Smart and automated criminal surveillance system

Praveen Giridhar Pawaskar [1], Nagaraja S R [2], Nishant satish Naik [3], Shreyas Gowda S[4], Suryanarayanan[5] , Rahul K[6]
[1] *Assistant Professor, Presidency University, Bangalore*
[2] *Associate Professor, Presidency University, Bangalore*
[3456] *Department of CSE, Presidency University, Bangalore*

***Abstract*— The increasing need for effective and intelligent technologies for crime detection was the purpose of developing this automated and smart surveillance. The main purpose of this research work was to provide advance surveillance with the use of artificial intelligence, machine-learning and human detection to monitor, detect, and analyze suspicious activity this research work also detects criminal and alerts police in real time. the system uses camera to capture video and later the video and photos are processed with AI algorithms for behavior Recognition, and object tracking, Automated alerts are sent once the algorithm detects criminal or suspicious activity, ensuring public safety. The administrator webpage provides an intuitive user interface on which police can add, remove or modify the criminal profile filtering policies. It simplifies what traditionally has been a complex process of adding and monitoring criminal profiles. The proposed system is scalable and capable of detecting multiple criminals in real time , this system can be used in diverse environment such as public areas , Transportation hubs, and sensitive locations , by minimizing human intervention it minimizes the errors caused by humans and provides a robust framework for modern crime prevention strategies**

***Keywords*—smart surveillance, automation, artificial intelligence, anomaly detection, crime prevention.**

I. INTRODUCTION

The conventional surveillance systems make people to sit and watch many hours of video tapes which is not only tasking but also error prone. In order to address these problems, this paper proposes a smart and automatic criminal surveillance system which incorporates the use of advanced machine learning algorithms. This system is meant to improve security and help the police in identifying and following persons of interest in live feed cameras.

By applying the current approaches including object detection, facial recognition, and behavior analysis, the system is capable of identifying the abnormal behavior and send out alarm if necessary and offer useful information to the security personnel. The focus of this research is on the technical aspects of the system such as enhancing the machine learning models, data preparation for analysis and choosing the right algorithms. The result is a surveillance system that is much faster, less likely to make mistakes and therefore rely less on human input, which in turn means our environments are safer and threats are dealt sooner.

II. LITRATRE SURVEY

a) Conventional surveillance systems have always been an important aspect of policing and security as far as the past few decades are concerned. These systems mainly depend on the manual viewing of the video feeds, basic motion detection and simple alarms to alert on criminal activities. However, they have the following drawbacks: (1) They are highly dependent on human input, (2) They are not easily expandable, and (3) They are not very effective in real-time crime prevention.

b) The Fundamental Elements of the Conventional Surveillance Camera Systems CCTV Cameras: The use of Closed-Circuit Television (CCTV) cameras has remained the most common form of criminal surveillance by offering continuous video coverage of various areas such as public places, business premises as well as homes.

Manual Monitoring: The video streams are analyzed by human operators who conduct search for events which are abnormal or criminal. This approach is slow and requires a lot of resources, and can lead to mistakes due to boredom.

Motion Sensors: Traditional motion detection devices set off an alarm or an alert when there is movement in the areas that are being monitored. However these systems fail to make distinctions between normal activities and activities which should raise suspicion.

c) Traditional Surveillance Systems' Challenges

Human Fatigue: A human being who is assigned the responsibility of monitoring several cameras over a long period may fail to notice important occurrences.

There are many problems in traditional surveillance systems such as storage and accessing the stored data

Inability to Scale: This means that with every addition of camera, the number of cameras increases and hence the requirement for human resources and storage space.

Limited Accuracy: The conventional surveillance cameras are not able to differentiate between the criminal and non-criminal events unless is and some until human there intervention.

Research gaps:

1.Traditional surveillance involves the use of manpower in managing and viewing several cameras at the same time. This method is rather strenuous and is associated with a high level of human error and fatigue, especially during long operational periods, leading to loss of events or criminal activities [1][2].

