Intelligent Social Surveillance System

A Project-II Report

Submitted in partial fulfillment of requirement of the

Degree of

BACHELOR OF TECHNOLOGY in COMPUTER SCIENCE & ENGINEERING

BY

DEV RAMCHANDANI – EN18CS301078
DHRUV BOTHRA – EN18CS301079
DHRUVANSH MOYAL – EN18CS301080

Under the Guidance of

DR. RATNESH LITORIYA

Department of Computer Science & Engineering
Faculty of Engineering
MEDI-CAPS UNIVERSITY, INDORE- 453331

JAN-MAY 2022

Report Approval

The project work **"Intelligent Social Surveillance System"** is hereby approved as a creditable study of an engineering subject carried out and presented in a manner satisfactory to warrant its acceptance as prerequisite for the Degree for which it has been submitted.

It is to be understood that by this approval the undersigned do not endorse or approved any statement made, opinion expressed, or conclusion drawn there in; but approve the "Project Report" only for the purpose for which it has been submitted.

Internal Examiner

Name:

Designation: Assistant Professor

Affiliation: Medi-Caps University, Indore

External Examiner

Name:

Designation:

Affiliation:

Declaration

I hereby declare that the project entitled **"Intelligent Social Surveillance System"** submitted in partial fulfillment for the award of the degree of Bachelor of Technology in 'Computer Science & Engineering' completed under the supervision of **Dr. Ratnesh Litoriya, Associate Professor Computer Science & Engineering,** Faculty of Engineering, Medi-Caps University Indore is an authentic work.

Further, I declare that the content of this Project work, in full or in parts, have neither been taken from any other source nor have been submitted to any other Institute or University for the award of any degree or diploma.

Dhruvansh Moyal

25th April, 2022

Certificate

I, **Ratnesh Litoriya** certify that the project entitled **“Intelligent Social Surveillance System”** submitted in partial fulfillment for the award of the degree of Bachelor of Technology by **Dhruvansh Moyal** is the record carried out by him under my guidance and that the work has not formed the basis of award of any other degree elsewhere.

______________________ ______________________

Dr. Ratnesh Litoriya

Computer Science & Engineering

Medi-Caps University, Indore

Dr. Pramod S. Nair

Head of the Department

Computer Science & Engineering

Medi-Caps University, Indore

Acknowledgements

I would like to express my deepest gratitude to Honorable Chancellor, **Shri R C Mittal,** who has provided me with every facility to successfully carry out this project, and my profound indebtedness to **Prof. (Dr.) Dilip. K. Patnaik,** Vice Chancellor, Medi-Caps University, whose unfailing support and enthusiasm has always boosted up my morale. I also thank **Prof. (Dr.) D K Panda,** Pro Vice Chancellor**, Dr. Suresh Jain,** Dean, Faculty of Engineering, Medi-Caps University, for giving me a chance to work on this project. I would also like to thank my Head of the Department **Dr. Pramod S. Nair** for his continuous encouragement for betterment of the project.

I express my heartfelt gratitude to my Guide, **Dr. Ratnesh Litoriya,** Associate Professor, Department of Computer Science & Engineering, MU, without whose continuous help and support, this project would ever have reached to the completion.

I would also like to thank to my team Dev Ramchandani and Dhruvansh Moyal who extended their kind support and help towards the completion of this project.

It is their help and support, due to which we are able to complete the design and technical report.

Without their support this report would not have been possible.

Dhruvansh Moyal
B.Tech. IV Year
Department of Computer Science & Engineering
Faculty of Engineering
Medi-Caps University, Indore

Executive Summary

In industry and research, big data applications are gaining a lot of traction and space. Surveillance videos contribute significantly to big unlabelled data. The aim of visual surveillance is to understand and determine object behaviour. It includes static and moving object detection, as well as video tracking to comprehend scene events. Object detection algorithms may be used to identify items in any video scene. Any video surveillance system faces a significant challenge in detecting moving objects and differentiating between objects with same shapes or features.

The primary goal of this work is to provide a quick overview of video analysis utilising deep learning algorithms to detect suspicious activity. In greater applications, the detection method is utilised to determine the region where items are available and the form of objects in each frame. This video analysis also aids in the attainment of security. Security may be characterised in a variety of ways, such as identifying theft or violation of covid protocols. The method of detecting anomalous (abnormal) human actions is known as suspicious human activity detection. We'll need to convert the video to MP4 format for this.

Table of Contents

<u>List of Figures</u>

List of Tables

Abbreviations

SRS	Software Requirement Specification
GUI	Graphical User Interface
BERT	Bidirectional Encoder Representations from Transformers
GENSIM	Generate Similar
CWE	Common Written Examination
IBPS	Institute of Banking Personnel Selection
AI	Artificial Intelligence
API	Application Programming Interface
CSV	Comma Separated Value

Chapter-1

Introduction

1.1 Introduction

In today's day and age, when people have begun to rely more on technologies that are called smart, meaning that they can operate and learn on their own at the command of humans and do the needed duties, we have achieved this new step with the assistance of computer vision and deep learning. In planning, operation, and sustainability of contemporary industrial and urban areas, video surveillance is a major factor. The efficiency, safety, security, and optimality of the region, infrastructures, persons, operations, and activities are all aided by video surveillance [1]. Autonomous equipment, cyber-physical systems, and energy-efficient architectures are becoming more common in industrial settings. With the rising use of multi-level structures, greater traffic, pedestrian, and crowd movements, urban landscapes are becoming more densely inhabited. In both industrial and urban contexts, this vertical and horizontal growth of asset and area utilisation has led to considerable growth in the implementation of closed-circuit television (CCTV) camera systems to ensure the safety of humans or assets and surveillance of activities. We have begun to perform analysis on photos, which was previously only confined to textual data. However, analysing static images was no longer sufficient; we now needed to examine videos that used a similar approach to still images. Giving a live video feed a real-time output is a new difficulty in the industry. During our investigation.

With the use of CCTV, one can watch the area 24/7, or the video may be retrieved when needed provided it is kept in a secure location. Although it could be used to prevent crime and assist law enforcement in identifying and solving crimes. With the help of YOLO, a new approach to perform detection with an extremely quick architecture, real-time image computation, Open-source computer vision (OpenCV) software, and a powerful library of image processing tools.

In our system, we have taken up that challenge and attempted to apply it as much as possible to the field of surveillance in order to tackle the challenges of the modern world. We've been looking for abnormalities in real-time footage. In our research, we selected to identify three real-world anomalies from today's world: violations of social distance norms, face mask usage monitoring, and theft detection to determine if a valuable object is being taken in real time and from where in the frame.

One of the main worries of today's citizens is security. In our system, we've combined three functionalities that have proven to be important in today's world. In order to try and address some problems of the new common world, we built a working prototype of a multi-purpose smart surveillance system that can be used to detect multiple anomalies in real-time and can be used anywhere from a crowded public space to a quiet museum with a reasonable number of people in the room.

1.2 Literature Review

In the field of video analysis and object detection, many researchers have done a substantial amount of work and shared their findings with the world. A few pieces of research of which are discussed below.

Redmon et al. [2] proposed a system using the YOLO algorithm. Classifiers developed for object detection are repurposed for per-form detection. Their design was lightning quick At 45 frames per second, their YOLO model analyses pictures in real-time. You just need to glance at an image once to guess what things are present and where they are with YOLO. YOLO is a simple and quick algorithm that uses complete photos to train the detection process. Because it doesn't employ a complicated pipeline. It uses a neural network to anticipate detections on a fresh picture at test time, allowing us to handle live streams.

The Model was simple to build and can be trained on complete photos. YOLO is the fastest object classification detector in the literature, and it pushes the boundaries of real-time object detection. YOLO is also adaptable to new domains, making it excellent for applications that need quick and reliable object recognition.

Suwarna Gothane [3] says in his research they can quickly detect and identify the items of our attention when looking at photographs or movies. Object detection is the process of passing this knowledge to computers. The YOLO model is quite precise and can recognise the things in the picture. YOLO takes an entirely different approach. Instead of focusing on individual regions, it employs a neural network to predict anchor boxes and their probability throughout the whole image [4]. YOLO uses a single deep neural network that divides the input image into grid size. Unlike image classification or face detection, each grid size in the YOLO algorithm has a related matrix in the output that informs us if a result is found in that grid size and class of that object.

