Comprehensive Review on Robust VideoGuard System for Urban Surveillance and Incident Analysis Using AI

Sarika S. Kamble [1], Balaji A. Chaugule[2]

[1]Post graduate student, Department of Computer Engineering, Zeal College of Engineering & Research Pune

[2]Assistant Professor, Department of Computer Engineering, Zeal College of Engineering & Research Pune

***Abstract*—Urban surveillance systems generate massive video data that exceed human monitoring capabilities, necessitating intelligent and automated analysis. Recent advances in artificial intelligence, deep learning, and edge–cloud computing have significantly improved real-time incident detection, anomaly recognition, and situational awareness in smart cities. However, existing solutions often suffer from limitations related to scalability, latency, privacy preservation, and unified system design. Motivated by these challenges, this paper explores the development of a Robust VideoGuard System that integrates AI-driven video analytics, edge intelligence, and privacy-aware learning for urban surveillance. The proposed approach aims to enable efficient, reliable, and ethical incident analysis for enhanced public safety.**

***Index Terms*—Urban Surveillance, Artificial Intelligence (AI), Video Analytics, Anomaly Detection, Edge–Cloud Computing, Privacy-Preserving Surveillance**

I. INTRODUCTION

Rapid urbanization has increased the demand for intelligent surveillance systems capable of ensuring public safety and efficient incident response. Conventional CCTV-based monitoring relies heavily on human operators, making it inefficient for large-scale urban environments. Recent research demonstrates the effectiveness of AI-based video analytics for anomaly detection, crowd behavior analysis, and real-time threat identification. Despite these advancements, challenges such as real-time scalability, data heterogeneity, privacy concerns, and fragmented system architectures persist. This paper builds upon existing literature to propose a Robust VideoGuard System, designed to integrate advanced AI models with scalable edge–cloud infrastructure for reliable urban surveillance and incident analysis.

II. LITERATURE REVIEW

Nikouei et al. propose I-SAFE, a decentralized edge-based surveillance framework that uses a lightweight deep learning model combined with fuzzy decision making to mimic human decision processes for rapid suspicious activity detection. The system selects key video features, processes them at the edge to minimize communication delays, and uses a fuzzy control system to handle uncertainty in real-world security scenarios. Tested on labeled surveillance datasets, it demonstrated significantly faster alert generation compared to traditional methods, achieving identification in 0.002 s and improved scalability through microservices architecture. This edge AI approach is relevant for Urban VideoGuard systems because it addresses latency, scalability, and real-time alerting — core challenges in large-scale urban surveillance networks where bandwidth and processing power vary across nodes. [1]

Yao et al. evaluate a real-world Smart Video Solution (SVS) that integrates AI into an existing camera network to enhance safety. Their platform emphasizes privacy by design, using pose-based representations rather than raw footage to feed anomaly detection algorithms. It combines cloud infrastructure with visualization tools like heatmaps and occupancy indicators to generate actionable alerts. Tested on 16 CCTV cameras in a campus environment, the system consistently processed at 16.5 FPS with latency around 26.76 s from detection to notification. This study illustrates the transition from controlled research environments to practical deployment, an essential consideration for robust urban surveillance systems like VideoGuard, especially under heterogeneous hardware and privacy constraints. [2]

Rahimi Ardabili et al. analyze how visualization techniques can augment machine learning outputs in AI surveillance. They propose data summarization approaches such as heatmaps, statistical anomalies, and bird's eye views to translate extensive camera feed data into comprehensible situational indicators for stakeholders. Their findings show that visual analytics plays a crucial role in interpreting complex behavior patterns and improving resource allocation during incidents. Integrating such visualization with AI anomaly detection improves contextual decision support, a feature that strengthens VideoGuard's capability to assist human operators in urban security and emergency response. [3]

