TRACKING PEOPLE'S BEHAVIORS FOR DETECTING AND UNDERSTANDING THEIR SUSPICIOUS ACTIVITIES IN SOCIAL ENVIRONMENTS

A Thesis

Submitted for the Partial Fulfillment of the Requirements for the Degree of Master of Philosophy (M. Phil.) in Information and Communication Engineering

Submitted by:

Partha Pratim Debnath

Examination Roll No: 1811177501
Registration No: 1811177501
Session: 2017-2018
University of Rajshahi

Supervisor:

Dr. Md. Golam Rashed

Associate Professor
Dept. of Information and Communication Engineering
University of Rajshahi

October, 2021

Certificate

This is to certify that the dissertation titled "**TRACKING PEOPLE'S BEHAVIORS FOR DETECTING AND UNDERSTANDING THEIR SUSPICIOUS ACTIVITIES IN SOCIAL ENVIRONMENTS**" is a unique work done by Partha Pratim Debnath, roll no-1811177501 under my guidance and supervision and is submitted to University of Rajshahi in partial fulfilment of requirement for degree of Master of Philosophy (M. Phil.) in Information and Communication Engineering during the academic year 2017-2018.

05/08/2021

Dr. Md. Golam Rashed
Associate Professor
Department of Information and
Communication Engineering
University of Rajshahi
Rajshahi- 6205
Bangladesh

01/07/2021

Dr. Dipankar Das
Professor
Department of Information and
Communication Engineering
University of Rajshahi
Rajshahi- 6205
Bangladesh

Declaration

I hereby declare that the dissertation titled "**TRACKING PEOPLE'S BEHAVIORS FOR DETECTING AND UNDERSTANDING THEIR SUSPICIOUS ACTIVITIES IN SOCIAL ENVIRONMENTS**" submitted to University of Rajshahi in partial fulfilment of requirement for the degree of Master of Philosophy (M. Phil.) in Information and Communication Engineering is a result of original research work done by us. It is further declared that the thesis report has not been previously submitted to any University for the award of degree of Master of Philosophy.

With regards and gratitude,

Partha Pratim Debnath
24 October, 2021

Dedicated

To my respectable parents

And

To my supervisor

Acknowledgements

I would like to thank my supervisor **Dr. Md. Golam Rashed**, Associate Professor, Department of Information and Communication Engineering, University of Rajshahi, Bangladesh, and my co-supervisor **Dr. Dipankar Das**, Professor, Department of Information and Communication Engineering, University of Rajshahi, Bangladesh for their kind help and guidance. I cannot only just say thanks for his tremendous support and help, I feel motivated and encouraged every time I attend their meeting. Without their proper guidance, this thesis work could not be materialized.

I am indebted to all of the faculty members of the Department of **Information and Communication Engineering, University of Rajshahi, Bangladesh** as well as to the members of examination committee for their help to accomplish my thesis work.

I would like to thank the ICT Division, Ministry of Post, Telecommunication and Information Technology, The People's Republic of Bangladesh for their research grants.

I owe many thanks to my classmates and friends to give me mental support and help me to solve many thesis oriented problems. Finally, I would like to show respect to my parents and my spouse for their proper care and support.

The Author

24 October, 2021

Abstract

This thesis represents an approach to develop a real-time autonomous invigilation and interrogation system suitable for different social environments to track people's cheating behaviors and to detect their lie. To this journey, a real-life smart class room scenario is primarily considered as social environment and we started with detecting the student's patterns of crimes by studying their psychology in smart classroom during examination period. In a smart classroom scenario, the built-in video camera of each computer will capture continuous video frames and feed it to our developed system for tracking student's suspicious behaviors. From the captured continuous video frames, we have detected target student's Visual Focus of Attention (VFOA) to track his/her suspicious gaze direction. A 3D head tracker is used to extract the facial regions from each video frame. From the extracted facial regions, face points and eye regions are detected. Using a Vector Field of Image Gradient (VFIG), the pupil of eye is pointed out from the detected face points and eye regions. In our approach, we have also used the face points to create a rectangle outside of the left eye. In our experiment, we have used the head movement along with the change of coordinate of one eye (left eye) with respect to the left eye rectangle to detect the gaze of suspect by applying different techniques under different lighting conditions and different distances from the camera and using different participants. Additionally, suspicious body movement have been tracked using Canny Edge Detector for slow, very slow and high dynamics with optimal accuracy for the same purposes. We also proposed an approach to track the fraud candidates by detecting the age and gender of the examinee in examination hall room by matching it with the pre-stored age and gender information. Texture variation of wrinkle density in the forehead, eye lids, and cheek area has been considered as key clues for age and gender classification. We have also included a real time emotion detection mechanism to our system. Besides VFOA tracking, change of shape of eyebrow has been detected using optical flow features to build a smart and robust autonomous interrogation system.

Furthermore, the viability of the proposed real-time autonomous interrogation system is demonstrated by experimenting with participants in a smart classroom under controlled environment. Finally, the system is tested to validate its effectiveness. Our obtained result shows that, we can efficiently track the focus of attention of the examinee (the average accuracy is 80%) from the coordinate of the pupil of the eye combining with head position and consequently we can detect the sustained attention together with transients very

effectively and control the attention by deploying an alarming system if inattention is detected. Canny edge detector shows best accuracy (>90%) when the examinee takes 3, 4 or 5 second for a proper movement. Our intelligent system works perfectly in happy, fear, surprise and neutral emotions detection and provides better accuracy to detect 21 to 35 and 36 to 50 age groups with standard deviation 5 and 7 respectively. It also shows optimal accuracy to classify examinee of different age and genders. It is also revealed that in controlled environment our system shows fair accuracy of lie detection (upto 100% accuracy for different types of suspicious symptoms detection). So the proposed system can detect different types of crimes based on the external symptoms provided by suspect in social environments like smart class room and interrogation room.

Keywords: *Suspicious behavior detection, Eye center localization, VFOA tracking, Lie detection.*

Index

List of Figures

List of Tables

Abbreviation and Acronyms

VFOA	Visual Focus of Attention
FOV	Field of view
VFIG	Vector Field of Image Gradient
ASM	Active Shape Model
CFV	Central Field of View
NPFV	Near Peripheral Field of View
RNPFV	Right Near Peripheral Field of View
LNPFV	Left Near Peripheral Field of View
FPFV	Far Peripheral Field of View
RFPFV	Right Far Peripheral Field of View
LFPFV	Left Far Peripheral Field of View
CNN	Convolutional Neural Network

Chapter 1
Introduction

This chapter outlines the research background, motivation, and objectives. A review of different works in the literature was also included. This chapter also gives a brief overview of the instruction to read.

1.1 Background

Human behavior detection and control of suspicious activities in social settings (e.g. Examination Halls, Museums, Conference Centers and Exhibition Centers, etc.) have been recognized for many years [1] as their actions reverberate with their interest, intentions, and preferences in the circumfluous area. This includes how people move, what they notice through different social spaces, where they focus their attention and distraction [2], [3], and finally what they capture. Such types of information can help the conveners make the typical social spaces more efficient, attractive, and safe. As a result, many types of research have already been carried out in social spaces, where researchers have given their attention to capturing the bodily actions of individuals by observing their different movements [4], [5], [6]. But very few works [7] have concentrated on identifying suspicious activities of people rather than simply monitoring their behaviors of this type. For example, it is very difficult for the surveillants to know how student cheats when writing on their answer scripts in an exam hall or to detect lies of the suspect in real-time invigilation system. We believe that it is possible to detect and understand suspicious activities if we track suspects' various bodily actions and other parameters like Age, Gender, and Emotion in different social spaces.

To achieve this task, in our approach, we have taken the help of different computer vision tools applied over the images captured by cameras (built-in webcam or external cameras connected to the system). The camera will capture real-time video and the system developed within each of the laptops will analyze each frame for suspicious behavior. We use eye pupil rotation detection [8], [9] combined with head pose detection to detect the Visual Focus of Attention (VFOA) of the selected person. In this process, the Haar cascade classifier detects the target object [10] using the AdaBoost algorithm. In a smart classroom scenario, Canny edge detectors [11] have been used to identify the random movement of the target person [12]. In recent years a number of motion detection techniques have been proposed (e.g., [13], [14], [15]). A human body motion detection method has been tested using a canny edge detector that

performs better in comparison to others. After detecting the examinee's random body movement, it will be merged into the examinee's VFOA controlling system and then the system will decide whether the target examinee is involved in cheating or not. After that, the age, gender, and emotion of the examinee are also detected to increase the system's accuracy. Because sometimes various individuals (with varying age or gender) play as a proxy candidate which is very common in Bangladesh. To control this illegal scenario, we may capture the real-time age and gender of the examinee and match it with the pre-stored ones to detect the false candidates. Moreover, when the normal behavior of the examinee changes slowly then he or she will be monitored closely because an examinee should be impartial and behave neutrally in the test space. Any deviation from his/her normal behavior gives a clue to the suspicious activities.

In our work, VFOA detection has also been used in a real-time interrogation system. Humans have a well-defined "rigid" skull structure and facial muscle structure - this means there are finite numbers of facial expressions a human can perform. These are called Facial Action Units (FACS) and there are 46 such Action Units. To this point, our proposed technique finds out the best parameters of lie detection from a visual domain, such as- Random Eye Movement, changes in the orientation of the head and shape of eyebrow changes, etc. Based on this, we shall create our knowledgebase to store how people reacted to different stimuli. This knowledgebase will help us to detect the phase of the questionnaire (performed while interrogation) and its corresponding possibility of the lie. A low-cost USB camera captures continuous video of the suspect while interrogation and feeds it to our proposed system. A multistage computer vision approach has been utilized to track the most critical parameters that show the symptoms of the lie.

Most of the suspicious behavior monitoring procedures in our country are manual. Large-scale observations of students in the examination hall using a few invigilators are very difficult and complex tasks. That's why we are proposing a new sensing system framework, to automatically analyze the suspect's behavior in the examination hall as well as real-time interrogation scenario and finally detect various types of suspicious activities immediately in an artificial intelligent environment. Our proposed technique is cost-effective and unbiased. It also provides proper video evidence of invigilation and interrogation and can be a helpful tool to the existing manual system.

1.2 Motivation

There were certain inspiring reasons triggered to do this thesis work in the field of Artificial Intelligence. Traditional examination hall monitoring and controlling system include a lot of manual works where dealing suspicious activities of students are one of the challenging issues. Although the vigilance team is always on duty to ensure the quality of the environment of any examination hall since a few decades ago, examination hall monitoring and the controlling system has been evolved which are based on Video Surveillance System [1], [2], [3] or Staged Matching Technique [4]. With such a system, often vigilance team can monitor the examination hall remotely. But real-time detecting and controlling students' suspicious activities within an examination hall isn't still possible with only the video surveillance system. Thus, it is crucial to develop an intelligent system particularly suitable for autonomously detecting and controlling students' suspicious activities within an examination hall. To meet the demand, so far, some researchers are dealing with this issue and trying to develop some system.

Moreover, no research has been conducted to detect the proxy student utilizing age and gender detection and to detect the student's suspicious behavior employing emotion detection. There have been several works for human motion and emotion detection [15], [16] which are generally very costly and the data collection procedure is not user-friendly to be applied in the examination hall. Moreover, no works have been done to detect the very slow movements that frequently occur in the examination hall.

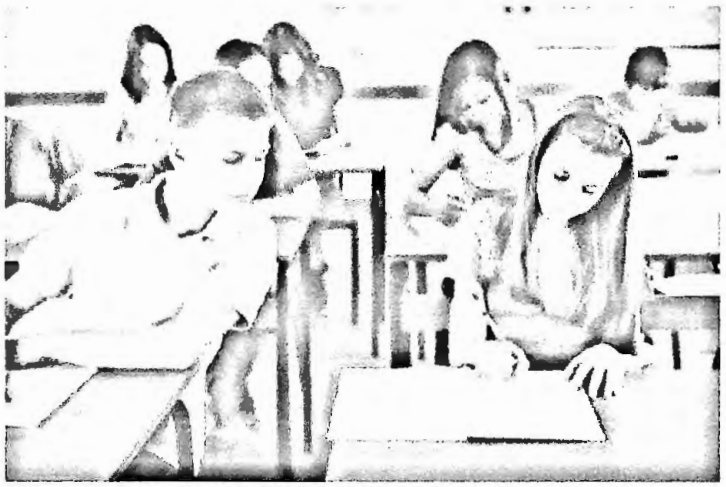


Figure 1.1: Student Cheating in an Examination Hall (Courtesy of Google).

When it comes to cheating detection, the most comprehensive survey that springs to mind is that of Dr. Donald McCabe and The International Center for Academic Integrity. It had been conducted throughout 12 years (2002-2015) across 24 high schools in the U.S. More than 70,000 students both graduates and undergraduates took part in it. And the results obtained were jaw-dropping, as 95% of the surveyed students admitted to cheating on a test and homework, or committing plagiarism [16].

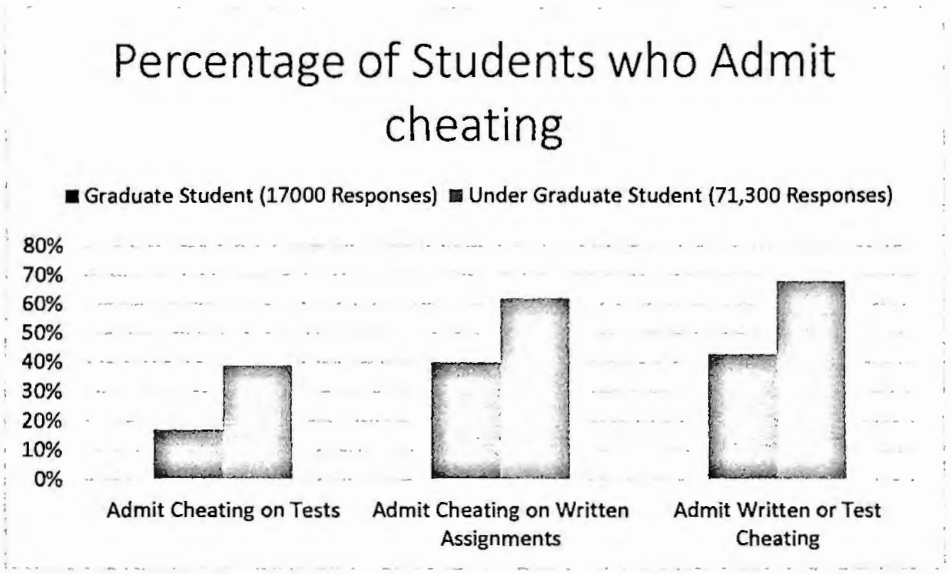


Figure 1.2: Cheating scenarios of students at different levels (Courtesy of Google).

So, we see that cheating in the examination hall is one kind of mental illness and a worldwide problem. For the contest of digital Bangladesh, a real-time and intelligent solution to this problem can be achieved, if we can deploy a cheating activities detection system in the examination halls. Generally, humans have very poor ability in detecting deceit and hostile intent, with an accuracy rate of 40-60% [5]. However, when one is being deceitful, she/he is making up something in the brain. This results in increased brain activity and rapid physiological responses which can be measured on the face (including facial blood flow pattern). Humans do not possess the ability to control these physiological responses to emotions. Stress causes abrupt changes in local skin temperature and distinctive facial signatures which provides the main key to lie detection- which is very hard to detect in real-time and the user interface is not friendly at all.

Considering these limitations in the social environment, we aim to build a robust system that can track suspicious behavior at a reasonable cost and with optimal accuracy.

1.3 Objective

We performed the whole research work in two different environments to make it generalized. One in a smart class scenario and the other in real time invigilation system.

The main purpose of this study in the smart classroom was to identify the student's cheating behavior in the examination hall. To accomplish this task, in our approach, we have used the built-in web camera installed in each of the laptops connected to our system's server via a network. Our proposed technique starts with the study of the psychological behavior of the examinee in a real-time environment to build a knowledgebase. This knowledgebase helps us to set the threshold values of crucial moments in different environments. A controlling signal from the server will automatically turn on the alarm if a cheat-prone environment is detected. For that,

- At first, we have tracked the visual focus of attention of the examinee with varying lighting conditions, distance from the camera, and movement speed.
- Then, we have detected the examinee's movement using a canny edge detector for slow and fast deviations.
- Nowadays, sometimes it happens that proxy candidates sit for an examination in the exam hall as a replacement of the original candidate. To avoid this problem, to identify a particular examinee, we have tried to detect the age and gender of the examinee and matched it with the pre-stored age and gender for each of the examinees.
- In general, students should be impartial and his/her behavior should be neutral in the exam hall. So, to detect suspicious behavior, we have tried to build a real-time emotion detection system. If an examinee communicates his or her abnormal behavior (rather than being neutral) then we carefully inspect the selected individual.
- In this study, we also contrasted the efficiency of the above-mentioned techniques.

In the second approach, we focused to detect and control the visual focus of attention of the suspect using his/her eye gaze and head movement direction to build up an automatic interrogation system- especially to detect lies. To this point-

- At first, we conducted psychological experiments on the sampled population to detect the different parameters connected with various symptoms when the suspect

tells lies and build our knowledgebase with the results. This knowledgebase helps the system to make strategic decisions and to optimize accuracy.

- Based on the knowledgebase, we have classified the crimes in different criteria and tried to prepared suitable questions for interrogation
- We classified interrogation conversations into different criteria and identified the fatal ones when the suspect starts to show symptoms.
- Random eye and head movement, change of eyebrow at the critical level of the questionnaire provide us the possibility of detecting lies
- Finally, experiments are conducted in a controlled environment to validate our psychological findings.

Since Bangladesh is a developing country, we aim to build a system based on computer vision tools that can track the gaze and body movement of the suspect and capture other parameters at a very low cost. Here in Bangladesh, when questions are uncertain, the students change their minds in the examination hall and commit cheating at that time- which is hard to track. Moreover, detection of lies in the invigilation room becomes tough when the suspect is smart and tries to bypass his guilt. Computer Vision based autonomous system are smart tools to solve this kind of real world problems in an unbiased manner. Moreover, it is a system that reduces the overall cost, stores the evidence and provides better accuracy.

1.4 Possible Implications

In our thesis work we are proposing an artificial and robust system that helps to track deceitful activities utilizing different computer vision tools. Our system is suitable for smart class room as well as real time interrogation system.

A smart classroom is technology-enhanced classrooms that foster opportunities for teaching and learning by integrating learning technology, such as computers, specialized software, audience response technology, assistive listening devices, networking, and audio/visual capabilities. Online remote real-time invigilation essentially recreates the exam hall experience online. Candidates sit the test online at a place of their choosing, such as the home or workplace, using their hardware or they may appear at a smart classroom. The exam is invigilated by a proctor in real-time. An automatic invigilation system will meet up all the lacks in our existing smart classroom scenario to take online examinations.

On the other hand, interrogation (also called questioning) is interviewing as commonly employed by law enforcement officers, military personnel, intelligence agencies, organized crime syndicates, and terrorist organizations to elicit useful information, particularly information related to a suspected crime. Interrogation may involve a diverse array of techniques, ranging from developing a rapport with the subject to torture. A vision based interrogation is the most suitable approach to ensure a torture free environment.

In this work, our goal is to detect suspicious behavior in a smart classroom to assist real-time invigilation as well as in real-time interrogation to detect deceitful manners and lies. In a real-time scenario, tracking the transient attention as well as the sustained ones is very tough here. The cost of delivery and processing as well as the cost of buying the camera/sensors should be affordable. But with a low-cost camera, it is very hard to work at night and on cloudy days because of the lack of proper lighting. Taking these limitations into consideration, our primary focus of this thesis work is to develop a system that can successfully track suspicious activities in social environments and control them by developing an external signaling or alarming system to alert if something fishy is detected. With the successful implementation of the proposed system, we shall march one step forward to acquire digital Bangladesh.

1.5 Research Contribution

The key objective of this research is to develop a computational framework for detecting suspicious activates of humans. The key contributions illustrate in the following:

- Developed a vision-based framework for detecting and understanding suspicious behavior of humans in social settings and finding the key features to relate it with suspicious activities.
- Developed an observational experiment paradigm for studying different external symptoms provided by the suspects while real-time interrogation.
- Developed a psychological experimental system for designing the knowledge base to control the real-time invigilation system.
- Proposed an emotion detection technique that can detect four basic emotions (i.e., happy, fear, surprise, and neutral) and investigated its integration in identifying suspicious activities.
- Investigated the several cues such as head and eye direction to estimate the visual focus of attention (VFOA) for predicting suspicious behaviors (such as copying in the examination halls, telling lies while interrogation etc.)

- Performed several experiments in real-time/ controlled settings to evaluate the effectiveness of the proposed model with appropriate evaluation measures.

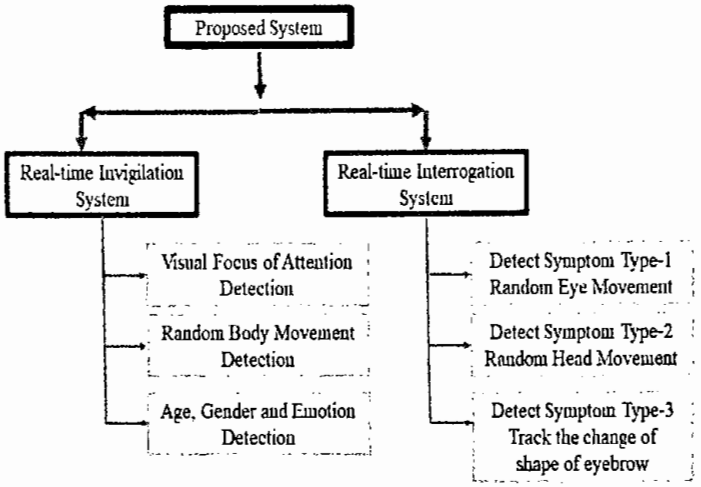


Figure 1.3: Overview of the proposed work.

1.6 Organization of Thesis

The rest of this book is organized as follows:

Chapter 2 (Literature Review): This chapter provides the interdisciplinary background and current state-of-the-art related to the different subjects treated in this dissertation.

Chapter 3 (Methodology): Provides a detailed description of the technical processes. In this chapter, we have described how the head and the iris center are extracted from the captured image and video. Then different gaze detection techniques combining with head pose and body movement have also been mentioned and clearly described how it is implemented in the field of Artificial Intelligence. The sustained and transient focus of attention techniques has also been included.

Chapter 4 (System Implementation): Describes the experimental environment and system setup in detail. It describes the hardware and software setup as well as the system model, parameter list, and programming flowchart.

Chapter 5 (Experimental Results): Provides the experimental results with corresponding scenarios and discussion. This chapter describes how data are collected from the video sequence and compares the performance of different detection methods.

Chapter 6 (Conclusion): Narrates the concluding remarks and the experience while doing the thesis. Some future works have also been included in this chapter

1.7 Summary

So in the conclusion, we may say that, when the crime occurs in a social space, there are always some clues or patterns left. So, it is necessary to track these patterns and build a robust system based on them. It will be a novel work in the field of Artificial Intelligence.

Chapter 2
Literature Review

This chapter outlines the studies concerning Human Behavior Detection and its controlling system. This chapter also gives a brief overview of Suspicious Activities Detection in Social Settings.

2.1 Important Terminologies

Some of the important terminologies related to suspicious behavior detection are discussed below-

- Visual Focus of Attention

  Visual attention can be directed to selected locations in the visual field, even covertly (i.e., in the absence of eye movements). Behavioral measurements have shown that stimuli at attended locations are processed faster and more accurate compared to stimuli at unattended locations. Modern imaging studies of attention have measured brain activity during spatial attention tasks and have documented an enhanced activity in retinotopic cortical areas representing the location of the focus of attention.

