Campus Surveillance with Violence Insight and Suspect Profiling: A Survey

Deepak N R[1,†], Santosh Kumar Paital[2,*,†],Umme Kulsum[3,†], Shaima Afreen[4,†] and Shrey Verma[5,†]

[1] *Professor, Department of Computer Science and Engineering, HKBK College of Engineering, Bengaluru, India*
[2345] *Department of Computer Science and Engineering, HKBK College of Engineering, Bengaluru, India*

Abstract

Campus safety is a paramount concern in educational institutions worldwide, especially in combating the prevalence of campus violence. This comprehensive review utilizes advanced image processing and computer vision technologies to proactively tackle security challenges. The appraised surveillance system integrates sophisticated methods, employing Convolutional Neural Networks (CNNs), bidirectional LSTM models, and the Convolutional 3D (C3D) network architecture. Leveraging transfer learning from ImageNet and strategic data augmentation, the system aims to refine its capabilities for robust performance. Through the integration of advanced techniques in image analysis and data processing, the system seeks to enhance security measures while respecting individual privacy. By leveraging an integrated surveillance framework, the emphasis is on delivering timely alerts to security personnel and providing comprehensive behavioural profiles. This research significantly contributes to the progression of security systems within educational contexts, emphasizing the significance of employing state-of-the-art computational methodologies. By prioritizing the creation of safer educational environments on a global scale, this approach stands as a pivotal stride in fortifying campus security measures.

Keywords

Computer vision, Convolutional Neural Networks, Image processing, Violence, LSTM

1. Introduction

The growing concern for safety within educational institutions globally, driven by the need to address security challenges such as violence and unauthorized access, has prompted a crucial shift. Integration of cutting-edge surveillance technologies, leveraging advancements in image processing and computer vision, has emerged as a pivotal strategy to fortify security measures within these environments. This exploration delves deeply into potential methodologies, such as Convolutional Neural Networks (CNNs), bidirectional LSTM models, and the Convolutional 3D (C3D) network architecture, aimed at enhancing

Symposium on Computing & Intelligent Systems (SCI), *May 10, 2024, New Delhi, INDIA*

* Corresponding author.

† These authors contributed equally.

deepaknrgowda@gmail.com (Deepak N R); santoshpaital@gmail.com (S. K. Paital); ummekulsum304162@gmail.com (U. Kulsum); shaimaafreen2001@gmail.com (S. Afreen); shreyverma9452@gmail.com (S. Verma)
0000-0002-2766-2990 (Deepak N R)

surveillance capabilities tailored specifically for security purposes. This investigation deeply explores the aspects of campus surveillance technologies, aligning closely with the primary objectives. These objectives encompass the development of a comprehensive surveillance system. The focus is on investigating methodologies and technological advancements essential for establishing a robust surveillance infrastructure, including incident detection, suspect identification, real-time suspect monitoring, and an integrated alert system. Moreover, the exploration aims to create an accurate facial recognition system. This involves exploring diverse methods and advancements in unauthorized face detection systems to strengthen security measures by implementing a precise system capable of efficiently identifying potential threats within the campus environment.

While prioritizing security measures, this study recognizes the delicate balance between safety and fostering a nurturing environment. It aims not only to explore surveillance technology capabilities but also to analyse their impact on the overall campus atmosphere. The overarching goal of these objectives is to serve as a valuable resource fostering an understanding of cutting-edge approaches in campus surveillance. Shedding light on their effectiveness, challenges, and potential consequences aims to significantly contribute to bolstering campus security measures and nurturing educational environments globally. Furthermore, there is a strong emphasis on continuously evaluating and adapting surveillance systems. Advocating for ongoing assessment and adaptation of surveillance technologies ensures their efficacy in addressing emerging threats, adapting to evolving landscapes, and meeting the changing needs of educational environments. This iterative approach fosters a culture of enhancement, ensuring the durability and effectiveness of campus surveillance technologies.

