Suspicious Activity Detection from Surveillance Video using Deep Learning Approach

Rohit Shinde, Sonali Suryavanshi, Akash Phad, Sarthak Kathe, Prof. S. S. Gunjal (Project Guide)

Dept. of Computer Engineering, Pune Vidyarthi Griha's College of Engineering & S.S. Dhamankar Institute of Management, Nashik.

***Abstract*– Video surveillance plays an important role in today's world. Artificial intelligence, machine learning, and deep learning entered his system, making the technology too advanced. Using a combination of the above, different systemsare positioned to help distinguish different suspicious behavior from live tracking footage. Human behavior is the most unpredictable and it is very difficult to tell if it is suspicious ornormal. Deep learning approaches are used to detect suspicious or normal activity in academic environments, sending alert messages to appropriate authorities when suspicious activity is predicted. Monitoring is often performed through consecutiveframes extracted from the video. The entire framework is divided into two parts. In the first part features are computed from the video image and in the second part the classifier predicts the class as suspect or normal based on the features obtained.**

Keywords **–Suspicious Activity, Video Surveillance, Deep Learning.**

I. INTRODUCTION

Detecting human behavior in real-world environments has numerous applications, such as intelligent video surveillance, shopping behavior analysis, and more. Video surveillance has a wide range of applications, especially indoors and outdoors. Surveillance is an important part of security. Today, security cameras have become a part of life for safety and security. E- Surveillance is one of his major agenda items for Digital India,a development program of the Government of India. Videosurveillance continues to be part of it. Advantages of video surveillance include effective surveillance, reduced manpowerrequirements, cost-effective audit capabilities, and his adoption of new security trends. There are currently people being tracked. Since we are dealing with a huge amount of video data, it is easy to get tired, and the manual work of can cause omissions. This greatly affects the efficiency of the system. This was resolved by automating video surveillance. Currently,it is not possible to manually monitor every event with a CCTV(closed circuit television) camera. Manually searching for the same event in recorded video wastes a lot of time, even if theevent has already occurred. Analyzing abnormal events using video is an emerging topic in the field of automated video surveillance systems. Artificial intelligence helps computers think like humans. In machine learning, a key component is learning from training data and making predictions about future data.

A combination ofcomputer vision and video surveillance ensures public safety.Computer vision techniques include the following stages:environment modeling, motion detection, moving objectclassification, tracking, behavioral understanding and description, and fusion of information from multiple cameras.This method requires a lot of preprocessing to extract featuresof various video sequences. Classification methods include supervised and unsupervised classification. Supervised classification uses manually labeled training data, while unsupervised classification is completely computer-controlled and does not require human intervention.

Deep Neural Networks is one of the best architectures uses to perform difficult learning tasks. A deep learning model automatically extracts features and creates a high-level representation of the image. This is more general because the feature extraction process is fully automated. A convolutional neural network (CNN) can directly learn visual patterns from image pixels. For video streams, Long Short-Term Memory (LSTM) modelscan learn long -term dependencies. An LSTM network can remember things. The proposed system uses footage obtained from CCTV cameras to monitor human behavior on campus and silently warn when suspicious events occur. The most important components of intelligent video surveillance are event detection and human behavior detection. Automaticallyunderstanding human behavior is a difficult task. On campus, various areas are under video surveillance and various activities are monitored. Neural Networks CNN and RecurrentNeural Networks (RNN).

A CNN is used to extract high-levelfeatures from an image so that the complexity of the input can be reduced. RNNs are used for classification purposes and are suitable for processing video streams. The proposed system uses a pretrained model called VGG-16 (Visual Geometry Group) trained on the ImageNet dataset.

Most of today'ssystems use video taken from CCTV cameras. In the event of crime or violence, this video is used for investigative purposes.However, it is more interesting and applicable to indoor and outdoor areas when considering systems that provide a mechanism to automatically detect anomalous or abnormalsituations in advance and alert relevant authorities. A proposedmethod is to design such systems in the academic field. This document is organized as follows: Section II presents -related research in the area of behavioral analysis to detect suspicious activity.

