Design and Development of a Framework for Human Behavioral Analysis

Project Work submitted to

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

in partial fulfillment of the requirements
for the award of degree of

BACHELOR OF ENGINEERING
in
COMPUTER SCIENCE AND ENGINEERING

Submitted by

ABHISHEK RAJU N	**1KG16CS001**
BHAVYA ASHOK K	**1KG16CS015**
JAYASHREE S	**1KG16CS039**
MOHAK N	**1KG16CS053**

Under the Guidance of

Prof. Deepa Yogish
Associate Professor
CSE
K S School of Engineering and Management

Department of Computer Science and Engineering
K.S. School of Engineering and Management
No. 15, Mallasandra, off Kanakapura Road, Bengaluru-560109

2019-2020

K.S. SCHOOL OF ENGINEERING AND MANAGEMENT
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the project work entitled **"Design and Development of a Framework for Human Behavioral Analysis"** is a bonafide work carried out by

ABHISHEK RAJU N	**1KG16CS001**
BHAVYA ASHOK K	**1KG16CS015**
JAYASHREE S	**1KG16CS039**
MOHAK N	**1KG16CS053**

in partial fulfillment for the award of **Bachelor of Engineering** in **Computer Science and Engineering** of **Visvesvaraya Technological University, Belagavi,** during the year 2019-2020. It is certified that all the suggestions indicated during internal assessment have been incorporated in the report and this report satisfies the academic requirement with respect to **Project Work (15CSP85)** prescribed for the degree.

Internal Guide	**Head of the Department**	**Principal**
Prof. Deepa Yogish Associate Professor Dept. of CSE	**Prof. Veena R S** HOD Dept. of CSE	**Dr. K. Rama Narasimha** KSSEM

Examiners

Name and Signature of Examiner-1

Name and Signature of Examiner-2

K.S. SCHOOL OF ENGINEERING AND MANAGEMENT
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DECLARATION

We

ABHISHEK RAJU N	**1KG16CS001**
BHAVYA ASHOK K	**1KG16CS015**
JAYASHREE S	**1KG16CS039**
MOHAK N	**1KG16CS053**

the students of **BE VIII Semester (Computer Science and Engineering)** declare that the project entitled **"Design and Development of a Framework for Human Behavioral Analysis"** is carried out by us as a partial fulfillment of academic requirement of degree under **Visvesvaraya Technological University.** The content in the report are original and are free from plagiarism and other academic dishonesty and are not submitted to any other University either partially or wholly for the award of any other degree.

NAME	USN	Signature
ABHISHEK RAJU N	1KG16CS001	
BHAVYA ASHOK K	1KG16CS015	
JAYASHREE S	1KG16CS039	
MOHAK N	1KG16CS053	

Date:
Place: Bengaluru

ACKNOWLEDGEMENT

The successful completion of this project was made possible with the help of guidance received from our faculty members. We would like to avail this opportunity to express our sincere thanks and gratitude to all of them.

We are grateful to our management for providing the necessary infrastructure and an ambience environment to work. We express our profound gratitude to **Dr. K Rama Narasimha, Principal and Prof Veena R S, Head, Department of Computer Science and Engineering, KSSEM, Bengaluru** for giving us opportunity and support to complete the project work.

We are grateful to my guide **Prof. Deepa Yogish Associate Prof., Department of Computer Science and Engineering, KSSEM, Bengaluru** for her valuable suggestions and advice throughout our work period.

We would also like to thank all the staff members of Department of Computer Science and Engineering for their support and encouragement. Finally, we would like to thank all of our friends without whose help and encouragement this project would have been impossible.

Definitely most, we want to thank our family. Words cannot express our thanks to our family members for all their love and encouragement.

ABHISHEK RAJU N	**1KG16CS001**
BHAVYA ASHOK K	**1KG16CS015**
JAYASHREE S	**1KG16CS039**
MOHAK N	**1KG16CS053**

ABSTRACT

Human Behavior analysis is a behavioral science discipline focused on how behavior's change or affected by the environment. Aim of action recognition is to automatically identify the action of a person based on some kind of sensor data. Also, we can detect the behavior among a group of individuals which can be achieved by the interpretation of data gained studying the natural movement of groups or objects. Masses of bodies, particularly human, are the subjects of these crowds tracking analysis that include how a particular crowd moves and when a movement pattern changes.

In a classroom environment, one of the biggest obstacles you may face is dealing with children with challenging behavior. Not only does the child affect you trying to teach a lesson and maintain control, but they also disrupt the productivity of the class as a whole. Children who display challenging behavior don't usually do so 'just because they want to'. There's often a reason behind their behavior or it might be their only way of telling you something's wrong. All behavior is a form of communication.

In this context, by using or applying technology to the classroom environment it helps the teacher's or professors in effective management of classroom by getting to know how many students are paying attention in the respective class or lecture and take necessary steps to make the lecture more interesting or engaging.

TABLE OF CONTENTS

LIST OF FIGURES

GLOSSARY

1) Affective computing	Is the collection of the methods used for affect recognition by considering physiological signals like blinking yawning etc., speech, gestures and facial expression cues for behavior detection.
2) Deep Learning	One of the latest trends in behavior recognition, emotion detection is an affective deep network.
3) Crowd analysis	A growing requirement for a smarter video surveillance of private and public space using intelligent vision systems which can differentiate what is semantically important in the direction of human observer as normal behavior and abnormal behaviors.
4) Object Tracking	Approach to recognize and detect the behavior in a crowd which identifying the position of each person in the same video sequence.
5) Behavior Recognition	Based on all the feature datasets obtain one can attempt to the fullest to analyze and train the system to detect the behavior.

Chapter 1

INTRODUCTION

Human Behavior analysis is a behavioural science discipline focused on how behavior's change or affected by the environment. Aim of action recognition is to automatically identify the action of a person based on some kind of sensor data. Also, we can detect the behavior among a group of individuals which can be achieved by the interpretation of data gained studying the natural movement of groups or objects. Masses of bodies, particularly human, are the subjects of these crowds tracking analysis that include how a particular crowd moves and when a movement pattern changes.

Affective computing is one of the latest trends in the field of emotion/sentiment recognition. Affective computing is the collection of the methods used for affect recognition by considering physiological signals, speech, gestures and facial expression cues for behavior detection

Human behavior modelling normally requires a predictive mechanism that can predict a future behavior of an individual, such as click, a buy, a call, or exercise. It can take the observed attributes of the individual and the social network as input and provide a predictive score as output. The higher the score, the more likely the individual will exhibit the predicted behavior. Given its open-ended nature, the application domains of human behavior prediction are very broad, e.g., healthcare, politics, e-commerce, psychology, personal life, classroom environment, etc

1.1 Problem Statement

In a classroom environment, one of the biggest obstacles you may face is dealing with children with challenging behavior. Not only does the child affect you trying to teach a lesson and maintain control, but they also disrupt the productivity of the class as a whole. Children who display challenging behavior don't usually do so 'just because they want to'. There's often a reason behind their behavior or it might be their only way of telling you something's wrong. All behavior is a form of communication. The safety of the student's well-being is paramount in their development of social ties with peers and their instructor. As education becomes more inclusive, teachers need to become more aware of how to organize groups of students and how the students are arranged can lead to a favorable environment. Wellorganized classrooms are an important component to classroom functions as it leads to more dialogue and formative assessment. Education becomes less of a chore and more enjoyable when students grow as a group which can lead to the reduction of students acting out destructively. In order to affect students, a teacher needs to monitor and modify the influence students have on one another.

1.2 Scope of the Project

In the 21st century, the use of information technologies [IT] has become a necessity in many everyday tasks. In fact, technology is used in almost all everyday tasks. The reflections of the intensive use of information technologies have made it necessary to make certain changes in the use of technology in education. Use of technology in education is of great importance for both realization of effective learning-teaching processes and raising individuals possessing competences necessary for being a member of the 21st century society. There are many variables that affect the use of technology in education. These variables include educational institutions, infrastructure facilities, curriculum, student, teacher count, competences of students and teachers related to technology use, skills related to classroom management and technology use, sufficient in-service training for teachers and adequateness of technical support. Effective classroom management, one of these variables, is a substantial precondition for providing meaningful technology integration. On the other hand, it can be said that IT has also some effects on classroom management. Teaching and management processes in classrooms, which may be defined as an environment specific to the teacher and the learner, change shape and direction with the design of instructional technologies.

To explore how human behaviour would impact a surrounding in a closed environment. The specific motivator was to evaluate the behaviour of an individual or a person in a group so that we can predict or diffuse any possible threats or deviance. In a classroom environment it becomes too difficult for the teacher to keep track of all the students in his/her class. Usually this often leads to a negative classroom environment because of some students who tend to disrupt the entire class due to which learning becomes a bit difficult to accomplish.

Teachers need to manage all resources in the classroom environment. Classroom management is one of the most important factors to increase the efficiency of education, ensure interaction, and reaching educational goals. Classroom management is concerned with principles, concepts, theories, methods and techniques related to planning, organization, application and assessment processes to reach educational goals. Classroom management seems to have multiple dimensions and may be addressed in five main dimensions; management of physical structure of the classroom, teaching management, time management, management of intra-classroom relationships, and behavior management.

Management of physical structure of the classroom involves ensuring the compatibility between the use of classroom tools and educational goals, and organizing factors related to the classroom environment. For this reason, to make learning process more productive, the physical structure should include motivating and interesting elements and allow learners to feel comfortable and peaceful.

Teaching management is to plan teaching methods and principles in accordance with the environment to achieve educational goals through plans, programs and activities. This planning must have a structure that supports the flow of learning-teaching activities and student engagement.

