GAZE TRACKING FOR EMOTIONAL DETECION

Egyetemi doktori (PhD) értekezés

a szerző neve:Alhamzawi Hussein Ali Mezher
témavezető neve: Attila Fazekas

DEBRECENI EGYETEM
Természettudományi és Informatikai Doktori Tanács
Informatikai Tudományok Doktori Iskola
Debrecen, 2020.

<<GAZE TRACKING FOR EMOTIONAL DETECTION>>

Dissertation submitted in partial fulfillment of the requirements for the doctoral (PhD) degree in in the scientific field of information technology

Written by: Alhamzawi Hussein Ali Mezher certified: Master of Science in Engineering.

In the framework of the informatics doctoral school of the University of Debrecen (Discrete mathematical, data processing and visualization program)

Dissertation advisor: Dr. Attila Fazekas

The comprehensive examination board:

Chairperson:	Dr. Vaszil György
Members:	Dr. Nagy Antal
	Dr. Adamkó Attila

The date of the comprehensive examination: 30/07/2018

The official opponents of the dissertation:

Dr. ……
Dr. ……
Dr. ……

The evaluation committee:

Chairperson:	Dr. ……
Members:	Dr. ……
	Dr. ……
	Dr. ……
	Dr. ……

The date of the dissertation defence: …… 20…

DECLARATION

Hereby I declare that I prepared this thesis within the Doctoral Council of Natural Sciences and Information Technology, Doctoral School of informatics, University of Debrecen in order to obtain a PhD Degree in Informatics at Debrecen University.

The results published in the thesis are not reported in any other PhD theses.

Debrecen, 2020 May 06. *Signature of the candidate*

Hereby I confirm that Alhamzawi Hussein Ali Mezher candidate conducted his/her studies with my supervision within the Discrete mathematical ,data processing and visualization Doctoral Program of the Doctoral School of informatics between 2015 and 2020 The independent studies and research work of the candidate significantly contributed to the results published in the thesis.

I also declare that the results published in the thesis are not reported in any other theses.

I support the acceptance of the thesis.

Contents

LIST OF FIGURES:

LIST OF TABLES:

Acknowledgements

Initially, I might want to offer my genuine thanks to my counselor Attila Fazekas for the consistent help of my Ph.D study and related research, for his understanding, inspiration, and gigantic information. His direction helped me in all the hour of research and composing of this postulation. I was unable to have envisioned having a superior counsel and tutor for my Ph.D study.

Besides my counselors, I might want to thank the remainder of my dissertation committee: Dr. Nagy Antal, Dr. Vaszil György, and Dr. Adamkó Attila, for their savvy remarks and consolation, yet additionally for the hard inquiry which incented me to broaden my exploration from different points of view.

And also I want to give big thanks for team of Wipro company (Hungary branch) that’s they support me and help me a lot,

Last yet not the least, I might want to thank my family: my parents and to my siblings and my lovely sister for supporting me profoundly all through composing this proposal and my life by and large.

INTRODUCTION

Emotional problems represent the human experience, including good or bad feelings, interpersonal interactions, mechanisms and intervention results. Human-computer Interaction (HCI) seeks to enhance user-experiences by enhancing the perception of the user's interests, including the user's emotional condition [1].Within the area of affective computing; various researchers have explored emotional aspects and methods of identifying, assessing and creating applications based on them. Those involve studies in calculating and predicting feelings, how individuals can effectively interact and whether they can change their lives. Appreciation of individual feelings is an integral aspect of learning and development. The moral perception of human nature, studies on mental and emotional health, analysis of social interaction, etc., is a pertinent topic. The function of the autonomous nervous system (ANS) is compounded, e.g., physiological amplitude, electrodermal function, controls the pupillary size and eye expressions with the subject's behaviors, and sentimental status. Such computerized emotional understanding can also aid in psychological research [2]. For example, such findings may have the foundation of a stable subject's emotional behavior, comparison with the others and help in identifying some psychological conditions such as melancholy.

Emotions are a crucial component in social experiences because they

empower individuals to comprehend the interests of others correctly and to encourage possible answers. Therefore, it is very necessary to understand the instincts of the others to build a relationship and mutual emotions. Accurate detection and analysis were used to support people to select whether they make socially appropriate comments and direct the choice of behavioral solutions or retirement strategies. Earlier experiments demonstrated various identification techniques, perception and assessment of emotions across a variety of emotional responses. It has examined a mixture of optimistic and pessimistic feelings. For instance, anxiety, sorrow, and joy are taking into consideration as the three top emotional conditions. In the recent studies laughter, anger, grief, and satisfaction considered as the mental states. The only example was where the analysis concentrated solely on a specific anger emotion [3].

Human eye is one of the key parts of body which is not only used for the vision but also to express the emotions and to communicate. Any irritability or false move can distract the audience and change the meaning of what is being communicated. Since it has a strong association to communication and credibility of a person, therefore, it is important to adopt extra precautions while drawing the human eye [4]. The study of eye movement got the interest of researchers and theorists in 1908 [5] where the initial researches connected the human eye movement as a key source of navigating the social

environment through cues, signals and facial expressions [6]. The researchers further identified that monitoring the direction of sight (gaze) often leads to concentration of attention by the observer. This effect appears very early in life, for example, in newborns up to the age of 3 months [7].

Besides that, one of the important signs to display feelings and emotions is the gaze [8]. Gaze behavior refers to the window of social cognition [9] where social cognition determines the quality and depth of social interaction between individuals on the basis of sympathetic behavioral analysis. Therefore, an effective contact plays a very important role in determining the quality of social interaction. However, if an individual has any disorders or psychiatric illnesses such as autism, eye gaze patterns and eye contact patterns become disturbed. Therefore, such individuals are unable to demonstrate an impactful social interaction [10]. Considering the behavioral aspects, gaze signals the behavioral tendencies of the expresser. Mehrabian model of Social Intimacy had evidenced that increased eye contact is one of the key signs which promotes and enhances psychosocial convergence in adults [11]. For example, in infants, it provides a means to approach or avoid others, and organize their emotional experience [10] which indicates that the behavior of gaze is innately prepared [12]. This instinctive ability to process gaze plays a key role in the development of “theory of mind" i.e. “by nonverbal behavior there is the ability to infer behavioral intentions” [13].

The researchers found that showcasing an emotional communicatory gaze is very tough because, any minor error can easily be noticed by the viewers which might destroy the illusion of life. This challenge of accuracy is intensified in the case where with interactive virtual computer-controlled characters interact with each other and human users. Unlike hand animated characters, interactive virtual characters must function in the virtual environment and interact with human users in real time. This also includes gazing as an arbitrary goal as it appears in the environment. For example, the environments in modern video games, such us Oblivion or World of War craft have more than thousand computer controlled objects interacting with the users. Creating such characters requires a lot of time, investment and concentration. “Expressive Gaze Model (EGM)” is identified as one of the effective method enabling virtual characters within the game to produce some highly reliable visual transitions ensuring that they convey the desired emotional expression. Likewise the expressive theory model consists of a hierarchical framework within which some simple behaviors are formed in the emotional expressive outlook with virtual characters that represent the itch of the trunk, head and eye [14]. Few experiments have been carried out to examine eye sensitivity to emotional stimulation. Another research measures the size of pupils while taking pictures to determine results, revealing that pupils become bigger as they perceive subjective stimuli

friendly and unfavorable. There are several forms of eye control; free head movement work has been done, underlying emotional responses. The research examined eye movement in reaction to positive and negative stimuli, finding more eye twitching and more corrugators when displaying unpleasant photos [15]. Throughout this analysis, the blinded eye reaction and corrugator behavior reduced when negative feelings were blocked. A behavioral function is proposed to better assess individual emotional perception [16]. Thus, in the current analysis, the emotional condition is being recognized not just by the student scale, but also by the observer. Pupil scale, though, is a brain function predictor. Previous experiments have demonstrated that cognitive therapy and effective knowledge affect individual pupils. For example, as cognitive load decreased, the measurements of the pupils were observed to increase information regarding the position and length of the pupil may be collected to draw other assumptions about the neurological system. Throughout the assessment of cognitive or affective disorders, the implementation of eye control will rely on two hypotheses: immediacy (the knowledge on people activities as seen) and eye-mentality (the eye staying locked on an item through processing). In this research, topics analyze the sensations and divide them into three bipolar parameters: mental valence, anticipation, and superiority. The element of valence ranges from pessimistic to optimistic emotional interactions.

Throughout our analysis and since the Arabic clip collection was small to be tested; we were encouraged to select two of the Arabic cartons for topics. We should use this technique to test the emotional perception of the participant in the clips. The purpose of this research is to examine the eye monitoring and pupil size and fastening as a reaction to stimuli with specific arousal material film clips.

1 AMIS AND OBJEACTIVE

1.1 Dissertation aim

The purpose of this dissertation is to track the gaze for distinguishing emotions by the direction of the gaze with the help of other parts of the face. To achieve this goal, the eyesight detector was developed and constructed and tested for analysis of sections that were photographed by volunteers who were assigned a specific gaze task. Taking into consideration both distance and lighting is different to check the accuracy of the technique as well as to obtain and find the best and most important parameters that will be used later. Besides, the accuracy of the method was checked separately for each eye gaze direction in eight different directions. From the analysis, it can be reported that accurate contactless measurement of gaze tracking in a two-dimensional color video is possible in real-life conditions.

1.2 Dissertation objective

The essential objective of this dissertation is to structure, actualize and assess a novel eye-tracking emotion recognition framework utilizing different measurable learning methods.

The aim and the goal will be achieved through the next objectives:

- To advance the knowledge on the emotion recognition systems, By applying a high level of machine learning methods and advanced image processing algorithms to identify the emotional state for the person.
- Developing a powerful computer vision system for recognition and analysis of human expressions from real time videos, by taking the advantages of the pupil size and analysis the direction of the gaze to improving the performance of our methods to get high-resolution results.
- To provide literature review of using gaze as an emotion indicator, by providing several types of researches related to the emotion detection study and compare among them, and explain them in such a brief way by taking advantage and disadvantages of every research.
- Optimizing and improving selection methods and feature extraction for human expression recognition, which morphological filtering and removals of noise are using which are very helpful in effective detection of the iris centre, for increasing learning precision and improving thesis result comprehensibility.

- To defeat several of the limitations and disadvantages of existing human emotional detection methods, and discuss them in such a deep manner by taking them attributes, implementation tool, accuracy , age and numbers of volunteers, the light of the environment.

1.3 Structure of the dissertation

Chapter 1: Describe the subject the objective and the aim and objective of this dissertation, and also including the dissertation structure.

Chapter 2: This chapter contents:

- Basic ophthalmology concepts
- Human eye anatomy and function of each part of the eye
- Explains literature on human eye anatomy by discussing its physiological and functional aspects followed by the technical implementations
- Brief summary for eye tracking methods

Chapter 3: This chapter contents:

- A gaze tracking models and also state of the art
- Gaze tracking in commercial areas
- Free and open source gaze tracking methods that used webcam

Chapter 4: This chapter contents:

- Discusses methodological interventions by listing all the datasets which are used for gaze estimation and head pose tracking along with self-recorded videos
- Explain all the methods and filters that used in our dissertation for estimate eye gazing
- Calibration system types was explained here

Chapter 5: This chapter contents:

- Setup our environment and make the experiments of estimation point of the gaze
- Test the calibration method which is used in our system

Chapter 6: In this chapter the final experiments has been done for evaluate the emotion using gaze estimation and pupil diameter, also this chapter contents comparison of visual assessment (gaze estimation) methods, comparison of emotional detection methods using gaze tracking (advantage and disadvantage) and summary of measurement techniques.

Chapter 7: The last chapter contents the main conclusion and research developments.

1.4 Published article related to the dissertation

1. Eye location (Chapter 2.6): A Journal paper on title
    - Location of eyes in images of human faces through analysis variance shine intensity
    - Faces and eyes Detection in Digital Images Using Cascade Classifiers.
2. Head Movement Compensation (Chapter 4.4): Control Mouse Cursor by Head Movement: Development and Implementation.
3. Emotion detection using gaze tracking (Chapter 6): A conference paper on:
    - Assessment of Patients Emotional Status According To iris Movement.
    - Real time gaze estimation for medical field using normally webcam with OpenCV.