Gupta and Devi [5] made research on the YOLOv2 method that is proposed for the identification of objects in pictures with geolocation and video recordings. The primary goal of this study is to detect things in real-time, i.e., live identification, utilising a camera and video recordings. COCO, a dataset with 80 classes, was employed in this work [6]. Using the YOLOv2 model, it is simple to recognise items with grids and boundary prediction, and it also aids in predicting exceptionally small things or objects that are far away in the image. Darknet makes it simpler to recognise moving objects in video recordings, and it generates.avi files containing detections.

Jayashri et al. [7] Suggest a Real-time system for monitoring social distance and avoiding congestion by employing the suggested critical social density. The team is committed to providing innovative, strategic advancements that protect persons and networks. Under the present circumstances of the COVID-19 pandemic, this work has practical value. The pipeline is capable of detecting persons with, without, and incorrectly wearing coverings with reasonable accuracy

Gupta et al. [8] proposes a simpler way to accomplish this goal by utilizing some fundamental deep learning tools such as TensorFlow, Keras, and OpenCV. The suggested technology successfully recognizes the face in the image/video stream and determines whether or not it is wearing a mask. It can recognize a face and a mask in motion as a surveillance task performance. optimal parameter settings for the CNN model in order to identify the existence of masks accurately.

Kakadiya et al. [9] suggested a system using deep learning to establish a Smart Cam that observes bank activities, can identify any suspicious conduct, and criminals can be followed based on mobility and weapon presence. The SmartCam immediately transmits a message to the safety committee if any suspicious weapons or actions are observed. The message specifies the sort of warning that has been issued, as well as the type of weapon and number of weapons identified, as well as a web link to a live picture that may be viewed by security personnel.

Patil et al. [10] proposed that theft is among the most widespread criminal activities, and it is on the ascent. It became one of the world's never-ending issues. Using a sophisticated algorithm like CNN which provides you an advantage over more standard algorithms such as RNN, SVM, and others. This program correctly recognises emotional expressions with a higher percentage of accuracy rate. The Keras toolbox is used for this. They arrived at the above-mentioned makeshift model after experimenting with various layer combinations and iterations.

1.3 Objective

One of the CCTV surveillance features, perceived at the majority of levels, was creating a safe environment. At the operational level (Level 1), the perception is that CCTV surveillance provided a sense of security for patrons and staff undertaking their work. The first-line management had a broader view and this linked to the goal of delivering good customer service and providing a safe atmosphere. This view was also supported at the senior level, where the perception is that the CCTV does provide a safe environment rather than creating the perception. The mid-level management was the only group that did not have the perception that CCTV created a safe environment. The majority view of creating a safe environment was summarised as reducing the fear of danger and crime.

Protection of Assets:

CCTV surveillance was perceived by all levels as a method to protect assets, with all levels in general agreement. The operational level considered CCTV as a method of “keeping a close eye on the important stuff”. This view can further be explained as monitoring money, people and procedures within the complex, in line with the mid-level view. At the first line manager and senior manager levels, it was perceived that the protection of assets covered a broader context than just material assets with money or possessions, extending to the protection of business reputation.

Legal Compliance:

The final overarching CCTV surveillance aim, perceived by the first to senior management levels (level 2 to 4), was the legal aspect. At first line management, it was claimed that the "objectives for an organisation like this 31 would be built around the requirements from the government". Such a requirement is defined by the Australian governmental in respect to the gambling and the sale of liquor. This view was supported at mid-level management, who stated that the organisational procedures are written in respect to these legislated requirements and therefore, "compliance is very heavy subject here". Finally, senior management stated that complying with these legal requirements is a significant purpose in supporting business objectives. Due to the type of business, namely gaming, the Australian government has certain expectations on evidence being provided by the organisation that only legal activity is occurring

1.4 Scope

This system has a wide range of uses in various fields, such as banking, forensic department, etc. The reason this system is quite useful is due to the fact that it is highly compact and it provides face detection and an instant notification about the same through email. In addition to this face recognition can also be tried in future. Recognition is the main part of any security system. Usually for a best recognition system, we require a well-trained database, which can provide the base for our recognition. So, to obtain the database, first collect the images of the subject individual for the recognition. Once we obtain and train our system, we can provide face recognition. We use the local binary pattern histogram (LBPH) for providing face recognition. This method helps us to provide a recognition model. The image is converted into a grey scale image. Then, the image pixels are compared with the neighbouring pixels in a clock-wise or anti-clock-wise manner. Histogram is performed and normalization is done and a feature vector is generated for every image. These feature vectors can now be processed with some algorithms to classify images which is used to identify the texture. Once the face is recognized, it is checked to see if the detected face is familiar or not. Thus, we integrate the face detection and recognition to provide a smart surveillance system for the domestic purposes in our everyday life.

1.5 Chapter Schema

Chapter 2: System Requirement Specification

This chapter include the information about system requirements like functional and non-functional requirement and specification of the software like software and hardware specifications.

Chapter 3: System Analysis & Design

This chapter include all the diagrams related to the system for better understanding of the system.

Chapter 4: Implementation

This chapter include the implementation phase of the system include proposed methods for the system and screenshots of interfaces and database.

Chapter 5: Testing

This chapter include the testing phase of the system. Several test cases are included in this chapter.

Chapter 6: Result and Discussion

This chapter include results drawn by the system.

Chapter 7: Summary & conclusion

This chapter includes summary about the project.

Chapter 8: Future Scope

This chapter include the future scope possible in the system

Chapter-2

System Requirement Specification

2.1 Proposed System

The suggested research is built using Python 3, OpenCV, and the flask framework of python. Using OpenCV's machine learning methods, we can educate the machine to discern between diverse user use cases and unauthorised offenders' unique undesired conduct, allowing us to take appropriate action based on the context. We are able to effectively use logic to execute the artificial intelligence idea at hand to recognise and classify the events that occur using image processing strategies and mathematical deductions. Additionally, the system is capable of taking action in response to the current occurrence [11].

Modules:

A. Theft Detection

After the motion is detected, we save the frame just before the motion and the next static frame after the motion is also taken. Then we calculate the image similarity score using the structural similarity function of open cv. If the similarity score is more than the desired threshold then we classify that nothing is stolen.

B. Face Mask Detection

If the input is a video stream, the picture or a frame of the video is initially delivered to the default face detection module for detection of human faces.

C. Social Distancing Detection

Apply people detection to detect all people in a video stream and draw a rectangular bounding box around each person. Find the centroid of the person detected on the frame.

2.2 Procedures Adopted

The primary goal of this system is to analyse collected video footage for human detection and then further analyse it for any anomalies. Anomaly can be anything the user sets the software to detect. A breach of social distancing laws, a person not wearing a mask, or the detection of theft on video footage could all be considered anomalies for this system.

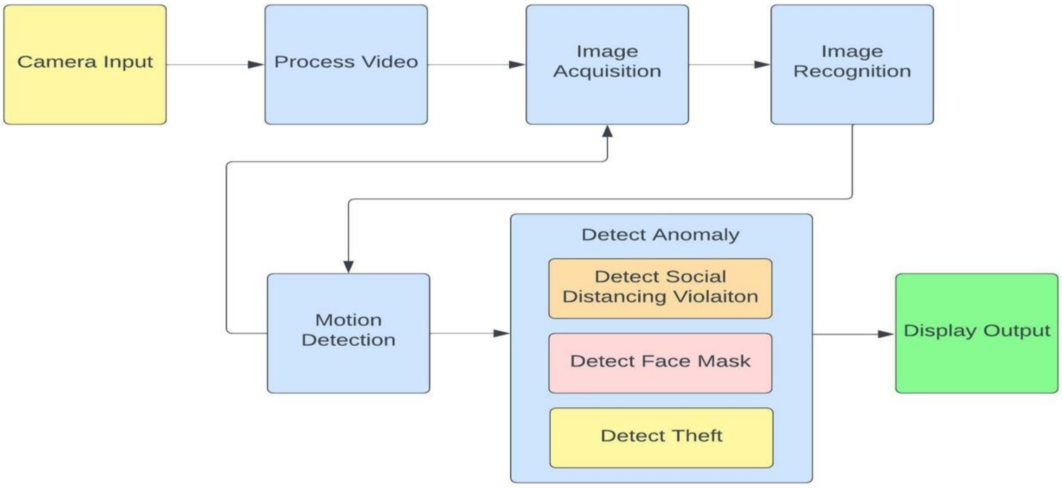

Figure 1: Block Diagram

As a result, the procedure begins by scanning each frame of a video stream one by one. This is depicted in Figure 1, which is also depicted in the block diagram with the entire sequence of actions.

