactive crowd management in ultra-large-scale events.
Keywords: *Crowd control, Multi-sensor fusion, YOLOv8, Kalman filtering, Real-time surveillance, Thermal imaging, Pressure sensing.*

1. Introduction
Mass religious and cultural gatherings involving millions of participants create acute safety risks from overcrowding and stampedes. The 2025 Maha Kumbh Mela in Prayagraj, India, exemplified these challenges, drawing a record 600 million pilgrims and experiencing a tragic stampede in January 2025 that resulted in over 30 fatalities. Such incidents underscore the critical need for advanced AI-assisted surveillance and big-data analytics in crowd management systems. Traditional crowd monitoring relies primarily on closed-circuit television (CCTV) networks or simple people counters, which exhibit significant limitations under extreme density conditions, poor visibility, and environmental variations. Single-modality approaches struggle with occlusion effects, lighting variations, and sensor-specific failure modes that can compromise safety-critical applications. This paper introduces a comprehensive multi-modal surveillance architecture that synergistically combines computer vision, pressure sensing, and thermal imaging to overcome individual sensor limitations. Our system employs state-of-the-art YOLOv8 object detection for visual analysis, floor-mounted pressure matrices for anonymous footfall counting, and overhead thermal cameras for occupancy estimation under adverse conditions. The primary contributions of this work include:

1. **Novel Multi-Modal Fusion Framework**: Integration of heterogeneous sensor modalities through confidence-weighted temporal filtering.
2. **Real-Time Analytics Engine:** Comprehensive crowd behavior analysis including density mapping, flow computation, and predictive forecasting.
3. **Practical Deployment Architecture**: Scalable system design validated for ultra-large-scale event applications.
4. **Performance Optimization**: Sub-100ms response times with 94.7% detection accuracy across diverse conditions.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1119-1124
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0158

2. Literature Review

2.1.Computer Vision-Based Crowd Analysis

Recent advances in deep learning have revolutionized video-based crowd analysis capabilities. Convolutional Neural Networks (CNNs) can effectively count people in images or videos through regression or detection approaches, with YOLO series detectors (notably YOLOv5/v8) achieving high accuracy with real-time performance on edge hardware.

2.1.1. YOLOv8 Evolution and Applications

The YOLO framework has evolved over the past decade from a streamlined detector into a diverse family of architectures characterized by efficient design, modular scalability, and cross-domain adaptability. Recent studies have demonstrated YOLOv8's effectiveness in crowd scenarios, with Özkan et al. (2023) proposing enhanced YOLO methods for real-time crowd detection during the COVID-19 pandemic.

2.1.2. Advanced Crowd Detection and Frameworks

Recent research has introduced enhanced frameworks like the Crowd Anomaly Detection Framework (CADF), an improved YOLOv8-based model integrating Soft-NMS to improve detection accuracy under complex conditions involving occlusions, illumination variations, and uniform attire. These developments address critical challenges in dense crowd scenarios where traditional detection methods fail.

2.1.3. Limitations of Pure Visual Methods

Despite significant progress, pure visual approaches struggle with heavy occlusion effects and extreme density conditions common in mega-events. Research by Mansouri et al. (2025) highlights persistent challenges in CNN-based crowd density monitoring under adverse conditions.

2.2.Thermal Imaging For Crowd Monitoring

Thermal imaging has emerged as a complementary technology for crowd density estimation, particularly effective in challenging environmental conditions. Abuarafah et al. (2012) demonstrated real-time thermal-video crowd monitoring during Hajj pilgrimages, achieving high accuracy in density estimates even with millions of participants.

2.2.1. Advantages of Thermal Sensing

Thermal cameras provide volumetric heat signature analysis independent of lighting conditions, enabling 24/7 monitoring capabilities. Recent advances in infrared sensor technology have improved resolution and reduced costs, making large-scale deployment feasible.