2 Basic Ophthalmology

2.1 Human eye physiology

Human eye is one of the five sensory organs which facilitate vision enabling both humans and animals to immediately see around. Human beings possess a high visual ability if compared other animals and have a very complex sight system. Approximately 30% of the human brain is engaged in the processing and interpretation of visual stimuli. The human brain is capable of, constructing physical images through the physical stimuli of light rays that is located in the human eye which is further converted into chemical and electrical signals. Identification of different colors is another process which is supported by retina in the visual system.

According to the Greek philosophers, eyes are “windows of the soul”. Through the eye we can read, interpret strong emotions including both happiness and intimidation. Pupil is another important parts of the visual system which expands or contracts according to the mental and physical state of human beings. For example, pupil expands when a person is focusing on something or stressed. On the other hand, it shrinks when a person is confronted to something revolting or observing a nearby object.

2.2 Human eye Anatomy

Human eye consists of a retina that is lined with a fibrous roller containing lens, the vitreous body and the aqueous humor [17]. The anatomy of human's eye and its parts are shown in Figure 2.1.

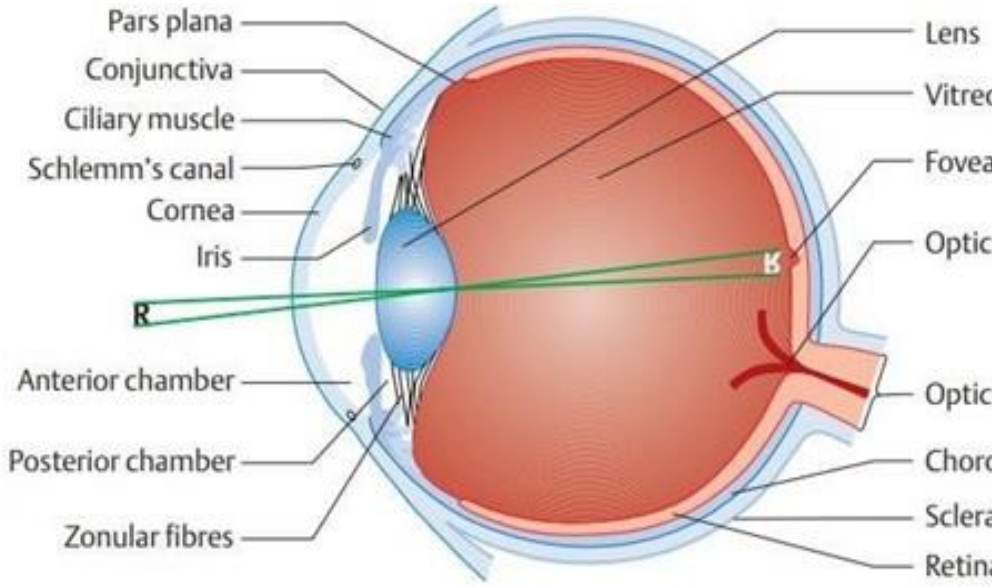

Figure 2.1 General anatomy of the human eye. [18].

Cornea: Cornea (cornea) is an arched transparent tissue found in the eye. It consists of five layers, separated by two membranes. Light breaks at the cornea before it enters the eye. It does not contain blood vessels and takes its oxygen directly from the air. It is the most anterior part of the eye.

Sclera: Sclera is also known as the white part of eye which is an opaque

outer layer of eye. It contains collagen and elastic fibers.

Conjunctiva: It is a transparent membrane lining which is present inside the eyelid and other parts of the eye.

Iris: It is the most visible part of the human eye, and consists of a circular and longitudinal muscle. Iris is used to control the amount of light passing into the retina through the human eye. Furthermore, it also controls the diameter and size of the pupil. The color of the iris distinguishes the color of the eye, including brown, blue or green.

Pupil: It is a hole in the center of the iris, which is circular in shape. Its size is changeable, and its function is to control the amount of light entering the eye. Two muscles in the iris automatically adjust the size of the pupil to the appropriate light level as the enlarged muscle expands the pupil in low light. In extreme light, the sphincter narrows the pupil to prevent intense light from entering the eye. The pupil also narrows when the eye looks at a nearby object to form a clear image.

Aqueous: The watery liquid that streams between the cornea and lens is termed as aqueous.

Lens: lens is a transparent structure of the eye. Its primary function is to contribute in the refraction of light towards the retina, which is completed with the help of the cornea. The lens changes its shape in adherence to changing the focal length of the eye enabling the individuals to focus on

objects at different distances, thus allowing to create a true picture of objects in the retina. The process in which shape of the lens is changed is known as accommodation. Given that this process is similar to the process of focusing the camera through the movement of the lens.

Schlemm's canal : also known as the scleral venous sinus, is a circular canal in the eye that collects the aqueous humor from the anterior chamber and sends it into the bloodstream through the anterior ciliary veins.

Vitreous body: It is a gelatinous and transparent liquid that fills the space between the inner surface of the retina and the posterior surface of the lens, is denser than the aqueous humor, which is in the space between the lens and the cornea.

Macula: Oval with much pigmented yellow spots near the center of the human retina. It is about 5 mm in diameter and considers the center of vision.

Choroid: vascular layer - containing blood vessels - in the eye, thickness ranging from 0.2 mm - 0.1 mm, its main function is to nourish layers outside the retina and provide oxygen.

Optic nerve: A conjugal nerve transmits visual information from the retina to the brain. The optic nerve in humans during the development of the fetus is derived from the optic stalks during the seventh week. The optic nerve consists of the axons of the retinal neurons and glial cells, extending from the optic disc to the optic chiasma and continues as an optic pathway until it

reaches the lateral geniculate nucleus pretectal nuclei, and superior colliculus.

Retina: It is the inner layer of the eye. The retina converts light rays into nerve impulses that are transmitted through the optic nerve to the upper brain centers in the light receptors. Retina is the main component of the eye and its most important function is photoreception [19]. All other human eye structures are subsidiary and work with retina on focusing image, to adjust the amount of the entering light to the eye or to supply motion, protection or nutrition.

Uveal tract: Immediately internal to the sclera, and between it and the retina, lies the uvea, a vascular tunic analogous to the pia-arachnoid of the central nervous system. Primarily, the uvea provides nutrients to the eye.

The ciliary body: The second portion of the uvea, lies just anterior to the choroid and posterior to the corneoscleral margin and provides nutrients by forming intraocular fluid, the aqueous humor. In addition, the ciliary body contains muscles which provide a supporting and focusing mechanism for the lens.

2.3 Function of Human Eye

The human eye functions like a camera. The inner part is considered as darkroom. The rays collected by the cornea and crystalline lens functions pass through the aperture of the pupil, i.e. the aperture in the camera, until they reach the retina that is the film facilitating focus. As a result, the processed image is formed /decoded in the brain [20].

As clearly represented in the Figure 2.1, iris is in the middle and has a round opening which is called pupil which allows light to pass through. Iris has muscles that cause constriction (miosis) or dilation (mydriasis), depending on lighting conditions, which is reflective of the autonomic nervous system. When there is a lot of light mitosis occurs, whereas when it is dark, the pupils dilate so as to permit more light to enter through them.

The cornea and iris form an angle, which is called anterior chamber. The anterior chamber is filled with a clear liquid, the aqueous humor. This liquid is responsible to maintain the setting in the anterior segment of the eye. Behind the iris is the crystalline lens which is approximately the size of a lentil. The cornea refracts and concentrates the light rays on the retina, which is focused on the macula. The lens may fluctuate its shape by means of a muscle and simultaneously varying refractive power, following the adjustment mechanism. Thus, "zooming" focuses the image, depending on

whether a person looks away or close to an object. Behind the lens is the posterior chamber of the eye which consists of the vitreous body which also facilitates the passage of light to the retina.

Finally, in the inner part of the human eye is the retina. The light stimulus is taken up by light- sensitive cells, cones and rods, and by a complex mechanisms converted into nerve impulses. Nerve cells take stimulus and carry it through the optic nerve. The ganglion cells in the human eye produce about a million axons to generate the optic nerve. The layer of the retinal nerve fibers consists of these axons, which converge to form the optic nerve. The cranial cavity is reached by the transfer of the orbital part of the nerve entering the bony foramen optic. The optic nerve is made up of visual fibers (80%) and afferent pupillary fibers (20%) [21].

Visual perception occurs when the optic nerve directs information through the thalamus to the cerebral cortex, but there are also two sites in the brainstem that require information for the vision mechanism to be transmitted through the optic nerve.

The pretectal is the first of these sites consisting of a group of cells (nuclei), which in turn control the size of the pupil in response to the intensity of light. The information that controls eye scanning and associated information related to moving targets is transferred to a second location in the brain stem,

called the superior colliculus nucleus. The site responsible for moving the eye in small jumps is superior colliculus. The collection of a series of still images enables the smooth visualization of the brain and this is allowed by saccades. Inordinate blurring problem that could be made because of the eye pan easily and without interruption or difficulty over a visual landscape, this problem has been solved by saccadic eye movement. If a person's eyes are seen staring through the room here the saccades will be easily observed. Most projections travel through the optic nerve from the retina to the thalamus, this is called lateral geniculate nucleus (LGN), in the deep center of the brain. The retinal inputs will be separated by the lateral geniculate nucleus into parallel streams, one containing motion and contrast, and the other containing fine structure and color. The top four of the six layers of the lateral geniculate nucleus consist of cells that process fine structure and color. The size of the small cells in these four layers is the reason why these layers are called parvocellular layers, the bottom layer of LGN consists of a group of cells responsible for processing both motion and contrast, and because of the large size of the cells it was called magnocellular layers.

All of the electrically encoded visual information transfer by the cells of the parvocellular and magnocellular layers to the brain's back and then to the primary visual region number six. The sensory system is able to measure wherever objects are, at any place through aligned cells in region six in

numerous ways. At first, area number six all the cells are aligned in retinotoptical format, which will detect point to point flowchart that exist between the first visual region and the membrane, and neighboring or adjacent areas within the corresponding overlay to adjacent areas number six. This makes region six to put the object in two dimensions of the visual reality, vertical and horizontal. By examining the signals that come from both eyes, the depth and proportions will be charted in this area, ocular dominance columns define these signals when refined in stacks of cells, a game form board contacts alternating between left and right eye. Any minor difference within the position of associate degree objects in relation to every eye let the depth to be measured by triangulation [22]. Columnar alignment of primary visual regions was Torsten Elie Wiesel and David Hubel, and they were awarded the 1981 Nobel Prize in this field. At last, in region six the orientation columns will be aligned, it is a group of cells that that are triggered efficiently by selected trend lines. The region six will be able to determine the sides of the objects with the visual reality and this will be done by help of orientation columns, so they will begin the advanced visual recognition project.

2.4 Human eye as an emotion indicator

Feelings and emotions are expressed differently in certain areas of the face and this has been shown by previous studies. For example and not limited, expression of the positive feeling or emotions can be show more in the lower portion for the face, while the negative emotions will be shown in the upper portion of the face. Sometimes an individual may come up with uncertain or accurate results about the environment if he/she fails to look closely at the relevant signals or integrate into what he/she sees fit. From visual processing the information will be gathered and also the technology of tracking the eye can help explain the nature of the impaired emotion perception. We can be more specifically, the eye region considered as main part for the emotional signals. Emotional reaction calculates by take the properties of the blink, size of the pupil and the gaze and analyzed theses data. The activation of the sympathetic nervous system is coupled with the size of the pupil and emotional reactions are known to be associated with varying size. Regulating the amount of light entering the eye is the main task of the pupil and the second task is to adjust the diameter of the eye. However, the size of the pupil is concerning to cognitive processing load and hue in visual stimuli or amount of the light makes the relationship so complex. Also the blink has been related to the reaction of the emotional, defensive reactions are the simplest example for this like emotion-modulated startle for eye blink. Joy,

disgust, disgust, sadness, surprise and fear considered one of the basic emotions, make a certain change in an element of the eye area (variation size of the pupil, furrow and brow movement and other eye area muscle differences).

The gaze can be influenced the (neutral and fearful) emotions that approve it by the [23], [24], [25] researches. These results suggest that facial expressions of fear are mainly decisive in influencing the direction of interest in people worried, as neutral, fearful, happy and angry expressions were all give to the one in an infrequent arranging. Consequently, the currently results show that it is not the passive and negative expressions in common (e.g. angry and fearful) that are significant. Almost orientation shows a strong impulse to be distinct from frightening expressions. This discovery produce sense in that according to the gaze of the eye and emotion can give a significant communal hint that out of danger. A direct look at another person's face with anger looking shows a threat, It is a huge difference where the angry face is looking somewhere else or when looking directly to someone face [26].

