IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 11, November 2024

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.524

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311031|

Suspicious Activity Detection

Prof. Aghav S.E[1], Vidya Dhiwar[2], Srushti Kekan[3], Sakshi Joshi[4], Vasudha Kale[5]

Department of Computer Engineering, S.N.D. COE & RC, Yeola, Maharashtra, India

ABSTRACT: Suspicious Activity is predicting the body part or joint locations of a person from live camera. This project will entail detecting suspicious human Activity from real-time Camera using neural networks. Human suspicious Activity is one of the key problems in computer vision that has been studied for more than 15 years. It is important because of the sheer number of applications which can benefit from Activity detection. For example, human pose estimation is used in applications including video surveillance, animal tracking and behavior understanding, sign language detection, advanced human-computer interaction, and marker less motion capturing

KEYWORDS: Suspicious activity detection, real-time camera, human pose estimation, neural networks, computer vision, video surveillance, behavior analysis, sign language recognition, human-computer interaction, motion tracking.

I. INTRODUCTION

We plan to build an application for detection of suspicious activity of people in public places in real time. Our application can be used in surveillance at places like emalls, airports, railway stations, etc. where there is a risk of robbery or a shooting attack. We will be using deep learning and neural networks to train our system. This model will then be deployed as a mobile and desktop app which will take real time camera as input and send an alert on the administrator's device if some suspicious pose is found. Human suspicious activity is related to identifying human body parts and possibly tracking their movements. Real life applications of it vary from gaming to AR/VR, to healthcare and gesture recognition.

II. BACKGROUND AND KEY CONCEPT

2.1 Background -Suspicious human activity detection refers to the process of identifying and analyzing unusual or potentially harmful human behaviors using computer vision techniques. One method to achieve this is by predicting body part or joint locations from a live camera feed, a task known as human pose estimation. This field has been extensively researched over the past 15 years due to its wide range of practical applications, including video surveillance, public safety, human-computer interaction, and even animal behavior monitoring. Advancements in neural networks and machine learning have enhanced the accuracy and efficiency of these systems, allowing for real-time detection of suspicious activities in dynamic environments. The capability to track human poses accurately without requiring physical markers has greatly expanded its usability in diverse fields.

2.2 Key Concepts:

- **Human Pose Estimation**: This technology predicts the positions of key human body parts (like joints) from images or videos, allowing machines to track human postures and movements. It forms the foundation for recognizing human actions and is widely used in surveillance, sports analysis, and motion capture.
- **Neural Networks**: These are machine learning models inspired by the human brain that learn from data patterns. In the context of suspicious activity detection, neural networks are used to automatically recognize and interpret human movements, making detection systems more accurate and reliable.
- **Real-Time Detection**: This refers to systems that can process and analyze data from live camera feeds instantly, without delays. Real-time detection is essential for applications like surveillance, where timely identification of suspicious behavior can prevent incidents or threats.
- **Computer Vision**: A subfield of AI that enables machines to interpret and make decisions based on visual data, such as images and videos. In suspicious activity detection, computer vision helps machines recognize and interpret human movements and actions automatically.
- **Video Surveillance**: The use of cameras to monitor and record human activity for security purposes. Suspicious activity detection systems enhance traditional video surveillance by automating the identification of unusual or dangerous behaviors in real time.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311031|

- **Behavior Understanding**: This involves analyzing human movements, gestures, and postures to interpret actions or intentions. It is important for systems that monitor human activity, helping them distinguish between normal and suspicious behavior.
- **Sign Language Detection**: An application of pose estimation that identifies hand and body movements to interpret sign language. It uses the same technology that tracks human poses to recognize gestures, helping improve communication for the hearing impaired.
- **Human-Computer Interaction (HCI)**: HCI refers to how humans interact with computers or machines, often through intuitive methods like gestures or movements. Advanced pose estimation and activity detection enhance HCI by enabling more natural, touchless interactions.
- **Suspicious Activity Detection**: This involves identifying abnormal or potentially dangerous human behavior by analyzing video footage, typically using AI techniques. Suspicious activities may include unauthorized access, unusual body movements, or behaviors that signal danger, all detected to enhance safety and security.
- **Markerless Motion Capture**: A method of tracking human movement without using physical sensors or markers on the body. It is useful in various fields such as gaming, entertainment, sports analysis, and surveillance, allowing for seamless tracking of motion in real-world environments.

