**VIDEO BASED ANOMALY DETECTION USING MACHINE LEARNING**

**SAHANA KUMARI B**

# Abstract

Machine learning anomaly detection systems help detect anomalies such as accidents. Detecting video anomalies is important, but an open issue. Today, it is very difficult to detect anomalous activity in a video in one day due to the ambiguity of the anomaly. The system automatically detects anomalies through image analysis from surveillance video. This system can track anomalous events in each frame and generate notifications about them. In our project we will implement software such that it is capable to detect any anomalies in the video and generate the notification. We have introduced a new unsupervised data mining anomaly detection algorithm that is suitable for large-scale multivariate time series analysis and can detect anomalous regions. The proposed K-means algorithm is relatively time efficient and superior to existing anomaly detection methods. The system only needs a data set of video images and is supplied as input to the anomaly detection model. The trained model automatically detects anomalies in time series of video frames and pop up the anomaly footage. Object tracking and anomaly detection can be improved by applying machine learning techniques. Our system ensures that anomalies are detected without data loss. This system is user-friendly and highly efficient in eliminating time-consuming manual inspections and detecting anomalous events. The dataset used here is a UCSD anomaly detection dataset with a fixed camera mounted at an altitude overlooking a pedestrian walkway. Common anomalies include cyclists, skaters, small carts, people walking across the sidewalk or in the grass surrounding the sidewalk.

## Chapter 1

**Introduction**

## Project Introduction

Today detecting Anomalous activity from the video is difficult due to the ambiguity of the anomaly. All important public places such as public gardens, supermarkets, university campuses and shopping malls are under video surveillance. It is very important to be secure and to monitor anomalous activity in these locations. CCTV has been widely used for monitoring, recording situation in a surveillance system. CCTV cameras are mostly used for a post video forensic process by allowing finding situation from the previous scene. CCTV cameras feed should be manually observed by human beings to find the abnormal event. In this we will be using K-means algorithm to detect the anomalies. In K-means algorithm number of observation or the dataset are divided into clusters and that clusters are represented as K-clusters. K-clusters have their own feature. K-clusters are put into the group based on the data points. As we know the K-means algorithm is unsupervised learning algorithm.

Many pattern recognition methods strive towards deriving models from complex and noisy data. Such models try to describe the prototypical normal behavior of the system being observed, which is hard to model manually and whose state is often not even directly observable, but only reflected by the data. They allow reasoning about the properties of the system, predicting unseen data, and assessing the “normality” of new data.In such a scenario, any deviation from the normal behavior present in the data is distractingand may impair the accuracy of the model. An entire arsenal of techniques has therefore been developed to eliminate abnormal observations prior to learning or to learn models in a robust way not affected by a few anomalies. Such practices may easily lead to the perception of anomalies as being intrinsically bad and worthless. Though that is true for random noise and erroneous measurements, there may also be anomalies caused by rare events and complex processes. Embracing the anomalies in the data and studying the information buried in them can therefore lead to a deeper understanding of the system being analyzed and to the insight that the models hitherto employed were incomplete or in the case of non- stationary processes outdated. A well-known example forthis is the discovery of the correlation. The use of anomaly detection techniques is not

limited to outlier removal as a pre-processing step. Anomaly detection also is an important task, since only the deviations from normal behavior are the actual object of interest in many applications. Besides the scenario of knowledge discovery mentioned above, fraud detection intrusion detection in cyber-security, fault detection in industrial processes, anomaly detection in healthcare and early detection of environmental disasters are other important examples. Automated methods for anomaly detection are especially crucial nowadays, where huge amounts of data are available that cannot be analyzed by humans. In our project we will implement software such that it is capable to detecting any anomalies in the video and generate the notification. Applying machine learning techniques can improve object tracking and anomaly detection.

## Problem Description

Today, most public places, such as supermarkets, public gardens, shopping malls, and university campuses, are under video surveillance. Such sites should take basic security measures and monitor for anomalous anomalies. Manual behavior detection operations are time consuming, tedious, and inefficient, especially in crowded areas. Need an automated system to streamline operational issues and immediately alert human operators.

## Application

College or University Surveillance: In a college environment anomalous activity may include students involving in a fight, abuse, ragging or dealing drugs.

Crowded Public Places Surveillance: An anomalous human activity recognition system can be used in banks, airports, railway stations, bus-stops, shopping malls and so on where there is a need for continuous monitoring of surveillance that is practically not possible.

