Dissertation on

"INTELLIGENT VIDEO SURVEILLANCE"

Submitted in partial fulfilment of the requirements for the award of degree of

Bachelor of Technology
in
Computer Science & Engineering

UE18CS390A – Capstone Project Phase - 1

Submitted by:

SIDDHANT KUMAR	**PES2201800129**
AYUSHI AGARWAL	**PES2201800053**
SHIVAM SINGH RAWAT	**PES2201800095**
AAKASH BALI	**PES2201800035**

Under the guidance of

Dr. Karthik Chandrashekar
Assistant Professor
PES University

January - May 2021

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
FACULTY OF ENGINEERING
PES UNIVERSITY
(Established under Karnataka Act No. 16 of 2013)
Electronic City, Hosur Road, Bengaluru – 560 100, Karnataka, India

PES UNIVERSITY

(Established under Karnataka Act No. 16 of 2013)
Electronic City, Hosur Road, Bengaluru – 560 100, Karnataka, India

FACULTY OF ENGINEERING

CERTIFICATE

This is to certify that the dissertation entitled

'Intelligent Video Surveillance'

is a bonafide work carried out by

SIDDHANT KUMAR	**PES2201800129**
AYUSHI AGARWAL	**PES2201800053**
SHIVAM SINGH RAWAT	**PES2201800095**
AAKASH BALI	**PES2201800035**

In partial fulfilment for the completion of sixth semester Capstone Project Phase - 1 (UE18CS390A) in the Program of Study -Bachelor of Technology in Computer Science and Engineering under rules and regulations of PES University, Bengaluru during the period Jan. 2021 – May. 2021. It is certified that all corrections / suggestions indicated for internal assessment have been incorporated in the report. The dissertation has been approved as it satisfies the 6th semester academic requirements in respect of project work.

Signature Dr**.** Karthik Chandrashekar Assistant Professor	Signature Dr. Sandesh B J Chairperson	Signature Dr. B K Keshavan Dean of Faculty

External Viva

Name of the Examiners	**Signature with Date**
1. _____	_____
2. _____	_____

DECLARATION

We hereby declare that the Capstone Project Phase - 1 entitled **"Intelligent Video Surveillance"** has been carried out by us under the guidance of Dr. Karthik Chandrashekar, Assistant Professor and submitted in partial fulfilment of the course requirements for the award of degree of **Bachelor of Technology** in **Computer Science and Engineering** of **PES University, Bengaluru** during the academic semester January – May 2021. The matter embodied in this report has not been submitted to any other university or institution for the award of any degree.

PES2201800129	**Siddhant Kumar**
PES2201800053	**Ayushi Agarwal**
PES2201800095	**Shivam Singh Rawat**
PES2201800035	**Aakash Bali**

ACKNOWLEDGEMENT

I would like to express my gratitude to Dr. Karthik Chandrashekar, Department of Computer Science and Engineering, PES University, for her continuous guidance, assistance, and encouragement throughout the development of this UE18CS390A - Capstone Project Phase – 1.

I am grateful to the Capstone Project Coordinator, Dr.Sarasvathi V, Associate Professor, for organizing, managing, and helping with the entire process.

I take this opportunity to thank Dr. Sandesh B J, Chairperson, Professor, Department of Computer Science and Engineering, PES University, for all the knowledge and support I have received from the department. I would like to thank Dr. B.K. Keshavan, Dean of Faculty, PES University for his help.

I am deeply grateful to Dr. M. R. Doreswamy, Chancellor, PES University, Prof. Jawahar Doreswamy, Pro Chancellor – PES University, Dr. Suryaprasad J, Vice-Chancellor, PES University for providing to me various opportunities and enlightenment every step of the way. Finally, this Capstone Project could not have been completed without the continual support and encouragement I have received from my family and friends.

ABSTRACT

This product should be in compliance with the modern world system and will be useful in monitoring the most crowded places. In order to have the best outcome possible leading to maximum accuracy and precision, we can implement this software using deep learning techniques such as Conv. 3D and Spatio-Temporal encoder decoder. This algorithm is best when it is compared to few of the statistical machine learning algorithms due to its outstanding performance. This algorithm is mostly preferred in case of sequence variation and it possesses the ability to be trained end-to-end directly on the source and target, handling variable length of input. Auto encoders have the ability to work with compressed forms of data. We will also be making use of LSTM for successful implementation of the software. The datasets used for the implementation are from UCSD dataset portal.

TABLE OF CONTENT

CHAPTER 1

INTRODUCTION

The application of video surveillance systems has been increasing day by day from the recent past decade. These systems are helpful in densely populated regions such as markets, banks,shopping malls, streets, etc. in order to safe-guard the public by automating the task of detection of various anomaly events related to crime,theft, and other such mishappenings. The history of Video Surveillance started from the era of the 1940s when it was first used to launch the V2 rocket in Germany.In the US, commercial surveillance applications began around 1947. In 1957 a number of companies such as General Precision Labs (GPL division), provided CCTV camera systems for education, medical and industrial applications.Despite the fact that video observation frameworks have been a necessary piece of general society and security areas for quite a long time, there is a critical interest in them outside of those businesses. This premium is generally because of expanded crime percentages and security dangers from one side of the planet to the other, which are driving a nonstop development of the video reconnaissance market.The anomalous activity are very much restricted to a particular domain or area of interest.Now the real question comes, what do we mean by anomaly? A very crisp and concise answer would be any mishappening or event happening that shows significant difference from the rest of the surrounding activities. When it comes to the task of video surveillance, it's quite a tedious job to manually perform the surveillance of the videos.It would require a huge manpower and chances of error(couldn't predict the mishappening). Here comes the most important role of Computer Vision,which is required for this automated task of surveillance.This automated version was to notify whenever there was a deviation from the normal activities(anomaly detected). This anomaly detection is completely different from normal analysis of videos, as the count of these anomaly videos are quite less in number compared to the normal videos and secondly,the anomaly videos may belong to various classes.Various techniques such as reconstruction,predictive models are