III. RECOMMENDATION SYSTEMS FOR SUSPICIOUS ACTIVITY DETECTION IN SURVEILLANCE FOOTAGE

To enhance the effectiveness of the suspicious activity detection system described in the abstract, a recommendation system could be integrated to provide real-time suggestions or alerts based on detected patterns. Here are some ways recommendation systems can complement and improve suspicious activity detection:

1. **Contextual Alerts for Security Personnel**: A recommendation system can analyze ongoing surveillance footage and recommend immediate actions for security personnel. For example, if a handgun is detected, the system could prioritize this feed for review and suggest dispatching security forces to the location, reducing response time for critical threats.
2. **Adaptive Threat Levels**: Based on historical surveillance data, the system can recommend the threat level of detected activities. It could flag certain patterns (e.g., abandoned luggage in a busy airport) with a higher priority, suggesting increased monitoring or lockdown measures depending on the location and time of day.
3. **Personalized Surveillance Focus**: The system could recommend areas of focus based on the environment's specific needs (e.g., urban vs. suburban settings). For example, the system may recommend more attention be placed on entrances in suburban areas prone to break-ins, or on busy public spaces in urban areas where abandoned luggage is more likely.
4. **Anomaly-Based Recommendations**: Using a machine learning algorithm, the system could learn normal behavior in different environments and recommend security personnel investigate anomalous behavior that deviates from the norm, even if it's not traditionally flagged as suspicious (e.g., unusual patterns of movement).
5. **Proactive Risk Mitigation**: Based on the frequency of suspicious activities detected in specific areas over time, the system can recommend preemptive security measures. For example, it could suggest increasing surveillance in areas with high risks of abandoned luggage or weapons detection at certain times (e.g., during events or busy hours).
6. **Training and Improvement Suggestions**: The recommendation system could track which activities are frequently missed or falsely flagged, recommending adjustments to the detection algorithms or additional training data to improve the system's accuracy.
7. **Crowd Behavior Analysis**: The system can recommend potential threats in crowded areas, such as unusual crowd behavior or movements, suggesting immediate surveillance or evacuation plans in large gatherings where threats like abandoned luggage or gun detection are more likely.

IV. LITERATURE REVIEW

4.1 Paper Name: Real-Time suspicious Detection and Localization in Crowded Scenes
Author: Mohammad Sabokrou , Mahmood Fathy
Description: In this paper, we propose a method for real-time suspicious detection and localization in crowded scenes. Each video is defined as a set of non-overlapping cubic patches, and is described using two local and global descriptors. These descriptors capture the video properties from different aspects. By incorporating simple and cost-effective

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311031|

Gaussian classifiers, we can distinguish normal activities and anomalies in videos. The local and global features are based on structure similarity between adjacent patches and the features learned in an unsupervised way, using asparse autoencoder. Experimental results show that our algorithm is comparable to a state-of-the-art procedure on UCSD ped2 and UMN benchmarks, but even more time-efficient. The experiments confirm that our system can reliably detect and localize anomalies as soon as they happen in a video.

4.2 Paper Name: Learning Temporal Regularity in Video Sequences
Author: Mahmudul Hasan Jonghyun Choi
Description: Perceiving meaningful activities in a long video sequence is a challenging problem due to ambiguous definition of 'meaningfulness' as well as clutters in the scene. We approach this problem by learning a generative model for regular motion patterns (termed as regularity) using multiple sources with very limited supervision. Specifically, we propose two methods that are built upon the auto encoders for their ability to work with little to no supervision. We first leverage the conventional handcrafted spatiotemporal local features and learn a fully connected auto encoder on them.

4.3 Paper Name: suspicious Detection in Video Using Predictive Convolutional Long Short-Term Memory Networks
Author: Jefferson Ryan Medel
Description : Automating the detection of anomalous events within long video sequences is challenging due to the ambiguity of how such events are defined. We approach the problem by learning generative models that can identify anomalies in videos using limited supervision. We propose end-toend trainable composite Convolutional Long Short-Term Memory (Conv-LSTM) networks that are able to predict the evolution of a video sequence from a small number of input frames. Regularity scores are derived from the reconstruction errors of a set of predictions with abnormal video sequences yielding lower regularity scores as they diverge further from the actual sequence over time.

