www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 12, December 2024, pp : 273-279

e-ISSN : 2583-1062
Impact Factor : 7.001

REAL-TIME SUSPICIOUS BEHAVIOR IN PUBLIC SPACES (THEFT DETECTION)

Gaurav Kumar Lakhera[1], Dr. Uday Pratap Singh[2]

[1]Student Dept. Artificial Intelligence and Data Science Poornima Institute of Engineering and Technology Jaipur, India.

[2]Dy HoD, Assistant Professor Dept. Artificial Intelligence and Data Science Poornima Institute of Engineering and Technology Jaipur, India

DOI: https://www.doi.org/10.58257/IJPREMS37513

ABSTRACT

AI and AR have revolutionized the field of theft detection and suspicious behavior recognition in public spaces, offering unprecedented precision and situational awareness. Beyond traditional surveillance benefits such as 24/7 monitoring and deterrence, these technologies enable real-time decision-making by analyzing complex behavioral patterns in large-scale video data. AI systems can automatically detect anomalies, classify suspicious activities, and track individuals across camera networks, ensuring enhanced security. Simultaneously, AR technology overlays dynamic visual cues onto live surveillance feeds, helping security personnel visualize threats, predict movement patterns, and respond effectively to critical scenarios like theft or aggressive behavior.

Despite significant advancements, challenges remain, including false positives, data privacy concerns, and computational resource demands. This paper presents innovative AI-driven detection techniques and AR-based visualization systems that integrate seamlessly to address these issues. By employing deep learning models for behavior classification and advanced AR displays for intuitive interaction, the proposed approach guarantees higher accuracy and operational efficiency in theft prevention systems.

This study provides a comprehensive overview of methodologies, results, and the potential future of AI and AR in reshaping public safety measures, paving the way for intelligent and proactive surveillance systems capable of revolutionizing theft detection practices in diverse environments.

Keywords: AI, Augmented Reality (AR), Suspicious Behavior Detection, Theft Prevention, Real-Time Analysis, Deep Learning, Public Safety Systems.

1. INTRODUCTION

The integration of Artificial Intelligence (AI) and Augmented Reality (AR) has revolutionized theft detection and public safety systems, offering precise and proactive surveillance solutions. The increasing complexity of public spaces demands high accuracy and efficiency in identifying and mitigating security threats. Traditional surveillance often struggles with issues such as limited visibility, inconsistent tracking, and delayed responses, creating a need for innovative technologies to address these challenges.

AR bridges this gap by providing real-time visual overlays on surveillance feeds, enabling enhanced situational awareness for security personnel. Simultaneously, AI leverages predictive analytics, behavioral modeling, and anomaly detection to identify potential threats with greater precision. Together, AI and AR enable systems to detect, analyze, and respond to suspicious behaviors dynamically, improving the reliability of theft prevention measures.

This paper introduces a novel AI-AR integrated system for real-time suspicious behavior detection in public spaces. By combining AI-driven behavior analysis with AR-enhanced visualization, the system offers intuitive tools for tracking individuals, predicting activities, and facilitating seamless interaction. The inclusion of gesture-based interfaces ensures efficient control in complex scenarios, enabling swift and informed decision-making.

The system addresses common challenges such as multi-camera tracking, real-time path prediction, and actionable visualization overlays. This paper presents the system architecture, methodologies, and experimental outcomes, demonstrating the transformative potential of AI and AR in enhancing public safety and theft prevention through scalable, efficient, and user-friendly solutions.

2. PROBLEM STATEMENT

Public safety and security have become increasingly critical in modern society, driving the demand for automated systems capable of identifying suspicious activities using real-time video surveillance. These systems are instrumental in detecting and addressing threats like vandalism, theft, violence, or trespassing, enabling swift responses to incidents. However,

creating efficient and reliable systems for suspicious activity detection involves several hurdles. A major challenge is accurately identifying and classifying suspicious behaviors while minimizing false alarms.