used for video anomaly detection.The main aim of these algorithms is to reconstruct the frame with minimum loss of precision.High reconstruction score is noticed in case of abnormal videos.The spatio- temporal pattern is one in which variation in spatial region and temporal is observed.In this case, the most popular model to deal with spatio temporal patterns is Long Short Term Memory(LSTM). LSTM is a deep-learning based recurrent neural network model.Contractionary to the case of Feed Forward Neural Network, LSTM also has feedback connections. It is considered to solve most of the sequence prediction problems.LSTMs have an edge over conventional feed-forward neural networks and RNN in many ways. This is because of their property of selectively remembering patterns for long durations of time. One of the major drawbacks of RNN is it can only remember the short-term dependencies.The major reason for this drawback is Vanishing Gradient. In order to overcome this drawback of RNN, we can make use of a tweaked version of RNN that is LSTM. The information flow in case of LSTM is through cells known as states.

Coming to the Architecture of LSTMs, a typical LSTM compromises of different memory blocks known as **cells.** When information is transferred to the next cell,there are 2 things that are forwarded to the next cell, **cell state** and **hidden state.** Now these memory blocks(cells) are responsible for remembering things and modifying them. Modification of these information by these cells are done by gates.There are a total of 3 types of Gates: **Forget Gate,Input Gate,Output Gate.** A **forged gate** is answerable for eliminating data from the cell state. The data that is not, at this point needed for the LSTM to get things or the data that is of less significance is taken out through augmentation of a channel. This is needed for improving the exhibition of the LSTM organization. A sigmoid function is responsible for generating a vector consisting of 0’s and 1’s corresponding to each input value.If the value generated by the sigmoid function is 0, for a particular input value then the LSTM wants to forget that piece **of data**.Coming, to the **input gate that** is **liable for** the addition of information to the cell state. This addition **of data is essentially** a three-step process.

- Controlling what esteems **ought to** be **another** to the cell state by **together with** a sigmoid capacity. This is essentially **essentially identical because the** neglect door and goes **concerning** as a channel for all **the information** from h_t-1 and x_t.
- Making a vector containing all potential qualities **that may** be **another** (as seen from h_t-1 and x_t) to the cell state. **this is often** finished utilizing the tanh work, **that** yields esteems from **- one** to +1.
- Increasing the worth of **the executive** channel (the sigmoid door) to the **created** vector (the tanh capacity) and **later on** adding this valuable **information** to the cell state

  And,the **preceding** gate, the **output gate** which is **to blame for choosing helpful data** from **the present** cell state **associate**d showing it out as an output **is finished** via the output gate.The functioning of **associate** output gate **will once more** be **lessened to a few** steps:

- Making a vector subsequent to applying tanh capacity to the cell state, in this way scaling the qualities to the reach - 1 to +1.
- Making a channel utilizing the upsides of h_t-1 and x_t, with the end goal that it can control the qualities that should be yielded from the vector made previously. This channel again utilizes a sigmoid capacity.
- Increasing the worth of this administrative channel to the vector made in sync 1, and sending it out as a yield and furthermore to the secret condition of the following cell.

Initially, the images have been preprocessed using OpenCV. OpenCV is used to deal with images and its properties.Using OpenCV we can rearrange,reorient and resize the images.OpenCV contains executions of in excess of 2500 calculations! It is unreservedly accessible for business just as scholarly purposes. Also, the delight doesn't end there! The library has interfaces for various dialects, including Python,Java, and C++.

CHAPTER 2

PROBLEM STATEMENT

Surveillance of security video feeds manually is a laborious process where the reaction to a trigger event must be immediate. In the past few decades, surveillance cameras, also known as Closed-circuit television (CCTV), have had a rapid growth in numbers around the world. Surveillance security is a very tedious and time-consuming job. In this project, we aim to build a system to automate the task of video surveillance. We will analyze the video and identify the abnormal activities such as violence and theft. The main areas of focus are health issues(like cardiac arrest),detect violence and find any suspicious objects.

CHAPTER 3

LITERATURE SURVEY

Paper 1

The main objectives that were inferred from this paper was to find a way to reduce the tiresome and laborious manual work as a result of which a lot of time and effort go wasted. Hence this project uses intelligent surveillance video analysis technique to attain maximum accuracy to reduce the possibility of false positives and negatives that might lead to false alarms to the concerned authorities. The response time should be as minimum as possible so that necessary actions can be taken as soon as possible. Estimation of movements comes handy in emergency situations like stampedes. It also needs to ensure that high volume of data is fed in the form of data frames so that high accuracy is achieved by the model.

Paper 2

Here we talk about the FaceNet technique with Multi-task Cascaded Convolutional Networks(MTCNN) which can throughput much higher accuracy than methods such as DeepFace and DeepID2+ which are but much faster than the earlier. We found that the Deep learning methods are better than the traditional machine learning methods since deep learning with small amount of data do not tend to perform quite well since it overfits the model and produces incorrect results and in this project, accuracy should be high since we are dealing with domains such as curbing the needy people in the time of a cardiac arrest, theft identification and monitoring various other abnormal activities happening in the surrounding.