Time management is to create a time plan for classroom activities.

Management of intra-classroom relationships is to ensure communication between students to allow for effective teaching in the learning environment. The warm relationship established between teacher and students within the framework of respect and mutual trust has critical importance to provide a positive classroom climate.

Behavior management is the organization of the classroom environment in a way that negative behaviour's in the classroom atmosphere are prevented. The observed undesirable student behaviour's generally can obstruct that it from being continued the learning-teaching process as healthy way. This situation can affect classroom management negatively. In this context, it is required that teachers control negative student behaviours in scope of class rules for an efficient classroom management.

1.3 Description of the project

It may be predicted that the structure and roles of classroom management will change as a result of changing classroom structure and interaction types due to integration of technology in education. In this context, by using or applying technology to the classroom environment it helps the teacher's or professors in effective management of classroom by getting to know how many students are paying attention in the respective class or lecture and take necessary steps to make the lecture more interesting or engaging.

1.3.1 Existing System

- Existing system consists of Human Behavior Detection Framework with multiple modalities.
- Indirect approach which is widely used for crowd analysis.
- Human action recognition method which combines the high-level pose features and the mid/low level action trajectory features.
- Efficient methodology of estimating the amount of variation on the local regions of a face image due to varying expression, non-uniform lighting and partial occlusions.

1.3.2 Proposed System

First the video is recorded using a webcam. The camera will be positioned in front of the students to capture the front face image. From the video, the frames are extracted. Face is detected in the frames. After detecting the face, facial landmarks like positions of eye, nose, and mouth are marked on the images. From the facial landmarks, eye aspect ratio, mouth opening ratio are quantified and using these features and machine learning approach, a decision is obtained about the drowsiness of the student. We are using Haar-Cascade algorithm. If drowsiness is detected, an alarm/buzzer will turn on.

Chapter 2

LITERATURE REVIEW

Survey on Human Behavior Recognition using Affecting Computing [1]

Affective Computing is one of the latest trends in the field of emotion or sentiment recognition. Human Detection can be achieved through the consideration of different behavioural cues such as physiological signals and psychological signals.

EXISTING RESEARCH

It mainly focuses on how well the recognition system is being analyzed in the new technologies of deep learning. Now let us see the different cues used for behaviour detection along with different approaches used.

Physiological Signals:

Every human behavior is characterized by two dimensions, first arousal and second is valence, where arousal is about the intensity of emotion while valence qualifies positivity or negativity of emotion felt by someone. The autonomic nervous system is responsible for controlling human body functions such as heart rate, respiration, digestion, arousal etc. These signals (Electroencephalography (EEG) Signals and Heart Rate) of the human body can be further used to detect human behavior

Psychological signals:

Facial expressions and appearance play an important role in behavior detection. Deep CNN along with an additional layer of global pooling average is trained for facial expressions, which is later used for behavior detection. The behavior like excess stress in humans lead to unwanted aggression, agony etc. hence detecting stress in the early stage is very much important. The person expresses his emotions and stress through a variety of gestures and speech modulations. Some nonverbal signs like gestures, facial expressions, postures and body language used in communication convey stress.

BEHAVIOUR RECOGNITION APPROACHES USING AFFECTIVE COMPUTING

Naïve Bayes Classifier: In behaviour recognition, it can be used for gesture recognition. The naïve Bayes classifier takes full video or sequence of videos as input for gesture detection. Necessary probabilities are estimated through the basic training process. Gesture recognition starts with the assumption that we are aware of the starting point of gesture. The process proceeds with extracting code-words from each sequence and determining the gesture.

K nearest neighbour (KNN): Most of the researchers have been using a KNN approach for detecting human depression using EEG data [19]. The process starts with capturing of the EEG signals with help of the Emotive EPOC+ device. Each captured signal is pre-processed using notch and high bandpass filter, later normalized to remove noise from the signals. Extracted data is trained as per the required classes. It uses the Euclidian distance formula to calculate the distance and classify the data.

Support Vector Machine (SVM): In the detection of human behaviour like anxiety, SVM can be used to classify the data over other classification algorithms. SVM also can be used to classify the facial expressions based on extracted features.

Deep Learning: One of the latest trends in behaviour recognition, emotion detection is an affective deep network. The foundation of the deep network is an artificial neural network, a neural network with a maximum number of neurons, complex connected network with automatic feature extraction.

Deep Belief Network (DBN): It is used for high-level feature ex-traction from the raw input vectors.

Convolutional Neural network (CNN): The effectiveness of the CNN in image processing is one of the main reasons why re-searchers prefer it in the field of behaviour recognition. CNN also can be used in sentiment analysis.

Recurrent Neural network (RNN): It can be used to process and analyze a stream of data. Many natural language processing applications use RNN.

CHALLENGES

We have now understood different behavior cues and approaches to effectively to recognize human behavior and emotions. Different approaches like deep learning, SVM, KNN and naïve Bayes classifier, which can be used in affective computing. Based on the accuracy and effectiveness over the large datasets, deep learning approaches are more suitable in the field of affective computing. The future trend in behavior recognition and analysis is towards a multimodal approach and effective use of deep learning approaches to optimize the results. However, there are some drawbacks in using them such as in:

Deep neural network- Slow learning process.

Deep belief network (DBN) -It is expensive as far as training is concern

Convolution neural network (CNN) - It requires the large dataset to train the network.

Recurrent Neural network (RNN) -It is applicable only in the situation where the output of current state depends on earlier state output.

A Review on Abnormal Crowd Behavior Detection [3]

Crowd analysis is a growing requirement for a smarter video surveillance of private and public space using intelligent vision systems which can differentiate what is semantically important in the direction of human observer as normal behavior and abnormal behavior's. There are two approaches for crowd detection such as direct and indirect approach.

Direct approach is a detection-based method that detects each individual in a scene using segmentation.

Indirect approach is a map-based approach where it detects visual features that are mapped to the number of people.

EXISTING RESEARCH

Different stages for abnormal crowd behavior detection are pre-processing, object tracking and behavior recognition.

Pre-processing is one of the important steps for feature extraction. Most of the researchers are intended to analyze and learn the pattern of abnormal in crowd scene.

Object Tracking approach to recognize and detect the behavior in a crowd which identifying the position of each person in the same video sequence.

Behavior Recognition: now based on all the feature datasets obtain one can attempt to the fullest to analyze and train the system to detect the behavior

CHALLENGES

Once abnormal behavior is detected, many times researchers face problem of mix behaviors when some abnormal crowd behaviors are associated with some others behavior. Those abnormal behaviors are happen at the same time.

Also, Indirect approach considered whole crowd as a global entity in analysis whereas direct approach has a problem of occlusion because crowd is analyzed by treating a collection of individuals. So that, indirect approach is widely used for crowd analysis.

Confidence-Based Human Action Recognition with Different-Level features [3]

Action Recognition is a challenging problem in a computer vision and is widely studied due to its wide range of applications. This work targets on interactive action between human and object in some complex real-world situation. In action recognition, actions of people can be categorized into several types Single Human Action, Crowd's, Human to Object Interaction and Interpersonal Action.

Previous human action recognition methods can be briefly divided into two groups. One uses **statistical representations** of local descriptors to recognize actions. The features extracted by these methods are often considered as low/mid-level features. For ex. propose a method which computes histogram of orientated gradient (HOG) of motion history image for differential frame images to recognize single-person actions (e.g. run, walk, wave). Histogram of optical flow (HOF) is also a histogram-based local video descriptor. The difference between HOF and HOG descriptors is that HOF uses orientation of optical flow to represent motion information in the frame at a given instant. Furthermore, HOG descriptor and HOF descriptor are combined to characterize both static appearance and local motion information in video sequences. However, since these methods do not consider spatio-temporal structure information, they may not be recognizing human-to-object interactive behaviors correctly. The second type of human action recognition methods focuses on **pose-based action recognition features**. They rely on human pose estimation which computes location of body's joints. Pose-based features are considered as high-level features.

EXISTING RESEARCH

Confidence-based Action Recognition with Different-Level Features Different-level features

Low/Mid-level Features:

Dense Trajectories (DT) feature which is one of the most efficient way to detect the low/mid-level features. Firstly, the feature points are densely sampled on each frame at several spatial scales. Secondly, Optical flow fields are applied to track feature points on each spatial scale separately. Two-frame motion estimation [23] is applied to compute optical flow fields. Concatenation of each feature point in subsequent frame forms a trajectory during the tracking process. Trajectories with a sudden large displacement will be considered as an error, and they are removed during the post-processing processes. After all trajectories are ready, 4 types of descriptors are computed for each trajectory. The first one is trajectory shape descriptor. The rest three types of descriptors are HOG, HOF and MBH. HOG extracts static appearance information while HOF and MBH capture motion information

High-level Features:

High-level pose features (HLPF) capture spatial and temporal relations derived directly from positions of human joint. HLPF is a typical high-level representation of action recognition. Many studies shown HLPF has good performance on classifying single-human actions and human-to-object interactions. HLPF is based on the 15-point human body model. Positions of the 15 body joints including shoulders, elbows, wrists, knees, ankles, hips, neck, face and belly are firstly normalized to reduce the influence of different human sizes in videos. After the pre-processing, nine kinds of features are extracted from the joint positions.

CHALLENGES

The objective of this paper was to compare whether and when the recognition system with DT achieves higher accuracy than the one with HLPF with complete and no pose knowledge.

For the method without pose knowledge, pose estimation is applied. The result suggests that our confidence-based method (50.17%) significantly outperforms the methods with DT (46.67%) and HLPF (32.15) on the whole dataset. It indicates that the system using HLPF or DT based on the confidence level is better than the one only using HLPF or DT.

