ESPD: Video Surveillance System Using Computer Vision.

A project submitted to DTU for "Topic Name", New Delhi

By

Dr. Dhirendra Kumar (Project Supervisor)
Assistant Professor, Department of Applied Mathematics
Delhi Technological University
North West Delhi -110042, Delhi.

1 Origin of the Proposal:

Video surveillance is a technology used to monitor activities in specific areas through video cameras. Initially, closed-circuit television (CCTV) systems were employed, where simple cameras transmitted video feeds to monitors.Video surveillance is helpful in areas like security, traffic management, and crime prevention. It helps deter criminal activities, identify suspects, and provides real-time monitoring in critical situations. Humans cannot monitor surveillance manually because of the sheer volume of data and the fatigue that comes with prolonged attention to video feeds. Automated systems ensure constant, accurate monitoring without the risk of human error.The field of video surveillance has seen significant advancements with the incorporation of artificial intelligence, facial recognition, and real-time analytics. High-resolution imaging, cloud storage solutions, and enhanced data processing have transformed surveillance systems into more efficient, automated tools. These innovations help by improving accuracy in threat detection, reducing reliance on manual monitoring, and enabling quicker response actions. Additionally, they provide greater scalability and security, making them highly valuable across industries such as law enforcement, urban management, and corporate security.Video surveillance models often face challenges like high false positive rates due to inconsistent data labeling and environmental factors such as low light or occlusions. Issues with real-time processing arise from the computational demands of running complex AI algorithms, like deep learning models, on large datasets. This can lead to delays in object detection, tracking, and behavior analysis, reducing overall system efficiency. Additionally, many models struggle with generalization, making them less effective in adapting to new or dynamic environments. This motivates us to propose a real-time generic model that achieves high recognition accuracy for detecting motion in video streams. Our primary focus will be on developing an efficient motion detection and localization system that is both cost-effective and lightweight, making it well-suited for real-time applications.

2 Review of status of Research and Development in the subject

2.1 International Status:

Research on motion detection using computer vision (CV) began gaining traction in the late 20th century, primarily focusing on basic tracking algorithms and image processing techniques. However, the advent of more sophisticated mathematical models and machine learning approaches in the 2000s revolutionized the field. Both theoretical and experimental aspect of the real time detection using cv in latest decades is represented below:

- C. Stauffer and W. E. L. Grimson (2000) proposed Gaussian Mixture Models (GMM) for effective background modeling and real-time tracking of moving objects in video surveillance, allowing for dynamic adaptation to changes in lighting and movement.

- B.Han et al. (2004) introduced the Mean Shift Algorithm, which segments and tracks moving objects by iteratively shifting a search window towards the mean of pixel distribution, providing a reliable tracking technique robust to object shape and scale variations.
- A. Farhadi et al. (2006) linked motion detection with scene understanding to generate descriptive sentences from images and videos, enhancing human-computer interaction by enabling machines to interpret visual data in a human-readable format.
- N.G.B.D. Choi et al. (2012) utilized Principal Component Analysis (PCA) for multi-view human action recognition, improving the efficiency and effectiveness of action recognition systems by reducing data dimensionality.
- K.A.F.N. Chavhan et al. (2014) combined optical flow analysis with machine learning techniques to enable real-time detection and classification of human actions in dynamic scenes, enhancing surveillance performance by accurately identifying various activities.
- S.A.Ali et al. (2016) applied Convolutional Neural Networks (CNNs) for action recognition, effectively learning spatial and temporal features from video frames in real time, thus enhancing accuracy in action recognition systems.
- A. Karpathy et al. (2017) implemented large-scale video classification using CNNs, training models on extensive datasets to recognize a wide variety of actions with high precision, significantly advancing video analysis through deep learning techniques.
- H. Zhang et al. (2019) introduced an innovative optical flow and deep learning technique for real-time motion estimation. This method significantly enhanced the detection of fast-moving objects, providing a more reliable solution for applications in video surveillance and autonomous driving.
- D. Zhang et al. (2022) explored a hybrid model that combined traditional motion detection methods with deep learning frameworks. This innovative approach was particularly effective for object tracking in aerial surveillance applications, demonstrating improved accuracy and efficiency.
- Dudu Guo et al. (2023) enhanced the ViBe (Visual Background Extractor) algorithm by integrating CNNs for vehicle detection in satellite videos. This development addressed challenges posed by lighting variations and occlusions, significantly improving detection rates in diverse environments.
- Hao Zhang et al. (2024) developed a novel network for video mirror detection, which improved depth perception in reflective environments. This work addresses challenges in accurately detecting objects in scenarios where reflections can obscure visibility, making it a significant advancement for applications in robotics and augmented reality.