2. Background

In the realm of computer vision, the journey from raw visual data to actionable insights is a meticulously orchestrated sequence of steps, each playing a pivotal role in extracting valuable information. Initiating the process, the capture of data through cameras or sensors marks the first step. Following this, crucial preprocessing steps come into play, acting as the foundation for subsequent analysis. These preprocessing tasks encompass a spectrum of operations, including but not limited to, noise reduction techniques to eliminate unwanted distortions, fine-tuning brightness levels for optimal clarity, and preparing images in a standardized format conducive to in-depth analysis. As the journey progresses, the preprocessed visual data traverses into the segmentation phase, where advanced techniques like edge detection partition the images into discernible regions or objects. This segmentation process acts as a precursor to detailed analysis by effectively isolating specific elements within the images, laying the groundwork for a deeper dive into individual components.

Advancing further, the process accentuates feature extraction, a critical phase where the system meticulously identifies and isolates crucial characteristics such as shapes or textures within the segmented regions. This intricate extraction process plays a pivotal role in distilling the most pertinent details essential for subsequent analysis and interpretation, setting the stage for insightful observations.

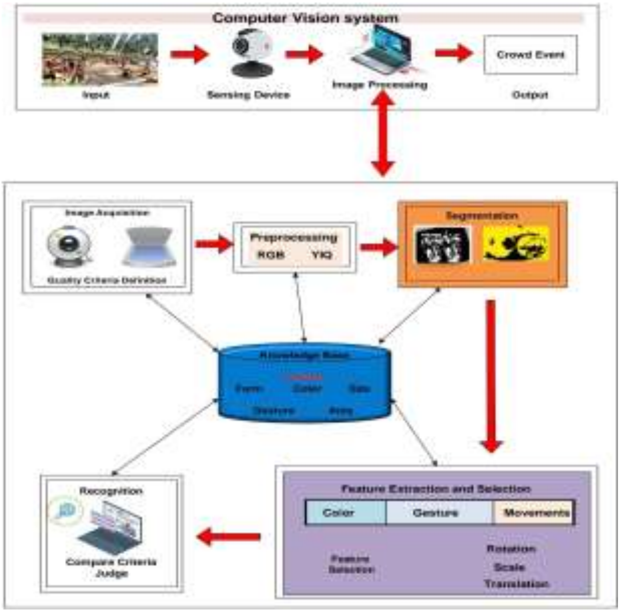

Figure 1: Basic Concept of Computer Vision with Image Processing.

Finally, culminating in the recognition stage, the interpreted data, derived from the extracted features, is employed to identify objects, faces, or patterns. Leveraging cutting-edge machine learning or deep learning algorithms, computers adeptly interpret visual information across multifaceted applications, spanning from healthcare to autonomous vehicles and surveillance systems. Illustrating these sequential steps in Fig. 1. The Basic Concept of Computer Vision with Image Processing provides a comprehensive visualization of the intricate process by which computer vision meticulously extracts, refines, and interprets insights from raw visual data, empowering its widespread application across diverse fields and industries.

3. Literature review

Abbass and Kang (2023) [1] conducted a thorough investigation into violence detection in surveillance videos using the UBI Fights dataset. Their research introduced six dis- tinct architectures, incorporating the Convolutional Block Attention Module (CBAM) with ConvLSTM2D or Conv2D and LSTM layers, as well as integrating CBAM with established models like ResNet50, VGG16, or MobileNet for efficient spatio-temporal feature extraction. Through rigorous evaluation and comparison with state-of-the-art methods on the UBI Fights dataset, their work showcased the superior performance of these architectures. Looking ahead, the authors plan to extend their approach to additional datasets, aiming for a more generalized capability in violence detection, thus contributing to the advancement of surveillance video analysis techniques.

Aktı et al (2022) [2] introduced the Social Media Fight Images (SMFI) dataset, a unique collection comprising samples gathered from media platforms and camera recordings. Their research underscored the model's impressive ability to accurately differentiate between images depicting fighting and non-fighting actions, with a particular emphasis on the model's resilience even when excluding a substantial portion (60%) of the dataset. Furthermore, experiments conducted on video-based fight recognition datasets demonstrated the model's effective classification using randomly selected frames. These cumulative findings position the SMFI dataset as a notable contribution to the field, demonstrating superior accuracy, especially on previously unexplored datasets, and indicating its potential to advance models in the domain of fight recognition.

