CROWD MONITORING AND SAFETY ANALYSIS

Submitted in partial fulfillment of the requirements

Of the degree of

(Bachelor of Engineering)

By

|  |  |  |
| --- | --- | --- |
| Name of student | Class | Roll No. |

1. Darshan Pramod Nemade BE-4 25
2. Ulkesh Sharad More BE-3 73
3. Paras Sameer Thakur BE-3 80

Under the guidance of

Prof. Vaishali Chavan

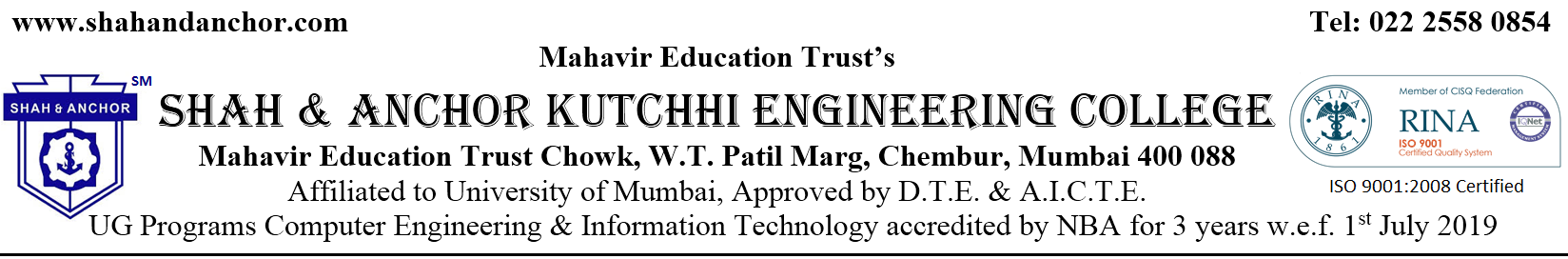


**DEPARTMENT OF COMPUTER ENGINEERING**

SHAH AND ANCHOR KUTCHHI ENGINEERING COLLEGE

**CHEMBUR, MUMBAI – 400088.**

**2020 – 2021**



Certificate

This is to certify that the report of the project entitled

**CROWD MONITORING AND SAFETY ANALYSIS**

is a bonafide work of

|  |  |  |
| --- | --- | --- |
| Name of student | Class | Roll No. |
| 1.Darshan Pramod Nemade | BE-4 | 25 |
| 2.Ulkesh Sharad More | BE-3 | 73 |
| 3.Paras Sameer Thakur | BE-3 | 80 |

submitted to the

**UNIVERSITY OF MUMBAI**

during semester VIII in partial fulfilment of the requirement for the award of the degree of

**BACHELOR OF ENGINEERING**

in

|  |  |
| --- | --- |
|  |  |

**COMPUTER ENGINEERING**.

---------------------------------------

(Prof. Vaishali Chavan)

Guide

|  |  |
| --- | --- |
|  |  |
| ------------------------------------------ | --------------------------------------- |
| (Prof. Uday Bhave) | (Dr. Bhavesh Patel) |
| I/c Head of Department | Principal |

**COMPANY’S LETTER HEAD**

**Date**

**To,**

**The Principal**

**Shah and Anchor Kutchhi Engineering College,**

**Chembur, Mumbai-88**

**Subject: Confirmation of Attendance**

**Respected Sir,**

**This is to certify that Final year (BE) students from your college**

**Darshan Pramod Nemade, Ulkesh Sharad More and Paras Sameer Thakur**

**have duly attended the sessions on the day allotted to them during the period from 2020 to 2021 for performing the Project titled CROWD MONITORING AND SAFETY ANALYSIS.**

**They were punctual and regular in their attendance. Following is the detailed record of the student’s attendance.**

**Attendance Record:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Date** | **Student1** | **Student2** | **Student3** | **Student4** |
|  | **Present/Absent** | **Present/Absent** | **Present/Absent** | **Present/Absent** |

**Signature and Name of External Guide**

|  |  |
| --- | --- |
|  | **Mahavir Education Trust's**  **SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE**  **Chembur, Mumbai - 400 088**  **UG Program in Computer Engineering** |

**Attendance Certificate**

**Date**

**To,**

**The Principal**

**Shah and Anchor Kutchhi Engineering College,**

**Chembur, Mumbai-88**

**Subject: Confirmation of Attendance**

**Respected Sir,**

**This is to certify that Final year (BE) students**

**Darshan Pramod Nemade , Ulkesh Sharad More and Paras Sameer Thakur**

**have duly attended the sessions on the day allotted to them during the period from 2020 to 2021 for performing the Project titled CROWD MONITORING AND SAFETY ANALYSIS.**

**They were punctual and regular in their attendance. Following is the detailed record of the student’s attendance.**

**Attendance Record:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Date** | **Student1** | **Student2** | **Student3** | **Student4** |
|  | **Present/Absent** | **Present/Absent** | **Present/Absent** | **Present/Absent** |

**Signature and Name of Internal Guide**

Approval for Project Report for B. E. Semester VII

This project report entitled PROJECT TITLE by Darshan Pramod Nemade, Ulkesh Sharad More and Paras Sameer Thakur is approved for semester VIII in partial fulfilment of the requirement for the award of the degree of Bachelor of Engineering.

Examiners

1.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Guide

1.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Date:

Place:

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

|  |  |  |  |
| --- | --- | --- | --- |
| Name of student | Class | Roll No. | Signature |

Darshan Pramod Nemade BE-4 25

Ulkesh Sharad More BE-3 73

Paras Sameer Thakur BE-3 80

Date:

Place:

**Abstract**

This project is differs from other implementations in quite a few aspects. Firstly you can input any video based data for training. The data will be pre processed and converted into numpy matrix which will be used for training. Converting data into matrix allows for more efficient retention of data. We have also tinkered with the layers and parameters of the model to make it more efficient and accurate. Our changes led to a boost in accuracy compared to baseline. We have also added 4 different ways to deploy the model. A person can use frames extracted from video to get accuracy, use real time video feed, use saved video or directly use a numpy file of the data to be classified and pass it through the model. All these changes makes our project flexible, adaptable and modular.

**Acknowledgement**

We wish to express gratitude to our principal Dr. Bhavesh Patel for allowing us to go ahead with this project and giving us the opportunity to explore this domain. We would also like to thank our Head of Department Prof. Uday Bhave for our constant encouragement and support towards achieving this goal. We would also like to thank the Review Committee for their invaluable suggestions and feedback without whom our work would have been very difficult. We take this opportunity to express our profound gratitude and deep regards to our guide Mrs. Vaishali Chavan for her exemplary guidance, monitoring and constant encouragement throughout the course of this project. The blessing, help and guidance given by her time to time shall carry us a long way in the journey of life on which we are about to embark. No project is ever complete without the guidelines of these experts who have already established a mark on this path before and have become masters of it. So, we would like to take this opportunity to thank all those who have helped us in implementing this project.

Table of Contents

[List of Figures 10](#_Toc71132746)

[List of Tables 11](#_Toc71132747)

[List of Abbreviations 12](#_Toc71132748)

[Chapter 1. Introduction 13](#_Toc71132749)

[1.1 VIDEO SURVEILLANCE 13](#_Toc71132750)

[1.2 VIDEO SURVEILLANCE ARCHITECTURE 13](#_Toc71132751)

[1.2.1 Automated Surveillance System 14](#_Toc71132752)

[1.3 INTRODUCTION TO ANOMALY DETECTION 16](#_Toc71132753)

[1.4 MOTIVATION 17](#_Toc71132754)

[1.5 ORGANIZATION OF THESIS 18](#_Toc71132755)

[Chapter 2. Literature Survey 20](#_Toc71132756)

[2.1 Survey Existing system 20](#_Toc71132757)

[2.2 Anomaly Detection Approaches In Video Surveillance 20](#_Toc71132758)

[2.3 Review on Detecting Anomalies In Video Sequences 22](#_Toc71132759)

[2.4 Review on Detecting Anomalies In Crowded Video Sequences 25](#_Toc71132760)

[2.5 Limitation of existing system or research gap 27](#_Toc71132761)

[2.5.1 Challenges In Anomaly Detection 27](#_Toc71132762)

[2.5.2 Gap Analysis on Detecting Anomalies 27](#_Toc71132763)

[2.6 Problem Statement and Objective 28](#_Toc71132764)

[2.7 Scope 28](#_Toc71132765)

[Chapter 3. Proposed System 30](#_Toc71132766)

[3.1 Algorithm 30](#_Toc71132767)

[3.2 Details of Hardware & Software 30](#_Toc71132768)

[3.2.1 Software Required 30](#_Toc71132769)

[3.2.2 Hardware required 31](#_Toc71132770)

[3.3 Design details 31](#_Toc71132771)

[3.3.1 An Efficient Spatio-Temporal Frequent Object Mining Method to Predict Abnormal Activities 31](#_Toc71132772)

[3.4 Methodology 32](#_Toc71132773)

[3.4.1 Preprocessing 32](#_Toc71132774)

