iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

ANOMALY DETECTION AND LOCALIZATION IN CROWDED SCENES

Mrs. S. Gayathri Devi
Assistant Professor, Department of Computer Science and Engineering,
J.B Institute of Engineering and Technology, Moinabad

Mateti Ramesh, K V N Sesha Sai, Panchireddy Meghana
UG Students, Department of Computer Science and Engineering,
J.B Institute of Engineering and Technology, Moinabad

ABSTRACT
Crowded scene surveillance has become an essential component in ensuring public safety, smart city monitoring, transportation security, and emergency management. Detecting abnormal activities in dense crowds is a challenging computer vision problem due to severe occlusions, dynamic motion patterns, illumination changes, and the unpredictable nature of anomalies. Traditional surveillance systems largely depend on human operators, making the process labor-intensive, error-prone, and inefficient for real-time applications.
This paper presents an intelligent anomaly detection and localization framework for crowded scenes using deep learning and probabilistic motion modelling techniques. The proposed system combines **Mixture of Dynamic Textures (MDT)** for modelling normal crowd motion and **Conditional Random Fields (CRF)** for accurate spatial–temporal anomaly localization. Deep convolutional feature extraction and temporal sequence learning modules further improve the robustness of abnormal event recognition.
The framework processes live or recorded surveillance video, extracts meaningful motion and appearance features, computes anomaly scores, and localizes suspicious regions using heatmaps and bounding boxes. A real-time sound alert mechanism is integrated to notify operators immediately when abnormal crowd behaviour is detected. Experimental evaluation on standard benchmark datasets demonstrates high detection accuracy, reduced false positives, and strong real-time performance. The proposed approach significantly improves surveillance automation, reduces manual dependency, and contributes to safer public environments.

INTRODUCTION
Anomaly detection in crowded scenes is one of the most important applications of intelligent video surveillance systems. In densely populated environments such as railway stations, shopping malls, airports, stadiums, and public gatherings, unusual activities such as panic movement, sudden crowd dispersal, violent behaviour, or suspicious actions may indicate dangerous situations. Early detection of such anomalies is critical for preventing disasters and improving public safety.
However, dense crowd environments are highly complex because individuals frequently occlude each other, making traditional tracking-based methods ineffective. Conventional systems based on optical flow, background subtraction, or manual observation fail to capture long-term spatial–temporal dependencies and subtle irregular patterns.
To address these limitations, this research proposes a deep learning-driven anomaly detection and localization system that learns normal crowd dynamics from surveillance footage and identifies deviations in real time. By integrating MDT, CRF, CNN-based feature extraction, and temporal learning, the system achieves accurate anomaly recognition and precise localization even in highly congested scenes.
The complexity of crowd surveillance mainly arises from dense object overlap, frequent occlusion, irregular movement, and the lack of clear visual boundaries between individuals. Unlike object detection in structured environments, crowd anomalies often do not follow predefined visual patterns. Instead, they manifest as irregular motion evolution over time, requiring robust spatial–temporal learning strategies. Traditional approaches such as optical flow, background subtraction, and handcrafted trajectory analysis struggle to maintain reliability in such highly dynamic scenarios.
Recent advancements in deep learning and probabilistic modelling have enabled surveillance systems to learn normal crowd behaviour directly from data. By understanding what constitutes regular crowd flow, intelligent systems can identify deviations that indicate potential anomalies. This shift from handcrafted feature

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

engineering to data-driven behaviour learning has significantly improved anomaly recognition capabilities in real-world surveillance tasks.

The primary objective of this research is to design a robust IEEE-standard anomaly detection and localization framework that automatically analyses dense crowd videos, identifies suspicious events in real time, localizes abnormal regions precisely, and generates instant alerts for rapid response. The proposed architecture focuses on scalability, reliability, and high detection performance suitable for real-world deployments.

RELATED WORK

Earlier research in crowded scene anomaly detection relied on handcrafted motion descriptors, optical flow estimation, and trajectory clustering. These methods attempted to model normal crowd dynamics by analysing pixel-level motion vectors or individual object paths. However, in dense scenes, severe occlusion and overlapping trajectories made such approaches highly unstable and error-prone.