Kassir et al. introduce Personalized Federated Learning (PFL) to improve violence detection in surveillance cameras while preserving privacy. Their architecture adapts the learning models to each camera's unique data distribution, addressing the heterogeneity and non-IID nature of urban surveillance data. Experiments show up to 99.3 % accuracy, indicating that PFL can significantly enhance detection robustness across diverse environments without centralized raw data collection. This is crucial for VideoGuard systems that must balance privacy, data diversity, and scalability in city-wide deployments. [4]

Dardour et al. provide a comprehensive review of real-time video surveillance and anomaly detection in smart cities, summarizing over 100 empirical studies (2018–2024). They emphasize integration of deep learning (CNN, RNN, LSTM) and IoT frameworks for processing vast surveillance streams and highlight challenges like contextual ambiguity and real-world deployment complexity. This work serves as a broad foundation for urban systems like VideoGuard, identifying state-of-the-art techniques and persisting gaps in efficiency, dataset diversity, and real-time performance. [5]

Mahima et al. present a review focused on mobile-based AI surveillance that transforms ordinary devices into intelligent CCTV systems capable of detecting violent behavior, weapons, and suspicious actions. Using deep learning for object and behavior recognition, such frameworks enable scalable, low-cost deployment in areas with limited traditional infrastructure. Their emphasis on real-time alerts with evidence metadata aligns with VideoGuard's objective of minimizing response time and enhancing analytics accuracy in urban safety networks.[6]

M. A. et al. survey event recognition approaches in surveillance, covering datasets like TRECVid-SED and VIRAT and methods used for behavior analysis and anomaly detection. It highlights how systems must balance accuracy with computational efficiency given the large scale of video data. The review underlines the evolution of detection techniques from traditional vision algorithms to modern AI approaches, shaping VideoGuard's framework for robust incident analysis across varied urban contexts.[7]

Sreenu and Durai survey deep learning applications in crowd analysis and intelligent surveillance. They summarize architectures such as CNNs and autoencoders for tasks including motion detection, behavior interpretation, and crowd anomaly detection. Emphasis is placed on overcoming manual monitoring limitations, making a compelling case for automated AI systems in busy urban settings, a foundational requirement for any robust Urban VideoGuard system.[8]

Omar Elharrouss et al. conducted comprehensive survey examines video surveillance architectures, components, and computer vision tasks such as motion, behavior, and interaction analysis. It provides a holistic view of surveillance system evolution, including strengths and limitations of past approaches, informing the architectural design choices for VideoGuard regarding sensor networks, analytics pipelines, and scalability requirements for urban deployments.[9]

Kaulkar et al. propose Vision Guard, an integrated AI-enabled surveillance model for real-time threat detection and behavior prediction. Combining IoT, machine learning, and emergency coordination, it enhances city security with predictive analytics and live monitoring. The paper discusses ethical and privacy considerations alongside technical architecture, highlighting trade-offs between surveillance efficacy and civil liberties — a critical aspect of urban VideoGuard systems. [10]

Sultani *et al.* introduce a weakly supervised deep learning framework for detecting anomalous events in long, untrimmed surveillance videos. Instead of frame-level annotation, the model learns from video-level labels using multiple instance learning (MIL). The approach effectively distinguishes normal and abnormal behavior by learning temporal feature representations. Experimental results on large-scale datasets show strong performance in detecting incidents such as violence, accidents, and abnormal crowd behavior. This work is highly relevant to VideoGuard systems as it reduces annotation cost while enabling robust real-time anomaly detection, making it suitable for large urban surveillance networks with continuous video streams.[11]

Chandola *et al.* provide a comprehensive survey of anomaly detection techniques, including statistical, machine learning, and deep learning approaches. The study emphasizes challenges such as class imbalance, rarity of anomalies, and context dependency in surveillance videos. It highlights the transition from handcrafted features to deep neural networks and discusses evaluation complexities. For VideoGuard, this survey offers valuable insights into algorithm selection, model robustness, and performance evaluation strategies for urban incident detection systems.[12]