Road -Accident Detection : The death rate caused by road accidents is gradually increasing. One possible way to help avoid death on road is to get timely help to the victims.

Patients’ Surveillance in Hospitals : Patients who are unconscious can be monitored using the Activity recognition system to detect any kind of motion rather than having an attender next to the patient for constant observation.

## Chapter 2

**Literature Review**

## General Introduction

Literary studies are an important activity that must be carried out while gathering information on a given topic. It helps you get the information and ideas you need for your work. The next section describes related tasks and issues in the area of energy consumption prediction and analysis in computer labs using machine learning algorithms.

## Literature Survey

Adal Nadjara Toosi and Mohsen Kahani [1] in the year 2010 Soft Computing approach for Anomaly detection. This method used the ANFIS(Adaptive neuro fuzzy interface system) network. This network is a kindof artificial network that applies logicalrulesto the knowledgebase toextract new information. It works with both neural networks and fuzzy logic.

Shekhar R. Gaddam [2] in the year 2017 “Hybrid Anomaly-based Detection Approach”. Here in this method Shekhar R.Gaddam introduced a method which helps detect anomalies that occurs in an environment by combining the Iterative Dichotomiser3 Decision Tree and K-means algorithm. Decisions from the dataset use the Dichotomiser3 iterative decision tree. The Dichotomiser3 iteration decision tree starts at root node S and iterates from one variable to another when the algorithm reaches the next step from the start step. The Dichotomiser3 Iterative Decision Tree selects unused attributes in the specified tree and calculates the entropy H (S) or information gain IG (S).

Latifur Khan [3] published in the year2016 Support Vector Machine based Anomaly Detection. The purpose of this white paper was to detect anomalies, especially in contracts with large datasets. This method used DGSOT.

Yu Guan [4] published on 2013 “Anomaly Detection System using K-means algorithm”.

Yu Guan was the first person to introduce this anomaly detection method using the K- Means algorithm. Different numbers of observations or datasets are divided into several clusters by the K-Means algorithm. Clusters are represented by K clusters. Each cluster has its own characteristics, based on which data points are grouped. Each cluster has a focus, which is the basic element of the cluster. This focal point has some characteristics (valuable information) that separate this particular cluster from the other clusters created. The K- means algorithm is unsupervised learning algorithm.

## Summary

Real-world problems have the following goals, such as time complexity, bad weather, real- world dynamics, inclusion, and object overlap. The existing method handled the problem individually. There is no way to treat all targets as a single suggestion feature. In order to handle effective real-time intelligent crowd video analysis, the method must be able to provide a solution to all these problems. Traditional methods cannot produce efficient and economical solutions in a limited period of time. The availability of high-performance computing resources such as GPUs enables the implementation of deep learning-based solutions for high-speed processing of big data. Existing deep learning architectures or models can be combined by removing unnecessary features, including great features.

## Chapter 3

**Problem Formulation**

## General

Before trying to solve a problem, we first need to formulate or define the problem. It is important to clearly define the problem you want to solve. Problem formulation is the act of addressing a problem, identifying the cause of the problem, and identifying a solution.

## Problem Statement

Time and efficiency are today's top priorities. To overcome the big problem of time consumption problem due to manual observation of cctv videos and server by human beings to detect the abnormal events. The proposed system is developed for automatic detection of abnormal events and to reduce the time consumption by using k- means algorithm.

## Objectives of the Present Study

The objectives of the proposed project are as follows:

1. To develop machine learning based spatio-temporal Anomaly detection.
2. To design and implement K means algorithm for abnormal and normal event cluster. 3.Unsupervised and Data mining algorithm is used for event detection.

## Summary

Unusual activity and anomaly detection has its immense utilization in many areas which are related to find the unexpected activities or variations in normal patterns of activities so that appropriate actions in against can take place on time. Unusual activity and anomaly detection can have great implications in various fields with some useful applications which are listed below:

* Security related applications such as Defense, Military, checking at airports.
* Healthcare, Visual Surveillance through Content Based Retrieval
* Network Security Threat Detection
* Assisting the sick and disabled patients
* Location based Services
* Home based rehabilitation
* For Ambient Assisted Living Environment

We have introduced a new unsupervised anomaly detection algorithm that can detect anomalous regions not only in time but also in spatiotemporal data from different domains, suitable for analysis of large multivariate time series. The proposed MDI algorithm is relatively time efficient and superior to existing anomaly detection techniques. Thanks to an efficient implementation and a new interval suggestion technique that excludes non- essential parts of the data from detailed analysis.