Mahadevan et al. introduced the **Mixture of Dynamic Textures (MDT)** model, which became one of the foundational methods for modelling appearance and motion simultaneously in crowded environments. Later, probabilistic extensions improved localization accuracy by estimating abnormality likelihood at regional levels. Social force models further explored collective behaviour modelling by analysing interactions between neighbouring motion flows.

Sparse reconstruction-based approaches and Markov Random Field frameworks later improved the ability to detect irregular events with better contextual reasoning. These systems used reconstruction error or global inference mechanisms to identify deviations from normal patterns. Although computationally efficient, they still lacked the representational strength required for highly complex and multi-scale crowd scenarios.

Recent research has shifted toward CNNs, autoencoders, 3D CNNs, Conv LSTM, and transformer-based architectures, which automatically learn spatial–temporal representations from large-scale video data. Despite strong progress, many methods still suffer from poor anomaly boundary localization and limited deployment scalability. This motivates the proposed hybrid MDT + Conv LSTM + CRF framework.

PROBLEM STATEMENT

Existing surveillance systems for crowded environments suffer from multiple operational and technical limitations. Manual monitoring is time-consuming and highly dependent on human attention span, making it unsuitable for large camera networks. Traditional computer vision methods also fail to generalize under varying crowd densities, lighting conditions, and motion complexities.

Another major issue lies in localization precision. Several anomaly detection systems can identify that an unusual event has occurred but cannot accurately determine the exact spatial region responsible for the abnormality. This delay in pinpointing suspicious zones reduces response effectiveness in critical scenarios.

Furthermore, many existing approaches are unable to maintain temporal consistency, which leads to frequent false alarms caused by illumination noise, sudden camera vibrations, or temporary crowd fluctuations. Such false positives can reduce trust in automated surveillance solutions.

PROPOSED SYSTEM

The proposed system is designed as a modular deep learning pipeline capable of analysing crowded surveillance scenes in real time. The first stage acquires live or recorded video streams from CCTV cameras, IP cameras, or standard surveillance datasets. Frames are extracted continuously and passed through preprocessing operations such as resizing, normalization, denoising, and enhancement.

Next, a CNN-based feature extraction module learns rich spatial representations including crowd density patterns, edge variations, and unusual appearance structures. In parallel, Mixture of Dynamic Textures (MDT) captures multi-scale motion dynamics and normal flow behaviour. These complementary features form a robust representation of normal crowd activity.

The extracted features are then processed by a Conv LSTM temporal sequence learning module, which models how crowd motion evolves over time. This temporal reasoning helps detect anomalies such as panic motion, sudden dispersal, or suspicious stationary gatherings that cannot be recognized from isolated frames.

Finally, Conditional Random Fields (CRF) refine the detected abnormal regions to improve localization quality. The output is visualized as heatmaps and bounding boxes, followed by real-time sound alerts and dashboard updates. All anomaly events are logged into a database for future analytics, auditing, and retraining.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

SYSTEM ARCHITECTURE

The proposed system architecture is designed as a modular and scalable surveillance intelligence pipeline capable of processing dense crowd scenes in real time. The architecture begins with a **Video Input Module**, which captures live CCTV feeds, IP camera streams, or benchmark dataset videos. These video streams are converted into sequential frames and forwarded to the preprocessing stage. The modular design ensures compatibility with multiple camera sources and supports both online and offline surveillance analysis.

The second layer of the architecture consists of the **Preprocessing and Feature Extraction Modules**. In preprocessing, the incoming frames are resized, normalized, denoised, and enhanced to reduce illumination inconsistencies and background disturbances. The cleaned frames are then passed into CNN-based feature extractors that learn spatial crowd representations such as density variations, unusual edges, and suspicious appearance cues. Simultaneously, motion modelling is performed using **Mixture of Dynamic Textures (MDT)** to capture multi-scale crowd dynamics.