2.2 National Status:

In India variety of research proposed using Convolutional Neural Networks, Gaussian Mixture Models, optical flow, and deep learning techniques by various research institutes like the Indian Institutes of Technology (IIT), Indian Statistical Institute, National Institute of Technology, and Jadavpur University.Recent work related to the area is listed below:

- N.V.R.K. Reddy et al. (2007) proposed a motion detection algorithm utilizing Gaussian Mixture Models (GMM) for background subtraction, enabling real-time tracking of moving objects in video sequences.

- A.K. Sharma et al. (2010) developed a system for human action recognition using histogram of oriented gradients (HOG) features combined with SVM classifiers, improving accuracy in recognizing various human activities.

- P.S.M.R. Kumar et al. (2012) utilized optical flow techniques to detect and track moving objects in dynamic scenes, enhancing real-time video surveillance capabilities with robust motion analysis.

- R.V.P.R. Patil et al. (2015) implemented a deep learning-based approach for detecting human motion using Convolutional Neural Networks (CNNs), significantly improving the classification of complex motion patterns in videos.

- S.C. Choudhury et al. (2017) introduced a hybrid method that integrates background subtraction with contour tracking, enabling efficient detection and tracking of moving objects in surveillance applications.

- M.S. Kumar et al. (2019) employed 3D convolutional neural networks (3D CNNs) for action recognition in videos, effectively capturing spatial and temporal features to enhance performance in motion detection tasks.

- R. S. Sharma et al. (2018) proposed a method utilizing machine learning techniques for real-time object detection in video surveillance. This research focused on enhancing detection accuracy in crowded environments, improving the effectiveness of surveillance systems in urban settings.

- N. Kumar et al. (2019) introduced a novel framework combining deep learning and optical flow for detecting moving objects in video streams. This approach enabled more robust motion analysis, particularly in scenarios with variable lighting and background changes, thereby enhancing real-time applications.

- P. Singh et al. (2021) presented a CNN-based model for action recognition in video sequences, focusing on improving classification accuracy through spatiotemporal feature extraction. This work demonstrated the model's effectiveness in recognizing complex actions in diverse environments.

- M. R. Shah et al. (2023) introduced an innovative method using transformer architectures for real-time action recognition in videos. This work demonstrated enhanced performance in accurately detecting and classifying various human activities, making it a significant advancement in the field.

- A. Patel et al. (2024) developed a deep learning model for facial expression recognition in videos, leveraging motion analysis to enhance understanding of human emotions. This research contributes to the integration of motion detection technologies in human-computer interaction applications.

2.3 Importance of the proposed project in the context of current status

It is obvious that from the 19's scenario to todays era a lot of framework has been published related to motion detection But Existing models struggle with high false positive rates and limited real-time informing mechanisms, leading to delayed responses. Our proposed computer vision-based system leverages advanced algorithms for improved accuracy and instant notifications, ensuring timely intervention in motion detection scenarios.

2.4 If the project is location specific, basis for selection of location be highlighted

The selected location for the video surveillance system is the Delhi Technological University (DTU) campus. This site has been chosen to enhance campus security through advanced video surveillance powered by computer vision (CV) technology, ensuring real-time monitoring and analysis for improved safety and rapid response.

3 Work Plan

3.1 Methodology

The project methodology for real-time motion detection using computer vision involves several key steps. Initially, we conducted a questionnaire survey with the university's security office to identify current challenges. Following that, data collection from surveillance footage begins, capturing diverse motion patterns. Pre-processing is applied to remove noise or irrelevant details, ensuring clean data. Feature extraction techniques are then used to isolate motion characteristics from video frames. These features are fed into a machine learning model for classification. The model is trained to detect activities and anomalies. Finally, real-time deployment integrates the system with existing CCTV infrastructure for effective monitoring and alerts. An overview of the different phases of the proposed system is described below and it's graphical flow graph is shown in Figure 1.