Aldehim et al (2023) [3] presents a groundbreaking technique called TSODL VD. This method aims to identify instances of violence, in surveillance videos, with precision and efficiency. By incorporating the TSO protocol as a hyperparameter enhancer for the residual DenseNet model, the researchers have improved the effectiveness of violence detection, in the TSODL VD procedure. Through experiments conducted on a violence dataset, the study demonstrates that this approach achieves accurate detection results, surpassing previous state-of-the-art methods.

Cao et al (2023) [4] contribute to the domain of supervised Video Anomaly Detection (VAD), with a specific focus on the NWPU Campus dataset, recognized as the largest dataset of its kind. This dataset stands out due to its emphasis on scene-dependent anomalies, providing a tailored frame- work for Video Anomaly Anticipation (VAA). The researchers propose an innovative scene-conditioned model adept at effectively addressing both Video Anomaly Detection (VAD) and Video Anomaly Anticipation (VAA) by giving priority to anomaly management across diverse scenes. Their commitment extends beyond short-term VAA, underscoring their dedication to advancing the understanding and application of anomaly detection and anticipation in video data, particularly emphasizing long-term anticipation.

Garcia-Cobo and Sanmiguel (2023) [5] introduce a cutting edge architecture for violence detection in surveillance videos, outperforming existing models in real-time operation through careful parameter optimization. The method seamlessly integrates pose estimation and dynamic temporal changes to identify violence instances autonomously. The authors stress the crucial significance of implementing this unique combination technique to improve overall system performance, highlighting the pivotal role of integration in violence detection methodologies. This pioneering approach represents a significant advancement in surveillance video analysis, offering a more effective and efficient solution to address contemporary challenges in the field.

Ghadekar et al (2023) [6] developed an advanced violence detection system capable of accurately identifying instances of violence in both textual content and videos. Notable for its commendable precision and adept avoidance of overfitting, this system emerges as a highly valuable tool for integration into surveillance systems and social media platforms, particularly where violence and harassment are prevalent. A distinctive feature of their work is the deliberate incorporation of diverse files into the dataset, strategically refining the system to capture various facets of violence. This enhancement significantly strengthens the system's versatility, establishing it as an effective means for proactively addressing potential forms of violence before they escalate into serious incidents.

Husz´ar et al (2023) [7] introduced two innovative architectures, FT and TL, designed for the classification of violence in video clips by utilizing action recognition features from the Kinetics 400 dataset. In the FT model, optimization is performed on X3D M parameters trained on Kinetics 400, while the TL model extracts spatio features without modifying these parameters and incorporates training multiple fully connected layers. Evaluation results, including dataset validation, highlight the TL model's superior generalization to unseen scenarios compared to the FT model. However, a recognized limitation in the evaluation method involves the use of non-overlapping four-second video segments, which may not adequately capture violent events spanning two segments. To address this limitation, the authors plan to implement strategies in future work, such as adjusting segment sizes or incorporating overlapping segments, aiming to strike a balance between the speed and accuracy of analysing violent events.

Kang et al (2021) [8] introduce a cutting-edge violence detection system, featuring attention modules that strategically target spatial and temporal dimensions to group frames effectively. Supported by comprehensive experiments, the paper underscores the seamless integration of these modules with efficient 2D CNN backbones, establishing their efficacy. The authors highlight a significant accomplishment with the successful deployment of a real-time violence recognition system in a resource-limited environment, showcasing the practical viability of their methodology. Looking ahead, they articulate plans to fortify the model's robustness through data utilization and explore data augmentation techniques. Furthermore, the researchers express their commitment to expanding their investigation into action recognition tasks, aiming to apply their versatile approach across a range of contexts.