- Eye or Gaze Direction

  Eye gaze direction refers **to the direction in which the eyes look.** Two typical types of gaze direction are the direct gaze, which is conceptualized as a model looking at viewers directly, and the averted gaze, which refers to a model who does not look at viewers directly, but looks in a different direction

- Classes of Emotion

  The classification of emotion varies according to the researchers, the general basic emotion found in most research studies includes happy, sad, anger, fear, disgust, surprise, where these emotions were based on a two-dimensional plane commonly called the valence-arousal plane. Positive emotions can be categorized as happiness or surprise, while negative emotions can be associated with sadness, anger, fear, and disgust. Emotional stress may be influenced by negative experiences over a long and continuous period. Emotion classification can be divided into two classes, primary emotion such as joy, sadness, anger, fear disgust, and surprise, and secondary emotion, which evokes a mental image that correlates to memory or primary emotion.

- Controlled Settings

  Controlled settings are artificially constructed environments that are developed for the sole purpose of conducting research. Laboratories, experimental centers, and research units are highly controlled settings often used for the conduct of experimental research.

- Attention Awareness

  Attention can be defined as the convolution of sensory processing with long- and short-term memory. Awareness, on the other hand, depends critically on recurrent processing. By combining the two, we understand why we seem only capable of conscious reports about what is in the focus of attention.

2.2 Human Behavior Detection: Psychological Approaches

There always exists some clues specific to emotion and cognition. Human intelligence is the key to track, understand, and prevent suspicious activities in social environments and therefore it is essential to know when the information obtained is false to identify the key features of deception. It is found that most people, unaided by technology, cannot detect lies from behavioral information, but that some groups do show significantly higher levels of accuracy—although more research is needed to understand why.

The correct judgment of abnormal human behavior analysis can serve as an effective warning signal before emergencies and hence minimize the damage caused by unexpected events. Many scholars have made outstanding contributions to the analysis of abnormal human behaviors in complex scenes. Lan et al. [17] proposed a multi-feature sparse representation framework for visual tracking, in which multiple sparse models were decomposed to distinguish different features. Most CNN-based trackers use hierarchical features to represent targets. However, the layered feature is not completely effective in separating the target from the background, especially in complex interference environments such as heavy occlusion and illumination variation. To solve the above problems, Qi et al. [18] proposed a CNN-based tracking algorithm that hedges deep features from different CNN layers to better distinguish target objects and background clutters. Many experimental results show that the target-tracking method is effective. For autonomous driving applications, traditional tracking methods usually distinguish between a target and a background by training a support vector machine (SVM). When the training sample is nonlinear, the tracker will not be able to track the target. To address this problem, Zhang et al. [19] proposed a distance-based distance learning method based on deep learning. Experiments have shown that even without model updates, the proposed method achieves favorable performance on challenging images. Zhang et al. [20] proposed a research

scheme using density-based clustering and a backtracking strategy to detect small dim targets. The algorithm extracted the trace with the highest clutter suppression ratio, increasing the accuracy of target detection. Rotation invariance was introduced into aerial images of urban scenes by ElMikaty and Stathaki [21]. A feature detection method that analyzes the principle of wavelet transform applied for signal filtering was proposed to obtain high detection performance for small targets of complex multimedia images [22]. Asif et al. [23] presented an object segmentation approach for indoor scenes using a perceptual grouping algorithm.

When it comes to establishing a multistage model, the Grey relational analysis method and an objective optimization model based on maximum entropy are utilized to make the decision-making more comprehensive [24]. Tammi et al. [25] proposed that using the clustering techniques with the amalgamation of the neural network had higher accuracy in detecting the attacks. To address the image forgery problem, an algorithm [26] was designed to classify the image blocks based on a feature present in multi-compressed JPEG images. Huang et al. [27] presented a bottom-up-based framework for salient object detection without any prior knowledge. The approach was more effective in highlighting the salient object and robust to background noise.

In terms of human behavior analysis, Iosifidis et al. [28] demonstrated a novel nonlinear subspace learning technique for class-specific data representation, which was obtained by applying nonlinear class-specific data projection to a discriminant feature space. Under the guidance of kinematics, Xiao et al. [29] proposed a data-driven approach to identify typical head motion patterns and stressed analyzing human behavioral characteristics via signal processing methods. Human dynamics provides new ways of researching human behavior by exploiting the statistical analysis method. The interval time series and the number series of individuals' operating behavior were investigated by a modified multiscale entropy algorithm, which provided insights for further understanding of individual behavior at different time scales [30].

Concerning abnormal human behavior detection, pedestrian behavior modeling and analysis in video surveillance is important for crowd scene understanding. Yi et al. [31] proved the reliability in judging normal and abnormal pedestrian behavior by simulating pedestrian behavior and established two large pedestrian walking route datasets for future research. Hu et al. [32] proposed to detect abnormal driving, which may incur tragic consequences for individuals and the public, by analyzing normalized driving behavior. The driver's anomaly

detection is very important in the field of automatic driving, which can prompt the driver to inform the danger. The detection of the driver is extremely difficult due to camera shake, a sharp change in the speed of the vehicle, and the like. In response to the above problems, Yuan et al. [33] proposed a spatial local constrained sparse coding method for anomaly detection of traffic scenes. From the experimental results, the proposed method is more efficient than the popular competitors and produces higher performance. By employing a 3D discrete cosine transform of the target in different frames, Yuan et al. [34] designed a multi-object tracker, which makes tracking analysis possible in high-density crowds. The experimental results of several public crowd video datasets verify the effectiveness of the proposed method. A taxonomy of novel classes of neighbor-based trajectory outlier definitions was designed to detect abnormal moving objects. The experimental result showed that the method worked well for high-volume trajectory streams in near real-time [35].

To sum up, scholars have proposed many efficient solutions in the analysis of abnormal human behavior. However, there are still some problems waiting to be resolved.

2.3 Human Behavior Detection: Computer Vision Approaches

In this study, we concentrate on related kinds of literature in the field of Artificial Intelligence that target salience detection and segmentation, object detection, visual attention detection, and body movement detection. Saliency detection approaches can be divided into biologically simulated techniques and theoretical model-oriented methods to monitor bottom-up, low-level visual saliency. The first categories of works [36] are generally based on the model proposed by Koch et al. [37]. The low-level information of an image such as color, the orientation of edges, or direction of motion is used in their approach. Itti et al.'s proposed method [36] uses a Gaussian model difference to evaluate the above-mentioned low-level features. The calculated saliency maps, however, are generally blurry in their approach, and often overemphasize small, purely local features.

In comparison, computational models may also be driven in biologically simulated methods, however, they are more closely related to typical computer vision and computer graphics applications. For example, the approaches suggested by authors in [38] measure a salience map based on the Fourier transform of an image's phase spectrum or amplitude. As a result, frequency-based technique-estimated salience maps maintain the high-level structure of an image [39]. Yet their approaches are suffering from undesirable blurriness and appear to estimate boundaries of objects rather than their entire area. Most of the approaches for saliency

detection methods [36] use only low-level feature values. They estimate the salient values in their strategies employing a center-surround technique. For example, if an image's pixel value is more significantly different from its surrounding pixels then a higher salient value is assigned to the center pixel. The frequency-domain feature for saliency detection has been analyzed by Hou et al. [38]. They find the frequencies with uncommon magnitude change in their process and select pixels with high response to these frequencies as salient regions. Now we're going to discuss segmenting or detecting object areas in an image to recognize the visual saliency map which attracts the attention of human beings. Some study uses pixel intensity-dependent threshold values in these areas to segment salient object-related regions [38]. On the other hand, some work uses a conditional random field model to segment the salient area, taking visual salience as one of the characteristics. Learned values of parameters have been used in their approach to detect a signal salient object in an image. However, in our approach, we use salient object regions based on automated adaptive threshold selection for segmentation purposes.

For most human-machine interactions visual attention [40] is important. A brief moment of loss of visual attention in safety-critical activities can cause serious consequences, such as accidents at travel time [41]. In immersive experiments with human-machine, eye gaze tracking is a very efficient method for assessing the visual attention of users quantitatively.

Visual attention assessment is currently available on several eye-tracking systems [42]. With near-infrared reference light, they use the red-eye effect to measure the precise target of the eye gaze. Typically, non-wearable eye trackers like "Tobii" and "SeeingMachine" are put before users without blocking their views. They are characterized as non-intrusive to participants, low sensitivity to head movement, and highly accurate (as high as 0.5L). More importantly, they output the coordinates concerning the world coordination system. These characteristics are highly suitable for tasks carried out in a sitting posture in a virtual environment. Head-mounted eye trackers, e.g., EyeLink and iView X, require the users to wear them on the head. Since the camera is fixed on the head, it captures the head direction-dependent scenes and then monitors the position of the scenes-dependent eye gaze. For many applications, head-mounted eye trackers are flexible in terms of field of view. Their performance, however, is linked to the moving scenes. With the world coordination scheme, they can't provide the coordinates. Hence, they are more suited for manual analysis post-experiment rather than automatic processing in real-time.

The non-wearable eye trackers which provide the correct coordinates for automatic processing typically have a restricted eye gaze tracking range. For example, the tracking range of Tobbi 950 is ±35 in the horizontal direction and 40L in the vertical direction. Such a small tracking range is probably enough for the application with a small field of view (FOV). It may not be wide enough for applications with wide FOVs. A complex system of multiple cameras has to be implemented to satisfy the applications with large FOVs. For example, the Smart Eye Pro 5.0 system up to six cameras to adapt to natural head motion (translation and/or rotation). With the growing demand for high fidelity, more and more virtual environment facilities are built to have wide FOVs.

It is difficult for the non-wearable eye trackers to track eye movements with large FOVs in the applications. The pupils may no longer be visible in the view of the eye-trackers due to the large head rotation in the big FOV scenarios. In such situations, the single camera-based eye trackers are unable to continuously track eye gaze. It is possible to use multiple eye trackers to solve this problem but it could be costly for many applications.

Literature is abundant about these two subjects separately; recent surveys on the position of the eye center and estimate of the head pose can be found in [43] and [44]. The eye position algorithms used in commercially available eye trackers share the sensitivity issue to head pose variations and allow the user either to be fitted with a head-mounted system or to use a high-resolution camera paired with a chin rest to restrict the permissible head movement. Furthermore, daylight applications are precluded due to the use of active infrared (IR) illumination to obtain accurate eye location through corneal reflection. The appearance-based approaches used by standard low-resolution cameras are considered less invasive and therefore more suitable in a wide variety of applications. Within the appearance-based methods for eye location proposed in the literature [45], [46], [47], [48] reported results to support the evidence that accurate appearance-based eye center localization is becoming feasible and that it could be used as an enabling technology for a various set of applications. Edge detection [11] plays a significant role in the processing of digital images and the functional aspects of our lives. In this paper, they examined multiple edge detection techniques such as operators Prewitt, Robert, Sobel, Marr Hildrenith, and Canny. Comparing them we can see that the canny edge detector performs better than all other edge detectors on different aspects as it is adaptive, performs better for noisy images, gives sharp edges, low probability of detecting false edges, etc.

2.4 Suspicious Activities Detection in Social Settings

Video surveillance systems and video processing are of important and critical importance in research in the current situation. The research proposed focuses on putting intelligence into an effective and efficient framework to prevent human intervention in identifying security threats [12] and to detect Suspicious Activities Detection in Social environments.

According to research conducted for almost three decades, face recognition has achieved better results [49], [50], and [51]. This study was conducted using facial images of an age range of 8-60 years consisting of both forms of gender and the age classification was done according to predefined age range [52].

Researchers have been involved in developing automatic emotion classifiers [53], [54], [55] in the last few years. Some expression recognition systems identify emotions such as happiness, sadness, frustration and other systems can recognize the movement of muscles on the face of the individuals. The Facial Action Coding System (FACS) [56] is one of the best psychological frameworks that describe the muscle movements the face can produce CNN's earlier work on Emotion 2015 shows better results in Yu and Zhang [57]. They used a group of Convolution Neural Networks (CNN) in their paper which had five Convolution layers. They achieved excellent performance in recognition and proposed data interference and voting methods which further improved the performance of recognition.

There have been a lot of psychological experiments and studies regarding the students' behavior in the examination hall [58], [59], [60] which ensures the necessity of a real-time cheating detection system. Some of the existing researches on real-time exam hall monitoring (For example, [61-62]) depend on feature detection, eye tracking, or head tracking process to detect suspicious activities of students in the examination hall. Some of them have also used one wear-cam and a microphone to monitor the visual and acoustic environment of the testing location [63] with the accuracy of 89%. Canny edge detection with Speeded Up Robust Feature (SURF) has been utilized in suspicious behavior detection with 80% accuracy [64]. Viola-Jones object detection and HAAR Feature Extraction resulted in 70% accuracy [65]. Some of the authors have utilized Artificial Neural Networks for body movement detection (90% accuracy) and face detection (100% accuracy) [66]. Convolutional Neural Network (CNN) for suspicious motion tracking with the accuracy of 90% [67].

To detect lies while real-time interrogation, some previous research on deceit detection systems used Polygraph [68], [69], [70]. It was developed in 1921 which measures and records several physiological indices such as blood pressure, pulse, respiration, and skin conductivity and tries to find a correlation between these measurements. However, it is highly invasive, very slow (requires several experts), and cannot be used in covert operations (where the suspect is unaware of the experiments.) At present, there are several techniques in vogue [71-73]. Facial Micro-Expressions Approach [71] resulted in 85% accuracy, Electroencephalogram (EEG) signals analysis by F-Score and Extreme Learning Machine resulted in 90% accuracy where Fuzzy reasoning approach resulted in 89.5% accuracy. But most of them use body sensors or high-cost thermal cameras. Moreover they cannot produce an instant output. So a cost effect system suitable for real life application with higher accuracy is highly needed.

2.5 Summary

So, we may conclude that there are a lot of works for eye gaze, head orientation detection along with emotion (including lie) tracking from the images captured from videos in different social environments. But a combination of these two (eye and head gaze) to determine the visual focus of attention of the examinee from images captured by the low-cost camera is a challenging and new concept. Besides that, edge detection of the target object and its application to detect the movement (slow, fast, and very fast) of the examinee along with the age, gender, and emotion detection to detect the fraud examinee will be novel work in detecting the suspicious activities of the students in the examination hall.

Chapter 3

Methodology

In this chapter, the proposed system entitled "cheating detection in a smart classroom environment and lie detection in real-time interrogation" has been described. The various components of the proposed system such as gaze detection, body movement detection, age, gender, and emotion detection in the field of Artificial Intelligence have also been described in detail.

3.1 Overview of the Proposed System

At first, we started creating a knowledgebase by conducting psychological experiments on the students in the examination hall. For suspicious behavior tracking, we need a camera or sensor to detect the visual focus of attention of the examinee. Then we shall store the captured data frame by frame in an intelligent system. The system will convert the image into grayscale because it provides easier processing and analyzes the data pixel by pixel. After then, these analyzed data will be used to detect the loss of attention if the examinee changes his/her visual focus of attention to another direction. How head direction changes at the time of loss of attention for different tasks, will be observed from the captured videos. The minimum deviation of the head orientation can be used as a clue to detect the loss of attention because when loss of attention occurs, examinees change the tilt/pan angle of their head based on the head location and its area from a 3-D head tracker. Then, we can roughly estimate the eye regions of the face using the Active Shape Model. After that, the vector field of the image gradient (VFIG) method is used to detect the iris center. Now the head position using the active shape model (ASM) and the localized iris center are used together for VFOA estimation. For detecting examinees' attention, we need to detect the examinees' body movement. For this purpose, a Canny edge detector is used to detect the edge of the target person. It can detect either he/she moves to the left or right. So, it detects examinees' body movement with the help of captured data that was previously received. Using IMDB-WIKI and UTKFace datasets, our intelligent system can detect the age, gender, and emotion of the examinees, which helps us to identify the fraud examinee. In this way, if inattention is detected, it is controlled by providing an alarm or external signaling system so that the examinee is alerted to pay proper attention in his/her answer script only. The overall process has been illustrated in Figure 3.1.

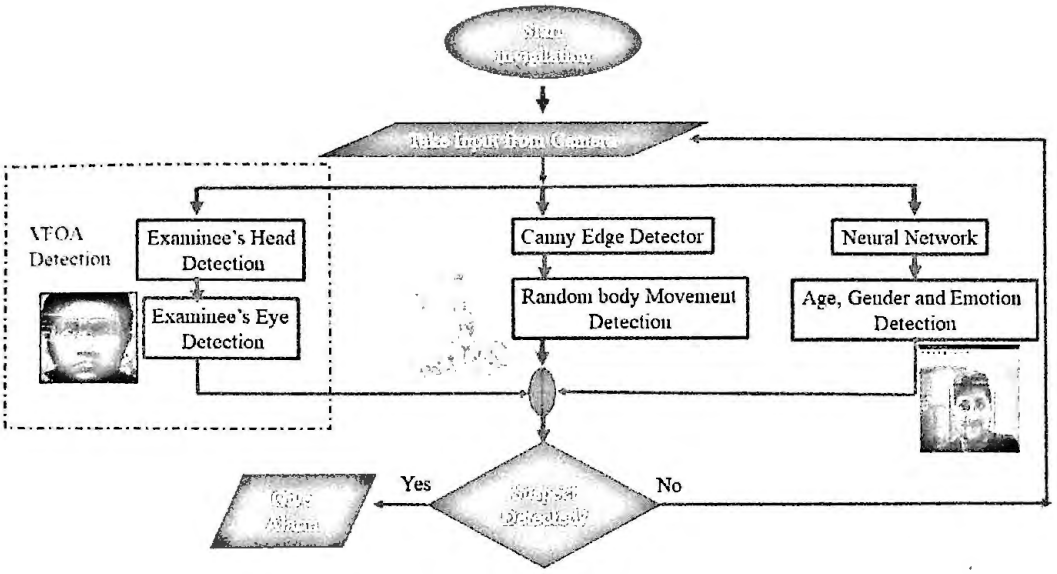


Figure 3.1: Overview of the proposed system for cheating detection in smart class room.

For lie detection at interrogation system, at first, we performed a different psychological experiment and classified different types of crime and other parameters of tracking. After that VFOA detection of the suspect has been used to identify lies while interrogation along with the other useful features which are depicted by the following figure-

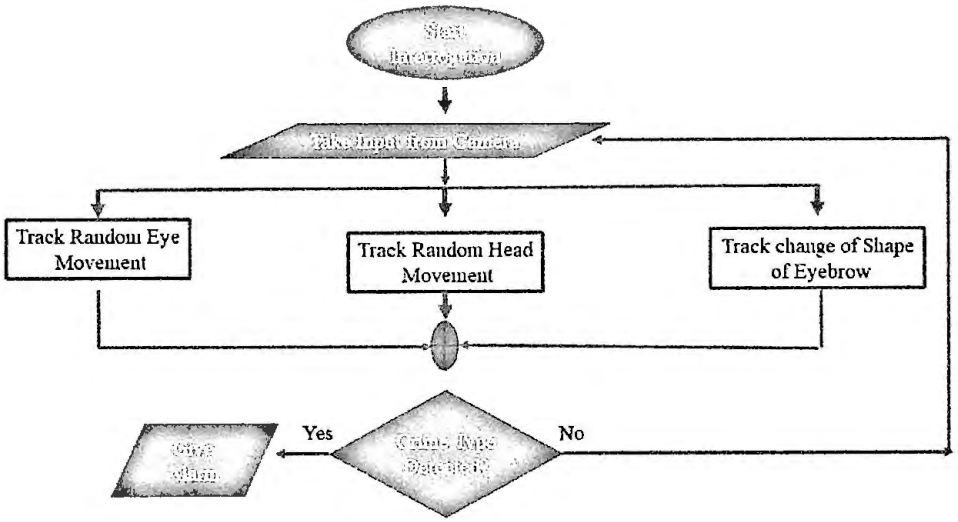


Figure 3.2: Overview of the proposed system for lie detection in real-time interrogation.

3.2 Image Capturing, Processing, and Storing

In this thesis work, we have used a 5-megapixel low-quality camera to capture continuous images. A low-quality camera will store the image at a low resolution. In this way, we can save a lot of memory spaces. The processing time of each of the frames is also reduced. The continuous images are then transferred to the system for processing. Then the images are converted into grayscale because it can process images easily. Histogram equalization is performed on each image so that we can continue our operation when there is a slight lack of proper lighting. 6 to 8 images are processed each second. After that, the frames are resized. The processed frames are saved as a continuous video file in AVI format at the rate of 7 fps (Frame per second). These stored video files are analyzed to get the results.

3.3 Knowledgebase Development

It is based on the psychological study of the examinee regarding the suspicious activities that happened in the examination hall. The main goal is the manual creation of a knowledge base connected with the system's server to detect the threshold value of suspicious behavior of the examinee that will automatically turn on the camera at the right moment. So, we collected video data from different conditions. Based on our study there are three parameters related to suspicious behavior-Exam hall, Type of Exam, and Time of Exam. We also classify the cheating activities in two major criteria-

Cheating Type-1: In this case, the victim copies from the sources such as books, summary papers, mobile phones, written answers on hands, etc. those exist hidden with him.

Cheating Type-2: The sources do not exist with the victim. He looks to and fro and tries to copy from his neighbors. Sometimes he also makes verbal communication.

Now based on the cheating tendency, we designate each of the exam halls with a weight, w_1 ranging from 0 to 5. Based on our study, each of the examinations is also given a weight, w_2 ranging from 0 to 5. Different time slots of the examination are assigned a weight, w_3 ranging from 0 to 3. These weights are stored in our knowledgebase per different parameters. Now the knowledgebase calculates the Cheating Weight, **W** for each of the parameters based on the following formulae

$$\mathbf{W} = \sum_{i=1}^{3} w_i \qquad (3.1)$$

Now, if **W** > threshold value, our system will automatically turn on the camera at the perfect time. We can also change the threshold value to provide very tight or loose invigilation.

3.4 Suspicious Behavior Detection in Exam Room

It is based on the sustained and transient [74] attention/distraction detection of the target person. After the frames are taken by the camera are fed to our proposed system for real-time processing. Then the system tries to detect the head followed by the face points. After that, the eyeball is detected and the fluctuation of the pupil within the eye area is tracked in an efficient way for the visual focus of attention detection. The random body movement is detected by edge detection. The age, gender, and emotion of the examinee are detected by the means of Neural Network. A multistage frame by frame analysis method is described below to detect the suspicious behavior of the examinee.

3.4.1 Iris Center Detection

A multistage approach [75], [76], [77] has been used to detect the iris. A 3-D head tracker [78] detects the head position and the rectangular area in the image. Then the facial feature points are extracted from the active shape model [79]. These points are used to roughly detect the eye regions from the face. The vector field of image gradient (VFIG) is used to detect the iris center.

3.4.1.1 Head Pose Detection by 3-D Head Tracker

A human head has non-planar geometry. Also, it has curved surfaces. A simple model for representing a human head such a plane would not track the 3D head motion accuracy. 2D planar model is simple but not effective for representing a human head because it cannot represent curved surfaces well. Therefore, 3D head tracking with a planar model is not robust to out-of-plane rotations. On the other hand, a complex actual head model would require a very exact initialization and suffer from computational burden. Therefore, to build a 3D model of a human head approximately, a cylinder or an ellipsoid has been often used [80] [81]. Among them, we adopt an ellipsoidal model for representing the human head.