2.3 System Feasibility

The feasibility of the project is analysed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential. Four key considerations involved in the feasibility analysis are Economic Feasibility, Technical Feasibility and Social Feasibility.

- ### Economic Parameter

  Economic feasibility analysis is the most frequently used method for evaluating the effectiveness of a new or proposed system. More commonly known as cost/benefit analysis, the procedure is to determine the benefits and savings that are expected from a candidate system and compare them with the costs needed to be spending for implementation and operation. In case, the benefits outweigh costs, the decision is made to design and implement the system. The determination of economic feasibility requires an identification of the potential costs associated with the implementation of the proposed system. When multiple layers are added to the existing devices, the cost may be increased by 50 percent to 100 percent, which is still a feasible price for a high security authentication device. This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus, the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

Our device is a combination of both hardware and software. The hardware components are easily available in the market at a very cheap

- ## Technical Parameter

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. The proposed model with proven results through the prototype shows that the system is technically feasible to be implemented as an independent authentication system. The camera system is compact and can be implemented with low cost. The implemented face detection algorithm (Haar like cascade classifier) is very effective, with an accuracy of 88.9 percent which can be increased further by effectively improving the illumination of the area. However, this system is connected with the help of an Ethernet cable to the laptop to communicate with the raspberry pi. This can be overcome by making the system wireless.

- ## Social Feasibility

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity.

2.4 Hardware Specification

- Processor Intel Core i5 -3230M 2.6GHz
- Memory 8GB DDR3 RAM
- Graphics NVIDIA GeForce GT 720M 2GB Graphics or above
- Display 14" HD Display (16:9) Wide Screen
- Camera Integrated Web Camera
- Communications Wi-Fi 802.11b/g/n, Ethernet Port
- Software Genuine Windows 10
- Webcam 1MP Fixed Focus CMOS camera on the laptop

2.5 Software Specification

- 4 gigabyte (GB) RAM (32-bit) or 2 GB RAM (64-bit)
- 16 GB available hard disk space (32-bit)
- Open cv 4.5.3
- NumPy 1.19
- Python 3.8 (64-bit) or above
- VS Code

2.6 Functional Requirement

A Functional Requirement (FR) is a statement that describes the service that the programmer must provide. It refers to a software system or a component of one. A function is nothing more than the inputs, behaviors, and outputs of a software system. A computation, data manipulation, business procedure, user interaction, or any other unique feature that determines what function a system is likely to execute can all be considered. Functional Requirements are also known as Functional Specification in Software Engineering.

2.6.1 Object Detection:

This programmer will use a human detection method to filter out humans from the video recording device's frames. By using human detection, the algorithm may ignore the remainder of the non-essential characteristics, lowering the amount of calculation required to calculate distance. OpenCV and the YOLOv3 algorithm will be used to create the person detecting unit.

2.6.2 Distance Calculation:

The programmer should be able to determine the centroid of each individual after knowing their position, as well as the pairwise distance between all of their centroids.

2.7 Non – functional requirements

These are basically the quality constraints that the system must satisfy according to the project contract. The priority or extent to which these factors are implemented varies from one project to other. They are also called non-behavioural requirements.

- **Efficiency in Computation:**
  When HGR is executing, the software shall utilize less system’s CPU resource and very little of system memory.
- **Extensibility:**
  The software shall be extensible to support future developments and add-ons to the HGR software. The gesture control module of HGR shall be extensible to allow new gesture recognition features to be added to the system.
- **Portability:**
  The HGR software shall be 100% portable to all operating platforms. Therefore, this software should not depend on the different operating systems.
- **Performance:**
  This software shall minimize the number of calculations needed to perform image processing and hand gesture detection.
- **Reliability:**
  The HGR software shall be operable in all lighting conditions. Regardless of the brightness level in user’s operating environment, the program shall always detect user’s hands.

- **Usability:**

  This software shall be easy to use for all users with minimal instructions. The graphical user interface (GUI) shall be intuitive and understandable by non-technical users too.

Chapter-3

System Analysis & Design

3.1 Use-Case Diagram

A use case diagram is used to represent the dynamic behavior of a system. It encapsulates the system's functionality by incorporating use cases, actors, and their relationships. It models the tasks, services, and functions required by a system/subsystem of an application. It depicts the high-level functionality of a system and also tells how the user handles a system.

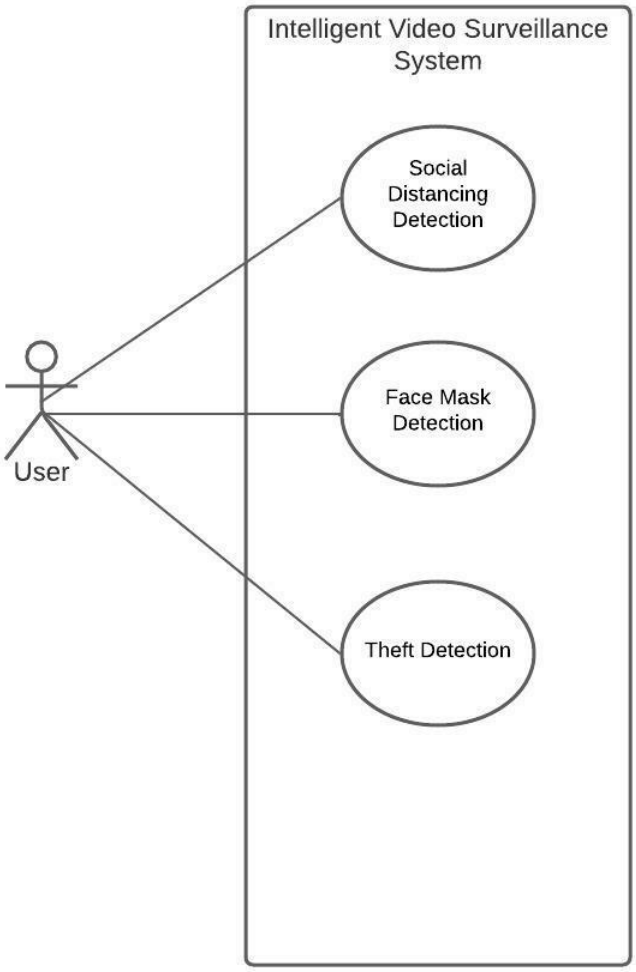

Figure 2 : Use Case Diagram

3.1 Activity Diagram

An activity diagram is a behavioral diagram i.e., it depicts the behavior of a system. An activity diagram portrays the control flow from a start point to a finish point showing the various decision paths that exist while the activity is being executed. We can depict both sequential processing and concurrent processing of activities using an activity diagram.