Mahareek et al (2023) [9] introduced an innovative model for detecting anomalies in surveillance videos. This model seamlessly integrates 3DCNN and ConvLSTM architectures, effectively addressing challenges in anomaly detection. The demonstrated performance includes flawless 100% training accuracy across five datasets. The model exhibits high recognition reliability, achieving evaluation scores of 98.5%, 99.2%, and 94.5%. When compared to the standalone 3DCNN, the integrated 3DCNN+ConvLSTM configuration consistently outperforms across all datasets. The authors underscore their dedication to research, emphasizing the model's proactive capabilities in anticipating anomalies in surveillance videos, representing a significant advancement in the field.

Perseghin and Foresti (2023) [10] introduce the School Violence Detection (SVD) system, proposing an innovative method that utilizes a 2D Convolutional Neural Network (CNN) for categorizing actions within educational settings. Their focus on cost-effectiveness and efficiency involves the utilization of school data to alleviate computational demands, resulting in an impressive 95% classification accuracy achieved through techniques like web scraping and transfer learning. Despite these accomplishments, challenges persist, particularly in addressing image noise and acquiring images of minors. To advance the SVD system, potential improvements may involve analysing frame time series, integrating DSB dataset data, and collaborating with experts in psychology and education. This dual emphasis on technological refinement and inter-disciplinary collaboration underscores the commitment of Erica Perseghin and Gian Luca Foresti to comprehensively address the intricacies of school violence detection.

Siddique et al (2022) [11] proposed an extensive framework designed to identify violence in sensitive areas by analysing video streams from surveillance cameras through computer vision and machine learning techniques. Their emphasis lies on the significant role played by Violent Flows (ViF) Descriptors in enhancing the accuracy of violence detection. The researchers conducted a thorough feasibility analysis and performance validation of diverse machine learning techniques, discovering that the combination of ViF with weighted averaging on Linear SVM, Cubic SVM, and Random Forest techniques produce outstanding results in detecting violence within video streams. They also recommend exploring the integration of cloud computing to expedite video stream processing and suggest refining the framework by incorporating a face detection module for identifying individuals involved in activities. In summary, Singh, Preethi, et al.'s framework employs cutting-edge technologies to ensure reliable violence detection in surveillance scenarios, addressing technical aspects and proposing practical enhancements for improved real- world applicability.

Singh et al (2020) [12] presents a groundbreaking methodology for anomaly detection in CCTV footage, integrating abnormal videos to enhance accuracy with a dual-network system and a labelled dataset. Their resulting anomaly detection model, validated on a dataset featuring 12 real- world anomalies, achieves an impressive accuracy rate of 97.23%, eliminating the need for laborious annotations in abnormal segments. Addressing overfitting concerns, their Threat Recognition Model categorizes anomalies into thirteen classes, employing techniques such as frame concatenation and categorical cross-entropy. This literature review underscores the substantial contributions of Singha, Singh, and their colleagues, emphasise the practical implications of their model and the importance of timely implementation while considering hardware constraints for computational efficiency and cost- effectiveness. Their work represents a significant advancement in the field of anomaly detection in CCTV footage, providing valuable insights for both researchers and practitioners.

Ullah et al (2023) [13] introduced a system for violence detection in surveillance videos, leveraging computer vision and AI techniques to enhance security. Their research, spanning from 2011 to the present, centres on the integration of neural networks (NNs) and emphasizes the significant contributions of artificial neural networks (ANNs) in advancing violence detection within computer vision. The study highlights the broad applications of these advancements, particularly in improving security and safety in urban environments and large-scale industries. Overall, their work underscores the evolving landscape of violence detection technologies and their potential impact on diverse sectors.

Vieira et al (2022) [14] conducted a comprehensive study focusing on the analysis and application of cost-effective techniques employing Convolutional Neural Networks (CNNs) for automated event recognition and classification. Their experiments revealed the effectiveness of specific architectural parameters, resulting in an impressive classification accuracy of up to 92.05%. To enable a detailed comparison of performance across various CNN models, the researchers devised a prototype for an intelligent monitoring system with a budget-friendly implementation, deployed on a Raspberry Pi as an embedded platform, achieving a real-time frame rate of up to 4.19 FPS. Going forward, the team aims to explore the implications of implementing mobile CNN architectures on embedded platforms. Furthermore, they plan to enrich their dataset by incorporating videos depicting non-violent actions in diverse settings like malls, air- ports, subways, parks, and sports stadiums, with the goal of significantly enhancing the models' applicability and accuracy in event recognition and classification.

