

Synopsis
on
Real Time Anomaly Detection in CCTV Surveillance

**Submitted in the partial fulfilment of the Degree of
Bachelor of Technology
(Computer Science and Engineering)**

Submitted by
Avdhesh Chaudhary (02315002719)
Amar Sinha (04415002719)
Chirag Aggarwal (04615002719)

Under the supervision of
Mr. Vikrant Shokeen
(Assistant Professor CSE)



Department of Computer Science and Engineering
Maharaja Surajmal Institute of Technology Janakpuri,
New Delhi.
2019-23

INDEX

1. Abstract	1
2. Introduction	2
3. Literature Survey	3
4. Objective	4
5. Research Methodology	5
6. Gantt Chart	6
7. References	8

Abstract

Closed-circuit television (CCTV) cameras-based surveillance and control system are used to monitor violent incidents all over the world; however, identifying such occurrences involves human personnel. This human-based continuous monitoring in CCTV surveillance systems is error-prone because it is not humanly possible to monitor the surveillance area throughout the day and night minutely. Thus, the surveillance system should be automated.

In this project we seek to improvise by making surveillance system more advanced and automated. For this we will use Deep Learning approach and classify video footage frames as anomalous and/or normal. To achieve aforesaid, we will be using two different neural networks CNN and RNN.

Also, we will be using InceptionV3 by Google - a pre-trained model to perform object detection to train our CNN via transfer learning. The further classified output of the CNN will be divided into segments and fed to RNN. The output of the RNN system gives the real-time detection of any suspicious activity. The time complexity is reduced to a great extent using this approach.

Introduction

With the increase in world population and unemployment ratio, criminal activities are increasing with each day. It is imperative to improve the conventional surveillance and security methods. The reactive approach of the conventional policing system begins investigations following the occurrence of robbery, snatching, and assault incidents. Reactive efforts, on the other hand, are insufficient to prevent violent events.

Closed-circuit television (CCTV) cameras-based surveillance and control system are used to monitor such violent incidents all over the world; however, identifying the occurrences involves human personnel. This human-based continuous monitoring in surveillance camera systems is error-prone because it is not humanly possible to monitor the surveillance area throughout the day and night minutely. Thus, the surveillance system should be automated.

Moreover, there is a need to show in which frame and which parts of it contain the unusual activity which aids the faster judgment of that unusual activity being abnormal or suspicious. This will help the concerned authorities to identify the main cause of the anomalies occurred meanwhile saving time and labor required in searching the recordings manually.

Anomaly Detection System can be seen as a real time surveillance program designed to automatically detect and account for the signs of threatening activities immediately. We plan to use two Deep Learning models to detect and classify levels of high movement in a video frame. We plan to treat videos as segments and will define Anomalous(threatening) and Normal(safe) segments. From there, a detection alert can be raised in the case of a threat, indicating the suspicious activities at an instance of time. Further, we will recognize the following 12 anomalous activities - Abuse, Burglar, Explosion, Shooting, Fighting, Shoplifting, Road Accidents, Arson, Robbery, Stealing, Assault, and Vandalism. Detecting these anomalies would provide better security to the individuals.

To solve the above-mentioned problem, we will apply deep learning techniques used which would create phenomenal results in the detection of the activities and their categorization. Here, two Different Neural Networks: CNN and RNN are proposed.

Initially, the video input will be divided into Frames, for this we will be using OpenCV Library of Python.

Next, aim will be to perform Image Classification using a CNN. For this we have two subtasks, first would be to perform Object Detection which can done by providing the video frames to InceptionV3, a pre-trained model of Google, it performs Object Detection with high accuracy. The inceptionV3 is selected by keeping in view that the modern models used for object recognition consider loads of parameters and thus take an enormous amount of time to completely train it. However, the approach of transfer learning would enhance this task by considering initially the previously learned model for some set of classified inputs which further can be re-trained based on the new weights assigned to various new classes. That is the last layer of InceptionV3 would be used in our CNN and final Image Classification into 12 anomalies and 1 normal frame will be done by our CNN only.

The output of CNN is used to create a feature map, which will be a collection of frames to form a Video segment which is fed to the RNN as input. RNN has one additional capability of predicting the next item in a sequence. Therefore, it essentially acts as a forecasting engine. Providing the sense to the captured sequence of actions/movements in the recordings is the motivation behind using this neural network in this project work. The output of the RNN will give the final classification of the video into the 13 groups (12 anomalies and 1 normal). The output of this system is used to perform real-time surveillance on the CCTV cameras of different organizations to avoid and detect any suspicious activity. Hence, the time complexity is reduced to a great extent.

Literature Survey

A lot of work has been done on “The Real Time CCTV Surveillance using Deep Learning Algorithms” with a wide range of different approaches. We went through a lot of Research papers to decide on this approach.

Firstly, the work on “AI Image Caption Bot” - which takes an image and produces a caption for that image – using Supervised Learning opened the doorway for Computer Vision enthusiasts to apply the Neural Network approach to Videos and thus CCTV surveillance.