Although satisfying performance is achieved by high-level pose features in human action recognition, the features significantly reply on accurate pose estimation. With poor estimation on pose, high-level pose features may perform worse than mid/low-level features. Therefore, a human action recognition method is proposed to combine the high-level pose features (i.e. HLPF) and the mid/low-level Dense Trajectory features (i.e. DT).

Behavior Detection and Analysis for Learning Process in Classroom Environment [4]

Classroom observations have been widely used in education over the past couple of decades to measure effective teaching practice. For example, to study the off-task behavior of elementary students; a protocol, namely BERI (Behavioral Engagement Related to Instruction) to describe the in-classroom behaviors of college students. It is notable that, these traditional observation methods rely on human observers, which is difficult to extend to large-scale or long-term measurement. With the development of intelligent teaching system, it became popular to analyze the in-system behavior of students.

EXISTING SYSTEM

In this paper, they have implemented a kind of automatic behavior measurement system, which utilizes the Microsoft Kinect devices to record the students' performance in classroom. Several Kinect devices are installed under the ceiling of one classroom. The facial images of attended students are collected and recognized. The typical gestures of students (such as sitting, raising hand, standing, sleeping and whispering) are also detected and recorded. A queue-based analysis engine is proposed to distinguish the meaningful learning behaviors from those pointless actions. Firstly, the data acquisition module controls the Kinect devices and collects the sensing data, including the depth image, the skeletal data. Then these data are fed into the behavior detection module. The facial images of attended students are collected and recognized. Finally, the behavior analysis module checks the temporal relations of these actions, and filters out the useful learning behaviors in the context of one learning process.

CHALLENGES

- When the system is extending the system to a larger scenario, more Kinect devices should be equipped and will cause higher deployment cost. It is not scalable to deploy more Kinect devices under one ceiling.
- The data collected by Kinect device is recorded in a coordinate system taking itself as the coordinate origin. It is required to convert the gesture data in oblique direction to the normal classroom coordinate system. What's more, it is difficult to distinguish some similar actions (such as bowing, writing and etc) because of the oblique direction.

Human Behavior Analysis from Video Data Using Bag-of-Gestures [5]

Feature extraction and gesture recognition from non-verbal language are of particular interest in the analysis of psychological factors. Using standard computer vision feature extraction and machine learning approaches, work is based on the definition of a large vocabulary of gesture units, which is said to be Bag-of Gestures (BOG), and a probabilistic modelling of human gestures.

EXISTING RESEARCH

For the feature extraction stage of a gesture recognition system, they have defined a set of simple visual features. These features are based on face detection, skin modelling, and feature tracking processes. They have used the Face Detector of Viola & Jones in order to detect the region of the face and defined the origin or coordinates, Inner pixels of the detected region are used to train a skill color model which is used to look for hand/arm candidate regions. Finally, those blobs connected with the highest density corresponds to the regions of interest, which are tracked using mean shift. All the spatial coordinates of the detected regions are computed in reference to the face coordinates and normalized using the face area. With the computed feature space, they have performed an initial experiment where 15 bachelor thesis videos were recorded. Using the social signal indicators defined in ,computed a set of activity, stress, and engagement indicators from the extracted feature space .Using the score obtained , they have categorized the videos in two levels: those with the lowest score, and those with the highest score, and trained a Discrete Adaboost binary classifier. Applying stratified ten-fold cross-validation, and obtained interesting results, showing high prediction performance of student score based on his/her non-verbal communication using the extracted features. Moreover, Adaboost margin in order to rank features by relevance. Furthermore, tested the system on different applications, such as Sign Language recognition using a novel multi-target dynamic gesture alignment, Attention Deficit Hyperactivity Disorder (ADHD), corporal physiotherapy analysis, and inpatient monitoring, with high success. Hence, could successively tried to get the similar gestures into one unit.

CHALLENGES

Computing the feature-space from a large set of videos involving several gestures and composing them into behavior vocabulary is a little complicated.

Chapter 3

HARDWARE AND SOFTWARE REQUIREMENTS

3.1 Hardware Requirements

- Octa Core (Processor).
- 4GB Ram Minimum
- 1GB Cache Memory
- Hard disk 50 GB
- Microsoft Compatible 101 or more Key Board
- Raspberry pi 3
- Webcam

3.2 Software Requirements

Language Used	:	Python (version 3)
Operating system	:	Windows 10/8/7 (64-bit)
Packages to be installed	:	pillow,opencv,imutils,scipy,playsound,dlib,pygame

Chapter 4

REQUIREMENTS SPECIFICATIONS

4.1 Objectives

The intent of our project is to develop a system to detect behavioral anomalies' which can be achieved through the following objectives–

1. Capturing input video.
2. Pre-processing the captured input data.
3. To extract the required features.
4. Recognizing the behavior based on the input taken.

4.2 Methodology

1) Methodology for objective-1

The video capture unit or webcam is used to record the video in real time of the frame containing the subject's face through a camera. The video is sampled with some frequency and the sampled frame is sent to the face detection unit.

```
# start the video stream thread
print("[INFO] starting video stream thread...")
vs = VideoStream(src=0).start()
time.sleep(1.0)
```

Figure 4.1: Snippet which shows the initialization of the video capture unit

2) Methodology for objective-2

Face detection unit receives the sampled video frame from the video capture unit. The images from the video capture unit are the RGB image and for the very dim light condition, we perform low-light image enhancement and noise elimination. For improving the accuracy of our system, we eliminate the noise of the image before amplifying it through contrast enhancement. The above process is divided into two subtasks. At first, for de-noising the image we apply the super pixel based adaptive noising and secondly, for amplifying the image we use luminance adaptive contrast enhancement method. We need to denoise the image before contrast enhancement, so that the noise has been eliminated before its amplification through contrast enhancement. The above method increases the accuracy of our system significantly as it eliminates heavy noise, texture blurring and over-enhancement from the image which is then processed as accordingly. The image is changed to the grayscale image because for face detection we do not need the color data. For face detection in the frame, we use the rapid object detection which uses the boosted cascade of the classifier by Viola-Jones that works with the Haar-like features.

3) Methodology for objective-3

For face detection in the frame, we use rapid object detection which uses Haar-cascade algorithm. The face detection method returns the abscissa, ordinates, length, and breadth of the rectangle boxed in the facial image. From the facial landmarks the eye aspect ratio and mouth aspect ratio are quantified.

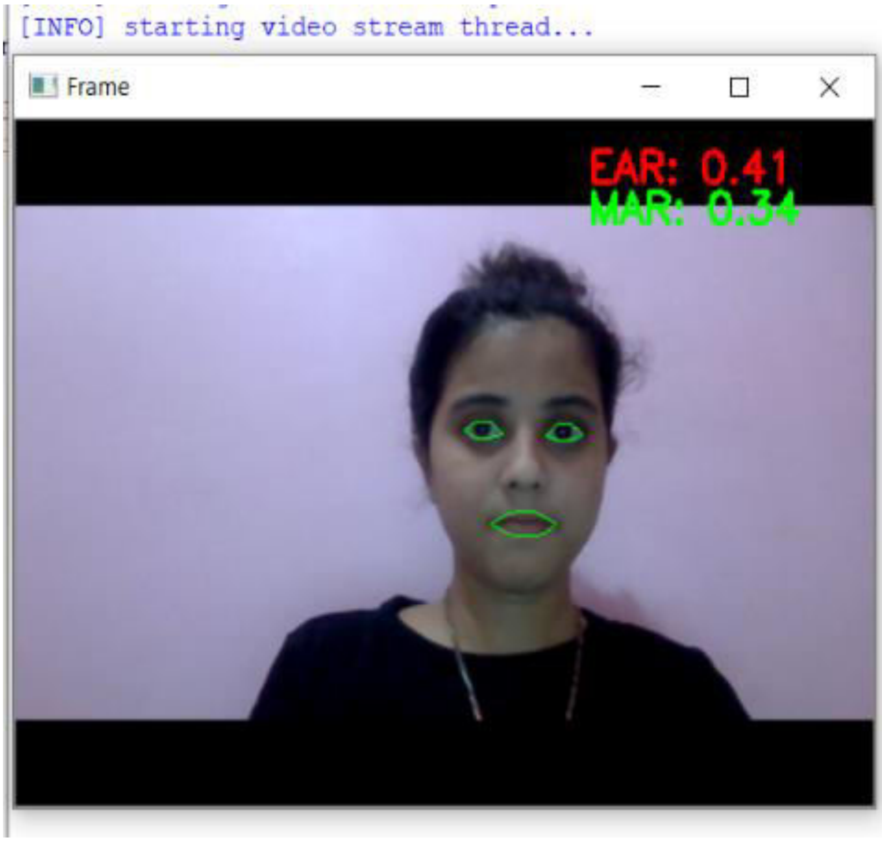


Figure 4.2: Shows facial features extraction

4) Methodology for objective-4:

From the face detection unit, we get a sequence of eye and mouth ratio of the person. Now from this extracted dataset, we can perform behavioral detection analysis on various facial features with the help of Haar-cascade algorithm. Also, we use optical flow and corner detection mechanisms to determine the head movements or the position of the subject's head in order to evaluate if the subject is paying attention or not.

Chapter 5

DESIGN

5.1 System Architecture:

The process begins with the user executing the program , the person faces the camera ,the computer stares taking the live input via webcam or an external camera unit ,from the obtained live input data ,it is converted into grey scale by using two operations ,one is to de-noise the image the next is to enhance to image ,this is done to increase the efficiency of the system , we use the rapid object detection which uses the boosted cascade of the classifier by Viola-Jones that works with the Haar-like features , From the face detection unit, we get a sequence of eye and mouth ten we recognize the behavior based on the input taken.