Table 1
Comparison of Various Existing System with our System

Ref. no.	Author	Violence detection	Face Identification	Event Place Alert	Live Location Alert	Unknown Face Alert
1	Abbass et al(2023)	Yes	No	No	No	No
2	Aktı et al (2022)	Yes	No	No	No	No
3	Aldehim et al (2023)	Yes	No	No	No	No
4	Cao et al (2023)	Yes	No	No	No	No
5	G. Cobo et al (2023)	Yes	No	No	No	No
6	Ghadekar et al (2023)	Yes	Yes	No	No	No
7	Husz´ar et al (2023)	Yes	No	No	No	No
8	Kang et al (2021)	Yes	No	Yes	No	No
9	Mahareek et al (2023)	Yes	No	No	No	No
10	Perseghin et al(2023)	Yes	No	Yes	No	No
11	Siddique et al (2022)	Yes	No	Yes	No	No
12	Singh et al (2020)	Yes	No	No	No	No
13	Ullah et al (2023)	Yes	No	No	No	No
14	Vieira et al (2022)	Yes	No	No	No	No
15	Ye et al (2021)	Yes	No	No	No	No

Ye et al (2021) [15] introduced an inventive method for campus violence detection by combining video sequence analysis with speech emotion evaluation. Through simulated scenarios, they utilized video samples from surveillance cameras representing both violent and non-violent situations, alongside speech samples from three databases for comprehensive testing. The study showcased an impressive recognition accuracy of 97%, outperforming existing methods when applied to the amalgamated video and emotional databases. A noteworthy contribution surfaced in the development of a fusion algorithm, showcasing a significant 10.79% enhancement in accuracy compared to previous fusion rules. In their future research agenda, Liang Ye and Tong Liu outlined plans to explore skeleton-based activity recognition methods for campus violence detection, aiming to compare them with established body-based approaches. The manuscript also proposed integrating emotion recognition to augment activity recognition and vice versa, indicating a promising avenue for continued research in this field.

The Table 1: Comparison of Various Existing System with our System presents a comprehensive snapshot of key components within security systems, encompassing violence detection, face identification, event place alert, live location alert, and unknown face identification. All existing systems, as exemplified in Table 1, have made notable strides in the realm of violence detection, showcasing diverse methodologies. These systems contribute valuable insights into the development of sophisticated algorithms and techniques for accurately identifying and responding to violent incidents. It takes Ghadekar et al (2023) [6] the lead in addressing face identification, shedding light on advancements in facial recognition technology. Moreover, event place alert systems are explored in Kang et al (2021) [8], Perseghin and Foresti (2023) [10] and Siddique et al (2022) [11], highlighting the significance of targeted threat response in specific locations. However, despite the progress in these areas, current existing systems exhibit a gap in the crucial domains of live location alert and unknown face identification. This underscores the need for innovative solutions, prompting our proposed objectives to pioneer real- time live location alerts and robust unknown face identification capabilities, thereby enhancing the overall efficacy of security measures in diverse contexts.

4. Proposed Work

Our proposed approach to address the gap identified in the literature review by devel- oping a surveillance system, for campuses. This system focuses on preventing entry and detecting activities through real-time analysis of CCTV camera footage. Firstly, we continuously extract frames from the CCTV feeds, allowing us to analyse the data in depth. For identifying entry, we employ object detection techniques to detect individuals entering the campus without authorization. Additionally, we utilize facial recognition algorithms to identify authorized personnel and promptly alert the system with information about when and where the unauthorized entry occurred.

Simultaneously, our methodology extends to analysing frames for detecting violence using advanced computer vision algorithms and anomaly detection models. When violent activities are detected, a series of steps are triggered for identification and recognition. Facial recognition algorithms play a role in identifying individuals involved in acts while tracking mechanisms monitor their movements across frames, creating a timeline of the incident. Real-time alerts are then activated, providing notifications to security personnel. It is crucial to note that these alerts contain information, about both the location where the violence took place and where suspected individuals are currently located, greatly enhancing the system's responsiveness.