Since a cylinder model does not represent vertically curved surfaces, compared with a 3D ellipsoid model, the latter is more suitable for representing a human head. A 3D ellipsoid itself is parameterized by the lengths of its major axes. We assume that the width r_x is 1. Thus, we only need to determine the ratios between the width r_x and to the height r_y and the width to the

depth r_z. In our approach, we statistically obtain these ratios from the sample data of human heads to represent curved surfaces of a human head more generally. Let the origin of an object coordinate frame be placed at the center of a human head and the frontal face looks at the positive Z axis of an object coordinate frame. Let $[X_o Y_o Z_o]^T$ be the coordinates of a surface point P_o in the object reference frame. Then, the surface point of a 3D ellipsoid can be easily defined like below-

$$X_o = r_x \sin\alpha \sin\beta \tag{3.2}$$

$$Y_o = r_y \cos\alpha \tag{3.3}$$

$$Z_o = r_z \sin\alpha \cos\beta \tag{3.4}$$

Compare with other ellipsoidal models, we only use the partial regions of a human head with a range of about $60° \leq \alpha \leq 135°$ and $-90° \leq \beta \leq 90°$ to express a human face more precisely and exclude disturbing regions such as hairs and background outliers.

3.4.1.2 Face Points Extraction by Active Shape Model

Our modeling method works by examining the statistics of the coordinates of the labeled points in the head rectangle. To be able to compare equivalent points from different shapes, they must be aligned concerning a set of axes. We achieve the required alignment by scaling, rotating, and translating the shape so that they correspond as closely as possible. In this technique, we aim to minimize a weighted sum of squares of distances between equivalent points on different shapes.

We first consider aligning a pair of shapes in the head rectangle. Let X_i be the vector describing the n points of the i^{th} shape in the set:

$$X_i = (x_{i0} y_{i0}, x_{i1} y_{i1}, \dots, x_{ik} y_{ik}, \dots, x_{i(n-1)} y_{i(n-1)})^T \tag{3.5}$$

Let $M(s, \theta)[x]$ be a rotation by θ and scaling by s. Given two similar shaper x_i and x_i, we can choose s_j, θ_j and translation (t_{xi}, t_{yi}) mapping x_i into $M(s_j, \theta_j)[x_j] + t_j$ to minimize the weighted sum

$$E_j = (x_i - M(s_j, \theta_j)[x_j] + t_j)^T \tag{3.6}$$

Where, $t_j = (t_{xj}, t_{yj}, \dots, t_{xj}, t_{yj})^T$ and W is the diagonal matrix for each point.

The weights can be chosen to give significance to those points which tend to be more stable in the set- the ones that move the least concerning the other points in the shape. We have used a weight matrix defined as follows- let R_{kl} be the distance between the points k and l in the shape and V_{Rkl} be the variance in this distance over the set of shapes. We can choose a weight W_k for the kth point as-

$$W_k = (\sum_{i=0}^{n-1} P)^{-1} \quad (3.7)$$

Where, $P = V_{Rkl}$

If a point tends to move around a great deal concerning the other points in the shape, the sum of the variances will be large and low weight will be given. If a point wants to stay fixed concerning the other, the sum of the variances will be small and a large weight will be given. We match such points in the shape to extract the face points.

3.4.1.3 Iris Center Detection by Vector Field of Image Gradient

The VFIG iris center detection technique is described as follows-

Let I_c be the possible iris center and I_{gi} is the gradient vector in position I_{xi}. If I_{di} is the normalized displacement vector, then it should have some absolute orientation as the gradient I_{gi}. We can determine the optical center I_c^* of the iris (darkest position of the eye) by computing the dot products of I_{di} and I_{gi} and finding the global maximum of the dot product over the eye image:

$$I_c^* = argmax\, I_c\{\frac{1}{N}\sum_{i=1}^{N}(P)\} \quad (3.8)$$

$$Where,\ P = (I_{di}{}^T\, I_{gi})^2 \quad (3.9)$$

$$I_{di} = \frac{I_{xi} - I_c}{(||I_{xi} - I_c||)^2}\ and\ i = 1, 2, \ldots, N$$

The displacement vector I_{di} is scaled to unit length to obtain an equal weight for all pixel positions in the image.

3.4.2 Examinee's Gaze Detection

The field of view (FOV) of the examinee is divided into three regions-

1. **Central Field of View (CFV):**

This FOV exists at the center of the human FOV. This zone is set to a 30° cone-shaped area (75° to 105°).

2. **Near Peripheral Field of View (NPFV):**

It is defined as the 45° fan-shaped area on both sides of CFV zones. At the right side of CFV (30° to 75°), it is defined as the right near peripheral field of view (RNPFV) and at the left-right side of CFV (105° to 150°) it is defined as the left near peripheral field of view (LNPFV).

3. **Far Peripheral Field of View (FPFO):**

This FOV exists on both sides at the edge of the human FOV. The right side of the RNPFV (-35° to 30°) is known as the right far peripheral field of view (RFPFV) and on the left side (-145° to 150°) is known as the left far peripheral field of view (LFPFV).

The examinees sitting in examination hall do not move his/her head if the focus is on CFV, LNPFV, or RNPFV. So, if the focus of attention of the examinees is in these regions, we shall track the pupil to get the gaze detection without considering the head movement. When the examinees' attentions are in the LFPFV and RFPFV that time we must have to consider the head movement for the proper focus of attention detection.

3.4.2.1 Gaze Detection by Coordinate of the Pupil

If we can track the pupil of the eye efficiently, then from the variation of the coordinate of the pupil of the driver we can detect on which direction he is focusing. From the face points, eye rectangle is created around the eye region. For simplicity, we shall consider the coordinate of the only left eye because we know that the pupil of the two eyes moves simultaneously. The width of the eye rectangle is divided into three regions: left, monitor and right. If the examinee changes his focus from the central field of view (CFV) to any of the near peripheral field of view (NPFV), the corresponding x coordinate of the pupil also changes. Within the eye rectangle, we shall detect on which region the x coordinate of the eye falls and thus detect the gaze of the examinee.

The followings are different approaches to define the threshold for Left, Monitor, and Right regions.

Equal Division of Eye Rectangle Width

In this approach, the width of the eye rectangle is equally divided among three regions. For example, if the width of the eye rectangle is 60 pixel and the x coordinate of the left eye rectangle starts from 300 and ends at 360, the threshold of three regions will be defined as follows-

Left = {X: X < 320 and X>300},

Monitor = {X: X <=340 and X> = 320},

Right = {X: X < 360 and X > 340}

But this technique is not very efficient and results in a lot of error. Because there are many points in the eye rectangle not covered by the eye. If we consider these coordinates where the x coordinate of the pupil never goes, will produce a lot of error.

We see from the above figure that most of the eye pixels are covered by the monitor region, left a very little number of pixels for left and right region. So, it can detect the monitor gaze, but not the left and the right.

Partitioning Eye Rectangle Leaving the Uncovered Regions

In this approach the total width of the rectangle is divided into five equal regions: Left uncovered, Left, Monitor, Right, Right uncovered. Let us again consider that the left eye rectangle starts from X = 300. For 60 pixels width of eye rectangle, the thresholds of these regions are defined as follows-

Left uncovered = {X: X>300 and X<312}

Left = {X: X>=312 and X < 324}

Monitor = {X: X >=324 and X <= 348}

Right = {X: X >=336 and X <= 348}

Right uncovered = {X: X > 348 and X < 360}

According to the threshold value, we can identify that the monitor region has been reduced and the left and the right region covers most of the left and right portion of the eye respectively. The "left uncovered" and "right uncovered" regions do not cover any portion of the actual eye. So, while gaze estimation, these two areas are discarder. But in this technique

the slight change in focus from Central Field of View (CFV) to the left or right cannot be detected very efficiently. Because a slight change in focus means a slight change in x coordinate of pupil and hence cannot be detected by the left or right region.

Unequal Partitioning of Eye Rectangle Leaving the Uncovered Regions

In this technique, we leave the uncovered regions but the width of the left, monitor, and right regions are not the same as we discussed before. To detect the slight change of focus from Central Field of View (CFV) to the left or right, the width of the monitor region is decreased and hence the width of the left, as well as the right regions, increases so that they can accommodate the slight change of focus in Near Peripheral Field of View (NPFV) and enhance the total accuracy. The threshold detection for left, monitor, and right regions are discussed below-

Let the width of the eye rectangle be 60 pixels and it is divided into five equal parts as before. A rough threshold for the five regions is defined as follows-

Left uncovered 1 = {X: X>300 and X<312}

Left = {X: X>=312 and X < 324}

Monitor = {X: X >=324 and X <= 336}

Right = {X: X > 336 and X <= 348}

Right uncovered = {X: X > 348 and X < 360}

Now the length of the monitor region is divided into 4 segments. If the length of the Monitor region is 12 pixels, the length of each of the segment, d = 12/4 = 3. Now the new threshold is estimated as-

Left uncovered = {X: X>300 and X<312}

Left = {X: X>=312 and X < (324+d)}

= {X: X>=312 and X<=334}

Monitor = {X: X >= (324+d) and X <= (336- d +1)}

= {X: X>=327 and X<327}

Right = {X: X >=(336-d+1) and X <= (348+d-1)}

= {X: X>=334 and X<= 350}

Right uncovered = {X: X > (348+d-1) and X < 360}

= {X: X>=350 and X<360}

3.4.2.2 Gaze Detection Using Head Rectangle

If the examinee's gaze is in the far peripheral field of view (FPFV) or near this zone, head movement occurs. In this case, we cannot detect the eyes for gaze estimation very accurately. So, depending on the head movement, we decide on which object the examinee is looking at. To accomplish this, the width of the head rectangle is divided into three regions: Front, Left side, and Right side. From the detected face points, the middle of the face is detected and it is the middle of the eye rectangle. Now if we move our head, this middle line of our face also moves concerning the head rectangle. We decide whether the examinee is looking on left, front, or right based on in which head region the middle line of the face falls. For example, let us consider that the width of the head rectangle is 180 pixels. It is divided into five equal segments. The length of each segment, s= 180/5 = 36 pixels. If the x coordinate of the head rectangle starts from 300, the threshold values for the three head regions are defined as follows-

Left side = {X: X > 300 and X < (300 + 2 * s)}

= {X: X > 300 and X < 372}

Front = {X: X >= (300 + 2 * s) and X <= (300 + 3 * s)}

= {X: X > 372 and X < 408}

Right side = {X: X > 408 and X < (300 + 5 * s)}

= {X: X > 408 and X< 480}

3.4.3 Sustained and Transient Focus of Attention Detection

Selective sustained attention, also known as focused attention, is the level of attention that produces consistent results on a task over time. However, the period of sustained attention varies with a different type of tasks. On the other hand, transient attention is a short-term response to a stimulus that temporarily attracts/distracts attention. The minimum time for sustained detection is 8 seconds [80] because the maximum time for transient attention is 8 seconds. Detection of sustained and transients are very important to ensure a cheating-free

examination hall. For a given period during examination, if we can detect that the transient attention of the examinee is occurring frequently, we can ensure the inattention of the examinee and give an alarm to alert the examinee. To ensure a cheating examination hall most of the attention must be sustained. No transients also mean inattention. Because during exam a student needs to look outside automatically which are also transient attention. The threshold values of time for transient and sustained focus of attention are given as follows-

Transient = {T: T > 0 and T < 3}

Sustained = {T: T > 8}

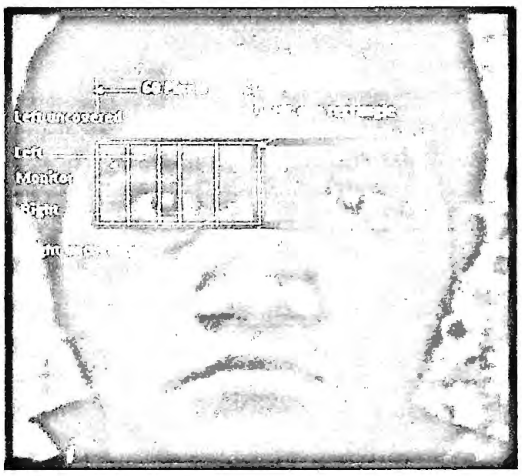


Figure 3.3: Cheating Type-1 detection based on eyeball fluctuation

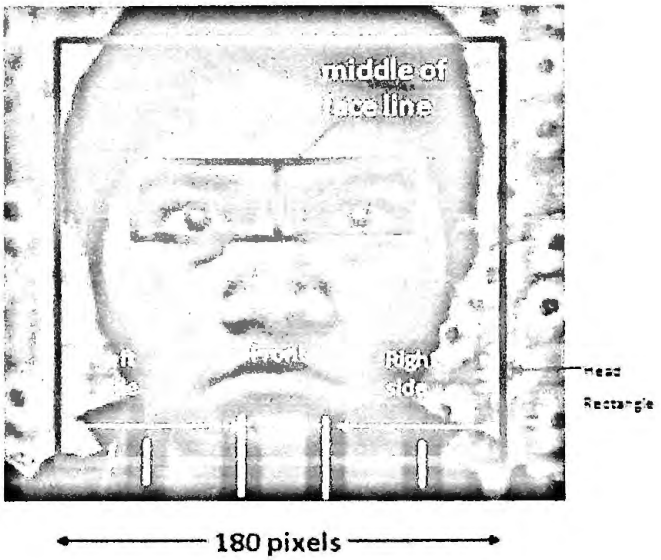


Figure 3.4: Cheating Type-2 detection based on head rotation detection

3.4.4 Examinee's Body Movement Detection

Examinee's random movement of body pixels is detected in the field of the neural network.

3.4.1.1 Edge Detection

An edge may be defined as a set of connected pixels that form a boundary between two disjoint regions. Edge detection is basically, a method of segmenting an image into regions of discontinuity. In our work, we use a canny edge detector because canny edge detector performs better than all other edge detectors on various aspects such as it is adaptive in nature, performs better for noisy image, gives sharp edges, low probability of detecting false edges, etc.

3.4.1.2 Canny edge detection algorithm

STEP-I: Noise reduction by smoothing

Noise contained in an image is smoothed by convolving the input image that is $I\ (i,j)$ with Gaussian filter G. Mathematically, the smooth resultant image is given by

$$F\ (i,j) = G * I\ (i,j) \tag{3.10}$$

Prewitt operators are simpler to an operator as compared to Sobel operator but more sensitive to noise in comparison with Sobel operator.

STEP-II: Finding gradients

In this step we detect the edges where the intensity shift in the grayscale is maximum. Required areas are calculated with the help of image gradient. Sobel operator is used to determine the smoothened image gradient at each pixel. Sobel operators in i and j directions are given as

$$D_i = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \quad \text{And} \quad D_j = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

These sobel masks [13] are convolved with smoothed image and giving gradients in i and j directions

$$G_i = D_i * F\ (i,j) \tag{3.11}$$

$$G_j = D_j * F\ (i,j) \tag{3.12}$$

Therefore, edge strength or magnitude of gradient of a pixel is given by

$$G = \sqrt{G_i^2 + G_j^2} \tag{3.13}$$

The direction of gradient is given by $\theta = arctan\ (\frac{Gj}{Gi})$ where, G_i and G_j are the gradients in the i- and j-directions respectively.

STEP III: Non-maximum suppressions:

Non-maximum suppression is carried out to preserves all local maxima in the gradient image, and deleting everything else results in thin edges. For a pixel M (i, j):

- First-round the gradient direction at nearest45 °, then compare the pixel gradient magnitude in positive and negative gradient directions. If the gradient direction is east then compare E (i, j) and W (i, j) respectively with the pixel gradient in the east and west directions.
- If the edge strength of pixel M (i, j) is largest than that of E (i, j) and W (i, j), then preserve the value of gradient and mark M (i, j) as an edge pixel, if not then suppress or remove.

STEP-IV: Hysteresis thresholding:

The non-maximum suppression output also includes the noise-creating local maxima. Instead of using a single threshold, two thresholds t_{high} and low- are used to avoid the problem of streaking.

For a pixel M (i, j) having gradient magnitude G following conditions exists to detect pixel as an edge:

- ✓ If G $<t_{low}$ then discard the edge.
- ✓ If G > t_{high} then keeps the edge.
- ✓ If t_{low} < G < t_{high} and any of its neighbors in a 3 ×3 region around it have gradient magnitudes greater than t_{high}, keep the edge.
- ✓ If none of the pixel (x, y)'s neighbors have high gradient magnitudes but at least one falls between t_{low} and, t_{high} search the 5 × 5 regions to see if any of these pixels have a magnitude greater than t_{high}. If so, keep the edge.

Else, discard the edge.

3.4.1.3 Body Movement Detection at Different Dynamics

To detect the random body movement (comprises of very high or slow dynamics) of examinee with optimal accuracy, we propose the following techniques-

Approach-1: Low and High Dynamics

At first, a video sequence of f frames is taken by the video camera to our system. Then the system started to compute their corresponding edges using Canny Edge Detector [13]. At the first stage, we detected the moving objects using detecting the moving edges. Let us consider that the moving edges are denoted by $Edge_m$. Now we considered three consecutive frames denoted by $\{I_{n-1}, I_n, I_{n+1}\}$ to perform a mathematical operation.

At first, we computed two different images by considering the central frame after subtracting from its two neighbors. That means, $(I_1 = |I_n - I_{n-1}|$ and $I_2 = |I_n - I_{n+1}|)$. We only considered those pixels that will result in positive value in these different images. Value zero is set for the negative pixels. Now from I_1 and I_2 we can easily get the moving pixels. But we may see, they may not match with each other and lacks finer details of the edges (see Figure 3.3). They are also having some background noises and δ edges which are related to the speed of the movement, see Figure 3.4 (bottom). We performed AND logical operation on these two resulting images to remove the δ edges which is expressed as follows -

$$\text{Final Image, } I_{final} = I_1 \cap I_2$$

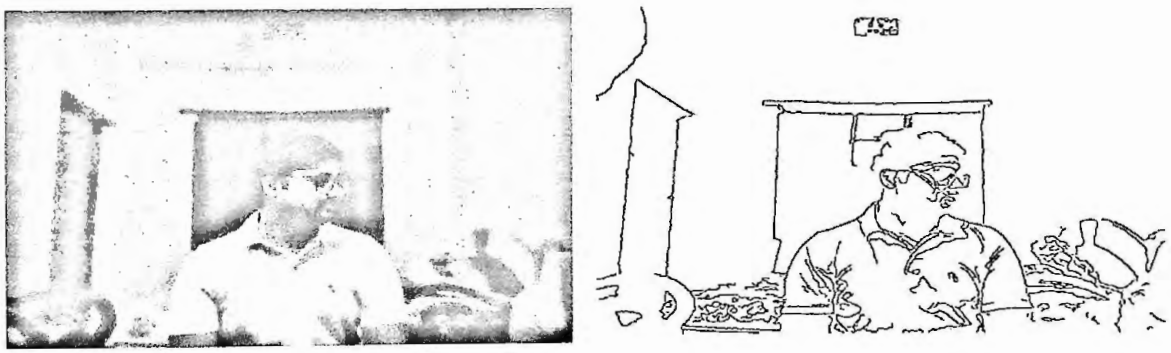


Figure 3.5: (left) Original frame. (right) Edge representation computed by the Canny edge detector.

Figure 3.6: Derivation of difference and final images with low dynamics.

Figure-3.4 (right) shows an illustration of the resulting I_{final}, corresponding to Figure 3.3, computed after merging Figure 3.4 (left) with Figure3.4 (center). During those experiments, the speed was very low. So the two different images I_1 and I_2 don't have any δ edges, but they are corrupted by some noisy edges.

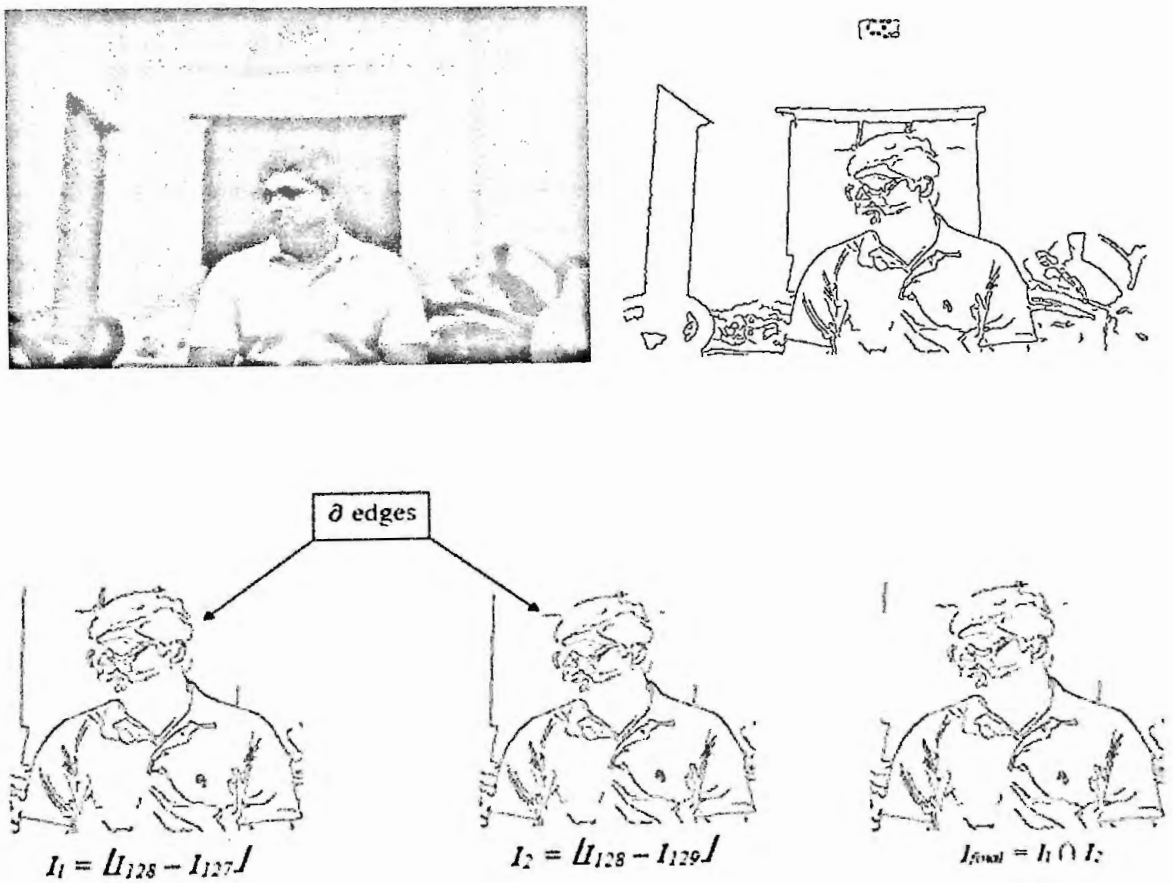


Figure 3.7: Derivation of difference and final images with high dynamics.

Figure 3.5 indicates the result obtained with a scene containing a movement having higher dynamics. From the figure, it is clear to us that now they are having some δ edges. As we discussed earlier, the final image I_{final} is again corrupted by noise.

Approach-2: Very Low Dynamics

Now we present a solution to the problem to detect the body movement with very low dynamics with the help of the Canny edge detector.

Let us consider that we have detected the moving edges for a particular frame I_n (located at nth position) with the process as described in approach-1. Now we set a variable, m=0, and also consider all the images located at n±m locations. Using the technique in approach-1, we computed the edges from all $I_{n\pm m}$ images. Now all of the resulting edge images are merged by OR operation to derive the missing edges caused by low dynamics and derive $I_{ultimate1}$. Now we increased the value of the variable m and perform the same operations as discussed above and derive the image, $I_{ultimate2}$ after OR operation. Next, we performed the OR operation on the above two ultimate images to detect the smallest movements with finer details. The whole operation is continued until no new information about the edge can be extracted from $I_{ultimate}$.