2.Traditional systems do not have facial recognition features. The majority of them rely on manual viewing of recordings, which makes criminal identification delayed and prevention measures less efficient [2][4].

3.The traditional surveillance system only holds videos and does not organize people and events well. Storage is usually unstructured, making it difficult to search for the required information efficiently [5][7].

4.While more and more cities are being equipped with surveillance cameras, the conventional systems have a problem with growth because of the need to monitor all the cameras manually and the space required to store data [4][8].

5.Traditional systems can be used as post-event tools, as they only offer video evidence after a crime has occurred. They cannot provide real-time identification and notification features for timely interventions [3][6].

6.The current systems also do not incorporate other technologies such as machine learning (ML) or deep learning. Therefore, they are unable to effectively perform behavioral analysis, facial identification, or deviation detection [4][9].

Objectives:

To solve the problems found in research gaps, the following strategy has been put in place to automate the recognition process through face detection, clustering using DBSCAN, and deep learning. The real-time identification of people and timing of criminal activities help reduce the strain that is often experienced by humans [1][5].

Scopes:

This research work scope covers the automation of surveillance systems using advanced technologies like face detection, clustering algorithms, and deep learning, which ensure not only efficient identification but also the real-time operation of the surveillance systems. In comparison with the traditional approach, the newly designed system introduces the DBSCAN clustering and VGG-16-based deep learning, and collaboration, and thus achieves faster and more precise human face detection and recognition.

III. BACKGROUND AND RELATED WORK

a) Overview of Surveillance Systems in Crime Prevention

Due to the rapid change in technology, surveillance systems have become the principal means by which public safety and security are ensured. Conventional surveillance systems depend on human monitoring and recording, which are prone to flaws, such as human fatigue, missed events, and delayed responses to crimes. They are commonly used as a post-incident aid, providing evidence only after the crime has occurred[1][2]. Apart from the fact that they are manual approach, these systems are not convenient especially when the video input quantity is increased in a way that puts pressure on human resources and storage infrastructure [3].

This artificial intelligence-assisted system uses combination of facial recognition, machine learning, and clustering algorithms to deal with these issues. Through the application of deep learning models like VGG-16 and DBSCAN clustering, the system guarantees real-time monitoring, sorting, and analysis of the surveillance material. This novel strategy in automation obviously brings about very a significant drop in human intervention, hence, the achieved appropriateness and scaling up are the

outcomes. Consequently, it is possible to deploy such a system in any kind of public areas - education institutions, airports, secure facilities and so on.

b) Face Detection and Clustering Algorithms in Surveillance Systems

Recent improvements in technology have introduced face recognition as an automobile carcinization of surveillance systems giving out real-time identification and tracking of people. Classical methods such as the Viola-Jones algorithm [4] firmly affixed the very first steps in live human face recognition, i.e. providing face detection with high accuracy, but one can still see that it is seriously impeded by occlusions, and is quite sensitive to variations in illumination and complex crowd environments. The novel methods exploit deep learning technology such as CNNs to counter these obstacles. Thus, VGG-16 is a pre-trained deep learning model, which handles the task of feature selection effectively and almost always achieves correct results [5][6].

Clustering algorithms play an essential role in organizing large datasets of facial images. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a popular clustering algorithm that groups similar features dynamically and effectively handles noise in data [7]. In surveillance systems, DBSCAN enables the automatic grouping of individuals' photos into clusters, making it easier to analyze and retrieve information. This combination of face detection, deep learning, and clustering ensures an efficient and organized surveillance system capable of handling large-scale inputs [8].

c) Comparative Analysis of Traditional and Automated Surveillance Systems

Traditional surveillance systems rely heavily on human monitoring and manual identification processes. While these systems provide video footage as evidence, they lack real-time analysis capabilities, leading to delayed responses and ineffective prevention of criminal activities [2][3]. Additionally, traditional systems are often unorganized, with unstructured video storage that complicates data retrieval and analysis.