These systems must effectively manage complex scenarios such as crowded environments, occlusions, and varying lighting conditions. Additionally, adaptability and scalability are vital to accommodate diverse settings, whether in bustling locations like train stations or secluded areas like warehouses. The variability of environments and behaviors further complicates system design. Moreover, developing activity recognition models requires extensive labeled datasets, making the process time-intensive and costly.

Efficient methods for labeling high-quality data are essential to train and validate these algorithms effectively. Suspicious activity detection systems serve multiple purposes. First, they enhance public safety by identifying threats and alerting security teams for prompt action.

They can be deployed in high-traffic public spaces like airports, malls, and train stations to prevent incidents such as theft, violence, or trespassing. Second, they strengthen security by providing continuous monitoring and identifying suspicious activities that human operators might miss, reducing security risks.

Third, these systems are cost-efficient, operating 24/7 without breaks, covering larger areas, and detecting threats more reliably than human observers. Fourth, they allow security personnel to concentrate on potential risks instead of passively monitoring video feeds, leading to better resource allocation and faster responses. Lastly, these systems can safeguard privacy by using techniques like face blurring or encryption, striking a balance between safety and privacy rights.

In conclusion, suspicious activity detection systems are crucial for improving public safety, bolstering security, optimizing resource allocation, and respecting privacy.

3. LITERATURE SURVEY

The study of Human Activity Recognition (HAR) has gained significant importance due to its broad applications in areas like healthcare, sports, and surveillance. HAR focuses on identifying and classifying human actions using sensor data such as accelerometers and gyroscopes. This survey explores the current state of HAR research, reviewing various methodologies and techniques, as well as examining the challenges and limitations in the field. These challenges include the need for extensive labeled datasets and the difficulty in accurately recognizing complex activities.

Additional challenges arise from handling large volumes of unlabeled sensor and video data, where factors like lighting, noise, and scale variation impact prediction accuracy. HAR using video sequences or images faces further obstacles such as background clutter, occlusion, and variations in viewpoint, lighting, and appearance. Many systems currently rely on sensor data or still images instead of video footage. Moreover, annotating behavioral roles requires domain-specific knowledge and is time-intensive. Similarities within and between movement classes add complexity, as does developing real-time visual models in the absence of adequate benchmark datasets.

Continuous monitoring of public areas is demanding, which highlights the need for intelligent video surveillance systems to categorize activities as normal or unusual and trigger alerts when required. By conducting this survey, the goal is to gain a comprehensive understanding of HAR research, identify existing gaps, and uncover opportunities for further study, thus advancing the field.

The approach proposed by Abobakr et al. [1] simplifies posture recognition by treating it as a pixel labeling problem, where labeled pixels are counted using a random forest algorithm. Arifoglu et al. [2] suggest generating synthetic data to simulate behaviors of dementia patients due to challenges in collecting real-world data. Atta et al. [3] review classification techniques for HAR using wearable inertial sensors placed on the chest, thigh, and ankle of healthy subjects.

Deep learning methods, including CNNs, RNNs, and CNN-LSTM hybrids, are employed by Kumar et al. [4]. Pradhan [5] proposes an activity recognition system using smartphone 3D accelerometers to extract time and frequency domain features. Bashar [6] describes a neural network model for classifying activities with hand-crafted features from embedded sensors, applicable in fields like surveillance and healthcare.

Advances in video action recognition often stem from image classification breakthroughs [7], [8]. CNNs, which achieve state-of-the-art results in image classification, have revitalized interest in deep learning for videos.

