JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

DEEP LEARNING APPROACH FOR SUSPICIOUS ACTIVITY DETECTION

[1]Kiran Yesugde ,[2]Vaishnavi Bankar, [3]Priya Raut, [4]Payal Salunkhe , [5]Shreya Sonawane,

[1]Professor , [2]Student , [3]Student , [4]Student , [5]Student

Department of Computer Engineering

1. Bharati Vidyapeeth College of Engineering For Women , Pune, India

Abstract: Due to the increased crime rate around the world , various organisations have started deploying surveillance systems at their locations with the help of CCTV cameras. Video surveillance plays a vital role in today's world. This system helps in detecting suspicious activities without human intervention. In this paper , we aim to automatically track people and detect unusual or suspicious activities from surveillance video and alert the shop owners. Hence, the main motive of this system is to take real-time videos from CCTV as an input and pass it to the CNN model created with the help of transfer learning and detect 'Shoplifting', 'Robbery' or 'Break-In' in the store and notify it to the owners as soon as it occurs. Monitoring of activities is performed through consecutive frames which are extracted from the video.

In general , we introduced a framework that processes raw data received from a camera which is installed at a particular location. Firstly, the proposed framework helps in detecting objects and tracking activities and then the activities are classified and it results in generating alerts to the authorised person.

Keywords : surveillance , suspicious activity, human intervention, monitoring, authorised person.

I. INTRODUCTION :

Suspicious activity is any observed behaviour that indicates a person may be involved in a crime or is about to commit a crime .Suspicious activities can be recognized from surveillance videos. Some common examples of suspicious activities include:

- A stranger loitering in your neighbourhood or a vehicle cruising the streets repeatedly.
- Someone peering into cars or windows. Someone loitering around schools, parks, or secluded areas.
- Open or broken doors and windows at an unoccupied residence.

Through the visual surveillance, human activities can be monitored in sensitive and public areas such as bus stations, railway stations, airports, banks, shopping malls, school and colleges, parking lots, roads, etc. to prevent terrorism, theft, accidents and illegal parking, vandalism, fighting, chain snatching, crime and other suspicious activities. It is very difficult to watch public places continuously, therefore an intelligent video surveillance is required that can monitor the human activities in real-time and categorise them as usual and unusual activities; and can generate an alert. In this paper, we demonstrate the overall progress of suspicious activity recognition from the surveillance videos. Number of efficient algorithms are available for the automatic detection of human behaviour in public areas.The combination of computer vision and video surveillance will ensure public safety and security. Computer vision methods involve the following stages: modelling of environments, detection of motion, classification of moving objects, tracking, behaviour understanding and description, and fusion of information from multiple cameras. This method requires a lot of pre-processing to extract features in different video sequences. The classification techniques are supervised and unsupervised classification. Supervised classification uses manually labelled training data whereas unsupervised classification is fully computer operated and does not require any human intervention.Video surveillance is a new application area for Artificial Intelligence, Machine Learning, and Deep Learning. Artificial intelligence assists computers in thinking like humans. Important components of machine learning are learning from training data and making predictions on future data. Because graphics processing units and large datasets are now available, the concept of deep learning is used.

Most of the current system uses the footage obtained from CCTV cameras. If any crime or violence happens, this video will be used for investigation purposes . But if we consider a system which will automatically detect any unusual or abnormal situation in advance and a mechanism to alert the respective authority is more interesting and which can be applied to indoor and outdoor places.

II. MOTIVATION :

The main motivation of this suspicious activity detection is to prevent various suspicious activities such as terrorism, theft cases, accidents, vandalism, fighting, crime and other suspicious activities by alert generation method.

This smart and intelligent video surveillance system can protect these sensitive areas from suspicious activities:

Airpots, Railway stations, Hospitals, Banks, Malls.

III. LITERATURE SURVEY:

The related work gives us an idea about different approaches which were used to detect the suspicious activity from video surveillance. The objective was to classify the activity as normal or suspicious.

Amrutha et al. , [1] describe the pre-trained model VGG-16. The system was designed to monitor the student's behaviour in examinations using neural networks and Gaussian distribution. In the first phase, the features were computed from video frames. And in the second phase, based on the obtained features, the classifier predicts the class as normal or suspicious. But the system was limited to academic areas which can be extended to the public as well as private sector. It can also be used in any scenario. Suspicious individuals can be suspected from suspicious activity.

Sathyajit et al. , [2] proposed the model in which the captured images were used for training purposes. The advantage of this method is that detection of abandoned baggage was computationally efficient. For each frame the computational time was considerably smaller. But improvement is needed to detect the guns in real time.

Guillermo et al. , [3] describe the 3D CNN and trained it under different approaches. For the purpose of training and classification, binary and multi-class approaches were used. The first stage was only able to classify input as normal or suspicious as it used binary training / binary classification (labelled as normal or suspicious). And the next stage was based on multi-class training. The disadvantage was that this method was unable to search the details of what type of crime it belonged to. It allows false positives between criminal classes.