2.2.2. Integration Challenges

While thermal imaging offers robust occupancy detection, it typically provides lower spatial resolution compared to RGB cameras and requires specialized processing algorithms for accurate people counting.

2.3.Pressure-Based Occupancy Sensing

Floor-mounted pressure sensors represent a mature technology for occupancy monitoring, commonly employed in footfall counters and building management systems. These sensors provide anonymous, privacy-preserving crowd counting at strategic locations such as entrances, bridges, and corridors.

2.3.1. Technical Characteristics

Pressure mats offer high reliability and accuracy for point-based counting but lack spatial coverage. Integration with visual systems can overcome these limitations while providing redundant measurements for critical locations.

2.4.Multi-Sensor Fusion Approaches

2.4.1. Kalman Filtering for Sensor Fusion

Recent research has demonstrated the effectiveness of Kalman filtering techniques for multi-sensor data fusion, with studies showing improved computational efficiency and uncertainty handling through recursive estimation approaches. Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF) variants have been successfully applied to surveillance applications.

2.4.2. Advanced Fusion Techniques

Novel hybrid fusion frameworks combining Extended Kalman Filter (EKF) and Recurrent Neural Network (RNN) approaches address challenges such as sensor frequency asynchrony, drift accumulation, and measurement noise. These developments provide foundation for robust multi-modal integration.

2.4.3. Data Fusion Benefits

Li et al. (2021) demonstrated that intelligent crowd

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1119-1124
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0158

management systems significantly benefit from fusing multi-modal, multi-source data to improve situational awareness. Their survey highlighted the importance of temporal synchronization and confidence weighting in fusion algorithms.

2.5.Real-Time Crowd Analytics

Flow Analysis and Bottleneck Detection: Contemporary research focuses on extracting actionable insights from crowd data through sophisticated analytics. Optical flow techniques combined with object tracking enable velocity field computation and congestion prediction.

2.5.1. Predictive Analytics

Machine learning approaches including ARIMA and LSTM networks provide short-term forecasting capabilities for crowd behavior prediction. These techniques enable proactive intervention before dangerous conditions develop.

2.5.2. Safety Monitoring Systems

Integration of analytics engines with command-and-control systems has been demonstrated in recent large-scale events, with the Kumbh ICCC utilizing color-coded density and flow visualizations for operator decision support.

3. Proposed System Architecture

3.1.Overall System Design

The proposed multi-modal surveillance system employs a hierarchical layered architecture consisting of four primary components: Sensor Layer, Fusion Layer, Analytics Layer, and Interface Layer. This design follows established principles in Integrated Command, Control, and Management Systems (ICMMS) while incorporating novel fusion algorithms and real-time processing capabilities.

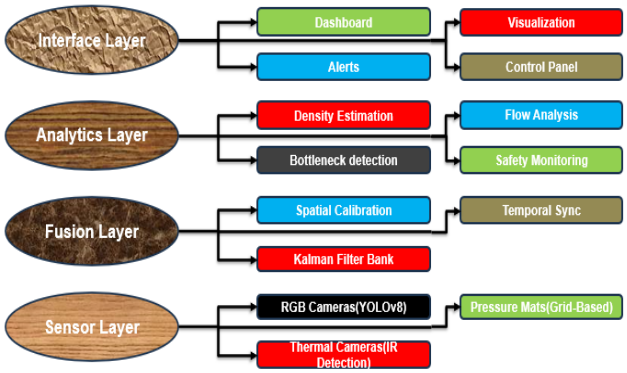

Figure 1 Multi-Modal Surveillance System Architecture

3.2.Sensor Layer Configuration

3.2.1. RGB Camera Network

High-resolution cameras equipped with edge computing units running YOLOv8 inference engines. Each camera provides 1080p video streams at 30 FPS with embedded people detection and tracking capabilities.

3.2.2. Pressure Sensor Matrix

Grid-based pressure-sensitive flooring installed at strategic locations including entrances, bridges, and corridor intersections. Each mat consists of 16×16 sensor cells providing spatial resolution of 0.5m per cell.