## Chapter 4

**Requirements and Methodology**

## Requirements

The proposed project consists of following modules:

Hardware requirements Software requirements

#### Hardware Requirements

The hardware requirements for the proposed project are depicted in the table below:

##### Table 4.1: Hardware requirements

|  |  |  |
| --- | --- | --- |
| **Sl. No** | **Hardware/Equipment** | **Specification** |
| 1 | laptop | * Processor Intel Core i5 * RAB 8GM * Hard disk 1TB * Input output console for interaction |

#### Software Requirements

The software requirements for the proposed project are depicted in the table below:

##### Table 4.2: Software requirements

|  |  |  |
| --- | --- | --- |
| **Sl. No** | **Software** | **Specification** |
| 1 | Operating System | Windows 10 |
| 2 | Scripting | Python 3.7 |
| 3 | Language | PYTHON ML |

## Methodology Used

Methodology is a systematic and theoretical examination of the methods applied to the field of study. This includes a theoretical analysis of methods and principles related to the field of knowledge. It usually includes concepts such as paradigms, theoretical models, phases, quantitative or qualitative methods.

* We have introduced a new unsupervised data mining anomaly detection algorithm that is suitable for large-scale multivariate time series analysis and can detect anomalous regions.
* The proposed K-Means algorithm is relatively time efficient and superior to existing anomaly detection methods.
* The proposed system only requires a dataset of video images and is supplied as input to the anomaly detection model.
* The proposed system eliminates manual time-consuming inspection.
* The trained model automatically detects anomalies in time series of video frames and pop up the anomaly footage.
* The proposed system makes sure of detecting of anomalies without any data loss.
* The system is user-friendly and much efficient in detecting abnormal events.

## Chapter 5

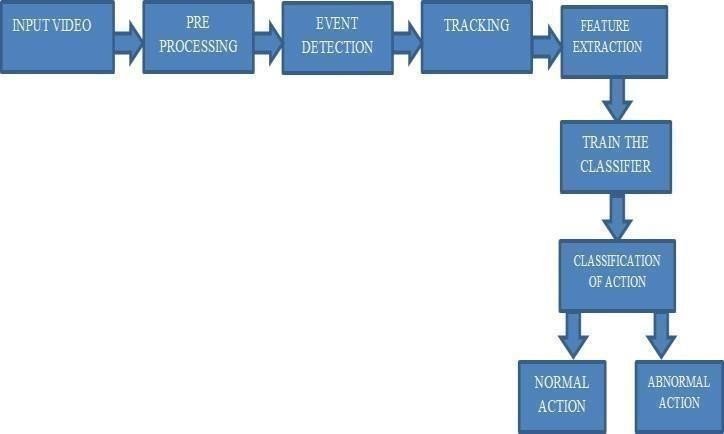
**System Design**

## System Design

System design is an important phase of software or system development. System design can be defined as a way to define the software or the various modules that a system needs to meet all its requirements.

#### Architecture of proposed system

The system design shows the overall design of the proposed human emotion detection system. This is shown in Figure 5.1 below.

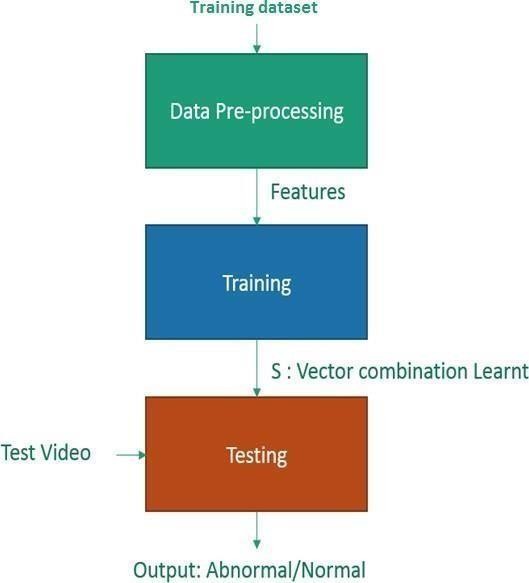


##### Figure 5.1: Design of the proposed anomaly detection system

The input video preprocessing is performed. It generates video frames and from that all occurred events are detected. These frames feature extracted. The model get trained by K- means algorithm. The classifier classifies the all events into normal event and abnormal event.

#### Sequence Diagram

The flowchart of the proposed system is shown in Figure 5.2 below.