II. LITERATURE SURVEY

A general overview of the proposed method is provided in Section III. Implementation details are described in Section IV, and completion and further work are described inSection V II.The purpose of the piece was to detect anomalousor suspicious events in video surveillance. Advanced MotionDetection (AMD) algorithms were used to detect unauthorizedintrusions into restricted areas [1]. In the first phase, objects arerecognized using background subtraction and objects are extracted from the frame sequence. The second phase was suspicious activity detection. The advantage of the system wasthat the algorithm worked in real-time video processing and itscomputational complexity was low. However, this system is limited to storage service, and you can also use hi-tech modeto record video in the surveillance area. A semantics-based approach was proposed in [2]. The captured video data were processed and foreground objects were identified by background subtraction. After subtraction, the objects areclassified as live or non-live using a Haar-like algorithm. Object tracking was done using the real-time Blob matching algorithm. Fire detection is also mentioned in this paper of hers.Suspicious activity detected in [3] based on movement characteristics between objects. We defined suspicious eventsusing a semantic approach. Object detection and correlation technique has been used for object tracking [2]. Events are classified based on motion characteristics and time information.The computational complexity of the given framework is now
less. Anomalous events from universities were detected by zoning, and the Lucas-Kanade method was used to estimate theoptical flow within each zone. Then I created a histogram of theoptical flow vectors of size. A software algorithm is used to analyze the content of the video and classify event as normal or abnormal [4]. A system was developed to distinguish abnormal from normal events based on the analysis of motioninformation from video sequences. The HMM method was used to learn the histogram of optical flow directions for video frames.Compare the captured video frames with the existing normal frames and identify similarities between her in those frames. The system has been evaluated and validated on different datasets, such as the UMN dataset and PETS [5]. Anomalous events in video recordings could be detected by tracking people. Humans are detected from video using the background subtraction method. Features are extracted using CNN and fed to DDBN (Discriminatory Deep Belief Network). Tagged videos of suspicious event are also fed intoDDBN and their feature is also extracted. A comparison of feature extracted from the sample videos labeled with CNN extracted feature was then performed using the DDBN, and various suspicious activities were detected from the given video [6].A real-time violence detection system was developed using deep learning to prevent violent acts byspectators or players in sports. frames were extracted from real-time video in the Spark environment. When the system detects football violence, it alerts security guards. To prevent riots, the system recognizes video actions in real time and alertssecurity forces. The VID dataset was used in and achieved 94.5% accuracy for detecting violence in football stadiums [7].
The proposed CNN and LSTM based model used the UT interaction dataset. One of theshortcomings of the system was
the difficulty in identifying similar human behaviors such as pointing and his blows [8]. Understanding Crowd Behavior Using a Deep SpatiotemporalApproach The approach classifies videos into predictions of future paths of pedestrians, target estimation, and overall crowd behavior. There are 3 different categories. Spatial information for video frames was extracted using convolutional layers. The LSTM architecture was used to learn or understand sequence time motion dynamics. The datasets used in the proposed system were PWPD, ETH, UCY, and CUHK. The accuracy of the system can be improved by using a deeper architecture [9]. Capture daily human activities from videos and classify these videos into home, work, care and helpSports-related activities are performed by deep learning. A CNN was used to obtain the input features and RNN was used forclassification purposes. We used the Inception v3 model, UCF101, and Activitynet as datasets. The accuracy achieved was 85.9% for UCF101 and 45.9% for Activitynet [10]. A system was developed to monitor student behavior on the examusing neural networks and Gaussian distributions. consists ofthree different phases: face detection, suspicious.

III. SYSTEM OVERVIEW

The proposed system uses footage from CCTV camerasto monitor student activity on the campus and send messages to appropriate authorities in the event of suspicious events.

A. System Phases

The architecture has various phases suchas video acquisition, video preprocessing, feature extraction, classification, and prediction. 1) Students using mobile phones on campus-questionable class 2) Students fighting or fainting on campus-questionable class 3) Walking, running - normal class.

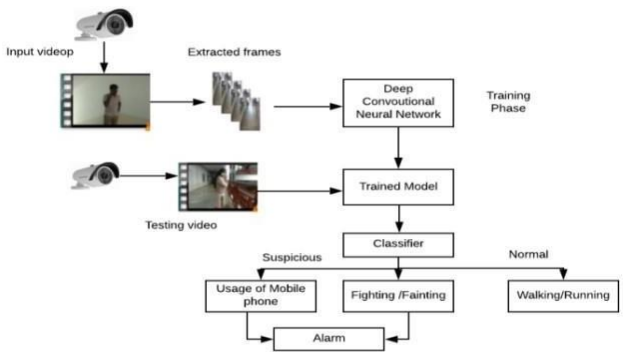

Fig-1: System Workflow

B. Video capture

Installation of CCTV camera and monitoring the footage is the initial step in video surveillance system. Various kinds of videos are captured from different cameras, covering the wholearea of surveillance. The processing in our implementation is carried out using frames, so the videos are converted to frames.