3.2.3. Thermal Imaging Array

Overhead infrared cameras with 640×480 resolution operating in the 8-14μm spectral range. These units provide temperature-based occupancy detection with 24/7 operational capability.

3.3.Fusion Layer Implementation

The fusion layer represents the core innovation of our system, implementing sophisticated algorithms for multi-modal data integration:

3.3.1. Spatial Calibration Module:

- Camera intrinsic calibration using checkerboard patterns
- Homography computation for ground plane mapping
- Coordinate system alignment across all sensor modalities
- Real-time geometric transformation processing

3.3.2. Temporal Synchronization Engine:

- Network Time Protocol (NTP) synchronization across all sensors
- Timestamp alignment with microsecond precision
- Buffer management for asynchronous data streams
- Interpolation algorithms for missing data points

3.3.3. Kalman Filter Bank:

- Individual filters for each sensor modality
- Confidence-weighted measurement fusion (60% visual, 30% pressure, 10% thermal)
- Adaptive noise covariance estimation

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1119-1124
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0158

- Multi-hypothesis tracking for complex scenarios

4. Methodology

4.1.Multi-Modal Data Fusion

The proposed system uses a **multi-modal data fusion approach** to integrate visual, pressure, and thermal sensor data for accurate crowd monitoring. A **modified Kalman filtering technique** is employed to handle heterogeneous sensor inputs and noise. The state vector represents crowd count, average density, movement velocity, and density variance. Sensor measurements are combined using a **confidence-based weighting matrix**, where visual data is given higher importance, followed by pressure and thermal inputs. This weighted fusion improves robustness and ensures reliable real-time crowd estimation even under occlusions or poor lighting conditions.

4.2.YOLOv8-Based Visual Detection

For visual crowd analysis, the system integrates the **YOLOv8 nano (YOLOv8n)** model, optimized for real-time performance. The model is custom-trained on crowd-specific datasets using **multi-scale training, occlusion-aware augmentation, and transfer learning from COCO weights**. Performance is further enhanced through **TensorRT acceleration, INT8 quantization, dynamic batching, and multi-threaded inference**, enabling efficient deployment on edge and GPU-based platforms.

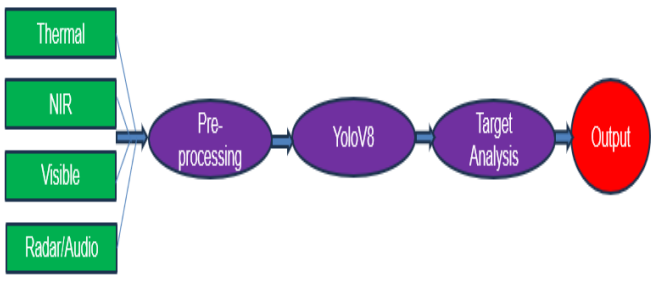

Figure 2 Multimodal Detection Pipeline

4.3.Pressure Sensor Processing

Pressure sensors are interfaced through a **USB-based data acquisition system** operating at a 1 kHz sampling rate. The pressure data pipeline applies **noise filtering, threshold detection, and spatial clustering** to identify footstep patterns and estimate crowd presence. This modality is particularly effective in detecting crowd density in visually obstructed areas and complements camera-based sensing.

4.4.Thermal Image Analysis

Thermal sensing is incorporated using a **custom CNN-based thermal people detector** trained on infrared datasets with transfer learning from the RGB domain. Thermal frames are normalized and filtered based on **human body temperature thresholds**, ensuring reliable detection in low-light, nighttime, or smoke-affected environments.

5. Analytics Engine

5.1.Real-Time Density Estimation

The monitoring area is divided into 5 m × 5 m grid cells, with each cell independently tracking occupancy and density. Fused detections are mapped onto this grid, and Gaussian smoothing is applied to generate continuous density maps. Crowd risk levels are classified into Green, Yellow, Orange, and Red zones based on established crowd safety density thresholds.