Figure 3.6 shows the Original frame (left), Moving edges extracted after one iteration(middle), and Moving edges extracted after two iterations (right).

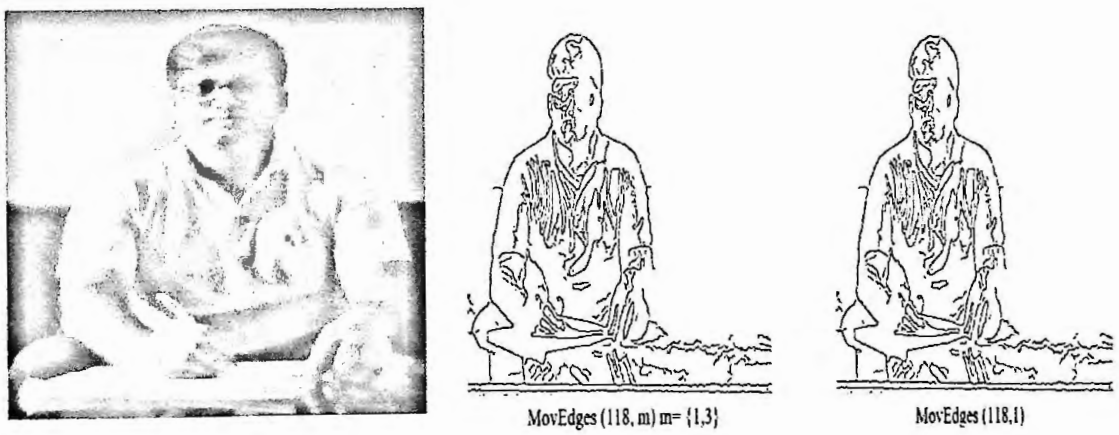


Figure 3.8: Illustration of Approach-2.

Examinee's Body Movement Detection

We combined Approach-1 and Approach-2 to build our system to track the examinee's suspicious body movement. At first, a frame is captured (starting with a timer=0) from the video sequences taken from the webcam of each examinee. We set a threshold value of black pixels remaining at the moving edges to be considered for a suspicious movement. Then we implemented approach-1 to detect the moving edges represented by black pixels in the white screen. We counted the number of black pixels and compared them with the threshold value. If it surpasses the threshold value, then we considered it a suspicious movement. If not, we implemented Approach-2 because the movement may be too slow to be detected by Approach-1. Again we counted the number of black pixels and compared them with the threshold value to detect the suspicious movement. If nothing is detected, we increased the timer and start from the beginning. The following flowchart illustrated in Figure 3.7 describes the above techniques.

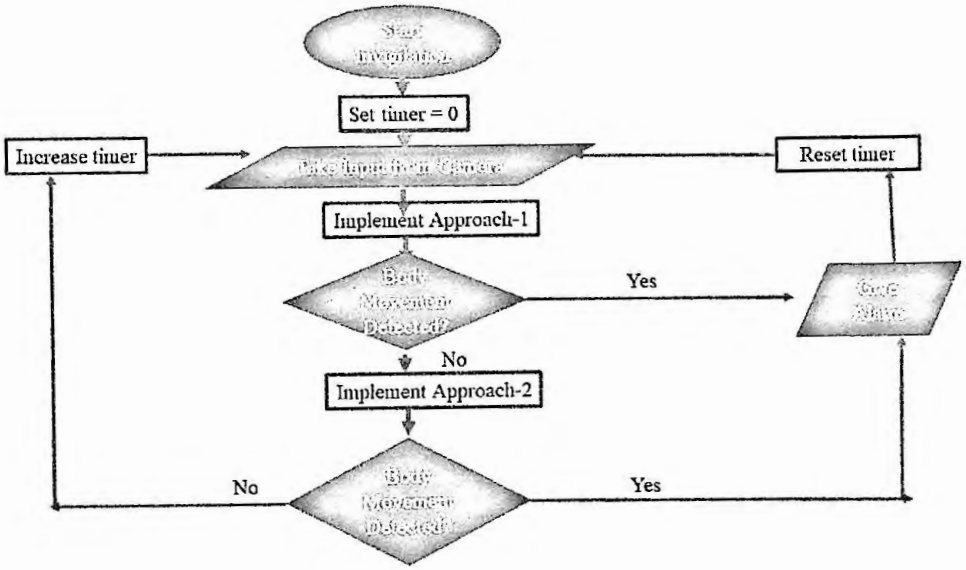


Figure 3.9: Programming flow chart to detect body movement of the examinee.

3.4.5 Age and Gender Detection

Automatic classification of human age/gender finds many real-world applications, such as target advertising, demographics analysis, visual surveillance, etc.

The main objective of the proposed algorithm is to use a specific set of facial features to classify the appropriate age range and gender from the human face images. Following figure shows the

flow of this proposed algorithm which consists of three main steps; Pre-processing, Feature Extraction, and Classification.

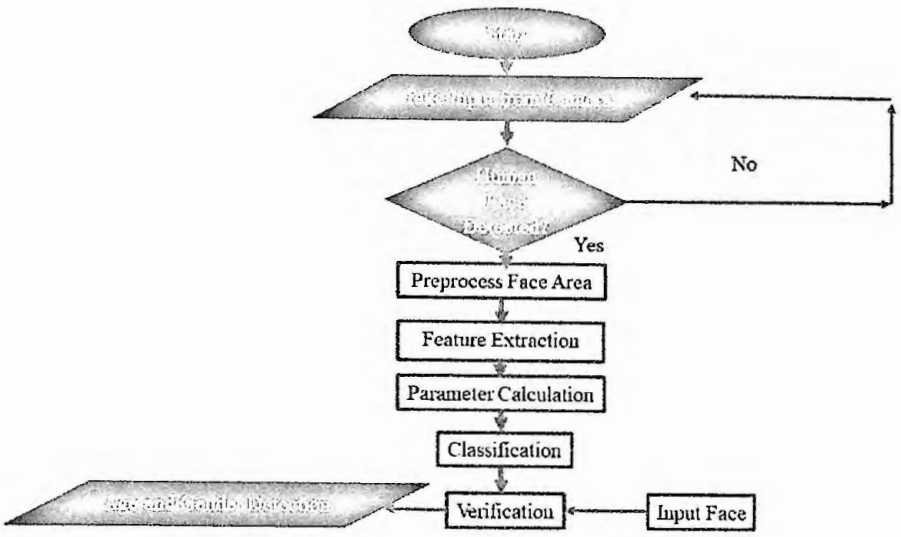


Figure 3.10: Main steps of the proposed Age and Gender Classifier.

A. Input Image

The input image is the source image intended to test with the age and gender classifier. User can input any type of image format like .jpg, .png, .tiff, and .bmp.

B. Face Area Detection

The first phase will continue to check whether or not the given input image contains a facial part. When there is no face area in the input image, the algorithm will reject the input image. Once the face area is detected, the classifier will extract the area and create a separate image per every face in the input image by cropping the background. The classifier was equipped with enough frontal, almost frontal, and rotated faces from 0 to 45 degrees, and non-face images. Detected face images are preprocessed to standardize the face images by converting them to a unique format.

C. Preprocessing

The photos used in the experiment are used in various conditions, such as the presence of noisy data, different lighting conditions, and different levels of intensity. Thus, before forwarding to the classification stage, detected face images must undergo a preprocessing step.

• *Resize Detected Face Image* - The images collected from the initial face detection are in various sizes. Therefore, the first step in preprocessing to standardize the data set is to modify each image into a standard width and height (e.g. 255 * 255 in this research).

• *Colour Conversion* - The images used in this research must be in standard format for color. Therefore, to overcome the complexity, all the images are converted into grayscale and finally do an equalization of the histogram to have a uniform distribution of the image intensity values. First, the red, green, and blue values of every pixel in the image are obtained and then the following formula is used to convert the RGB image into a grayscale image:

$$G\ (x, y) = 0.21R + 0.7G + 0.07B \tag{3.14}$$

• *Noise Reduction* - Dirt on camera lenses, imperfections in-camera flash lighting that result in natural images taken from digital cameras creating noise. The transformed color images are sent to the filter for noise reduction. Gaussian smoothing is used in pictures to eliminate the noise. Gaussian smoothing function f(x, y) can be expressed [13] as,

$$\mathrm{G\ (x,\ y)} = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/(2\sigma^2)} \tag{3.15}$$

This function can be used to calculate the weight for each pixel in the image. Assume the center point of the weight matrix is (0, 0). Then the nearest coordinate values can be represented as,

$(-1,1)$	$(0,1)$	$(1,1)$
$(-1,0)$	$(0,0)$	$(1,0)$
$(-1,-1)$	$(0,-1)$	$(1,-1)$

Then the weight matrix is calculated by setting a value to σ. The output weight matrix for a sample data set by using σ =1.0.

0.0453	0.0556	0.0453
0.0556	0.0707	0.0556
0.0453	0.0556	0.0453

The sum of the weighted matrix is calculated using the formula:

$$sum\ (w) = \sum_{i=1}^{9} \mathrm{w(i)} \tag{3.16}$$

Then the weighted average of the nine points is calculated by.

$$\text{avg(w)} = \frac{w(i)}{sum(w)} \tag{3.17}$$

The Gaussian blur for each point in the matrix is calculated by multiplying the color value of each point by the weighted average. Each color value is between 0-255.

$$G(i) = int\ (i) * avg\ (w) \tag{3.18}$$

These values of the matrix help to calculate the Gaussian blur value for the center point.

$$blur(middle) = \sum_{i=1}^{9} \text{G(i)} \tag{3.19}$$

By repeating the above steps for all the points in the image, the Gaussian blur for the face images can be calculated.

• *Face Identification* - Identifying the face area from the image input plays an important task, as some image input may contain unnecessary data other than the face area. To make this efficient, a classifier is trained using several images which include face images (positive set of face images) as well as non-face images (negative set of face images). Separate features of the objects are extracted during the training process. Those extracted features are later used to classify the objects in the images that are not seen. Hair wavelets are square waves of single wavelengths that contain a high interval and a low interval. In two dimensions, two adjacent rectangles represent a square wave, in which the pair contains one light color and one dark color. The determination process of the presence of a Haar feature is done by subtracting the average dark region pixel value from the average light-region pixel value. During the learning process, a threshold value is provided, and it checks whether or not the difference is greater than the threshold value. If it is higher than the threshold value then the feature is present.

D. Feature Extraction

The human face contains 66 feature points of landmarks according to the research done by using images from the FG-NET database. Relevant features important for the classification should be extracted from the face images. According to human gender variation, there is a considerable difference in geometric size, shape, and distance variations in the facial content. Some of these variations are described in Table 3.1.

Table 3.1: Facial Characteristic Variation between Male and Female.

Facial Feature	Male	Female
Hairline	Usually has higher peeks on the sides and tend to be 'M' shaped	Rounded shape
Eyebrows	Usually thick and just under the orbital rims	Generally, sit higher and more arched
Eyes	Appear small	Appear large
Distance between eye and eyebrow	Lower	Higher
Nose	Relatively bigger and smaller	Small and short
Lips	Distance between the nose and the upper lip is large	Distance between the nose and the upper lip is small
Chin	Wider chin	Rounded chine
Cheeks	Hollow cheek	More rounded
Face shape	Square appearance	Heart shape appearance

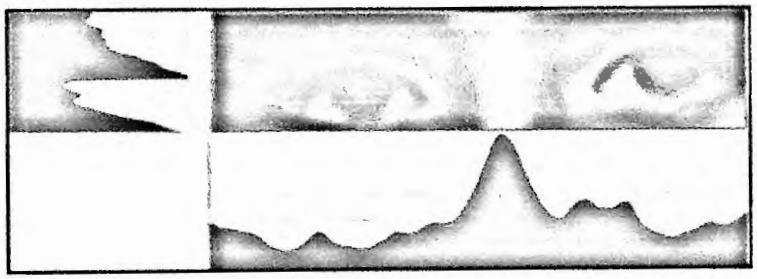


Figure 3.11: Projections used to locate eyes.

A classifier is used to identify areas of the photos in the nose. The classifier has worked with pictures of the nose and of the non-nose. It is smart enough to identify the areas of the nose and represent a square zone, to use the wide enough area to locate the tip of the nose. Feature point locating is done by the horizontal and vertical projection to the cropped big enough area. The crossing point of valleys of the horizontal and vertical projection can be used to locate the nose tip as shown in the figure 3.10.

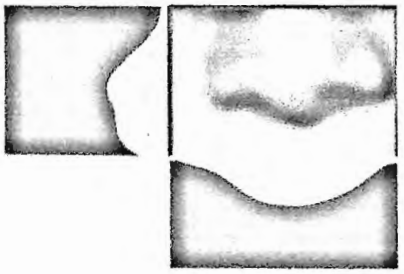


Figure 3.12: Nose tip location.

Another classifier has been used to identify the area of the mouth in the images. The classifier has been trained with images with and without mouth areas. It is intelligent enough to locate mouth areas and represent them as a square region and to locate the endpoints the mouth as well. Feature point locating is again done by the horizontal and vertical projections.

E. Locating Secondary Features - Wrinkles on the face are becoming clearer as people get older. Hence, wrinkles are the best feature defining an adult person's age. The general area of wrinkle is; Forehead, Eye Corners, and Cheek. Use Sobel edge magnitude to identify the measurement of wrinkles in corresponding facial images after location of the wrinkle areas. Wrinkles can definitely be identified from normal skin because the Sobel edge magnitudes of the wrinkles are greater. For the age classification, algorithm used wrinkle feature variation from young age to old age. The wrinkle areas are located manually and the Sobel edge magnitude is calculated for each area.

The description of the parameters is given below:

- ✓ Height of Eye = (X1 + X2) / 2
- ✓ Distance between Eye and Eye brow = (Y1 + Y2) /2
- ✓ Height of Nose (N) = b – (d + f) / 2
- ✓ Distance between Lip and Nose = h – b
- ✓ Width of Eyebrow = Q
- ✓ Distance between Eyes (E) = e – c
- ✓ Eye to Upper Lip Distance (L) = h – (f + d) / 2
- ✓ Ratio1 = Eye Distance /Nose Distance = E / N
- ✓ Ratio2 = Eye Distance / Eye to Upper Lip Distance = E / L

The calculated parameter vector for the gender classification consists of the Eye height (H E), Distance between eye and eye brow (Dist E & EB), Height of the nose (H N), Width of the eye

brow (W EB) and Distance between upper lip and nose top (Dist N & L). Similarly, the calculated parameter vector for the age classification contains two ratios: ratio between eyes to nose distance and eyes to mouth distance and ratio between distance from eyes to mouth and distance between the two eyes. Two wrinkle parameters from each wrinkle area are also calculated:

- ✓ Wrinkle Density ($W_{density}$) = No of all wrinkle pixels/ No of all pixels in the area
- ✓ Wrinkle Depth (W_{depth}) = Total Sobel Magnitude of wrinkle pixels/Total no of pixels in the area.

F. Classification

Classification is carried out in two principal steps. The gender classifier will first identify the corresponding gender of the query image. The image is then moved to the age classifier to identify the respective age group. Gender classification is basically done using variations in the shape of the features on the face and the age classification is based on variations in the texture of the wrinkle areas. Classification is performed using calculated neural network parameters. The neural networks are trained using data taken from nearly 10000 images of the two gender groups and different age groups which are images from frontal and nearly frontal faces. Following figure shows the structure of the neural network used for age classification.

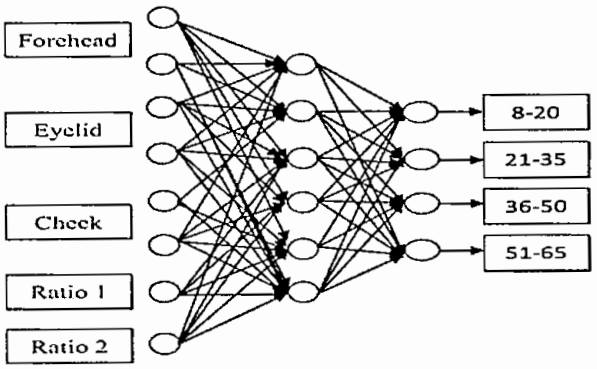


Figure 3.13: Neural network structure used for Age Classification.

Calculated parameters are sent to the corresponding neural network to classify the images according to age and gender. Neural network used for gender classification contains five input nodes and two output nodes 0 and 1 corresponding to the two gender groups, male (0 – 0.5) and female (0.5 – 1.0). Output value from the neural network is used to classify the image into one of the gender groups. The neural network used to classify the given input image in to a corresponding age group has eight input nodes and the output layer contains four nodes

namely 0, 1, 2 and 3 to represent the age groups. The neuron with the highest response is considered as the output of the age classification neural network.

3.4.6 Emotion Detection Approach

Developing a real-time emotion detection system, we had to get different components working independently. We also analyzed several research paper to identify the basic problem and how we could improve our results accuracy. Our general approach summarizes below:

• ***Build Dataset***: We gathered data sets from multiple sources from coded facial expressions and translated their labels and images into a common format.

• ***Pre-process Images***: We've run software for facial detection to extract the face in each image. We then re-scaled the croppings, and removed bad images manually. We also applied a Gaussian filter to the images as a preprocessing step for the CNN, and subtracted the mean image from each image of the training set. We also expanded the images to include reflections and rotations of each image in order to get more out of our limited training data, hoping that this would increase robustness.

• ***Construct CNN:*** In Caffe on AWS, we used pre-trained versions of AlexNet and LeNet, where we retrained the first and last layers. We also had to experiment with different methods and parameters of the learning rate in order to produce a non-divergent model.

• ***Develop real-time Interface***: OpenCV has allowed us to pick up images from the webcam of our laptop. We then extracted the face as before, pre-processed the image for the CNN, and with the help of Visual Geometry Group and Keras model, we got a prediction and the results would be shown in my pc webcam.

• ***Dataset Development:*** The first step in the development of our emotion-detection system was to gather data to train our classifier. We sought to find the largest data set we could have, and with the help of IMDB-WIKI dataset we created the UTKFace dataset. This data set is composed of over 100 individuals portraying 6 different labeled emotions: angry (1), fear (2), happy (3), Neutral (4), sad (5) and surprise (6). One thing we really liked about this data set is that the directory includes 10 to 30 images for each person expressing an emotion, showing the progression of that individual from a neutral expression to the target emotion. Initially we chose to take the first two photos from each series and mark them as neutral, and the last three as the target emotion. Nonetheless, we found this significantly restricted the size of our training package, as we were left with less than 1000 training photos. To combat this, we looked at the

images more closely and decided to take as the target emotion the last third of each sequence, as opposed to just the last three.

Our emotion detection system follows the given model to express human emotion.

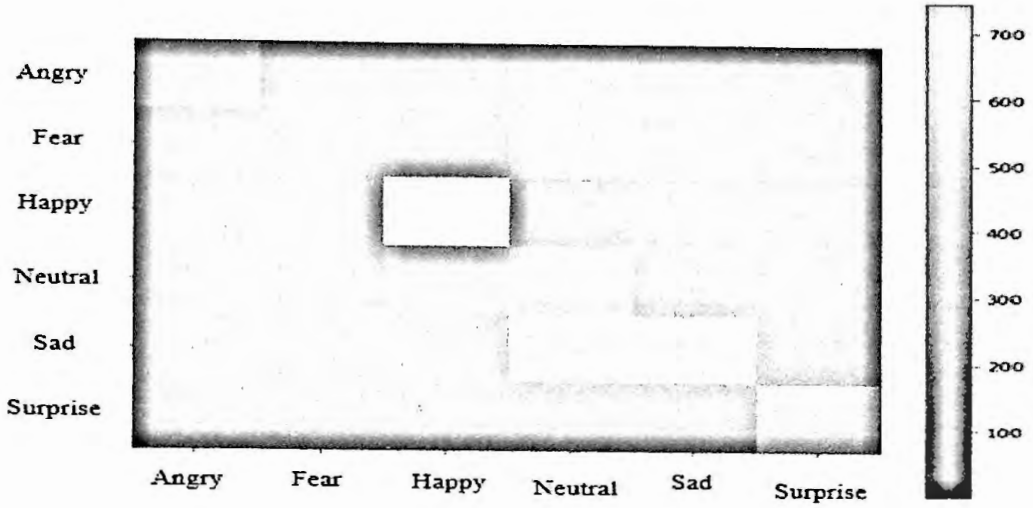


Figure 3.14: Overview of the trained model.

The following data Table 3.2 shows the accuracy of our trained model-

Table 3.2: Confusion Matrix and Classification Report.

Confusion Matrix					
[[216 3 39 130 86 17]					
[119 19 49 173 94 74]					
[30 1 743 61 33 11]					
[95 6 161 203 98 63]					
[60 2 47 240 238 7]					
[28 8 46 48 7 279]]					
	Classification Report				
	Angry	0.39	0.44	0.42	491
	Fear	0.49	0.04	0.07	528
	Happy	0.68	0.85	0.76	879
	Neutral	0.24	0.32	0.27	626
	Sad	0.43	0.40	0.41	594
	Surprise	0.62	0.67	0.64	416
	micro avg	0.48	0.48	0.48	3534
	macro avg	0.48	0.45	0.43	3534
	weighted avg	0.48	0.48	0.45	3534

Our system can detect all 6 classes of emotions and we had good accuracy on Happy, Fear, Neutral, sad, and Surprise but bad accuracy on angry emotion. The following figures show different types of emotions.

Figure 3.15: Illustration of Neutral emotion.

Figure 3.16: Illustration of Fear emotion.

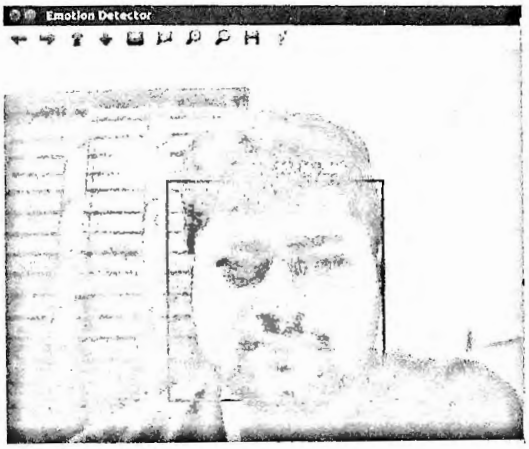


Figure 3.17: Illustration of Sad emotion.

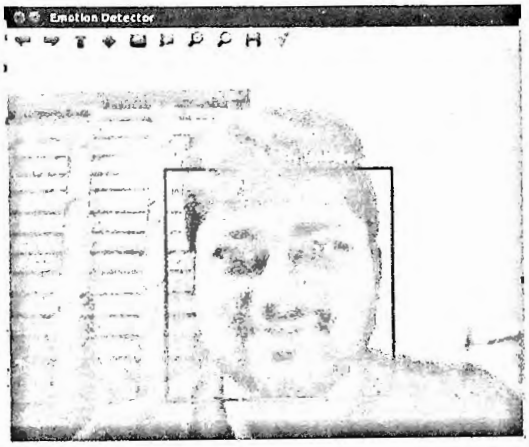


Figure 3.18: Illustration of Happy emotion.

Figure 3.19: Illustration of Surprise emotion.

3.5 Lie Detection at Real-time Interrogation System

Human characteristics are very peculiar in nature. It is very tough to detect the insights of a human, especially when he is telling lies. In the case of an interrogation environment, the problem becomes tougher because the suspect keeps himself ready for some common question. In our approach, we classify the answers of the victim in four major criteria as follows

Criteria (1, 1): The suspect has done it and also admits it. It would be considered as "**True.**" For example, if the suspect has stolen and tell that "Yes, I have done it."