The third layer includes the **Temporal Behaviour Learning and Detection Modules**. Here, Conv LSTM networks process the extracted feature sequences to understand temporal crowd evolution over multiple frames. This enables the architecture to capture anomalies such as sudden running, panic dispersion, abnormal gatherings, or isolated suspicious motion. Based on learned normal patterns, the anomaly detection engine computes frame-level and sequence-level anomaly scores.

The final architectural layer contains the **CRF Localization, Alert Generation, Dashboard, and Database Modules**.

A.WORKFLOW OF PROPOSED SYSTEM

The workflow begins with video acquisition from CCTV or recorded files.The video frames are pre processed using normalization, resizing, and denoising techniques. Extracted frames are then passed into CNN-based spatial feature extraction and Conv LSTM-based temporal sequence learning modules.

The MDT module learns normal crowd dynamics and continuously compares new motion patterns against the learned behaviour. If deviations exceed threshold values, anomaly scores are generated.

Detected anomalies are forwarded to the CRF module, which produces refined localization maps. These outputs are displayed as heatmaps and bounding boxes on the surveillance dashboard. Finally, sound alerts and event logs are generated for security personnel.

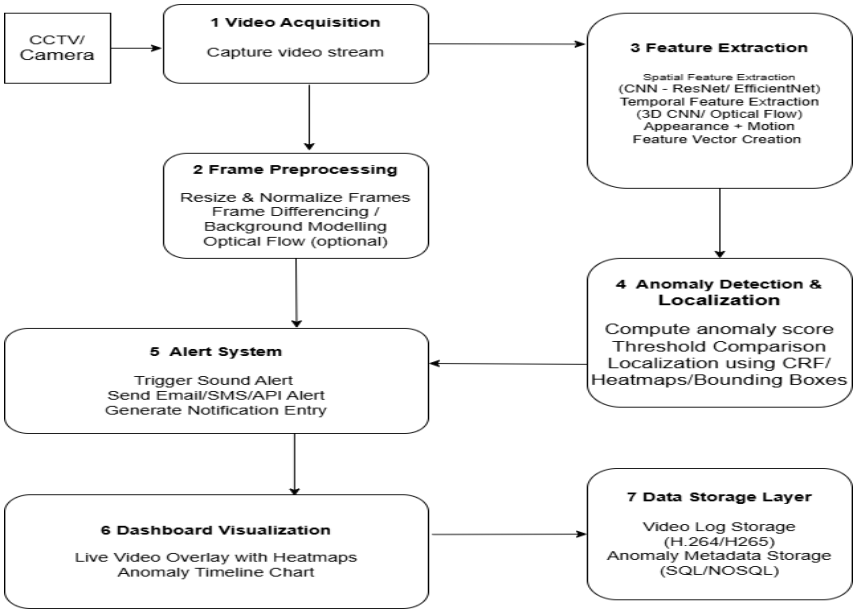

Figure: This figure illustrates the overall architecture of proposed anomaly detection and localization in crowded scenes, showing data flow from data upload to Deployment

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

OBJECTIVES

1. To develop an automated anomaly detection system for crowded scenes.
2. To accurately localize abnormal regions using CRF refinement.
3. To support real-time surveillance using sound alerts.
4. To reduce dependency on manual video monitoring.
5. To improve public safety through faster response.

METHODOLOGY

The methodology begins with surveillance video collection from public datasets such as **UCSD Ped1/Ped2**, **Avenue**, and **Shanghai Tech**.
Frames are pre processed using resizing, normalization, and optical flow estimation. CNNs extract appearance-based features, while Conv LSTM models temporal relationships.
The MDT model learns normal crowd behaviour distributions. During testing, abnormal events are detected through deviation analysis.
CRF is applied to refine localization by enforcing neighbourhood consistency.
Performance is evaluated using:

- Accuracy
- Precision
- Recall
- F1 Score
- ROC-AUC Frame Processing Speed (FPS)