4.4 Paper Name: Real-Time suspicious Detection and Localization in Crowded Scenes
Author: Mohammad Sabokrou , Mahmood Fathy
Description: In this paper, we propose a method for real-time suspicious detection and localization in crowded scenes. Each video is defined as a set of non-overlapping cubic patches, and is described using two local and global descriptors. These descriptors capture the video properties from different aspects. By incorporating simple and cost-effective Gaussian classifiers, we can distinguish normal activities and anomalies in videos. The local and global features are based on structure similarity between adjacent patches and the features learned in an unsupervised way, using asparse autoencoder. Experimental results show that our algorithm is comparable to a state-of-the-art procedure on UCSD ped2 and UMN benchmarks, but even more time-efficient. The experiments confirm that our system can reliably detect and localize anomalies as soon as they happen in a video.

4.5 Paper Name: Learning Temporal Regularity in Video Sequences
Author: Mahmudul Hasan Jonghyun Choi
Description: Perceiving meaningful activities in a long video sequence is a challenging problem due to ambiguous definition of 'meaningfulness' as well as clutters in the scene. We approach this problem by learning a generative model for regular motion patterns (termed as regularity) using multiple sources with very limited supervision. Specifically, we propose two methods that are built upon the auto encoders for their ability to work with little to no supervision. We first leverage the conventional handcrafted spatiotemporal local features and learn a fully connected auto encoder on them.

4.6 Paper Name: Real-Time suspicious Detection and Localization in Crowded Scenes
Author: Mohammad Sabokrou , Mahmood Fathy
Description: In this paper, we propose a method for real-time suspicious detection and localization in crowded scenes. Each video is defined as a set of non-overlapping cubic patches, and is described using two local and global descriptors. These descriptors capture the video properties from different aspects. By incorporating simple and cost-effective Gaussian classifiers, we can distinguish normal activities and anomalies in videos. The local and global features are based on structure similarity between adjacent patches and the features learned in an unsupervised way, using asparse autoencoder. Experimental results show that our algorithm is comparable to a state-of-the-art procedure on UCSD ped2 and UMN benchmarks, but even more time-efficient. The experiments confirm that our system can reliably detect and localize anomalies as soon as they happen in a video.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

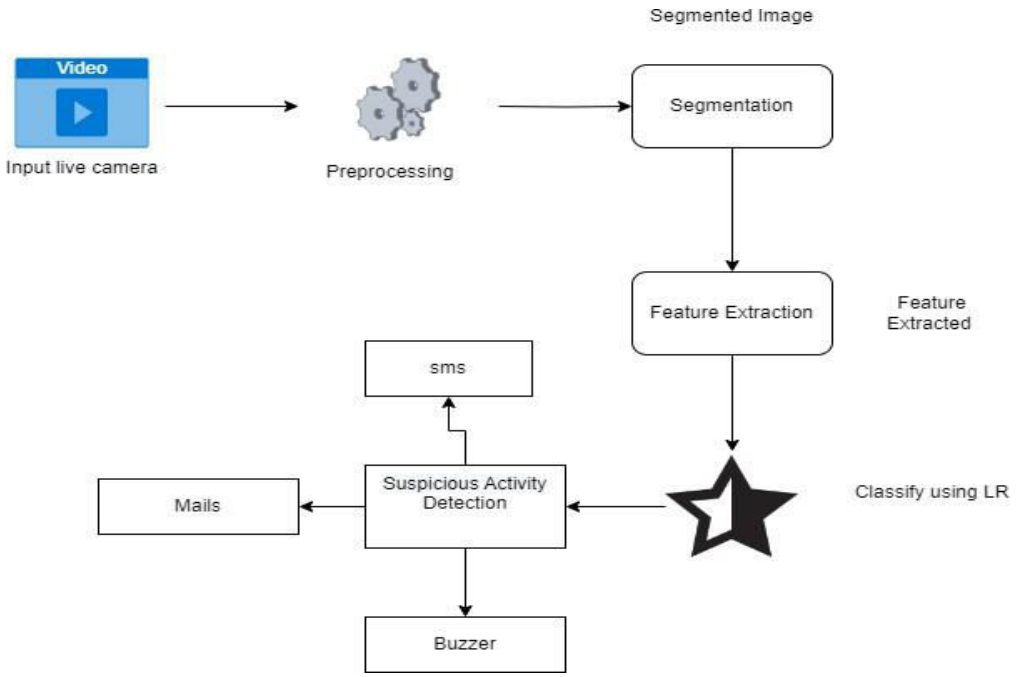
|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311031|