C. Dataset Description

The KTH dataset is a standard dataset which has collection of sequences representing 6 actions and each action class has got 100 sequences. Each sequence has got almost 600frames and the video is shot at 25 fps [14]. The model is traine

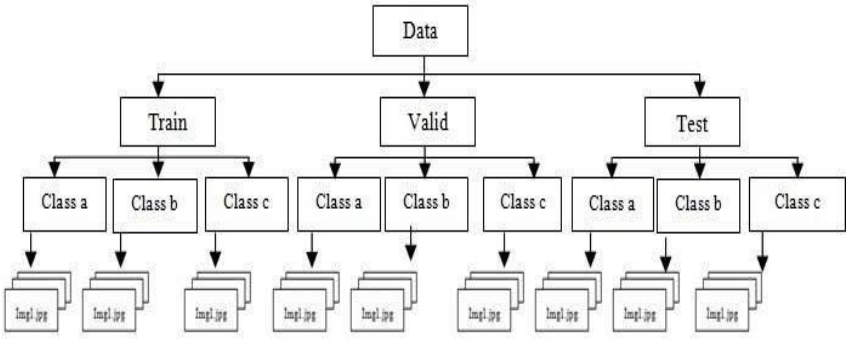

Fig-2: Data Description

on this dataset for normal behavior (running and walking). CAVIAR dataset, videos taken from campus and YouTube videos are used for training suspicious behavior (mobile phoneusing inside the campus, fighting and fainting). Around 7035 frames are extracted from different videos. The whole dataset is manually labelled and separated into 80% for training set and20% for validation set. The directory structure of dataset is as shown in Fig.2. A combination of KTH, CAVIAR, YouTube videos and videos captured from campus are used in our system.

D. Video pre-processing

A deep learning network is using in our proposed systemfor suspicious activity detection from video surveillance. By deep learning architectures, the accuracy obtained can be higher and it also works better with large datasets. A detailed design overview is represented in Fig.3.

The input videos are taken from existing and createddatasets. As part of pre-processing, frames are extracted from the captured videos. Based on the videos, three labelled foldersare created and stored the frames in it. The entire video is converted to 7035 frames and the frames are stored as images in jpg format. Each frame is then resized to 224 ×224 to suite 2D CNN architecture and stored them. The testing video is alsoconverted to frames and resized to 224 ×224 and stored in folder. OpenCV library in python is used for video pre- processing.

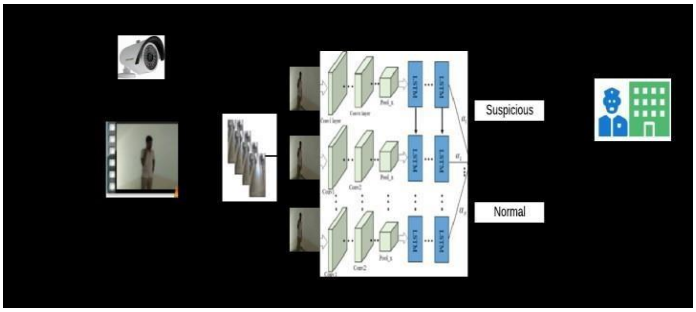

Fig-3 (A): Video Pre-Processing

In image feature extraction, a pre-trained CNN model knownas VGG-16 is trained on ImageNet dataset. VGG-16 architecture is shown in Fig.4. VGG-16 neural network [15] has convolution layers of size 3×3, max pooling layers of size2×2 and fully connected layers at end, which makes a total of16 layers was the deep learning architecture used here. The input image should be in the size 224×224×3 RGB form.

Representations of the various layers which include convolution layers, ReLU (Rectified Linear Unit) layer i.e. activation function, max pooling layers, fully connected dense layers and normalization layers. The model can fine tune as perour requirement and the last layer of this model is removed.

Then the model is trained on LSTM architecture. LSTM networks are a kind of RNN capable of learning order dependence in sequence prediction problems. We have ReLU activation function, dropout layer and fully connected dense layers. The count of neurons in the last layer is equal to the count of classes that we have and hence the number of neuronshere is three.

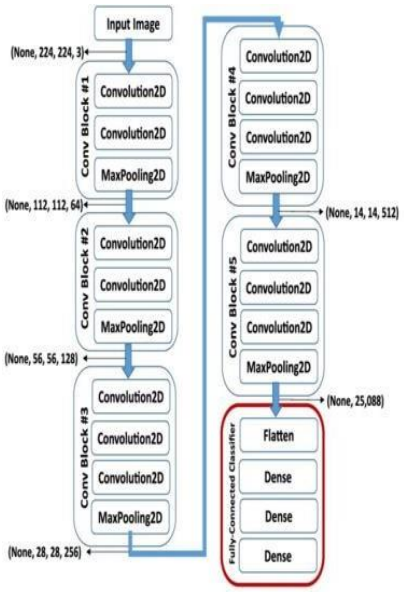

Fig-3 (B)

The system classifies the videos as suspicious (students using mobile phone, fighting, fainting) or normal (walking, running). In the case of suspicious behavior, an SMS (Short Message Service) will be send to the respective authority.

IV. RESULT ANALYSIS

The aim of the project is monitoring the suspicious activities in a campus using CCTV footages and alerts the security when any suspicious event occurs. This was done by extracting features from the frames using CNN. After the extraction was done, LSTM architecture is used to classify the frames as suspicious or normal class. Fig.5 shows the Suspicious and Normal videos sequences.