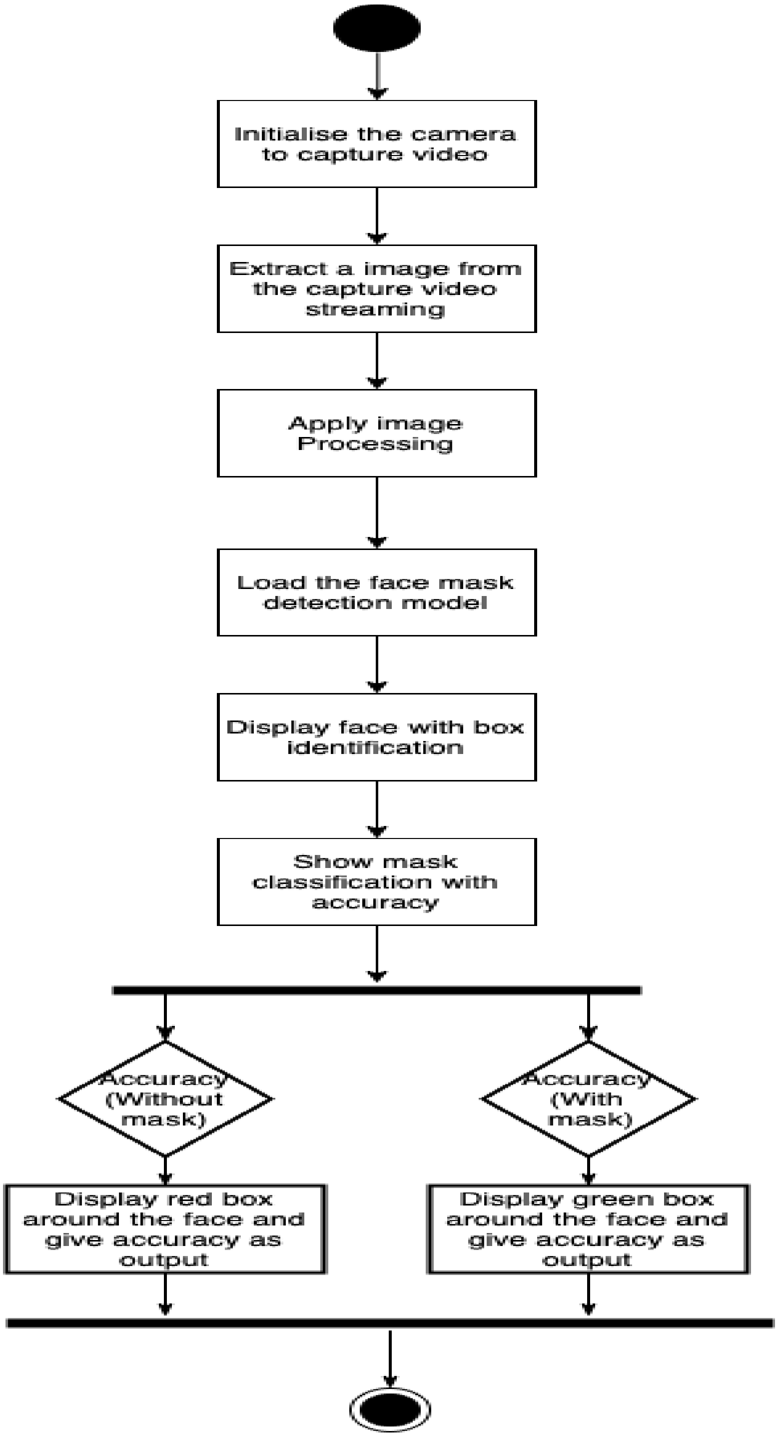

Figure 3: Activity Diagram for face mask detection

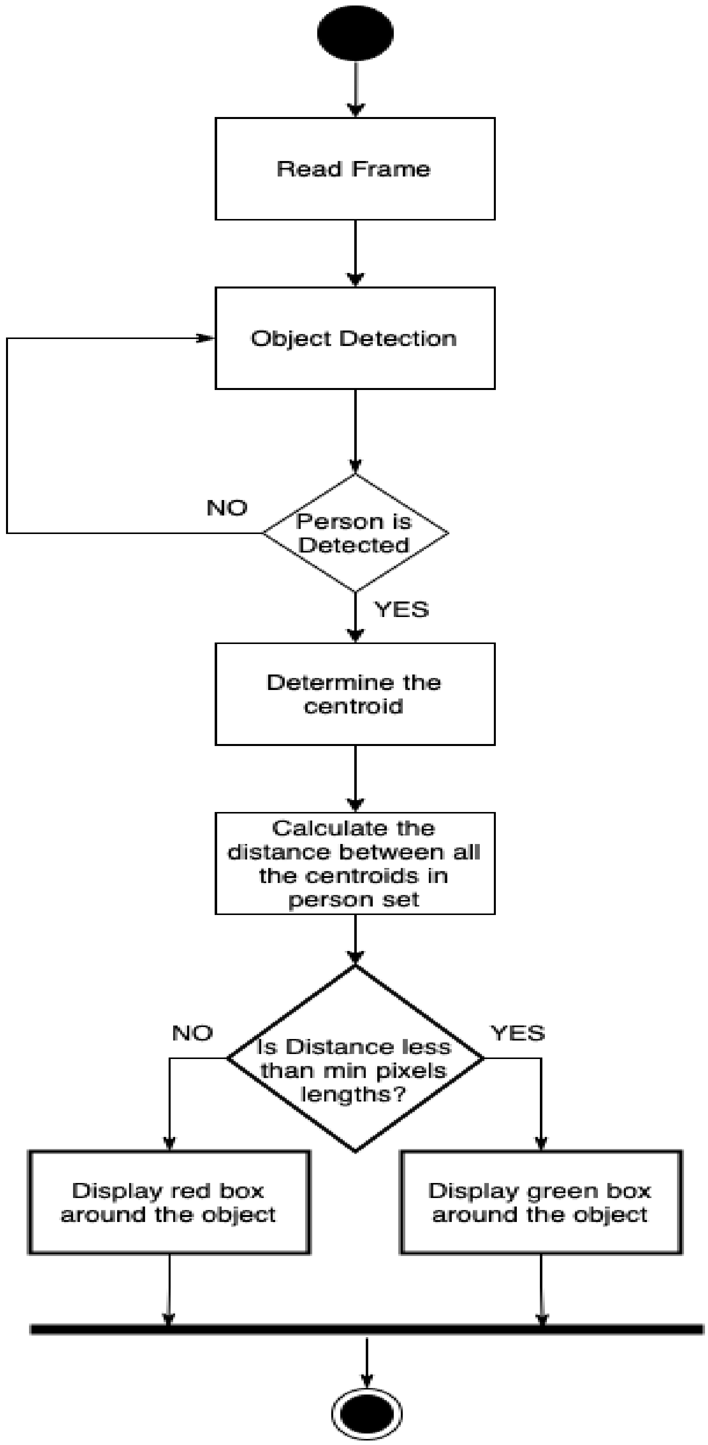

Figure 4: Activity Diagram for social distancing

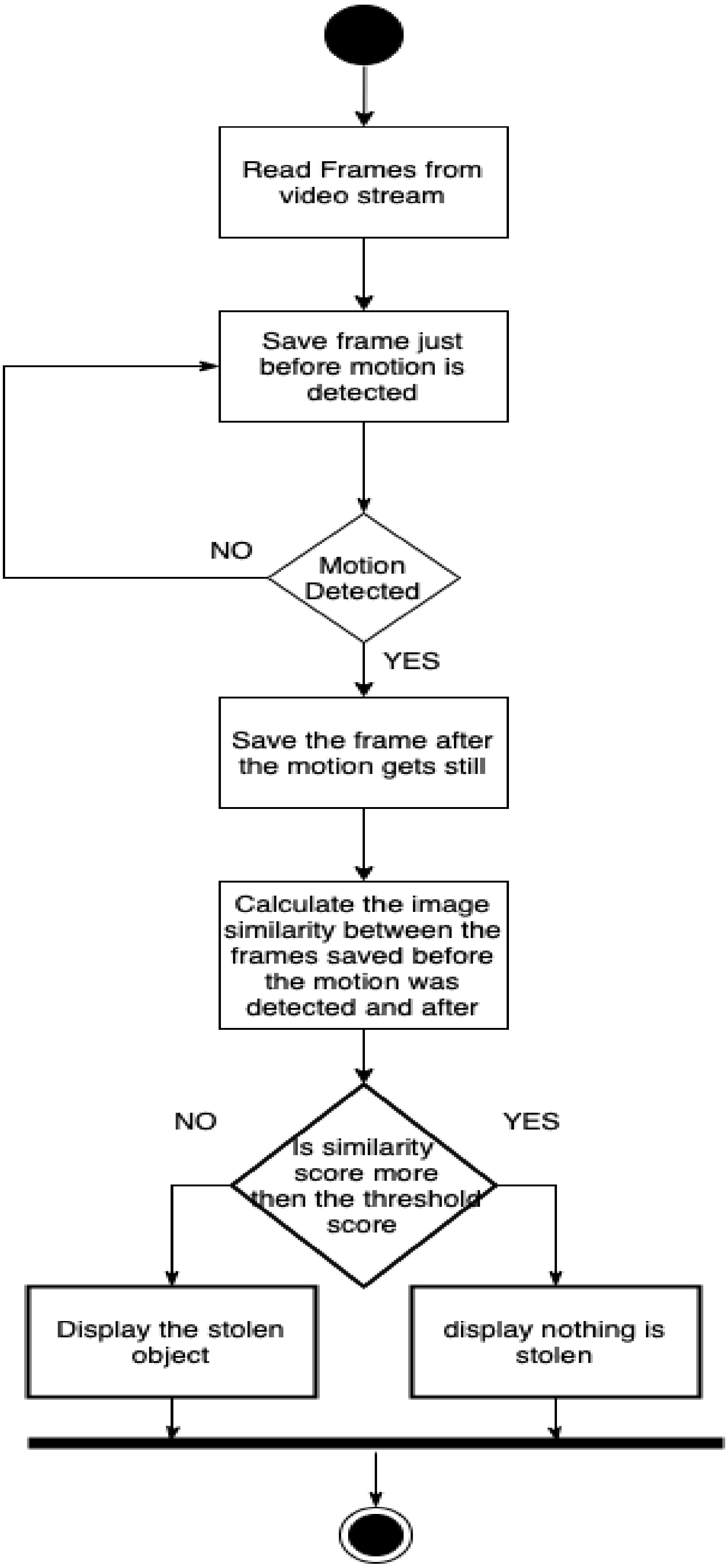

Figure 5: Activity diagram for theft detection

3.1 Sequence Diagram

A sequence diagram is a Unified Modelling Language (UML) diagram that illustrates the sequence of messages between objects in an interaction. A sequence diagram consists of a group of objects that are represented by lifelines, and the messages that they exchange over time during the interaction.