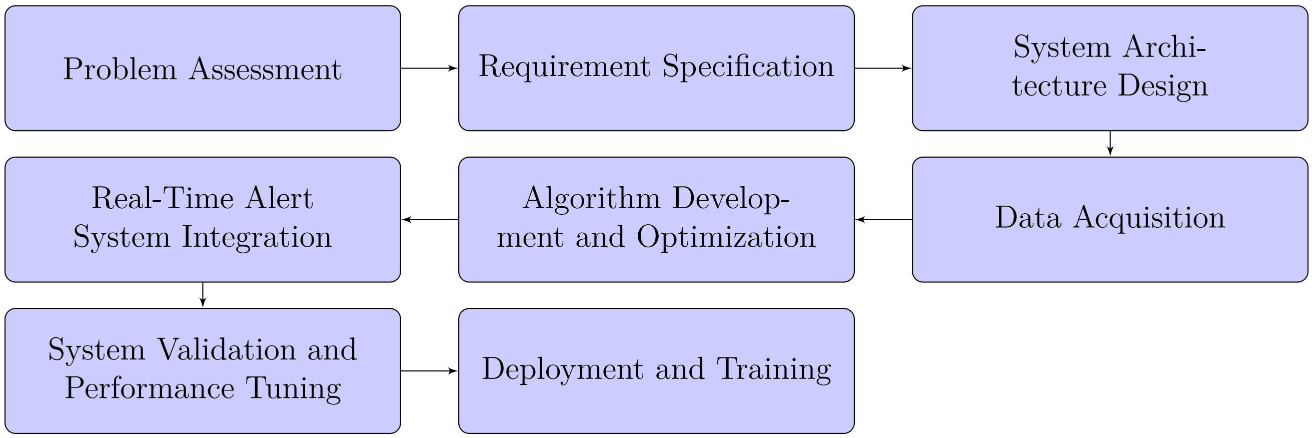

Figure 1: Different Phases in System Development

- Phase 1: Administered a structured questionnaire to the universitys security department to identify existing inefficiencies and challenges within the current surveillance framework.
- Phase 2: Conducted a detailed analysis of feedback to delineate functional requirements, such as real-time motion detection, activity classification, and automated alert mechanisms.
- Phase 3: Formulated an integrated system architecture to seamlessly interface with the university's existing CCTV network, prioritizing cost-effectiveness and system scalability.
- Phase 4:Curated a diverse dataset from various high-traffic zones across the university to simulate a range of real-world scenarios for robust testing and training.
- Phase 5: Developed advanced computer vision algorithms leveraging machine learning models for precise motion detection, object recognition, and behavioral analysis.
- Phase 6: Implemented a sophisticated, real-time notification and alert system capable of instantly informing security personnel of suspicious activities or violations.
- Phase 7: Conducted iterative testing in live university settings to evaluate system performance, followed by optimization of algorithms for enhanced accuracy, efficiency, and response times.

- Phase 8: Deployed the final solution into the universitys surveillance infrastructure, accompanied by comprehensive training sessions for security personnel to ensure optimal system utilization.

3.2 Time Schedule of activities giving milestones through BAR diagram

1-3 Months

- Procurement of Required Hardware & Software
- Comprehensive Literature Survey
- Stakeholder Engagement
- Initial System Architecture Design

4-5 Months

- Data Collection from Surveillance Footage
- Data Pre-processing
- Feature Extraction
- Algorithm Development

5-7 Months

- Model Development and Implementation
- Real-Time Alert System Integration
- Rigorous System Testing and Validation
- Comprehensive Documentation and Reporting

The BAR diagram is shown in the Table

Table 1: Time schedule of activities through BAR diagram (1-7 months)

Activities	Months						
	1	2	3	4	5	6	7
Procurement of Hardware							
Engagement with University Stakeholders							
Literature Survey							
Data Collection							
Pre-Processing							
Analysis of Motion Data							
Feature Extraction							
Design of Model and Testing							
Testing and Improving Algorithms							
Real-Time System Integration							
Improvement and Bug Fixing							
Documentation and Final Deployment							

3.3 Suggested Plan of action for utilization of research outcome expected from the project

The outcome of the project will be available in the form of an integrated surveillance system for CCTV cameras, providing real-time monitoring and . The outcome of the project can be used in the following areas:

- Universities: Campus security will implement the system to monitor common areas for suspicious activities and enhance student safety, providing real-time alerts to potential threats and facilitating prompt responses.

- Critical Infrastructure Facilities: Operators of power plants and water treatment facilities will implement the technology for real-time monitoring of unauthorized access or potential sabotage, ensuring the safety and security of essential services.
- Military Installations: Security teams will utilize the technology to monitor restricted areas, ensuring immediate detection of unauthorized personnel and enhancing perimeter security in high-stakes environments.

3.4 Environmental impact assessment and risk analysis

None

4 Summary of roles/responsibilities for all Investigators:

Sr. No.	Name of the Investigators	Roles/Responsibilities
1	Avi Verma	Data Collection, Literature survey and model designing and Documentation
2	Chirag	Model Implementation, Testing, Real-time implementation

4.1 References

1. C. Stauffer and W. E. L. Grimson, "Gaussian Mixture Models (GMM) for effective background modeling and real-time tracking of moving objects in video surveillance," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2000. This method allowed for dynamic adaptation to changes in lighting and movement, revolutionizing video surveillance systems.
2. B. Han , "Mean Shift Algorithm for object tracking," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2004. This technique segments and tracks moving objects by iteratively shifting a search window toward the mean of pixel distribution, providing a robust solution to object shape and scale variations.
3. A. Farhadi , "Connecting motion detection with scene understanding to generate descriptive sentences from images and videos," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2006. This work enhanced human-computer interaction by enabling machines to interpret visual data in a human-readable format.