Varol et al. [9] introduce Long-term Temporal Convolution (LTC) models for extended temporal inputs, while Simonyan and Zisserman [10] propose a two-stream CNN architecture for extracting appearance and motion features, combining them via average pooling or linear SVMs.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
www.ijprems.com
editor@ijprems.com
Vol. 04, Issue 12, December 2024, pp : 273-279
e-ISSN : 2583-1062
Impact Factor : 7.001

4. METHODOLOGY

A. Proposed Model

Human Activity Recognition (HAR) systems rely on machine learning to analyze data from various sensors, including accelerometers, gyroscopes, and magnetometers, to classify human activities like walking, running, cycling, or sitting. Machine learning algorithms process sensor data to identify patterns that match specific activities. A significant challenge in HAR is selecting the right features that effectively represent the relevant information from sensor data. These features must remain robust and unaffected by variations in sensor orientation or noise. Another difficulty lies in designing accurate, efficient algorithms that can handle complex situations, such as recognizing activities involving multiple people or adapting to different environments.

Wearable devices with embedded sensors track physical activity and provide users with feedback on fitness levels and daily activity. In sports, these sensors analyze athletes' performance and detect potential anomalies that could cause injuries. In security applications, sensors identify suspicious behaviors and alert personnel for immediate action.

To implement HAR, several datasets were utilized. A combination of the KTH dataset for walking and running activities and another dataset featuring fight sequences from Kaggle.com was selected for training the model. Once the datasets were collected, the next step involved data preprocessing, which included frame-by-frame analysis to reduce noise and enhance activity detection accuracy. The dataset attributes were visualized to identify correlations among features, and relevant features were extracted for further analysis.

Following preprocessing, the model was trained. Among the available options, Long-term Recurrent Convolutional Networks (LRCN) were chosen. After building the model, it was trained, and its performance was evaluated using accuracy metrics to ensure its effectiveness in detecting human activities.

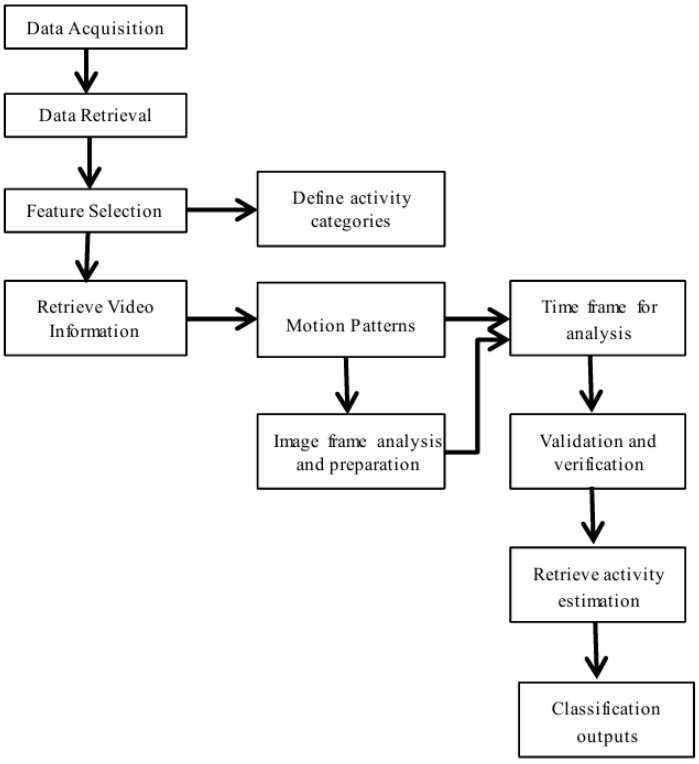


Fig. 1. Architecture Diagram

B. System Methodology

The proposed model utilizes a labeled dataset comprising video clips of individuals engaged in various activities. The KTH dataset includes two types of human actions—walking and running—while the Kaggle dataset features video sequences depicting fighting. All sequences in the study were recorded against consistent backgrounds using a fixed camera. Additionally, supplementary videos from movies and YouTube were incorporated to enhance the training process. The video clips were converted into frames represented as numpy arrays, enabling the model to capture crucial aspects for efficient preparation. Preprocessing techniques were applied to remove noise, correct camera motion, and adjust lighting and contrast in the video clips.

www.ijprems.com
editor@ijprems.com
INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 12, December 2024, pp : 273-279
e-ISSN : 2583-1062
Impact Factor : 7.001

The preprocessing steps included:

- **Frame Extraction:** Frames were converted into 3D NumPy arrays where the dimensions correspond to RGB channels.
- **Frame Resizing:** Adjusting the total number of pixels in an image, either increasing or decreasing its size.
- **Normalization:** Scaling pixel values by dividing them by 255, ensuring they fall within a range of 0 to 1, which aids in faster learning and better feature extraction.