Criteria (1, 0): The suspect has done it but denying it. It would be considered as "**False.**" For example, if the suspect has stolen and told that "No, I have not done it."

Criteria (0, 1): The suspect has not done it but pretending that he has done it. It would be considered as "**False.**" For example, if the suspect does not have a redshirt but if he says "Yes, I have."

Criteria (0, 0): The suspect has not done it but is denying it. It would be considered as "**True.**" For example, if the suspect does not have a redshirt and if he says "No, I don't have."

In our approach Criteria (1, 1) and Criteria (0, 0) are considered TRUE and hence, they are to be bypassed by the system. But Criteria (1, 0) and Criteria (0, 1) are considered as FALSE and must be tracked by the system.

Now when the interrogation begins, even the weakest liar can overcome our test without showing any external symptoms because at first s/he is ready to fight with some of the ready answers. However, interrogation means a chain of questions and to support one lie the suspect must tell another lie. Over time it becomes tougher for the suspect to answer with his/her ready wit. So, weak liars cannot sustain with the "Chain of Lie." and shows some external symptoms. And the symptoms are very obvious for the criteria (0, 1). The reason behind this fact is that, if anyone claims to do something that s/he has not done, he needs to imagine that situation at first before answering about it. While imagination there is a change in body language showing symptoms of a possible lie. Our main goal is to track those symptoms to detect a lie. To this point, we have classified the symptoms in three major criteria-

Symptom Type-1: The suspect provides a very random movement of his eye which is a clear indication of his nervousness. S/he also moves his eyeball to the upper right corner of his eye area as an indication that s/he is thinking of something.

Symptom Type-2: The suspect provides a head movement. In most cases for the weak liar, it is tough for him/her to continue the conversation with direct eye contact with the interrogator. Too much loss of direct eye contact is a clear indication of telling a lie. To avoid direct eye contact, the suspect moves his/her head downward. Sometimes s/he also makes a frequent head movement.

Symptom Type-3: The suspect changes the shape of his eyebrow as an indication of his worry. Now if we can detect Symptom Type-1, 2, or 3, we can say that there is a possibility that the suspect is telling a lie. To confirm the possibility, we need to compare it with our knowledgebase. The knowledgebase is the storehouse of symptoms provided by different suspects at different levels of questionnaire obtained from our manual interrogation. Each of the questions in the questionnaire is marked as Phase 1, 2,... ,N, etc. The symptoms offered by the suspects are also age variant. The teenagers are generally impatient and cannot remain calm and hence provide useful and clear as well as useless symptoms while interrogations. But the aged groups provide very low levels of symptoms which are very hard to track. So to build our knowledgebase to store how people react while telling a lie, we need to consider our samples from different age groups. The statistics from the samples will help us to get the threshold value to confirm a lie. Since human emotion is the prime parameter of our lie detection strategy; we classify our crime pattern as follows

Crime Type-1: Emotion is connected with crime. In this situation, there is no economic profit. Any type of revenge or avenge (for example killing for an extramarital relationship) can be considered as a Type -1 crime. Here the suspects show more symptoms of telling a lie.

Crime Type-2: The crime is done only for economic profit, there is no emotional issue. Any type of fraud activities such as stealing, cheating, abduction, mugging, etc, can be considered as crime Type -2. Here the suspects show fewer symptoms. A questionnaire for this type of crime (for stealing in this case) may be as following Table 3.3.

Table 3.3: Questionnaire for Crime Type-2.

Phase	Possible Question	Possible Answer	Symptom Possibility
1	Have you done this?	No	NO
2	Where was your last day?	At my village home	LOW
3	What were you doing at 10.00 am?	Playing cards	LOW
4	Who was with you?	My other cousins.	NO
5	What does your cousin do?	They study in school.	NO
6	The last day was not a vacation. What were they doing with you in school time?	They had a fever and did not go to school.	HIGH
7	Both of them had a fever?	May be	HIGH

Now we see that the suspect provides external symptoms at different levels of questions. The changed focus of attention offered by the suspect is our key to lie detection. The overall system is depicted by the following flowchart illustrated in Figure 3.18.

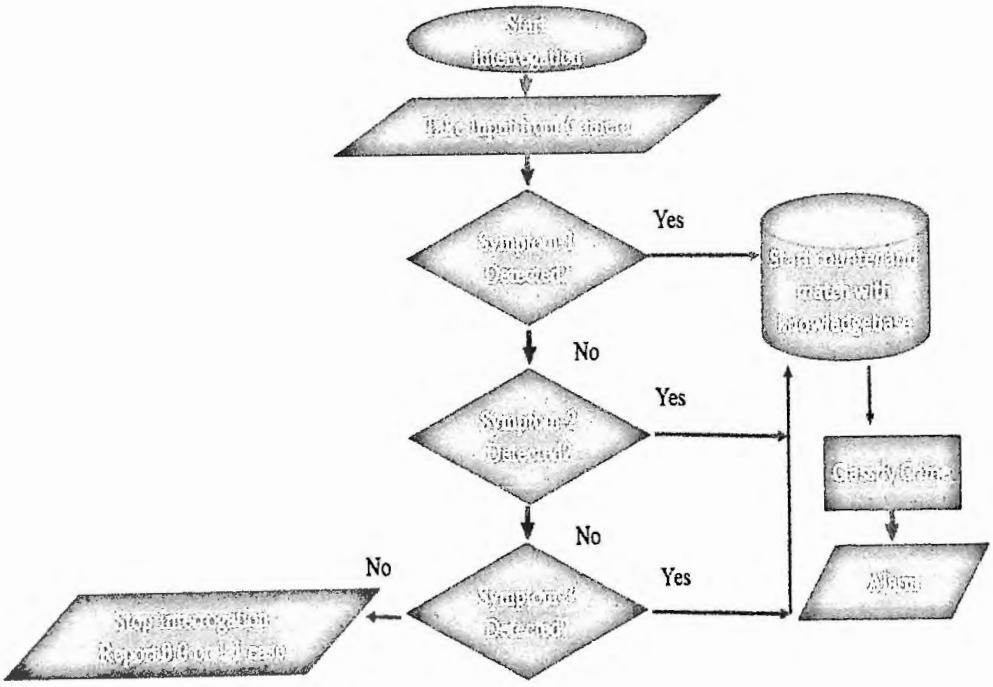


Figure 3.20: Programming Flow Chart of Lie Detection System.

To detect the lie of the target person, at first we have detected the head from the captured image followed by face point extraction by Active Shape Model. After that, the random fluctuation of the eyeball is detected using the VFIG method as discussed earlier. We have also tracked the eyebrow movement. To estimate the optical flow of the eyebrow we need the sequences of the ordered images. This approach tries to evaluate the motion between the two consecutive images taken at time t and t+ Δt. In our approach, we have used the Lucas-Kanade method [82]. For a pixel under consideration, it considers the optical flow constant in its neighborhoods. Using the least square criterion solves all the optical flow equations for all of the neighborhood pixels. To distinguish among different image elements, we have used The Shi and Tomasi corner detection algorithm [83]. In our proposed work, the eye area has been considered the Region of Interest (ROI), where the change of the shape of the eyebrow is detected.

3.6 Summary

In the end, we say that for gaze detection of the target suspect we have tracked the random fluctuation of the head (detected by Haar cascade classifier) and eyeball (detected by vector field of image gradient). The body movement has been detected by Canny Edge Detector with some advanced approaches. Age, gender, and emotion detection are performed with the help of Artificial Neural Network on collected datasets. Finally being merged with the previous techniques, optical flow feature was introduces for lie detection in real time interrogation.

Chapter 4
System Implementation

In this chapter, we have described the System model and parameter list needed to create a system environment in which we conducted all of our experiments. Hardware and software setup has also been added in the last portion of this chapter.

4.1 System Modeling

The system model generally describes how the system will be implemented sequentially also describes what parameters are chosen during implementation. It provides a clear concept about our technical approach and supports standard quality programming.

4.1.1 System Representation

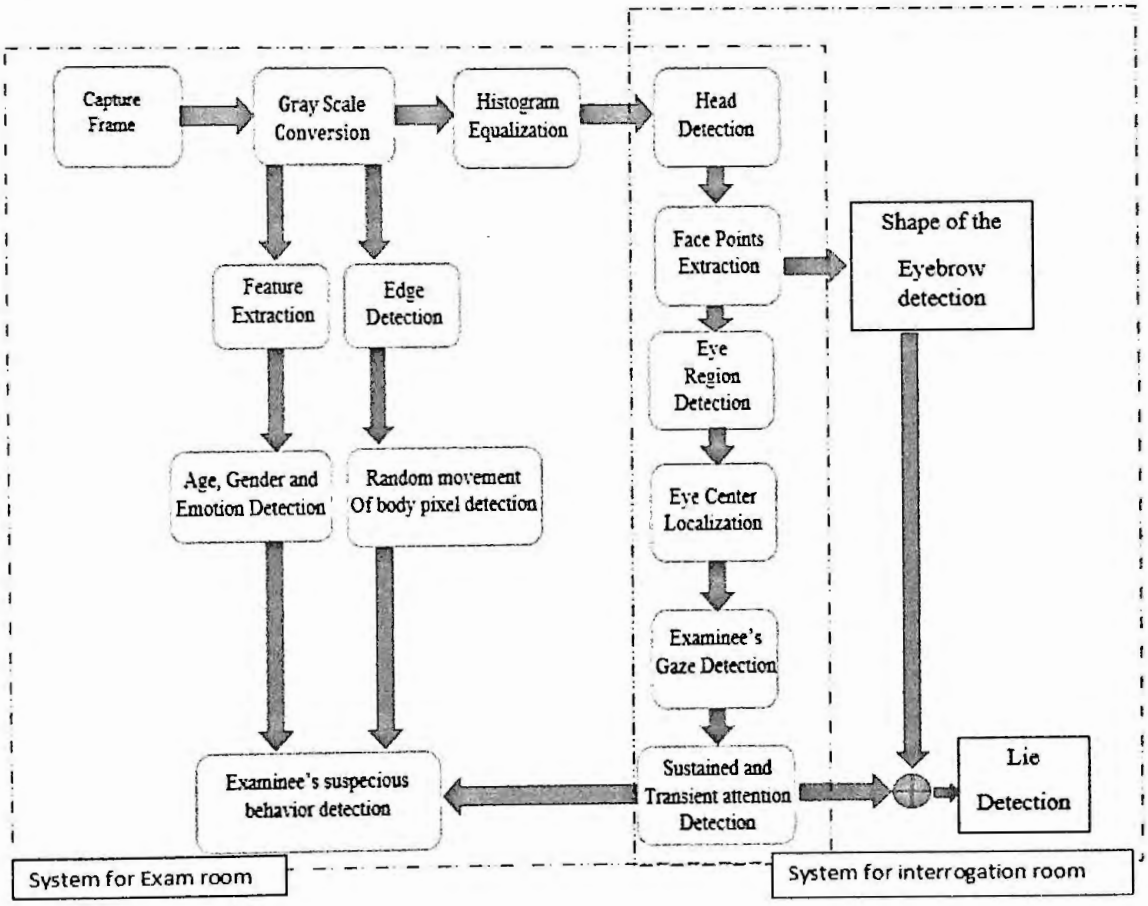


Figure 4.1: Block diagram representation of the combined system.

4.1.2 Description of the Proposed System

A step-by-step analysis of different modules used as a part of our system is given below.

Frame Capturing: At first, we need to capture the continuous frames of the examinee from the camera. The camera is placed in front of the examinee in such a way that it can detect the examinee's head and eye area efficiently.

Gray Scale Conversion: The captured frame by frame continuous images then transferred to the system and it converts the image into the grayscale so that they may be processed easily comparing other techniques.

Histogram Equalization: Histogram equalization is performed on each of the images so that they may be processed in low lighting conditions.

Head Detection: Haar cascade classifier and 3-D head tracker is used to detect the head of the examinee.

Face Points Extraction: When the head is detected from the continuous frame by frame images, face points are extracted using the active shape model.

Eye Region Detection: Eye regions are detected with the help of face point extraction.

Eye Center Localization: Vector field of image gradient is used to detect the eye center.

Pseudocode:

```
Step-1: Start by taking the input image as out and facepoints as Ptr

Step-2: Initialize cy = 0, cx=0;

Step-3: For cy<=EyeRectangle.rows, increase cy

    Define *Or = out.ptr<double>(cy);

    Step-4: For cx<= EyeRectangle.col, increase cx

        X=cx, Y=cy;

        Step-5: create a  displacement vector from the possible center to the gradient
        origin

            define dx = x - cx;

            define dy = y - cy;
```

Step-6: normalize d

*define magnitude = sqrt((dx * dx) + (dy * dy));*

dx = dx / magnitude;

dy = dy / magnitude;

Step-7: calculate dot product with the image gradient

*dotProduct = dx*gx + dy*gy;*

dotProduct = std::max(0.0,dotProduct);

Step-8: square and multiply by the weight defined by blurred and inverted image

if (kEnableWeight) {

*Or[cx] += dotProduct * dotProduct * (weight/kWeightDivisor);*

} else {

*Or[cx] += dotProduct * dotProduct;*

}

Step-7: End of For loop

Step-8: End of For loop

Step-9: Compute the Global Maxima of the dot product

Gaze Detection of the Examinee: The gaze (VFOA) is detected based on the coordinate of the pupil and the head position in the head rectangle.

Pseudocode to track the fluctuation of Head:

Step-1: start by taking the input of the width of face rectangle

Step-2: divide the width in five equal parts s

Step-3: Define Left side, Front and Right side

Step-4: compare and get in which defined side the middle of face line is located

Step-5: define the gaze of the target person

Pseudocode to track the fluctuation of Eyeball:

Step-1: start by taking the input of the width of eye rectangle

Step-2: divide the width in five unequal parts u

Step-3: Define Left uncovered, Left, Monitor, Right, Right uncovered areas

Step-4: compare and get in which defined areas the x-coordinate of the left pupil is located

Step-5: IF gaze is found in Left uncovered || in Right uncovered

Continue;

ELSE

Define the gaze of the target person

Step-6: Normalize the decision from 20 consecutive frames

Sustained and Transient Attention Detection: Sustained and Transient Attention is detected based on the proper threshold value of time.

Pseudocode

Step-1: Start by initiating counter, t=0

Step-2: While (t>0)

Step-3: Take the defined Gaze, G_D of the target person

Step-4: For 0<t<3, calculate the Gaze, G_s of the target person

Step-5: IF $G_D > G_s$, define the transient attention

ELSE continue

Step-6: end of FOR

Step-7: For t>8, calculate the Gaze, G_t of the target person

Step-8: IF $G_D > G_t$, define the sustained attention

ELSE continue

Step-9: end of FOR

Step-10: IF t=20

Break;

Age, Gender, and Emotion Detection: Examinee's Age, Gender, and Emotion are detected using a neural network.

Pseudocode:

Step-1: Load the classifier and pre-trained model

Step-2: Define our model parameters

Step-3: Get our weight file

Step-4: Initialize Webcam and capture image, I

Step-5: while (I. faceclassifier>0):

    Step-6: perform preprocessing (resize and grayscale conversion)

    Step-7: make a prediction for Age and Gender

    Step-8: make a prediction for Emotion

    Step-9: Overlay our detected emotion on our Image

Step-10: Release frames

Random Movement of Body Pixel Detection: Examinee's random movement is detected based on the result of the Canny edge detection method.

Pseudocode:

Step-1: Initialize times, n=0, Threshold-1, T_{h1} and Threshold- 2, T_{h2}

Step-2: While (n>0)

Step-3: capture frames I_n

Step-4: calculate the two-difference image, $I_1 = \lfloor I_n - I_{n-1} \rfloor$ and $I_2 = \lfloor I_n - I_{n+1} \rfloor$

Step-5: calculate $I_{final} = I_1 \cap I_2$

Step-6: calculate the number of black pixels, B_{p1} on I_{final}

Step-7: IF $B_{p1} > T_{h1}$,

Provide alarm;

BREAK and reset n;

ELSE

Step-8: initialize m=0

Step-9: while ($B_{p1} < T_{h1}$)

Step-10: perform operations from step-4 and step-5 on $I_{n\pm m}$ images and calculate I_{final_n}, I_{final_n+m}, I_{final_n-m} images

Step-11: calculate $I_{ultimate1} = I_{final\_n} || I_{final\_n+m} || I_{final\_n-m}$

Step-8: increase m and repeat step-10 and step-11 to calculate $I_{ultimate2}$

Step-9: calculate $I_{ultimate} = I_{ultimate1} || I_{ultimate2}$

Step-10: calculate the number of black pixels, B_{p1}

Step-11: IF $B_{p1} > T_{h2}$

Provide alarm;

BREAK and reset n;

ELSE

Continue;

The shape of the Eyebrow detection: We have used optical flow features to track the change of the shape of the eyebrow.

Pseudocode:

Step-1: Initialize parameters for corner detection

Step-2: Initialize Parameters for lucas kanade optical flow

Step-3: Take first frame and find corners in it

Step-4: Create a mask image for drawing purposes

Step-5: while (1):

Step-6: calculate optical flow

Step-7: Select good points

Step-8: draw the tracks

Step-9: break;

Step-10: Updating previous frame and points

4.1.3 Parameter List

In an attempt to use realistic and typical values, the following parameters depicted in Table 4.1 were used. Using these parameters and their corresponding values, we conducted all of our experiments.

Table 4.1: List of parameters.

No.	Parameter	Parameter Value
1	Distance from the camera	30 cm, 40 cm, 50 cm, 60 cm, 70 cm, 80 cm, 90 cm, 100 cm
2	Lighting condition	50 lux, 100 lux, 150 lux, 200 lux
3	Age of participants	8 to 65
4	Head detection technique	Optical flow feature
5	Iris center detection	3-D head tracker, Active Shape Model (ASM), Vector Field of Image Gradient (VFIG)
6	Gaze detection of examinee using the coordinate of pupil	Equal division of eye rectangle width, partitioning eye rectangle leaving the uncovered region, unequal partitioning of eye rectangle leaving the uncovered region

Table 4.1: List of parameters (Continue).

No.	Parameter	Parameter Value
7	Gaze detection using the head pose	Partitioning the head rectangle
8	Head rotation time (Second)	1, 2, 3
9	Sustained focus duration (Second)	Minimum 8 seconds
10	Transient focus duration (Second)	1, 2, 3
11	Video length	Average 2 minutes for object detection experiments, for real-time observation average is 4 minutes 43 seconds
12	Edge detection technique	Canny edge detector
13	Age Detection Framework	TensorFlow
14	Gender Detection Framework	TensorFlow
15	Emotion Detection Model	AlexNet

4.2 System Environment Creation

A suitable system creation depends on proper hardware and software installation which has been described properly in this section.

4.2.1 Hardware Setup

To run head, face detection, and random movement of body pixel detection in neural network field programs we need a computer with good processing speed. In our experiments, we have used a laptop with the following specification shown in Table 4.2.

Table 4.2: Laptop Specification.

Model	HP ProBook 430 G3
RAM	8 GB
Processor	2.30 GHz
Hard drive	1 TB
Display	14.0 inch

We have used a low-cost camera. The camera specification is showed in Table 4.3.

Table 4.3: Camera specification.

Model	Logitech c170
Photo quality	5 MP
Video quality	HD 720p
Focus type	40 cm and beyond
Auto light correction	Premium

4.2.2 Software Setup

We need the following software to run the program.

1. **Anaconda version 3.6.**
2. **Visual Studio 2010.**
3. **OpenCV 2.4.9.**

Anaconda is a free and open-source [84] distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. Package versions are managed by the package management system conda [85] The Anaconda distribution includes data-science packages suitable for Windows, Linux, and MacOS.

Microsoft Visual Studio is an integrated development environment (IDE) from Microsoft. It is used to develop computer programs, as well as websites, web apps, web services, and mobile apps. Visual Studio uses Microsoft software development platforms such as Windows API, Windows Forms, Windows Presentation Foundation, Windows Store, and Microsoft Silverlight. It can produce both native codes and managed code.

OpenCV (Open source computer vision) is a library of programming functions mainly aimed at real-time computer vision [86] originally developed by Intel, it was later supported by Willow Garage then Itseez (which was later acquired by Intel [87]). The library is cross-platform and free for use under the open-source BSD license.

4.3 Summary

In concluding remarks, we see that high-end PC with a low-end cameras have been utilized to implement our system. We need to keep in mind that the real invigilation/ interrogation system is much more complicated. Simulations are made based on some acceptable system environments because of some difficulties in hardware installation and data collection in a real environment.

Chapter 5
Experiments, Results and Discussion

This chapter comprises three segments: data collection, results, and performance evaluation in terms of accuracy of detection. Data are collected using different participants and under different conditions. Screenshots from the detection results were included in this chapter's first section. Results are obtained through the analysis of the experimental outlines. The accuracy of our method has been shown by the histogram in a particular detection approach and discussed in the third segment of this chapter.

5.1 Experiments to Detect Cheating Behavior in Exam Hall

To detect the cheating behavior in examination hall, we performed several experiments by considering different parameters like eye and head rotation, body movement detection, age, gender and emotion detection. All the experiments with accuracy measures are described below

5.1.1 Experiments for Gaze Detection

5.1.1.1 Data Collection

We've gathered real data from different people and their ages and gender are different as well.

Participants

The participants in my study are both male and female. Participants' ages range from 8 to 65 years. There were a total of 150 participants for testing purposes.

Procedure of Gaze Detection Based on Object Detection

The data collection technique is defined as follows for different experiments-

Head and Eye Center Detection

The participants were asked to sit at different distances from the camera under different lighting conditions during head and eye pupil monitoring tests. We stood away from the camera at 30 cm, 40 cm, 50 cm, 60 cm, 70 cm, 80 cm, 90 cm, and 100 cm. Different lighting conditions were given like 50, 100, 150, and 200 lux.

Visual Focus Detection

These participants were asked to sit 70 cm away from the camera for gaze identification by applying different techniques and to look at different target items with varying head rotation

times (1, 2, 3 seconds). These artifacts lie 0.5 m apart from each other. One of the three objects is in the Central View Field (CFV) and the other two are in the Near Peripheral View Field (NPFV). The average recording time for the video had been 2 minutes.

Sustained and Transient Focus of Attention Detection

The participants were asked to stare at the left object and the right object for a very short duration in continuous and intermittent focus detection tests so that they could be viewed as transients. During various experiments, the duration of the transients was also varied. The total processing time for the video had been 2 minutes.

5.1.1.2 Results

In these experiments, the participant looked at different objects with different duration of observation and different head rotation time. Gaze of the participant is extracted from object detection.

Head and Eye center Detection

The accuracy of head and eye center detection depends on the distance from the camera and proper lighting condition. The head and eye center detection for different distances from camera has been showed in **Table 5.1** and for different lighting condition has been showed in **Table 5.2.** For different lighting condition, the distance between the bulb and the participant is kept 2 meters.

Table 5.1: Head and eye center detection with the varying distances from the camera.

No. of Trial	Distance from the camera (cm)	Detection
1	30	No
2	30	No
3	30	No
4	30	No
5	30	Yes
6	30	Yes
7	30	No
8	30	No
9	30	No
10	30	Yes
11	40	Yes
12	40	No
13	40	No
14	40	No
15	40	Yes
16	40	Yes
17	40	No
18	40	Yes
19	40	No
20	40	Yes
21	50	Yes
22	50	Yes
23	50	Yes
24	50	Yes
25	50	Yes
26	50	Yes
27	50	No
28	50	No
29	50	Yes
30	50	No
31	60	Yes
32	60	Yes
33	60	Yes
34	60	Yes
35	60	Yes

Table 5.1: Head and eye center detection with the varying distances from the camera (Continue).