[3.4.2 Feature Learning 32](#_Toc71132775)

[3.4.3 Regularity Score 33](#_Toc71132776)

[Chapter 4. Implementation Details 34](#_Toc71132777)

[4.1 Modules & Description 34](#_Toc71132778)

[4.1.1 AutoEncoder 34](#_Toc71132779)

[4.2 Snapshot 36](#_Toc71132780)

[4.2.1 DATASET DESCRIPTION 36](#_Toc71132781)

[Chapter 5. Testing 37](#_Toc71132782)

[5.1 Testing 37](#_Toc71132783)

[5.1.1 OUTPUT OF ANOMALY DETECTION 40](#_Toc71132784)

[5.2 Results 41](#_Toc71132785)

[Chapter 6. Results & Analysis 43](#_Toc71132786)

[6.1 PERFORMANCE METRICS 43](#_Toc71132787)

[6.1.1 True Positive Rate 43](#_Toc71132788)

[6.1.2 Accuracy 43](#_Toc71132789)

[6.1.3 Precision 43](#_Toc71132790)

[6.1.4 Recall 43](#_Toc71132791)

[6.1.5 Information Gain Ratio 43](#_Toc71132792)

[6.1.6 Regularity Score 43](#_Toc71132793)

[6.2 Results & Analysis 44](#_Toc71132794)

[Chapter 7. Conclusion and Future Scope 47](#_Toc71132795)

[7.1 APPLICATIONS OF ANOMALY DETECTION 47](#_Toc71132796)

[Chapter 8. References 49](#_Toc71132797)

# List of Figures

[Figure 1 Architecture of a simple video surveillance system 13](#_Toc71131768)

[Figure 2 Architecture of automated surveillance system 14](#_Toc71131769)

[Figure 3 Anomaly detection process in real time video sequence 15](#_Toc71131770)

[Figure 4 anomalies observed during object or human behavior 16](#_Toc71131771)

[Figure 5 literature review work done towards 19](#_Toc71131772)

[Figure 6 Different approaches used for anomaly detection 21](#_Toc71131773)

[Figure 7 Algorithm of proposed system 29](#_Toc71131774)

[Figure 8 Frame Sequence of a segmented video 36](#_Toc71131775)

[Figure 9 installing required python library 37](#_Toc71131776)

[Figure 10 run python code 37](#_Toc71131777)

[Figure 11 User interface to open test data 38](#_Toc71131778)

[Figure 12 User interface define data location 38](#_Toc71131779)

[Figure 13 Test data under analysis 39](#_Toc71131780)

[Figure 14 Real time detection result 40](#_Toc71131781)

[Figure 15 Man throwing bag in air 40](#_Toc71131782)

[Figure 16 Small Boy Jumping 40](#_Toc71131783)

[Figure 17 Man Running Man 41](#_Toc71131784)

[Figure 18 Running In Opposite Direction 41](#_Toc71131785)

# List of Tables

[Table 1 System Accuracy for Different Data Samples Taken 43](#_Toc71131759)

[**Table 2 System Precision for Different Data Samples Taken** 43](#_Toc71131760)

[Table 3 System Recall for Different Data Samples Taken 44](#_Toc71131761)

[Table 4 Accuracy Comparison for Different data 44](#_Toc71131762)

# List of Abbreviations

|  |  |  |
| --- | --- | --- |
| AI | : | Artificial Intelligence |
| ANN | : | Artificial Neural Network |
| API | : | Application Programming Interface |
| BG | : | Background |
| BN | : | Batch Normalization |
| CDNET | : | Change Detection .net |
| CNN | : | Convolutional Neural Network |
| CPU | : | Central Processing Unit |
| FC | : | Fully connected |
| FP | : | False Possitive |
| FN | : | False Negetive |
| TP | : | True Possitive |
| TN | : | True Negetive |
| FG | : | Foreground |
| FOV | : | Field Of View |
| GPGPU | : | General purpose computing on graphics processing units |
| ILSVRC | : | ImageNet Large Scale Visual Recognition Challenge |
| ML | : | Machine Learning |
| MLP | : | Multilayer Perceptron |
| MoG | : | Mixture of Gaussian |
| NN | : | Nearest Neighbor |
| PTZ | : | Pan Tilt Zoom |
| RGB | : | Red Green Blue |
| ROI | : | Region Of Interest |
| SFO | : | Static Foreground Object |
| SGD | : | Stochastic Gradient Descent |
| VDAO | : | Video Database of Abandoned Objects |

# Introduction

## Video Surveillance

Video Surveillance is defined as the continuous monitoring of various activities and behavior of objects in the vicinity of the area covered for monitoring. Surveillance also involves monitoring the changing behavior observed across objects prevailing in the monitoring area. The observation of events and recording in a surveillance system is typically done using Closed Circuit Television (CCTV) cameras. Surveillance systems are deployed for monitoring various premises like homes, banks, offices and across public gathering places such as airports, railway stations and theatres to prevent the occurrence of any mishaps or untold incidents. There are different categories and types of surveillance cameras available in the market such as dome camera, bullet camera, c-mount camera, Pan Tilt Zoom (PTZ) camera and Day-Night camera and based on the contextual nature of the surveillance environment a suitable camera is deployed for monitoring.

## Video Surveillance Architecture

A typical video surveillance system takes the input from a video camera. Depending of the environment and the purpose for which the video surveillance system is needed, different types of input cameras are used to capture the footage. The footages are encoded and sent to the Network Video Recorder (NVR) through the network switch. The recordings are displayed in a monitor which is typically a TV screen or PCs. Nowadays IP cameras are used to capture the footage and send it across the internet for remote monitoring. These video footages are in general analyzed constantly by a LAN administrator present in the base station where the system is installed or by a WAN administrator who is present in a remote location. If any abnormal activities or anomalies are spotted during the monitoring the administrator or the security officer takes immediate actions based on the severity of the security incident. A simple video surveillance system architecture using analog and IP camera is shown in Figure 1.

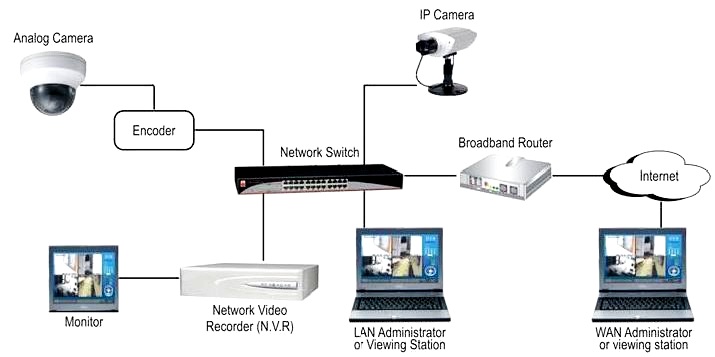
****

Figure 1 Architecture of a simple video surveillance system

### Automated Surveillance System

The primary drawback in using traditional video surveillance system is the huge dependency on human capabilities to analyze and make decisions based on the monitored video footage. The essence of real time monitoring system diminishes if any key events are missed due to human errors and sometimes this may lead to security breaches as well. The primary benefits of using digital video are the ability for computer processing of frames and easy analysis of video which is also referred to as video analytics. This involves using a set of computer intensive algorithms that identify and monitor the changes even at pixel level in surveillance area. The intra frame changes observed are used to identify the movement of objects or human, recognize people activity or people as a grouping and also to detect anomalous events in the footage without any human intervention. A typical architecture of a video analytics based surveillance system is depicted in Figure 2.

As depicted in the above figure, the input video is captured and segmented to frames. Then various image analysis techniques which include pre-processing, background subtraction, foreground extraction, object detection and tracking algorithms are applied over the frames. The inter-frame features are compared using various algorithms and the events are classified using any of the classification techniques such as Support Vector Machine (SVM), Machine Learning, Convolution Neural Network (CNN) and Deep Learning. The anomalous events are identified and then suitable alarms are generated using event generators thus eliminating the need for human intervention in analyzing the video footage and making decisions. As a result of this automation, the time taken to analyze and make decisions and in turn the time taken to generate and trigger alert is minimized to a great extent. Several applications of video analytics include the following:



Figure 2 Architecture of automated surveillance system

1. People counting in crowded places
2. Anomaly detection
3. Motion tracking
4. Object detection
5. Object tracking
6. Direction based tracking

This report work is focused towards developing enhancement of algorithm used to detect different types of anomalies in real time video.

## Introduction To Anomaly Detection

An anomaly refers to a deviation from the normal routine. They may be any abnormal events or activities identified during surveillance. Anomaly detection is defined as the task of finding abnormal patterns in the behavior of objects that do not conform to expected behavior. A typical video based anomaly detection system as depicted in Figure 3 involves extracting the relevant features from training videos and these relevant features are mapped from feature space to the actual behaviors. Each behavior is described by a set of feature descriptors. The feature descriptors are modeled based on the training and learning. The actual video sequence is given as the test input and is evaluated with the extracted features. A classifier is used to classify normal events and abnormal events in video sequence. This classifier uses different classification algorithms such as SVM, CNN, machine learning, deep learning etc. The selection of classification algorithm is dependent on the nature of the surveillance application and also on the nature of the data that will form as input to the surveillance system. The efficiency of the anomaly detection system relies highly on the accuracy of the classification algorithm.