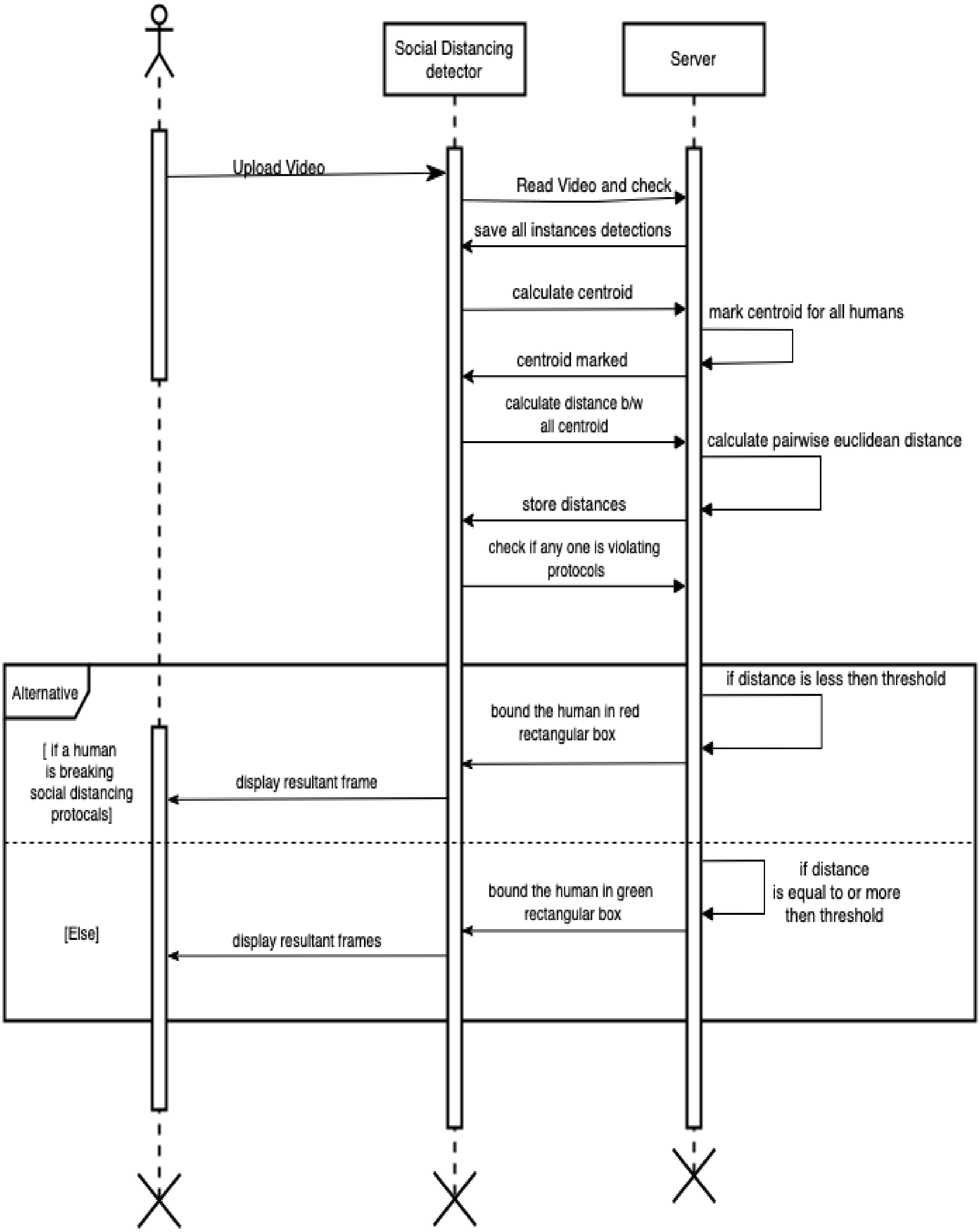

Figure 6: sequence diagram social distancing

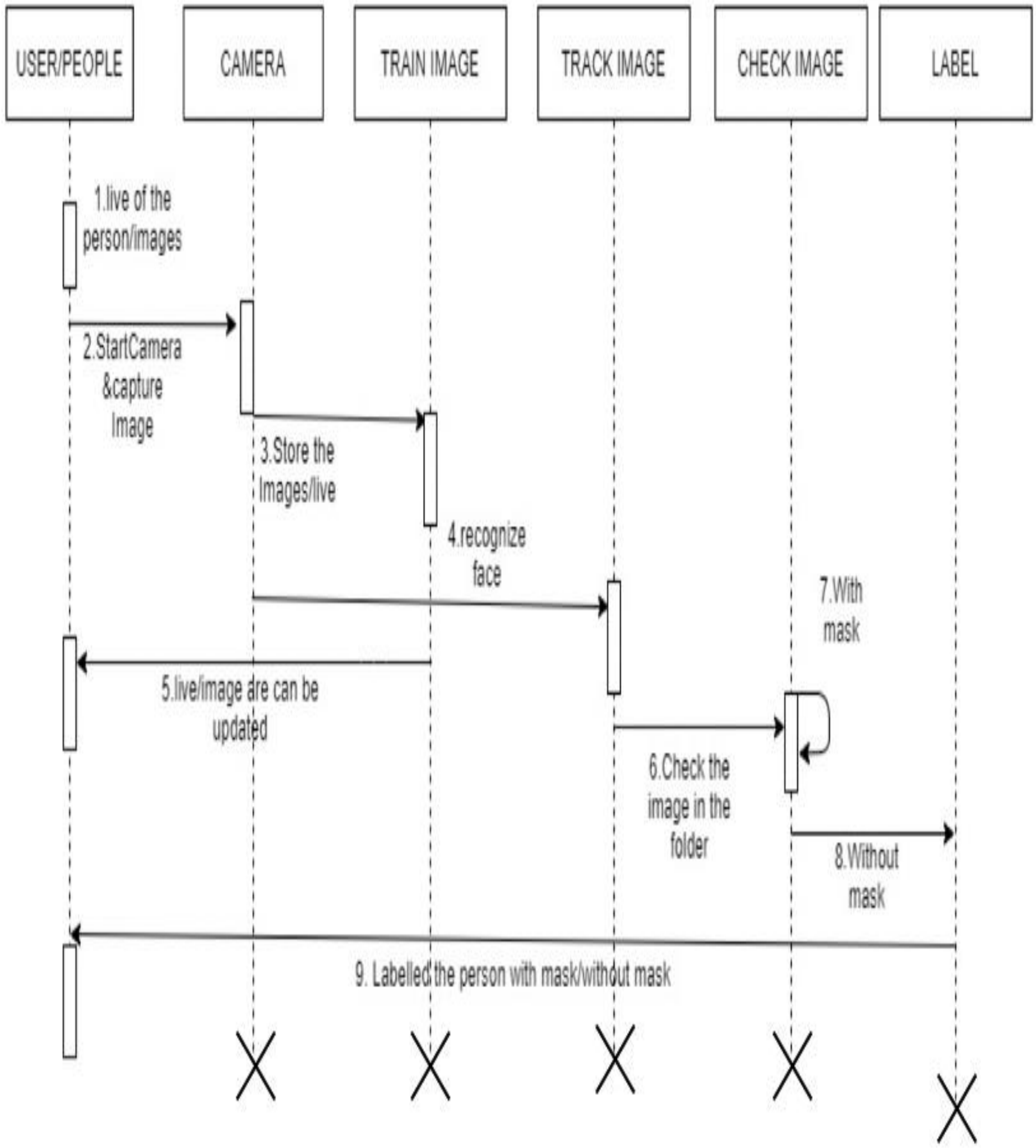

Figure 7: sequence diagram for face mask detection

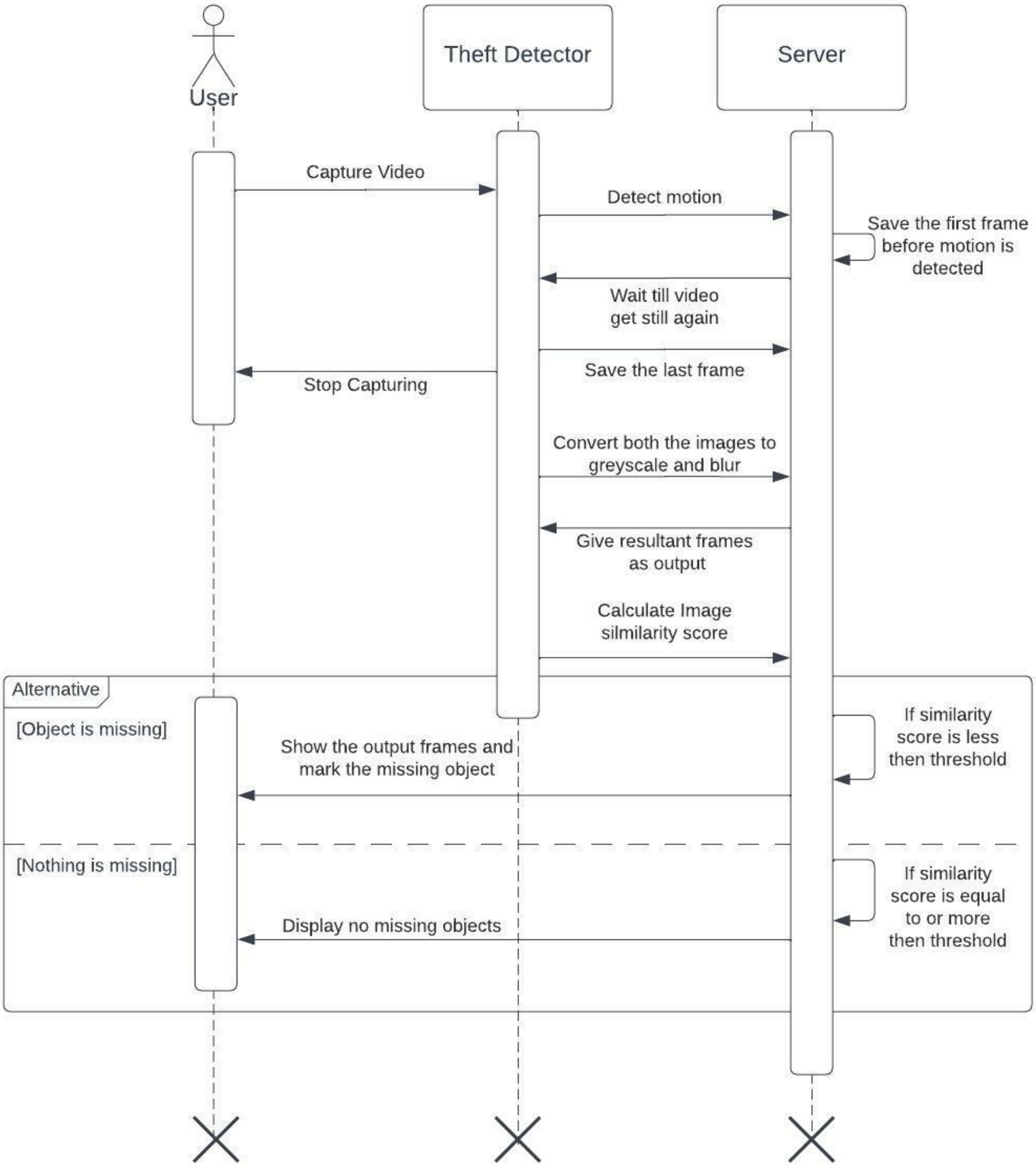

Figure 8: sequence diagram for theft detection

Chapter-4

Implementation

4.1 Procedural Description

The suggested system is built using Python 3, OpenCV, and the flask framework of python. Using OpenCV's machine learning methods, we can educate the machine to discern between diverse user use cases and unauthorised offenders' unique undesired conduct, allowing us to take appropriate action based on the context. We are able to effectively use logic to execute the artificial intelligence idea at hand to recognise and classify the events that occur using image processing strategies and mathematical deductions. Additionally, the system is capable of taking action in response to the current occurrence [11].

The primary goal of this system is to analyse collected video footage for human detection and then further analyse it for any anomalies. Anomaly can be anything the user sets the software to detect. A breach of social distancing laws, a person not wearing a mask, or the detection of theft on video footage could all be considered anomalies for this system.

As a result, the procedure begins by scanning each frame of a video stream one by one. This is depicted in Figure 1, which is also depicted in the block diagram with the entire sequence of actions.

The object detection framework is the most essential aspect of this research. This is due to the study's component that focuses on establishing a person's position from the input frame. As a result, selecting the most appropriate object detection model is critical in order to prevent any issues with recognising people [12].

Monitoring: As soon as the system is turned on, its initial inclination is to scan its surroundings for any movement that could occur in the situation under consideration. The primary goal of motion analysis is to minimise large duplicate activity storage. The recording of the movement begins as soon as the camera detects any unknown creature approaching the target.

Masking Frame: Masking is an image processing technique in which a tiny picture fragment is defined and used to affect a bigger image. Setting part of the pixel values of an image to zero and another backdrop value is known as masking [13]. The picture will be isolated. For example, a video is a collection of images that are played in a specific order over a period of time. To build ROI for each frame of the input frame, the OpenCV masking approach will be employed in this study.