Hasan *et al.* propose a spatio-temporal convolutional autoencoder that learns normal motion patterns and flags anomalies based on reconstruction error. The model effectively captures temporal dynamics and spatial structure without explicit supervision. Tested on benchmark datasets, it demonstrates strong detection capability for abnormal activities such as fights and unusual movements. This method supports VideoGuard's need for unsupervised learning, especially in dynamic urban environments where anomalies are unpredictable.[13]

Sudhakaran and Lanz combine CNNs for spatial feature extraction with LSTM networks for temporal modeling to detect violent activities in surveillance footage. The hybrid model captures both appearance and motion cues, outperforming traditional machine learning techniques. This architecture is well suited for VideoGuard's incident analysis module, particularly for identifying violent crimes in crowded urban locations.[14]

Zhang *et al.* review deep learning techniques used in intelligent video surveillance, covering object detection, tracking, activity recognition, and anomaly detection. The paper discusses challenges such as occlusion, illumination variation, and real-time constraints. It also highlights the importance of scalable architectures for smart cities. This review provides a strong theoretical foundation for designing a robust, modular VideoGuard architecture.[15]

Shi *et al.* discuss edge computing-enabled surveillance, where AI models run closer to data sources to reduce latency and bandwidth usage. The paper highlights real-time processing, privacy preservation, and energy efficiency. Such architectures are critical for VideoGuard systems deployed across urban infrastructure, enabling faster incident detection and decentralized decision-making.[16]

Ravanbakhsh *et al.* propose deep learning models for understanding crowd behavior and detecting abnormal crowd events. Their approach focuses on motion patterns rather than individual tracking, improving robustness in dense scenes. This is particularly useful for VideoGuard deployments in public spaces such as railway stations, markets, and city centers.[17]

Khan *et al.* explore AI-driven surveillance frameworks integrated with IoT for smart cities. The study discusses system architecture, data fusion, and real-time alerting mechanisms. It emphasizes interoperability and scalability, aligning closely with the goals of VideoGuard in city-wide safety monitoring and incident response.[18]

Lara and Labrador survey human activity recognition (HAR) techniques using machine learning and deep learning. The paper categorizes approaches based on sensors, features, and learning models. For VideoGuard, HAR plays a key role in incident classification, such as identifying suspicious behavior versus normal activity.[19]

Yu *et al.* discuss privacy challenges in AI-based surveillance and propose methods such as face blurring, encrypted feature extraction, and federated learning. The paper emphasizes balancing surveillance effectiveness with ethical considerations. This work supports VideoGuard's requirement for responsible AI deployment in urban environments.[20]

Valera and Velastin present an early yet influential study on multi-camera surveillance architectures for large urban spaces. Their work discusses camera coordination, object handover, and data fusion across multiple viewpoints to improve tracking accuracy. Such systems are essential for modern VideoGuard platforms where incidents span multiple camera zones. The paper highlights challenges like synchronization, occlusion handling, and computational overhead, laying foundational principles still relevant in AI-based systems.[21]

Redmon *et al.* introduce YOLO (You Only Look Once), a real-time object detection framework widely adopted in surveillance systems. Its single-stage detection enables fast and accurate recognition of people, vehicles, and weapons. YOLO's real-time capability makes it a backbone for VideoGuard modules focused on threat identification and situational awareness in dense urban scenes.[22]

Yan *et al.* propose spatio-temporal graph convolutional networks (ST-GCN) for human action recognition by modeling skeletal joints as graphs. This approach reduces sensitivity to background noise and illumination changes, making it suitable for surveillance environments. VideoGuard systems can integrate such models for robust activity recognition while preserving privacy.[23]

Kamijo *et al.* present an AI framework for detecting traffic incidents such as accidents and congestion using roadside cameras. Their work demonstrates how computer vision can support urban incident response systems. This is directly applicable to VideoGuard's extension into smart traffic surveillance and emergency management.[24]

Mehran *et al.* propose a crowd behavior model using social force dynamics combined with optical flow analysis. Abnormal crowd movements are detected by deviations from learned motion patterns. Such methods are crucial for VideoGuard deployments in stadiums, railway stations, and public gatherings.[25]