Also, deep learning methods tend to outperform in ways that it learns after being trained and it becomes quite handy to deal with complex algorithms with the use of Deep learning

methods in such scenarios. The limitation incurred from such a method is that it requires high resolution image sequences to achieve higher accuracy.

Paper 3

It becomes very essential to identify violence happening in the environment and automatic video surveillance system makes it efficient and possible to achieve it. This can be achieved using the background subtraction method proposed by Datta et al. but it becomes even harder to use it in case when there is collection of people present in each region under surveillance. So, we use the method of irregular motion information where we consider a vector of motion coordinates of the pedestrian in given 8 directions.

Paper 4

In this we talk about target tracking in which the detection of moving targets in the video and considering various other useful information such as position and trend of the tracked target by using various tracking algorithms. It mainly includes many other advanced technologies like sensor technology, pattern recognition, image processing and so on.
It is based on the particle filter which is a nonlinear filtering algorithm, which is based on non-parametric Monte Carlo and Bayesian estimation.

Paper 5

This reviewed through many research papers related to surveillance present online, showing research trends over the years and application areas, network architectures and frameworks, hence listing methods widely applied in such models.

This only listed the methods proposed over time but did not provide a detailed review of the techniques used but just summarised its application and was affected by illumination, camera movement and weather changes.

Paper 6

This paper focuses upon the effective way to scale up the intelligent video surveillance and this is achieved through the framework called SIAT. SIAT focuses on the distributed computing technologies to ensure features such as scalability, effectiveness, fault tolerance and availability. But this paper fails to explain the components such as resource utilization, privacy and security that need to be ensured so that individuals can access the system with zero threat.

Paper 7

Spatiotemporal technique is suitable for intelligent video surveillance since it considers the kernels that are both space and time oriented in order to process the videos into the data frame. It is able to process the nonlinear functions compared to PCA (Principal Component Analysis) technique which also performs the dimensional reduction.

The model is built by incorporating the implementation using three layers of the convolutional layers. This consists of the temporal encoder-decoder and Convolutional LSTM gated with tanh and sigmoid activation functions. The spatial decoder further has three layers of deconvolutional layers. There is loss of data while performing the dimensional reduction compared to PCA method. The training is carried out on only with the normal data classes.

Paper 8

This research paper talks about big data solutions to aid the increasing storage redundancy and large space consumption worldwide while developing the system to consider the increasing demand for the surveillance of suspicious activities. Based on the Sensor Web Enablement framework in the backend Dey et al. made use of the big data and cloud-based framework to get better retrieval results. This also ensures to make use of the plus points of a cloud-based system such as efficiency, availability, and fault tolerance.

CHAPTER 4

DATA

When the videos are split up into frames it is stored as an additional directory. Our dataset was downloaded from http://www.svcl.ucsd.edu/ in which a camera was installed at an elevation to monitor human movements.

We plan to use the Avenue dataset (776 MB) and UCSD(707 MB) dataset in our project which contains 16 training and 21 testing video clips. There was a total of 30562 frames collected from the CUHK campus. The model is trained on the normal events while the testing of the model is performed on both normal and abnormal/anomalous events and find the happening of an anomaly activity in the given region.

The UCSD dataset is equipped with a camera for carrying out the surveillance on the campus that monitors the activities of the pedestrians which is further subdivided into two sets i.e., Peds1 and Peds2 dataset. The dataset included the crowd variability ranging from meagre to quite high.

The following is the description of the two sets:

Peds1: This dataset consists of people walking to and fro from the camera. It contains 34 training video samples and 36 testing video samples.

Peds2: This dataset has clips of pedestrians walking parallel to the camera plane. It contains 12 testing video samples and 16 training video samples.

CHAPTER 5

PROJECT REQUIREMENTS SPECIFICATION

Product Features

5.1.1 Health issue detection

Incase of sudden heart attacks or seizures this model can detect a subject collapsing and can immediately raise an alarm so that immediate help can be delivered at the time of the need.

5.1.2 Theft Identification

Intrusion or trespassing can be easily detected and the security department can be informed for quick action..

5.1.4 Lost Commodity alert

There are times that in a rush people tend to forget even important things in public areas ,such things get easily misplaced or stolen in some cases. To prevent that this system can detect lost commodities and lost and found authorities can be alerted regarding the same

5.1.5 Traffic collision detection

Traffic policemen cannot be at duty at all times due to intense pollution on the road these systems can help detect if there has been a disruption or collision in the traffic so authorities can be informed instead of manual surveillance at all times.

User Classes and Characteristics

Various user classes for this product can be :-

5.2.1 Security Organizations

Specific organizations providing security solutions can use this as a product to enhance their security suite for better customer satisfaction and trust

5.2.2 Customers

Normal day to day consumers can directly access the web UI and use the footage from their surveillance fed to detect the various anomalies.

5.2.3 Developers

Consumers with knowledge of machine learning algorithms and coding can also undertake this project and adapt it to their use along with any useful extra features that they could implement.

Operating Environment

This project should be able to run in all environments irrespective of the operating system or machine architecture as this uses google collab which is a jupyter notebook environment that runs entirely on the cloud.

General Constraints, Assumptions and Dependencies

The following might limit the choices for developers :-

- **Regulatory policies**

  Any surveillance system must follow the country/region's privacy policy for storing and monitoring video records. Every organisation or workplace has different security policies listed in their documents and that should be analysed before implementing this.