To improve our system continuously, we integrate machine learning algorithms ConvLSTM, CNN Abbass and Kang (2023) [1] that enhance the accuracy of detecting entries and violent activities. This involves updating our database of personnel and refining recognition models through ongoing analysis. We prioritize privacy and ethics by implementing measures to comply with standards, such, as anonymizing data and restricting access to information. Our methodology also emphasizes documentation and reporting generating incident reports maintaining audit logs and taking an approach to continuously enhancing campus surveillance. This comprehensive methodology represents an advancement in surveillance effectively addressing the identified gap and establishing a proactive approach, to campus security.

The strategy to overhaul campus safety involves creating a sophisticated surveillance system (depicted in Fig. 2. Proposed Methodology Flowchart) that utilizes advanced image processing and computer vision technologies. This theoretical analysis responds to the urgent need to address campus violence in educational institutions worldwide. The proposed surveillance system incorporates a fusion of sophisticated methods including Convolutional Neural Networks (CNNs) Abbass and Kang (2023) [1] Kang et al (2021) [8] Perseghin and Foresti (2023) [10] Vieira et al (2022) [14], bidirectional LSTM models, Convolutional 3D (C3D) networks Mahareek et al (2023), Campus Block based Location with GPS, Haar Cascade for face detection, and LBPH algorithm for person recognition. LSTMs Abbass and Kang (2023) [1] are a type of recurrent neural network (RNN) designed to process and analyse sequences of data, effectively capturing long-term dependencies. In the context of campus safety, LSTMs can be utilized for behavioural analysis and pattern recognition within video sequences.

Unlike standard RNNs, LSTMs Abbass and Kang (2023) possess a more sophisticated memory cell structure, allowing them to retain and selectively forget information over extended periods. In this surveillance system, bidirectional LSTM Abbass and Kang (2023) [1] Mahareek et al (2023) [9] models are employed to understand and analyse the temporal patterns in video data collected from CCTV cameras across the campus. This model excels in comprehending sequences of frames, identifying behavioural patterns, and detecting anomalous activities that may signify potential threats or violent behaviour. By leveraging LSTM's ability to capture temporal dependencies, it aids in recognizing suspicious behavioural sequences, thus contributing to the violence detection and suspect identification processes. CNNs Abbass and Kang (2023) [1] Kang et al (2021) [8] Perseghin and Foresti (2023) [10] Vieira et al (2022) [14], are specialized neural networks primarily used for image processing tasks, such as object recognition and feature extraction. They consist of multiple layers, including convolutional layers, pooling layers, and fully connected layers, enabling them to learn hierarchical representations of visual data. In the context of campus safety, CNNs Abbass and Kang (2023) [1] Kang et al (2021) [8] Perseghin and Foresti (2023) [10] Vieira et al (2022) [14], play a crucial role in the violence detection phase. The surveillance system employs Convolutional Neural Networks to analyse the frames extracted from the video data. These networks utilize convolutional layers to extract meaningful spatial features from each frame, recognizing patterns indicative of violent behaviour. The integration of transfer learning from ImageNet and strategic data augmentation techniques further refines CNN's ability to identify violence-related cues accurately.

By integrating both LSTM Abbass and Kang (2023) [1], and CNN Abbass and Kang (2023) [1] Kang et al (2021) [8] Perseghin and Foresti (2023) [10] Vieira et al (2022) [14], architectures, the surveillance system can effectively process and interpret video data in a holistic manner. The LSTM Abbass and Kang (2023) [1] models contribute to understanding temporal dynamics and behavioural patterns, while CNNs Abbass and Kang (2023) [1] Kang et al (2021) [8] Perseghin and Foresti (2023) [10] Vieira et al (2022) [14], excel in extracting spatial features and identifying violent behaviour within individual frames. Their combined functionality enhances the system's capability for violence detection, suspect identification, and real-time alert generation, contributing significantly to campus safety measures. The methodology commences with the acquisition of data through CCTV cameras installed across the campus.