Some approaches to Action Recognition in Video Surveillance for Threat Detection include:

- i) Detecting a specific Anomalous Event, like a Traffic Accident. It has a lack of generalization.
- ii) Sparse Coding Based Approaches, in this approach we assume initial frames of video contains normal events for building normal event dictionary and the anomaly is detected when there is an unexpected change in the frames. This system ends up giving too many false alarms and does not handle environment changes very well.
- iii) Prediction Based Model, in this approach we use a future video frame prediction-based anomaly detection method. If a frame agrees with its prediction, it potentially corresponds to a normal event. Otherwise, it potentially corresponds to an anomalous event. But it faces constraint on intensity and gradient, optical flow gloss.

The approach we are going to use in this project is the Classification Based Approach and we will use a CNN and RNN to classify the video feed frames into different categories using Object detection. From the so far literature review, it has been observed that the maximum number of researches have designed methodologies for learning distribution of ordinary movements from the training done using available recordings.

Deep learning has resulted in being best for image classification and hence, is found suitable for video activity classification.

Objective

The main objective of our project is to develop real time Anomaly detection in CCTV surveillance.

Objectives Pointwise:

1. To reduce probability of error in anomaly detection in CCTV surveillance.
2. To reduce time of finding the video segment in which anomalous activities happen by making the process real time.
3. To increase accuracy of automatic threat detection in CCTV surveillance.
4. To increase the reliability of the system by making it more generalized and training it on 12 anomalous activities videos data.
5. To increase operational efficiency.

Research Methodology

Our approach can be divided into 6 steps:

Dataset Preparation:

The dataset we are going to need are the video footages which can be used to train our model to differentiate between various anomalous activities and normal activity. We have gone through available datasets and found that “UCF Crime Dataset” is the best suited for the purpose, it consists of 13 Categories of Anomalous activities and Normal activities footages. This will lead to major generalization of our model. The UCF Crime Dataset contains 1800 videos and is made using web scrapping these videos from websites like LiveLeak and YouTube with minor alterations for each anomaly, individually.

Video Preprocessing:

In the video preprocessing step, we will have to convert the Video into frames, so that we can perform object detection in frames so that we can recognize motion in a segment of frames.

Frame Resizing and Object Detection:

For object detection in the frames, we will not train our model from the beginning, we intend to apply transfer learning from a pretrained model, InceptionV3 by Google, so that we can save on time for object detection. InceptionV3 is learned on the ImageNet dataset. It is a large dataset released in Visual Recognition competition. The model attempts to classify entire dataset into 1,000 categories. InceptionV3 have a standard input dimension of 299 x 299 pixels thus, we will resize frames to 299 x 299 pixels for input of InceptionV3.

Convolution Neural Network Processing:

The output of the Inception model is passed to the input of the CNN which isn't the final classification model. Rather, the outcome of the last pooling layer is extracted which is a vector containing 2,048 features to feed as an input to RNN. The vector is referred to as a high-level feature map. The overall learning process requires less computational resources and training time.

Grouping of Feature Maps into a Single Pattern:

To give the framework a sense of the sequence, multiple prepossessed frames are considered. And a chunk of frames will be used to make the final classification. Such a temporal segment of the video can provide a sense of motion. For this, some feature maps are stored which are predicted by the inception model (CNN), generated in that fixed period of the video. A single combined feature map is then passed to the RNN.

Recurrent Neural Network Processing:

The input of the RNN is the concatenated collection of high-level feature maps generated in the previous step. The result of RNN will give the final classification.

Gantt Chart

Real Time Anomaly Detection in CCTV Surveillance



Gantt Chart



Real Time Anomaly Detection in CCTV Surveillance



Gantt Chart



References

1. Andrej Karpathy and Lei Fei-Fei. “*Deep Visual-Semantic Alignments for Generating Image Descriptions*”, Department of Computer Science, Stanford University, 2015.
2. Vipin Shukla, Gaurav Kumar Singh and Dr. Pratik Shah. “*Automatic Alert of Security Threat through Video Surveillance System*”, Department of Nuclear Engineering, Department of Electrical Engineering Pandit Deendayal Petroleum University, India, July 2013.
3. Utkarsh Contractor. “*Thread Detection in Surveillance Videos*”, SPARK-AI Summit 2019.
4. Atif Jan and Gul Muhammad Khan. “*Real World malicious event recognition in CCTV recording*”, January 2022.
5. Virender Singh, Swati Singh and Dr. Pooja Gupta. “*Real Time Anomaly Recognition Through CCTV Using Neural Networks*”, 2020.
6. Abdul Jaleel, Syed Khaldoon Khurshid, Rehman Mustafa, Khalid Mehmood Aamir, Madeeha Tahir and Ahmad Ziar. “*Towards Proactive Surveillance through CCTV Cameras under Edge-Computing and Deep Learning*”, March 2022.