- **Hardware limitations. E.g. - signal timing requirements**

  The area to be monitored should have decent quality cctv cameras to avoid resolution issues during the analysing phase. Should be adequately lit and if implemented in real time should have minimum or no delay if possible.

- **Interfaces to other applications**

  This project has a web frontend and can be integrated into a suite of security applications.

- **Parallel operations**

  It has the capability to detect multiple types of anomalies ranging from burglary in apartment complexes or accidents on road and fights .

- **Criticality of application**

  This application has high priority as the security and wellbeing of the subjects are dependent on its correct and useful functioning. Any error or downtime is neither tolerable nor affordable as it can cost not only money but also lives.

- **Safety and security consideration**

  The system has been designed to detect all possible incidents and report them appropriately but it will always be better to have a person overlooking the process since even though the chances of the model failing is very low it will never be zero.

Risks

There is always a possibility of failure in any innovation related to any field hence there will always be risks either related to business or in general.

Some risks associated with this project are listed below :-

1. **Risk of life** - If the system is not manually monitored and some anomaly slips through the system there will be a threat to either someone's well being or risk of not being able to provide help in a timely manner.
2. **Risk of finance** - If this system is integrated in an environment where failures are not tolerable , any malfunction can easily translate to loss of property or money.
3. **Risk of compromise** - Network facilitated devices such as these could easily be exploited by attackers if proper security measures are not in place or there is fault in the firmware or design which can lead to the data being stolen
4. **Risk of losing privacy** - Footage of sensitive areas like people's houses or safes cannot be stored or leaked to the public since this can easily lead to a planned attempt for a crime.

CHAPTER 6

SYSTEM REQUIREMENTS SPECIFICATION

Functional Requirements

The input is video which is broken down into frames, following this data is read from each individual frame but first all the frames are scaled to the same size followed by getting gray scale values. We follow a process which is a spatiotemporal approach which first breaks down the input with respect to space and then looks for change in value with respect to time and if it finds an irregular change then it triggers an alarm depending on the kind of change.

The output of the software is generated only in case of an anomaly being detected .

The error that is possible is false alarms being raised which need to be taken care of by analyzing the trend in the false alarms and further training the software.

Another possible error is when the system fails to detect an anomaly.

External Interface Requirements

User Interfaces

The frontend of this application will be a website hosted where an alert will be generated if anomalous activity is detected , basic documentation pertaining to usage of application along with complete procedure will also be available to the consumer.

Hardware Requirements

This project will require footage of the area under observation which can be captured through a CCTV or any recording device and can be processed on any machine with

internet connectivity as the model is hosted on google collab which provides collaboration on Cloud Level.

If the user wants to run it on their own systems following are the minimum hardware requirements :-

1. Intel Core i5 or AMD Ryzen 3 Processor
2. 8GB RAM

Software Requirements

Following are the software requirements necessary for building the model :-

- Python 3.6+

  Python needs to be installed in order to run this project on local host, if the model code is being compiled from scratch
- Microsoft Windows 10 ,Linux or Mac

  An operating system is mandatory for this project to be compiled and executed , best efforts have been put to make the code portable.
- Web Browser

  If the user is not interested in undertaking the hassle of compiling the code themselves they can access the finish product interface directly from the browser by accessing the website
- Libraries

  Numpy 1.9.1 , keras ,Scipy 0.14
- Tools

  Terminal , command prompt or powershell.

Communication Interfaces

These communication interfaces are important for this project :-

1. If the project is being used for real time or close to real time purpose then the database server should be connected to the CCTV via high speed wireless communication channels or optic fibre cables.
2. If the project is being used for existing recordings where fast computation is not required then adequate and fast storage space is necessary to store the vast number of frames generated after processing of video recordings.
3. If the project is being run on online cloud platforms like google collab(which makes the monitoring more dynamic) then fast internet services or fibre is recommended but not compulsory.

Non-Functional Requirements

Performance Requirement

The project is reliable in the domain it is meant to operate unless the staging environment changes drastically this will function normally and can recover from any incidents if they occur.

The project is robust to a great extent when it comes to handling failures as the hardware and software specifications as stated are easy to go with on any system. Also care is taken that the accuracy in detection of the applied areas are maximum with less number of False Positives.

This project can be implemented using the principle of Convex polygon which states that the area can be maximized with the minimum number of CCTV required.

Safety Requirements

There are no safety requirements as such but when it comes to the safety of the public being monitored by the intelligent video surveillance system is concerned, this system is meant to report crimes and health abnormalities it has been trained to detect and not

substitute an actual police system where the presence of armed policemen is enough to stop a crime or reduce the intensity of a crime taking place in the vicinity of a law enforcement officer.