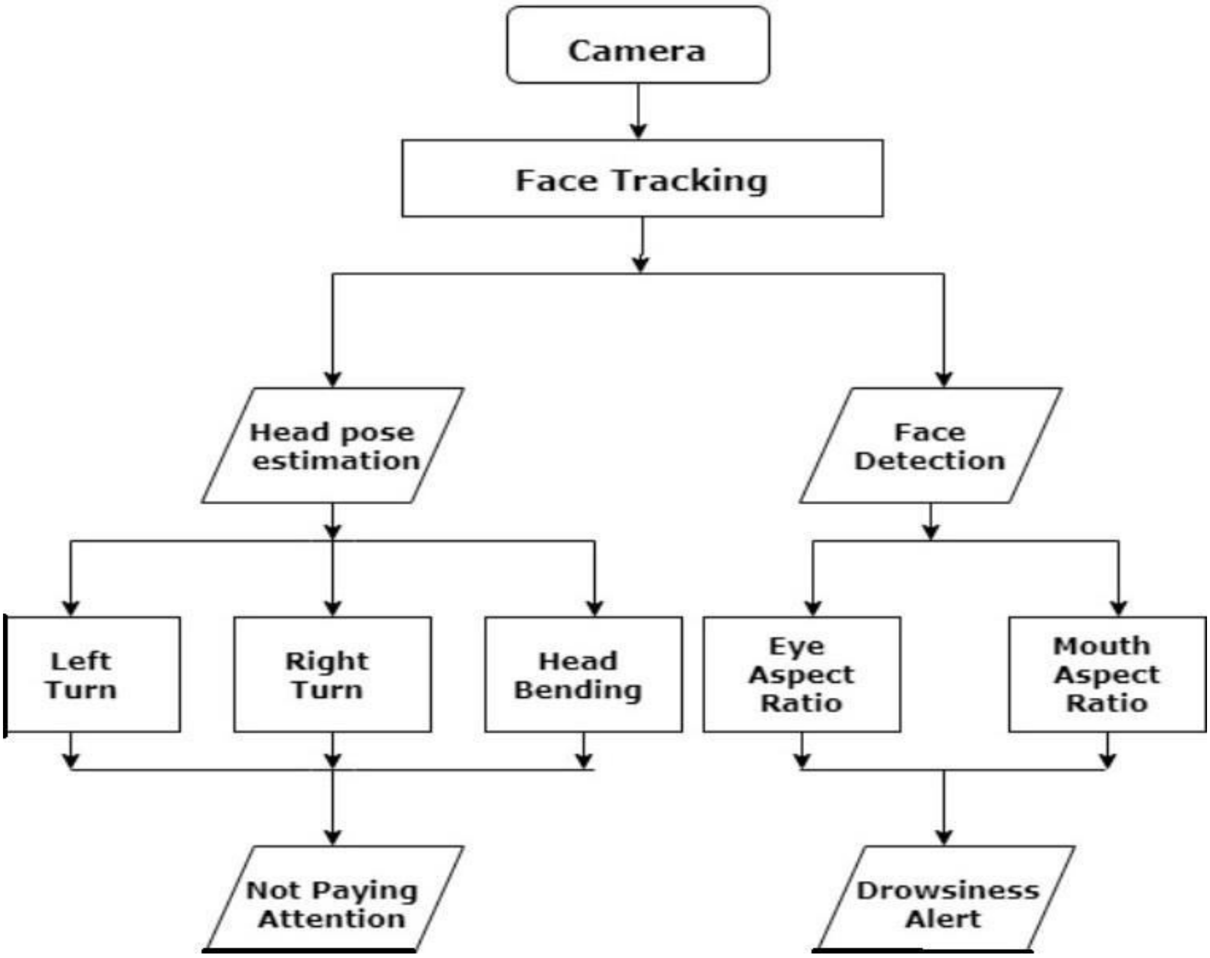


Figure shows the system architecture

5.2 Flow Diagrams:

This section shows the flow carts of two models implemented ,Eye and Mouth detection ,head bending detection , For all the models we start with execution of the program ,The real time video input is taken via webcam ,the obtained image or video is converted to grey scale For improving the accuracy of the system ,the grey scale image is pre-processed in face detection unit ,with the help of haar Cascade Classifier ,we extract the facial features like eye aspect ratio and mouth aspect ratio ,we also extract the positional feature like head bending ratio ,we calculate the eye ,mouth and head aspect ratio ,if the ESR and MSR are greater or less than the given threshold value and if the head bending ratio exceeds the threshold an alert message is being displayed.