No. of Trial	Distance from the camera (cm)	Detection
36	60	Yes
37	60	Yes
38	60	Yes
39	60	Yes
40	60	Yes
41	70	Yes
42	70	Yes
43	70	Yes
45	70	Yes
46	70	Yes
47	70	Yes
48	70	Yes
49	70	Yes
50	70	Yes
51	80	Yes
52	80	Yes
53	80	Yes
55	80	Yes
56	80	Yes
57	80	Yes
58	80	Yes
59	80	Yes
60	80	Yes
61	90	Yes
62	90	No
63	90	Yes
65	90	Yes
66	90	Yes
67	90	Yes
68	90	Yes
69	90	Yes
70	90	Yes
71	100	No
72	100	No
73	100	Yes
74	100	Yes
75	100	Yes
76	100	Yes

Table 5.2: Head and eye center detection with varying lighting conditions.

No. of trial	Illumination (Lux)	Detection
1	50	yes
2	50	no
3	50	no
4	50	yes
5	50	no
6	50	yes
7	50	no
8	50	no
9	50	yes
10	50	yes
11	100	no
12	100	yes
13	100	yes
14	100	yes
15	100	yes
16	100	yes
17	100	yes
18	100	yes
19	100	no
20	100	yes
21	150	yes
22	150	yes
23	150	yes
24	150	yes
25	150	yes
26	150	yes
27	150	yes
28	150	yes

Table 5.2: Head and eye center detection with varying lighting conditions (Continue).

No. of trial	Illumination (Lux)	Detection
29	150	yes
30	150	no
31	200	yes
32	200	yes
33	200	yes
34	200	yes
35	200	yes
36	200	yes
37	200	yes
38	200	yes
39	200	yes
40	200	yes

Gaze Detection from the Coordinate of Pupil

For gaze detection with only coordinate of pupil of the left eye, we conducted 3 separate experiments with variable duration through applying different gaze detection techniques .**Table – 5.3, 5.4** and **5.5** shows the result for some sample successive frames applying **Equal Division of Eye Rectangle Width, Partitioning Eye Rectangle Leaving the Uncovered Regions** and **Unequal Partitioning of Eye Rectangle Leaving the Uncovered Regions** techniques respectively.

Table 5.3: Gaze detection from the coordinate of pupil using equal division of eye rectangle width.

Frame no	X coordinate	Regions			Looking At	
		Left	Monitor	Right	Simulation	Original
1	299	283-297	298-305	306-321	monitor	left
2	307	286-300	301-308	309-324	monitor	left
3	306	284-298	299-306	307-322	monitor	left
4	288	286-300	301-308	309-324	left	left
5	295	283-297	298-305	306-321	left	left
6	305	289-303	304-311	312-327	monitor	left
7	304	282-296	297-304	305-320	monitor	left
8	305	283-297	298-305	306-321	monitor	left
9	299	283-297	298-305	306-321	monitor	left
10	290	282-296	297-304	305-320	left	left
11	307	286-300	301-308	309-324	monitor	right
12	321	284-298	299-306	307-322	right	right
13	299	283-297	298-305	306-321	monitor	right
14	299	283-297	298-305	306-321	monitor	right
15	299	283-297	298-305	306-321	monitor	right
16	290	282-296	297-304	305-320	left	right
17	308	284-298	299-306	307-322	right	right
18	312	286-300	301-308	309-324	right	right
19	304	282-296	297-304	305-320	monitor	right
20	313	289-303	304-311	312-327	right	right
21	307	286-300	301-308	309-324	monitor	monitor
22	302	286-300	301-308	309-324	monitor	monitor
23	307	291-305	306-313	314-329	monitor	monitor
24	305	283-297	298-305	306-321	monitor	monitor
25	304	282-296	297-304	305-320	monitor	monitor
26	305	283-297	298-305	306-321	monitor	monitor
27	313	289-303	304-311	312-327	right	monitor
28	307	291-305	306-313	314-329	monitor	monitor
29	305	289-303	304-311	312-327	monitor	monitor
30	290	282-296	297-304	305-320	left	monitor

Table 5.4: Gaze detection from the coordinate of pupil using partitioning eye rectangle leaving the uncovered regions.

Frame no	X coordinate	Regions			Looking At	
		Left	Monitor	Right	Simulation	Original
1	294	291-305	306-313	314-329	left	left
2	287	286-300	301-308	309-324	left	left
3	303	291-305	306-313	314-329	left	left
4	288	286-300	301-308	309-324	left	left
5	295	283-297	298-305	306-321	left	left
6	305	289-303	304-311	312-327	monitor	left
7	304	282-296	297-304	305-320	monitor	left
8	305	283-297	298-305	306-321	monitor	left
9	299	283-297	298-305	306-321	monitor	left
10	290	282-296	297-304	305-320	left	left
11	307	286-300	301-308	309-324	monitor	right
12	323	284-298	299-306	307-322	right	right
13	321	291-305	306-313	314-329	right	right
14	323	284-298	299-306	307-322	right	right
15	324	286-300	301-308	309-324	right	right
16	307	291-305	306-313	314-329	monitor	right
17	305	283-297	298-305	306-321	monitor	right
18	304	282-296	297-304	305-320	monitor	right
19	304	282-296	297-304	305-320	monitor	right
20	313	289-303	304-311	312-327	right	right
21	290	282-296	297-304	305-320	left	monitor
22	305	289-303	304-311	312-327	monitor	monitor
23	299	283-297	298-305	306-321	monitor	monitor
24	307	291-305	306-313	314-329	monitor	monitor
25	303	286-300	301-308	309-324	monitor	monitor
26	305	283-297	298-305	306-321	monitor	monitor
27	313	289-303	304-311	312-327	right	monitor
28	307	291-305	306-313	314-329	monitor	monitor
29	305	289-303	304-311	312-327	monitor	monitor
30	305	289-303	304-311	312-327	monitor	monitor

Table 5.5: Gaze detection from the coordinate of pupil using unequal partitioning of eye rectangle leaving the uncovered regions.

Frame No	X coordinate	Regions			Looking At	
		Left	Monitor	Right	Simulation	Original
1	299	283-297	298-305	306-321	right	right
2	307	286-300	301-308	309-324	monitor	right
3	328	291-305	306-313	314-329	right	right
4	320	289-303	304-311	312-327	right	right
5	312	286-300	301-308	309-324	right	right
6	315	285-299	300-307	308-323	right	right
7	304	282-296	297-304	305-320	monitor	right
8	314	283-297	298-305	306-321	right	right
9	290	289-303	304-311	312-327	left	left
10	292	291-305	306-313	314-329	left	left
11	287	286-300	301-308	309-324	left	left
12	299	283-297	298-305	306-321	monitor	left
13	298	285-299	300-307	308-323	left	left
14	297	285-299	300-307	308-323	left	left
15	296	291-305	306-313	314-329	left	left
16	299	283-297	298-305	306-321	monitor	left
17	294	291-305	306-313	314-329	left	left
18	287	286-300	301-308	309-324	left	left
19	303	291-305	306-313	314-329	left	left
20	305	289-303	304-311	312-327	monitor	monitor
21	299	283-297	298-305	306-321	monitor	monitor
22	307	286-300	301-308	309-324	monitor	monitor
23	306	284-298	299-306	307-322	monitor	monitor
24	286	285-299	300-307	308-323	left	monitor
25	305	289-303	304-311	312-327	monitor	monitor
26	299	283-297	298-305	306-321	monitor	monitor
27	307	291-305	306-313	314-329	monitor	monitor
28	303	286-300	301-308	309-324	monitor	monitor
29	301	285-299	300-307	308-323	monitor	monitor
30	307	283-297	298-305	306-321	right	monitor
31	308	284-298	299-306	307-322	right	right
32	312	286-300	301-308	309-324	right	right
33	315	284-298	299-306	307-322	right	right
34	322	289-303	304-311	312-327	right	right
35	284	283-297	298-305	306-321	left	right
36	302	286-300	301-308	309-324	monitor	right
37	321	291-305	306-313	314-329	right	right

Table 5.5: Gaze detection from the coordinate of pupil using unequal partitioning of eye rectangle leaving the uncovered regions (Continue).

Frame No	X coordinate	Regions			Looking At	
		Left	Monitor	Right	Simulation	Original
38	323	284-298	299-306	307-322	right	right
39	324	286-300	301-308	309-324	right	right
40	285	283-297	298-305	306-321	left	right
41	307	282-296	297-304	305-320	right	right
42	311	284-298	299-306	307-322	right	right
43	286	285-299	300-307	308-323	left	left
44	288	286-300	301-308	309-324	left	left
45	295	283-297	298-305	306-321	left	left
46	307	291-305	306-313	314-329	monitor	left
47	293	282-296	297-304	305-320	left	left
48	301	291-305	306-313	314-329	left	left
49	299	286-300	301-308	309-324	left	left
50	299	283-297	298-305	306-321	monitor	left
51	290	282-296	297-304	305-320	left	left
52	301	285-299	300-307	308-323	monitor	monitor
53	307	291-305	306-313	314-329	monitor	monitor
54	305	289-303	304-311	312-327	monitor	monitor
55	304	282-296	297-304	305-320	monitor	monitor
56	305	283-297	298-305	306-321	monitor	monitor
57	313	289-303	304-311	312-327	right	monitor
58	302	286-300	301-308	309-324	monitor	monitor

Gaze Detection from the Coordinate of Pupil Combining with Head Pose

For gaze detection combining with head pose and coordinate of pupil, we conducted 30 separate experiments with three varying duration: 3 seconds, 2 seconds and 1 second -which is showed in **Table 5.6**.

Table 5.6: Gaze detection from the coordinate of pupil combining with head pose.

No. of Trial	Rotation Time (Second)	Monitor to left detection	Monitor to Right detection
1	3	yes	yes
2	3	yes	yes
3	3	yes	yes
4	3	yes	yes
5	3	yes	yes
6	3	yes	yes
7	3	yes	yes
8	3	yes	yes
9	3	yes	yes
10	3	yes	yes
11	2	yes	yes
12	2	yes	no
13	2	yes	no
14	2	no	yes
15	2	yes	yes
16	2	yes	yes
17	2	yes	yes
18	2	yes	no
19	2	no	yes
20	2	yes	yes
21	1	yes	no
22	1	no	yes
23	1	no	no
24	1	yes	yes
25	1	no	no
26	1	yes	no
27	1	yes	no
28	1	no	yes
29	1	no	no
30	1	yes	yes

Sustained and Transient Focus of Attention Detection

Table 5.7 shows the detection of transients with varying duration. For sustained and transient focus detection we conducted 3 experiments.

Table 5.7: Detection of transients of varying duration.

No. of trial	Transient Duration (Seconds)	Detection	
		Left transient	Right transient
1	1	no	yes
2	1	no	no
3	1	no	no
4	1	no	yes
5	1	yes	yes
6	1	no	no
7	1	yes	no
8	1	no	no
9	1	no	no
10	1	no	no
11	2	yes	no
12	2	yes	yes
13	2	no	yes
14	2	no	yes
15	2	no	no
16	2	yes	yes
17	2	no	yes
18	2	yes	no
19	2	no	yes
20	2	yes	no
21	3	yes	yes
22	3	yes	yes
23	3	yes	yes
24	3	yes	yes
25	3	yes	yes
26	3	yes	yes
27	3	yes	yes
28	3	yes	yes
29	3	yes	yes
30	3	yes	yes

5.1.1.3 Performance Evaluation

Quality appraisal reflects how effective our program is in measuring examinee's visual attention. They need to keep in mind that the bulk of the findings were taken to help the tests in an artificial environment. The findings may not fully match the real-world tests collected. The output is evaluated as follows in various experiments using the object detection technique to identify the examinee's gaze:

Performance Evaluation for Head and Eye Center Detection

We did various tests with equal distance from the camera and different lighting conditions. The sensitivity of head and eye center identification is described as follows for the best lighting condition-

$$\textbf{Accuracy} = \frac{\text{Total number of trials in which head and eye center is detectd}}{\text{Total number of trials at a particular distance from camera}} \times 100\% \qquad (5.1)$$

The output of identification of head and eye center also depends on the illumination state. The detection precision for a given distance from the camera is expressed as follows-

$$\textbf{Accuracy} = \frac{\text{Total number of trials in which head and eye center is detectd}}{\text{Total number of trials at a particular lighting condition}} \times 100\% \qquad (5.2)$$

Figure 5.1 and Figure 5.2 display accuracy tests with different camera lengths and different lighting conditions, respectively.

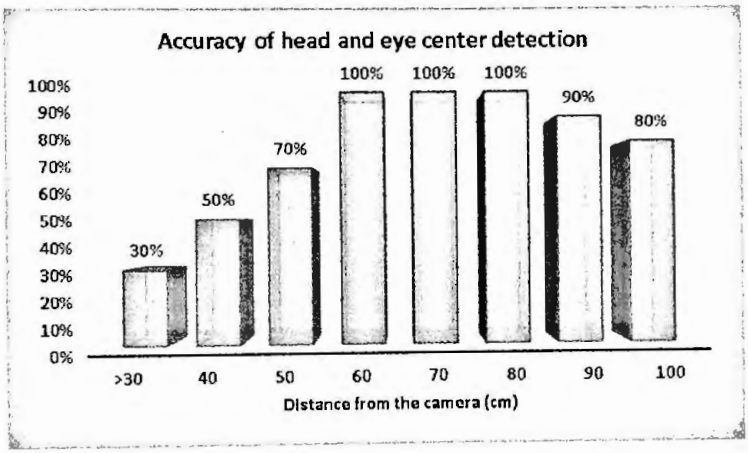


Figure 5.1: Accuracy of head and eye center detection for different distances from camera.

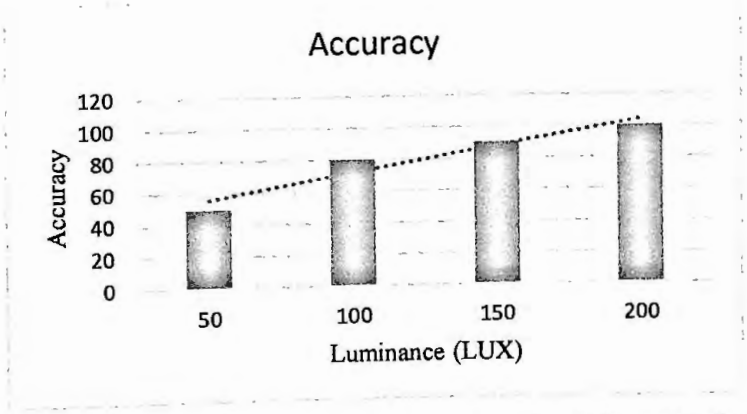


Figure 5.2: Accuracy of head and eye center detection for different lighting conditions.

Performance Evaluation for Gaze Detection with Coordinate of Pupil

While the gaze is sensed from the pupil's coordinate, the result is evaluated in terms of "accuracy" which is defined as follows-

$$\textbf{Accuracy} = \frac{\text{Total number of gaze detection in a particular direction}}{\text{Total number of gaze occured in that particular direction}} \times 100\% \qquad (5.3)$$

We carried out 3 different experiments using the following three techniques-

- Equal division of eye rectangle width
- Partitioning eye rectangle leaving the uncovered regions
- Unequal Partitioning of eye rectangle leaving the uncovered regions

The overall accuracy value for detecting the left, display and right look with different detection techniques is shown in the following histogram (Figure 5.3).

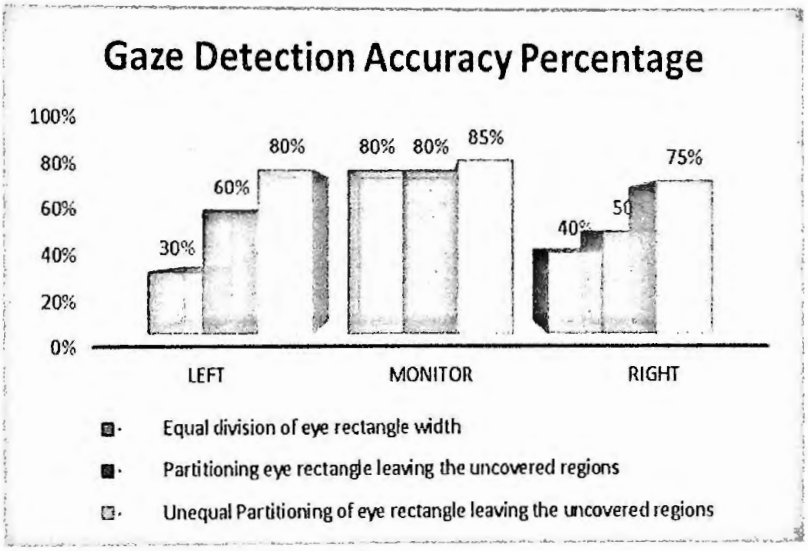


Figure 5.3: Accuracy percentage of different gaze detection using the coordinate of pupil and applying different detection techniques.

5.1.2 Experiments for Body Movement Detection

5.1.2.1 Data Collection

We performed two types of experiments that are conducted in two separate environments. The 3 unpaid participants (3 male) were students at Bangladesh Army University of Engineering and Technology mean age of 20 years and standard deviation of 0.775 years.

Experiment-1

We collected data from three different examination halls running the semester final exams. In this experiment, data is collected from an examination room where a quiz test is running. The quiz test is in the form of multiple-choice questions to be filled up on the screen of the computer with the help of a mouse provided to each of the examinees. Each of the exams comprises 1.00 hours, total 30 students were seated for examination into each of the considered Examination halls. Apart from looking behind, the victim is capable of copying from its neighbor.

Experiment-2

Based on the previous cheating records, we identified 3 different victims. Without their conscious, they were monitored very closely in the different examinations and their cheating profiles were updated based on it.

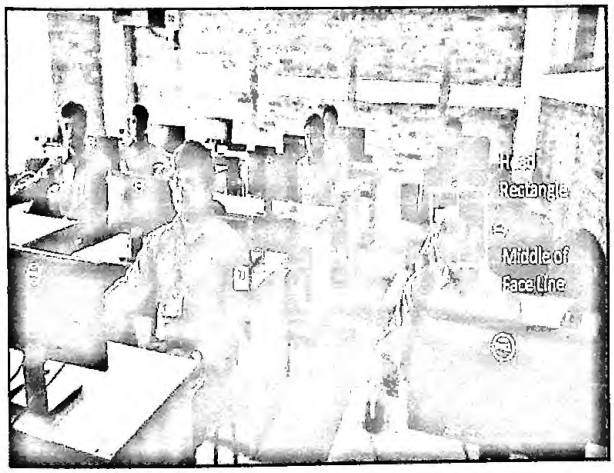


Figure 5.4: Scenario of cheating Type-2 in smart class room.

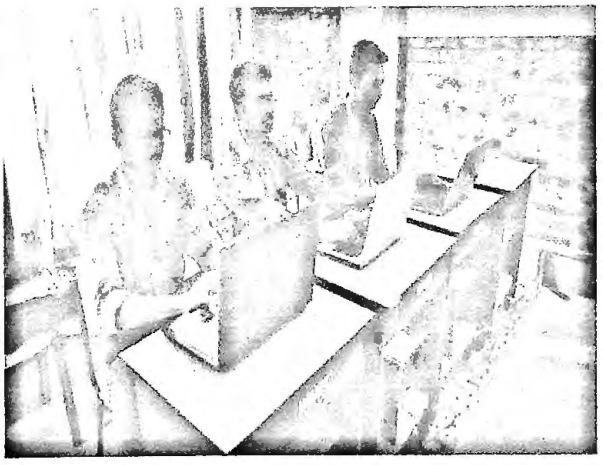


Figure 5.5: Experimental setup in controlled environment.

5.1.2.2 Results

All the experiments are stored as video files. Results are collected by analyzing these videos and represented in different tables. Our following tables summarize the studies from different halls according to different time slots. Each of the students were very closely monitored to detect the suspicious behavior in terms of random head movement or eyeball fluctuation. Table 5.8 and Table 5.9 reveals that cheating activities is crucial in final 20 minutes for all the Examination halls. Here data is collected from Experiment-1.

Table 5.8: Cheating Type -1 detection from different Examination hall in different time slots.

	Cheating Type-1		
	Hall -1	Hall-2	Hall-3
First 20 minutes	0	1	0
Second 20 minutes	0	2	1
Final 20 minutes	1	2	3

Table 5.9: Cheating Type -2 detection from different Examination hall in different time slots.

	Cheating Type-2		
	Hall -1	Hall-2	Hall-3
First 20 minutes	7	10	8
Second 20 minutes	15	13	17
Final 20 minutes	25	27	28

From Experiment-2, different victims with copy prone attitudes were monitored closely. We updated their cheating profile as in Table 5.10 to Table 5.12.

Table 5.10: Cheating profile of the victim -1.

	Cheating Type-1			**Cheating Type-2**		
	Exam			Exam		
	1	**2**	**3**	**1**	**2**	**3**
First 20 minutes	0	0	1	3	1	2
Second 20 minutes	0	1	2	4	3	5
Final 20 minutes	1	2	4	5	5	6

Table 5.11: Cheating profile of the victim -2.

	Cheating Type-1			Cheating Type-2		
	Exam			Exam		
	1	2	3	1	2	3
First 20 minutes	1	0	0	1	1	2
Second 20 minutes	0	0	2	4	4	5
Final 20 minutes	1	2	4	5	5	5

Table 5.12: Cheating profile of the victim -3.

	Cheating Type-1			Cheating Type-2		
	Exam			Exam		
	1	2	3	1	2	3
First 20 minutes	1	4	4	5	2	3
Second 20 minutes	1	5	5	6	4	5
Final 20 minutes	4	6	6	7	5	7

5.1.2.3 Performance Evaluation for Edge detection technique

$$\textbf{Accuracy} = \frac{\text{Total number of body movement detected}}{\text{Total number of body movement occured}} \times 100\% \qquad (5.4)$$

Canny edge detector accurately senses the random movement of the examinees when the examinee moves his / her body taking 3, 4 and 5 seconds. The accuracy degrades in 1 and 2 seconds because our device camera can't accurately calibrate or focus during this time. Figure 5.6 explains the accuracy in different times of canny edge detector.

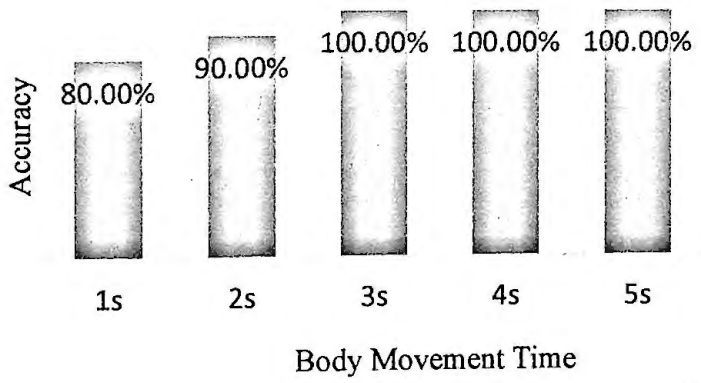


Figure 5.6: Canny edge detector accuracy.

5.1.3 Experiments for Age, Gender and Emotion Detection

Data were collected from different social spaces including schools, colleges, universities and social gatherings like shopping malls and tea stalls and experiments were done in real life and controlled environment.

5.1.3.1 Data collection

Real data were collected from various locations. Both males and females engaged in the system's testing. The participants are aged 8 to 65 years old.