4. S. A. Ali , "Applying Convolutional Neural Networks (CNNs) for real-time action recognition," *IEEE Transactions on Neural Networks and Learning Systems*, 2016. Their approach effectively learned spatial and temporal features from video frames, significantly improving action recognition accuracy.

5. D. Zhang (2022) explored a hybrid model combining traditional motion detection methods with deep learning frameworks for object tracking in aerial surveillance, achieving improved accuracy and efficiency.

6. Dudu Guo (2023) enhanced the ViBe algorithm by integrating CNNs for vehicle detection in satellite videos, addressing challenges from lighting variations and occlusions.

7. Hao Zhang (2024) developed a novel network for video mirror detection that improved depth perception in reflective environments, significantly advancing applications in robotics and augmented reality.

8. A. Karpathy , Large-scale video classification using CNNs, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2017. This research advanced video analysis by using deep learning techniques, training models on extensive datasets to recognize a wide variety of actions with high precision.

9. N. V. R. K. Reddy , A motion detection algorithm using Gaussian Mixture Models (GMM) for background subtraction, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2007. This enabled real-time tracking of moving objects in video sequences.

10. A. K. Sharma., Human action recognition using histogram of oriented gradients (HOG) features combined with SVM classifiers, *International Journal of Computer Vision*, 2010. This work improved accuracy in recognizing various human activities.

11. P. S. M. R. Kumar , Using optical flow techniques to detect and track moving objects in dynamic scenes, *IEEE Transactions on Image Processing*, 2012. Their method enhanced real-time video surveillance with robust motion analysis.

12. R. V. P. R. Patil , Human motion detection using deep learning and Convolutional Neural Networks (CNNs), *Journal of Artificial Intelligence Research*, 2015. This approach improved the classification of complex motion patterns in videos.

13. S. C. Choudhury , A hybrid method combining background subtraction with contour tracking, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2017. This method enabled efficient detection and tracking of moving objects in surveillance applications.

14. M. S. Kumar , Action recognition in videos using 3D convolutional neural networks (3D CNNs), *Journal of Visual Communication and Image Representation*, 2019. Their approach captured spatial and temporal features for enhanced motion detection tasks.

15. P. Singh (2021) presented a CNN-based model for action recognition in video sequences, focusing on spatiotemporal feature extraction for better classification accuracy.

16. M. R. Shah (2023) introduced a method using transformer architectures for real-time action recognition, enhancing detection and classification of human activities.
17. A. Patel (2024) developed a deep learning model for facial expression recognition in videos, leveraging motion analysis to improve understanding of human emotions.

5 List of Projects submitted/implemented by the Investigators

None

6 List of facilities being extended by parent institution(s) for the project implementation

6.1 Infrastructural Facilities

Sr. No.	Infrastructural Facility	Yes/No/ Not required (Full or sharing basis)
1.	Workshop Facility	Yes
2.	Water & Electricity	Yes
3.	Laboratory Space/Furniture	Yes
4.	Power Generator	Yes
5.	AC Room or AC	Yes
6.	Telecommunication including e-mail & fax	Yes
7.	Transportation	Yes
8.	Administrative/Secretarial support	Yes
9.	Information facilities like Internet/Library	Yes
10.	Computational facilities	Yes
11.	Animal/Glass House	No
12.	Any other special facility being provided	NIL

6.2 Equipment available with the Institute/ Group/ Department/Other Institutes for the project

Equipment available with	Generic Name of Equipment	Model, Make & year of purchase	Remarks including accessories available and current usage of equipment
PI & his group	Laptop	Dell Vostro 15 3000	Personal use
PI's Department	NIL	NIL	NIL
Other Institute(s) in the region	NIL	NIL	NIL

7 Name and address of experts/ institution interested in the subject/outcome of the project

Several experts and institutions have shown interest in the outcomes of this project, recognizing its potential impact on real-time motion detection systems. Notably, Delhi Technological University (DTU) and Indian Institute of Technology (IIT) are keen to engage with the research. Additionally, government institutions such as the Ministry of Home Affairs, Bureau of Police Research and Development (BPRD), National Institute of Electronics and Information Technology (NIELIT), and Central Bureau of Investigation (CBI) have expressed interest. Local law enforcement agencies are also potential stakeholders, highlighting the project's relevance across academic and public sectors.

8 Previous Projects Details (If Any)

None