In the testing's with direct gaze, support for this belief was found to have been slow to respond in people with severe anxiety or anger, as opposed to in neutral or happy expressions where the response was rapid. An analysis of

programmed present that is lateness on experiment with direct angry expressions of the gaze spread from frightened face with a straight gaze.

Kleck and Adams reported these outcomes are logical, which features that straight gaze boost the understanding of angry emotion and has no effect on the perception of frightening expressions [27]. His study investigated the role of happy, neutral, angry and frightening facial expressions in enhancing and improving the direction of vision. Where portraits of faces with side and direct looks were presented. A target letter (T or L) look alike unexpected to the right or the left of the human face, either 700ms or 300ms after changed the direction of the gaze. In congruent conditions, response times were faster, for example, where the eyes looked at the exactly target Compared to inappropriate conditions, for example, the eyes stare away from the desired letter or target.

Reaction times will be affected in facial expression, but these effects were qualified by individual differences in self-reported anxiety. In contrast, when the eyes stared direct ahead, Anxiety was associated with personality with a slowdown in response when facial expressions portrayed anger. Thus, in people who are anxious, attention is likely to be drawn by expressing anger, while attention is more effectively channeled through frightening facial expressions.

However, as in the previous study, monitoring the facial expression of fearful with an avoid look (gaze) output in increased Indoctrination power for people with (not low) but high Anxiety trait levels [28].

Leppanen and Hietanen [29] after making six experiments they concluded that face features are not affecting the direction of attention caused by the direction of the look (gaze). This research confirms the weight while taking into account individual differences in stress and critical to the ability of emotional expressions cause greater focus on results, although the other research and studies have failed to discovery effects of the face features to Look the effects of indoctrination In unspecified samples (for example, Mathews et al [23]).

There are other researchers and studies supporting this inference only the frightening and neutral speech in the task of outlook memorization [30], [31]. The facial expression of the fear is especially important in affecting the attention of people in distress, for example the neutral, fear, happy and angry expressions were provided with all the same participants in random order. The result of his study shows that not only the scared and angry expressions (negative expressions) that important, actually, the enhanced guidance appears to be specific to express fear. This makes a logical conclusion that according to eye gaze and emotion can supply a significant social cue to the source of danger. A direct looking of the angry face at someone clearly

indicates where the menace is coming from, as well as a scary face looking elsewhere.

A new method of detection and measuring the emotion of the human have been introduced by Jakob de Lemos et al [32] by analyzing the properties of the eye through the platform of the eye tracking. They experiment was done by using the planned a test environment and images displayed in the screen. Eye tracking data was used in this method and make it possible to detect and measure the real senseless and emotional responses uncontrollable sooner than they are perceived cognitively, explain and interpreted and, and biased from our brain.

The system (IET system) they made is automating manner similar to those used before for example witness experts, psychologists and the others, who qualified to discover subtle differences in facial expressions, to find emotional signals from persons. More detailed, emotional cues are based primarily on the eye area. Jacob says that the pupil's size, flashing characteristics and gaze are analyzed and used to figure calculate the emotional reaction index, used the equipment of modern eye tracking and software developed for this purpose. The emotional reactions are related to the pupil size. Dilation of the pupil has been linked with activation of the sympathetic nervous system. Here the relationship will be so complicated

because of the size of the pupil is also linked to the processing of cognitive load and the amount of light or color gradation in visual motivate. The emotional state and reaction also related to the blink, such as the reactions of the defensive for example modulated emotion for eye blink startle. Finally, perception patterns have been linked to emotional effects. The human brain is able to gather accurate information in a way that can reveal people's feelings.

2.5 The types of eye movement

The eye is always moving even when we are in sleep time. Movement of the eye can be divided into several various classes, and the most popular ones are explained below:

2.5.1 Saccade

The eye makes some rapid movement when jumping form point to other this movement is called Saccade movement. Saccades generally have high velocity and acceleration. Saccades are studied as a movement of the eye among fixation points. Saccade movements duration usually between 40 ms and 120 ms [33].

2.5.2 Fixation

Fixation or visual fixation is one of the basic eye movements and usually maintains the direction of vision in a stable location. The eye usually stay fixated on a specific spot to obtain the information from view; almost all the information is obtained during the fixation. A small random drifting may be involved through the fixation of eye movement. The overall duration of fixation on a point its ranges from 120 ms to 1000 ms, the typical frequency of the fixation is less than 3 Hz [33].

2.5.3 Pursuit:

Pursuit is involuntary movements. The pursuit movements happen when the eyes keep track of a moving target in order to save it in focus. Normally eyes usually track a moving object smoothly, but sometimes there are some rapid saccades were performed by the eye to concentrate and focus on the target [33].

2.6 Eye-tracking Methods types

The term eye tracking refers to the process of following the movement of eye and deciding what an individual wants to see. There are variety of principles which are utilized in mensuration of eye movements, together with measurements of electrical and physical phenomenon signals, following variety of visual options within the image of the attention, mensuration of

relative reflection of infra-red (IR) light-weight, and mistreatment of either mechanical or optical levers or a field.

The next section is an approximate classification of eye tracking methods:

2.6.1 Electrooculography (EOG):

Electrooculogram (EOG) is a test that involves placing small electrodes near the eye muscles to measure their movement. This test is used in polysomnography. Under usual conditions there is a potential difference of approximately 0.4mV to 5mV between the cornea and the Bruch membrane located at the back of the eye. The origin of this difference is found in the retinal pigment epithelium and allows considering the presence of a dipole, which can be represented by a vector whose arm coincides with the anteroposterior axis of the eyeball, where the cornea corresponds to the positive end and the retina at the negative end of said dipole

2.6.2 Video Image Analysis:

The basic concept in this method is use a video camera to monitor the user's eyes. Some image-processing techniques after that used to analyze the image and track the features of the eye. The system will determine where the user is looking at the moment based on calibration.

2.6.3 Mechanical Method (Coil Systems):

In this method, a magnetic coil is used to obtain data for estimates eye movement by integrated as a part of the contact lens. Before the experiment, the user must wear a contact lens. This method not suitable enough, because of the coil can be fragile and also sensitive to noise.

3 Gaze tracking

3.1 State of the art

This chapter will contain the outline of the state of the art in gaze -tracking concepts which concentrate on videography because this is the most pertinent to our dissertation. At first, some models are presented for tracking the gaze, which covers the location of the pupil, the point estimates of the gaze, and estimates the head position. After that, solutions for gaze tracking will be presented for the commercial purpose, which covering remote VR gaze tracking and mobile, while some models for the gaze tracking that using webcam and analysis software are also presented. They are also considered the best solution for eye tracking which using at the current time for the market purpose. After giving a brief for the eye-tracking solutions in the commercial field, several solutions will be compared. All of them almost focus on gaze tracking which based on the webcam, and this what our dissertation's aim is about.

3.2 Gaze tracking models

Two kinds of models are required for tracking the gaze in videography, one model for detecting and estimating the point of the gaze while the other is to

detect the center of the pupil. Since in this dissertation the volunteers have given them the freedom to move their heads toward any direction, so we need a third model that can estimate the position of the head. In this chapter, some of these models will be explained in brief.

For the detection of the pupil, (Hansen and Ji 2010) [34] approved that it is possible to recognize between the appearance and the shape, almost all of them work on the images containing the area of the eye and also the small area around the eyes. The models (shape-based) suppose that, the iris is an object with the circular shape and ellipses will add optionally around this object (iris) to model the eyelids or sclera. For finding the centers of the pupil, most of the methods are used a circular Hough transform algorithm for example in [35], [36].

Other models have used the random sample consensus or expectation maximization to determine the eyes pupils by performing the local fits of the ellipses or circles into the image [37]. The models that based on the appearance seeking other image properties, for example, the color to eliminate the white spot of the sclera in various color spaces (Periketi 2011) [38]. While other models ignore the candidate pixels in the eye center if these pixels have entropy with high values, stating that, the differences between the illumination and the blood vessel of the sclera drive to it ((Fini et al. 2011) [39].

Some models that related to machine learning such us, SVM (super vector machine), have been used to detect and determine facial properties like eye, mouth and other facial features (Park et al. 2002) [40]. Too many models are used infrared light to estimate the point of the gaze, and these models are used to tracking the first frame image, the source light reflection of the cornea, it represents as first Purkinje image (Ohno and Mukawa 2004) [41].

Others are used multiple cameras for the estimation of the approximate center of the eyeball and also for estimating the 3D head position to cast rays from the eye-ball during pupil detection (L Tian, Z Zhang 2017) [42]. Some methods need some kind of calibration for gaze estimation (Hansen and Ji 2010) [43], however, a few studies have been done by used free calibration methods, such as what is used in (Xiong et al. 2014) [44] in the research they used Gaussian mixture models, while (Nagamatsu et al. 2010) [45] used the geometric calculations. But, to apply the free calibration methods there are some qualifications. Nagamatsu et al. (2009), for keeping the head focusing on the four cameras and also to avoid the movement of the head, a chinrest was used for this purpose, where Xiong et al. (2014) only used differentiating between the right and the left of the head.

Face detection always was the problem in the computer vision, this problem can be solved by using and applying Haar features (Viola and Jones 2001) [46]. In 2013, unified schema was proposed to compare with the detection

results. The adverse effects, all the methods that depended on the landmarks supply simple wherewithal to extract the image which contains the eyes region, which leads us to use these features for processing the pupil detection methods. This will lead us to remove the necessity for eye detection specifically, like the example that has been done by Rosenfeld and Sirohey (2011) [47].

3.3 Gaze tracking solutions for the Commercial aim

The gaze tracking solution for the commercial has come in a magnificent variety. There are some hardware systems with one or multiple cameras, this hardware is coming with their specialized computer hardware, and also some of them have their software that used to analyze and visualize the recorded data, whereas other hardware depended on other software.

Most manufacturers concentrate on webcam solutions, few of them have been built strong specialized hardware for this reason. The price is wildly different; it is ranging from 100 USD up to 10000+ USD [48] for example the cheap one (Tobii Eye Tracker 4C) while the expensive eye tracker device is Tobii Pro Eye Tracking Glasses 2 – 50 / 100Hz (Wireless).

Most of the gaze tracking solutions for the commercial purpose is focusing on the degree of the visual angel. At the viewing distance 57.3 cm, the gaze tracker will have an error of about 1 cm with an accuracy of 1°.

In this chapter, just some segment of available solutions will be listed. There are some cheaper remote gaze tracker which is used for gaming and not using for the research such us Tobii Eye Tracker 4X1 and Tobii EyeX models.

A developer kit for Windows computers it has come with Tobii Hardware and also can buy it with the built-in gaze tracker or the standalone for a modern gaming laptop. Their competitor only with the low-cost class, in late 2016, Oculus was gained The Eye Tribe2 which is the selling tracker for 102 EUR to 168 EUR (Constine 2016), and this item no longer available on their website [49].

Also, there is another gaze tracker device with a price below 1000 EUR, which is coming without any software GP33; it comes with a frame rate of 60 Hz and accuracy between 0.5° to 1°.

The Tobii gaze tracker offers a higher price, a device with an accuracy of 0.4° and a frame rate of up to 600 Hz. SR Research provided another remote gaze tracker device with similar accuracy and up to 1000 Hz, it is named EyeLink 1000. Another device found with present multiple monitors with multi-camera setups, the number of cameras that used up to eight cameras. They have between 60 Hz and 120 Hz frame rate and 0.5° of accuracy. LC Technologies produced another tracker device with a multi-camera setup [50].

Lab [51] using a unique solution for a mobile eye-tracking pupil, all their software as well as their hardware can be built manually because of all their software are open sources. They present their gaze tracking glasses with frame rate with 200 Hz and price (2967 USD); also these glasses have add-ons for HoloLens, an HTC Vive, and a VR kit.

Fovea [52] represented as another choice of the VR with less price (599 USD), this gaze tracking device is developed specifically for VR solution: To improve the graphics renderings performance (Patney et al. 2016) [53], they specialize in foveated rendering. However, pupil labs and Fovea are not just the devices in the market for the VR gaze tracking purpose.

Another hardware kit Put up in markets, Ergoneers [54], is designed to be integrated into helmets beside that can be adapted to VR. A wide range of remote, AR, Movile, And VR solutions are offered by SensoMotoric Instruments [55] with high-price ranges till these Instruments bought by Apple company in 2017. Some companies are built their software for analyzing the data of gaze tracking, although too many hardware manufacturers are shipping their independent software.