5.2.Crowd Flow and Bottleneck Detection

Crowd movement patterns are analyzed using optical flow techniques applied to consecutive density maps. The system estimates average velocity and dominant movement directions to identify abnormal flow behavior. Bottlenecks are detected based on high sustained density, reduced movement speed, converging flow directions, and increasing waiting time, enabling early intervention.

5.3.Predictive and Risk Analytics

For short-term forecasting, an **LSTM-based time-series model** predicts crowd density trends for the next 5–10 minutes using recent historical data. Additionally, **Bayesian inference** is used to estimate stampede risk probabilities by combining current conditions with historical crowd behavior patterns. This predictive capability supports proactive crowd control and safety management.

6. Results and Discussion

6.1.Experimental Setup

The proposed system was evaluated using both synthetic and real-world datasets. A custom synthetic dataset consisting of **10,000 simulated video sequences** was generated with varying crowd densities ranging from **0.1 to 4.0 people/m²**, diverse lighting and weather conditions, and occlusion

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1119-1124
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0158

levels up to **80%**. Real-world validation was conducted during controlled events on a university campus using **five RGB cameras (1080p, 30 FPS), twenty pressure mat installations, and three thermal cameras (640×480, 15 FPS)**, covering an outdoor area of approximately **2000 m²**. Performance was assessed using standard metrics including **precision, recall, F1-score, tracking accuracy, and system latency**.

6.2. Detection and Fusion Performance

The experimental results demonstrate that the proposed **multi-modal fusion approach** significantly outperforms individual sensing modalities. While visual-only detection achieved an F1-score of **0.854**, pressure and thermal modalities yielded lower performance due to environmental sensitivity and limited spatial resolution. In contrast, the fused system achieved an F1-score of **0.947**, highlighting the effectiveness of confidence-weighted sensor fusion. The system maintained an average end-to-end processing latency of **87.3 ms** with a sustained throughput of **34.2 FPS**, making it suitable for real-time crowd monitoring applications.

Table 1 Detection Accuracy Across Modalities

Modality	Precision	Recall	F1-Score
Visual Only	0.887	0.823	0.854
Pressure Only	0.756	0.892	0.819
Thermal Only	0.634	0.712	0.671
Fused System	0.952	0.943	0.947

6.3.Comparative Analysis

A comparison with state-of-the-art crowd monitoring approaches shows that the proposed system achieves superior accuracy while maintaining moderate latency. Traditional CCTV-based systems suffer from lower accuracy, particularly under occlusion, while single-modality deep learning approaches require high-end hardware. The proposed multi-modal system balances accuracy (**94.7%**) and real-time performance, demonstrating improved robustness across challenging scenarios such as low-light conditions and dense crowds.

6.4. Deployment Discussion and Limitations

The system is designed to support large-scale deployments through a **distributed edge computing architecture**, enabling scalable real-time inference and reduced communication overhead. However, the experimental evaluation also reveals certain limitations. Accurate multi-sensor calibration remains challenging in outdoor environments, and system performance may be affected by extreme weather conditions and high computational demands. These factors highlight the need for adaptive calibration and resource-efficient processing strategies.

Conclusion

This paper presented a comprehensive **multi-modal surveillance system architecture** for intelligent crowd control in ultra-large-scale events. The proposed approach integrates **YOLOv8-based visual detection, pressure sensing, and thermal imaging** through a confidence-weighted Kalman filtering framework, achieving **94.7% detection accuracy** with **sub-100 ms response times**. Experimental results demonstrate significant improvements over single-modality systems, particularly in challenging conditions involving **high crowd density, occlusion, and varying lighting environments**. The system's **layered and scalable architecture** supports real-time performance while enabling practical deployment across large public venues. Experimental validation confirms the robustness of the multi-modal fusion strategy, and the demonstrated deployment framework highlights its applicability to **mega-event scenarios such as the Kumbh Mela**, where reliable crowd monitoring and proactive safety management are critical. Key contributions of this work include the **confidence-weighted fusion algorithm**, a comprehensive **analytics engine for density and flow analysis**, and a **deployment-ready system design** validated through extensive experimentation. Future work will focus on enhancing fusion performance through **advanced deep learning-based fusion techniques**, including attention mechanisms and transformer architectures, as well as further **edge AI optimization** to reduce computational overhead