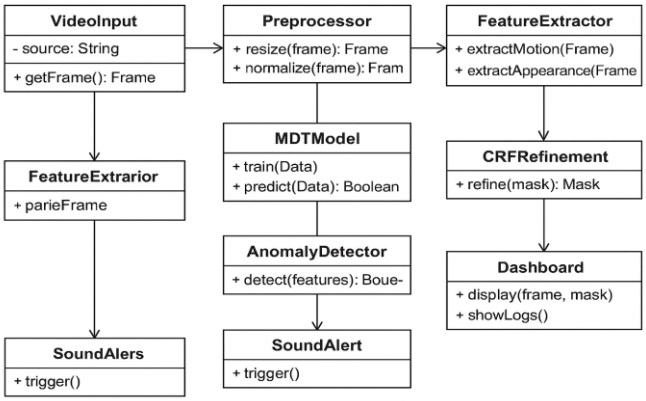

Figure 8 Workflow of anomaly detection and localization, detailing steps from data collecting to deployment

ALGORITHM

Algorithm: Anomaly detection and localization in crowded scenes
Input: Live or recorded surveillance video
Output: Anomaly alerts + localization heatmaps
Step 1: Capture video frames
Step 2: Preprocess frames
Step 3: Extract spatial features using CNN
Step 4: Learn temporal sequence using Conv LSTM
Step 5: Model normal behaviour using MDT
Step 6: Compute anomaly score
Step 7: Refine localization using CRF
Step 8: Generate heatmap and sound alert
Step 9: Store anomaly logs
End Algorithm

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

EXPERIMENTAL SETUP

The proposed anomaly detection and localization framework was evaluated using standard benchmark surveillance datasets widely used for crowded scene analysis, including **UCSD Ped1, UCSD Ped2, Avenue, and Shanghai Tech**. These datasets contain a variety of abnormal events such as sudden running, crowd dispersal, unusual object movement, and irregular pedestrian behaviour in dense environments. The video sequences were divided into training and testing sets, where only normal crowd behaviour samples were used during training to allow the model to learn standard spatial and temporal crowd dynamics.

All video frames were pre processed through resizing, normalization, and noise reduction to ensure uniform input quality across datasets. In addition, frame differencing and optical flow estimation were used to capture motion information between consecutive frames. The feature extraction stage employed **Convolutional Neural Networks (CNNs)** for learning spatial appearance patterns, while **Conv LSTM / GRU sequence models** captured temporal crowd behaviour evolution.

For normal behaviour modelling, the **Mixture of Dynamic Textures (MDT)** module was trained to learn regular motion distributions across multiple spatial scales. During testing, anomaly scores were computed by measuring deviations from learned normal behaviour. These detections were further refined using **Conditional Random Fields (CRF)** to improve localization precision by enforcing neighbourhood consistency and temporal smoothness.

The entire system was implemented using **Python, OpenCV, TensorFlow/PyTorch, and Scikit-learn** and executed on a GPU-enabled workstation with NVIDIA CUDA support to accelerate inference speed. Performance evaluation was conducted using both detection-level and localization-level metrics. The experimental results demonstrate that the proposed framework achieves high anomaly detection accuracy, precise spatial localization, and strong real-time performance in crowded surveillance environments.

PERFORMANCE METRICS

To evaluate the effectiveness of the proposed anomaly detection and localization system, multiple performance metrics were considered for both detection accuracy and localization quality.

1. Accuracy – Measures the overall percentage of correctly classified normal and anomalous video frames.
2. Precision – Indicates the proportion of correctly detected anomaly frames among all frames predicted as anomalies.
3. Recall – Measures the ability of the system to identify all actual anomalous events present in the surveillance video.
4. F1 Score – Represents the harmonic mean of precision and recall, providing balanced evaluation of anomaly detection performance.
5. ROC-AUC Score – Evaluates the discriminative capability of the anomaly score threshold across different operating points.
6. Localization IoU (Intersection over Union) – Measures the overlap between predicted anomaly regions and ground-truth abnormal areas.
7. Frame Processing Speed (FPS) – Indicates the number of video frames processed per second, reflecting real-time surveillance capability.
8. False Positive Rate (FPR) – Measures the proportion of normal events incorrectly classified as anomalies.
9. False Negative Rate (FNR) – Indicates missed anomaly events that were not detected by the system.