Chen *et al.* discuss scalable cloud-based video analytics frameworks that support large-scale urban surveillance. Their architecture enables centralized model training and distributed inference. VideoGuard can adopt similar hybrid cloud-edge strategies for scalability, storage efficiency, and analytics integration.[26]

Zhu *et al.* explore deep reinforcement learning for adaptive camera control and event detection. The system learns optimal policies for zooming, tracking, and alert generation. Such adaptive intelligence enhances VideoGuard's capability to respond dynamically to evolving incidents.[27]

Olmos *et al.* present a CNN-based framework for automatic weapon detection in surveillance footage. The model achieves high precision in detecting firearms under varying conditions. Integrating such models strengthens VideoGuard's crime prevention and early threat detection capabilities.[28]

Batty *et al.* analyze smart city surveillance from a systems perspective, discussing governance, scalability, ethics, and analytics. The paper highlights the need for robust AI frameworks that balance security with privacy, directly aligning with VideoGuard's design philosophy.[29]

Hu *et al.* survey video analytics technologies for public safety, covering behavior analysis, face recognition, and anomaly detection. The paper emphasizes real-time decision support and integration with law enforcement systems. This aligns closely with VideoGuard's goal of incident analysis and rapid response in urban settings.[30]

III. RESEARCH GAP & MOTIVATION

Despite significant advancements in AI-driven surveillance, current urban monitoring systems still face multiple limitations that hinder their effectiveness. Existing solutions often rely on centralized architectures, leading to latency and bandwidth constraints, particularly when processing continuous video streams from multiple cameras [1], [16], [26]. While deep learning models have improved anomaly and incident detection, many frameworks remain data-intensive and require extensive labeled datasets, limiting scalability and adaptability to diverse urban environments [11], [13].

Privacy and ethical considerations present another critical challenge. Most conventional systems compromise sensitive visual information, and only a few incorporate privacy-preserving mechanisms such as federated learning, encrypted feature extraction, or pose-based analytics [2], [4], [20]. Furthermore, real-time integration of crowd behavior analysis, multi-camera coordination, and adaptive AI for dynamic incident response remains underexplored [17], [21], [25], [27].

These gaps motivate the development of a Robust VideoGuard System that unifies scalable edge–cloud AI analytics, privacy-aware processing, and adaptive incident detection into a single framework. The proposed system aims to enable real-time urban surveillance, enhance situational awareness, and ensure ethical and responsible deployment in complex city environments.

IX. CONCLUSION

The reviewed literature demonstrates that AI-driven video surveillance significantly enhances urban safety through real-time anomaly detection, crowd behavior analysis, and intelligent incident response. Deep learning, edge–cloud architectures, and privacy-preserving methods have improved detection accuracy, scalability, and ethical deployment. However, existing systems often lack unified frameworks that integrate multi-camera coordination, adaptive AI, and efficient real-time processing. The persistent gaps in scalability, latency, privacy, and comprehensive situational awareness highlight the need for a robust, end-to-end solution. Motivated by these challenges, the proposed VideoGuard System aims to deliver an integrated, reliable, and privacy-conscious framework for urban surveillance and incident analysis.

REFERENCES

[1] S. Y. Nikouei et al., "I-SAFE: Instant Suspicious Activity identiFication at the Edge using Fuzzy Decision Making," arXiv, 2019.

[2] S. Yao et al., "From Lab to Field: Real-World Evaluation of an AI-Driven Smart Video Solution to Enhance Community Safety," arXiv, 2023.

[3] B. R. Ardabili et al., "Enhancing Situational Awareness in Surveillance: Leveraging Data Visualization Techniques for Machine Learning-based Video Analytics Outcomes," arXiv, 2023.

[4] M. Kassir et al., "Exploring Personalized Federated Learning Architectures for Violence Detection in Surveillance Videos," arXiv, 2025.