V.CONCLUSION AND FUTURE WORK

Fig-5: Sample Videos

The steps for building the complete system are collect video sequences from CCTV footage, extraction of frames fromvideos, pre-processing of the images, and preparation of training and validation sets from the datasets, training and testing. In the case of suspicious activity, the system sends an SMS to the respective authority. The system has been developed in an open source platform using python. Sending of SMS is done by creating an account in Twilio and installed the twilio library in python. Twilio allows programmatically make and receive phone calls, send and receive text messages.

A. Training and Testing

The input videos are taken from CAVIAR dataset, KTH dataset, YouTube videos and videos taken from campus. Around 300 videos of different suspicious and normal behaviorvideos are collected. As part of pre-processing, frames are extracted from the captured videos. The pre-trained model usedin our system is VGG-16 and take its learnings to solve our problem. The last layer of this model is removed based on our requirement and LSTM architecture is used for classification. Our dataset is trained on it. CCTV video footages of different scenarios are taken from our campus for testing and it is converted into frames. The stored frames are given to thetrained model and finally the classifier classifies the video intosuspicious or normal behavior.

B. Results

The accuracy of the training phase is 76% for the initial 10 epochs. The accuracy of the model can be improved by increasing the number of iterations. The frames are extracted from videos and stored in a single folder for the purpose of testing. Using our trained model, the system predicts the framesas suspicious (mobile phone using inside the campus, fighting or fainting) or normal (walking, running) class. In the case of suspicious activity, a message will be sent to the correspondingauthority with the predicted class. The accuracy achieved is 87.15%. The confusion matrix is as shown in Table I.

	Prediction M	Prediction F	Prediction N
Actual M	45	3	2
Actual F	2	18	1
Actual N	2	3	30

In present world, almost all the people are aware of the importance of CCTV footages, but most of the cases these footages are being used for the investigation purposes after a crime/incident have been happened. The proposed model has the benefit of stopping the crime before it happens. The real time CCTV footages are being tracked and analyzed. The resultof the analysis is a command to the respective authority to takean action if in case the result indicates an untoward incident is going to happen. Hence this can be stopped.

Even though the proposed system is limited to academicarea, this can also be used to predict more suspicious behaviorsat public or private places. The model can be used in any scenario where the training should be given with the suspiciousactivity suiting for that scenario. The model can be improved by identifying the suspicious individual from the suspicious activity.

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 andTechnology (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) 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 ofmovement information of video sequence" ,Article- Optik,vol152,January-2018.
6. Elizabeth Scaria, Aby Abahai T and Elizabeth Isaac, "Suspicious Activity Detection in Surveillance Video using Discriminative Deep Belief Netwok", International Journal of Control Theory and Applications Volume 10, Number 29 -2017.
7. Dinesh Jackson Samuel R,Fenil E, Gunasekaran Manogaran, Vivekananda G.N, Thanjaivadivel T , Jeeva S , Ahilan A, "Real time violence detection framework for football stadium comprising of big data analysis and deep learning through bidirectional LSTM",The International Journal of ComputerandTelecommunications Networking,2019.
8. Kwang-Eun Ko, Kwee-Bo Sim"Deep convolutional framework for abnormal behaviour detection in a smart surveillance system."Engineering Applications of Artificial Intelligence ,67 (2018).
9. Yuke Li "A Deep Spatiotemporal Perspective for Understanding Crowd Behavior", IEEE Transactions on multimedia, Vol. 20, NO. 12, December 2018.
10. Javier Abellan-Abenza, Alberto Garcia-Garcia, Sergiu Oprea, David Ivorra-Piqueres, Jose Garcia-Rodriguez "Classifying Behaviours in Videos with Recurrent Neural Networks", International Journal of Computer Vision and Image Processing,December 2017.
11. Asma Al Ibrahim, Gibrael Abosamra, Mohamed Dahab "Real-Time Anomalous Behavior Detection of Students in Examination Rooms UsingNeural Networks and Gaussian Distribution", International Journal of Scientific and Engineering Research, October 2018.
12. G. Sreenu and M. A. Saleem Durai "Intelligent video surveillance: a review through deep learning techniques for crowd analysis" , Journal Big Data ,2019.
13. Radha D. and Amudha, J., "Detection of Unauthorized Human Entity in Surveillance Video", International Journal of Engineering and Technology (IJET), 2013.
14. K. Kavikuil and Amudha, J., "Leveraging deep learning for anomaly detection in video surveillance", Advances in Intelligent Systems and Computing,2019.
15. Sudarshana Tamuly, C. Jyotsna, Amudha J, "Deep Learning Model for Image Classification", International Conference on Computational Vision and Bio Inspired Computing (ICCVBIC),2019.