After preprocessing, the dataset was split into a training set and a testing set in a 75:25 ratio. To ensure reproducibility, a specific seed was used to generate a consistent sequence of random numbers, allowing the same randomized splits to occur across runs. Randomness in the split prevents the model from memorizing data, contributing to more effective training. To avoid overfitting, an Early Stopping callback was implemented during training. This mechanism halts the training process once the monitored accuracy metric no longer shows improvement. Additionally, it restores the model's weights to those from the epoch with the best performance on the monitored metric. This strategy enhances the model's ability to generalize and perform effectively on unseen data. Instrument trajectories and guidance paths adapt dynamically based on tissue deformation, surgeon input, or tool positioning.

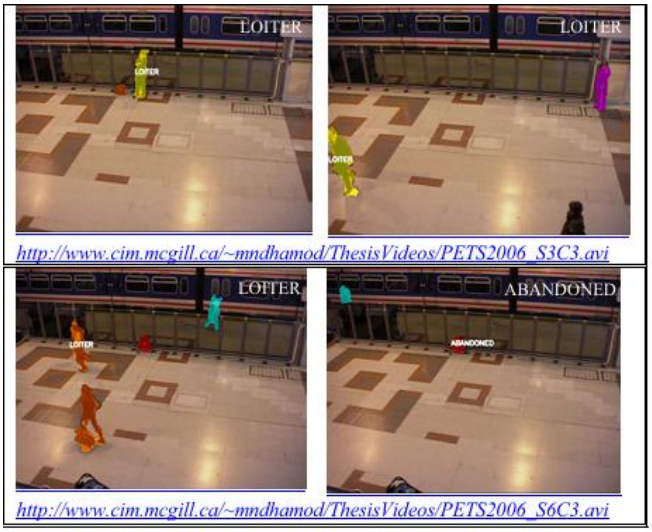
5. RESULT

Evaluating behavior recognition presents significant challenges due to the inherent complexity of human actions and the presence of clutter and other distractions in test environments [36]. Furthermore, the literature offers an insufficient number of comprehensive datasets for

analysis. Assessing the performance of behavior recognition algorithms is a complex task [36] because of (1) the lack of standardized evaluation metrics, (2) the difficulty in assigning hit and miss weights, and (3) the challenges involved in constructing accurate ground truth data. These factors contribute to inconsistencies in experimental results reported across different studies. Most publicly available datasets are low-quality, unprofessional, and fail to present adequately challenging scenarios. For instance, nearly all papers cited in [15] report high accuracy rates due to the overly simplistic nature of the datasets used. Despite these limitations, standardized public datasets were carefully selected for testing our framework, given the lack of better alternatives. The datasets used include BEHAVE[16], CAVIAR (PETS 2004)[35], and PETS 2006[30].

To ensure meaningful comparisons, our experiments were only compared to studies that used these standard datasets. However, this approach was restrictive since many papers rely on private, low-quality datasets that fail to offer significant challenges. This lack of robust datasets limits the advancement of the field and underscores the need for greater dataset standardization. Our framework demonstrated the capability to detect all behavior types, with the recognition stage requiring just 1 millisecond per frame on average. The system also proved robust to variations in camera specifications. For the CAVIAR dataset, events such as "meeting" and "walking together" were short-lived, making real-time detection challenging. Nevertheless, these events were successfully recognized qualitatively.