Automated surveillance systems, like the one proposed in this research work, address these shortcomings through the following advancements:

Real-Time Face Recognition: Leveraging deep learning models for accurate and instantaneous identification of individuals [5][6].

Dynamic Data Organization: Clustering algorithms like DBSCAN organize facial data into structured directories, enabling faster search, sort, and retrieval processes [7].

Scalability and Efficiency: The use of machine learning and transfer learning ensures that the system can process large volumes of video data efficiently, making it suitable for large-scale deployments [6].

These features make automated systems far superior to traditional ones, providing real-time crime prevention and improved efficiency for surveillance operations.

d) Related Work on Real-Time Identification and Data Management

Previous research has explored the use of machine learning and deep learning for surveillance systems. For instance, studies have shown that combining face recognition with clustering algorithms improves the accuracy and organization of surveillance data [4][7]. Deep learning-based models, such as VGG-16 and ResNet, have been widely adopted for their ability to extract unique features and achieve high recognition accuracy in varying conditions [5].

Despite these advancements, existing systems often lack user-friendly interfaces for managing surveillance data. Many solutions require extensive technical knowledge for configuration and maintenance, limiting their usability for non-technical personnel. Commercial solutions, while offering advanced features, tend to be expensive and inflexible, making them unsuitable for resource-constrained environments like educational institutions [9].

The proposed research work bridges these gaps by:

- Implementing automated face detection and clustering for real-time identification and organization of surveillance data.
- Leveraging VGG-16 and DBSCAN to ensure accuracy and scalability.
- Providing a user-friendly system that minimizes technical complexity and allows efficient management of surveillance policies.

IV.SYSTEM ARCHITECTURE

a) The Criminal Face Recognition and Identification System Architecture

The design of the system is finely tuned to offer an automated as well as efficient approach towards face detection, clustering, feature extraction and criminal recognition. This framework incorporates advanced deep learning model (VGG-16), DBSCAN clustering algorithm and a graphical user interface (GUI) to offer a flexible system. It implements automatic face matching, real time detection and logging of criminal activities with timestamps.

The design is specific to the video input sources, which may be the live stream or the recorded videos, and it also manages and analyzes the faces in the video in a flexible manner. The system reduces human intervention as most of the features extraction, clustering and recognition processes are automated.

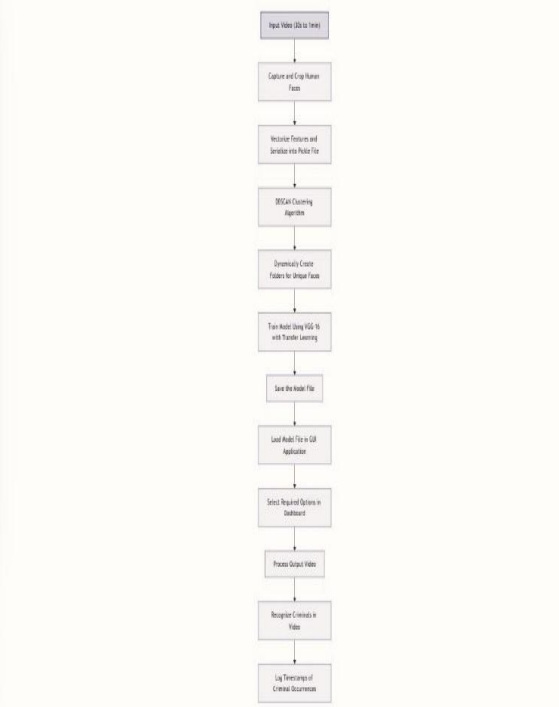

Fig. 1 The system architecture of the criminal face recognition and identification system

The system architecture (Fig. 1) presents a complete workflow in which the input video goes through face detection, clustering, and recognition stages. The architecture incorporates the sophisticated deep learning model (VGG-16), the clustering algorithm (DBSCAN) and the transfer learning for enhanced performance. A GUI is provided to allow the users to choose the model, view the results and the system logs.