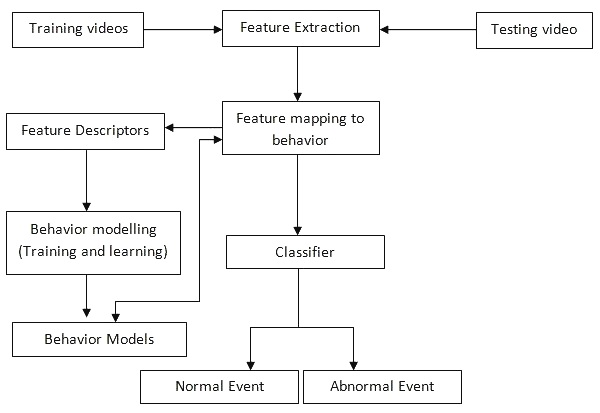


Figure 3 Anomaly detection process in real time video sequence

Anomaly detection involves continuous monitoring of the behavior of both human as well as objects in the area under surveillance. The normal activities involving objects and normal human behavior are not significant in terms of video surveillance perspective. However, the abnormal events that may occur from objects are highly significant. Some of these anomalous or abnormal events observed while monitoring the objects may be objects that are being stolen, objects being misplaced, objects that are abandoned intentionally etc. Similarly human behavior can also be classified as normal and abnormal behavior. Normal behavior includes walking, eating, talking etc, and these are less significant from the point of anomaly detection. Abnormal activities performed by human may be individual activities such as repeated actions, masked faces, lifting or touching protected objects or things, unwanted gestures, unexpected falls etc, or they may be group behaviors that involve more than one individual such as quarrels and fights, sudden crowd scattering or running etc. Figure 4 depicts some of abnormal events that occur during the surveillance of objects and human behavior.

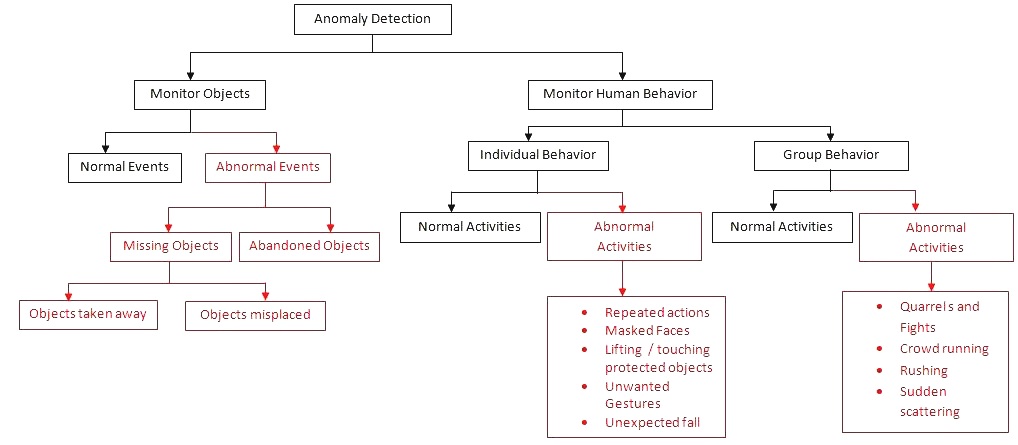
****

Figure 4 anomalies observed during object or human behavior

## Motivation

It has been observed that in spite of the presence of several secured surveillance systems in place, several untoward incidents such as thefts, robbery, terror plots etc take place. The video footage in such cases are mostly used for post-mortem analysis rather than preventing the untoward incident from happening. This serves as a motivation to develop a computer vision based smart system to detect and alert anomalous activities instantaneously and automatically without any human intervention. With this motivation a Computer Vision Based Anomaly Detection System (CVADS) is proposed which automatically detects anomalies such as presence of masked faces, anomalous activities that may result in security threat and detecting abandoned objects. This system automatically detects the anomalies without any intervention of security persons and sends instantaneous security alerts.

## Organization Of Thesis

The organization of the thesis is discussed chapter wise as below:

Chapter 1 presents an introduction to surveillance system, types of anomalies, significance of anomaly detection and the necessity to develop a computer vision based anomaly detection for surveillance systems. It also discusses in detail on the advantages of developing intelligent anomaly detection systems. Chapter 2 discusses the various existing works related that have been carried out towards detecting different types of anomalies in surveillance systems using computer vision. The review covers the relevant works carried out across detecting masked faces, abnormality detection and detecting abandoned objects. Chapter 3 presents a computer vision based approach that uses a set of pivotal points to detect the presence of partially occluded faces or masked faces and the results have been shared. In Chapter 4, an Intelligent Video Analytics Model which uses spatio-temporal aspects to detect abnormal event occurrence in real time videos has been presented. Chapter 5 present results and chapter 6 result analysis and concluded in chapter 7.

# Literature Survey

## Survey Existing system

A detailed literature review was carried out to understand the existing approaches used towards detecting anomalies in surveillance videos and the details of the same are presented in this chapter. Also, the literature survey was carried out to identify the various datasets, tools and classifiers that were used to detect different types of video anomalies from surveillance perspective. The review discusses about the literature regarding: detecting masked or partially occluded faces, detecting anomalous object movements, detecting anomalous activities in crowded environment and detecting abandoned objects. A detailed analysis and comparative study of various methods used for detecting different anomalies has also been performed and presented in this chapter. Figure 5 shows the high level taxonomy of the literature work carried out as part of this work towards anomaly detection.

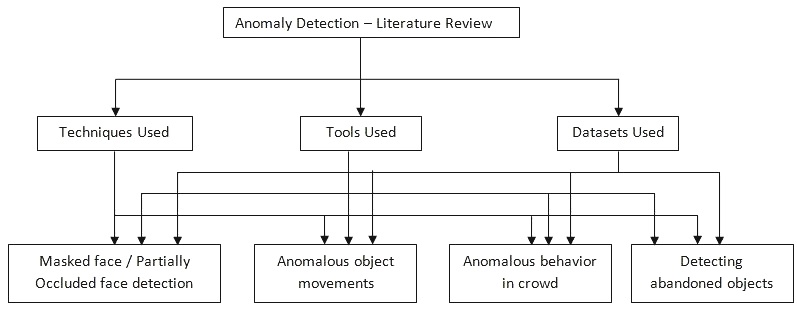


Figure 5 literature review work done towards

## Anomaly Detection Approaches In Video Surveillance

Various approaches to detect anomalies in surveillance video have been proposed. The choice of a suitable approach is dependent on the nature of data available and also on the environmental characteristics where the surveillance application is deployed. In applications where there is a known behavioral pattern, anomaly detection becomes easier and can be implemented using some rule based approaches where a set of pre-defined rules are coined and fed to the surveillance system. Any deviation from those rules is categorized as anomaly. A rule based anomaly detection approach to detect the anomalies in ship was proposed by (Liu et al 2015). Various parameters such as longitude, latitude, speed and direction were considered to frame the rules that determine the trajectory of movement of ship. An optimal decision rule based approach was proposed by (Saligrama et al 2012) to determine local anomalies and a probabilistic framework was developed. When the common behavioural pattern is unknown, training based approaches are preferred for detecting anomalies. Training based approaches involve the usage of some set of data to train the system to understand the common behavioural pattern and there by classify any abnormal activities as anomalies. Standard classifier based approaches such as Random Forest, SVM and other classification mechanisms are used to classify anomalies. When the training data is not balanced an ensemble of classifiers are deployed to balance the training data. When the data is auto-correlated, time series based approaches or Recurrent Neural Network based approaches are used. However, the training data may not be available at all times. In such cases, anomaly detection can be accomplished using semi-supervised or unsupervised learning. It may be applying some point based anomaly approaches such as percentiles and histograms or applying some collective anomaly approaches. If the data is univariate in nature, Markov chain based approach or any model based approaches can be deployed to detect anomalies. When the data is multivariate and ordered, a combination of clustering and Markov chain based approaches can be used. If the data is multivariate and un-ordered, any of the clustering based approaches or K-nearest neighbour based approaches can be used. Figure 6 shows the taxonomy of different approaches to anomaly detection.

This literature review was performed over several works related to detecting anomalies such as detecting masked or partially occluded faces, anomaly detection in video sequences, detecting anomalies in crowded area and detecting abandoned objects in video.