Data Collection for Training

To train our system, we used a total of 28000 data. Our intelligent system training data was broken down into 6 categories. All of the categories are described below with sample pictures:

a) **Angry:** some image samples are given below-

b) **Fear:** some image samples are given below:

c) **Happy:** some image samples are given below:

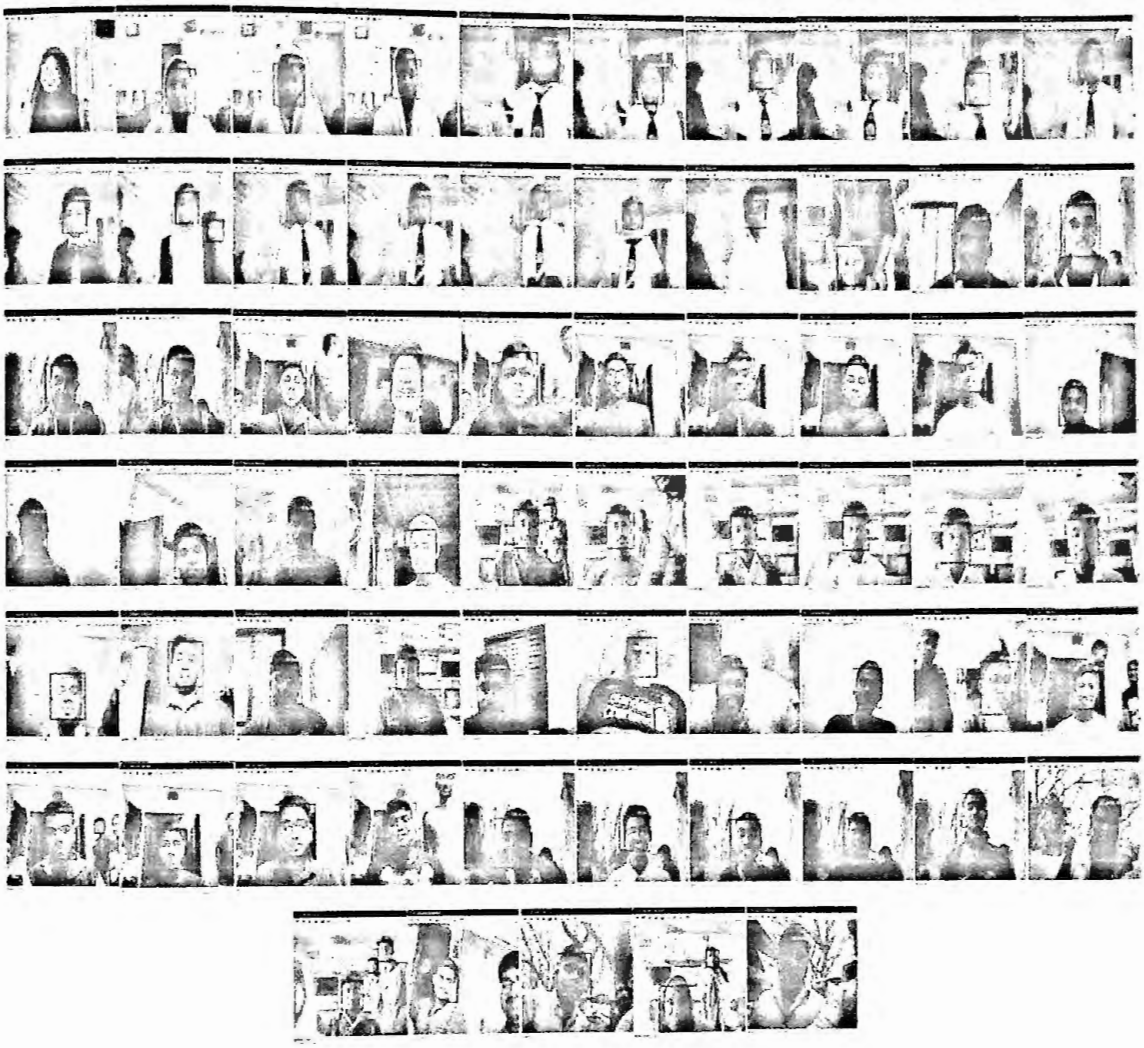
d) **Neutral:** some image samples are given below:

e) **Sad:** some image samples are given below:

f) **Surprise:** some image samples are given below:

Data Collection for Testing Age and Gender in Real Environment

We gather some real-time data while our system was being tested. The samples are given below:

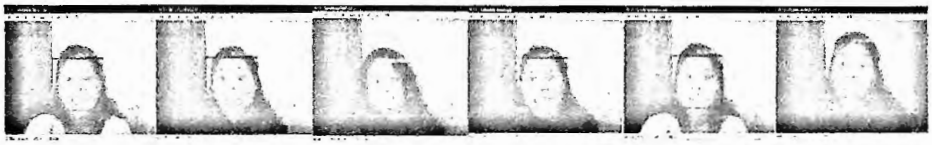


Data Collection for Testing Emotion Detection in Controlled Environment

a) **Angry:**

b) **Fear:**

c) **Happy:**



d) **Neutral:**

e) **Sad:**

f) **Surprise:**

5.1.3.2 Performance Evaluation

Age detection technique

The standard deviation measures the spread of the data about the mean value. This is the formula for Standard Deviation:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x-\mu)^2} \tag{5.5}$$

here,

σ = standard deviation of the detected age from the mean value

N=the size of the age group

X = each value for detected age

μ = the mean of the age group

Figure 5.7 illustrates the standard deviation for different age groups.

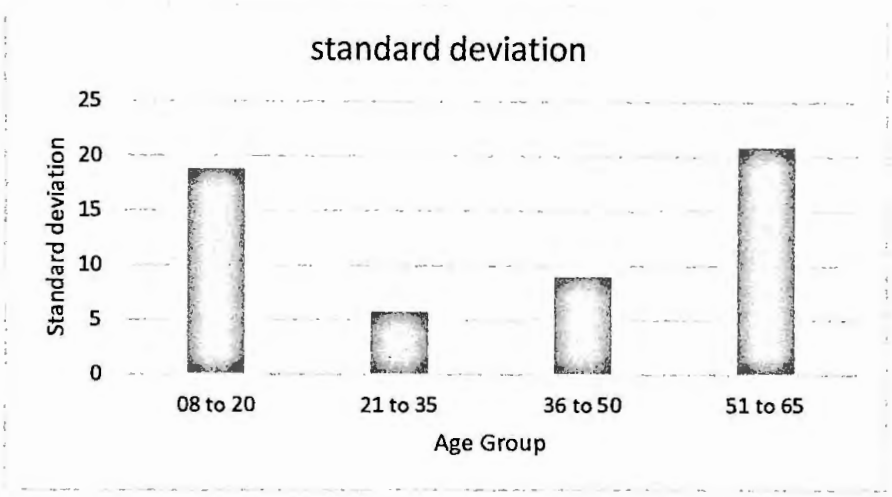


Figure 5.7: Standard deviation for different age groups.

Gender detection with Confusion Matrix and ROC

The confusion matrix is a useful visualization tool that provides analysis on the true negative, false positive, false negative and true positive made by our model. Beyond a simple accuracy metric, we should also look at the confusion matrix to understand the performance of the model.

The definition of true negative, false positive, false negative and true positives in our experiment are as follows:

- **True negative:** Actual class is negative (Male) and the model predicted negative (Male).
- **False positive:** Actual class is negative (Male), but the model predicted positive (Female).
- **False negative:** Actual class is positive (Female) but the model predicted negative (Male).
- **True positive:** Actual class is positive (Female) and the model predicted positive (Female).

Clearly, we want our false positive and false negative numbers to be as low as possible and for the true negative and true positive numbers to be as high as possible.

Confusion matrix result is given below:

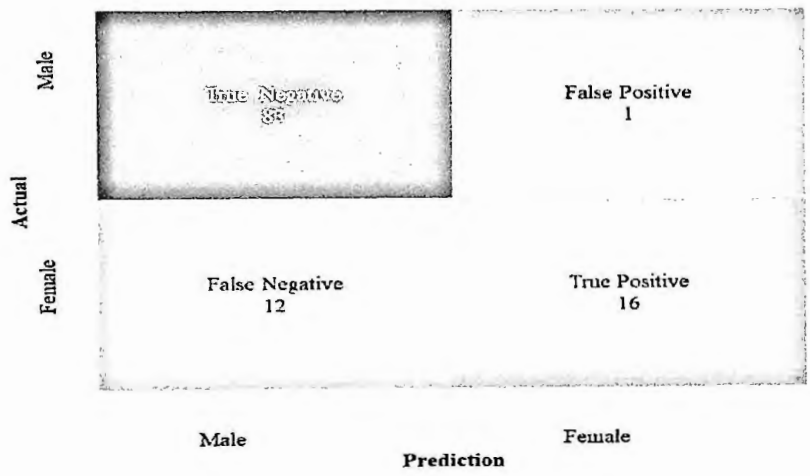


Figure 5.8: Performance evaluation using Confusion Matrix.

For classification tasks, we should also look at the ROC curve to evaluate our model. The ROC curve is plot with the True Positive Rate (TPR) on the y axis and the False Positive Rate (FPR) on the x axis. TPR and FPR are defined as follows:

$$\text{True Positive Rate (TPR)} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{5.6}$$

Here,

True Positive = 16

False Negative = 12

So, True Positive Rate (TPR) $= \frac{16}{16+12} = 0.571$

$$\text{False Positive Rate (FPR)} = \frac{\text{False Positive}}{\text{True Negative + False Positive}} \quad (5.7)$$

Here,

False Positive = 1

True Negative = 83

So, False Positive Rate (FPR) $= \frac{1}{83+1} = 0.011$

Emotion Detection

The accuracy for emotion detection is defined as follows

$$\textbf{Accuracy} = \frac{\text{Total number of actual predection}}{\text{Total number of trial for a particular emotion}} \times 100\% \quad (5.8)$$

Figure 5.9 describes the emotion detection accuracy of our intelligent system.

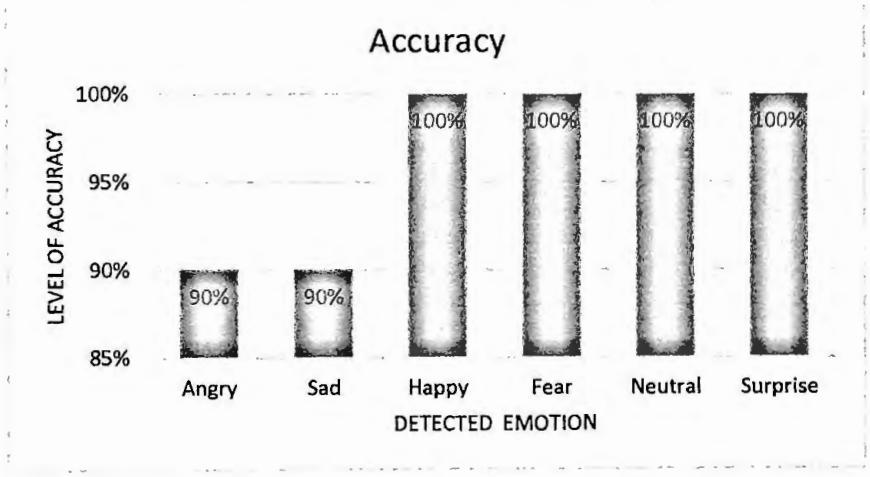


Figure 5.9: Emotion Detection Accuracy.

5.1.4 Comparison with Existing Works

A comparison of our proposed model with the recent existing work for suspicious behavior in exam hall detection has been given in the following table-

Table 5.13: Comparative performance to detect suspicious behavior in exam hall

	SL		1	2	3	4	5	6
	Year		2015	2018	2013	2012	2020	Proposed
	Method		binary SVM classifier	Canny edge detection + Speeded Up Robust Feature (SURF)	Viola-Jones object detection + HAAR Feature Extraction	Gabor feature extraction +ANN	CNN	**VFIG + modified Canny Edge Detection + LittleVGG**
VFOA detection	Eye movement	Sustained	no	no	no	no	no	yes
		Transients	no	no	no	no	no	yes (65%)
	Head Movement	Sustained	no	no	yes (90%)	yes (85%)	no	yes (100%)
		Transients	no	no	no	no	no	yes (75%)
Body Movement	Slow		no	yes (100%)	no	yes (90%)	no	yes (100%)
	Fast		no		no		no	yes (55%)
Suspicious Candidate Detection	No face/ multiface		no	yes (100%)	yes (100%)	yes (100%)	yes (100%)	yes (100%)
	Age		no	no	no	no	no	yes (avg SD 12.5)
	Gender		no	no	no	no	no	yes
	speech Detection		yes (89%)	no	no	no	no	no
	Object exchange /suspicious motion Tracking		no	yes (85%)	yes (70%)	no	yes (90%)	no
Emotion Detection			no	no	no	no	no	yes (80%)
Adaptive Load Control			no	no	no	no	no	**yes** (reduces the network load by 35%)

5.1.5 Concluding Remarks on Suspicious Behavior in Exam Hall Experiments

From the results derived from experiments we can see that for suspects' detection by means of gaze and body movement, we get optimal result. Age, gender and emotion detection is a new concept for proxy candidate tracking, which needs to be more accurate.

5.2 Experiments to Detect Liar while Interrogation

Several experiments were done on the sampled population to detect lie while real interrogation. Step by step procedures has been discussed below

5.2.1 Data Collection and Experiments for Lie Detection

People of different ages show different levels of body language while telling lies. To validate it, we collected data with different age groups. There was a total of 20 participants. We divided them according to their age into four consecutive groups as follows:

Group-1 (10 to 14 years old): Four members with the age 11, 11, 12 and 13 years, respectively.

Group-2(15 to 19 years old): Four members with the age 16, 16, 18 and 19 years, respectively.

Group-3(20 to 24 years old): Four members with the age 23, 24, 24 and 24 years, respectively.

Group-4(25 to 29 years old): Four members with the age 25, 26, 27 and 28 years, respectively.

There were 3 unpaid participants (2 male, 1 female) who were students at Bangladesh Army University of Engineering and Technology (mean age 20 years and standard deviation 0.775 years.). They were interrogated with the same questionnaire and the external symptoms (rotation of head, eye and change of eyebrow) were monitored very carefully.

To validate the lie detection in terms of eye rotation detection with our system, the participants were asked to look within the LNPFV and RNPFV with keeping their head in stable state that means in CFV with varying duration. The illumination of the room was 200 Lux and the distance from the camera was 0.5 meter. After that, to track the head rotation, they were asked to rotate their head with varying rotation time i.e. 1, 2 and 3 seconds respectively. They were also asked to move their eyebrow (shrink or expand it) so that it can be tracked by

the optical flow feature. The average video duration is 3 minutes with average12 frames per second.

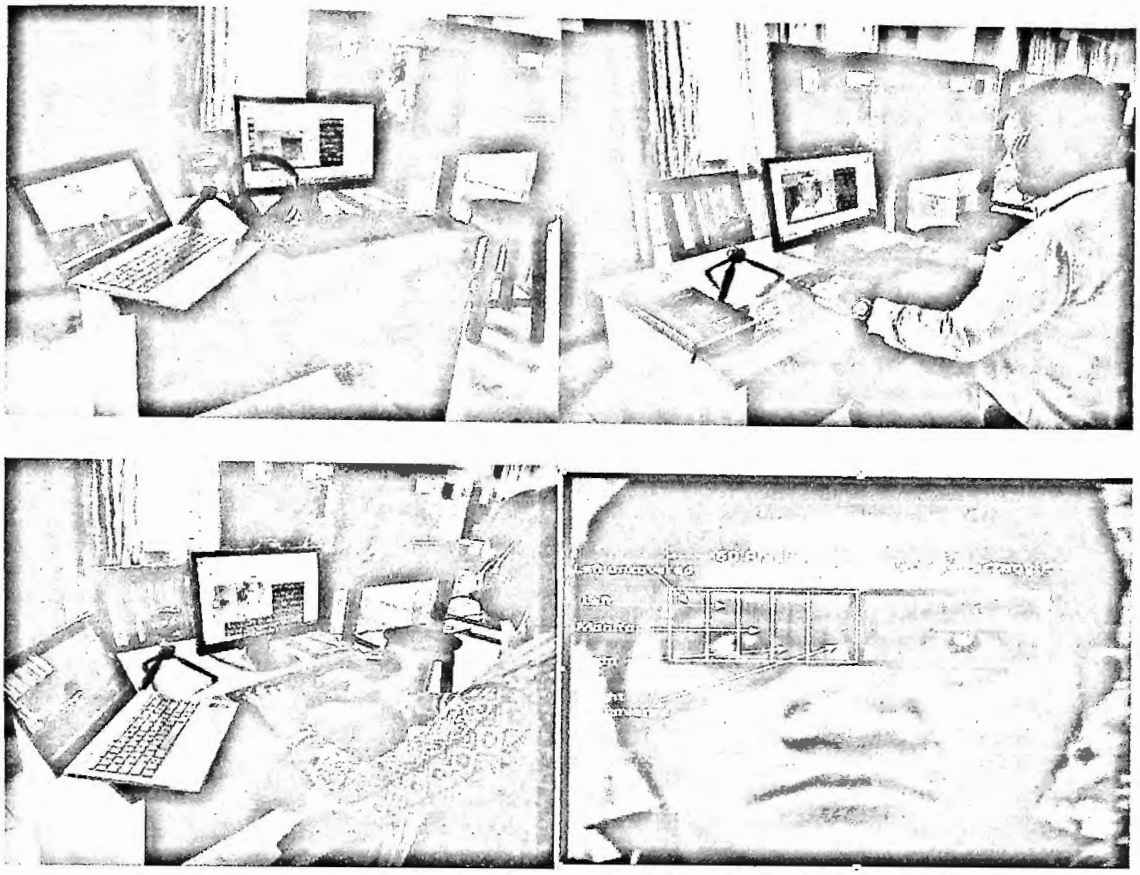


Figure 5.10: Experimental Setup of interrogation system.

5.2.2 Performance Evaluation of Lie Detection

The participants were interrogated with the same questionnaire (priorly unknown to them) and instructed them to tell lies instantly. The resulting symptoms were tracked very carefully. The summarized results of age variant symptom analysis is given in Table 5.13.

Table 5.14: Age variant symptom analysis

	Head Rotation	Eye Rotation
Group 1	3	2
Group2	3	1
Group3	2	2
Group4	1	0

We conducted several experiments to detect the level of questionnaire (Phase) at which the suspect shows external symptoms. We collected data as in Table 5.14.

Table 5.15: External Symptoms Detection.

	Suspect 1	Suspect 2	Suspect 3
Phase 1			
Phase 2	1	1	0
Phase 3			1
Phase 4			
Phase 5	2	1	2
Phase 6	1	1	1
Phase 7	2	1	1

To trace the relationship between Crime Type-1 and 2 with emotion, we conducted two separate experiments in the controlled environment. We set the threshold value of sustained symptoms to be greater than 3 seconds. Less than 3 seconds will be considered as transient symptoms. The summary of the collected data is given in Table 5.15.

Table 5.16: Crime Type 1 and 2 detections.

		Suspect 1	Suspect 2	Suspect 3
EXP 1	Transient	1	0	1
	Sustained	2	1	2
EXP 2	Transient	0	1	1
	Sustained	1	1	0

In our lie detection technique, we detected Symptom Type-1 (st1) based on random fluctuation of eye movement. The Accuracy of detection while interrogation is expressed as-

$$\textbf{Accuracy}_{st1} = \frac{\text{Total number of eye movement tracked}}{\text{Total number of eye movement occurred}} \times 100\% \qquad (5.9)$$

Based on equation (5.9), we get the Figure 5.11 which shows the eye rotation detection accuracy for varying eye rotation time. Similarly, Symptom Type-2 is tracked based on random head movement or change in the orientation of head. The accuracy is expressed as:

$$\textbf{Accuracy}_{st2} = \frac{\text{Total number of head movement tracked}}{\text{Total number of head movement occurred}} \times 100\% \qquad (5.10)$$

Based on equation (5.10), we get the Figure 5.12 which shows the head rotation detection accuracy for varying head rotation time. Symptom Type-3 is tracked with the change in the shape of the eyebrow by implementing Optical Flow Feature in the ROI. The detection accuracy is expressed as:

$$\text{Accuracy}_{st2} = \frac{\text{Total number of eyebrow movement tracked}}{\text{Total number of eyebrow movement occurred}} \times 100\% \quad (5.11)$$

The accuracy of Symptom Type-3 has been illustrated in Figure 5.13.

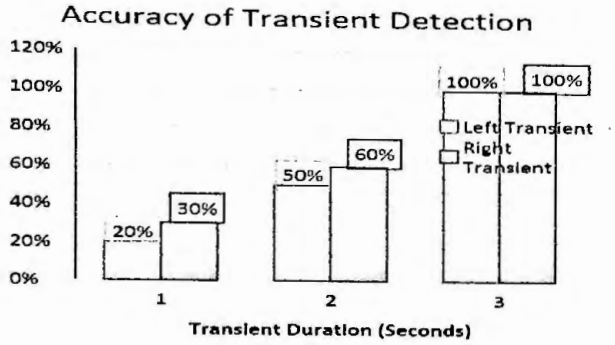


Figure 5.11: Accuracy Measurement of Symptom Type -1 Detection.

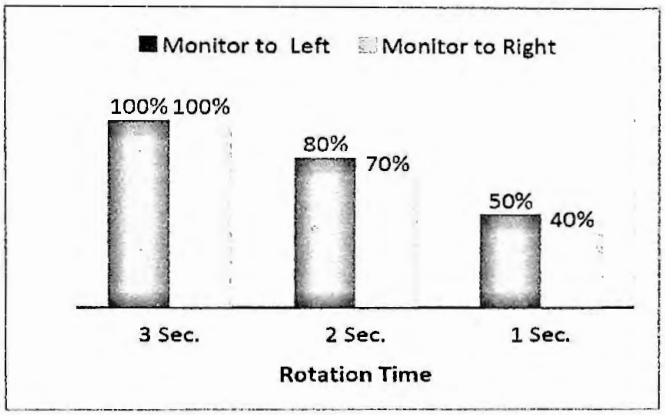


Figure 5.12: Accuracy Measurement of Symptom Type -2 Detection

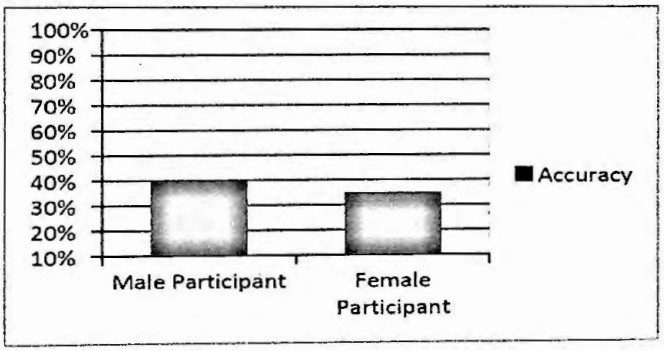


Figure 5.13: Accuracy Measurement of Symptom Type -3 Detection.

5.2.3 Comparison with Existing Works

A comparison of our proposed model with the recent existing work for suspicious behavior in interrogation room has been given in the following table-

Table 5.17: Comparative performance to detect liar

	SL		1	2	3	4	5	6
	Year		2019	2014	2012	2013	2018	Proposed
	Proposed Method		Blood Flow Pattern Analysis	Eulerian Video Magnification (EVM) technique	Facial Micro-Expressions	electroence phalogram (EEG) signals analysis by F-Score and Extreme Learning Machine	Fuzzy reasoning approach	VFIG + Optical Flow
VFOA detection	Eye movement	Sustained	no	no	yes	no	no	yes
		Transients	no	no	yes	no	no	yes (65%)
	Head Movement	Sustained	no	yes	yes	no	no	yes (100%)
		Transients	no	yes	yes	no	no	yes (75%)
Fast eyebrow movement			yes	yes	no	no	no	yes (35%)
Leap			yes	yes	no	no	no	no
Frequent swallowing			yes	yes	no	no	no	no
Blink rate			yes	yes	no	no		no
Brain Signal			no	no	no	yes	yes	no
Sensors			Thermal Camera	FHD camera	FHD camera	electroence phalogram	FHD camera	very low cost camera
Real time			no	no	no	no	no	yes
Accuracy			70%	80%	85%	90%	90%	**60%**

5.2.4 Concluding Remarks on Lie Detection Experiments

Symptom Type -1 and Symptom Type -2 have been detected with optimal accuracy. But for Symptom Type -3 detection, the accuracy is not optimal.