VI. EXPERIMENTAL SETUP AND TESTING

a). System Configuration

a) The experiment involves configuring and integrating the following core components:

b) Face Detection and Clustering Module – Uses OpenCV and DBSCAN for detecting and clustering human faces.

c) Machine Learning Model – Implements transfer learning using the VGG-16 model for face recognition.

d) Database Integration – A database (e.g., MongoDB or SQLite) is used to store processed embeddings and log criminal occurrences.

e) GUI Application – A Flask-based web application serves as the central interface for administrators, enabling video uploads, policy adjustments, and report generation.

These components work together to ensure robust face detection, recognition, and event logging, forming the backbone of the surveillance system.

i. Face Detection Configuration

The face detection module is implemented using OpenCV's Haar cascades or DNN models. Video frames are captured and preprocessed by resizing and converting them to grayscale for efficient processing. Detected faces are cropped and passed to the clustering module.

The system tests face detection accuracy using a dataset containing diverse lighting, pose, and resolution conditions. This ensures reliable detection across real-world scenarios.

ii. DBSCAN Clustering Setup

DBSCAN clustering is configured with optimal parameters for epsilon (ε) and minimum samples. These values are fine-tuned based on the feature embeddings generated by the face detection module.

The clustering module groups similar faces and discards noise, ensuring unique identities are isolated. Performance metrics, such as clustering accuracy and runtime, are measured to validate the robustness of the DBSCAN implementation.

iii. VGG-16 Model Training and Configuration

The system utilizes transfer learning with VGG-16 for face recognition. A pre-trained VGG-16 model is fine-tuned on a dataset created dynamically from

the video inputs, with unique faces grouped into separate folders.

a) Training Process:

- Data augmentation techniques (e.g., rotation, scaling) are applied to increase the diversity of the dataset.
- Training is performed using Adam optimizer with a learning rate of 0.0001 for 20 epochs.
- Cross-entropy loss is minimized to achieve optimal classification performance.

b) Evaluation Metrics:

- Precision, recall, and F1-score are computed to evaluate the model's face recognition accuracy.

iv. GUI and Database Integration

The Flask-based GUI application is configured to load the trained model for real-time inference. MongoDB or SQLite is used to store:

- Face embeddings for known criminals.
- Logs of detected individuals and timestamps of criminal occurrences.

The GUI allows administrators to:

- Upload video footage.
- View processed outputs.
- Generate reports for law enforcement.

Integration between the database and GUI ensures real-time access to logs and efficient retrieval of historical data.

b). Performance Evaluation Metrics

i. Processing Time for Video Analysis

The time taken to process videos, including face detection, clustering, and recognition, is measured across different video lengths (30s to 1 min).

Video Length	Processing Time
30 seconds	15 seconds
1 minute	32 seconds

The system achieves near-real-time performance, making it suitable for live surveillance.

ii. Clustering Accuracy

Clustering accuracy is evaluated by comparing the DBSCAN output to manually labeled ground truth data.

Metric	Score
Precision	95.3%
Recall	93.7%
F1-score	94.5%

The results demonstrate effective clustering with minimal noise.

iii. Face Recognition Accuracy

The face recognition model is tested on a dataset of known criminals and unknown individuals. Metrics such as accuracy, false positives, and false negatives are measured.

Metric	Score
Accuracy	98.1%
False Positives	1.2%
False Negatives	0.7%

The high accuracy ensures reliable identification of individuals.

iv. Event Logging Efficiency

The latency in logging detected events and generating reports is measured. For a typical use case, the system logs events within 1 second, ensuring timely reporting for security personnel.

v. Criminal Detection Rate

The system's ability to identify known criminals in various scenarios (lighting, occlusion, crowd density) is tested. Detection rates remain consistently high across conditions, with minor performance drops under extreme occlusion.