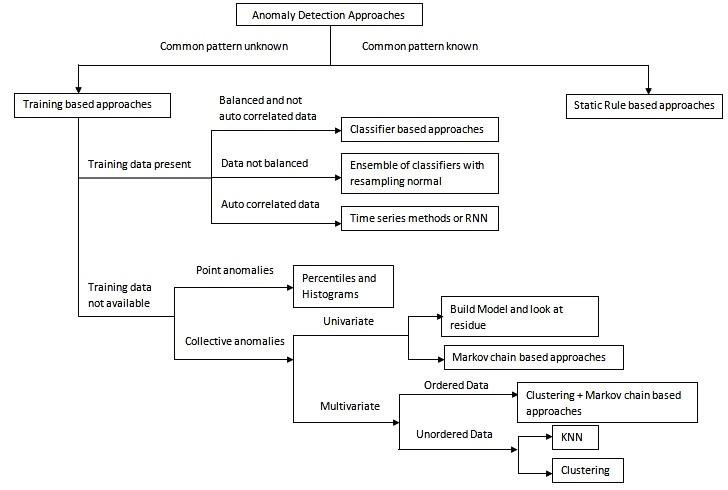


Figure 6 Different approaches used for anomaly detection

## Review on Detecting Anomalies In Video Sequences

Kim & Reddy (2006) had proposed a network based measurement approach which can spontaneously identify and detect attacks and anomalous traffic by monitoring packet headers passively. Saligrama *et al.*(2010) proposed a family of unsupervised approaches to anomaly detection in videos based on statistical activity analysis.(Li *et al.* 2012) have addressed the automatic anomaly detection problem for surveillance applications by devising a general framework for anomalous event detection in un-crowded sequences. Tran *et al.*(2014) proposed a solution to search for spatio-temporal paths for detecting events in video which can detect and locate video events accurately in cluttered space and at the same time produced stable results to camera motions.

Hu *et al.*(2018) proposed a modified LBP called as squirrel cage LBP (SCLBP) that can encode the motion information effectively and was robust to noise and unwanted disturbances caused by dynamic background and lighting changes. Piciarelli *et al.*(2008) proposed an approach based on single-class Support Vector Machine (SVM) clustering, where the SVM classifier was used for the identification and detection of anomalous trajectories. Piciarelli & Foresti (2011) have worked towards semantically interpreting video sequences to detect anomalous, dangerous or forbidden situations. Leyva *et al.*(2017) proposed an approach that used a compact set of highly descriptive features, which was extracted from a new cell structure which helped to define supportive regions from coarse to fine fashion.

Sabokrou *et al.*(2016) introduced two novel cubic patch based anomaly detector approaches where one worked based on power of an auto encoder on reconfiguring an input video patch and another one was based on the sparse representation of an input video patch. Using this, a fast and precise video localisation and anomaly detection method was presented. Laxhammar & Falkman (2014) proposed a sequential Hausdorff Nearest Neighbor Conformal Anomaly Detector (SHNN-CAD) for online learning and sequential anomaly detection in trajectories. This algorithm was having less input parameters and offered a well formed approach to calibrate the anomaly threshold. Mo *et al.*(2014) developed a new joint model based on sparse representation for anomaly detection that enabled the joint anomalies detection involving more than one objects. A greedy pursuit technique was deployed to solve the continuous sparsity problem.

Xiang & Gong (2008) proposed a new framework for automatic behaviour profiling and online detection of anomalies without any manual labelling of the training data set with the aim to address the modelling video behaviour problem captured in surveillance videos for the application of anomaly detection and online normal behaviour recognition. Thomaz *et al.*(2018) developed a family of algorithms based on sparse decompositions that detect anomalies in video sequences obtained from slow moving cameras to restrict search space to the most relevant subspaces search spaces. Cheng *et al.*(2015) presented a hierarchical framework for detecting local and global anomalies via hierarchical feature representation and Gaussian process regression (GPR) which was fully non-parametric and robust to the noisy training data, and supported sparse features.Hu *et al.*(2016) proposed a deep incremental slow feature analysis (D-IncSFA) network which was constructed and applied to directly learning progressively abstract and global high-level representations from raw data sequence. The D-IncSFA network had the functionalities of both feature extractor and anomaly detector that make AD completion in one step.

Ying Zhang *et al.*(2016) proposed a novel anomaly detection approach based on Locality Sensitive Hashing Filters (LSHF), which hashed normal activities into multiple feature buckets with Locality Sensitive Hashing (LSH) functions to filter out abnormal activities. (Emmanu Varghese *et al.*) proposed a new supervised algorithm for detecting abnormal events in confined areas like ATM room, server room etc. (Siqi Wang *et al.* 2018) proposed a novel approach to detect and localize video anomalies automatically. Video volumes were jointly represented by two novel local motion based video descriptors, SL-HOF and ULGP-OF. Sovan Biswas & Venkatesh Babu(2017) proposed a novel idea of detecting anomalies in a video, based on short history of a region in motion based on trajectories. Maying Shen *et al.*(2018) proposed a Nearest Neighbour (NN) based search with the Locality-Sensitive B-tree (LSB-tree) to detect anomalies, which helped to find the approximate NNs among the normal feature samples for each test sample. Dan Xu *et al.*(2014) proposed an approach to detect anomalies based on a hierarchical activity pattern discovery framework, comprehensively considering both global and local spatio-temporal contexts. Tian Wang *et al.*(2018) proposed an algorithm to solve abandoned object detection efficiently based on an image descriptor which encodes the movement information and the classification method.

Huorong Ren *et al.*(2017)proposed an anomaly detection approach based on a dynamic Markov model. This approach segmented sequence data by a sliding window. Also, an anomaly substitution strategy was proposed to prevent the detected anomalies from impacting the building of the models and keep anomaly detection continuously. Fan Jiang *et al.*(2011) proposed a hierarchical data mining approach where frequency-based analysis was performed at each level to automatically discover regular rules of normal events. Events deviating from these rules were identified as anomalies. Shifu Zhou *et al.*(2016)coupled anomaly detection with a spatial–temporal Convolutional Neural Networks (CNN) to capture features from both spatial and temporal dimensions by performing spatial–temporal convolutions, thereby, both the appearance and motion information encoded in continuous frames were extracted.

## Review on Detecting Anomalies In Crowded Video Sequences

In most of the existing video surveillance systems, objects that are in motion alone are identified and tracked. The actions that lead to the movement of object are not tracked. However, it is equally important to track the person’s movement as well for detecting any abnormal activities. Also, most of the surveillance systems only record the actions. The classification of abnormal events is generally performed by human intervention where security personnel identify the abnormal events manually. This manual intervention has to be removed and the surveillance system should be intelligent enough to recognize abnormal events on its own and report to the concerned authorities automatically. In the first phase, the proposed system recognizes the regions that have been subject to changes. In the second phase, the system computes the relevant data pertaining the changed region. The data include computing the speed of motion, acceleration, trajectory of movement and accordingly a representation of the current state of the object is provided. In the last stage, the video is examined by comparing the state constraints with the prestandardized constraints. This provides the details of the unusual activities as discussed by Geng-yu & Xue-yin (2010) and Sudo *et al.*(2007). With the help of an already fixed criterion, the outline, movement and also additional data pertaining to the objects have been extorted from image series. The intermediary state representation and replication is performed using Hidden Markov Model (HMM). This state representation retrieved from the image series is compared with a standard action model as discussed in (Matern *et al.*2011, and Bouttefroy *et al.* 2010) and if the resultant values are not equivalent, the action is deemed to be anomalous. Conversely, due to the occurrence of unusual actions, the organized learning process gets disrupted. An extensive range of unusual actions cannot be easily stated because of its intricacy in occurrence and movements as described in (Li & Zhao 2012 and Li *et al.* 2011).

Zhang *et al.*(2005) and Reddy *et al.*(2011), based on certain rules the input video files are separated into certain sections and from every sub section of video the attributes are also extorted. This sub section of video is presented by creating vectors. The grouping technique and even the resemblance measures are applied to those vectors and once it is processed, the sub-video actions might be deemed as irregular only if the sub video had very less resemblance. In real time, it is very complex to recognize the unusual actions. To verify irregular performance like burglary, fight and chasing (Jian-hao&Li 2011) projected a technique which identifies the actions based upon the turmoil of speed and also the path of movement. However, the three unusual actions cannot be differentiated by this technique. (Cheng *et* *al.*2011) projected a method which could identify the cyclic activities and also distinguishes the cyclic motion of a flexible moving object such as an instance of finding the running behaviour of the human. Moreover, to recognize the human running behaviour, a descriptor resulted from cyclic action depiction is utilized. To gratify the real-time presentation sequentially in the surveillance method, a technique has been projected in this work which identifies the unusual running action in surveillance tape in accordance with spatiotemporal constraints. Firstly, the objects presented in the foreground are extorted from video segments associated with Gaussian Mixture representation and also frame subtraction calculation as discussed in (Xin *et al.* 2008 and Chen *et al.* 2010) is performed. The input images are converted into binary images. The nonlinear structures are entailed in extorted foreground object detection algorithm as discussed in the works done by Liao *et al.*(2011), Hu. *et al.*(2011) and Liao *et al.*(2010).