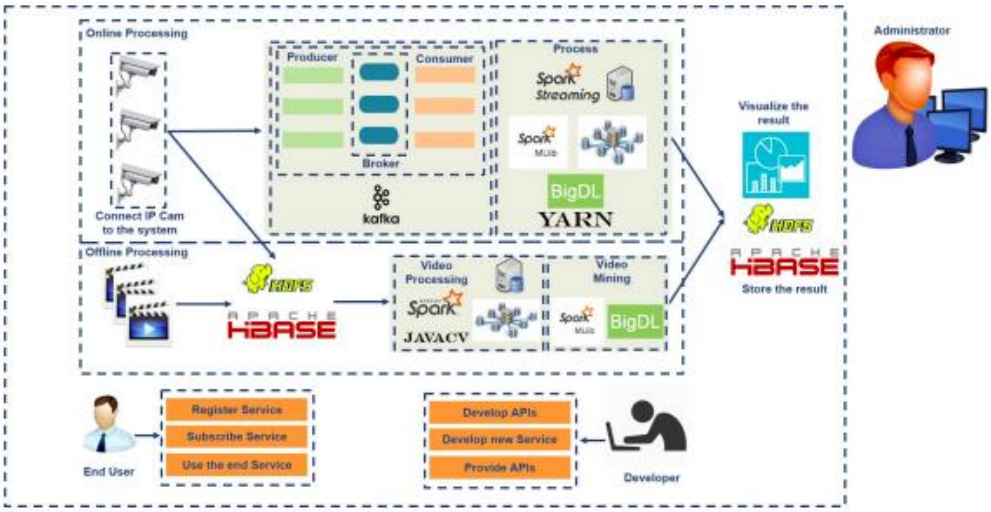
Security Requirements

A security requirement that is a must is that there must not be fake cases being enacted or false alarms being triggered from the software which would make it a tedious task for the agencies and make the software a burden rather than a help.

Other Requirements

Scalability

Scalability from the demo presented to a real life deployment would be a really complex time because a real deployment would involve really intensive computing, data storage, data flow and many other factors which we haven't encountered in a simulation like the one we're presenting. A real deployment would have a set-up similar to the set-up shown below:



Portability

The software is portable but needs to be trained with no abnormalities for a given feed followed by a set up as shown in the diagram above.

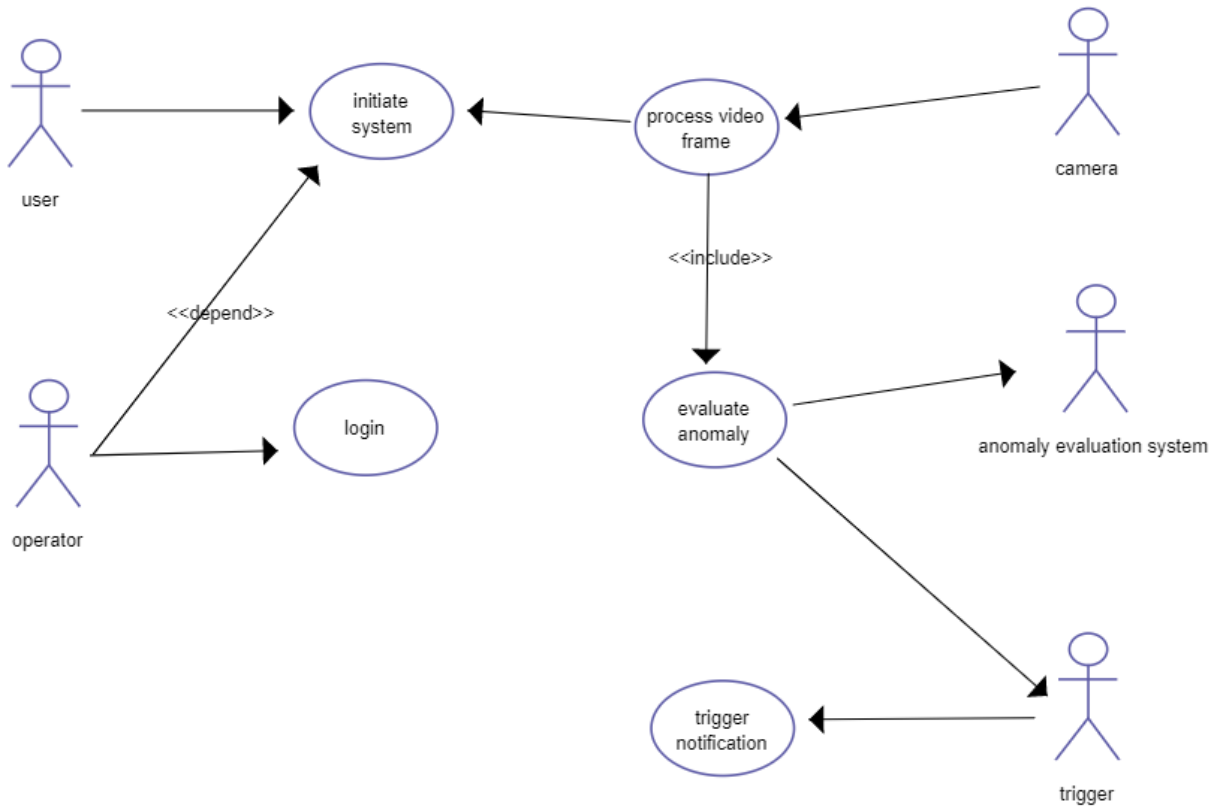
Maintainability

The software would be scheduled for regular software updates and would be updated with the latest algorithms that we come across and find fit for the situation where it is deployed.