4.7 Paper Name: Learning Temporal Regularity in Video Sequences
Author: Mahmudul Hasan Jonghyun Choi
Description: Perceiving meaningful activities in a long video sequence is a challenging problem due to ambiguous definition of 'meaningfulness' as well as clutters in the scene. We approach this problem by learning a generative model for regular motion patterns (termed as regularity) using multiple sources with very limited supervision. Specifically, we propose two methods that are built upon the auto encoders for their ability to work with little to no supervision. We first leverage the conventional handcrafted spatiotemporal local features and learn a fully connected auto encoder on them.

V. SYSTEM ARCHITECTURE



The entire system is structured into five key steps:

1. **Image Acquisition**: The first step involves gathering live video feeds from a camera source to capture human activity in real time. The system ensures that the video is processed at an appropriate frame rate to maintain clarity and detail.
2. **Pre-processing**: This stage is crucial for enhancing the quality of the captured video. It involves converting the video frames from RGB to grayscale and applying techniques to reduce noise. Key body parts and joints are identified using algorithms such as OpenPose or similar pose estimation methods.
3. **Feature Extraction**: In this step, important features related to human pose and movement are extracted and represented as vectors during both training and testing phases. The system focuses on critical body joints (e.g., shoulders, elbows, knees) that are indicative of human activity. Techniques like Principal Component Analysis (PCA) or neural network feature extraction are employed.
4. **Activity Recognition**: A classifier (e.g., a recurrent neural network or convolutional neural network) is utilized to analyze the extracted features and classify the detected activity (e.g., walking, running, loitering). The system assigns an activity label based on the recognized patterns in the body movements.
5. **Alert and Response**: The final step involves generating alerts or responses based on detected suspicious activities. If a suspicious action is detected, the system can trigger notifications or further investigation protocols, integrating with other security systems if necessary.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311031|

VI. EVALUATION METHODOLOGY

6.1 Datasets

Selecting appropriate datasets is crucial for evaluating the performance of the suspicious activity detection system. Commonly used datasets in this field include:

- **UCF101**: A dataset containing a wide range of human activities, useful for training models on various actions and movements.
- **HMDB51**: A large dataset of realistic human activities, providing diverse action samples to enhance model training.
- **KTH**: A dataset featuring human activities captured in different scenarios, allowing for robust testing of activity detection algorithms.

These datasets offer a diverse range of activities and scenarios, making them suitable for testing and training your detection models.

6.2 Evaluation Metrics

To assess the effectiveness of the system, several metrics can be utilized:

- **Accuracy**: The percentage of correctly identified activities out of all predictions made.
- **Precision**: The ratio of true positive detections to the sum of true positives and false positives, indicating the accuracy of detected suspicious activities.
- **Recall (Sensitivity)**: The ratio of true positives to the sum of true positives and false negatives, reflecting how many actual suspicious activities were detected.
- **F1-Score**: The harmonic mean of precision and recall, providing a balanced measure useful when class distributions are uneven.
- **Confusion Matrix**: A table displaying the number of correct and incorrect predictions for each type of activity, offering insights into which behaviors are more challenging to classify.

6.3 Cross-Validation

To ensure that the model generalizes well to unseen data, k-fold cross-validation can be performed:

- **Procedure**: The dataset is divided into k subsets (folds).
- **Training and Testing**: The model is trained on k-1 folds while testing on the remaining fold.
- **Repetition**: This process is repeated k times, with results averaged to provide a reliable estimate of the model's performance.

6.4 Comparison with Baselines

You should compare your proposed model's performance against baseline models to demonstrate improvements. Common baselines include:

- **Single-modality Models**: Evaluating performance using only visual data from cameras.
- **Traditional Machine Learning Methods**: Such as Support Vector Machines (SVM) or Decision Trees applied to video data.
- **Existing Multimodal Systems**: Analyzing performance against recent literature that employs similar detection techniques.