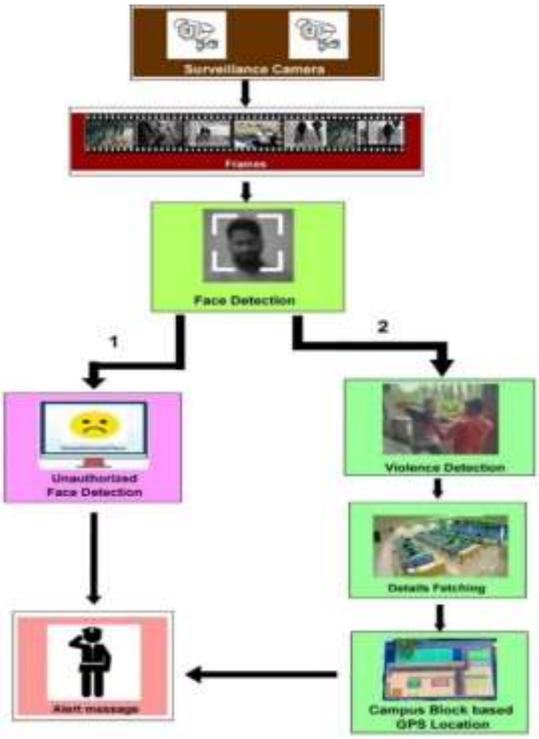

Figure 1: Proposed Methodology Flowchart.

The collected video data undergoes a frame extraction process, subsequently subjected to violence detection algorithms via sophisticated image processing techniques. Upon detecting violence, the system triggers suspect identification processes employing Haar Cascade for face detection and the LBPH algorithm for person recognition. Upon successful suspect identification, the system generates real-time alerts notifying security personnel about the suspect presence in the particular camera feed. Incorporating the IPStack algorithm facilitates GPS Location data retrieval, contributing to comprehensive suspect monitoring. Iterative refinement of the system involves continual data enhancement and model optimization. Emphasis is placed on seamless integration with existing surveillance infrastructure to ensure efficient real-time suspect tracking and campus-wide security reinforcement.

The design methodology underscores the system's commitment to delivering timely alerts to security personnel and constructing comprehensive behavioural profiles. It navigates a delicate balance between heightened security measures and safeguarding individual privacy rights. This approach represents a significant stride in employing state-of-the-art computational methodologies to fortify campus security measures globally.

The culmination of this methodology manifests in the development of a sophisticated surveillance framework capable of swiftly detecting, identifying, and alerting security personnel to potential threats within educational settings. By prioritizing the creation of safer educational environments, this approach stands at the forefront of enhancing campus safety on a global scale.

5. Discussion

Exploring methodologies for detecting violence on college campuses through computer vision and image processing techniques reveals notable progress. However, as we pivot from conventional approaches, a distinct gap emerges, warranting attention in our project survey paper. Firstly, prior studies predominantly focused on anomaly detection, employing algorithms like 3DCNN Mahareek et al (2023) [9], LSTM Abbass and Kang (2023) [1], and CNNs Abbass and Kang (2023) [1] Kang et al (2021) [8] Perseghin and Foresti (2023) [10] Vieira et al (2022) [14] on embedded platforms. While effective in identifying actions, these methods often lacked the ability to recognize individuals involved in real-time during such incidents. Our enhancement involves the introduction of real-time surveillance-based person recognition, advancing beyond mere activity detection to immediate identification of individuals, significantly enhancing system responsiveness to security incidents. Secondly, existing alert systems issued alerts about detected anomalies without providing information about the violent occurrence places or the current whereabouts of suspects, hindering effective responses. Our proposed alerting system addresses this limitation, offering precise details about where violence happened and the current location of suspects. This capability empowers security staff to respond quickly and precisely, reducing the time it takes to address situations. The key distinction in present approaches lies in the transition from a system primarily focused on detection to one combining real-time person recognition and advanced alerting capabilities. Past methods successfully identified anomalies, but the lack of real-time person recognition limited intelligence during security incidents. The advanced alerting system bridges this gap by providing crucial information for a fast and targeted response, thereby improving the overall effectiveness of surveillance systems.

