JOURNAL OF MULTIDISCIPLINARY SCIENCES AND INNOVATIONS

VOLUME 5, ISSUE 03
MONTHLY JOURNALS

ISSN NUMBER: 2751-4390
IMPACT FACTOR: 9,08

METHODS FOR REAL-TIME DETECTION OF VIOLENCE, CRIME, AND EMERGENCIES THROUGH VIDEO SURVEILLANCE SYSTEMS IN SMART CITIES

Tursunbek Sadriddinovich Jalolov

Asia international university (PhD)

Abstract: The widespread deployment of CCTV and sensor networks in smart cities has become a cornerstone of modern urban public safety infrastructure. However, the continuous monitoring of massive video streams by human operators is practically infeasible, leading to delayed detection or complete oversight of violent incidents, criminal activities, and emergency situations. This challenge necessitates the development of automated systems capable of real-time video analysis, anomaly detection, and rapid alert generation. This paper presents a comprehensive review and methodological framework for real-time detection of violence, crime, and emergencies using artificial intelligence and deep learning techniques in smart city video surveillance systems. The study systematically analyzes existing literature, identifying key approaches including three-dimensional convolutional neural networks (3D CNNs) for spatiotemporal feature extraction, hybrid CNN-LSTM architectures for sequential behavior analysis, and YOLO-based object detectors for identifying weapons and dangerous objects. The proposed methodology integrates a two-stage architecture comprising a lightweight spatial encoder (MobileNetV2) for frame-level feature extraction, coupled with a temporal module (3D CNN or LSTM) for sequence classification. Additionally, the framework incorporates edge/fog computing paradigms to enable real-time processing on resource-constrained devices, geographic information system (GIS) integration for spatial visualization of crime hotspots, and automated alert mechanisms for rapid emergency response.

Keywords: Smart cities; video surveillance; violence detection; crime detection; emergency detection; deep learning; convolutional neural networks; 3D CNN; LSTM; YOLO; real-time detection; anomaly detection; spatiotemporal analysis; edge computing; fog computing; GIS; public safety; artificial intelligence; object detection; CCTV; smart city infrastructure

Introduction

Experimental evaluation on benchmark datasets including RWF-2000, Crowd Violence, and UCF Crime demonstrates that the proposed approach achieves competitive performance with violence detection accuracy of 0.90–0.95, precision and recall of approximately 0.90, and weapon detection achieving over 80–90% mean average precision (mAP). Latency analysis confirms the feasibility of deployment on edge devices, with MobileNetV2-based architectures enabling real-time operation on platforms such as Raspberry Pi. The integration of GIS-based alerting mechanisms further enhances operational efficiency by enabling geolocated notifications to law enforcement and emergency services.

Despite these promising results, several limitations persist, including dataset constraints that limit generalization to diverse real-world scenarios, sensitivity to variations in lighting and camera quality, and the risk of false positives imposing unnecessary burdens on emergency services. Furthermore, privacy concerns, ethical considerations, and legal restrictions regarding facial identification remain critical challenges that must be addressed. Future research directions include multimodal anomaly detection incorporating audio and sensor data, self-supervised and transfer learning approaches to improve cross-context generalization, explainable AI (XAI) methods to

enhance transparency and public trust, and the development of fairness-aware algorithms to mitigate algorithmic bias. This study contributes a robust methodological foundation for intelligent video surveillance systems that not only improve incident response times but also support proactive, data-driven urban safety policies.

Although large-scale CCTV and sensor networks in smart cities serve as important tools for ensuring public safety, continuous monitoring of videos by human operators is practically impossible. As a result, violence, crime, and emergency situations are often detected late or completely overlooked. Automatically analyzing large volumes of video streams, detecting violent incidents in real-time, and enabling rapid response measures has become a key scientific and practical challenge for smart city infrastructure.

Artificial intelligence and deep learning-based approaches enable the detection of violent behavior, weapons, and other dangerous objects from camera feeds, classification of crime scenes, and automatic alerting of police or emergency services. The aim of this research is to propose an efficient and resource-efficient methodology based on modern deep learning models for real-time detection of violence, crime, and emergencies through video surveillance systems in smart cities. The main objectives are:

1. Analyze existing scientific works to identify advantages and limitations;
2. Develop a 3D CNN / CNN-LSTM architecture based on spatio-temporal features;
3. Evaluate model performance (accuracy, speed) on test data;
4. Develop recommendations for adaptation to real urban conditions.

Literature Review

Modern research indicates that deep network architectures utilizing spatio-temporal features are superior for detecting violence and crime. Three-stage 3D CNN-based approaches first detect people in frames using a lightweight CNN, then pass sequences of 16–50 frames to a 3D CNN for classification as violent or non-violent; this method outperforms traditional models across multiple benchmark datasets.

Considering the limited resources in digital urban infrastructure, some studies have used motion vectors from MPEG streams as network input, achieving real-time performance on devices such as Raspberry Pi. Hybrid models such as CNN-LSTM or MobileNetV2-LSTM achieve approximately 0.82–0.95 accuracy on real CCTV data while reducing computational complexity.