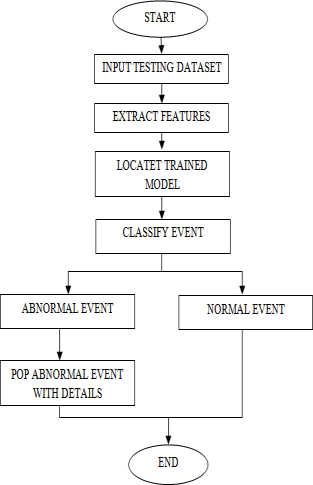
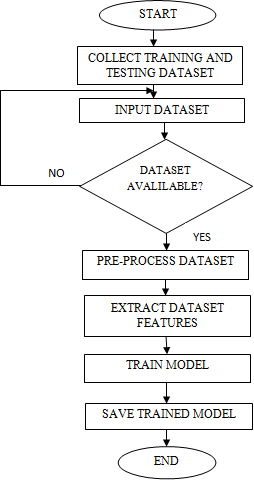


##### Figure 5.2 Sequence Diagram of proposed anomaly detection system

The Data pre-processing step includes importing the video frames and making it reading for training. It also involves feature extraction which is the input to the training algorithm. Our module takes the dataset as input for the training, applies the algorithm and starts extracting the features of all input dataset. Testing module involves again the pre- processing stage where it normalizes all the testing input dataset and feed forward it to the predicting model.

#### System Flowchart

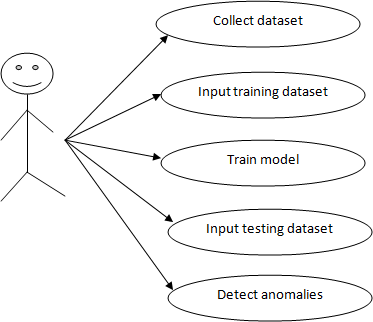
The flow chart of the proposed system is shown in Figure 5.3 below.



##### Figure 5.3: Flowchart of the proposed system

#### Use Case Diagram

The use case diagram of the proposed system is shown in Figure 5.4 below.



##### Figure 5.4: Use Case Diagram

Use case diagrams are UML dynamic or behavioral diagrams. Use case diagrams model the functionality of the system based on actors and use cases. Use cases are a set of actions, services, and functions that a system needs to perform.

## Disadvantages of Existing System

* Existing system consists of manual monitoring.
* One should sit back and inspect/watch complete video of footages.
* Need careful inspection.
* Much time consuming.
* Might miss some anomalies in footages.

## Proposed System

* We have introduced a novel unsupervised and data mining algorithm for anomaly detection that is suitable for analyzing large multivariate time-series and can detect anomalous regions.
* The proposed K-means algorithm outperforms existing anomaly detection techniques, while being comparatively time efficient.
* The proposed system just requires a dataset of video frames and is fed as input to the anomaly detecting model.
* The proposed system eliminates manual time consuming inspection.
* The trained model automatically detects anomalies in time series of video frames and pop up the anomaly footage.
* The proposed system makes sure of detecting of anomalies without any data loss.
* The system is user-friendly and much efficient in detecting abnormal events.
* The proposed model is flexible in multiple domains for many climate change detection, crime detection and many abnormal events from the video frames.

## Chapter 6

**Implementation**

The proposed project consists two implementation modules.

## Pseudo code

Pseudo code is a step-by-step draft of code that can be gradually transcribed into a programming language.

Input:

D={t1,t2, tn} //Set of element

K //Number of desired clusters Output:

K //Set of clusters K-Means algorithm:

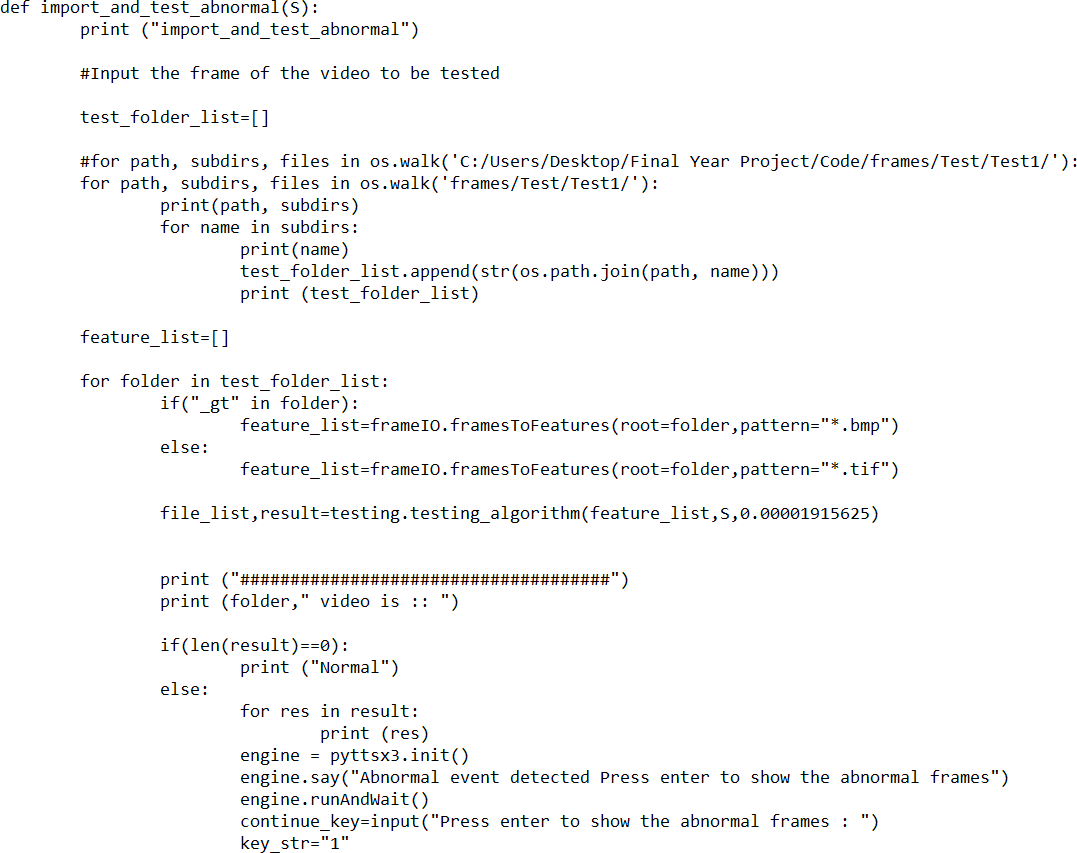
Assign initial values for m1,m2, mk repeat

assign each item ti to the clusters which has the closest mean; calculate new mean for each cluster;

until convergence criteria is met;