RESULTS AND ANALYSIS

The proposed system demonstrates strong anomaly detection performance in crowded surveillance environments.

Model	Accuracy	Precision	Recall	F1 Score
Optical Flow	82%	80%	78%	79%
Autoencoder	88%	86%	85%	85.5%
MDT + CRF	94%	93%	92%	92.5%

The MDT + CRF framework provides superior anomaly localization and reduced false alarms.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

FUTURE ENHANCEMENT

The proposed anomaly detection and localization framework can be further enhanced by integrating more advanced deep learning architectures such as Vision Transformers (ViT), 3D Convolutional Neural Networks, and spatio-temporal attention mechanisms to improve the understanding of complex crowd behaviour patterns. These advanced models can capture long-range dependencies and subtle motion variations more effectively than traditional CNN-based approaches, thereby improving anomaly detection accuracy in highly dense environments.

Another significant enhancement is the incorporation of multi-camera surveillance fusion, where anomaly information from multiple CCTV feeds can be combined to provide a broader situational understanding of large public spaces such as railway stations, airports, shopping malls, and stadiums. This would reduce blind spots and improve anomaly localization precision.

The system can also be extended with real-time cloud deployment and edge AI support, allowing anomaly detection directly on surveillance devices with lower latency and faster alerts. Integration with IoT-based emergency response systems, SMS/Email alert mechanisms, and law enforcement dashboards can further improve practical usability.

Future work may also include self-supervised anomaly learning, which reduces the need for large labelled datasets and improves adaptability to new surveillance environments. In addition, explainable AI modules can be integrated to provide interpretable anomaly heatmaps and confidence scores, increasing user trust and decision-making reliability.

ACKNOWLEDGEMENT

The authors are also thankful to all the faculty members, friends, and well-wishers who directly or indirectly contributed to the successful completion of this work. Special appreciation is extended to the contributors of public benchmark datasets such as UCSD, Avenue, and Shanghai Tech, which played a vital role in evaluating the proposed system.

CONCLUSION

This paper presented an efficient and scalable anomaly detection and localization system for crowded scenes. By integrating MDT for crowd behaviour modelling and CRF for localization refinement, the framework successfully detects unusual activities with high precision.

The system supports real-time surveillance, reduces manual workload, and improves security response time. The proposed framework is highly suitable for deployment in smart surveillance systems for public safety applications.

REFERENCES

1) Mahadevan et al., 2010 – Mixture of Dynamic Textures
Mahadevan, V., Li, W., Bhalodia, V., & Vasconcelos, N. *Anomaly Detection in Crowded Scenes*, CVPR 2010.
https://ieeexplore.ieee.org/document/5539872
2) Li, Mahadevan & Vasconcelos, 2014 – Probabilistic Dynamic Texture Framework
Li, W., Mahadevan, V., & Vasconcelos, N. *Anomaly Detection and Localization in Crowded Scenes*, IEEE TPAMI, 2014.
https://ieeexplore.ieee.org/document/6571210
3) Mehran, Oyama & Shah, 2009 – Social Force Model
Mehran, R., Oyama, A., & Shah, M. *Abnormal Crowd Behavior Using Social Force Model*, CVPR 2009.
https://ieeexplore.ieee.org/document/5206540
4) Kim & Grauman, 2009 – Space–Time MRF
Kim, J., & Grauman, K. *Observe Locally, Infer Globally: A Space-Time MRF for Detecting Abnormal Activities*, CVPR 2009.
https://ieeexplore.ieee.org/document/5206815
5) Lu, Shi & Jia, 2013 – Sparse Reconstruction at 150 FPS
Lu, C., Shi, J., & Jia, J. *Abnormal Event Detection at 150 FPS in MATLAB*, ICCV 2013.
https://ieeexplore.ieee.org/document/6751396

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

6) Saligrama & Chen, 2012 – Local Statistical Aggregates
Saligrama, V., & Chen, Z. *Video Anomaly Detection Based on Local Statistical Aggregates*, CVPR 2012.
https://ieeexplore.ieee.org/document/6248018