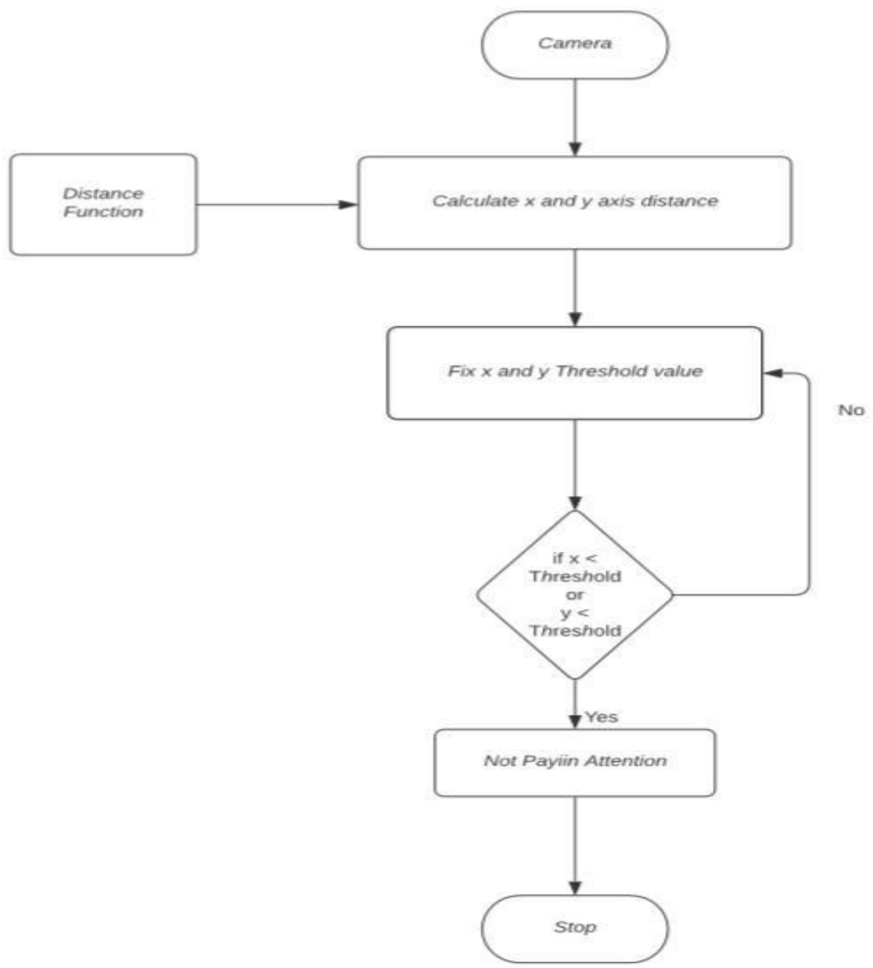


Figure shows the flowchart of head movement

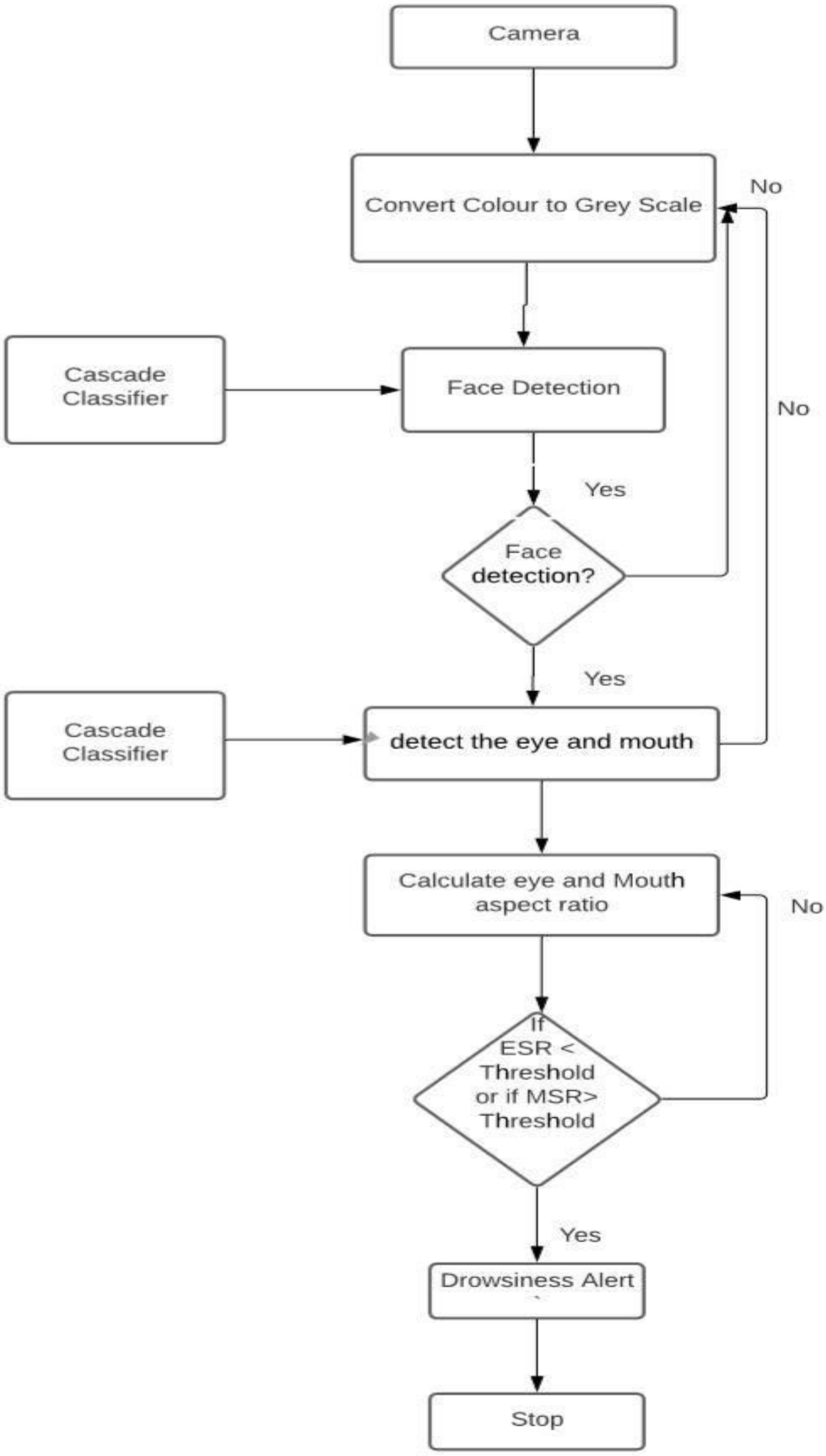


Figure shows the flowchart of drowsiness alert by facial features exctraction

5.3 Use Case Diagram: -

Actor interacts with the initial system and the camera is responsible for all the activities that involve in capturing a video steam of the subject, from the obtained images from the video ,features such as eye and mouth features are extracted ,and output is predicted based on the input taken

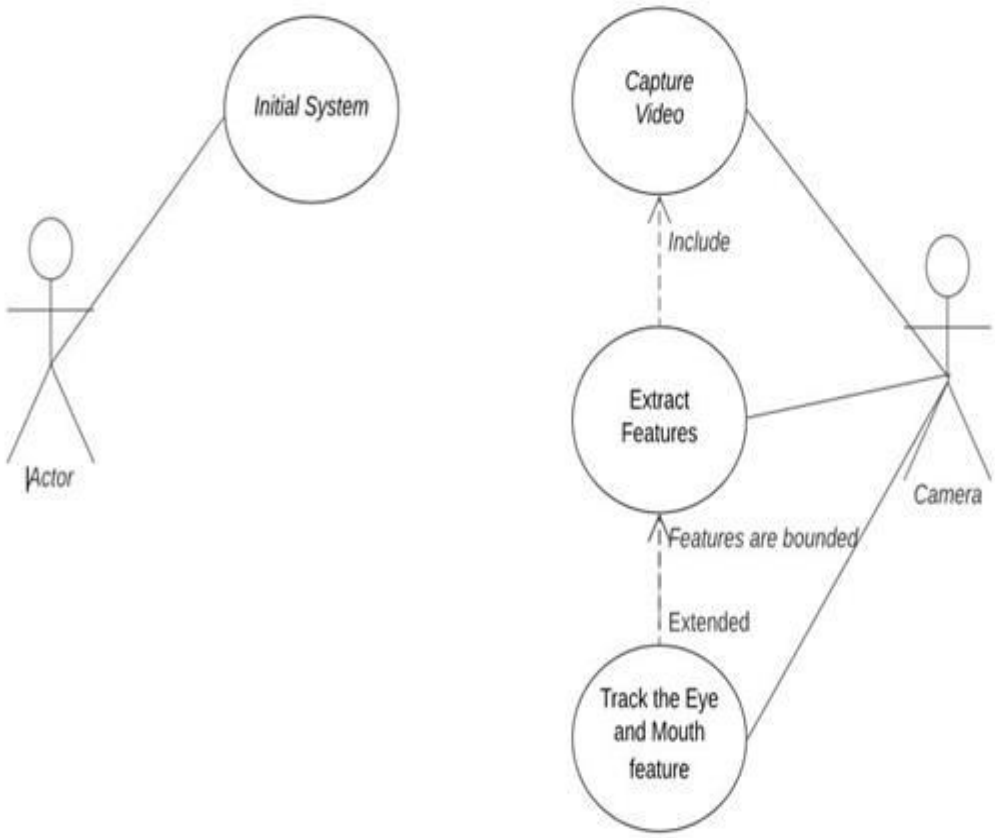


Figure shows the use case diagram for the framework

Chapter 6

IMPLEMENTATION

Capturing video stream

- We would take the video in real time by using VideoStream library. Using the code snippet below we instruct the webcam to start capturing the video and display a message while this process is taking place.

```
# start the video stream thread
print("[INFO] starting video stream thread...")
vs = VideoStream(src=0).start()
time.sleep(1.0)
```

Figure 6.1: Code snippet for video stream

Pre-Processing the input data:

- The video we obtain from the webcam is in RBG form. So, as soon as we get the input data, we convert it into grey scale since facial features extraction is very difficult in RGB form.

```
# loop over frames from the video stream
while True:
        # grab the frame from the threaded video file stream, resize
        # it, and convert it to grayscale
        # channels
        frame = vs.read()
        frame = imutils.resize(frame, width=450)
        gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)
        frame_gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)
        faces = face_cascade.detectMultiScale(frame_gray, 1.3, 5)
        for (x,y,w,h) in faces:
            cv2.rectangle(frame,(x,y),(x+w,y+h),(255,0,0),2)

        # detect faces in the grayscale frame
        rects = detector(gray, 0)
```

- Using the code snippet, we convert RGB image or video into grey-scale for easy facial features extraction.

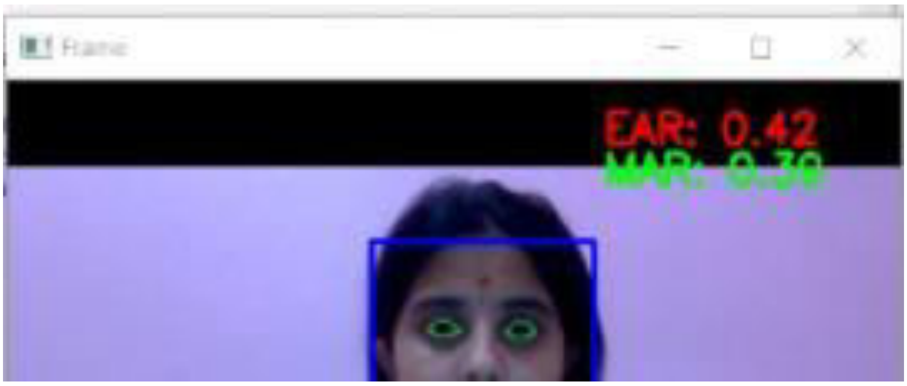
Tracking facial features

- Upon receiving the input data i.e., the video from the webcam, we track the noticeable facial features like the eyes and mouth and using corner analysis we track the face.
- We compute the Euclidean distances between two sets of vertical eye landmarks i.e. the vertical positions for the left and the right eye. This step gives the height of each eye.
- Since our both of our eyes lie on the same horizontal plane or axis, we calculate the Euclidean distance for the single horizontal eye landmark. This gives the width of the eye.

```
#to compute the distance between two set of vertical and horizontal eye landmark
def eye_aspect_ratio(eye):
        A = dist.euclidean(eye[1], eye[5])
        B = dist.euclidean(eye[2], eye[4])
        C = dist.euclidean(eye[0], eye[3])

        # compute the eye aspect ratio
        ear = (A + B) / (2.0 * C)
        return ear
```

- Based on the above values we compute the eye aspect ratio.
- As for Eye Aspect Ratio, it is used to detect the eye blinks (eyes open and close) using the ratio formula based on the eye's width and height.
- Since eye blinking is performed by both eyes synchronously the aspect ratio of both eyes is averaged.

- After successful tracking of the eyes we get the below result or output.



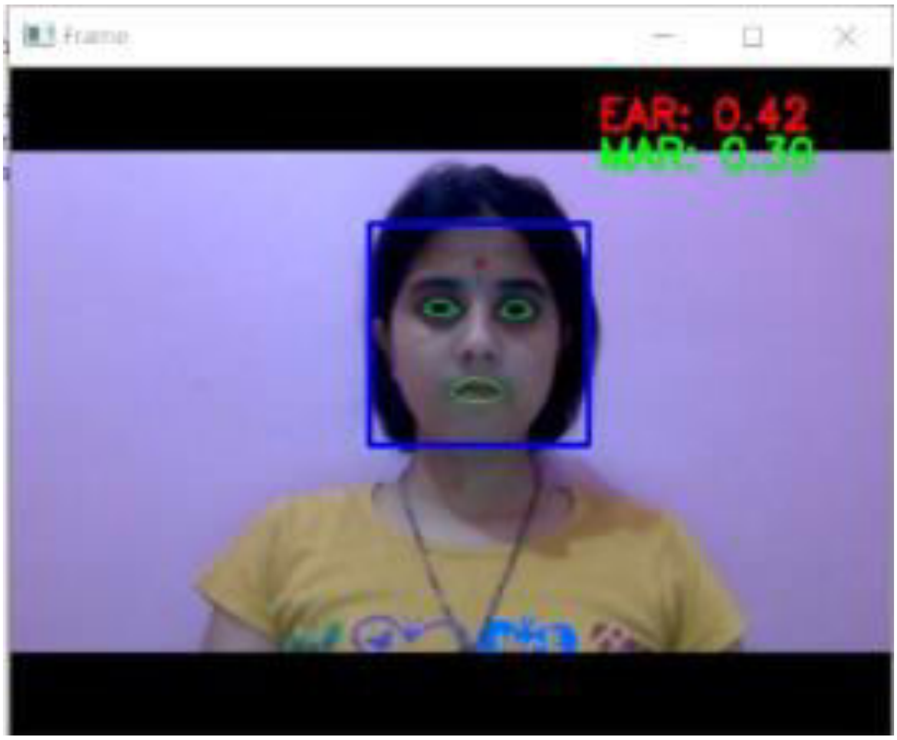
- Similarly, like the eye aspect ratio to determine the yawning parameter the aspect ratio of the mouth is calculated.

```
#compute the mouth acpect ratio
def mouth_aspect_ratio(mouth):

    A = dist.euclidean(mouth[2], mouth[10])
    B = dist.euclidean(mouth[4], mouth[8])
    C= dist.euclidean(mouth[3], mouth[9])
    D = dist.euclidean(mouth[0], mouth[6])
    mar = (A+B+C) / (3.0 * D)
    return mar
```

- When the mouth is closed the mouth aspect ratio is almost zero. When the mouth is slightly open the mouth aspect ratio increases slightly. But when the mouth aspect ratio is significantly higher than the threshold value, we have defined then it is clear that the mouth is wide open most probably for yawning.