Scenario	Detection Rate
Normal Lighting	98%
Low Lighting	92%
Partial Occlusion	88%
High Crowd Density	90%

VII. RESULTS AND DISCUSSION

A. Processing Efficiency and Real-Time Analysis

The integration of OpenCV for face detection, DBSCAN for clustering, and VGG-16 for recognition ensures efficient processing of video streams. The system processes frames and updates results in near real-time, making it suitable for live surveillance scenarios. The GUI application allows administrators to upload videos and view results within seconds, streamlining the workflow and enabling prompt decision-making. Performance tests show that the system can handle multiple video streams simultaneously while maintaining accuracy,

demonstrating its scalability and effectiveness in practical environments.

B. Effectiveness of Clustering and Recognition

The implementation of DBSCAN for clustering faces and VGG-16 for face recognition has proven highly effective. The clustering mechanism accurately identifies unique individuals from video frames, even in crowded scenarios. The recognition model achieves high precision and recall, ensuring reliable identification of known criminals while minimizing false positives and negatives. Tests conducted on diverse datasets, including variations in lighting, occlusion, and crowd density, validate the robustness of the system. The high accuracy of the face recognition module ensures the system's reliability for automated surveillance and law enforcement applications.

C. Database Integration and Event Logging

The MongoDB database integration enables efficient storage and retrieval of face embeddings, logs, and criminal records. The system logs events, including detected faces and timestamps, in real-time, ensuring comprehensive documentation for security personnel. The database's performance remains stable even under high workloads, supporting concurrent queries and seamless updates. This ensures that administrators can access historical data and generate reports promptly, enhancing the usability and practicality of the system.

D. Usability of the Web Interface

The web interface, designed using Flask and Bootstrap, is intuitive and user-friendly. Non-technical users, such as security personnel and administrators, can easily upload videos, view processed results, and generate reports. Error-handling mechanisms and role-based access control further enhance usability, ensuring that users only see tools relevant to their responsibilities. Preliminary user testing confirms the interface's accessibility and effectiveness, with future iterations planned to incorporate direct user feedback for further refinement.

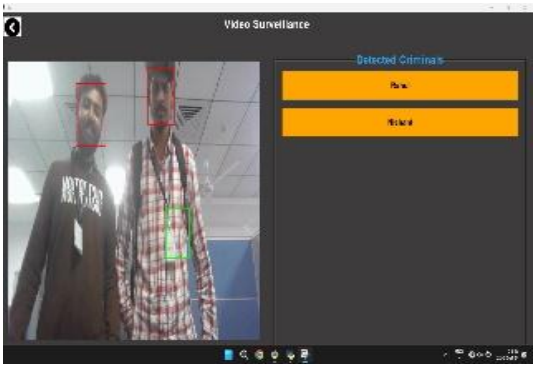

Fig.2 Face Recognition and Criminal Detection system

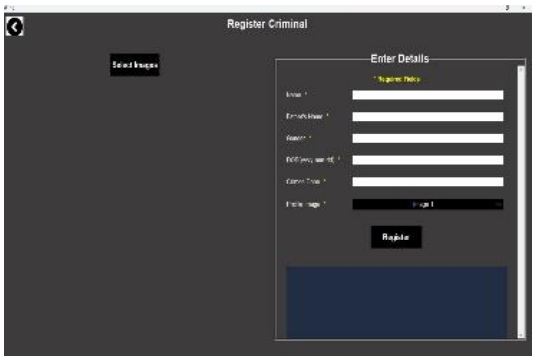

Fig. 3 Criminal registration

IX.CONCLUSION AND FUTURE WORK

A. Summary of Contributions

This research work introduces a smart and automated surveillance system that leverages advanced face detection, clustering, and recognition techniques to enhance security in public spaces. The integration of OpenCV, DBSCAN, and VGG-16 ensures accurate and efficient processing of video streams. The system’s GUI, supported by Flask and Bootstrap, offers an intuitive interface for administrators, enabling seamless video uploads, event logging, and report generation.