Motion Detection: The surveillance system stays silently watching until it detects an unknown creature coming, at which point the camera begins recording the scene in question, which includes a clear view of the item.

4.2 Module Description

A. Theft Detection

After the motion is detected, we save the frame just before the motion and the next static frame after the motion is also taken. Now both the frames are converted to greyscale and blurred to make a comparison. Then we calculate the image similarity score using the structural similarity function of open cv. If the similarity score is more than the desired threshold then we classify that nothing is stolen. On the other hand, if the similarity score is less than the desired threshold then it suggests that the two frames have structured dissimilarity and something is stolen or missing. Furthermore, in the second case, we again use the grey scaled images, apply thresholding to them and construct a rectangle box where the dissimilarity exists.

B. Face Mask Detection

If the input is a video stream, the picture or a frame of the video is initially delivered to the default face detection module for detection of human faces. This is accomplished by first enlarging the picture or video frame, then identifying the blob inside it [14]. The face detector model receives this identified blob and outputs just the cropped human face without the backdrop. This face is used as model input that we previously trained. This determines whether or not a mask is present. [15].

C. Social Distancing Detection

Apply people detection to detect all people in a video stream and draw a rectangular bounding box around each person. Find the centroid of the person detected on the frame using the below formula.

Centroid of rectangle:($(x_1 + x_2) / 2$, $(y_1 + y_2) / 2$)

Where,

x-Centre = $(x_1+x_2) / 2$

y-Centre = $(y_1 + y_2) / 2$

Moving on after finding the centroid, compute the pairwise Euclidian distances between all detected people.

Euclidian distance: $d = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]}$

where,

- (x_1, y_1) are the coordinates of one point.
- (x_2, y_2) are the coordinates of the other point.
- d is the distance between (x_1, y_1) and (x_2, y_2).