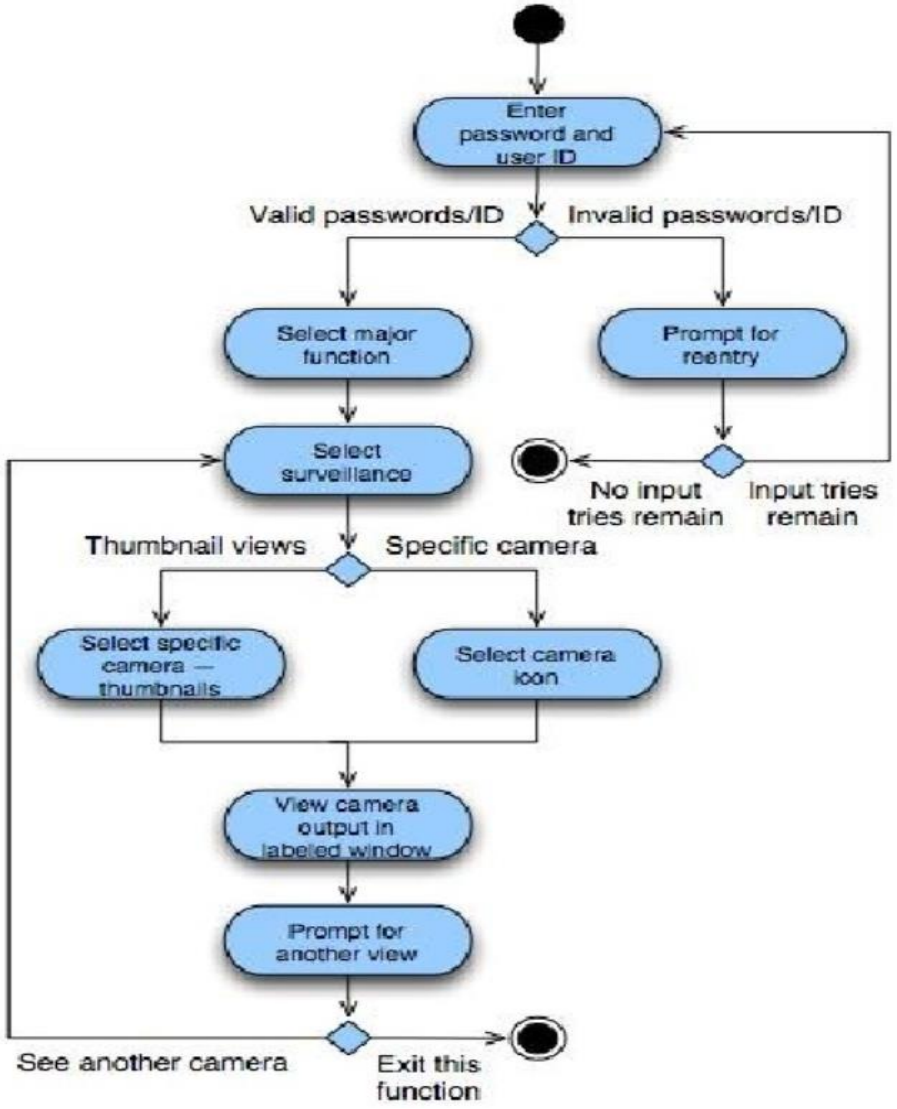
CHAPTER 7

SYSTEM DESIGN

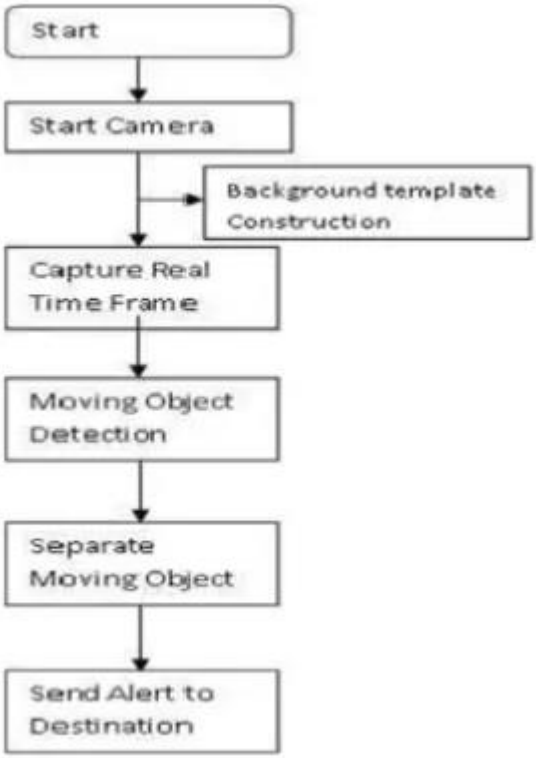
1. Conceptual or logical diagram



2. Activity Diagram .



3. Modules



The front-end part has a login feature that takes a login id and a password and has all necessary functions to meet the user as well as the project requirements. It gives the look and feel to the user and is quite user friendly. The main code runs that runs at the backend process the video fed into the system in data frames of a given pixel and is further trained and tested. It ensures that the system has a good accuracy in terms of its capability to find the anomalies in the video when there was actually one by reducing the number of false positives/negatives based on the regularity and the anomaly score calculated by the model.

CHAPTER 8

IMPLEMENTATION AND PSEUDOCODE

Front-End Implementation

The project consists of a web based frontend which is built on technologies like HTML, CSS and Javascript further this is wrapped up by Flask a python micro framework, which is a third-party Python library used for developing web applications. The goal is to build an API that will serve the machine learning model. Another alternative to Flask was Django but Flask is faster due to its less complicated and minimal design, also Flask is a good choice for a lightweight codebase.

Pseudocode for the front-end :-

1.) Display login page and ask for credentials, either login or register
2.) Verify credentials from credential database
3.) Provide options to the end-user for multiple features of the model through the welcome page.
4.) Accept the options selected by the user and deliver the appropriate feature web page.
5.) Prompt user to input link to the video data that needs to be processed
6.) Fetch the video data from the input link and send it to the backend(model).
7.) Receive frame from the backend when the anomaly occurs.
8.) Generate an alert to the respective authority with relevant details.