Although various strategies and object handling methods are utilized in real life to promote tracking in crowded area, more difficulties emerge while tracking the scenes in crowd area rather than the small sequences. For instance, It is highly difficult to recognize a targeted object in crowded area due to the size of the targeted object and other scenarios such as occlusion, relative movement of other objects etc. To overcome these difficulties, various outcomes are projected in (Li *et al.* 2011) where the researchers have reinstated those by tracking each unit of the targeted object. Some researchers have projected the algorithm by removing the foreground as suggested by Liao *et al.*(2011). The plan for recognizing and observing the temporal strategy for a crowded area is represented. Initially, various attributes recover the substances of every lead frameworks involved in the operation. Once every object is identified, the Gaussian Mixture algorithm (GMM) is used. In this segment, we describe the recognition of unusual performance in wider aspect, for instance, the unpredicted actions of a person. The researchers try to expand several methods that are usually utilized for video surveillance. If there are any unexpected transformation in scenes like lighting or change in weather and difficulties such as identifying the action are addressed using Gaussian Mixture Model (GMM). Individual events are identified in this series based on identifying the action of every person. Then “vision.BlobAnalysis” object is used for analyzing the individual objects. Before performing blob analysis, segmentation of the objects from the background is performed using GMM and then morphological operations are applied for removing noise and extracting the boxes containing the connected components.

## Limitation of existing system or research gap

### Challenges In Anomaly Detection

Computer Vision Based Anomaly Detection System (CVADS) takes the video feed as input. The major challenges faced while analyzing the video feed are:

1. Automatic anomaly detection without any human intervention
2. Instantaneous alerting with minimal time delay
3. Handling occlusion scenarios
4. Minimal latency
5. Illumination changes

In terms of computational efficiency, other challenges include:

1. Accuracy
2. Minimal response time
3. Minimal space and time complexities

### Gap Analysis on Detecting Anomalies

From the above research works on anomaly detection, it could be found that most of the activities were focused only towards detecting abnormal events. It is highly essential that the anomaly detection is carried out with minimal errors and high degree of accuracy. As part of this research a new spatio-temporal approach towards anomaly detection has been proposed. The salient feature of this approach is that it not only provides high degree of accuracy in detecting anomalies, but also comprises of very minimal errors.

**Gap Analysis on Detecting Anomalies in Crowded Spaces**

Most of the anomaly detection related works were focused towards detecting anomalies in video sequences. But it is highly complex to detect anomalies when the surveillance space is crowded. Also, the human behavior is difficult to track in crowded spaces.

## Problem Statement and Objective

The primary objective of this research work is to design a Computer vision based anomaly detection system using smart anomaly detection algorithms to promote better and smart surveillance system without any human intervention. The secondary objectives of this work include:

1. To provide high degree of accuracy in anomaly detection.
2. To maintain minimal misclassification rate.
3. To improve response time in terms of both anomaly detection and alerting.
4. To improve anomaly detection in occlusion conditions.

## Scope

The four contributions as part of designing the computer vision based anomaly detection system are:

1. A pivotal point based approach for detecting partially occluded or masked faces in videos was developed to detect partially masked faces in video frame. The primary advantage of this approach when compared to the existing approaches is the quick turnaround time in detecting masked or partially occluded faces.
2. A new approach based on spatio-temporal parameters has been designed to detect anomalies in video sequences and alert the anomalies detected. The salient aspect of this approach when compared to the previous approaches is that the anomaly detection is carried out using spatial segmentation of video frames which in turn improves the accuracy of detection with minimal errors.
3. An improved block based strategy using discrete cosine transform co-efficient and entropy has been proposed to detect anomalies in video sequences involving crowded space. The prominent feature of this approach when compared with existing works is its efficiency in detecting anomalies in crowded sequences.
4. A new strategy to detect abandoned objects based on blob analysis has been proposed. The striking feature of this approach is that the abandoned object classification is consistently carried out even under occlusion scenarios.

# Proposed System

## Algorithm

This project is differs from other implementations in quite a few aspects. Firstly you can input any video based data for training. The data will be pre processed and converted into numpy matrix which will be used for training. Converting data into matrix allows for more efficient retention of data. We have also tinkered with the layers and parameters of the model to make it more efficient and accurate. Our changes led to a boost in accuracy compared to baseline. We have also added 4 different ways to deploy the model. A person can use frames extracted from video to get accuracy, use real time video feed, use saved video or directly use a numpy file of the data to be classified and pass it through the model. All these changes makes our project flexible, adaptable and modular

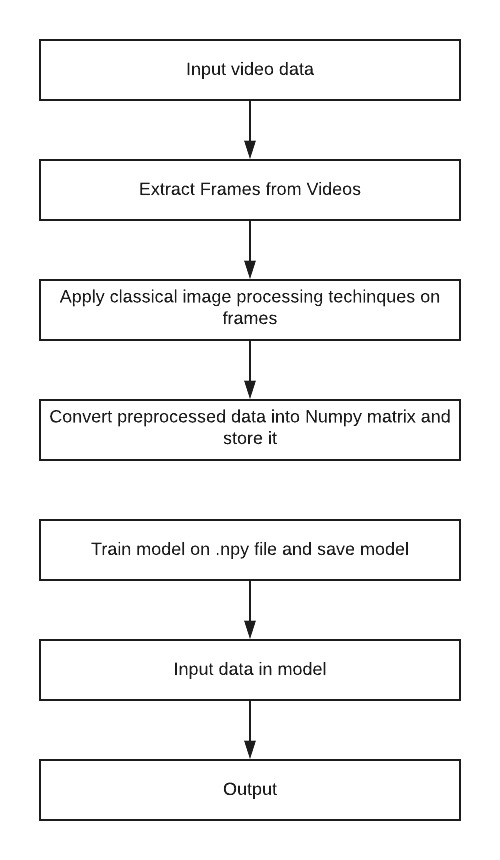


Figure 7 Algorithm of proposed system

## Details of Hardware & Software

### Software Required

1. Visual Studio 2012

Various Dependencies and library based Dependencies such as

1. ffmpeg for Video frame extraction.
2. numpy
3. sklearn
4. keras
5. tensorflow
6. h5py
7. scipy
8. OpenCV

### Hardware required

1. Computer with windows OS: for simulation and training and code compilation purpose
2. Camera: record real time anomaly activity performed by subject
3. Pen drives: transfer data from one device to another device

## Design details

### An Efficient Spatio-Temporal Frequent Object Mining Method to Predict Abnormal Activities

A sequential Pattern Mining (SPM) (Li and Fu, 2014) was proposed to predict the human activity. In SPM, the frequent actions are determined by using Apriori algorithm. But there are issues with the Apriori algorithm with regard to memory consumption as well as time taken to find the results. So in the first phase of this research work, Frequent Pattern-growth (FP-growth) is introduced to determine the recurrent actions in the video surveillance data. Initially, knowledge on space, size and motion association among objects in the video frames is collected and then partial filter technique is applied to track the movement of objects in the video frames. These identified and tracked objects are converted into complex symbolic sequence and the frequent pattern is found from the complex symbolic sequence by using FP-Tree. The frequent itemsets are classified as normal activities, whereas the infrequent itemsets are classified as abnormal activities. The whole process is named as Spatio-Temporal Frequent Object Mining (STFOM).

## Methodology

The method described here is based on the principle that when an abnormal event occurs, the most recent frames of video will be significantly different than the older frames. Inspired by [5], we train an end-to-end model that consists of a spatial feature extractor and a temporal encoder-decoder which together learns the temporal patterns of the input volume of frames. The model is trained with video volumes consists of only normal scenes, with the objective to minimize the reconstruction error between the input video volume and the output video volume reconstructed by the learned model. After the model is properly trained, normal video volume is expected to have low reconstruction error, whereas video volume consisting of abnormal scenes is expected to have high reconstruction error. By thresholding on the error produced by each testing input volumes, our system will be able to detect when an abnormal event occurs. Our approach consists of three main stages:

### Preprocessing

The task of this stage is to convert raw data to the aligned and acceptable input for the model. Each frame is extracted from the raw videos and resized to 227 x 227. To ensure that the input images are all on the same scale, the pixel values are scaled between 0 and 1 and subtracted every frame from its global mean image for normalization. The mean image is calculated by averaging the pixel values at each location of every frame in the training dataset. After that, the images are converted to grayscale to reduce dimensionality. The processed images are then normalized to have zero mean and unit variance. The input to the model is video volumes, where each volume consists of 10 consecutive frames with various skipping strides. As the number of parameters in this model is large, large amount of training data is needed. Following [5]s practice, we perform data augmentation in the temporal dimension to increase the size of the training dataset. To generate these volumes, we concatenate frames with stride-1, stride-2, and stride-3. For example, the first stride-1 sequence is made up of frame {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}, whereas the first stride-2 sequence contains frame number {1, 3, 5, 7, 9, 11, 13, 15, 17, 19}, and stride-3 sequence would contain frame number {1, 4, 7, 10, 13, 16, 19, 22, 25, 28}. Now the input is ready for model training.

### Feature Learning

We propose a convolutional spatiotemporal autoencoder to learn the regular patterns in the training videos. Our proposed architecture consists of two parts spatial autoencoder for learning spatial structures of each video frame, and temporal encoder-decoder for learning temporal patterns of the encoded spatial structures. As illustrated in Figure 1 and 2, the spatial encoder and decoder have two convolutional and deconvolutional layers respectively, while the temporal encoder is a three-layer convolutional long short term memory (LSTM) model. Convolutional layers are well-known for its superb performance in object recognition, while LSTM model is widely used for sequence learning and time-series modelling and has proved its performance in applications such as speech translation and handwriting recognition.