An interactive mind [56] is the company which is works closely together with LC Technologies manufacturer. Other companies that provided the software for tracking the gaze and also offers software for EEG

Electrocardiography (ECG), and other biometric analysis such us iMotions [57] .

Other Companies such as the Für Wahrnehmungsforschung Institute [58] provided the studies of gaze tracking; these studies focus on the advertisement and marketing purpose. Eye zag [59] proposed (SaaS) approach and this approach presented as modern software; this software offers service that used to track the gaze for the website, by using the webcam for the user. The related xLabs and EyesDecide [60] are similar for the SaaS platform, they also used for recording and analysis the gaze for the user on website depended on the webcams.

3.4 Free and open-source webcam gaze tracking software

Some of the Free and open-source webcam gaze tracking software are trying to get the best performance for gaze tracking. All of them have different requirements and guidelines: They require modified webcams, regular Webcams, or webcams with specific properties and also they need images for full-face or just a single image that contains the eyes. Several of these projects only used for tracking the eyes. There is a comparing table overview of all the software presented in this chapter as shown in Table 3.1.

Software and Authors	License	Input	Tracks
Webgazer.js (Papoutsaki et al. 2016)	GPL 3.0	face	gaze
OpenGazer (Ferhat 2012)	GPL 2.0	face	gaze
gazr (Lemaignan et al. 2016	Apache 2.0	face	head
webcam-eyetracker (Dalmaijer 2015)	GPL 3.0	infrared eye	eyes
deep gaze (Patacchiola and Cangelosi 2017	MIT	not available	head
iTracker (Krafka et al. 2016)	Research-only	face, eyes, mask	gaze
eyeLike (Hume 2012)	MIT	face	eyes
Gaze	MIT	face	eyes

Table 3.1: Comparison among all of the open-source software for gaze tracking, regrettably, not all the software has been tested, thus results were dependent on the author's reports results [61].

Another software used that also for tracking the gaze, it was implemented by Tristan Hume called eyeLike [62] for detecting the eye center the researcher used (Timm and Barth 2011) [63] algorithm, the side effect of this, it is not appropriated to be integrated into other software. A more flexible interface was implemented in this method.

Piotr Zieliński developed software; it is called Open gazer [64] which used for tracking the gaze after some of the few calibration steps by using just

regular webcam without any other additional equipment. This work published at the end of 2010, and the last update was 2013, however, this project continued in 2013 and maintained till 2016 by Onur Ferhat [65].

This fork was able to improve the original project can achieve error decreasing by about 17.5 % less than the original project by 1.5°, and about 1° for eye light gaze tracker (Tobii X1). PyGaze [66] is proposed as one of the most examples; the primary task for this software was to perform the gaze tracking and the eye-tracking analysis (Dalmaijer et al. 2014) [67]. This was maintained by Sebastiaan Mathôt and Edwin Dalmaijer, and the main program that used for this purpose was a python.

The reason why they used Python for an application:

1. Python a program that free and open source.
2. The source file is available on the website, and this will help the other developers to maintain this code or integrated it with other software.

The infrared light is used to detect and track the pupil for one eye, and this the side effect of this method, because we cannot use it with all webcams, because not all webcam has a built-in infrared filter. Dalmaijer mentioned in his report that he must turn off all the regular lights because this camera comes with specific lighting rules.

Séverin Lemaignan he integrates his package Gazr [68] with operating system for robot (ROS) [69]. The Gazr package is used to Human-computer interaction (HCI) and also use for human-robot communication; they also used the capabilities of gaze tracking to done this work, it is used webcams or RGB-D cameras for estimation of the head positions (Lemaignan et al. 2016) [70]. Kyle Krafka and his colleagues at the start of 2016 published the Gaze Capture dataset beside the pre-trained CNN tracker (Krafka et al. 2016) [71]. The gaze tracking performs using iPad and iPhone cameras.

Another gaze tracking software is using a browser for tracking the gaze online and this software has been done by Alexandra Papoutsaki using WebGazer.js (Papoutsaki et al. 2016) [72]. They provided some examples that make it so easy to integrate this tool into the websites. The WebGazer.js software using the webcams, and then the calibration method will be done by asking the user to focus on some points on the screen.

Massimiliano Patacchiola uses for his software Python package deep gaze [73] CNN's neural network to implement a different HCI (Human-computer interaction) tasks. One of its features is to tracking the gaze, the researcher mentioned in his report, but in the current version, this functionality does not offer yet. But, this software is able to estimate the position's head (Patacchiola and Cangelosi 2017) [74]. Since this work was published to

estimate the head position from 2017, a gaze estimation feature maybe will be added in the future.

4 Methodologies

4.1 Overview of the work

In Figure 4.1, the high-level description of our eye-tracking can be seen, gaze estimation and emotional detection system. While Figure 4.2 provides a comprehensive explanation of the extracted features in the first framework. In Figure 4.3 the required steps for eye tracking are defined while in Figures 4.4 and Figure 4.5 the steps for head movement compensation and estimation of gaze are defined.

4.1.1 Feature Extraction in the First Frame

Detection of the eye is the very first task to be performed for the eye-tracking and estimation of gaze. The user must configure the estimation position of his eyes in the first frame of the video that used in this application.

The initialization of positions by the input of the user is a very popular and common approach for the tracking of eyes. These initialization positions should be ideally in or close to the iris of the eye. The next job is to locate the pupil, which is in the middle of the iris. In Figure 4.2.2 the detailed method to find the center of the iris is explained. Figure 4.2.1 is the precondition of Figure 4.2.2 in which morphological filtering and removals

of noise are described which are very helpful in effective detection of the iris center.

4.1.2 Morphological Filtering

The user initialized points as the centers are regarded with a small window of shape like rectangular in every eye. The eyes are in both windows now needed for the extraction of the features.

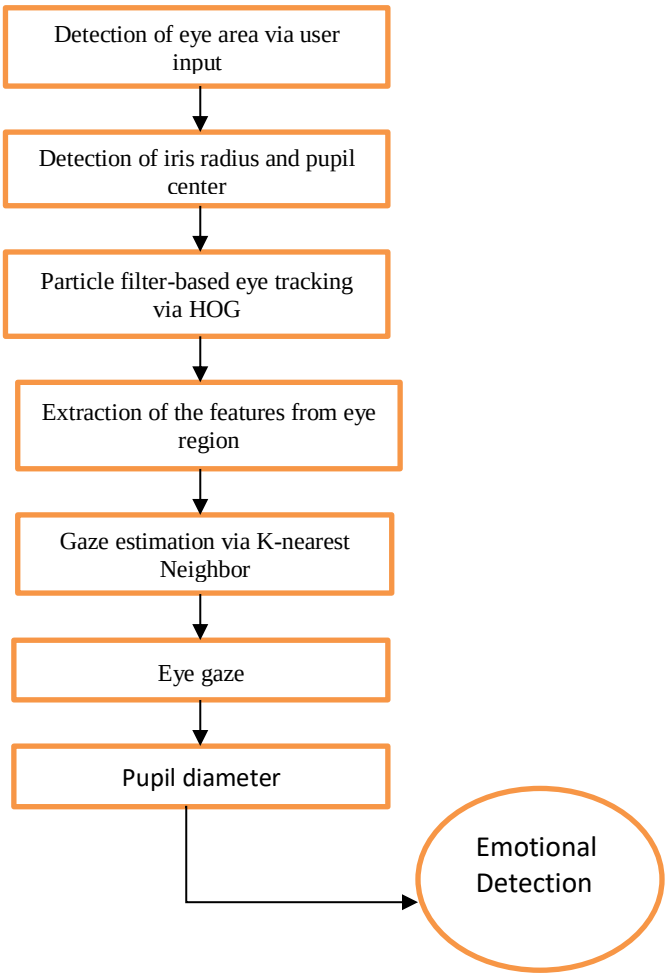

Figure 4.1: Overview of the current system

The right eye image is shown in Figure 4.2. The detection of the radius and iris center is carried out only on the RGB image only on the red screen. This is because the iris normally shows a very low red meaning in dark as well as light irises according to [75]. While the underlying skin and sclera produce considerably higher red values. The red channel image of Figure 4.2 is shown in Figure 4.3. The eye area pictures are very noisy. If we use these noisy pictures then we are unable to get better results. To get better results we need to filter these images to eliminate the noise before starting the work with pictures. A median filter is used to remove the noise of the pictures. Median filtration is an image filtration process that is nonlinear and the filtration is only performed on the eye picture red channel value.

Figure 4.2: Region of eye

In this method a small area of the picture is taken through the median filter and substitutes for the median value of neighbor with the central pixel.

During the process if any noisy value is detected then pixel value is replaced with a median value which is more often than noisy value by using median filters. The composition of the 5×5 neighborhood is discussed here. The median filter is not only used for the replacement of the noisy value it also preserved the edges. A filtered image of the eye is shown Figure 4.3:

Figure 4.3: Applied morphological filtering on red channel eye image (Left), the effect of the filter is shown on the right side

Iris is detected as a circular shape region in the filtered image but still the nosiness is present in the image. There are many possibilities that we can detect the many noisy edges in the area of iris if we take the edge map of the picture and these noisy edges are the bases of an inaccurate result. When any source of light is present in front of the user then the cornea reflects some light, which is shown in Figure 4.5 as a glint.

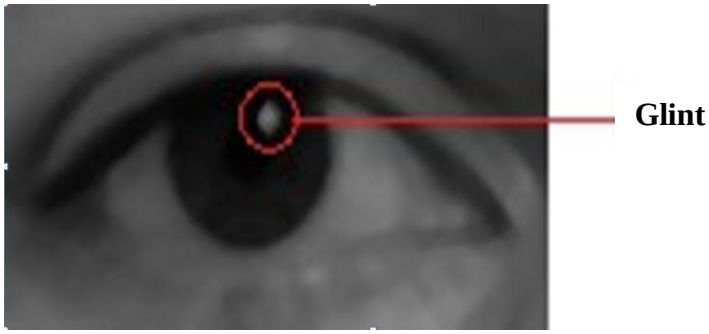

Figure 4.4: Glint in Eye Image.

Dilate operations and morphologically erode are used in this image to eliminate the glint. Erosion is used 3×3 square kernel [76]. The shape and size of the kernel are not fixed but it can be any shape and any size. The kernel point is described as an anchor point and kennel is scanned through the area of an image and replace minimum region value with the value of anchor point that is overlapped with a new image generator kernel. While the dilation process is opposite to erosion operation. In the dilation process, the kernel is scanned through the area of an image and replace maximum region value with the value of anchor point that is overlapped with a new image generator kernel. The following equation is used to define the dilation and erosion process on a grey level image.

$$erod(x,y) = \min_{(x',y') \ \epsilon \ kernel} image(x+x', y+y') \quad (4.1)$$

$$dilate(x,y) = \max_{(x',y') \ \epsilon \ kernel} image(x+x', y+y') \quad (4.2)$$

Let's say A is the area of the image and B is the kernel then the erode and dilation process can be seen as A and B tumult. The dilate process is referred to as A⊕B and the eroding process is referred to as AΘB.

The morphological erode process is referred to as a filtered image of the eye. By using the erosion operations glint is removed, because glint is brighter than the whole iris. Due to the erosion operation the region of the iris is expanded since the region around the iris is lighter than the region of iris. After this dilation operation is performed on the resultant image to minimize the effects of erosion operations. The area of the iris is diminished by the dilation since the sclera is darker than the iris. The same size of the kernel is used for the erosion and dilation operations, now the glint is removed and the iris is back in its original form. In Figure 4.6 resulting picture following the application of eroding and expansion is shown:

Figure 4.5: After Morphological Operations.

This process is used to calculate the number of edge pixels in the annulus region. At each different point the value of r_0 is different. The center (x_0, y_0) and radius r_0 are used for the determination of the parameters of irises. The result in the selected area where there are maximum edge pixel numbers is shown in Figure 4.9 (left). While in Figure 4.11(right) pupil and iris detection result is shown. In this scenario the pupil and the iris center are equal to the center of the half-circle mask.

4.2 Eye-tracking

Particle filters are used to achieve real-time eye tracking which is a sophisticated Bayesian model estimation method. Particle sets or discrete hypotheses are used to find the approximate probability distribution of objects. In an image every particle shows as the center of the region.