- The below figure shows how the mouth is tracked.

- So, the final result or output we see on the screen after successful facial features extraction is as given below.



Detecting drowsiness

- We give specific threshold value for both the eyes and the mouth. We have also defined a counter which is used to compare with the consecutive number of frames.
- We have also defined a threshold value for consecutive frames for both the eye and the mouth.
- The eye aspect ratio is used to detect blinking action and the mouth aspect ratio is used to determine the yawning parameter.

```
# check to see if the eye aspect ratio is below the blink
# threshold, and if so, increment the blink frame counter
if (ear < EYE_AR_THRESH) or (mar > MOUTH_AR_THRESH):
	COUNTER += 1

	# checking the eyes and mouth aspect ration is below the thershold value if not drowiness alert will be displayed
	if (COUNTER >= EYE_AR_CONSEC_FRAMES) or (COUNTER >= MOUTH_AR_CONSEC_FRAMES):
		if not ALARM_ON:
			ALARM_ON = True
			t = Thread(target=sound_alarm,args=('alarm.wav',))
			t.deamon = True
			t.start()
		cv2.putText(frame, "Drowsiness ALERT!", (10, 30),
			cv2.FONT_HERSHEY_SIMPLEX, 0.7, (0, 0, 255), 2)

# otherwise, the eye and mouth aspect ratio is not below the threshold, so reset the counter and alarm
else:
	COUNTER = 0
	ALARM_ON = False
```

- If the eye aspect ratio is below the eye threshold and if the mouth aspect ratio is above the mouth threshold as we have defined then a counter is incremented.
- If the counter is more than the threshold number of frames for both the eye and the mouth then the alarm or buzzer is made to or it is set off.

Head movement tracking:

- As mentioned before only the eye and mouth aspect ratio cannot be used in all scenarios to accurately predict drowsy or alertness of an individual or a group.
- To overcome this, we take into account the head movements also.
- We mainly use two algorithms like Lucas kanade optical flow and Shi Tomasi corner detection.
- Optical flow is the pattern of apparent motion of image objects between two consecutive frames caused by the movement of object or camera.
- A corner can be defined as the intersection of two edges.
- Corner detection is an approach used within computer vision systems to extract certain kinds of features and infer the contents of an image.
- Corner detection is frequently used in motion detection, object recognition etc.

```
#dinstance function
def distance(x,y):
    import math
    return math.sqrt((x[0]-y[0])**2+(x[1]-y[1])**2)

#capture source video
cap = cv2.VideoCapture(0)

#parameter for ShiTomasi corner detection algorithm
feature_params = dict( maxCorners = 100,
                       qualityLevel = 0.3,
                       minDistance = 7,
                       blockSize = 7 )

# Parameters for lucas kanade optical flow algorithm
lk_params = dict( winSize  = (15,15),
                  maxLevel = 2,
                  criteria = (cv2.TERM_CRITERIA_EPS | cv2.TERM_CRITERIA_COUNT, 10, 0.03))
```

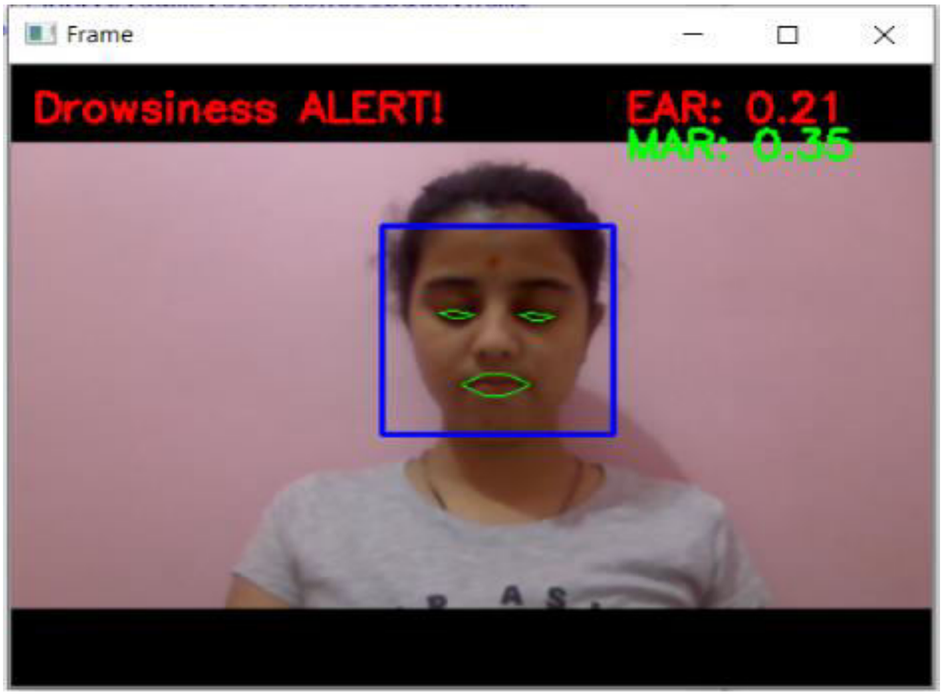
- To figure out the vertical and horizontal position of a person's head we use distance function.
- Then we define the parameters for both Shi Tomasi corner detection and Lucas kanade optical flow algorithm.
- Using both the above-mentioned algorithms we track the head movements continuously.
- The below figure shows the code snippet of how this is achieved.

```
x_movement = 0
y_movement = 0
gesture_show = 60 #number of frames a gesture is shown

while True:
    ret,frame = cap.read()
    old_gray = frame_gray.copy()
    frame_gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)
    p1, st, err = cv2.calcOpticalFlowPyrLK(old_gray, frame_gray, p0, None, **lk_params)
    cv2.circle(frame, get_coords(p1), 4, (0,0,255), -1)
    cv2.circle(frame, get_coords(p0), 4, (255,0,0))

    #get the xy coordinates for points p0 and p1
    a,b = get_coords(p0), get_coords(p1)
    x_movement += abs(a[0]-b[0])
    y_movement += abs(a[1]-b[1])

    text = 'x_movement: ' + str(x_movement)
    if not gesture: cv2.putText(frame,text,(50,50), font, 0.8,(0,0,255),2)
    text = 'y_movement: ' + str(y_movement)
    if not gesture: cv2.putText(frame,text,(50,100), font, 0.8,(0,0,255),2)

    #comparing the x and y movement with threshold
    if x_movement > gesture_threshold:
        gesture = 'Not Paying Attention'
    if y_movement > gesture_threshold:
        gesture = 'Drowsiness Detected'
    if gesture and gesture_show > 0:
        cv2.putText(frame,'Status: ' + gesture,(50,50), font, 1.2,(0,0,255),3)
        gesture_show -=1
    if gesture_show == 0:
        gesture = False
        x_movement = 0
        y_movement = 0
        gesture_show = 60 #number of frames a gesture is shown
```