Based on these measurements, determine whether any two persons are fewer than N pixels away, which is the accepted threshold pixel distance. The minimum pixel distance varies depending on the camera's height and angle. The social distance protocols are broken if the distance between two centroids is smaller than the threshold distance, and vice versa.

Chapter-5

Testing

5.1 Unit Testing

Unit testing is the first level of testing and is often performed by the developers themselves. It is the process of ensuring individual components of a piece of software at the code level are functional and work as they were designed to. Developers in a test-driven environment will typically write and run the tests prior to the software or feature being passed over to the test team. Unit testing can be conducted manually, but automating the process will speed up delivery cycles and expand test coverage. Unit testing will also make debugging easier because finding issues earlier means they take less time to fix than if they were discovered later in the testing process.

As previously stated, unit testing is only done for individual components, and because our project is limited in scope owing to its primary focus on anomaly detection, we only had a few components to test. Our functions performed admirably and passed the test. Individual components such as routines loading web pages, doing simple calculations, and obtaining relevant files were tested and found to be successful during the unit testing phase.

5.2 Integration Testing

After each unit is thoroughly tested, it is integrated with other units to create modules or components that are designed to perform specific tasks or activities. These are then tested as group through integration testing to ensure whole segments of an application behave as expected (i.e, the interactions between units are seamless). These tests are often framed by user scenarios, such as logging into an application or opening files. Integrated tests can be conducted by either developers or independent testers and are usually comprised of a combination of automated functional and manual tests.

After we completed unit testing in our product, it was time to move on to integration testing. Various functions in our small-scale programme work together and call each other to produce the required outcome. We can declare our system passed this testing approach since we obtained our intended output at the end and all of the functions called each other at different times and functioned without any disruption or resource imbalance.

5.3 Validation Testing

The process of determining if software meets stated business requirements throughout the development process or at the conclusion of the development process. Validation Testing guarantees that the product satisfies the demands of the customer. Inter-rater reliability, intra-rater reliability, repeatability (test-retest reliability), and other features can be used to test/validate test validity, which is normally done by running the test numerous times and comparing the results. Validation is crucial for two reasons: quality assurance and cost minimization. Validation ensures that the product is fit for its intended purpose.

One of the main problems in our predecessor software's or as we can say one of the main challenges in video analysis was of processing a video in different environment. Many a times if a shadow was introduced in a frame if was harder to classify it for many software's and it finally end up classifying it wrongly sometimes as humans sometimes as different object but in our software, we were able to classify it rightly. As you can see in the below figure that while running and processing a video, our software was more consistent in classifying shadow as itself and not letting it hinder our working, proof of which being that our software was designed to detect if humans were following social distancing rule or we can say it was only interested in detection of humans and ignore another object. If the shadow was wrongly classified as humans, our program would have drawn a rectangular box around it but as we can see our software overcame this problem well and only at a minute instance made a false classification which still was acceptable.

Fig 9 Shadow Check Validation

5.4 White Box Testing

White box testing is also known as glass box testing, structural testing, clear box testing, open box testing, and transparent box testing. It examines a software's core code and architecture with an emphasis on comparing present inputs to expected and intended results. It centers around internal structure testing and is centered on the inner workings of a program. The design of test cases in this form of testing necessitates programming expertise. The main purpose of white box testing is to concentrate on the flow of inputs and outputs via the software while also ensuring its security.

White box testing, like the rest of our software, was critical. It gave us confidence in how our program would respond if it were given unknown inputs. However, our software operated admirably. As intended, it took mp4, mov and mkv files and produced an avi file as output.

5.5 Performance Testing

The technique of examining how a system performs in terms of responsiveness and stability under a specific workload is known as performance testing. Performance tests are commonly used to evaluate application performance, robustness, dependability and size.

Our software was designed only for smaller prerecorded files with sizes less than equal to 64 mbs. Any file larger than that will result in our software rejecting it and requesting a file with smaller size. Depending on the hardware this threshold of size can we changed.

Chapter-6

Result & Discussion

6.1 Result and Discussion

6.1.1 Theft Detection:

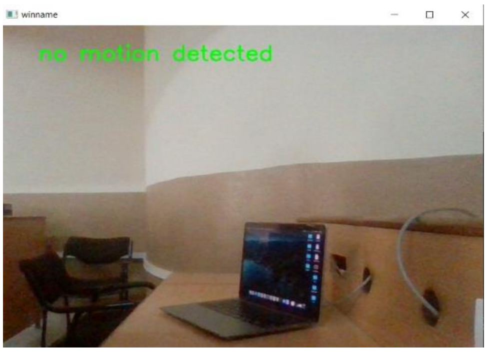

Figure 10: Frame with object

Figure 11: Frame without object

This method is used to find whether an object is moved from its place or not. If the object is moved or is in motion the next static frame after motion will be taken to check whether the object is moved or not. If the object goes missing in the static frame after motion then a box will be made on the initial frame taken for comparison. The full procedure is depicted in Figure 10-12. Figure 10 shows an object resting on the table with no motion detected at this time. However, when the object is moved, as shown in Figure 11, motion is detected, and our system provides us with information about the object being moved as well as from where the object is moved, as shown in Figure 12.

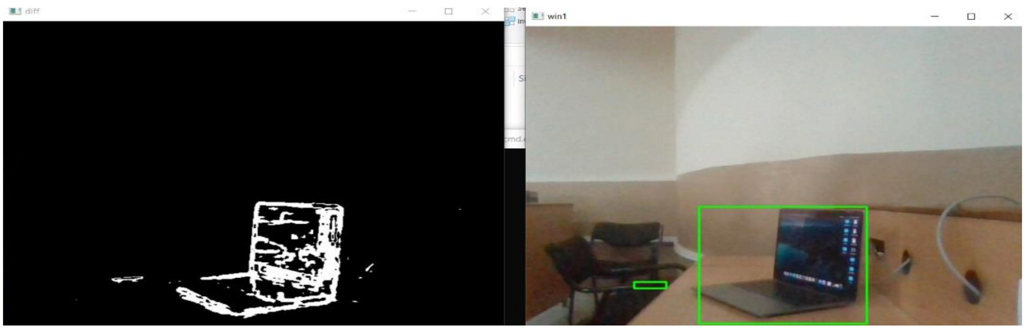

Figure 12: Missing object highlighted in green boundary and grey scale image of missing object

In our theft detection function, motion was identified 100% of the time in decent lighting, but only 91 percent of the time in low or dim illumination. Each time an object was taken or moved in the frame, it was detected with 100% accuracy and the precise location of the missing object was noted. Although the software's recognition of the object's shape was not flawless, it was able to accurately indicate the shape of the missing object more frequently than not.

6.1.2 Face Mask Detection:

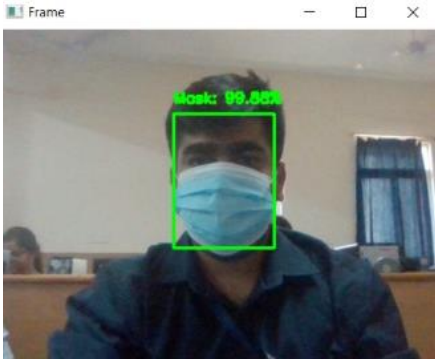

Figure 13: Person with mask on

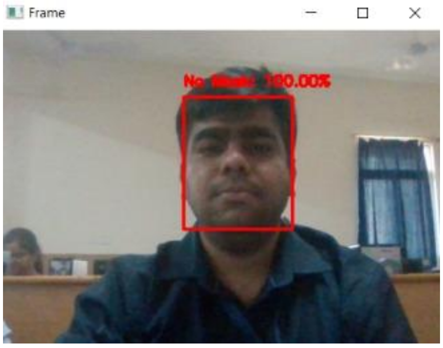

Figure 14: Person without mask

In the above Figure 13-14, if a bounding box is drawn around the ROI as and check whether the person is wearing a mask or not. The green colour bounding box depicts the person is wearing a mask and the red bounding box depicts a person is without mask. Figure 14 shows that when a person successfully wears a mask, our system correctly predicts that the person is wearing the mask with an accuracy of 99.88 percent, but when the person is not wearing the mask, it predicts that accurately with 100% accuracy as shown in figure 14.