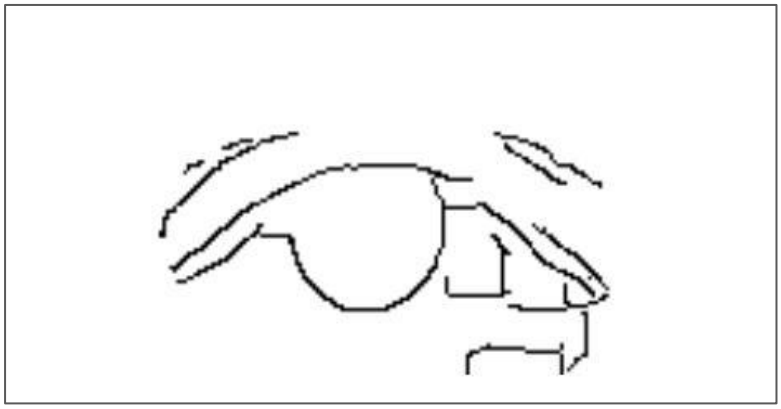

Figure 4.6: The edge map for the eye after applied Canny Edge Operator.

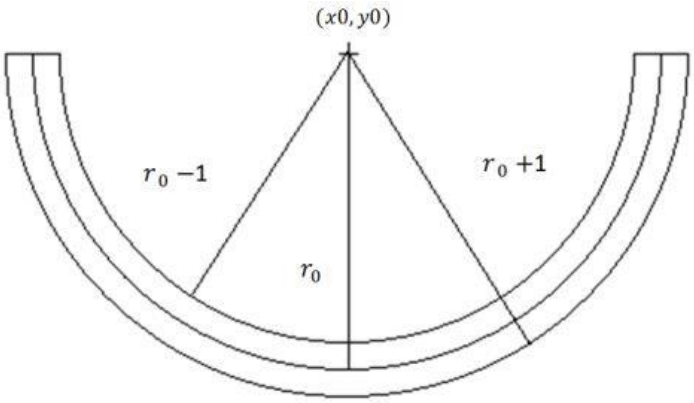

Figure 4.7: Mask for Half Circular.

The comprehensive follow-up method is listed in this section. In section 4.2.1 the comprehensive procedure of tracking the eye is defined by a particle filter. In section 4.2.5 Support Vector Machine method is defined which is used to determine the probability of an eye at the center with a particle filter. In section 4.2.2 the training set that is used for the training of

support vector machine is defined. The SVM classifier defines histograms of directed gradients as the function in Section 4.2.3. Section 4.2.4 explains the integrated image technology (this technology try to enhance access points for both advanced material and introductory in technology field), a very efficient form of feature computation.

4.2.1 Particle Filtering

Particle filtering [78] is an algorithm which based on Bayesian estimation algorithm from the theory of samples, which combines the Bayesian theory with the sequential Monte Carlo (SMC) method together. We can set the vector parameter of the target at *t* time, be indicated by X_t, while observation by Z_t. The observation history starts from time 1 to *t* is indicated by Z_t=[Z_1,Z_2, …… Z_t]. The formulation for Bayesian of particle filtering can be calculated as:

$$p(x_t|Z_t) \propto p(z_t|x_t).p(x_t|Z_{t-1})$$

$$= p(z_t|x_t) . \int p(x_t|x_{t-1}).\ p(x_{t-1}|z_{t-1})\ dx_{t-1} \qquad (4.3)$$

Search and find a set of random samples in the state of the posterior probability density is the basic concept behind the particle filter $p(x_t|Z_t)$

Rounded and replaced $E[g(z_t|x_t)]$ with the mean of the sample to obtain a case of estimating the lowest variance.

The main point of implementing a particle filter is to find random samples from a distributed obey $p(z_t|x_t)$ and this called particles. The sampling points (N) will be discovered and extracted from the posterior probability density separately; the weighted sum will be representing the posterior probability density. An easy sample will be used by Bayesian importance sampling (BIS) of know distribution $(z_t|x_t)$. For sampling, we need to replace the posterior probability density and this can be done by weighing the sampling particles of the significant function $p(z_t|x_t)$, we can obtain from the Bayes theory the following:

$$E\big(g(x_t)\big) = \int g(x_t)\, \frac{p(z_t|x_t)\, p(x_t)}{q(z_t|x_t)\, q(x_t)}\, q(z_t|x_t)dx_t$$

$$= \int g(x_t)\, \frac{\omega_k(x_t)}{p(z_t)}\, q(z_t|x_t)dx_t$$

$$= \frac{\int \big(g(x_t)\, \omega_k(x_t)\big)\, q\,((x_t|z_t)\, dx_t}{\int\, \omega_k\,(x_t)\, q\,(x_t|z_t)\, dx_t} \qquad (4.4)$$

The importance function will be used for sampling and after that the mathematical anticipation of the target state vector will be:

$$\overline{E(g(x_t))} = \frac{(1/N)\ \Sigma_{i=1}^{N} g\left(x_t^{(i)}\right)\ \omega_k\ (x_t^{(i)})}{(1/N)\ \Sigma_{i=1}^{N}\ \omega_k\ (x_t^{(i)})}$$

$$= \sum_{i=1}^{N} g\left(x_t^{(i)}\right)\ \overline{\omega}_k\ (x_t^{(i)}) \qquad (4.5)$$

$\overline{\omega}_k\ (x_t^{(i)}) = \omega_k\ (x_t^{(i)}) / \sum_{i=1}^{N}\ \omega_k\ (x_t^{(i)})$ represented weights normalized while $x_t^{(i)}$ is a Particle sampling that extracted from $q\ (x_t|z_t\ )$.

Some techniques are used often for sampling called resampling, this technique used to defeat the lack of importance sampling. The basic concept behind the resampling is to put down or delete the small particle's weight. According to the weights, big particle's weights are replicated. Particle filtering widely used for tracking purpose with dynamic propagation.

Every particle is the center of a region within the picture in our implementation process. Every phase weight the particles by the output of the vector support classifier [79].

Figure 4.8: Result of the iris and Pupil pupil Detection

4.2.2 Training Set

The BioID face database pictures [80] are used to train the support vector machine classifier. More than one thousand and five hundred twenty-one face images of twenty-three different persons are presented in BioID database (all pictures were in the grayscale format and with 384 × 288 pixels). Face pictures were taken at various locations during several sessions. The database is publicly available [80]. Figure 4.10 shows several examples of BioID face database. In this platform eye positions set by manually is also available. From the database the good eye pictures are choose. Two eye pictures from every face image act as examples for the eye, so for the Eye classifier, a total of 3042 eye pictures are taken as positive examples. For every case, the patient has a square window with a side length of 1/3, making the training examples scale-invariant. It is chosen because it often includes the iris but is small enough to rule out other confusing characteristics,

including eyebrows or eyes. Figure 4.11 displays eye image examples. Likewise, the non-eye images were created with the exception of the arbitrary positions in the picture except the eyes. There are also 3042 non-eye images. Figure 4.12 provides several examples of non-eye images.

Figure 4.9: Set of images from the database of BioID

Figure 4.10: Eye Images Example Images (top row: right eye images, bottom left eye images)

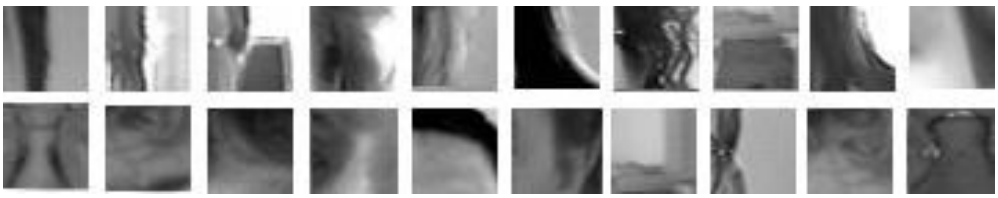

Figure 4.11 : Non eye image examples

4.2.3 Histograms of Oriented Gradients

For object detection and image processing in computer vision Histograms of Oriented Gradients (HOG) are used as feature descriptors. The presence and structure of local artifacts can also be well described by the distribution of gradients of local intensity. A histogram of strength gradient directions with a certain number of fixed bins, which are weighted by gradient, can be used to achieve the implementation of the HOG descriptor. For human detection in [81] and [82], Dalal and others use histograms for directed gradients (HOG). The eye-tracking HOG is used here. HOG is very helpful for the effective detection of eyes, particularly the sclera and iris region. For each

square window in the training set, a HOG function vector is measured. The picture window is divided into four sub-windows; for each sub-window a histogram of width gradient directions is separately determined by gradient magnitude. An 8-bar histogram is computed for every sub-window. Therefore the total direction of the gradient is divided into 8 bits. For measure histograms values the gradient magnitudes of these directions are summarized. The 8-bar histogram then normalized. Here the method of MIN MAX normalization is used. The four 8-bar histograms of the four sub-windows, after normalization, will be connected into a 32-element vector of that classification range.

Figure 4.13 indicates the determination of the magnitude and direction of the gradient of an image of the sample eye. In the very first step the gradient of the image is measured horizontally and vertically. The magnitude and direction of the gradient images are obtained from these pictures. The method of measuring a HOG in Figure 4.14 is shown after measurement of the gradient magnitude and direction pictures. As the training example of the classifier, HOG function vectors are both calculated for eye and non-eye images.

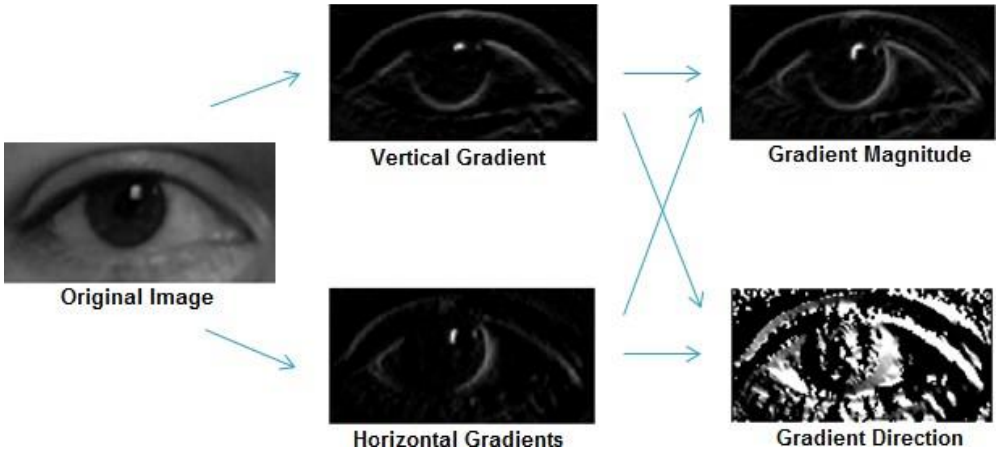

Figure 4.12: Direction calculator and Gradient Magnitude

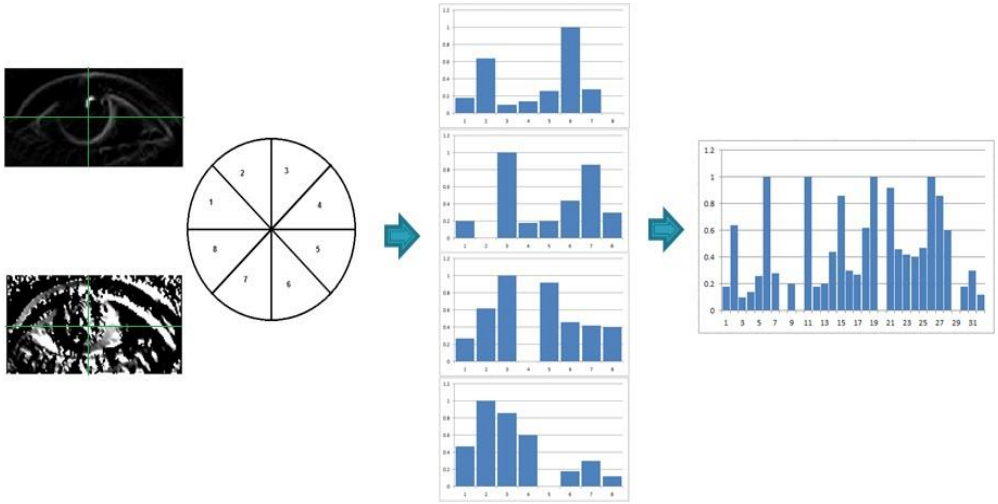

Figure 4.13: Oriented Gradient Histograms Calculation Four classifier.