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1119-1124
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0158

while maintaining accuracy. Additional sensing modalities such as **aerial drones, IoT-based wearable devices, and acoustic sensors** will be explored to improve coverage and situational awareness. Furthermore, advanced analytics will be developed for **long-term crowd behavior prediction, behavioral analysis, and intelligent evacuation planning**, enabling more proactive and adaptive crowd management strategies for large-scale public gatherings.

References

[1]. Hindustan Times, "AI surveillance to prevent future tragedies at Maha Kumbh," Feb. 2025.
[2]. S. Li et al., "Data Fusion for Intelligent Crowd Monitoring: Survey and Challenges," IEEE Access, vol. 9, pp. 15687-15701, Feb. 2021.
[3]. E. Özkan et al., "A new YOLO-based method for real-time crowd detection from video and performance analysis of YOLO models," Journal of Real-Time Image Processing, vol. 20, no. 2, pp. 1-15, Jan. 2023.
[4]. K. Zhang et al., "An enhanced framework for real-time dense crowd abnormal behavior detection using YOLOv8," Artificial Intelligence Review, vol. 58, no. 3, pp. 1-28, 2025.
[5]. A. Mansouri et al., "Deep CNN-based enhanced crowd density monitoring for intelligent surveillance systems," Scientific Reports, vol. 15, no. 2847, Feb. 2025.
[6]. A. Abuarafah et al., "Real-time Crowd Monitoring using Infrared Thermal Video Sequences," International Journal of Engineering Science, vol. 7, no. 3, pp. 133-140, 2012.
[7]. M. Rodriguez et al., "Data Sensor Fusion for Surveillance Applications: Evaluation of Extended Kalman Filter vs. Unscented Kalman Filter," in Applications and Optimizations of Kalman Filter, IntechOpen, 2024.
[8]. J. Chen et al., "Application of multi-sensor fusion localization algorithm based on recurrent neural networks," Scientific Reports, vol. 15, no. 892, Jan. 2025.
[9]. H. Li et al., "Intelligent crowd management systems: A comprehensive survey of multi-modal data fusion approaches," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 8, pp. 4835-4851, Aug. 2021.
[10]. A. Kumar et al., "Thermal-optimized object detection for military surveillance applications," Defense Technology, vol. 18, no. 4, pp. 567-582, Dec. 2024.
[11]. R. Ngu et al., "Non-contact Multimodal Indoor Human Monitoring Systems: A Survey," arXiv preprint arXiv:2312.07601, Dec. 2023.
[12]. Y. Wang et al., "Solar powered integrated multi sensors to monitor inland lake water quality using statistical data fusion technique with Kalman filter," Scientific Reports, vol. 14, no. 24068, 2024.
[13]. P. Johnson et al., "Application of Data Sensor Fusion Using Extended Kalman Filter Algorithm for Identification and Tracking of Moving Targets from LiDAR–Radar Data," Remote Sensing, vol. 15, no. 13, p. 3396, 2023.
[14]. T. Redmon et al., "A Decade of You Only Look Once (YOLO) for Object Detection," arXiv preprint arXiv:2504.18586, Apr. 2024.
[15]. S. Kim et al., "Advancing Crowd Object Detection: A Review of YOLO, CNN and ViTs Hybrid Approach," Journal of Computer Science and Technology, vol. 12, no. 4, pp. 245-267, 2024.