6. Conclusion

The incorporation of real-time person recognition enhances the surveillance system, ensuring the swift identification of individuals involved in incidents. The proposed alerting system, providing details on the event location and suspects' where-abouts, facilitates prompt response and intervention. These advancements bridge the gap between detection and action, offering a comprehensive approach to campus security. The shift from anomaly detection to live person recognition reflects addressing security challenges and adapting to evolving environments. This initiative not only aids in identifying violence but also establishes a platform for evolving intelligent surveillance systems. The integration of real-time person recognition and advanced alerting capabilities underscores our focus on creating a responsive and proactive security infrastructure for campus environments.

References

[1] Abbass, M. A. B., Kang, H. S. (2023). Violence Detection Enhancement by Involving Convolutional Block Attention Modules into Various Deep Learning Architectures: Comprehensive Case Study for UBI-Fights Dataset. IEEE Access.

[2] Aktı, S¸., Ofli, F., Imran, M., Ekenel, H. K. (2022). Fight detection from still images in the wild. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (pp. 550-559).

[3] Aldehim, G., Asiri, M. M., Aljebreen, M., Mohamed, A., Assiri, M., Ibrahim, S. S. (2023). Tuna Swarm Algorithm with Deep Learning Enabled Violence Detection in Smart Video Surveillance Systems. IEEE Access.

[4] Cao, C., Lu, Y., Wang, P., Zhang, Y. (2023). A New Comprehensive Benchmark for Semi-supervised Video Anomaly Detection and Anticipation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 20392-20401).

[5] Garcia-Cobo, G., SanMiguel, J. C. (2023). Human skeletons and change detection for efficient violence detection in surveillance videos. Computer Vision and Image Understanding, 233, 103739.

[6] Ghadekar, P., Agrawal, K., Bhosale, A., Gadi, T., Deore, D., Qazi, R. (2023). A Hybrid CRNN Model for Multi-Class Violence Detection in Text and Video. In ITM Web of Conferences (Vol. 53). EDP Sciences.

[7] Husz´ar, V. D., Adhikarla, V. K., N´egyesi, I., Krasznay, C. (2023). Toward Fast and Accurate Violence Detection for Automated Video Surveillance Applications. IEEE Access, 11, 18772-18793.

[8] Kang, M. S., Park, R. H., Park, H. M. (2021). Efficient spatio-temporal modeling methods for real-time violence recognition. IEEE Access, 9, 76270-76285.

[9] Mahareek, E. A., El-Sayed, E. K., El-Desouky, N. M., El-Dahshan, K. A. (2023). Detecting anomalies in security cameras with 3DCNN and ConvLSTM.

[10] Perseghin, E., Foresti, G. L. (2023). A Shallow System Prototype for Violent Action Detection in Italian Public Schools. Information, 14(4), 240.

[11] Singh, K., Preethi, K. Y., Sai, K. V., Modi, C. N. (2018, December). Designing an efficient framework for violence detection in sensitive areas using computer vision and machine learning techniques. In 2018 Tenth International Conference on Advanced Computing (ICoAC) (pp. 74-79). IEEE.

[12] Singh, V., Singh, S., Gupta, P. (2020). Real-time anomaly recognition through CCTV using neural networks. Procedia Computer Science, 173, 254-263.

[13] Ullah, F. U. M., Obaidat, M. S., Ullah, A., Muhammad, K., Hijji, M., Baik, S. W. (2023). A comprehensive review on vision-based violence detection in surveillance videos. ACM Computing Surveys, 55(10), 1-44.

[14] Vieira, J. C., Sartori, A., Stefenon, S. F., Perez, F. L., De Jesus, G. S., Leithardt, V. R. Q. (2022). Low-cost CNN for automatic violence recognition on embedded system. IEEE Access, 10, 25190-25202.

[15] Ye, L., Liu, T., Han, T., Ferdinando, H., Sepp¨anen, T., Alasaarela, E. (2021). Campus violence detection based on artificial intelligent interpretation of surveillance video sequences. Remote Sensing, 13(4), 628.