### Regularity Score

Once the model is trained, we can evaluate our models performance by feeding in testing data and check whether it is capable of detecting abnormal events while keeping false alarm rate low. To better compare with [5], we used the same formula to calculate the regularity score for all frames, the only difference being the learned model is of a different kind. The reconstruction error of all pixel values I in frame t of the video sequence is taken as the Euclidean distance between the input frame and the reconstructed frame:

Equation 3.1

where fW is the learned weights by the spatiotemporal model. We then compute the abnormality score sa(t) by scaling between 0 and 1. Subsequently, regularity score sr(t) can be simply derived by subtracting abnormality score from 1:

Equation 3.2

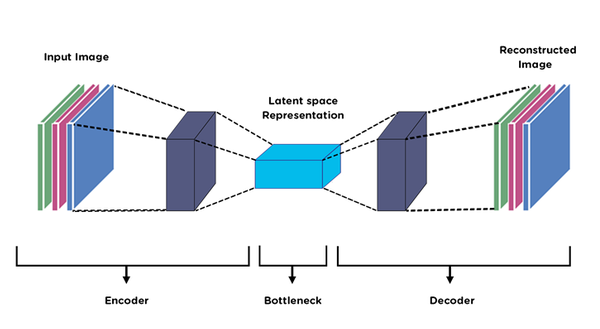
Equation 3.3

# Implementation Details

## Modules & Description

### AutoEncoder

Autoencoder is an unsupervised artificial neural network that learns how to efficiently compress and encode data then learns how to reconstruct the data back from the reduced encoded representation to a representation that is as close to the original input as possible. Autoencoder, by design, reduces data dimensions by learning how to ignore the noise in the data.

[](https://github.com/ShrishtiHore/Anomaly-Detection-in-CCTV-Surveillance-Videos/blob/master/visualizations/auto.png)

#### Autoencoder Components:

Autoencoders consists of 4 main parts:

1. Encoder: In which the model learns how to reduce the input dimensions and compress the input data into an encoded representation.
2. Bottleneck: which is the layer that contains the compressed representation of the input data. This is the lowest possible dimensions of the input data.
3. Decoder: In which the model learns how to reconstruct the data from the encoded representation to be as close to the original input as possible.
4. Reconstruction Loss: This is the method that measures measure how well the decoder is performing and how close the output is to the original input. The training then involves using back propagation in order to minimize the network’s reconstruction loss.
5. Architecture : The network architecture for autoencoders can vary between a simple FeedForward network, LSTM network or Convolutional Neural Network depending on the use case.

Proposed system modules run in following steps

1. **Data Pre-Processing**
2. Download the videos ie; 16 training videos and 12 testing videos and divide it by frames.
3. Images with random objects in the backgorund.
4. Various background conditions such as dark, light, indoor, oudoor, etc.
5. Save all the images in a folder called images and all images should be in .jpg format.
6. Use Argprase parser to add argument to the file names.
7. Divide each and every video into frames and save the frames in a directory separated by the type of anomaly or situation as well as resize the images to scale.
8. Reshape and normalize the images.
9. Clip negative values and remove buffer directory.

**Step 2. Loading the Keras Models**

* 1. Import the three models given below:
  2. Convolutional 3DConvolutional LSTM 2D
  3. Convolutional 3D Transpose
  4. Using Sequential define filters, padding and activation of these models. I am choosing Relu.
  5. Let the optimizer be Adam and metric loss be Categorical Crossentropy.

**Step 3: Training the Model**

1. train.py which runs the training process
2. pipeline\_config\_path=Path/to/config/file/model.config
3. model\_dir= Path/to/training/
4. If the kernel dies, the training will resume from the last checkpoint. Unless you didn’t save the training/ directory somewhere, ex: GDrive.
5. If you are changing the below paths, make sure there is no space between the equal sign = and the path.
6. And use early Callbacks to stop the training if it goes out of hand.

**Step 4: Export the Trained Model**

1. The model will save a checkpoint every 600 seconds while training up to 5 checkpoints. Then, as new files are created, older files are deleted.
2. A file called model.h5 is created which will be used while testing later.
3. Epochs were used as arg.epoch and batch size for training was 32.
4. Another file called training.npy would be created it contains the array form of all the coordinates required while testing. SO here no frozen inference graph or pdtxt file is created.

**Step 5: Testing the Detector**

1. Load the model.h5 file and training.npy file.
2. Test the Videos as: Anomalous Bunch of Numbers, Whether it is normal or abnormal

## Snapshot

### Dataset Description

A Multicamera Avenue Dataset for Abnormal Event Detection is used in this experiment. Avenue dataset collects a large body of human action data using 8 cameras. It consists of 17 action classes such as WalkFall, ClimbLadder, JumpOverGap, PullHeavyObject, Kick, ShotGunCollapse, LookInCar, PickupThrowObject, WalkTurnBack, DrunkWalk, CrawlOnKnees, WaveArms, DrawGraffiti, JumpOverFence, RunStop, SmashObject, and Punch which are performed by 14 actors. Also, a few real time videos are also used in the experiment, which is collected using CCTV cameras that are installed in highly crowded areas for video surveillance purposes.

|  |  |  |  |
| --- | --- | --- | --- |
| **J:\darshan Nemade\projectoutputscreenshot_lastscreenshotshowsatwhatf - Copy\Screenshot (172).png** | **J:\darshan Nemade\projectoutputscreenshot_lastscreenshotshowsatwhatf - Copy\Screenshot (173).png** | **J:\darshan Nemade\projectoutputscreenshot_lastscreenshotshowsatwhatf - Copy\Screenshot (174).png** | **J:\darshan Nemade\projectoutputscreenshot_lastscreenshotshowsatwhatf - Copy\Screenshot (175).png** |
| **J:\darshan Nemade\projectoutputscreenshot_lastscreenshotshowsatwhatf - Copy\Screenshot (176).png** | **J:\darshan Nemade\projectoutputscreenshot_lastscreenshotshowsatwhatf - Copy\Screenshot (177).png** | **J:\darshan Nemade\projectoutputscreenshot_lastscreenshotshowsatwhatf - Copy\Screenshot (178).png** | **J:\darshan Nemade\projectoutputscreenshot_lastscreenshotshowsatwhatf - Copy\Screenshot (179).png** |
| **J:\darshan Nemade\projectoutputscreenshot_lastscreenshotshowsatwhatf - Copy\Screenshot (180).png** | **J:\darshan Nemade\projectoutputscreenshot_lastscreenshotshowsatwhatf - Copy\Screenshot (181).png** | **J:\darshan Nemade\projectoutputscreenshot_lastscreenshotshowsatwhatf - Copy\Screenshot (182).png** | **J:\darshan Nemade\projectoutputscreenshot_lastscreenshotshowsatwhatf - Copy\Screenshot (183).png** |

Figure 8 Frame Sequence of a segmented video

# Testing

## Testing

Implemented system is subjected for the test various datasets available on internet such as avenue data set, [UCSD Anomaly Detection Dataset](http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm), [University of Minnesota crowd activity datasets](http://mha.cs.umn.edu/Movies/), [Anomalous Behavior Data Set](http://vision.eecs.yorku.ca/research/anomalous-behaviour-data/), [Virat video dataset](http://www.viratdata.org/) and [McGill University Dominant and Rare Event Detection Data](http://www.cim.mcgill.ca/~javan/index_files/Dominant_behavior.html) is used for test

# C:\Users\jaybh\OneDrive\Pictures\Screenshots\Screenshot (1).png

Figure 9 installing required python library

Step first install tqdm library for the simulation with help of pip python then after installing tqdm run python code with help of run key in visual studio 2012 as shown in figure 9 and figure 10.