Back-End Implementation

The project tries to extend deep neural networks to 3 dimensions for learning spatio-temporal features of the video. A spatiotemporal autoencoder is used which utilises the concept of 3 dimensional convolutional neural network. The encoder part extracts the spatial and temporal information, and then the decoder reconstructs the frames. The abnormal events are identified by computing the reconstruction loss using Euclidean distance between original and reconstructed batch.

Pseudocode for the back-end:-

1.) Video is received from the front-end to the back-end that is the model.
2.) Video is divided into frames by the model.
3.) Frames are further compressed using INTERPOLATION_AREA of OpenCV and stored.
4.) Model analyzes the frames for anomaly using euclidean distance
5.) Once the threshold value for abnormal events,and regularity score is crossed , the abnormal event is detected
6.) Extracts the relevant frame and generates an alert to the front-end with the relevant frame.

CHAPTER 9

CONCLUSION OF CAPSTONE PROJECT PHASE-1

As we reach the conclusion of the phase 1 of the capstone project we realise that this phase encompassed deciding the problem statement which we would work upon and improve our understanding in the field. It also involved thorough review of multiple literatures that had been released previously regarding this topic which gave us an insight into the process of intelligent video surveillance , its various algorithms and approaches undertaken by researchers of the same field by reading such research papers published by accomplished and capable individuals. This also led us to realise the feasibility of the project and its various implementations or features. Progressing in the project we developed the front-end using suitable technologies and decided upon the required dataset , pre-processing it and making it usable for our consumption. At each stage of the project we enhanced our knowledge by the vast amount of resources present on the internet to ensure optimum output and compatibility with real life implementations.

CHAPTER 10

PLAN OF WORK FOR CAPSTONE PROJECT PHASE-2

In the next phase of the project we will build the actual model which will work as a standalone project and integrate it with the front-end using flask as stated above. Overall the software will evolve from a computer code to a working product that will see its usefulness on the field. The research work done by the team will reflect on the upcoming product through the modifications in design choices and functionalities. The next phase will involve implementation of the deep learning algorithms and produce fruitful results from the raw data collected in this phase. This will be succeeded by intensive testing , bug fixing and optimizations according to needs.

REFERENCE / BIBLIOGRAPHY

1. Kardas K, Cicekli NK. SVAS: surveillance video analysis system. Expert Syst Appl. 2017;89:343–61.

2. Wang Y, Shuai Y, Zhu Y, Zhang J. An P Jointly learning perceptually heterogeneous features for blind 3D video quality assessment. Neurocomputing. 2019;332:298–304 (ISSN 0925-2312).

3. Tzelepis C, Galanopoulos D, Mezaris V, Patras I. Learning to detect video events from zero or very few video examples. Image Vis Comput. 2016;53:35–44 (ISSN 0262-8856).

4. Fakhar B, Kanan HR, Behrad A. Learning an event-oriented and discriminative dictionary based on an adaptive label-consistent K-SVD method for event detection in soccer videos. J Vis Commun Image Represent. 2018;55:489–503 (ISSN 1047-3203).

5. Luo X, Li H, Cao D, Yu Y, Yang X, Huang T. Towards efficient and objective work sampling: recognizing workers' activities in site surveillance videos with two-stream convolutional networks. Autom Constr. 2018;94:360–70 (ISSN 0926-5805).

6. Paper 1: Intelligent video surveillance: a review through deep learning techniques for crowd analysis by G. Sreenu

7. Paper 2: A deep learning approach to building an intelligent video surveillance system by Jie Xu

8. Paper 3: Violence Detection for Video Surveillance System Using Irregular Motion Information by Jinsol Ha, Jinho Park, Heegwang Kim, Hasil Park, and Joonki Paik

9. Paper 4: An improved target tracking algorithm and its application in intelligent video surveillance system by Nana Zhang & Chunxue Wu & Yan Wu & Neal N. Xiong

10. Paper 5: A Systematic Review of Intelligence Video Surveillance: Trends, Techniques, Frameworks, and Datasets by Guruh Fajar Shidik

11. Paper 6: SIAT: A Distributed Video Analytics Framework for Intelligent Video Surveillance by Md Azher Uddin

12. Paper 7: Video Anomaly Detection using Convolutional Spatiotemporal Autoencoder by Umesh Chandra Pati and Santos Kumar Das

13. Paper 8: Smart Monitoring Cameras Driven Intelligent Processing to Big Surveillance Video Data by Zhenfeng Shao, Jiajun Cai, Zhongyuan Wang

APPENDIX A DEFINITIONS, ACRONYMS AND ABBREVIATIONS

1. 3D Convolution – A type of convolution where kernel slides in 3 dimensions and a 3 dimensional filter is applied to the dataset which moves in 3 directions (x,y,z) to calculate the low level feature representations.

2. Spatio-temporal – Involves data collected from both space and time and is used to describe a phenomenon according to certain location and time.

3. Encoder – Converts an input into a feature map/vector/tensor which hold the information, the features, that represents the input.

4. Decoder - A network (usually the same network structure as encoder but in opposite orientation) that takes input from the encoder, and gives the best closest match to the actual input or intended output.