4.2.4 Integral Image

Viola and Jones use integral image techniques in [83] and [84], which determine calculations of characteristics very rapidly within a rectangular area. In any position (x, y), the full image is determined by taking the sum of all pixel values including the pixels which are above or on the left. Let suppose that the original image is denoted by *Img* and the respective merged image isIntglmg.

$$IntgImg(x,y) = \sum_{x' \le x, y' \le y} Img(x', y') \tag{4.6}$$

Now, the integral image can only be determined with one pass over the original image with equations 4.7 and 4.8. For the summation of total rows, another CRSum array is used. The following is defined by CRSum $(x, -1) = 0$ and by IntgImg $(-1, y) = 0$.Figure (4.15) shown the calculation of integral image.

$$CRSum(x,y) = CRSum(x, y - 1) + Img(x,y) \tag{4.7}$$

$$IntgImg(x,y) = IntgImg(x - 1, y) + CRSum(x,y) \tag{4.8}$$

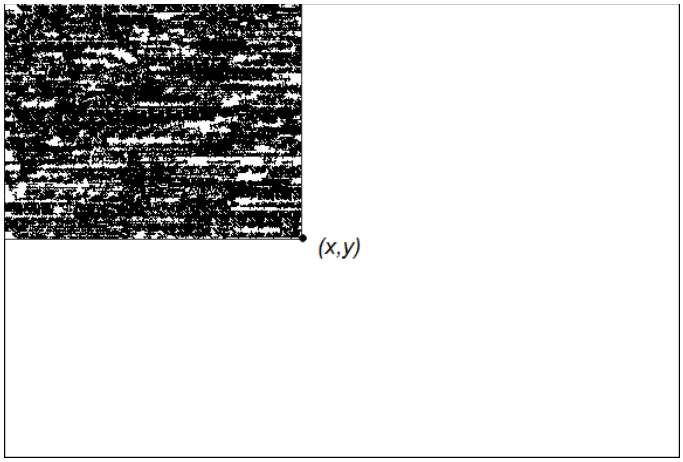

Figure 4.14: Integral Image, the Point (x,y) value comes from sum all values for the pixels above this point.

In the original image the summation of pixels values of the rectangular area is measured in constant time after the merged image is determined. Suppose the computed full picture appears in Figure 4.16. Now it is possible to measure from 4 points P1, P2, P3 and P4 the total pixel value of area C. The P1 value is the sum of all area pixel values of region A. Just like A the P2 value is the summation of the both A and B region pixels and P4 value is the summation of both regions A and D. In constant time, the equation 4.9 is used for calculating the summation of area C pixels values.

$$RegionC = P3 - (P2 + P4) + P1 \quad (4.9)$$

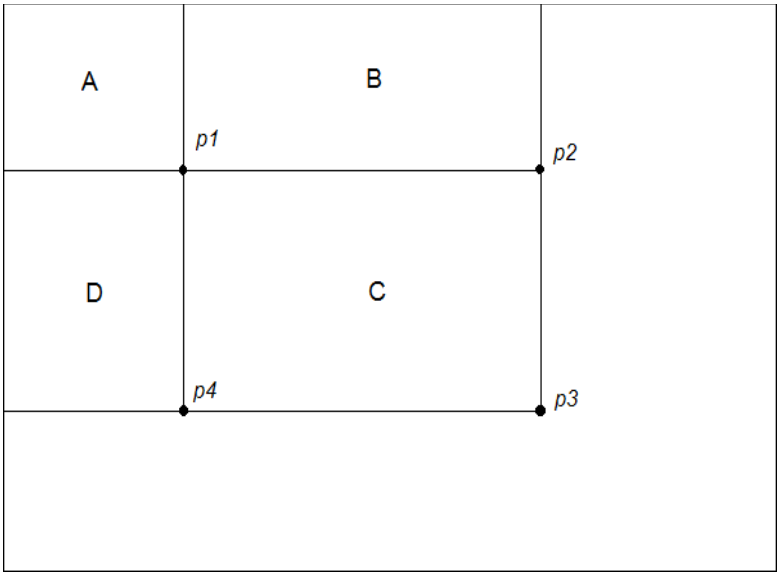

Figure 4.15: The Sum of Pixels in Region C is Calculated from the Integral Image by P3 + P1 − P2 − P4

In specific area the histogram of the oriented gradient is calculate by using the integral image technique. Histogram of every bin is measured by using the technique which is called integral image technique. The entire directional gradient is divided into eight bins, as eight bins are considered during the computation of HOG. After that for every component eight magnitude gradient images are measured. Only those magnitude values whose directions are fallen in that range are taken into account in every picture. Eight separate integral images, one for each component, are calculated from these images. For the specific area, after the determination of integral images the HOG is measured in constant time by using integral images. With this implementation, after the calculation of initial cost for the integral pictures it

is possible to measure HOG values for many regions with a broader image in constant time. The HOG calculation is defined in Figure 4.17 and in Figure 4.18. The value of HOG of any rectangular region in the image can be measured in constant time following this computation. Thus, a 32-element functional vector can also be generated in constant time in any field.

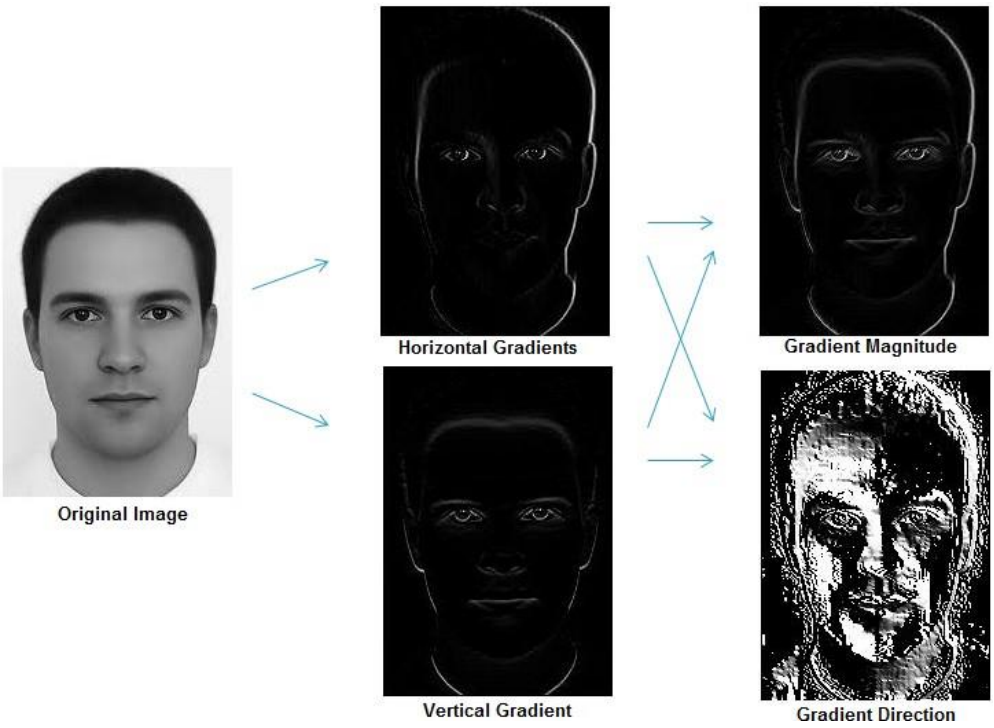

Figure 4.16: Gradient Magnitude and Director Calculation [80].

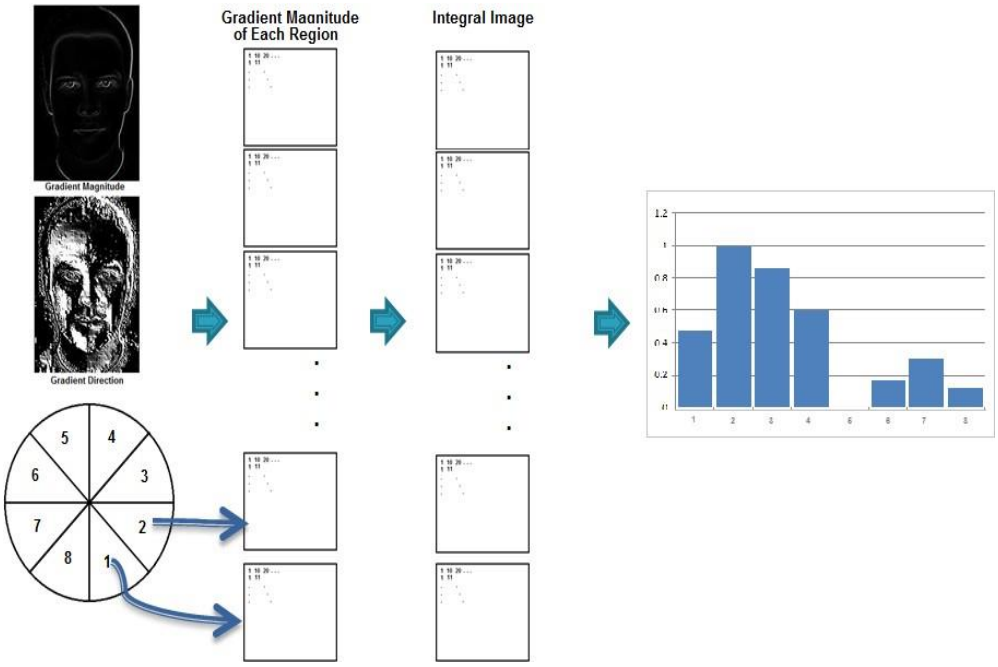

Figure 4.17: HOG calculate (Normalize the input image and then applied the horizontal and vertical gradient, then integral the image)

4.2.5 Support Vector Machine

For the classification of objects, a supervised machine learning technique is used which is called support vector machine (SVM). First of all, the training of the machine is required which included different object classes. By generating a hyperplane or group of hyperplanes in a multi-dimensional setting, the SVM produces a model which best distinguishes the characteristics of the classes with these training data. It also aims to split the various classes to ensure the largest possible distance between different

classes. New tests are categorized into this model space and the class is forecast depending upon which class the point falls into the training class.

Hsu, Chang, and Lin [85] are presenting an excellent guide to SVM. In Figure 4.19, two classes of objects are present. An SVM plots the examples on a feature plane and tries to get the best plane that can split up the two classes. In this case, a line will suffice in the two-dimensional plane. The solid line divides the classes.

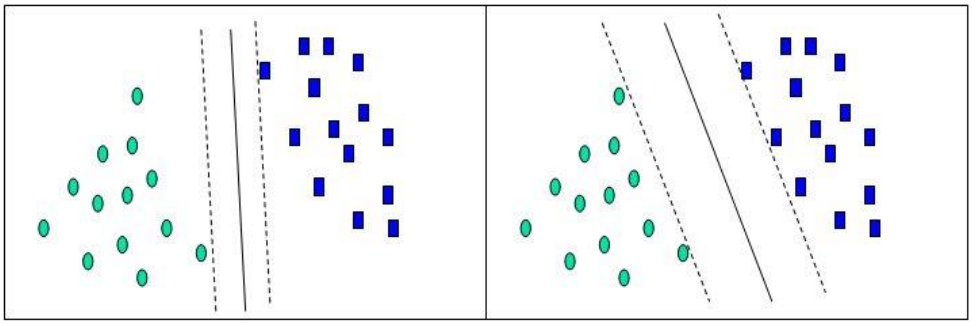

Figure 4.18: A Two-Dimensional SVM

After computing the HOG feature vectors for each example, a SVM (Support Vector Machine) classifier is trained with two classes of objects, eyes, and non-eyes. Instead of a binary classification value, a continuous output is considered whose value increases with the likelihood of that class. The classification boundary helps us to calculate SVM by returning the Euclidean distance of the test vector.

4.2.6 Tracking the Eyes with Particle Filter

Two particle filters track both eyes separately. Particle filters are advantageous for tracking because they maintain multiple hypotheses about the location of the tracked object. A window around the particle position is considered as a candidate. On each frame, some hypotheses about the eye location (particles) may be on the eye while others are not exactly on the eyes. In subsequent frames, as more information becomes available, the algorithm can discard the particles of the poor hypotheses, and propagate only the most likely hypotheses to the next iteration. In each iteration (frame), the particle filter maintains a fixed number of particles. Generally, more particles result in better performance, but the number of particles is limited by the constraint of running in real time. Each particle is evaluated by running its HOG value through the SVM prediction described in section 4.2.5. The SVM output value, which is the distance to the classification boundary, is converted into a particle weight by first normalizing by the minimum value so that all values are positive, then adding 1, then taking an exponent to exaggerate the difference between higher and lower values. The particles for time $t + 1$ are produced by sampling particles at time t with probability proportional to their weight. Each sampled particle is then propagated to the next generation, with the addition of a noise term. Noise considered here is Gaussian noise. A technique is used to introduce adaptive

behavior into the particle filter. The noise term in particle propagation is set to the sum of a base level, proportional to the distance between the pupils, and a variable level, proportional to the number of particles with high absolute scores. Typical particle filters normalize the particle weights, making them insensitive to absolute confidence levels; however, this implementation increases the noise when confidence is low that the particle cloud is tracking an eye. As a result, the particles are produced in a wider area when they lose the eye. Again, as in the training examples, the edge length of the sub-windows is 1/3 the distance between the pupils. This takes advantage of the constrained situation of the face to relieve the burden of determining the proper scale.