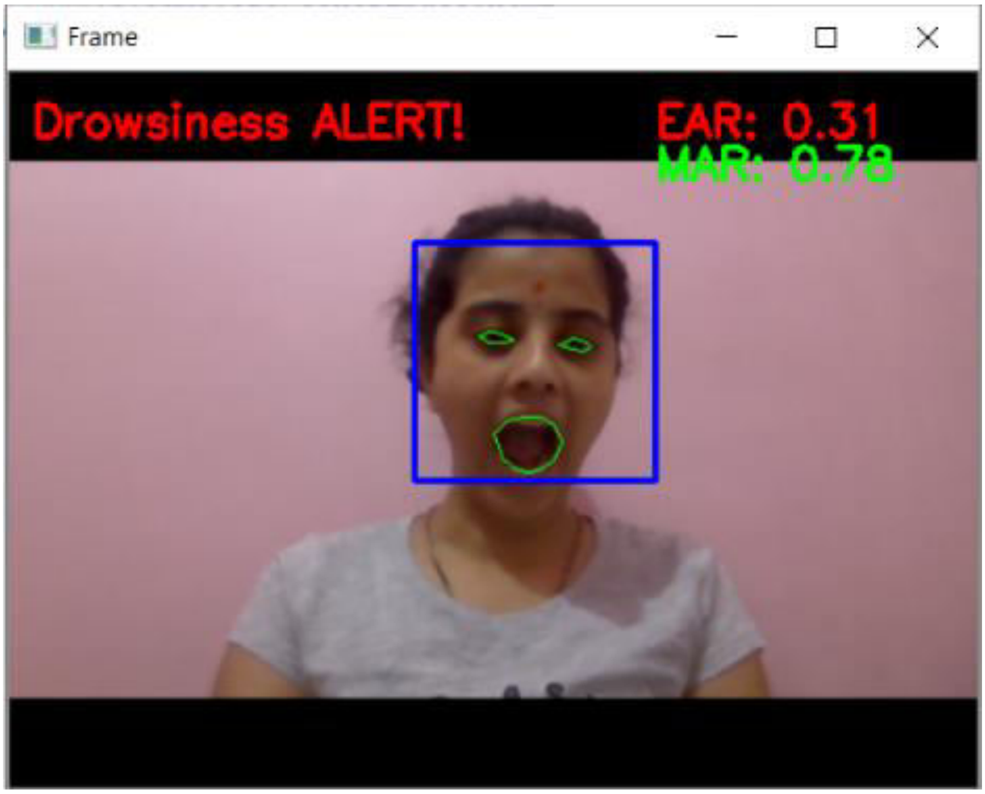
- We also define head movement threshold i.e. for turning the head left and right and also bending movement of the head in a downward manner.
- Based on the threshold values and using the vertical and horizontal position of a person's head we are able to predict if a person is paying attention or not.
- If there is an increase in the x and y co-ordinate of the head then that value would be compared with the initial threshold value of x and y. Then there would be a text display telling that person is not paying attention.

Results

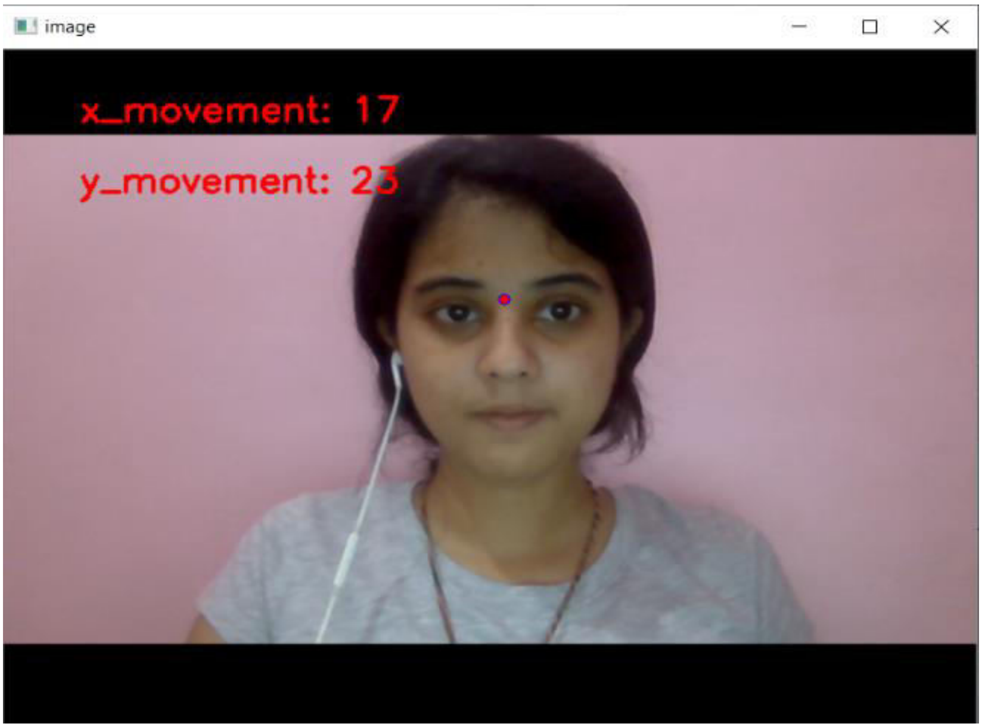
1) Drowsiness detection when eyes are closed



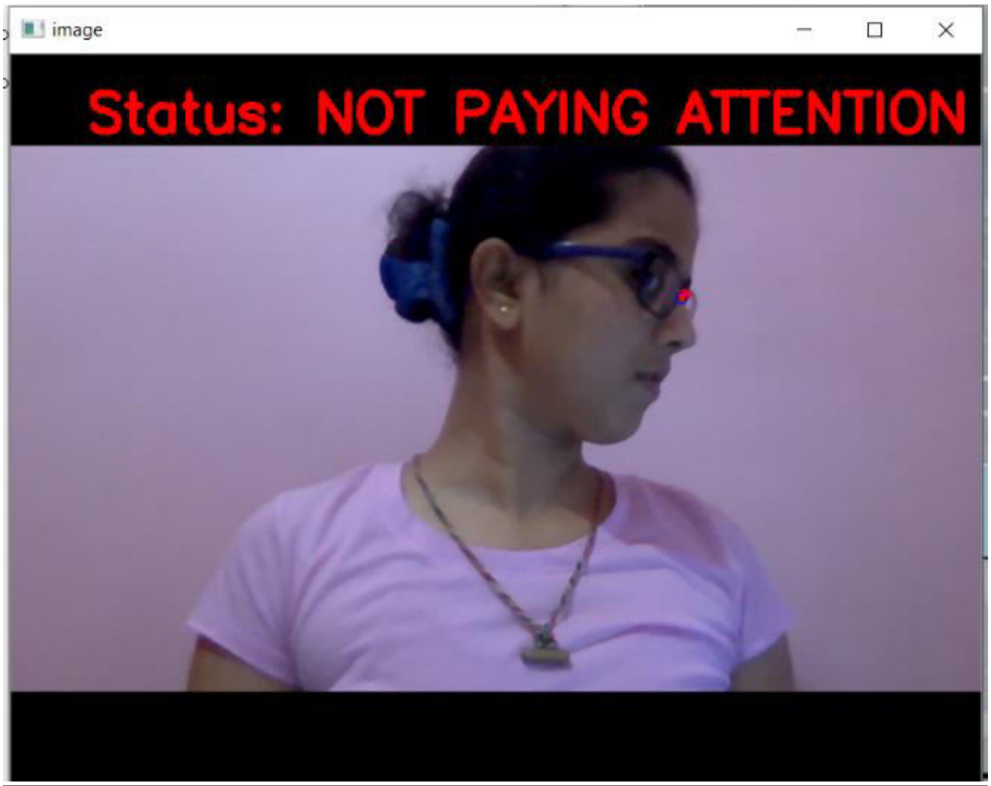
2) Drowsiness detection when mouth is opened