# C:\Users\jaybh\OneDrive\Pictures\Screenshots\Screenshot (2).png

Figure 10 run python code

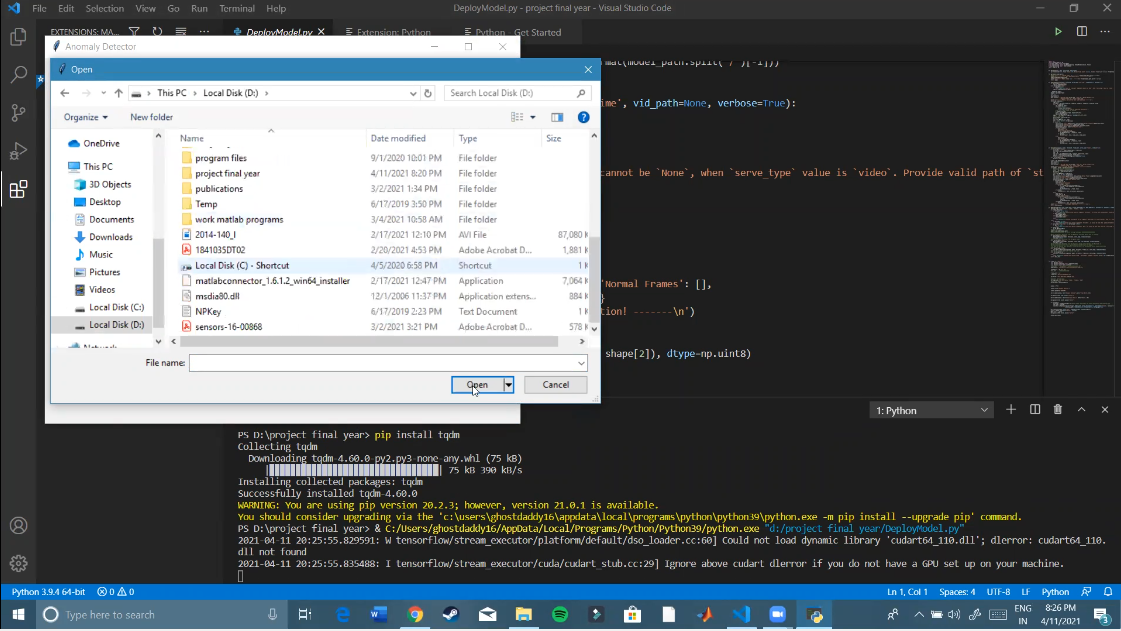
When anomaly detector dialog box is open, then with help of open key present in dialog box open directory where test data is stored in local computer as shown in figure 11 and figure 12.

# C:\Users\jaybh\OneDrive\Pictures\Screenshots\Screenshot (3).png

Figure 11 User interface to open test data



Figure 12 User interface define data location



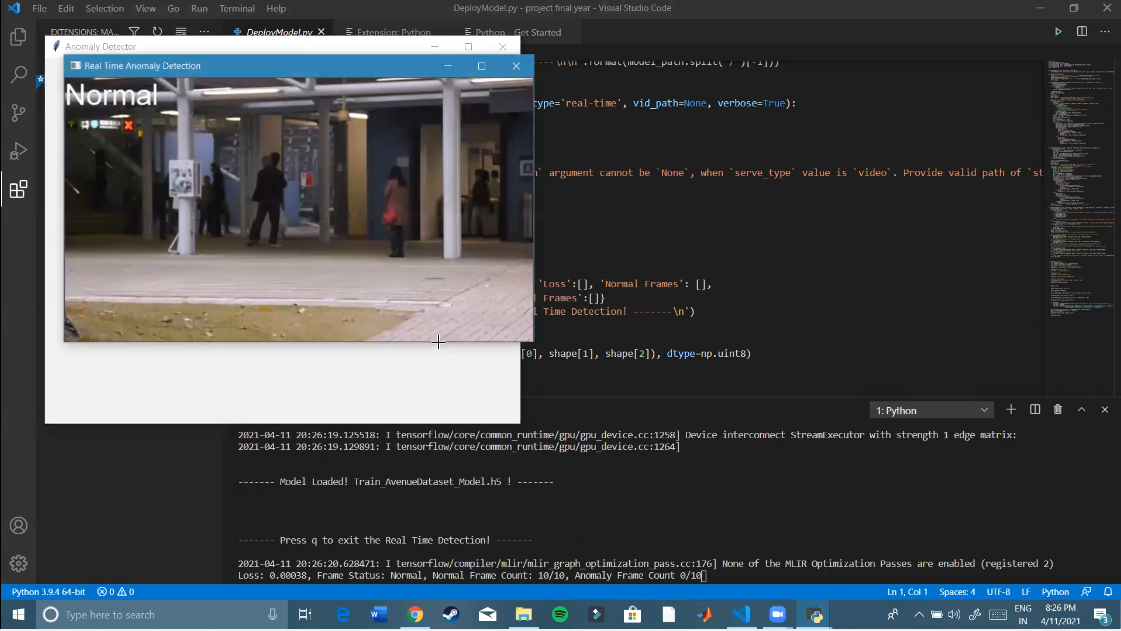


Figure 13 Test data under analysis

Figure 13 shows test data under analysis, normal frame is analyzed in screen short then final output is shown in the form of numbers in figure 14

### Output Of Anomaly Detection

Anomaly detection have two types of output techniques:

1. Scores: Scoring techniques assign an anomaly score to each instance in the test data depending on the degree to which that instance is considered an anomaly as shown in figure 14.
2. Labels: Techniques in this category assign a label (normal or anomalous) to each test instance as shown in figure 15 to figure 18.

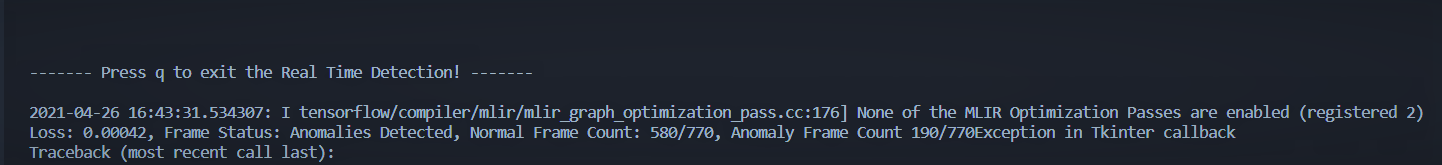


Figure 14 Real time detection result

## Results



Figure 15 Man throwing bag in air



Figure 16 Small Boy Jumping



Figure 17 Man Running Man



Figure 18 Running In Opposite Direction

# Results & Analysis

## Performance Metrics

The performance of the existing and the proposed methods for human activity prediction in the VSS are analyzed on the basis of accuracy, precision, recall, information gain ratio, and true positive rate.

### True Positive Rate

True Positive Rate is described as the rate of human abnormal activity predicted as human abnormal activity in videos.

### Accuracy

It is the fraction of true results of human activity prediction (true positive and true

negative) among the total number of cases analyzed. It is calculated as,

Equation .

where, if the class label is positive and the human abnormal activity prediction outcome is positive, then it is TP. If the class label is negative and the human abnormal activity prediction outcome is negative, then it is TN. If the class label is negative and the human abnormal activity prediction outcome is positive, then it is FP. If the class label is positive and the human abnormal activity prediction outcome is negative, then it is FN.

### Precision

It is the fraction of the number of suspicious faces that are appropriately recognized to the sum of the count of correctly recognized suspicious faces and the wrongly recognized suspicious faces.

Equation .

### Recall

It is the fraction of the number of suspicious faces that are appropriately recognized to the sum of the count of correctly recognized suspicious faces and the wrongly recognized non-suspicious faces.

Equation .

### Information Gain Ratio

It is defined as a quantity of knowledge obtained during the prediction of human activities in the videos.

### Regularity Score

Once the model is trained, we can evaluate our models performance by feeding in testing data and check whether it is capable of detecting abnormal events while keeping false alarm rate low. To better compare with [5], we used the same formula to calculate the regularity score for all frames, the only difference being the learned model is of a different kind. The reconstruction error of all pixel values I in frame t of the video sequence is taken as the Euclidean distance between the input frame and the reconstructed frame:

Equation .

where fW is the learned weights by the spatiotemporal model. We then compute the abnormality score sa(t) by scaling between 0 and 1. Subsequently, regularity score sr(t) can be simply derived by subtracting abnormality score from 1:

Equation .5

Equation .6

The unusual circumstances comprise of various volunteers suddenly dancing, running and pushing in a crowded place. Overall there are six kinds of unusual or abnormal circumstances which take place in 12000 frames of video series. The usual screening quality for a video surveillance is 720 576 by 29 frames per second, which is the spatial motion of a novel video frame. Moreover, a different strategy which is projected by the authors in (Jian-hao&Li 2011) is contrasted with the presentation of this strategy and the resulting outcome is evaluated. Each frame is divided into four segments in our present research work. The number of segments per frame is customizable. The entropy of DCT coefficients is computed for every segment also the median rate for the first 500 frames is calculated. In relation to this research and analysis, the threshold median entropy is positioned to 3 times than the median rate to categorize the abnormal happenings. If there are any unusual happenings in any of the segment, in such cases an unusual indicator raises for the entire structure. Table 1 represents the set of all frames extracted from a segmented video. The duration of the segmented video is one minute and is customizable.