5. LSTM - Long Short-Term Memory networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems.

6. Feed Forward Neural Network - Artificial neural network with connections between the nodes that do not form a cycle.

7. Recurrent Neural Network(RNN) - Type of neural network where the output from previous step are fed as input to the current step

8. Vanishing gradient - When a deep feed-forward network or RNN is unable to propagate gradient information from out to layers near the input.

9. FaceNet - A face recognition system developed in 2015 by researchers at Google that achieved then state-of-the-art results on a range of face recognition benchmark datasets.

10. Multi-task cascaded convolutional neural network (MTCNN) - a human face detection architecture which uses a cascaded structure with three stages (P-Net, R-Net and O-Net).

11. Representation learning - learning representations of input data by transforming it or extracting features from it(by some means), that makes it easier to perform a task like classification or prediction.

12. InterpolationArea or INTER_AREA – Python method which uses pixel area relation to increase or decrease the size of an image.

APPENDIX B USER MANUAL (OPTIONAL)

We will be providing the User Manual for the end users to properly understand the working and configuration of the software.

1. Description of the product

1.1. General Information

This product has been designed to automate the task of surveillance that otherwise may cost huge amounts and the negligence in which might cost even more. This will be done by monitoring targets in the scene along with analysis of behaviour and prediction of trajectory which correspond to unwanted behaviour like accidents or violence.

1.2. Topology of the system

The system has been described by the diagrams and models above. In normal operation, the user, web interface and the machine learning model communicate with one another through the product. These relations are setup by default in the programming of the product which can be accessed through an internet connection. The web-based front end also provides a uniform access across platforms.

1.3. Operating Environment

This model is setup to operate in any environment according to situation it may be operated where security is quintessential like home apartment complex, traffic lights, and areas which are prone to crowd and accidents. According to hardware and software environments the backend can run independently on the cloud using Jupyter Notebook environment like google collab notebooks which let the user leverage backend hardware through abstraction thus hiding unnecessary complicated setups. If the model is to be run as a whole with the backend and frontend integrated together it can be done locally in any operating system by downloading the respective dependencies or the user could

utilise the web-based front-end that has been provided with the project which will be hosted online(tentative).

1.4. Operation Flow

Overall the project when the front-end and back-end is integrated together will follow the flow given in the report above which can be restated in a concise way that the front-end is supposed to receive the video input that needs to be processed through a link . The video is then extracted from the link and sent to the machine learning model in the backend which then processes it into frames , analysing , detecting and alerting the front-end about various anomalies that might have occurred. The back-end also sends the relevant frame which is proceeded by an alert from the front-end along with the relevant details and frame . All this will be secured by a login interface which will require credentials.

1.4.1. Setup

The project if being setup locally can be done by installing the various dependencies required that consist mainly of the python modules that have been used. A live server plugin can be installed to host the model locally. Overall this project requires minimal setup because of the web interface which takes care of the various dependencies that otherwise might have been necessary.

1.4.2. Use Instructions

End-users can easily make use of the functionality provided by the project via its easy-to-use web interface , the various steps involved in it would be selecting the option that best suits the kind of anomaly to be detected and providing the link to the video feed once prompted with the relevant page. Rest will be handled by the project itself that will include detection and alarm.

2. Introduction to the software

The various micro-features of the projects that might not be actively stated in the above document are :-

1. Preview of the video

Whatever link that user provides ,a preview of that will be displayed on the relevant web page

2. Video Options

Various options related to the video will be provided to the user depending on the link of the video, these options are generally linked with the video service provider that the link uses and might contain volume toggle, resolution settings, and full screen settings.

3. Electronic Preview Zooming

The user can zoom in or out of the video as they feel comfortable using the web interface.

4. Video Playback

The video that the user enters the link to will be displayed on the screen in form of a square and it will display the video on the screen.

5. Search for a recording

The relevant authority will be informed when an anomaly will be detected and same could be found in the logs whatever form they may be along with the relevant frame .

6. Play a recording

The video that the user enters the link to will be displayed on the screen in form of a square and it will display the video on the screen and the user can play/pause according to need.

7. Clipping of Recording Files

The relevant frame will be extracted from the video hence the video recording will be clipped on the point the anomaly occurs and the frame will be stored.

8. Downloading of Recording Files

The relevant frame can be downloaded from the alert generated.

9. Zoom on the video

The page and video both can be zoomed in and out of according to user needs and comfort.

10. Video Preview

A frame from the video will be displayed initially which will act as a preview of the video.

11. Recording Playback

The incident will be recording in the form of frames which could be viewed at a later point of time.

12. Operation Log

There will be a log of anomalies generated which can be viewed by the customer.

13. Alarm

An alarm will be triggered in the form of an alert when an anomalous activity occurs in the video input.

14. System Configuration

The system should have a source of recording which will be fed into the model so a recording of the incident is necessary which would require the need of a CCTV camera but is not explicitly needed.

15. User Management

The project requires the user to login or register , once the credentials of the user is verified they are granted resources on the basis of conditional access.

16. New User

New user can register on the service using the login page that is prompted when the webpage first loads, this will later stored in a database and verified accordingly.

17. User Right

Each user has different resource access on the website and that will be determined through the login credentials. The user can either be admin or a normal user.

18. Modify User

The admin will have the right to modify user data if needed.

19. Remove User

The admin will also have a right to remove a user from the database , this also may or may not be provided to the users for themselves.

20. Change password of registered user

This right will be provided to the user themselves and may/may not be provided to the admin taking into account privacy laws and concerns of the users.

21. Settings for alarms

The user can decide on the method of alarms by tweaking a code little bit , but this is mainly for developer's interests.