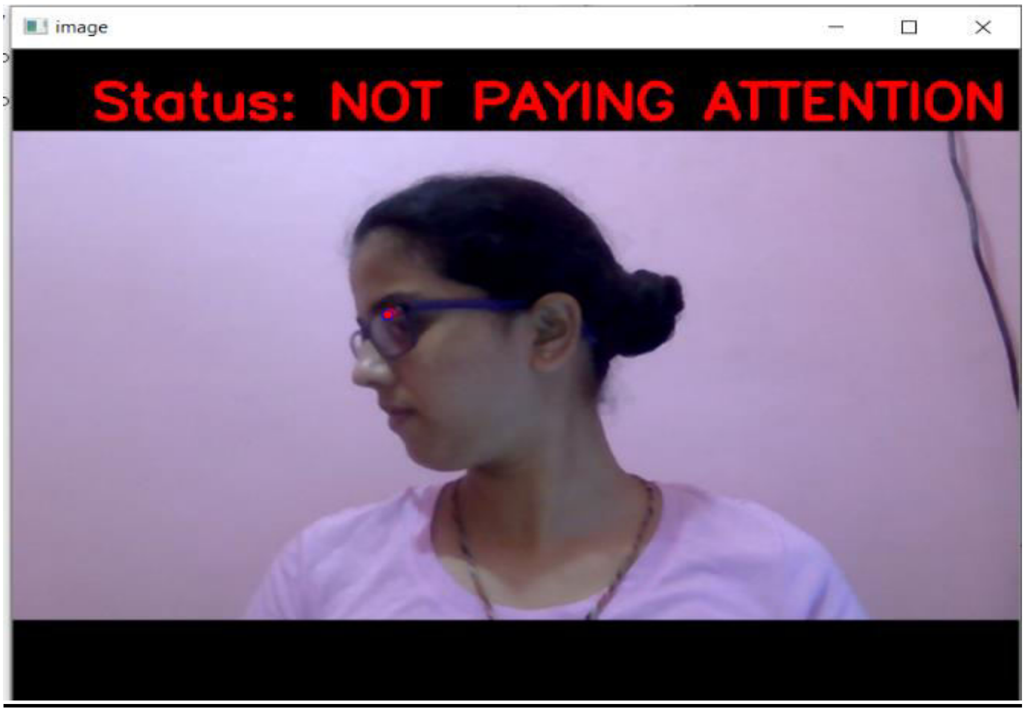
3) **Initial position of head**



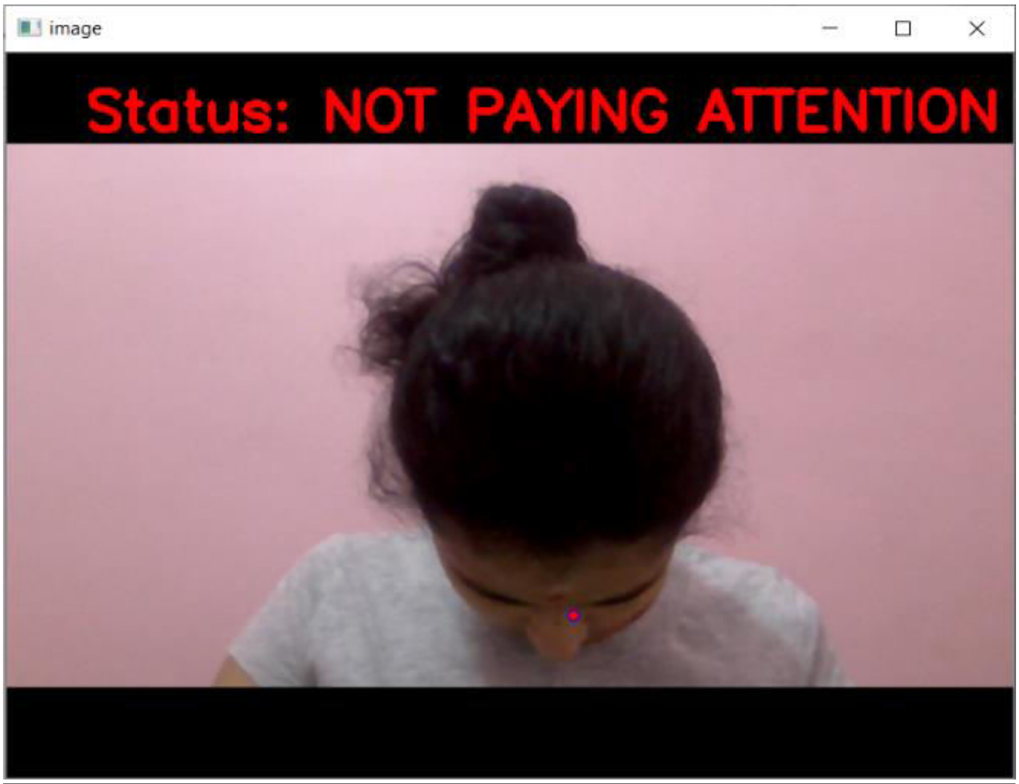
4) **Alert message displayed when head is tilted to the left**



5) **Alert message displayed when head is tilted to the right**



6) **Alert message displayed when head is bent down**



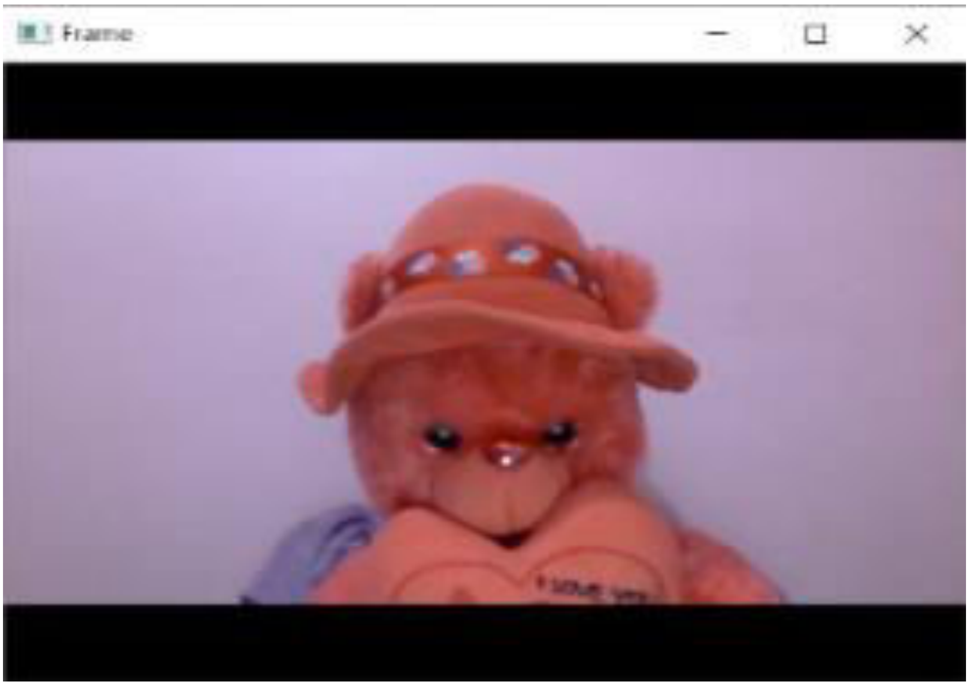
Chapter 7

TESTING

1) Recognizing human face:

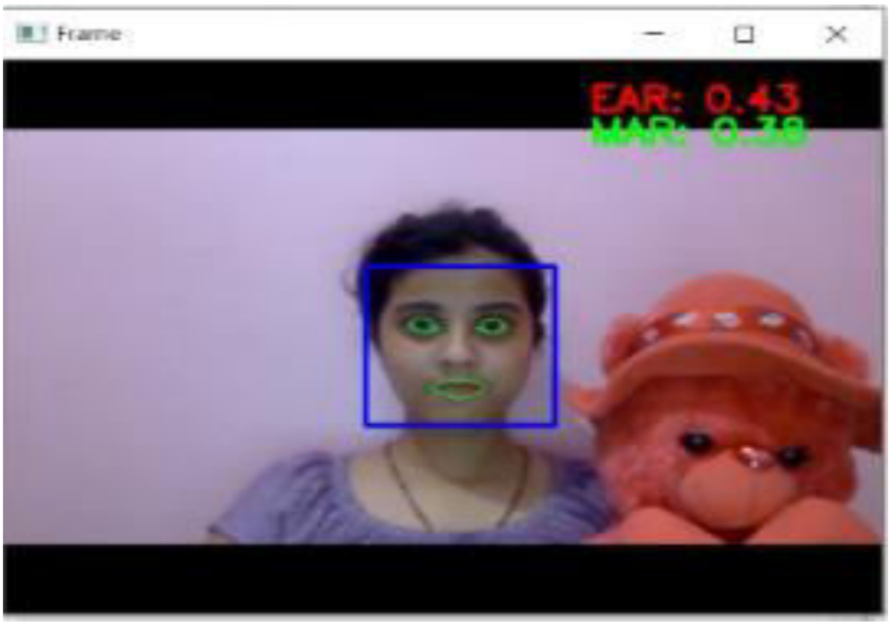
- As you can see in the figure above a human face is detected and the facial features such as eyes and mouth are extracted.

2) Not recognizing human face:



- From the above figure we can see that a non-human face is not detected.

3) Recognizing only human face when the frame consists of both human and non-human face:



Chapter 8

CONCLUSION

- Determining the complex human behaviour by focusing on analysis of a particular individual or to find deviant behaviour in a group.
- The system detects drowsiness whenever the eyes are closed or the mouth of the subject is opened wide and a buzzer is made to buzz during such an event.
- It also detects if a person is paying attention or not by detecting head movements.
- So, our project can be used by educator’s to better understand the classroom environment by understanding the state the students are in the classroom.
- We have taken into account the classroom as a dataset but our project can be incorporated in many wide areas like medicine, or to stop crime etc.

Chapter 9

FUTURE ENHANCEMENTS

- We wish to extend our project so that we can analyze an entire class or student data. This helps the faculty head to get an idea about the mental state of the students in the class and how they can improve to deliver the best teaching experience.
- As of now our project works only on python3 language, in the future we would like to make our project portable so that it can be implemented on a variety of programming languages and on different machines.
- We wish to create a compact and compressed mobile application.
- We would also like to incorporate our project in school buses to check if the bus driver is feeling drowsy or not and send timely updates to the representatives at the school about the situation of the driver.

BIBLIOGRAPHY

[1]. Nagesh Jadav ,Rekha Sugandhi 2018 –IEEE Global Conference on Wireless Computing and Networking (GCWCN) survey on human behavior recognition using affective computing.

[2]. Krutika Rohit ,Ms.Kinjal Mistree ,Mr.Jenish Lavji ,2017 International Conference on invention in information ,embedded and communication systems. A view on Abnormal crowd detection.

[3]. S.Greene, H.Thapliyal and A.Cabsn-Holt, "A Survey of Affective Computing for Stress Detection:Evaluating technologies in stress detection for better health", in IEEE Consumer Electronics Magazine,vol. 5, no. 4, pp.44-56, oct 2016.

[4]. I. Lefter, G.J.Burghouts and L.J.M.Rothkrantz, "Recognizing Stress Using Semantics and Modualation of Speech and Gestures",in IEEE Transaction of Affective Computing, vol. 7, no.2, pp. 162-175, April-June 1 2016.

[5] Satyanadh Gundimada, Vijayan Asari Electrical and Computer Engineering Department Old Dominion University, Norfolk – "Adaptive Confidence Level Assignment to Segmented Human Face Regions for Improved Face Recognition" IEEE 2005

[6] Xian Hu and Yongchang Jing – "Confidence based human action recognition with different level features" IEEE 2016

[7] Ahmad Jalal, Shaharyar Kamal and and Daijin Kim – "Individual Detection-Tracking-Recognition using Depth Activity Images" IEEE 2015

[8] Adri Priadana, Muhammad Habibi – "Face Detection using Haar Cascades to Filter Selfie Face Image on Instagram" IEEE 2019

[9] Dinah K. Ulfa & Dwi H. Widyantoro – "Implementation of Haar Cascade Classifier for

Motorcycle Detection" IEEE 2015

[10] Wenxiang yu, Jiapeng xiu, Chen Liu, Zhengqiu Yang – "A Depth Cascade Face Detection Algorithm Based on Adaboost" IEEE 2016

[11] Yun Yuan, Zhenjiang Miao, Shaohai Hu – "Real-Time Human Behavior Recognition in Intelligent Environment" IEEE 2006

[12] Zhang WeiWei, Li Wei – "A Deep Reinforcement Learning based Human behavior Prediction Approach in Smart Home Environments" IEEE 2019