Om et al. , [4] proposed the system to take videos as an input and then pass it to the CNN model which was created with the help of transfer learning. This system was used to detect 'Shoplifting', 'Robbery' or 'Break-In' in the store and as soon as possible the owners were notified. The first step was to extract the frames from the video. Second step was to pass the frame to a trained CNN model. Third step was to push the predicted label for each frame to Queue. The fourth step was to repeat step 3 for 'k' frames. The fifth and final step was to select the label with the highest probability corresponding to the mean of the last 'k' predictions. But, there was the flickering effect which can be reduced.

Guillermo et al. , [5] proposed system in which the suspicious activities are detected where a semantic based approach was used to define & detect the suspicious activities . The approach is based on the motion features between the different objects. The system's framework consisted of defining suspicious activity, background subtraction, objects detection, tracking & classification of activities. And with the help of motion features the activities were classified as normal or suspicious. The disadvantage was that datasets were not easily available, so, standard public dataset was used to test the proposed framework.

Sandesh et al. , [6] proposed a complete semantics-based behaviour recognition approach which was dependent on object tracking. The various stages of the system model were low-level pre-processing (stage 1), object detection (stage 2), tracking (stage 3), and classification of activities (stage 4). The aim was to obtain the foreground objects by using background subtraction. Then these objects were classified into people (animate) and inanimate objects (luggage). But as we know that human movements are random in nature, the reliable classification of suspicious human movements became very difficult.

Mohannad et al. , [7] proposed a paper in which the framework was introduced which processed the raw video data received from a fix colour camera installed at a particular location, which made the real-time inferences about the observed activities. The approach was totally dependent on object tracking. The objects were translated which were obtained by background segmentation into semantic entities in the scene. Then these objects were tracked in 2D and further classified as being either animate (people) or inanimate (objects). Then, their 3-D motion features were calculated and recorded in the form of historical records.

Ajeet et al. , [8] used the deep learning model which had the ability to classify and localise the activities detected using a Single Shot Detector (SSD) algorithm with a bounding box, which is explicitly trained to detect usual and unusual activities for security surveillance applications]. But, in case of real time implementation the video needs to be processed immediately. The critical operations include input and output functions.

Weiming et al. , [9] proposed the paper in which a general framework was implemented for visual surveillance systems. The model focused on following tasks - detection, tracking, understanding and description of behaviours, personal identification for visual surveillance, and interactive surveillance using multiple cameras. It had a wide spectrum of promising applications, which included access control in special areas, human identification at a distance, crowd flux statistics and congestion analysis, detection of anomalous behaviours, and interactive surveillance using multiple cameras, etc. But, to describe object behaviours by natural language in accord with human habits was challenging.

Mohannad et al. , [10] introduced an approach which detected semantic behaviours based on object and inter-object motion features . To demonstrate the capabilities of this system, a number of interesting types of behaviour have been selected. But the disadvantage was that there had been much effort to devise automated real-time high accuracy video surveillance systems.

IV. SYSTEM ARCHITECTURE :

The proposed system will use footage obtained from CCTV cameras for monitoring activities in a particular area and send alerts to corresponding authorities when any suspicious event occurs.The system architecture contains various different phases like A deep learning network is used in our proposed system for suspicious activity detection from video surveillance. By deep learning architectures, the accuracy obtained can be higher and it also works better with large datasets.The CNN algorithm's primary role in this case is to detect human behaviours and events. Understanding human behaviour is a difficult task. Robbery, shoplifting, and the use of guns and knives occur in places such as malls, shops, and railway stations, among others, which is dangerous and must be controlled. This proposed system will aid in the prevention of such activities in the future.The system architecture contains various different phases like capturing videos , pre-processing ,feature extraction.As shown is fig.1. The system is classified in two classes. 1] suspicious activity – Fighting , knife detection 2]Normal Activity – Walking , Talking.The initial step is to installation of CCTV camera and covering entire area in it.Then convert videos into frames.The dataset is taken which contain collection of Sequences of action.Each sequence has various frames, the frames are stored as images in jpg format. The testing video is also converted to frames and resized to 224 ×224 and stored in a folder .The whole dataset is labelled manually and it is separated as 80% for training and 20% for testing.