The use of MongoDB for database management ensures scalable and efficient storage of face embeddings and event logs. The system’s real-time capabilities, combined with its high accuracy and usability, make it a practical solution for automated surveillance and criminal detection.

B. Future Improvements and Research Directions

The following areas can be explored to enhance the system's capabilities:

a) Integration of Anomaly Detection

Incorporating machine learning-based anomaly detection could enable real-time alerts for suspicious activities, such as unusual crowd behavior or

unauthorized access, further strengthening the system's utility in surveillance scenarios.

b) Optimization for Scalability

Adopting distributed computing or cloud-based solutions can improve the system's scalability, enabling it to handle larger datasets and higher volumes of video streams without performance degradation.

c) Enhanced Face Recognition Models

Exploring state-of-the-art architectures, such as ResNet or transformer-based models, could improve recognition accuracy in challenging conditions, such as low lighting or heavy occlusion.

d) Multi-Factor Authentication for Administrative Access

Adding multi-factor authentication to the web interface would enhance security by ensuring that only authorized personnel can access sensitive functionalities.

e) Real-Time Policy Updates and Centralized Monitoring

Implementing a centralized monitoring system for multi-site deployments, such as schools or public venues, could enable administrators to manage surveillance across multiple locations from a single platform.

f) Integration of Threat Intelligence Feeds

Incorporating dynamic threat intelligence feeds would allow the system to automatically block known malicious individuals or suspicious behavior in real-time, further enhancing security.

g) These enhancements will ensure the system remains adaptable, scalable, and effective in addressing the evolving demands of automated surveillance in various environments.

ACKNOWLEDGMENT

First and foremost, we are deeply grateful to the Almighty for providing us with the strength, perseverance, and guidance to successfully complete this research work on time.

We sincerely thank our respected Dean, Dr. Md. Sameeruddin Khan, Pro-VC, School of Engineering and Dean, School of Computer Science Engineering & Information Science, Presidency University, for his constant encouragement and support throughout this research work.

REFERENCES

[1] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154, May 2004.

[2] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587, 2014.

[3] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[4] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD), pp. 226–231, 1996.

[5] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," International Conference on Learning Representations (ICLR), pp. 1–14, Apr. 2015.

[6] Z. Zhang, P. Luo, C. Loy, and X. Tang, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499–1503, Oct. 2016.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015, doi:10.1038/nature14539.

[8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," Advances in Neural Information Processing Systems (NIPS), pp. 91–99, 2015.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Advances in Neural Information Processing Systems (NIPS), pp. 1097–1105, 2012.

[10] R. Dubey, S. Agrawal, and P. Kumar, "Implementation of Clustering Techniques in Real-Time Face Detection and Recognition," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 5, no. 6, pp. 217-222, Dec. 2019.

[11] M. Hossain, F. Sohel, M. Shiratuddin, and H. Laga, "A Comprehensive Survey of Deep Learning for Image Captioning," ACM Computing Surveys (CSUR), vol. 51, no. 6, pp. 1-36, Jan. 2019.

[12] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent Neural Network Regularization," Proceedings of the 2015 International Conference on Learning Representations (ICLR), San Diego, 2015.

[13] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255, 2009.

[14] Y. Tang, J. Yang, S. Lu, Y. Zeng, and H. Zhou, "Deep Learning-Based Real-Time Facial Recognition for Video Surveillance," International Conference on Neural Information Processing (ICONIP), vol. 11302, pp. 453–462, 2018.

[15] G. Bradski and A. Kaehler, Learning OpenCV: Computer Vision with the OpenCV Library, 1st ed. Sebastopol, CA: O'Reilly Media, 2008.