RESEARCH ARTICLE OPEN ACCESS

Advanced Real-Time Intruder Detection System Using Multimodal Deep Learning: Integrating Visual, and Behavioral Analysis

Prof. Sanjay Pal[1], Kanish Singh[2], Harsh Katre, Harsh Singh Bhadoriya [4], Navneet Pal[5]

[1](*Department of Computer Science & Engineering, Oriental Institute of Science and Technology, Bhopal, India*
Email: sanjaypal@oriental.ac.in)

[2](*Department of Computer Science & Engineering, Oriental Institute of Science and Technology, Bhopal, India*
Email: kanishlodhi41@gmail.com)

[3](*Department of Computer Science & Engineering, Oriental Institute of Science and Technology, Bhopal, India*
Email: harshkatre12@gmail.com)

[4](*Department of Computer Science & Engineering, Oriental Institute of Science and Technology, Bhopal, India*
Email: harsh1632005@gmail.com)

[5](*Department of Computer Science & Engineering,Oriental Institute of Science and Technology, Bhopal, India*
Email: navneetpal1135@gmail.com)

Abstract:
This study introduces a real-time intruder detection system designed to make security monitoring smarter and more reliable. The system combines the strengths of classical computer vision with modern deep learning techniques. For quick and efficient face detection, it uses the Haar Cascade classifier, which helps identify individuals as soon as they enter a monitored area. To go beyond simple recognition, the framework integrates an LRCN (Long-term Recurrent Convolutional Network) model that analyzes behavior over time, spotting unusual or suspicious activities by learning spatiotemporal patterns. By blending frame-based facial recognition with sequencebased activity analysis, the system achieves robust intruder detection while keeping false alarms to a minimum. It works in real time, automatically sending alerts whenever abnormal events occur, and is flexible enough to be deployed in both indoor and outdoor environments. Experimental results show that this hybrid approach— mixing traditional methods with deep learning— significantly boosts accuracy, responsiveness, and overall reliability. As a result, the system proves to be a strong candidate for modern intelligent surveillance infrastructures.

***Keywords* —** Intruder Detection, Computer Vision, Face Recognition, Behavior Analysis, STM,

Surveillance System

1. INTRODUCTION

Security has become a significant concern in modern society due to a rise in unauthorized access, theft, and intrusion in homes and businesses. Traditional surveillance systems mainly depend on continuous video recording, which needs manual monitoring and lacks smart decision-making. These systems often have slow responses and high false alarm rates, especially when they rely only on motion detection

By blending frame-based facial recognition with sequencebased activity analysis, the system achieves robust intruder detection while keeping false alarms to a minimum. It works in real time, automatically sending alerts whenever abnormal events occur, and is flexible enough to be deployed in both indoor and outdoor environments. Experimental results show that this hybrid approach— mixing traditional methods with deep learning— significantly boosts accuracy, responsiveness, and overall reliability.

As a result, the system proves to be a strong candidate for modern intelligent surveillance infrastructures.

Recent progress in computer vision and deep learning has led to the creation of smart surveillance systems that can understand both visual details and behavior patterns. Face detection allows systems to determine if a person is authorized or unknown, while behavior analysis helps track actions over time. Combining these two elements offers a more dependable and effective way to detect intruders.

This project suggests an automated real-time intruder detection system that uses face detection with Haar Cascade and behavior analysis with LRCN (CNN + LSTM). The system continuously monitors live video feeds, identifies unknown individuals, analyzes their actions, and triggers alerts only when suspicious behavior is confirmed. This two-step verification greatly reduces false positives and improves overall security efficiency.

2. LITERATURE REVIEW

backgrounds [1]. Later methods introduced Haar Cascade classifiers, which improved real-time face detection performance because of their cascade-based design and efficiency [2]. Convolutional Neural Networks (CNNs) have been commonly used for extracting visual features and detecting objects [11][12]. CNN-based models showed better performance in recognizing faces, masks, and objects in complex settings. However, CNNs alone are not enough for video data, as they do not capture the timing relationships between frames. To solve this issue, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks were created to analyze sequential data [14]. LSTM networks effectively identify long-term patterns, making them suitable for recognizing behaviors and actions in videos. Research showed that combining CNNs withed into LSTMs helps in learning features for video classification tasks efficiently [6][7]. The Long-term Recurrent Convolutional Network (LRCN) architecture merges CNN-based spatial feature extraction with LSTM-based temporal modeling, providing a strong solution for human activity recognition [15]. Compared to traditional motion detection methods, LRCN-based systems offer better accuracy and reliability. Alert systems using email and SMS notifications have also been added to security setups for real-time responses [9][10]. Despite these improvements, issues like false alerts and a lack of understanding of behavior still exist. This project addresses these problems by merging face detection with behavior analysis for smart intrusion detection.