6.5 Ablation Study

An ablation study is conducted to understand the contribution of different components of your system. For example:

- **Testing with Only Visual Data**: Evaluate the system using just the camera feed to assess its standalone performance.
- **Testing with Only Joint Data**: Assess the system's performance based solely on joint location data.

6.6 Computational Efficiency

It's important to assess the computational cost of your system, especially for real-time applications. You may include:

- **Inference Time**: The time it takes for the system to detect suspicious activities and provide alerts.

Resource Usage: Evaluate CPU/GPU memory usage and the complexity of the model in terms of the number of parameters involved.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311031|

Combining Modalities: Analyze performance when integrating both visual and joint data to illustrate improvements from a multimodal approach

VII. CHALLENGES IN SUSPICIOUS ACTIVITY DETECTION

7.1. Data Quality and Diversity

- **Limited Data Availability**: High-quality, labeled datasets specifically for suspicious activity detection are often scarce. The lack of comprehensive datasets that encompass a variety of activities across different demographics can hinder the training of effective models.
- **Data Imbalance**: Datasets may exhibit an imbalance in the number of samples for different types of activities, leading to biased models that excel in detecting frequently represented activities while struggling with less common or more nuanced behaviors.

7.2. Real-Time Processing

- **Latency**: Achieving low-latency processing for real-time detection of suspicious activities is challenging, particularly when processing video feeds and analyzing joint locations simultaneously. Ensuring the system provides immediate feedback is crucial for effective monitoring and user experience.
- **Resource Constraints**: Real-time applications often face limitations on hardware resources (CPU/GPU), necessitating optimization of neural network models to balance performance and resource consumption effectively.

VIII. FUTURE DIRECTIONS

8.1. Expanding Dataset Diversity

Encourage the creation of larger and more diverse datasets that include a variety of suspicious activities across different cultures and environments. This will enhance the model's ability to generalize.

8.2. Addressing Ethical Concerns

Develop privacy-preserving techniques, such as differential privacy and federated learning, to ensure user data is protected during suspicious activity detection processes.

8.3. Real-World Application Enhancements

Enhance alert systems by considering contextual information (user behavior patterns and environmental factors) to tailor responses more effectively.

8.4. Advancements in Model Interpretability

Invest in research on explainable AI to improve transparency in how suspicious activity detection systems make decisions, fostering user trust and understanding.

IX. CONCLUSION

This paper discusses the significance of detecting suspicious human activity using real-time camera systems, particularly in enhancing security measures. While there have been significant advancements in this field, challenges remain, such as the need for diverse and high-quality data and concerns about privacy and bias. Moving forward, research should focus on improving the accuracy of activity recognition and making these systems easier for users to understand and trust. By effectively using real-time data to detect suspicious activities, we can develop technologies that are more responsive and reliable, ultimately enhancing public safety.

REFERENCES

[1]P.Bhagya Divya, S.Shalini, R.Deepa, Baddeli Sravya Reddy,“Inspection of suspicious human activity in the crowdsourced areas captured in surveillance cameras”,International Research Journal of Engineering and Technology (IRJET), December 2017.

[2] Jitendra Musale,Akshata Gavhane, Liyakat Shaikh, Pournima Hagwane, Snehalata Tadge, “Suspicious Movement Detection and Tracking of Human Behavior and Object with Fire Detection using A Closed Circuit TV (CCTV)

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311031|

cameras ", International Journal for Research in Applied Science & Engineering Technology (IJRASET) Volume 5 Issue XII December 2017.

[3] U.M.Kamthe,C.G.Patil "Suspicious Activity Recognition in Video Surveillance System", Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018.

[4] Zahraa Kain, Abir Youness, Ismail El Sayad, Samih Abdul-Nabi, Hussein Kassem, " Detecting Abnormal Events in University Areas ", International conference on Computer and Application,2018.

[5] Tian Wanga, Meina Qia, Yingjun Deng, Yi Zhouc, Huan Wangd, Qi Lyua, Hichem Snoussie, "Abnormal event detection based on analysis of movement information of video sequence" ,Article-Optik,vol-152,January-2018.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details