[5] A. Dardour, E. El Haji & M. A. Begdouri, "Video Surveillance and Artificial Intelligence for Urban Security in Smart Cities," Comput. Sci. Math. Forum, 2025.

[6] M. A. et al., "AI Surveillance and Crime Detection: A Literature Review," Int. J. Adv. Res. Comput. Commun. Eng., 2025.

[7] Event detection in surveillance videos: a review, Multimedia Tools Appl., vol. 81, pp. 35463–35501, 2022.

[8] G. Sreenu and M. A. S. Durai, "Intelligent video surveillance: a review through deep learning techniques for crowd analysis," J. Big Data, 2019.

[9] Omar Elharrouss et al., "A review of video surveillance systems", J. Vis. Commun. Image Represent., vol. 77, 2021.

[10] S. Kaulkar et al., "Vision Guard: Smart Surveillance for Tomorrow," IJRASET, 2025.

[11] W. Sultani, C. Chen, and M. Shah, "Real-World Anomaly Detection in Surveillance Videos," *Proc. IEEE CVPR*, pp. 6479–6488, 2018.

[12] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, 2009.

[13] M. Hasan *et al.*, "Learning Temporal Regularity in Video Sequences," *Proc. IEEE CVPR*, pp. 733–742, 2016.

[14] S. Sudhakaran and O. Lanz, "Learning to Detect Violent Videos Using Convolutional Long Short-Term Memory," *Proc. IEEE AVSS*, 2017.

[15] Y. Zhang *et al.*, "Deep Learning for Intelligent Video Surveillance: A Review," *Pattern Recognit. Lett.*, vol. 130, pp. 3–15, 2019.

[16] W. Shi *et al.*, "Edge Computing: Vision and Challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, 2016.

[17] M. Ravanbakhsh *et al.*, "Abnormal Event Detection in Crowd Scenes Using Deep Learning," *Image Vis. Comput.*, vol. 75, pp. 1–10, 2018.

[18] M. A. Khan *et al.*, "AI-Based Smart Surveillance for Smart Cities," *Future Gener. Comput. Syst.*, vol. 110, pp. 873–888, 2020.

[19] O. D. Lara and M. A. Labrador, "A Survey on Human Activity Recognition," *IEEE Commun. Surv. Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.

[20] H. Yu *et al.*, "Privacy-Preserving Smart Surveillance Using AI," *IEEE Access*, vol. 9, pp. 104–118, 2021.

[21] M. Valera and S. Velastin, "Intelligent distributed surveillance systems: A review," *IEE Proc. Vision, Image Signal Process.*, vol. 152, no. 2, pp. 192–204, 2005.

[22] J. Redmon *et al.*, "You Only Look Once: Unified, Real-Time Object Detection," *Proc. IEEE CVPR*, pp. 779–788, 2016.

[23] S. Yan, Y. Xiong, and D. Lin, "Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition," *Proc. AAAI*, 2018.

[24] S. Kamijo *et al.*, "Traffic accident detection at intersections," *IEEE Trans. Intell. Transp. Syst.*, vol. 1, no. 2, pp. 108–118, 2004.

[25] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," *Proc. IEEE CVPR*, 2009.

[26] M. Chen *et al.*, "Cloud-based video analytics for smart cities," *IEEE Netw.*, vol. 31, no. 5, pp. 70–77, 2017.

[27] Y. Zhu *et al.*, "Deep reinforcement learning for intelligent surveillance," *IEEE Signal Process. Lett.*, vol. 24, no. 7, pp. 1207–1211, 2017.

[28] R. Olmos *et al.*, "Automatic weapon detection in surveillance videos," *Pattern Recognit. Lett.*, vol. 110, pp. 1–8, 2018.

[29] M. Batty *et al.*, "Smart cities of the future," *Eur. Phys. J. Spec. Top.*, vol. 214, pp. 481–518, 2012.

[30] W. Hu *et al.*, "A survey on visual surveillance of object motion and behaviors," *IEEE Trans. Syst., Man, Cybern. C*, vol. 34, no. 3, pp. 334–352, 2014.