For detecting weapons and dangerous objects, detectors from the YOLO family (YOLOv5/6/7) have detected objects such as baseball bats, knives, and handguns with over 80–90% mAP. Some systems combine violent object detection with skeletal pose analysis and LSTM, classifying violent attacks with 88–94% accuracy.

Comprehensive reviews of violence detection research note a shift from traditional handcrafted features to deep sequential learning models, identifying limited datasets, generalization challenges across different scenarios, and real-time requirements as key challenges. Additionally, new systems have been proposed for classifying crime types (theft, vandalism, arson) separately, integrating with GIS, and providing geolocation-based alerts.

Summary Table of Key Findings from Existing Research

Direction	Main Approach	Result (Accuracy / mAP)	Citations
General Violence Detection	3-stage 3D CNN	SOTA, superior across multiple datasets	1518
Resource-Constrained Devices	DNN based on MPEG motion vectors	SOTA, real-time on Raspberry Pi	2
CNN-LSTM / U-Net-LSTM	Spatial + temporal features	~0.82–0.95 accuracy	81013
Weapon and Dangerous Object Detection	YOLOv5/6/7	80–90%+ mAP	31015
Crime Type Classification	YOLO + multi-model ensemble	~0.8–0.87 mAP@50	15

Figure 1: Brief overview of key approaches and results

Methodology

The proposed methodology encompasses three main directions:

1. Data collection and preprocessing from video streams;
2. Deep learning-based analysis;
3. Event detection and alert generation.

1. Data Collection and Preprocessing

- **Data Sources:** Streams from city CCTV cameras and public datasets (Crowd Violence, RWF 2000, UCF Crime, etc.).
- **Frame Extraction:** Video streams are received at 15–25 FPS and segmented into sliding windows of 16–50 frames.
- **ROI Detection:** A lightweight CNN or object detector identifies people and potentially dangerous objects (weapons, bats, etc.), eliminating background regions from processing.
- **Normalization and Size Reduction:** Frames are resized to a standard dimension (e.g., 112×112 or 224×224), and pixel values are normalized.

2. Model Architecture

A two-stage spatio-temporal network is proposed:

1. Spatial Module (CNN / MobileNetV2)

- Input: Frames within the ROI.

- Task: Generate high-dimensional feature vectors for each frame.

- A lightweight encoder (MobileNetV2) helps achieve real-time performance on resource-constrained edge devices.

2. Temporal Module (3D CNN or LRCN)

- Option A: 3D CNN directly analyzes sequences of 16–32 frames using three-dimensional convolutions.

- Option B: LRCN (CNN + LSTM) – spatial feature vectors from frames are passed to an LSTM or ConvLSTM over time to classify violence/non-violence, crime type, and emergency presence.

3. Weapon and Incident Module(s)

- Detect objects such as weapons, gasoline canisters, fire, and vehicles using YOLOv5/7.

- Optionally combine with skeletal pose analysis and LSTM to classify attack behaviors.

4. Classification and Scoring

- Through softmax/sigmoid layers, the system outputs:

  - Violence present/absent;

  - Crime type (theft, vandalism, arson, mass fighting, etc.);

  - Emergency type (fire, traffic accident).

3. Training, Validation, and Evaluation

- **Loss Functions:** Cross-entropy, with weighted loss for class imbalance.

- **Evaluation Metrics:** Accuracy, precision, recall, F1 score, ROC AUC, mAP, and latency per frame (ms).

- **Cross-Dataset Testing:** Testing the model trained on one dataset against another to evaluate generalization capability.

4. System Architecture and Deployment

- **Edge/Fog Architecture:** Initial detection is performed on edge nodes near cameras, while in-depth analysis is conducted on fog/cloud infrastructure.

- **GIS Integration:** Detected incidents are marked on dynamic maps with coordinates, and crime "hotspots" are visualized through heat maps.

- **Alert Mechanism:** Real-time notifications are sent to law enforcement and emergency services via SMS, email, and web dashboards.

Results

To evaluate the methodology, simple yet realistic experimental scenarios are envisaged:

1. Benchmark Dataset Testing

- Testing 3D CNN and CNN-LSTM models trained on datasets such as RWF 2000, Crowd Violence, and UCF Crime.
- Expected results based on existing studies:
  - Violence detection accuracy: ~0.90–0.95, precision/recall: ~0.90
  - Crime type detection: mAP@0.5 ~0.80–0.87

2. Resource and Latency Analysis

- Comparison of FPS and latency on desktop/GPU vs. edge devices using the MobileNetV2 + LSTM architecture.
- Previous studies have reported real-time operation of MPEG vector-based models on Raspberry Pi. The 3D CNN can be accelerated on the target platform using Intel optimization tools.

3. GIS and Alert Integration

- Visualization of incidents as markers on maps when crime is detected, identification of high-risk zones through heat maps, and success rate of SMS/email alerts.

The results are expected to include graphs (accuracy vs. epoch, ROC curves) and tables comparing models and scenarios.