4.3 Feature Extraction in Each Frame

The particle filter tracks the eye, but does not always pinpoint the pupil location. The support vector machine does not reliably distinguish pupil points from other eye points near the pupil. For this reason, the eye features need to be extracted by other methods in each frame.

4.3.1 Searching for the Dark Region

The mean or median of the particle cloud also does not locate the pupil because some particles lag behind the eye after eye or head movements. Instead, it is assumed that the median particle position is at least within a certain radius *r* of the pupil. Within this radius, the intensity value of the darkest pixel is searched. It is supposed that the other dark parts of the faces, e.g. eye brows, hair, are not present in this region. Any pixels with intensity values that differ from the darkest by less than a threshold are labeled as candidates for the pupil pixels, and the median location of candidate pupil pixels is the final candidate for pupil location.

4.3.2 Finding the Iris Center

In section 4.3.1 a method has been described for obtaining the final candidate for pupil location. A small region surrounding this location is searched by using the same method described in section 4.1.1 for the iris center. If the distance between the camera and the eyes does not change, then diameter of the iris will not be changed. Thus the radius calculated in the first frame can be used. So, the iris is searched for within a two-pixel margin of the initial radius. If the iris is not found, then the iris is searched for using a larger range of radii.

4.4 Head Movement Compensation

The method described in Section 4.2 can reliably track both eyes and detect the eye features in each frame. Eye-tracking is also robust to head movements. However, by detecting the number of head movements accurately in each frame, it is possible to estimate the user's gaze more accurately, as it is tough for the users to keep their head still. Although significant rotations of head, changes in distance from the camera, or large in-plane translations will still disrupt the calibration, small in-plane translations can be compensated for with this method.

The head-movement compensation works by an updating template matching method. The midpoint between the irises is chosen in each frame. Then, a small window is chosen as the 'between-the-eyes' region, keeping the midpoint of the irises in the center of this region. This point is chosen as this is a very stable area of the face. Usually, the between the eyes area is brighter than its two sides, as the eyes are on both sides of the region and this area is the nose and the forehead region. Thus, this region is easier to track [86]. Then, in each frame, the 'between-the eyes' region is stored. In the next frame, the closest match to this template, within a fixed distance from the previous location is searched for. The difference between the two locations is the estimated head translation for that time step. Then the midpoint between

the irises is calculated again for that frame and used to update the 'between-the-eyes' template.

4.5 Gaze Estimation

Efficient eye tracking is a precondition for the gaze assessment. For the estimation of gaze point, gaze calibration is required. For the experiment's user needs to look on the fixed screen locations with their heads fixed. Then the position of the corresponding screen and pupil is gathered. A quick calibration is performed after the noisy data is removed. The k-nearest neighbors algorithm is used for this simple calibration. Section 4.5.1 explains the thorough calibration process.

4.5.1 Calibration Procedure

The calibration procedure is simple. The calibration panel, that is, the monitor of the laptop is divided into some vertical regions. The monitor screen is only divided into a few vertical regions. If these regions are divided by horizontal regions, then as the distance of two adjacent regions in the screen becomes very small, making gaze estimation from the iris data very difficult. For the gaze calibration, the user is required to focus at some predefined points on the monitor; and these points are at the centers of these

regions. The user is required to focus on each point for not a long time. Through this not long period of time, the pupil positions from both eyes are taken for several frames. Figure 4.20 shown an example of three calibration points panel. While Figure 4.21 presents an example of the data acquisition process. By using the method that has been described in Section 4.5.2, to remove the outliers noisy, and the best candidate points are determined. K-nearest neighbor algorithm will do the gaze estimation by using these best points, as described in Section 4.5.3 and Section 4.5.4.

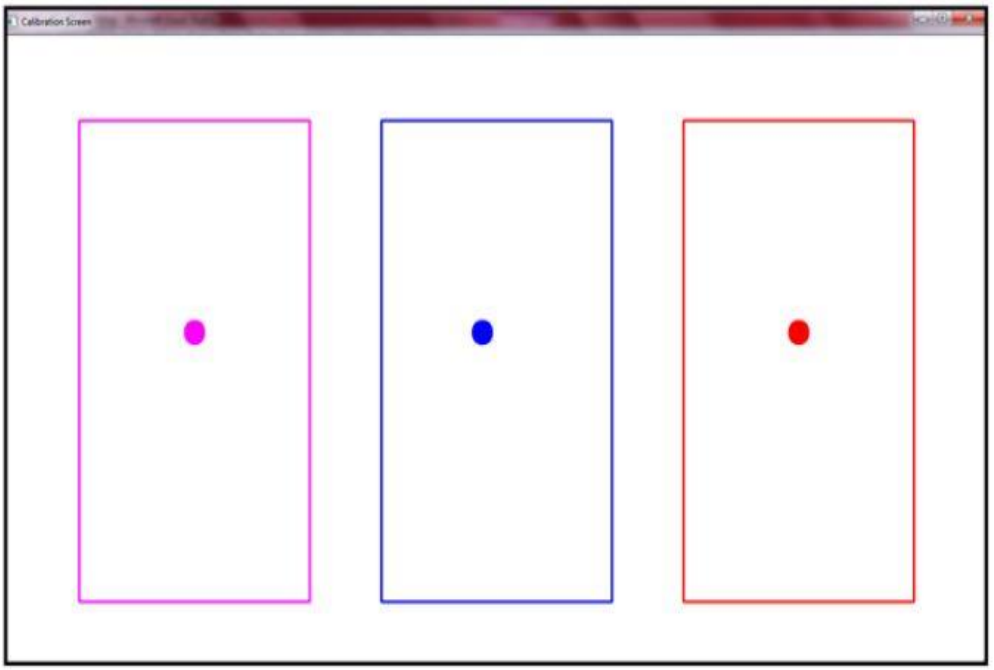

Figure 4.19: Calibration Panel

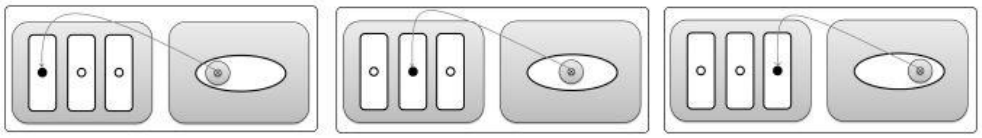

Figure 4.20: Calibration Process

4.5.2 Selecting Best Candidate Points

A two-step process is used to select the best candidate pupil points for every calibration region of each eye. The first goal is to remove a certain percentage of the total set of points as outliers. From that result, a certain percentage of points should be considered as the best candidate. So, in the first step, the center position is calculated from all the points. The outliers are classified from this set based on their distance from the center. The outliers are then removed from this point set. After the outliers removal, the center of the points is calculated again to remove the effect of the outliers on the center. From the rest of these points, the points are ranked based on their distance to the new center. From this ranked list, certain percentages of points are considered as the best candidate points.

4.5.3 K-Nearest Neighbor Algorithm

K-Nearest Neighbor algorithm (KNN) is an algorithm of lazy analysis. It is based on voting of the closest samples in the feature space. Any feature is classified by the majority voting of its neighbors. K is the number of neighbors considered in the classification. The metric to measure closeness is also important. When training the classifier, only the sample features are stored, no computation is made. During the classification, all the computation is done. The training time is very small, but the classification time is relatively large rely on the number of sample's training for KNN. Let us suppose we have the dataset which plotted in Figure 4.22.

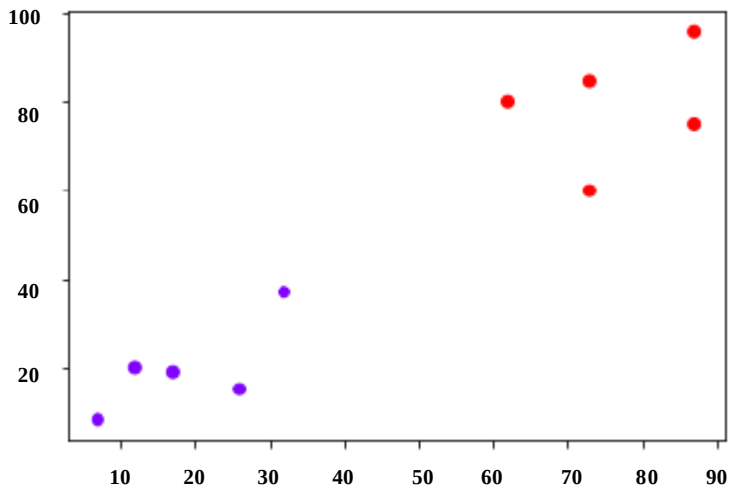

Figure 4.21: dataset have two classes red and blue dots

Let us suppose we need to classify new data (black dot) at the point (60,60), and also we suppose that we have k=3, the nearest three points will be found as shown in the Figure 4.23.

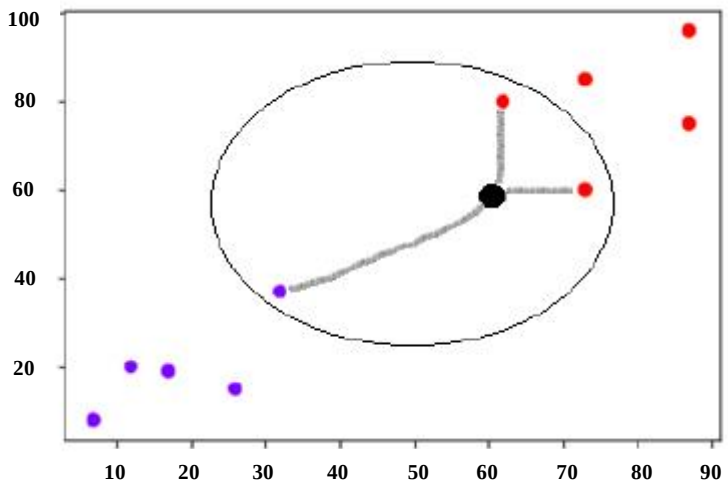

Figure 4.22: Nearest points to black dot

In the diagram above we can see the three closest neighbors to the data point with a black point. Of these three, two are located in the red layer, so the black point will also be set in the red layer.

For gaze estimation, the user's iris center positions are used as the test examples. The test point is now classified using both KNN classifiers. In the ideal case, both classifiers should classify the test example as the same class and give the same output point on the screen. However, sometimes deviation from the ideal case is observed. So, another voting scheme is introduced. For

each classifier, the weight of that classifier is assigned by the nearest number of points that vote positive for the output class. The estimated gaze region is determined by the output of the classifier with maximum vote. Suppose, the classifier of the left eye classified one example as output region 2, and the classifier of the right eye classified it as output region 3. Here 11 nearest neighbors are considered. For the left eye, the region 2 is classified with 6 positive votes for region 2. So, the weight of this classifier will be 6. On the other hand, suppose the right eye classifier classified that position as region 3 with 9 positive votes for this region. So the weight of this classifier will be 9. This point will be classified as region 3, from the output of the right eye classifier, and region 3 is shown on the screen.

4.5.4 Pseudo-code of Proposed Gaze-estimation Method

Pseudo-code will be used for gaze-estimation method as it is described below.

Input the follow: (*Status of _calibration, Calibration points_number, Exm training number , best Exm)*

If (Status of_ calibration == false) **then**

For m=1 → Calibration points _number **do**

Small circle should be drawing depended on the position of *m* on the screen

For k=1→ *Exm training number* **do**

DL(k) = coordinate of left pupil as training data

LBL(k) = *m*, as training label of left pupil coordinate

DR (k) = coordinate of right pupil as training data

LBR(k) = *m*, as training label of right pupil coordinate

End for

End for

For each calibration region as outliers from (DL,LBL) the *outlierPr* percent points will be remove (DR,LBR) based on maximum distance from the centroid of *DL* and *DR* respectively.