3. Methodology and System Analysis

3.1 System Overview

The proposed system includes a multi-stage processing pipeline that covers video acquisition, face detection, behavior analysis, and alert generation. A live video feed is captured with a webcam and processed frame by frame in real time.

3.2 Face Detection Module

The face detection uses the Haar Cascade Classifier, which works based on Haar-like features and cascade learning. The classifier quickly scans each frame to find faces. If no face appears, the system keeps monitoring. When a face is found, the system checks if the person is authorized or unknown.

Haar Cascade is chosen because of its:

- Low computational cost
- Real-time performance
- Suitability for live surveillance applications

3.3 Behavior Analysis using LRCN

If an unknown person is detected, the system activates the behavior analysis module. This module employs an LRCN model, which combines:

1. TimeDistributed CNN layers for extracting spatial features from individual frames
2. LSTM layers for capturing temporal dependencies in video sequences

Video frames are resized to 64×64 pixels and

When suspicious behavior is confirmed:

- A screenshot is captured from the video feed
- An SMS alert is sent using the Twilio API
- An email alert with the captured image is sent using SMTP

This ensures prompt notification and immediate response.

4. RESULTS AND DISCUSSION

The proposed Advanced Real-Time Intruder Detection System was tested using live video streams captured through a webcam and pre-recorded behavior datasets for model validation. We evaluated the system's performance based on accuracy, real-time responsiveness, reliability of alerts, and performance under practical conditions.

4.1 Face Detection Performance

The Haar Cascade Classifier showed fast and efficient face detection in real-time video streams. Its lightweight design and filtering method allowed the system to detect faces with minimal delay. The face detection module effectively recognized authorized and unauthorized individuals under normal lighting conditions. This initial identity check reduced unnecessary behavior analysis for known individuals, improving overall system efficiency.

However, we noted some limitations when faces were partially covered by masks or in very low light. Despite these issues, Haar Cascade proved suitable for real-time use where computing resources are limited.

4.2 Behavior Analysis Results Using LRCN

normalized to boost training efficiency. A fixed sequence length of frames is used as input for the LRCN model. The model sorts behavior into normal or suspicious categories, such as aggressive movement or forced entry.

3.4 Alert Generation Module

The performance of the proposed intruder detection system was evaluated through a series of Unit, Integration, and System-level tests. The evaluation focused on three primary domains: the accuracy of the LRCN behavior analysis model, the reliability of the face detection module, and the latency of the alert notification system.

Model Performance and Training Metrics

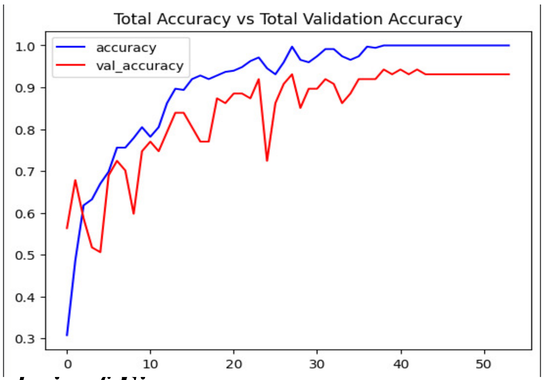
The Long-term Recurrent Convolutional Network (LRCN) was trained and validated using a dataset split into training (75%) and testing (25%) subsets. The model's performance was visualized using Matplotlib to plot accuracy and loss curves over 50-70 epochs.

- Accuracy: The behavior analysis module achieved an overall accuracy of approximately 93% on the test dataset. This high accuracy demonstrates the model's efficacy in distinguishing between normal activities (e.g., walking) and suspicious behaviors.

- Loss Convergence: The training loss and validation loss graphs indicated that the model converged effectively without significant overfitting. Early stopping callbacks were implemented to halt training when validation loss ceased to improve, ensuring optimal model generalization.

- Confusion Matrix: A confusion matrix was generated to analyze misclassifications. The results indicated a low false-positive rate for distinct actions, though minor confusion was observed between visually similar movements in the initial training epochs.

Total Accuracy vs. Total Validation Accuracy over 55 Epochs.