5.3 Chapter Summary

We see that, at first, psychological data is collected from various sources concerned to social environments to have a robust insight about the crime patterns. After that results is gathered and experiments have been done. Finally, we see that, our system has been utilized fully to track the patterns with fair accuracy.

Chapter 6
Conclusion

In this chapter, we will discuss the research summary extracted from different experimental results. Different experiences related to the implementation of our research works, limitations and a short description of our future work have also been included in this chapter.

6.1 Research Summary

- At first, we created a knowledgebase gleaned from manual experiments to define different parameters of suspicious behavior for real-time interrogation and examination room. We tried to track the head and eye center from the captured continuous image with variable distances from the camera and in different lighting conditions. When the distance from the camera is between 60cm to 80cm, we get the best tracking performance because we need to ensure a certain distance from the camera to the target object to maintain the proper focusing. Moreover, a minimum of 200 Lux of illumination must be ensured to separate the regions of interest (the head and the body) from the background.
- For gaze detection from the coordinate of a pupil, we applied different techniques. However, "Unequal partitioning of Eye Rectangle Leaving the Uncovered Regions" produces the best result in gaze tracking. In gaze detection combining with the head pose, we conducted different experiments with variable head rotation time. When the head rotation time is 3 seconds, it shows the best performance in gaze tracking.
- With our proposed technique, our system can detect examinees' random movement with the help of a Canny edge detector. Our intelligent system provides 100% accuracy when the examinee takes 3, 4, and 5 seconds for a proper movement.
- It is seen that the accuracy of detection depends on the elapsed time. This is because our low-cost camera needs some time for proper focusing if the object is in motion. Without proper focusing, the captured frames result in improper data which degrades the overall performance.
- Our intelligent system provides better detection accuracy for 21 to 35 and 36 to 50 age groups in real environments. But we have obtained a degraded performance for the age group 8-20 and 51-65.

- While performing tests for emotion detection, our intelligent system can detect happiness, fear, neutral, and surprise perfectly and provide 100% accuracy. For the rest (angry, sad) the detection accuracy is 90%.

- For lie detection, from the experimental results, it can be concluded that the accuracy of Symptom Type-1 and 2 is dependent on transient duration. And the accuracy increases linearly with the duration. However, Symptom Type-3 detection accuracy depends on the gender of the suspect. The reason is that the width of the female suspect's eyebrow is less than that of the male suspect. So, it becomes harder to track the change in the shape of eyebrows using optical flow features. We conducted our experiments in a controlled environment which may not match with the results of real-life environments. We leave this for our future research.

6.2 Limitations

We wanted to build up a unique system merged with VFOA and body movement, age, gender, emotion estimation of the target examinee and suspects. But we see that VFOA detection is illumination depended and a certain distance must be ensured between camera and object for proper focusing which makes the system harder to implement in real life situation. It was very difficult for us to detect examinees' random body movements with different dynamics if there is a lot of objects in the background.

Our smart system can detect examinees' age and gender with the help of a neural network classifier trained by IMDB-WIKI and UTKFace datasets to identify fraud candidates in the examination hall. The fact is that a neural network requires a huge data set to be trained (in our case 28,000 data) properly. The collection of this huge dataset for the native Bangladeshi people is a tremendously hard job. So, we have trained the network with images mostly taken from foreigners and tested the effectiveness of the network with the locally taken images. So the performance is not up to the mark in real environment of Bangladesh.

To detect the human emotion and lie of the suspect while interrogation, the problem we faced is that sometimes it becomes harder to distinguish among different emotion levels, even for a well-trained network because human emotion expression is not always well defined and it varies from man to man depending on his age, culture, customs, education, social status, economic backgrounds. In the future, we would like to conduct our experiments in real environments.

6.3 Future Improvements

In the near future some of the other experiments will be done. Some of them may be:

- Developing a new technique for edge detection with more accuracy of detection.
- Developing a technique that can track the gaze with optimum accuracy with the head rotation time for 1 or 2 seconds.
- Developing a unique system that can detect all human emotions with 100% accuracy.
- Developing an intelligent system that can detect our local people properly.
- Developing a system that can track the head and eye in low illumination.
- Apply our proposed techniques in a real-life environment and increase the real-time suspicious behavior detection performance.

6.4 Summary

This thesis focuses on creating a smart system for suspicious behavior detection in the field of Artificial Intelligence based on the suspect's eye gaze detection, head detection, and random movement of body pixel detection. The system also considers other parameters like age, gender, and emotion tracking and an advanced part of the system is capable of detecting lies in the interrogation room. Due to the Covid-19 pandemic situation, we couldn't fulfill our desire and didn't get optimal accuracy in all respects. But we are hopeful to increase the accuracy in future.

References

[1] Zhang, H., Smith, M. R., & Witt, G. J. (2006). Identification of real-time diagnostic measures of visual distraction with an automatic eye-tracking system. *Human factors*, *48*(4), pp. 805-821.

[2] Judd, T., Ehinger, K., Durand, F., & Torralba, A. (2009). Learning to predict where humans look. In *2009 IEEE 12th international conference on computer vision, IEEE,* pp. 2106-2113.

[3] Goferman, S., Zelnik-Manor, L., & Tal, A. (2011). Context-aware saliency detection. *IEEE transactions on pattern analysis and machine intelligence*, *34*(10), pp. 1915-1926.

[4] Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature reviews neuroscience*, *2*(3), pp. 194-203.

[5] Khandagale, P., Chaudhari, A., & Ranawade, A. (2013). Automated Video Surveillance to detect suspicious Human Activity.

[6] Gowsikhaa, D., & Abirami, S. (2012). Suspicious Human Activity Detection from Surveillance Videos. *International Journal on Internet & Distributed Computing Systems*, *2*(2).

[7] P. Deraiya, J.Pandya. (2016) "A Survey for Abnormal Activity Detection in Classroom," *International Journal of Innovative Research in Science, Engineering and Technology*, *5*(12), pp. 20511– 20515.

[8] Asteriadis, S., Nikolaidis, N., Hajdu, A., & Pitas, I. (2006). An eye detection algorithm using pixel to edge information. In *Int. Symp. on Control, Commun. and Sign. Proc.*

[9] Kim, S., Chung, S. T., Jung, S., Oh, D., Kim, J., & Cho, S. (2007). Multi-scale gabor feature based eye localization. *World Academy of Science, Engineering and Technology*, *21*, pp. 483-487.

[10] Mustafa, R., Min, Y., & Zhu, D. (2014). Obscenity detection using Haar-like features and gentle adaboost classifier. *The Scientific World Journal*,

[11] Zhao-Yi, P., Yan-Hui, Z., & Yu, Z. (2010). Real-time facial expression recognition based on adaptive canny operator edge detection. In *2010 Second International Conference on MultiMedia and Information Technology, IEEE.* (2), pp. 154-157.

[12] Mei, Y., & Yu, J. (2010). An Algorithm for Automatic Extraction of Moving Object in the Image Guidance. In *2010 International Conference on Intelligent System Design and Engineering Application, IEEE.* (1), pp. 226-230.

[13] Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6), pp. 679-698.

[14] Xin, G., Ke, C., & Xiaoguang, H. (2012). An improved Canny edge detection algorithm for color image. In *IEEE 10th International Conference on Industrial Informatics IEEE.* pp. 113-117

[15] Anandhi, M., Josephine, M. S., Jeyabalaraja, V., & Satthiyaraj, S. COMPARISON OF CANNY AND SOBEL EDGE IN DETECTION TECHNIQUES. International Journal of Engineering Sciences & Research Technology

[16] Dans, P. E. (1996). Self-reported cheating by students at one medical school. *Academic Medicine*, *71*(1), S70-2.

[17] X. Lan, S. Zhang, P. C. Yuen, and R. Chellappa. (2018). "Learning common and feature-specific patterns: a novel multiplesparse-representation-based tracker," *IEEE Transactions on Image Processing*, *27*(4), pp. 2022–2037.

[18] Y. Qi, S. Zhang, L. Qin et al. (2019). "Hedging deep features for visual tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(5), pp. 1116–1130.

[19] S. Zhang, Y. Qi, F. Jiang, X. Lan, P. C. Yuen, and H. Zhou. (2018). "Point-to-set distance metric learning on deep representations for visual tracking," *IEEE Transactions on Intelligent Transportation Systems, 19* (1), pp. 187–198.

[20] H. Zhang, J. Bai, Z. Li, Y. Liu, and K. Liu, (2017). "Scale invariant SURF detector and automatic clustering segmentation for infrared small targets detection," *Infrared Physics & Technology, (83),* pp. 7–16.

[21] M. ElMikaty and T. Stathaki, (2017). "Detection of cars in highresolution aerial images of complex urban environments," *IEEE Transactions on Geoscience and Remote Sensing*, *55*(10), pp. 5913–5924.

[22] H. Zheng, J. F. Liu, J. L. Gao, and Q. Lu, (2017) "Feature detection method for small targets of complex multimedia images in cloud environment," *Multimedia Tools and Applications*, *76*(16), pp. 17095–17112.

[23] U. Asif, M. Bennamoun, and F. Sohel, (2016). "Unsupervised segmentation of unknown objects in complex environments," *Autonomous Robots, 40*(5), pp. 805–829,

[24] D. Luo and Y. Li, (2015) "Multi-stage and multi-attribute risk group decision-making method based on grey information," *Grey Systems: Theory and Application, 5*(2), pp. 222–233, 2015.

[25] W. Matin, Z. Nasim, N. Ahmed, K. Zannat, and F. Muhammad, (2015), "Artificial neural network based system for intrusion detection using clustering on different feature selection," *International Journal of Computer Applications, 126*(12), pp. 21–28,.

[26] G. Bhartiya and A. S. Jalal, (2017) "Forgery detection using feature clustering in recompressed JPEG images," *Multimedia Tools and Applications, 76*(20), pp. 20799–20814.

[27] K. Huang, Y. Zhang, B. Lv, and Y. Shi, (2017). "Salient object detection based on background feature clustering," *Advances in Multimedia*, vol. 2017, Article ID 4183986,

[28] A. Iosifidis, A. Tefas, and I. Pitas, (2015) "Class-specific reference discriminant analysis with application in human behavior analysis," *IEEE Transactions on Human-Machine Systems, 45*(3), pp. 315–326.

[29] B. Xiao, P. Georgiou, B. Baucom, and S. S. Narayanan, (2015) "Head motion modeling for human behavior analysis in dyadic interaction," *IEEE Transactions on Multimedia, 17*(7), pp. 1107–1119.

[30] J. Pan, H. Hu, X. Liu, and Y. Hu, (2016). "Multiscale entropy analysis on human operating behavior," *Entropy, 18*(1), pp. 3.

[31] S. Yi, H. Li, and X. Wang, (2016). "Pedestrian behavior modeling from stationary crowds with applications to intelligent surveillance," *IEEE Transactions on Image Processing, 25*(9), pp. 4354–4368.

[32] J. Hu, L. Xu, X. He, and W. Meng, (2017). "Abnormal driving detection based on normalized driving behavior," *IEEE Transactions on Vehicular Technology, 66*(8), pp. 6645–6652.

[33] Y. Yuan, J. Fang, and Q. Wang, (2015). "Online anomaly detection in crowd scenes via structure analysis," *IEEE Transactions on Cybernetics, 45*(3), pp. 548–561.

[34] Y. Yuan, D. Wang, and Q. Wang, (2017). "Anomaly detection in traffic scenes via spatial-aware motion reconstruction," *IEEE Transactions on Intelligent Transportation Systems*, *18*(5), pp. 1198–1209.

[35] Y. Yu, L. Cao, E. A. Rundensteiner, and Q. Wang, (2017) "Outlier detection over massive-scale trajectory streams," *ACM Transactions on Database Systems*, *42*(2), pp. 1–33.

[36] Koch, C., & Ullman, S. (1987). Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of intelligence, Springer, Dordrecht,* pp. 115-141.

[37] Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, *20*(11), pp. 1254-1259.

[38] Hou, X., & Zhang, L. (2007). Saliency detection: A spectral residual approach. In *Proc. IEEE Conference on computer vision and pattern recognition,* pp. 1-8.

[39] Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. In *Advances in neural information processing systems*, pp. 545-552.

[40] Xu, F., Xu, B., Ding, H., & Qian, W. (2000). Virtual automobile driver training simulator. *International journal of virtual reality. 4*(4).

[41] Cai, H., Lin, Y., & Mourant, R. R. (2007). Evaluation of driver visual behavior and road signs in virtual environment. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *51*(27), pp. 1645-1649.

[42] Zhang, H., Smith, M. R., & Witt, G. J. (2006). Identification of real-time diagnostic measures of visual distraction with an automatic eye-tracking system. *Human factors*, *48*(4), pp. 805-821.

[43] Hansen, D. W., & Hansen, J. P. (2006). Eye typing with common cameras. *In Proceedings of the 2006 symposium on Eye tracking research & applications*, pp. 55-55.

[44] Murphy-Chutorian, E., & Trivedi, M. M. (2008). Head pose estimation in computer vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, *31*(4), pp. 607-626.

[45] Kim, S., Chung, S. T., Jung, S., Oh, D., Kim, J., & Cho, S. (2007). Multi-Scale Gabor Feature Based Eye Localization. *World Academy of Science, Engineering and Technology*, 21, pp. 483-487.

[46] Kroon, B., Hanjalic, A., & Maas, S. M. (2008). Eye Localization for Face Matching: Is it Always Useful and Under What Conditions? In *Proc. 2008 international conference on Content-based image and video retrieval,* pp. 379-388.

[47] Niu, Z., Shan, S., Yan, S., Chen, X., & Gao, W. (2006). 2d Cascaded Adaboost for Eye Localization. In *Proc. 18th International Conference on Pattern Recognition, IEEE,* pp. 1216-1219.

[48] Valenti, R., & Gevers, T. (2008). Accurate Eye Center Location and Tracking Using Isophote Curvature. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* pp. 1-8.

[49] Lawrence S., Giles C. L., Tsoi A. C., and Back A. D. (1997). Face recognition: A convolutional neural network approach. *IEEE Transactions on Neural Networks, (8),* pp. 98-113.

[50] Rowley H. A., Baluja S., and Kanade T. (1998). Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine intelligence, (*20), pp. 23-38.

[51] Chang K.I., Bowyer K.W., and Flynn P. J. (2003). Face recognition using 2d and 3d facial data. *ACM Workshop on Multimodal User Authentication,* pp. 25-32.

[52] Kalansuriya, T. R., & Dharmaratne, A. T. (2014). Neural network based age and gender classification for facial images. *ICTer, 7*(2).

[53] Raghuvanshi, A., & Choksi, V. (2016). Facial expression recognition with convolutional neural networks. *CS231n Course Projects*, pp. *362.*

[54] M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, (2006). Fully automatic facial action recognition in spontaneous behavior. In *Proc. IEEE Conference on Automatic Facial and Gesture Recognition.*

[55] M. Pantic and J.M. Rothkrantz, (2004) Facial action recognition for facial expression analysis from static face images. *IEEE Transactions on Systems, Man and Cybernetics, 34*(3).

[56] M.S. Bartlett, G. Littlewort, M.G. Frank, C. Lainscsek, I. Fasel, and J.R. Movellan. (2006). Automatic recognition of facial actions in spontaneous expressions. *Journal of Multimedia.*

[57] Zhiding Yu and Cha Zhang, (2015). Image based static facial expression recognition with multiple deep network learning. In *Proc. 2015 ACM on International Conference on Multimodal Interaction*, pp.435–442.

[58] Song, Y., Guo, Y., & Thuente, D. (2016). A quantitative case study on students' strategy for using authorized cheat-sheets. In *2016 IEEE Frontiers in Education Conference (FIE)*, pp. 1-9..

[59] Kaiiali, M., Ozkaya, A., Altun, H., Haddad, H., & Alier, M. (2016). Designing a secure exam management system (SEMS) for M-learning environments. *IEEE Transactions on Learning Technologies*, *9*(3), 258-271.

[60] N. Mathew, (2016). Student preferences and performance: A comparison of open-book, closed book, and cheat sheet exam types. In *Proc. The National Conference on Undergraduate Research (NCUR)*, pp. 1243-1260.

[61] N. Soman, M. R. Devi, G. Srinivasa, (2017). Detection of anomalous behavior in an examination hall towards automated proctoring. In *Proc. Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, pp. 1-6.

[62] T. Ujbanyi, J. Katona, G. Sziladi, A. Kovari, (2016). Eye-tracking analysis of computer networks exam question besides different skilled groups. *In: Proc. IEEE International Conference on Cognitive Info communications (CogInfoCom)*. pp. 000277-000282.

[63] Y. Atoum, L. Chen, A.X. Liu, S. D. Hsu, X. Liu, (2017). Automated online exam proctoring. *IEEE Transactions on Multimedia. (19)*, pp. 1609-1624.

[64] G. Nandini. B. Mathivanan, R. S Nantha Bala, P Poornima (2018). Suspicious human activity detection. *International Journal of Advance Research and Development*, 3(4), pp.12–14.

[65] P. Khandagale, A. Chaudhari , A. Ranawade , P.M.Mainkar (2013), Automated Video Surveillance to Detect Suspicious Human Activity, *International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)*, 4(2), pp.170- 173.

[66] D. Gowsikhaa, S. Manjunath Abirami (2012), Suspicious Human Activity Detection from Surveillance Videos, *International Journal on Internet and Distributed Computing Systems*. 2(2), pp. 141-148.

[67] N. Tanjila, A. Saha, K. Akter, S. Ahemd, (2020), A Proposed Architecture to Suspect and Trace Criminal Activity Using Surveillance Cameras, *Proc. 2020 IEEE Region 10 Symposium (TENSYMP),* pp. 100-102.

[68] J. E. Reid, F. E. Inbau, Williams and Wilkins Co. & United States of America, TRUTH AND DECEPTION-THE POLYGRAPH ('LIE-DETECTOR') TECHNIQUE, 2D ED (1977).

[69]. M. R. Bhutta, M. J. Hong, Y. H. Kim & Hong, K. S (2015), Single-trial lie detection using a combined fNIRS-polygraph system, *Frontiers in Psychology*, *6*, pp. 709.

[70]. National Research Council(2003).. The polygraph and lie detection. *National Academies Press.*

[71]. M. Owayjan, A. Kashour, N. Al Haddad, M. Fadel & G. Al Souki, The design and development of a lie detection system using facial micro-expressions, (2012), *Proc. 2012 2nd international conference on advances in computational tools for engineering applications (ACTEA)*, pp. 33-38.

[72]. J. Gao, Z. Wang, Y. Yang, W. Zhang, C. Tao, J. Guan & N. Rao, (2013), A novel approach for lie detection based on F-score and extreme learning machine. *PloS one*, *8*(6), e64704.

[73]. Y. F. Lai, M. Y. Chen & H. S. Chiang, (2018). Constructing the lie detection system with fuzzy reasoning approach, *Granular Computing*, 3(2), 169-176 doi:10.1007/s41066-017-0064-3

[74] David Cornish, M., & Dukette, D. (2009). The essential 20: *twenty components of an excellent Health Care Team.* Dorrance Publishing.

[75] Mateo, J. C., San Agustin, J., & Hansen, J. P. (2008). Gaze Beats Mouse: Hands- Free Selection by Combining Gaze and Emg. In. *CHI'08 extended abstracts on Human factors in computing systems,* pp. 3039-3044.

[76] Asteriadis, S., Nikolaidis, N., Hajdu, A., & Pitas, I. (2006). An Eye Detection Algorithm Using Pixel to Edge Information. *In Int. Symp. On Control, Commun. and Sign. Proc.*

[77] Ba, S., & Odobez, J. M. (2007). From Camera Head Pose to 3D G;obal Roomhead Pose Using Multiple Camera Views. *In Proc. Int'l. Workshop Classification of Events Activities and Relationships.*

[78] An, K. H., & Chung. M. J. (2008). 3D Head Tracking and Pose-Robust 2d Texture Map-based Face Recognition Using A Simple Ellipsoid Model. In *Proc IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 307-312.

[79] Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active Shape Models-Their Training and Application. *Computer vision and image understanding*, *61*(1), pp. 38-59.

[80] Liu, X., & Chen, T. (2005). Pose-Robust Face Recognition Using Geometry Assisted Probabilistic Modeling. *In Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (1)*, pp.502-509.

[81] La Cascia, M., & Sclaroff, S. (1999). Fast, Reliable Head Tracking Under Varying Illumination. In, *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (1).*

[82] Horn, B. K., & Schunck, B. G. (1981). Determining optical flow. *Artificial intelligence*, *17*(1-3), pp. 185-203.

[83] Bansal, M., Kumar, M., Kumar, M., & Kumar, K. (2021). An efficient technique for object recognition using Shi-Tomasi corner detection algorithm. *Soft Computing*, *25*(6), pp. 4423-4432.

[84] Sheppard, K. (2012). Introduction to Python for econometrics, statistics and data analysis. *Self-published, University of Oxford*, version, 2.

[85] Snihovyi, O., Ivanov, O., & Kobets, V. (2018). Cryptocurrencies prices forecasting with anaconda tool using machine learning techniques. *In CEUR Workshop Proceedings,* pp. 453-456.

[86] Bradski, G., & Kaehler, A. (2008). Learning OpenCV: Computer vision with the OpenCV library. " O'Reilly Media, Inc.".

[87] Kaehler, A., & Bradski, G. (2016). Learning OpenCV 3: computer vision in C++ with the OpenCV library. " O'Reilly Media, Inc.".

Appendix I

List of Published Paper

Journal:

1. Partha Pratim Debnath, Md. Golam Rashed, Dipankar Das and Md. Rubel Basar, **"Smart Interrogation System by Detection of Visual Focus of Attention,"** *Rajshahi University Journal of Scientific Research, Volume 13, Issue 1*, 2021, pp. 47-58.

2. Partha Pratim Debnath, Md. Golam Rashed, Dipankar Das, Mehedi Hasan Imran, Mitaly Roy, **"Suspicious Behavior Detection in Smart Class Room: Computer Vision Approach,"** *BAUET Journal, Volume 02, Issue 02,* 2020, pp. 255-267.

3. Partha Pratim Debnath, Md. Golam Rashed, Dipankar Das, **"Detection and Controlling of Suspicious Behavior in the Examination Hall,"** *International Journal Of Scientific and Engineering Research(IJSER), Volume 09, Issue 7, 2018, ISSN 2229-5518,* pp. 1227-1232.

Conference:

1. Partha Pratim Debnath, Md. Golam Rashed, Dipankar Das, **"Smart Interrogation System: Computer Vision Approach,"** In Proc. *7th International Conference on Engineering Research, Innovation and Education (ICERIE), SUST, Sylhet, Bangladesh,* January, 2021, pp. 791-797.

2. Partha Pratim Debnath, Md. Golam Rashed, Dipankar Das, **"Automatic Examination Hall Invigilation System to Detect Examinees Suspicious Behaviour,"** In Proc. *5th International Conference on Engineering Research, Innovation and Education (ICERIE), SUST, Sylhet, Bangladesh,* January, 2019, pp. 416-421.