## Results & Analysis

The number of frames extracted from a segmented video is 95. Some of the frames are given in Figure 5.4. Initially all the images are considered as a normal images and the process started. In each frame, following object detection, the entropy is calculated from the DCT values and compared with the threshold values already calculated and stored in a database from ground truth images. Ground truth images are suggested from programming experts. The variation of the entropy and matching with the ground truth objects helps to classify the objects are normal or abnormal. Out of six abnormalities, five abnormalities are identified using the proposed approach. Also the performance of the proposed approach is evaluated by computing the computational time complexity and accuracy in classification. To do that, time taken for video to segments, segment into frames, frames into objects and object classification is computed in the experiment and the obtained results is shown in table 1 to table 3

Table 1 System Accuracy for Different Data Samples Taken

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Data (video) | Frames Classification | | | | Total frames | Accuracy in Percentage |
| True Positive | False Positive | True Negetive | False Negetive |
| Boy Jumping | 435 | 430 | 65 | 70 | 500 | 0.5035 |
| Man Running | 16 | 10 | 314 | 330 | 330 | 0.49253731 |
| Man Running Opposite | 634 | 630 | 136 | 140 | 770 | 0.5024 |
| Throwing Bag | 563 | 560 | 537 | 540 | 1100 | 0.5045 |

**Table 2 System Precision for Different Data Samples Taken**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Data (video) | Frames Classification | | | | Total frames | Precision |
| True Positive | False Positive | True Negetive | False Negetive |
| Boy Jumping | 435 | 430 | 65 | 70 | 500 | 0.50289 |
| Man Running | 16 | 10 | 314 | 330 | 330 | 0.615385 |
| Man Running Opposite | 634 | 630 | 136 | 140 | 770 | 0.501582 |
| Throwing Bag | 563 | 560 | 537 | 540 | 1100 | 0.501336 |

Table 3 System Recall for Different Data Samples Taken

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Data (video) | Frames Classification | | | | Total frames | Recall |
| True Positive | False Positive | True Negetive | False Negetive |
| Boy Jumping | 435 | 430 | 65 | 70 | 500 | 0.861386 |
| Man Running | 16 | 10 | 314 | 330 | 330 | 0.046243 |
| Man Running Opposite | 634 | 630 | 136 | 140 | 770 | 0.819121 |
| Throwing Bag | 563 | 560 | 537 | 540 | 1100 | 0.510426 |

Also, the average time taken for the entire system model to compute unusual actions is 74.25 ms which is inclusive all intermediate stages. In terms of classification the accuracy is calculated and the obtained results are given in table-1. From the results it’s clear that percentage of accuracy is 50%. Table 1 shows the comparison of classification accuracy of proposed system with other contemporary approaches. It can be found that the proposed system has higher classification accuracy than other contemporary methods.

Table 4 Accuracy Comparison for Different data

Table 5 Precision Comparison for Different data

**Table 6 Recall Comparison for Different data**

# Conclusion and Future Scope

We have successfully applied deep learning to the challenging video anomaly detection problem. We formulate anomaly detection as a spatiotemporal sequence outlier detection problem and applied a combination of spatial feature extractor and temporal sequencer ConvLSTM to tackle the problem. The ConvLSTM layer not only preserves the advantages of FC-LSTM but is also suitable for spatiotemporal data due to its inherent convolutional structure. By incorporating convolutional feature extractor in both spatial and temporal space into the encoding-decoding structure, we build an end-to-end trainable model for video anomaly detection. The advantage of our model is that it is semi-supervised the only ingredient required is a long video segment containing only normal events in a fixed view. Despite the models ability to detect abnormal events and its robustness to noise, depending on the activity complexity in the scene, it may produce more false alarms compared to other methods. For future work, we will investigate how to improve the result of video anomaly detection by active learning having human feedback to update the learned model for better detection and reduced false alarms. One idea is to add a supervised module to the current system, which the supervised module works only on the video segments filtered by our proposed method, then train a discriminative model to classify anomalies when enough video data has been acquired.

## Applications Of Anomaly Detection

1. **Intrusion detection:** Intrusion detection refers to detection of malicious activity. The key challenge for anomaly detection in this domain is the huge volume of data. Thus, semi-supervised and unsupervised anomaly detection techniques are preferred in this domain.
2. **Fraud Detection:** Fraud detection refers to detection of criminal activities occurring in commercial organizations such as banks, credit card companies, insurance agencies, cell phone companies, stock market, etc. The organizations are interested in immediate detection of such frauds to prevent economic losses.
3. **Medical and Public Health Anomaly Detection:** Anomaly detection in the medical and public health domains typically work with patient records. The data can have anomalies due to several reasons such as abnormal patient condition or instrumentation errors or recording errors. Thus the anomaly detection is a very critical problem in this domain and requires high degree of accuracy.
4. **Industrial Damage Detection:** Such damages need to be detected early to prevent further escalation and losses.
5. **Image Processing:** Anomaly detection techniques dealing with images are either interested in any changes in an image over time (motion detection) or in regions which appear abnormal on the static image. This domain includes satellite imagery.
6. **Anomaly Detection in Text Data:** Anomaly detection techniques in this domain primarily detect novel topics or events or news stories in a collection of documents or news articles. The anomalies are caused due to a new interesting event or an anomalous topic.
7. **Sensor Networks:** Since the sensor data collected from various wireless sensors has several unique characteristics.

# References

1. Zhao Bin, Li Fei , and Xing E. P. “Online detection of unusual events in videos via dynamic sparse coding,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 3313–3320, 2011.
2. Cong Yang, Yuan Junsong, and Liu Ji “Sparse reconstruction cost for abnormal event detection,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 3449–3456, 2011.
3. Chen Zhu, and Saligrama V. “Video anomaly detection based on local statistical aggregates,” Computer Vision and Pattern Recognition, pp. 2112–2119, 2012.
4. Zhou Xu Gang, Zhang Li Qing “Abnormal Event Detection Using Recurrent Neural Network,” International Conference on Computer Science and Applications,pp. 222–226, 2015.
5. Sabokrou M., Fathy M., and Hoseini M. “Video anomaly detection and localisation based on the sparsity and reconstruction error of auto-encoder,” Electronics Letters, vol. 52, no. 13, pp. 1122–1124, 2016.
6. Xu Dan, Ricci Elisa, Yan Yan, Song Jingkuan, and Sebe Nicu “Learning Deep Representations of Appearance and Motion for Anomalous Event Detection,” BMVC , 2015.
7. Hasan Mahmudul, Choi Jonghyun, Neumann Jan, Roychowdhury Amit K., and Davis Larry S. “Learning Temporal Regularity in Video Sequences,” Computer Vision and Pattern Recognition , pp. 733–742,2016.
8. Ravanbakhsh Mahdyar, Nabi Moin, Sangineto Enver, Marcenaro Lucio, Regazzoni Carlo, and Sebe Nicu “Abnormal Event Detection in Videos using Generative Adversarial Nets,” International Conference on Image Processing, 2017.
9. Yong Shean Chong, and Yong Haur Tay “Abnormal Event Detection in Videos Using Spatiotemporal Autoencoder,” International Symposium on Neural Networks, pp. 189–196, 2017.
10. Patraucean Viorica, Handa Ankur, and Cipolla Roberto “Spatio-temporal video autoencoder with differentiable memory,” Computer Science, vol. 58, no. 11, pp.2415– 2422, 2015.
11. Sutskever, Ilya, Vinyals, Oriol, Le, and Quoc V, “Sequence to sequence learning with neural networks,” In Advances in Neural Information Processing Systems, vol. 4, pp. 3104–3112, 2014.
12. Srivastava, Nitish, Mansimov, Elman ,Salakhutdinov, and Ruslan “Unsupervised Learning of Video Representations using LSTMs,” International Conference on Machine Learning, pp. 843–852, 2015.
13. Ji Yangfeng, Cohn Trevor, Kong Lingpeng, Dyer Chris, and Eisenstein Jacob “Document Context Language Models,” Computer Science, 2016.
14. He Kaiming, Zhang Xiangyu, Ren Shaoqing, and Sun Jian “Deep Residual Learning for Image Recognition,” Computer Vision and Pattern Recognition , pp. 770– 778, 2015.
15. Huang Gao, Liu Zhuang, Weinberger Kilian Q, and Laurens Van Der Maaten “Densely Connected Convolutional Networks,” Computer Vision and Pattern Recognition, 2016.
16. Shi Xingjian, Chen Zhourong, Wang Hao, Yeung Dit Yan , Wong Wai Kin, Woo Wang Chu “Convolutional LSTM network: A machine learning approach for precipitation nowcasting,” NIPS, pp. 802–810, 2015.
17. Graves, and Alex “Generating Sequences With Recurrent Neural Networks,” Computer Science, 2013.
18. Jefferson Ryan Medel, Andreas E. Savakis “Anomaly Detection Using Predictive Convolutional Long ShortTerm Memory Units,” CoRR abs/1612.00390, 2016.
19. Y. Kozlov, and T. Weinkauf “Persistence 1D: Extracting and filtering minima and maxima of 1d functions,” http://people.mpiinf.mpg.de/ weinkauf/notes/persistence1d.html
20. V.Mahadevan, W.Li, V.Bhalodia, and N.Vasconcelos “Anomaly detection in crowed scenes,” IEEE International Conference on Signal Processing, pp. 1975– 1981,2010.
21. C.Lu, J.shi, and J.Jia “Anomaly event detection at 150fps in matlab,” IEEE International Conference on Computer Vision, pp. 2720–2727, 2013.
22. Amit Adam, Ehud Rivlin, Ilan Shimshoni, and David Reinitz “Robust Real-Time Unusual Event Detection using Multiple Fixed-Location Monitors,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 3, pp. 555–560, 2008.
23. Wang Tian, and Snoussi H. “Histograms of optical flow orientation for abnormal events detection,” IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS 2013), vol. 5, no. 9, pp. 13–18, 2013.