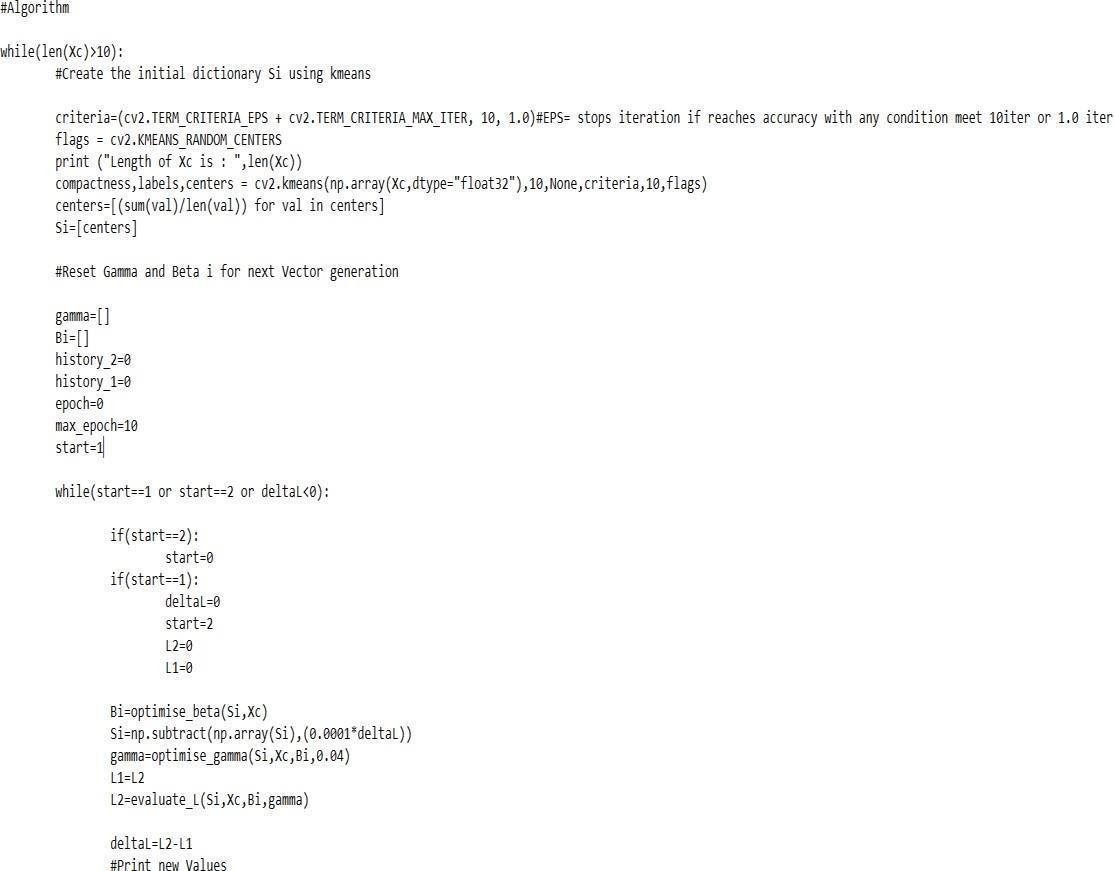
## Implementation Code

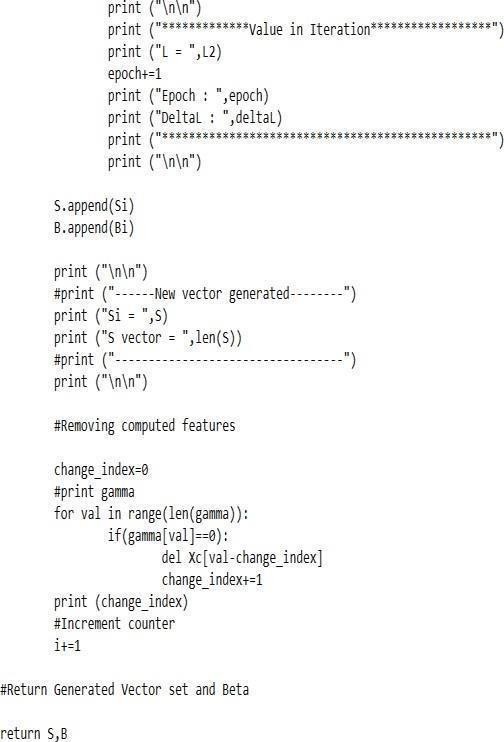


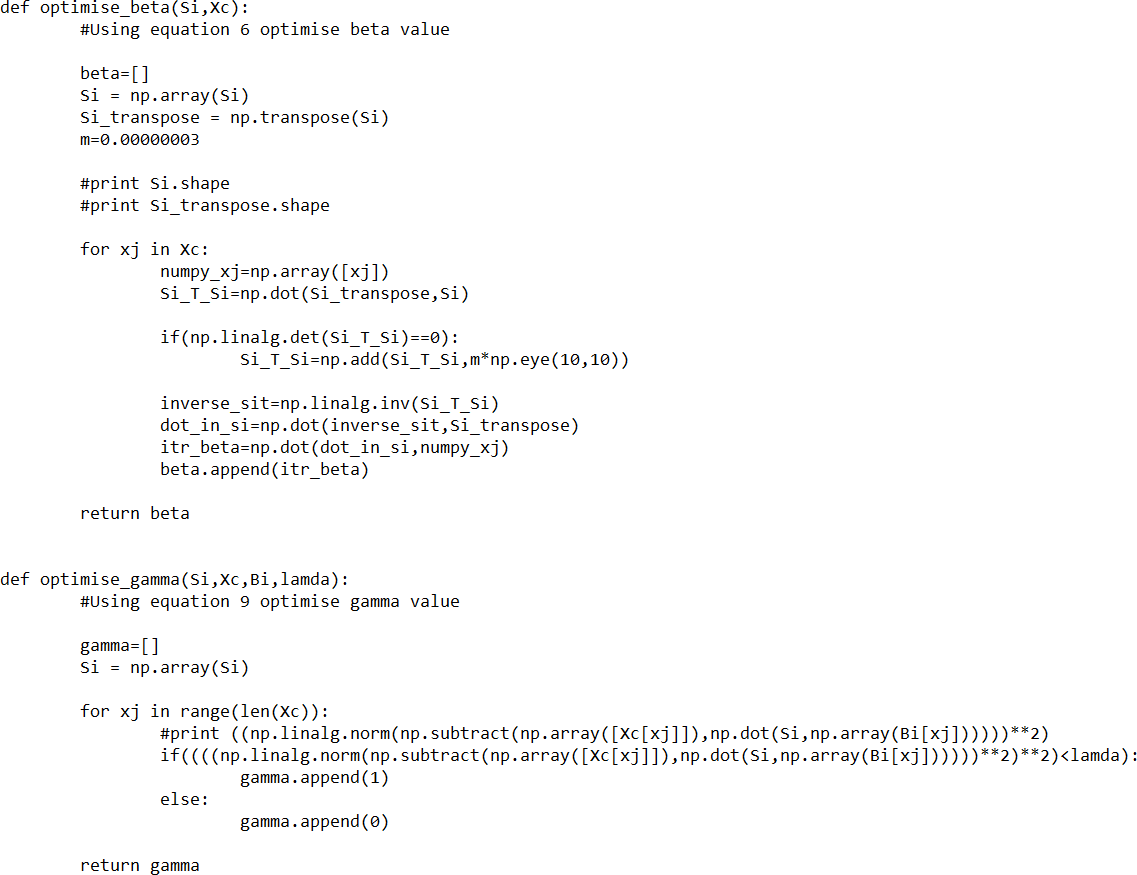


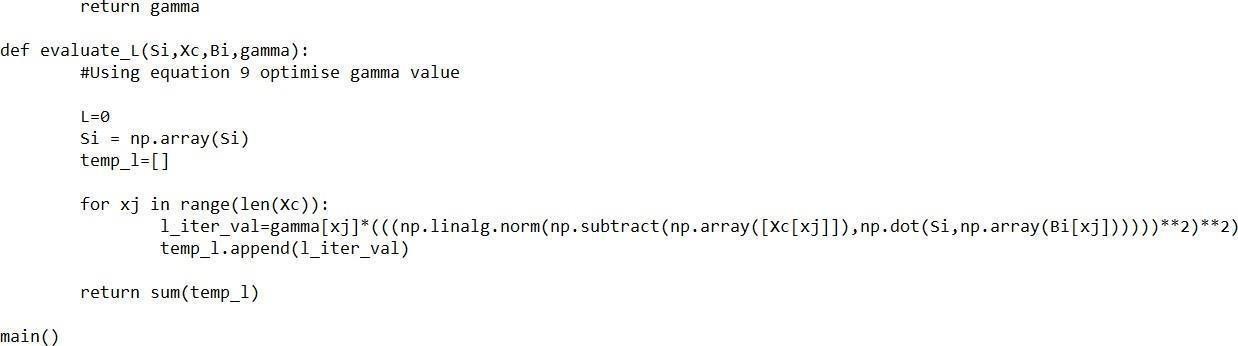












## Chapter 7

**System Testing, Results and Discussion**

## System Testing

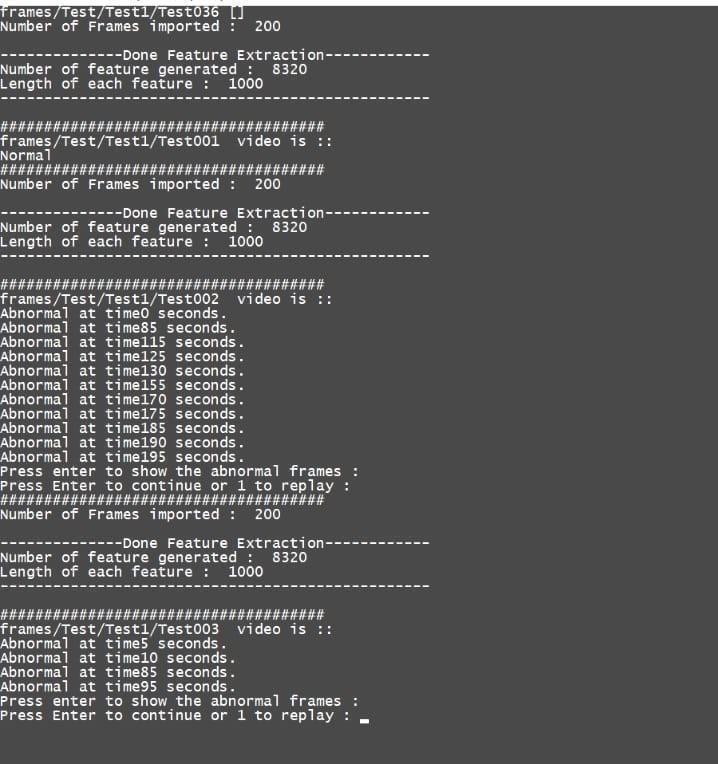
The purpose of the test is to find bugs. Testing seeks to find all possible flaws or weaknesses in the work product. Provides a way to validate the functionality of components, subassemblies, assemblies, and / or final products. This is the process of running software to ensure that the software system meets its needs and user expectations and does not fail in an unacceptable way.

software testing ensures that the developed system works according to the original goals and requirements. Once the program is created and the data structures associated with the document are designed, the software testing process begins. Software testing is essential to fix bugs. Otherwise you will not be able to complete the project.