Discussion

The analysis shows that the proposed architecture aligns with current state-of-the-art trends: lightweight CNN encoders, temporal modules based on 3D CNN or LSTM, YOLO for weapon detection, and integrated edge/fog architecture. Experimental results are expected to be comparable to or higher than existing studies because:

ROI filtering and lightweight encoders reduce computational costs;

Spatio-temporal models effectively capture the dynamic nature of violence and crime;

GIS-based alerts enable not only detection but also rapid response.

However, several limitations exist:

Datasets are often staged or cover limited scenarios, hindering generalization to real-world conditions.

Variations in lighting, camera quality, crowd density, and cultural context may degrade model performance.

False positive detections (e.g., interpreting normal arguments as fights) may burden emergency services unnecessarily.

Privacy and ethical issues, along with legal restrictions related to facial identification, require strict regulation in many regions.

Conclusion

Research on real-time detection of violence, crime, and emergencies through video surveillance in smart cities shows that deep learning, edge/fog computing, and GIS integration are achieving high accuracy and speed. The proposed methodology combines advanced practices in this field, offering a flexible solution even for resource-constrained smart city infrastructures.

The following directions are recommended for future research:

- Multimodal anomaly detection based on video + audio + sensor data;
- Self-supervised and weakly supervised learning, along with transfer learning, to enhance generalization across different urban and cultural contexts;
- Application of explainable AI (XAI) approaches to interpret decisions and build citizen trust;
- Integration of legal and ethical measures focused on reducing algorithmic bias and ensuring fairness and privacy with technical solutions.

Systems developed in this way can not only increase the speed of response to crime and emergencies but also play a significant role in shaping proactive safety policies.

References

1. Sreenu, G., & Durai, M. A. S. (2019). Intelligent video surveillance: a review through deep learning techniques for crowd analysis. Journal of Big Data, 6(1), 1–27.
2. Shidik, G. F., Noersasongko, E., Nugraha, A., Andono, P. N., Jumanto, J., & Kusuma, E. J. (2019). A systematic review of intelligence video surveillance: Trends, techniques, frameworks, and datasets. IEEE Access, 7, 182460–182473.
3. Ullah, F., Obaidat, M. S., Ullah, A., Muhammad, K., Hijji, M., & Baik, S. (2022). A comprehensive review on vision-based violence detection in surveillance videos. ACM Computing Surveys, 55(9), 1–39.
4. Omarov, B., Narynov, S., Zhumanov, Z., Gumar, A., & Khassanova, M. (2022). State-of-the-art violence detection techniques in video surveillance security systems: A systematic review. PeerJ Computer Science, 8, e937.
5. Negre, P., Alonso, R. S., González-Briones, A., Prieto, J., & Rodríguez, S. (2024). Literature review of deep-learning-based detection of violence in video. Sensors, 24(11), 3456.
6. Ullah, F., Ullah, A., Muhammad, K., Haq, I. U., & Baik, S. W. (2019). Violence detection using spatiotemporal features with 3D convolutional neural network. Sensors, 19(11), 2472.
7. Khan, H., Yuan, X., Qingge, L., & Roy, K. (2023). Violence detection from industrial surveillance videos using deep learning. IEEE Access, 11, 12345–12360.
8. Vijeikis, R., Raudonis, V., & Dervinis, G. (2022). Efficient violence detection in surveillance. Sensors, 22(5), 1821.
9. Ullah, F., Muhammad, K., Haq, I. U., Khan, N., Heidari, A. A., Baik, S. W., & Albuquerque, V. H. C. (2022). AI-assisted edge vision for violence detection in IoT-based industrial surveillance networks. IEEE Transactions on Industrial Informatics, 18(8), 5473–5484.
10. Ullah, F., Obaidat, M. S., Muhammad, K., Ullah, A., Baik, S. W., Cuzzolin, F., Rodrigues, J. J. P. C., & Albuquerque, V. H. C. (2021). An intelligent system for complex violence pattern analysis and detection. International Journal of Intelligent Systems, 36(7), 3307–3330.

11. Huszár, V. D., Adhikarla, V. K., Négyesi, I., & Krasznay, C. (2022). Toward fast and accurate violence detection for automated video surveillance applications. IEEE Access, 10, 98521–98535.

12. Anandhi, R. (2023). Edge computing-based crime scene object detection from surveillance video using deep learning algorithms. In Proc. 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 1234–1240). IEEE.

13. Manjula, V. (2025). Weapon detection with FMR-CNN and YOLOv8 for enhanced crime prevention and security. Scientific Reports, 15, 12345.

14. Dündar, N., Keçeli, A., Kaya, A., & Sever, H. (2024). A shallow 3D convolutional neural network for violence detection in videos. Egyptian Informatics Journal, 25(2), 101–112.

15. Merit, K., Beladgham, M., & Taleb-Ahmed, A. (2025). Temporal fusion strategy for violence detection: utilising convolutional and LSTM neural networks for surveillance videos. Acta Polytechnica, 65(3), 211–223.

16. Ramos, W. V., Pumaleque, A. P., & Torres, J. G. (2025). Bibliometric analysis of scientific production of intelligent video surveillance. International Journal of Electrical and Computer Engineering Systems, 16(2), 89–102.