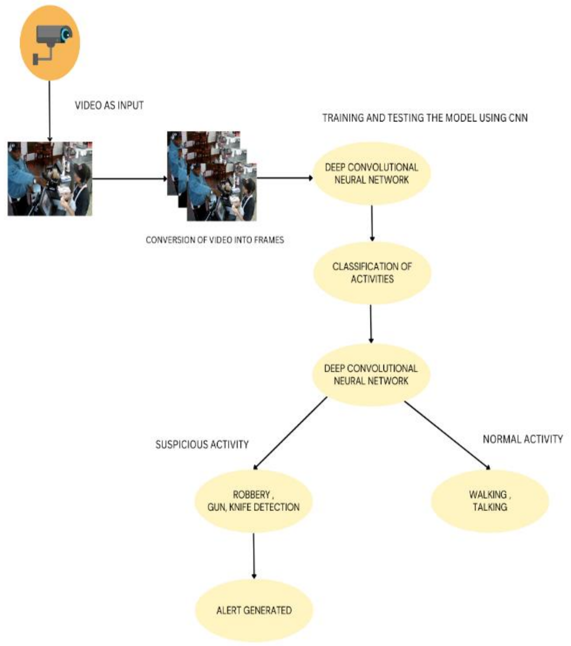

Fig.1 System Architecture

A. Data Pre-processing

The CCTV will provide input to the system, which will be used to detect suspicious activity. We used a dataset from Kaggle that consisted of three types of images: shoplifting, robbery, and gun use. Because the input images are not in proper format, various image preprocessing techniques are used to improve the image's quality. We collected approximately 16853 videos and images, of which 9676 are normal and 7177 are suspicious.

B. Training and Testing of Model

The input videos are from the DCSASS dataset, the KTH dataset, YouTube videos, and shop videos. A total of 16853 videos of suspicious and normal behaviour have been collected. Frames are extracted from captured videos as part of the pre-processing process. The model is divided into five distinct layers:

1.Convolutional layer

2.Pooling Layer

3.Fully Connected Layer

4.Dropout Layer

5.Activation Function

Real-time videos are fed into the system as input. The frames are taken from a live video stream. After that, the extracted frame is pre-processed by resizing it. The model predicts the class label for each frame extracted and pre-processed from real-time video. The same procedure is followed for each frame of the real-time video. The predicted class and probability for each input frame are displayed on the output frame. As a result, the output video is a series of frames labelled with the class label.

Our dataset is trained on it. CCTV video footage of different scenarios are taken from shops for testing and it is converted into frames. The stored frames are given to the trained model and finally the classifier classifies the video into suspicious or normal behaviour.

It is used to train our dataset. CCTV video footage of various scenarios is collected from stores for testing and converted into frames. The stored frames are fed into the trained model, which then classifies the video as suspicious or normal behaviour. Colour histograms are used to track objects in the video frame. Three histograms are calculated for each detected object, one for each of the object's red, green, and blue components. These histograms, along with any previous histograms, are associated with the respective object. The objects detected in the current frame and those detected in the previous frame must be matched for each frame that is processed. This matching of objects allows us to track a particular object across multiple frames.

C. Defining the Activity

Using our trained model, the system predicts whether the frames are suspicious (cell phone use on campus, fighting, or fainting) or normal (walking, running). In the event of suspicious activity, a message with the predicted class will be sent to the appropriate authority.

V. FUTURE SCOPE :

Even though the proposed system is limited to specific areas, this can also be used to predict more suspicious behaviours at public/private places.
This model can be used in any scenario where the training should be given with the suspicious activity suiting that scenario.

VI. CONCLUSION :

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

VII. REFERENCES :

[1] Amrutha C.V, C. Jyotsna, Amudha J. : "Deep Learning Approach for Suspicious Activity Detection from Surveillance Video" Second International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2020) IEEE Xplore Part Number: CFP20K58-ART; p:335-339.
[2] Sathyajit Loganathan, Gayashan Kariyawasam, Prasanna Sumathipala : "Suspicious Activity Detection in Surveillance Footage " 2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA).
[3] Guillermo Arturo Martınez, Mascorro, Jose ,CarlosOrtiz-Bayliss, Hugo Terashima-Marın : "Detecting Suspicious Behavior on Surveillance Videos: Dealing with Visual Behavior Similarity between Bystanders and Offenders" University of Gothenburg
[4] Om M. Rajpurkar, Siddesh S. Kamble, Jayram P. Nandagiri and Anant V. Nimkar : "Alert generation on detection of suspicious activity using transfer learning" 11th ICCCNT - IIT – Kharagpur.
[5] Guillermo Arturo Martınez, Mascorro, Jose ,CarlosOrtiz-Bayliss, Hugo Terashima-Marin:" IEEE - Detecting Suspicious Behavior on Surveillance Videos: Dealing with Visual Behavior Similarity between Bystanders and Offenders" 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA).
[6] Sandesh patil,kiran talele:" Suspicious movement detection and tracking based on color histogram" 2015 International Conference on Communication, Information & Computing Technology (ICCICT) (15-17).
[7] Mohannad Elhamod,Martin D. Venine:" Automated Real-Time Detection of Potentially Suspicious Behavior in Public Transport Areas" IEEE Transactions on Intelligent Transportation Systems .p:688-699.
[8] Ajeet Sunil , Manav Hiran Seth , Shreyas E:" Usual and Unusual Human Activity Recognition in Video using Deep Learning and Artificial Intelligence for Security Applications" 2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT).
[9] Weiming Hu, Tieniu Tan,Liang Wang:" A survey on visual surveillance of object motion and behaviors" IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews).p:334-352.
[10] Mohannad Elhamod,Martin D. Venine:" Real-Time Semantics-Based Detection of Suspicious Activities in Public Spaces"2012 Ninth Conference on Computer and Robot Vision.p:268-275.