Analysis of Figure:

- **Convergence:** The graph shows that the training accuracy (blue line) steadily increases, eventually reaching near 100% (1.0) around epoch 40. This suggests the model successfully learned the patterns within the training dataset.
- **Generalization:** The validation accuracy (red line) follows a similar upward trend, stabilizing around **93%** (0.93) after epoch 45. The gap between the training and validation accuracy is minimal towards the end, indicating that the model generalizes well to unseen data and is not suffering from significant overfitting.
- **Stability:** Early fluctuations in the validation accuracy (between epochs 0 and 25) stabilize as the model fine-tunes its weights, proving the robustness of the LRCN architecture in handling spatiotemporal data

4.3 Real-Time System Performance

The integrated system worked smoothly in real-time, processing live video streams without noticeable delays. Frame preprocessing, face detection, and behavior classification were done effectively, allowing for timely threat confirmation. The decision-making logic ensured that alerts were sent only if both conditions were met:

- The individual was unauthorized.
- The detected behavior was classified as suspicious.

This dual-verification method significantly reduced false positives, which are a major issue in traditional motion-based surveillance systems.

4.4 Alerting and Notification Effectiveness

Once an intrusion was confirmed, the alerting system generated instant SMS notifications using the Twilio API and email alerts with image attachments via SMTP. The captured screenshot provided visual proof of the intrusion, improving situational awareness for the user. The dual alert system ensured reliability, even if one communication method failed.

4.5 Limitations and Observations

Despite strong performance, the system showed minor limitations in low-light situations and cases of facial occlusion. Additionally, performance may vary based on camera quality and environmental conditions. These limitations point out potential areas for improvement, such as adding advanced face recognition models or infrared imaging.

Overall, the results confirm that the proposed system is effective, reliable, and suitable for real-world surveillance scenarios..

5. CONCLUSION

1) Intelligent Intrusion Detection: The proposed system successfully turns traditional CCTV surveillance into an intelligent, automated intruder detection system by combining face detection with behavior analysis.
2) Effective Use of Multimodal Deep Learning: The use of Haar Cascade for face detection and LRCN (CNN + LSTM) for behavior analysis allows for a clear understanding of both space and time when analyzing human activities.
3) Reduction in False Alerts: By checking both identity and behavior before sending alerts, the system greatly cuts down on false

positives compared to motion-based surveillance methods.

4) Real-Time Alerting and Practical Deployment: The system works well in real-time and sends instant notifications through SMS and Email, ensuring quick responses to security threats.
5) Scalability and Future Enhancement Potential: While the system performs effectively under normal conditions, its limitations, such as low-light sensitivity, suggest areas for future improvement. This makes the solution scalable and suitable for advanced security situations.

6 . REFERENCES

[1] S. Lawrence et al., “Face Recognition: A Convolutional Neural Network Approach,” IEEE, 2019.
[2] X. Zhang et al., “Real-Time Face Detection and Recognition in Complex Backgrounds,” IEEE, 2018.
[3] R. Bayir, “Deep Learning Based Mask Detection in Smart Home Entries,” 2020.
[4] J. Hershey et al., “CNN-based Audio Classification for Security Applications,” IEEE, 2017.
[5] T. Gemmeke et al., “Deep Learning for Acoustic Event Detection,” IEEE, 2017.
[6] A. Ravanbakhsh et al., “Anomaly Detection Using Recurrent Neural Networks,” IEEE, 2017.
[7] M. Sabokrou et al., “Unsupervised Anomaly Detection in Surveillance Videos,” ICCV, 2018.
[8] J. Carreira et al., “Fusion of Face and Behavior Analysis for Security,” IEEE, 2018.
[9] R. Singh et al., “Ethical Considerations in AI-driven Surveillance,” IEEE, 2020.
[10] S. Agrawal et al., “Privacy-Preserving Surveillance Technologies,” IEEE, 2020.
[11] L. Li et al., “Face Detection Using CNNs,” IEEE TPAMI, 2015.
[12] S. Zhang et al., “Real-Time Face Detection in Complex Backgrounds,” CVPR, 2018.
[13] Y. Jain et al., “Deep Learning-based Mask Detection,” IEEE, 2020.
[14] S. Naseer et al., “Enhanced Network Anomaly Detection Using Deep Neural Networks,” 2015.
[15] J. Donahue et al., “Long-Term Recurrent Convolutional Networks for Visual Recognition,” CVPR, 2015.