Figure 15: Two people with and without mask

This method works on a group of people in the frame as well, as we can see in figure 15-16. Accordingly ROI is created around each face and bounding boxes are assigned green box will only be assigned if a person is wearing the mask correctly in all other cases red bounding box is assigned. Figure 15 shows that when a person is not wearing a mask properly, it is accurately recognised as such with a high degree of accuracy, whereas Figure 16 shows that our system can recognise faces from a fair distance and correctly classify several persons wearing masks or not.

Figure 16: Group of people with and without mask

Table 1: Outcomes of Face Mask Detection Method

Sr.no.	People in Frame	Lighting Conditions	Face Detected	Correct prediction
1	1	Standard lighting	1	1
2	1	Dim lighting	1	1
3	2	Standard lighting	2	2
4	2	Dim lighting	2	2
5	3	Standard lighting	3	3
6	3	Dim lighting	3	3
7	5	Standard lighting	5	5
8	5	Dim lighting	5	5
9	7	Standard lighting	7	7
10	7	Dim lighting	6	6
11	8	Standard lighting	8	8
12	8	Dim lighting	8	8
13	10	Standard lighting	10	10
14	10	Dim lighting	9	9
15	11	Standard lighting	11	11
16	11	Dim lighting	11	10
17	12	Standard lighting	12	12
18	12	Dim lighting	11	11
19	15	Standard lighting	15	15
20	15	Dim lighting	14	13

We have exhibited the results of our trials, as well as a table of observations titled table 1 that we used to calculate the accuracy of our system. When a human face was in an acceptable distance from the camera, the face mask detection algorithm was able to detect a human face with an accuracy of 97 percent and determine whether the person was wearing a mask or not based on the features that could be seen. After repeated tests, the mask identified and undetected accuracy ranged from 92 percent to 100 percent. The optimal setting to run the system was bright light since it produced the best accuracy of up to 100 percent, but when the light was dim, the accuracy dropped and usually stayed between 86 percent and 95 percent. Overall, the technology correctly predicted whether or not people were wearing masks 96 percent of the time.

6.1.3 Social Distancing Detection:

Figure 17: Frame with crowed people applied with social distancing module

Figure 18: Social distancing module used at a non-inclined angle

Table 2: Outcome for Social Distancing Method

Sr no.	People in Frame	Lighting Conditions	Humans Detected	Correct Prediction
1	10	Standard lighting	10	10
2	10	Dim lighting	10	10
3	12	Standard lighting	12	12
4	12	Dim lighting	12	12
5	18	Standard lighting	18	18
6	18	Dim lighting	18	18
7	22	Standard lighting	22	22
8	22	Dim lighting	22	21
9	29	Standard lighting	29	29
10	29	Dim lighting	29	29
11	32	Standard lighting	32	32
12	32	Dim lighting	31	30
13	35	Standard lighting	34	33

14	35	Dim lighting	33	33
15	40	Standard lighting	40	40
16	40	Dim lighting	37	37
17	44	Standard lighting	42	40
18	44	Dim lighting	43	43
19	50	Standard lighting	50	49
20	50	Dim lighting	48	48

As seen in the figure 17, our system was able to recognise human figures in video frames rather effectively, even when given a frame of crowded individuals, and assess whether humans are obeying or breaching social distance regulations. Experiment was conducted by positioning camera from a top-down perspective and observation table is shown as table 2. Our system was able to recognise practically all human figures from a decent height with a 98 percent efficiency. Although we have seen that its efficiency only drops by 2%, to 96 percent, when the same camera is put at the same angle but under different environmental conditions, in this example when the lights are dim. We also put our surveillance system to the test against shadows, which might look to be human-like from afar. In the case of shadows, our algorithm performed admirably, with a 97 percent accuracy in identifying between a genuine human and a shadow. Even though our system is able to detect humans affectively when the camera is not at a top-down approach and is directly in front of humans, as shown in figure 18, we can see that it is not able to recognise the distance correctly because it does not factor in the depth and thus gives us bad predictions. As a result, it is recommended.

Chapter-7

Summary & Conclusions

7.1 Summary & Conclusion

One of the most significant precautions in avoiding physical contact that could contribute to the spread of coronavirus is social distancing. Viral transmission rates will be increased as a result of non-compliance with these rules. To implement the proposed features that are crucial to stop the spread of coronavirus i.e., social distancing and wearing face masks, a system was created using Python and the OpenCV library. In the first and second features, we also employed the YOLO method to recognize humans and classify faces, respectively. The current research examines whether or not people were wearing face masks. Real-time video streams and Images were used to test the models. The model's optimization is a continual process, and we're fine-tuning the hyperparameters to provide a very accurate answer.

Chapter-8

Future Scope

8.1 Future Scope

Because of its high precision and low error rate, the proposed method can be easily implemented in real scenarios, such as schools, public places like airports, bus stations, banks, tourist attractions, museums, and many more. We can check that a safe distance is maintained by monitoring the space between two persons, which can assist us in containing the virus. However, it does not work as well for all camera angles or setups. A top-down camera angle is the best-advised camera angle. The system can be designed to work for different camera angles as well.

References

[1] H. Liu, S. Chen and N. Kubota, "Intelligent Video System and Analytics: A Survey," *IEEE Transactions on Industrial Informatics,* vol. 9, no. 3, pp. 1222-1233, 2013.

[2] J. Redmon, S. Divvala, R. Grishick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788, 2016.

[3] D. S. Gothane, "A Practice for Object Detection Using YOLO Algorithm," *International Journal of Science Research in Computer Science, Engineering and Information Technology,* vol. 7, no. 2, pp. 268-272, 2021.

[4] B. Qiang, R. Chen, M. Zhou, Y. Pang, Y. Zhai and M. Yang, "Convolutional Neural Networks-Based Object Detection Algorithm by Jointing Semantic," *Segmentation for Images, Sensors,* 2020.

[5] S. Gupta and D. T. U. Devi, "YOLOv2 Based Real Time Object Detection," *International Journal of Computer Science Trends and Technology* , vol. 8, no. 3, 2020.

[6] G. S., "Real-Time Object Detection with Yolo," *proceedings of the International Journal of Engineering and Advanced Technology (IJEAT)* , 2019.

[7] T. K. M, V. P. M., Y. B., J. S. and L. Dr. K., "Video Analytics on Social Distancing and Detecting Mask - A detailed Analysis," *International Journal of Advanced Engineering Research and Science (IJAERS),* vol. 8, no. 5, 2021.

[8] S. Gupta, V. Dhok, A. Chandrayan and S. Tiwari, "Facemask Detection using OpenCv," *International Journal of Advanced Research in Computer and Communication Engineering,* vol. 10, no. 6, 2021.

[9] R. Kakadiya, R. Lemos, S. Mangalan, M. Pillai and S. Nikam, "AI Based Automatic Robbery/Theft Detection using Smart Surveillance in Banks," *Proceedings of the Third International Conference on Electronics Communication and Aerospace Technology,* 2019.

[10] S. Patil, M. Shidore, T. Prabhu, S. Yenare and V. Somkuwar, "Theft detection using computer vision," *International Journal of Advance Research, Ideas and Innovations in Technology,* vol. 5, no. 1, pp. 567-569, 2019.

[11] P. K. Mishra and G. P. Saroha, "A Study on Video Surveillance System for Object Detection and Tracking," *IEEE,* 2016.

[12] A. Bari, S. Waseem, S. and S. , "Social Distancing Through Image Processing, Video Analysis, and CNN," in *International Conference on Computational Intelligence and Emerging Power System*, 2022.

[13] A. H. Ahamad, N. Zaini and M. F. A. Latip, "Person Detection for Social Distancing and Safety Violation Alert based on Segmented ROI," *IEEE International Conference on Control System, Computing and Engineering (ICCSCE2020),* 2020.

[14] D. k. R. S. M. P. V. R. D. P. R. Harish Adusumalli, "If the input is a video stream, the picture or a frame of the video is initially delivered to the default face detection module for detection of human faces. This is accomplished by first enlarging the picture or video frame, then identifying the blob ins," *Proceeding of the third international conference on intelligent communication technologies and virtual mobile networks,* 2021.

[15] H. Adusumalli, D. Kalyani and R. K. Sri, "Face Mask Detection Using OpenCV," *Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV 2021).,* 2021.