Figure 2 provides key frames from tested scenarios using the selected datasets, with web links to the full videos. In one scenario (CAVIAR_Meet_WalkSplit), detection of "walking together" was delayed by one second. This highlights the time window issue discussed in Section III, where certain behaviors experience minor detection delays in real-time applications.



www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 12, December 2024, pp : 273-279

e-ISSN : 2583-1062
Impact Factor : 7.001

Fig:2

Table IV compares the performance of our framework against the dataset ground truth by presenting precision and recall scores for the selected scenarios. An incorrectly raised alarm is classified as a false positive, while a missed detection is classified as a false negative. The results show that the detection of abandoned luggage, theft of luggage, loitering, and fainting events is highly accurate. For most behavior types, precision scores tend to be higher than recall scores, indicating fewer false alarms. This approach is preferable to having more missed detections at the expense of false alarms, as seen in [1].

The low performance of "meeting" and "walking together" in the CAVIAR dataset [35] is due to the brief duration of these activities and the challenges posed by the nonlinear camera calibration model.

To further evaluate our framework, we compared it against a limited number of quantitative results based on publicly available datasets from the literature. Table V presents a comparison with two other publications that provide such results. When compared to [37], our framework demonstrates a higher precision while still achieving a relatively high recall.

It's also important to mention that some parameters defining the semantic behaviors—such as acceptable walking speeds and meeting distances—must be selected based on logical and physical reasoning. Although they might seem arbitrary at first glance, many of these parameters can be determined through physical and logical considerations. For more challenging scenarios like fighting, additional experimentation was needed, and some parameters were very sensitive to variations in video footage. Determining how much movement qualifies as fighting before an event is flagged remains a complex challenge.

6. CONCLUSION

In conclusion, using LRCN for recognizing suspicious human activity presents a promising method for improving video surveillance systems. LRCN, which combines CNN and RNN, is effective in detecting both spatial and temporal features of human actions, making it a robust technique for identifying anomalous events. The model achieves an accuracy of around 83%, which demonstrates strong performance based on the selected dataset. Numerous studies have highlighted the high accuracy and efficiency of LRCN in detecting suspicious activities across various environments, including airports and shopping centers. However, there is still potential for enhancement and further exploration.

A significant challenge remains the need for large labeled datasets for training and evaluating the LRCN model. Collecting such datasets is often labor-intensive and expensive. Additionally, the performance of the LRCN approach is highly dependent on the quality of the video data, which can be influenced by factors like lighting conditions and camera angles.

Future research in suspicious human activity recognition using LRCN should aim to overcome these obstacles. For instance, improving video quality by using multi-view cameras or combining data from different sources could boost the accuracy of the LRCN model. Moreover, exploring techniques to minimize the amount of labeled data required for training, such as transfer learning or unsupervised learning, could help reduce the time and cost associated with dataset creation.

In summary, utilizing LRCN for suspicious human activity detection offers great potential to enhance public safety and security and remains an active area of study. Continued research and development in this field may lead to more precise and efficient video surveillance systems, contributing to a safer and more secure environment.

7. REFERENCES

[1] Abobakr A, Hossny M, Nahavandi S (2018) A skeleton free fall detection system from depth images using random decision forest. IEEE Syst J 12(3):2994–3005. https://doi.org/10.1109/JSYST.2017.2780260

[2] Arifoglu D, Bouchachia A (2017) Activity recognition and abnormal behaviour detection with recurrent neural networks. Procedia Comput Sci 110:86–93.

[3] Attal F, Mohammed S, Dedabrishvili M, Chamroukhi F, Oukhellou L, Amirat Y (2015) Physical human activity recognition using wearable sensors. Sensors (Switzerland) 15(12):31314– 31338.