## Result Analysis

The input video USCD dataset is provided to this system. The feature extraction is done and displays number of frames imported. Then it shows Number of feature generated and length of that each frame. If no anomalous activities are detected prints normal. The figure

7.1 shows the system processing in command prompt interface. Or else any anomalous activities are detected then it will be displayed at the time from where the abnormal event is detected and also shows the video clip of that event. The figure 7.2 shows video clip of abnormal event.



##### C:\Users\Megha\AppData\Local\Packages\Microsoft.Office.Desktop_8wekyb3d8bbwe\AC\INetCache\Content.Word\Screenshot (149).pngFigure 7.1 Command prompt user interface

##### Figure 7.2 Abnormal event video frame.

## 7.3 Summary

The System takes video as an input it performs the testing. If it detects any anomalous activities it will show the exact time from where the abnormal activity is detected and also it shows the video clip of abnormal event. In the presence of no abnormality, user should be displayed with a message that the video is normal. After the whole video is processed the user should be given an option to see the portion of the video where abnormal event was detected. Only the frames to be suspected as abnormal should be played as a video.

## Chapter 8

**Conclusion and Scope for Future Work**

## Conclusion

Technology has played a very important role in our lives. We use it almost everywhere and always. The clear and rapid developments we discover every day prove to us that it makes no sense to give up and struggle with obstacles in our lives. Technology offers many important solutions to our problems and shortcomings. Our job is to use them properly to achieve a level of success that benefits individuals, society, and the country as a whole.

## Scope for Future Work

The proposed K-Means algorithm is relatively time efficient and superior to existing anomaly detection methods. Thanks to an efficient implementation and a new interval suggestion technique that excludes non-essential parts of the data from detailed analysis. Though our model can perform very well with the large dataset it still needs to be improvised based on future requirements and needs to increase the performance by optimizing the algorithm to produce faster results.

**References**

1. Tadashi Ogino, “Detection System for Video data Using Machine Learning”, International Journal of Knowledge Engineering , Volume 2, No. 2, June, 2016.
2. B.V Emmanu Varghese, Jaison Mulerikkal and Smitha Mathew, “Video Anomaly Detection in Confined Areas”, 2017.
3. Amuthan prabakar Muniyandi, R. Rajeswari and R. Rajaram, “Network Anomaly Detection by Cascading K-Means Clustering and C4.5 Decision Tree algorithm”, International Conference on Communication Technology and System Design (ICCTSD), Volume 30, pp. 174-182, May 2012.
4. Sandhya Rani Sahoo, Ratnakar Dash, Ramesh Ku. Mahapatra and Baishnabi Sahu, “Unusual Event Detection in Surveillance Video using Transfer Learning”, International Conference on Information Technology (ICIT), :IEEE, 2019.
5. Jinghua Wang and Guoyan Zhang, “Video Data Mining based on K-means Algorithm for Surveillance Video”, Computer Science and Technology Department

, Hua Zhong Normal University :IEEE, 2011.

1. K. Meena, A. Viji, J. Joshan Athanesious and V. Vaidehi, “Detecting Abnormal Event in Traffic Scenes using Unsupervised Deep Learning Approach”, Department of Electronics, Madras Institute of Technology: IEEE, July 2020.
2. Aiswarya Mohan, Meghavi Choksi and Mukesh A Zaveri, “Anomaly and Activity Recognition Using Machine Learning Approach for Video Based Surveillance”, 10th International Conference, Communication and Networking Technologies (ICCNT): IEEE, 2019.
3. Nguyen Thanh Van and Tran Ngoc Thinh, “An anomaly based network intrusion detection system using deep learning”, International Conference on System Science and Engineering (ICSSE), pp. 210-214: IEEE,2017.
4. Claudio Piciarelli, Christian Micheloni and Gian Luca Foresti, “Trajectory Based Anomalous Event Detection”, IEEE Transaction on Circuits and Systems for Technology, volume 17, No. 11, November 2008.

Adal Nadjara Toosi and Mohsen Kahani, “Support Vector Machine based Anomaly Detection”,2016.