[4] Akash Kumar, Varshini Shenoy, Puneet Tiwari Human Activity Recognition, http://14.99.188.242:8080/jspui/bitstream/1234567 89/14759/1/1NH17CS701.pdf

[5] Pinki Pradhan Human Activity Recognition using Smartphones https://www.niser.ac.in/~smishra /teach/cs460/202 1/project/21cs660_group22/

[6] Bashar SK, Al Fahim A, Chon KH (2020) Smartphone based human activity recognition with feature selection and dense neural network. Annual international conference of the ieee engineering in medicine and biology society EMBS, vol. 2020- July, pp 5888–5891, 2020.

[7] H. Weiming, T. Tieniu, W. Liang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol. 34, pp. 334-352, 2004.

[8] G. L. Foresti, C. Micheloni, L. Snidaro, P. Remagnino, and T. Ellis, "Active video-based surveillance system: the low-level image and video processing techniques needed for implementation," Signal Processing Magazine, IEEE, vol. 22, pp. 25-37, 2005.

[9] N. Firth. (2011 18 August) Face recognition technology fails to find UK rioters. NewScientist. Available:http://www.newscientist.com/article/mg21128266.000 face-recognition-technology-fails-to-find-uk-rioters.html

[10] L. M. Fuentes and S. A. Velastin, "Tracking-based event detection for CCTV systems," Pattern Anal. Appl., vol. 7, pp. 356-364, 2004.

[11] M. Elhamod, "Real-Time Automated Annotation of Surveillance Scenes," M. Eng, McGill University, Montreal, 2012. Available: http://www.cim.mcgill.ca/~mndhamod/my_eThesis.pdf

[12] H. Bouma et al., “Behavioral profiling in CCTV cameras by combining multiple subtle suspicious observations of different surveillance operators,” 2013, no. May, [23] doi: 10.1117/12.2015869

[13] Z. Shao, J. Cai, and Z. Wang, “Smart Monitoring Cameras Driven Intelligent Processing to Big Surveillance Video Data,” IEEE Trans. Big Data, vol. 4, no. 1, pp. 105–116, 2017, doi: 10.1109/tbdata.2017.2715815.

[14] S. Hommes, R. State, A. Zinnen, and T. Engel, “Detection of abnormal behaviour in a surveillance environment using control charts,” 2011 8th IEEE Int. Conf. Adv.

[15] Video Signal Based Surveillance, AVSS 2011, pp. 113–118, 2011, doi: 10.1109/AVSS.2011.6027304.

[16] J. Iskander, M. Hossny and S. Nahavandi, "A Review on Ocular Biomechanic Models for Assessing Visual Fatigue in Virtual Reality," in IEEE Access, vol. 6, pp. 19345-19361, 2018, doi: 10.1109/ACCESS.2018.2815663

[17] Masakazu Hirota, Hiroyuki Kanda, Takao Endo, Tomomitsu Miyoshi, Suguru Miyagawa, Yoko Hirohara, Tatsuo Yamaguchi, Makoto Saika, Takeshi Morimoto & Takashi Fujikado (2019) Comparison of visual fatigue caused by head-mounted display for virtual reality and two dimensional display using objective and subjective evaluation, Ergonomics, 62:6, 759-766, DOI: 10.1080/00140139.2019.1582805

[18] K. Schindler, and L.J.V. Gool, “Action snippets: how many frames does human action recognition require?”, In: CVPR, 2008.

[19] A. Yao, J. Gall, and L. Van Gool, “A hough transform-based voting framework for action recognition”, In: CVPR, 2010.

[20] S. Sadanand, and J.J. Corso, “Action Bank: a high-level representation of activity in video”, In: CVPR, 2012.

[21] Cheoi, K.J. Temporal Saliency-Based Suspicious Behavior Pattern Detection.Appl.Sci.2020,10,1020.https://doi.org/10.3390/app1003102 0

[22] Verma, K.K., Singh, B.M. & Dixit, A. A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system. Int. j. inf. tecnol. (2019). https://doi.org/10.1007/s41870-019-00364-0