Select *bestExm* number of points for each calibration region as best training examples based on minimum distance from the new centroid point of *DL* and *DR.*

By using best training examples, data *DL* and label *LBL* the classier (*KNN*)*L* will be train.

By using best training examples, data *DR* and label *LBR* the classier (*KNN*)*R* will be train .

Status of _calibration=True.

Elseif (*Status of _calibration==True)* **then**

Get the coordinate of left and right pupil in *LP* and *RP* respectively

OutR = classication result from classier (*KNN*)*R* for test example *RP*

WR = number of neighboring points that vote for test example *RP in* previous step

OutL = classication result from classier (*KNN*)*L* for test example *LP*

WL= number of neighboring points that vote for test example *LP* in previous step

if (*OutR == OutL*) **then**

Highlight *OutL* region on screen as gaze-estimation

Elseif (*WR< W L)* **then**

Highlight *OutL* region on screen as gaze-estimation

Else

Highlight *OutR* region on screen as gaze-estimation

Endif

Endif

4.6 Pupil Diameter Measurement

Human detection condition can be done by using pupillary changes. Several researchers are focusing on their studies on the change of pupil diameter to study the emotion of a human, such as stress, annoyance, and also seeing exciting products, or the pleasant and unpleasant image's reaction. Human emotions detection can be detected from pupillary changes when resolving mathematical issues using patterns or separately. However, from medical point of view, pupil diameter can be helpful in early detection of diabetes. Detection of the eye and the pupil is the first and essential steps to measure pupil diameter.

After detecting the image that contained pupil as was described in section 4.3.1, the following process will measure the pupil diameter. There are different methods for pupil diameter measurement such as using ellipse equation [88], and then the algorithm of Hough Transform [89], after this, a circular formula will be applied as shown in equation 4.10 where is the S is a ring era in the form of pixels which have a black color in the pupil territory. The measurement of pupil's diameter can also use an algorithm (geometry feature-fitting) combined with least squares method that can effectively be implemented with an acceptable price [90].

$$r_p = \sqrt{\frac{S}{\pi}} \tag{4.10}$$

Ellipse fitting algorithm using linear least-squares with error criteria to measure ellipse diameter. Zhu [91] mentioned in his report that this algorithm is not complicated and subtle as long as enough data are available. This algorithm used by Mokhayeri [92] for pupil diameter's calculation, depended on formula 4.11 to find out both of the minor and major axis, the centre point of the ellipse is (x0, y0), while both of A, B, C are coefficients. The length is a pupil's diameter.

$$a = \sqrt{\frac{2(Ax_0^2 + Bx_0\, y_0 + Cy_0^2) - F}{(A+B)^2 - \sqrt{B^2\,(A-B)^2}}} \tag{4.11}$$

Another method used to find the pupil diameter by using a Hough Transform algorithm, and this method done by Jomier [89]. Hough Transform algorithm is quite applied to detect the circle shape. The advantage of this algorithm is easy to improve the performance, reliability and storage if we compared with other methods. The S era can be the pupil detection era as Kawai declared in his research. Because of this, equation 4.10 can be used to find (rp) Kawai pupil diameter. Another easiest method that we can through it detect the

pupil diameter, is least square algorithm by applied circle equation or ellipse equation.

In this dissertation, center of mass method detection has been applied [93], by searching and finding the point which is located in the center of the pupil. In this method, searching for the white pixels, while the information that related to the point is storied in matrix. Additional, from point's information, height and length values can found by calculating the difference between the point that which storied in the matrix and the center point. The next step will be used the least-square method to locate the circle radius by applied algebraic equations 4.12, the pupil center is represented by a and b point. The pupil radius will be the maximum value of the radius. For creating a pupil circle, Hough Transform technique has been used.

$$R = \sqrt{(x-a)^2 + (y-b)^2} \tag{4.12}$$

Measurement of pupil diameter described in Figure 4.24.

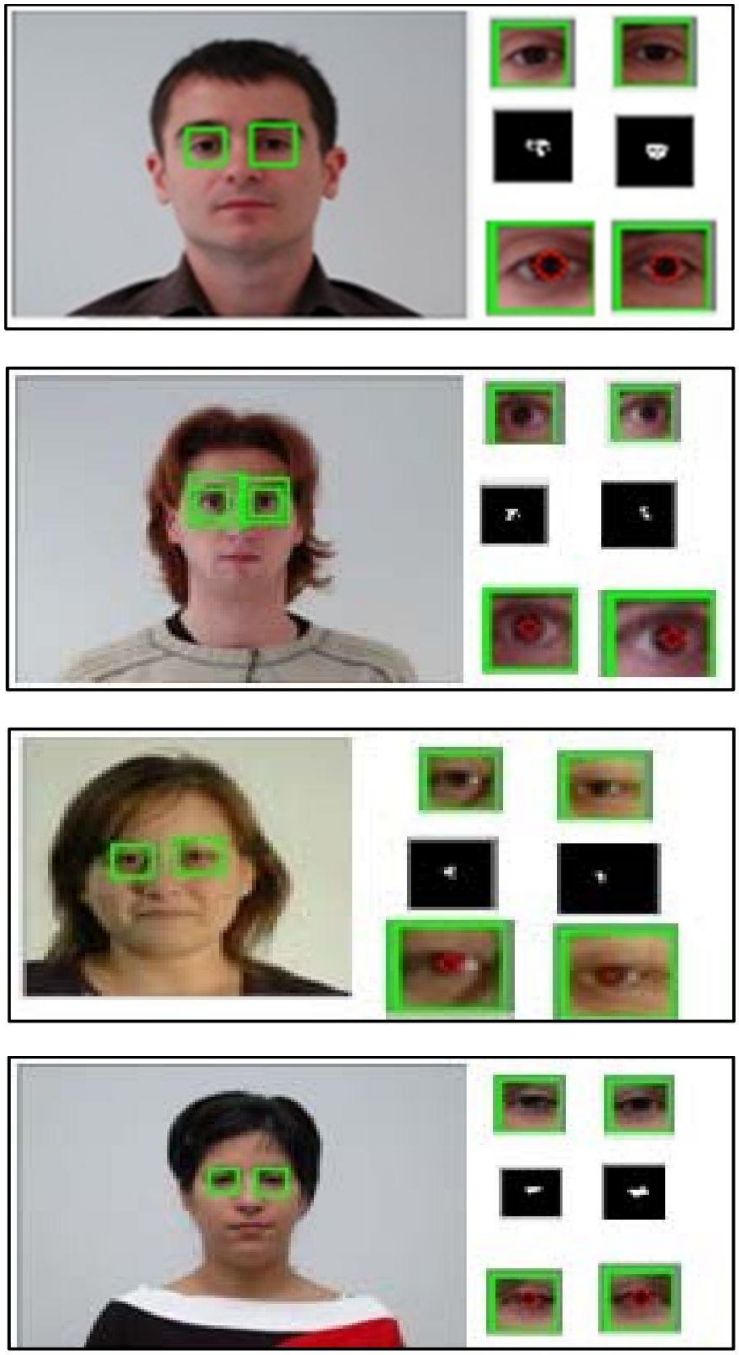

Figure 4.23: Measuring pupil diameter on images from GVTA database.

5 Experimental Results For gaze tracking system

5.1 Setup and Environment

The existing system developed under windows 10 operating system by using visual C ++ programming language and OpenCV library. Open Source Computer Vision (OpenCV) is programming functions library which developed by Intel company for computer vision in real-time aim. It is free open source for both commercial and academic. The images are captured with 'hp laptop webcam', with 640 × 480 pixels of resolution. The system is implemented on an Intel(R) Core(TM) i3 processor at 2.4GHz with 4GB of RAM.

5.2 Test Result of the Eye-tracker

The user needs to sit in front of the computer, keeping 30 to 60 cm of distance from the camera. In the top of the monitor of the computer the camera was placed. After the system starts capturing the video sequence, the user needs to click on the video window to start the eye-tracking. The user needs to click on both eyes in the image to locate the eyes for the tracker in that frame. The points do not need to be centered exactly on the pupils, but points near the iris regions are preferred. Then the system starts to track both

eyes simultaneously in the following frames. An example of experimental setup is shown in Figure 5.1.

Figure 5.1: Experimental Setup (under environment light, not need any additionally light).

To quantitatively analyze the performance, the pupil positions detected by the eye tracker are stored. The video of the user is also stored for hand labeling purposes. Then the pupil positions are hand labeled from the video. The disparity is defined here as the space between the pupil positions detected by the tracker and hand labeled pupil positions. The eye regions are around 75 × 50 pixels when the user sits around 45 cm from the camera.

Although the radius of the iris is different from person to other, it is found that the radius is within 10 to 14 pixels range if the user sits in this distance.

5.3 Experiments

In the first experiment, the user sits around 30 cm from the camera, facing the monitor. After the tracker starts tracking the eyes, the user is free to move his/her head. The data obtained from the eye tracker is recorded in this experiment. The video is also recorded and hand-labeled later. The length of the first video sequence is 500 frames, in a real time, frame rate of 10 frames per second.

Figure 5.2: Right Eye (Row) Positions data that extracted from the eye tracker

The row positions of the right pupil in each frame for both tracked and hand-labeled pupil positions are shown in Figure 5.2. Figure 5.4 presents the row positions of the left pupil in each frame for both tracked and hand-labeled pupil positions. The column positions of both pupils in each frame for both tracked and hand-labeled pupil positions are presented in Figure 5.3 and Figure 5.5. Figure 5.6 and Figure 5.7 show the row and column difference in pixels between the hand labeled pupil positions and the data obtained from the tracker for the right pupil respectively. Figure 5.8 and Figure 5.9 show the row and column disparity in pixels for the left pupil respectively.

Figure 5.10 shows the Euclidean distance (error) between the tracked and hand-labeled pupil positions in each frame for the right eye. The Euclidean distance (error) between the tracked and hand-labeled pupil positions in each frame for the left eye is presented in Figure 5.11.

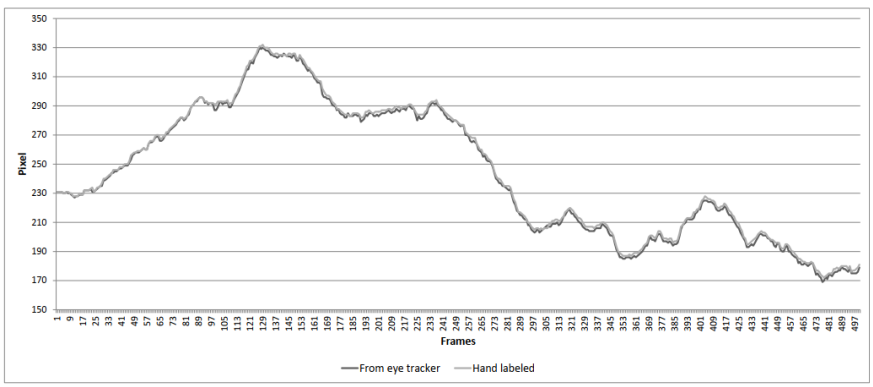

Figure 5.3: Right Eye (column) Positions data that extracted from the eye tracker

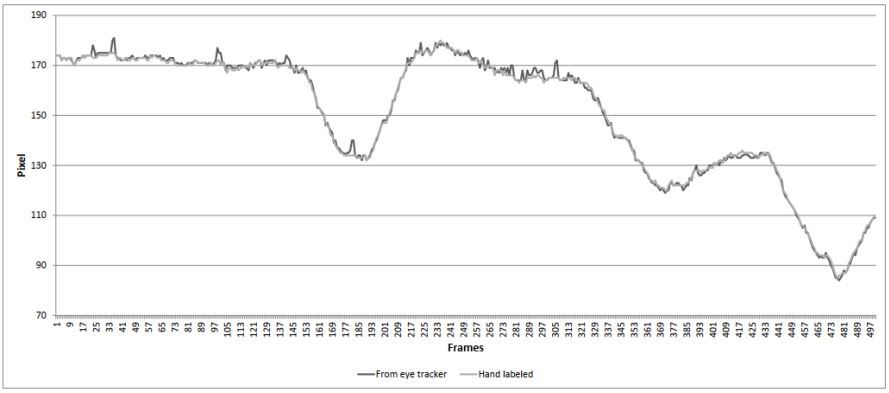

Figure 5.4: Left Eye (column) Positions data that extracted from the eye tracker

Figure 5.5: Left Eye (Row) Positions data that extracted from the eye tracker

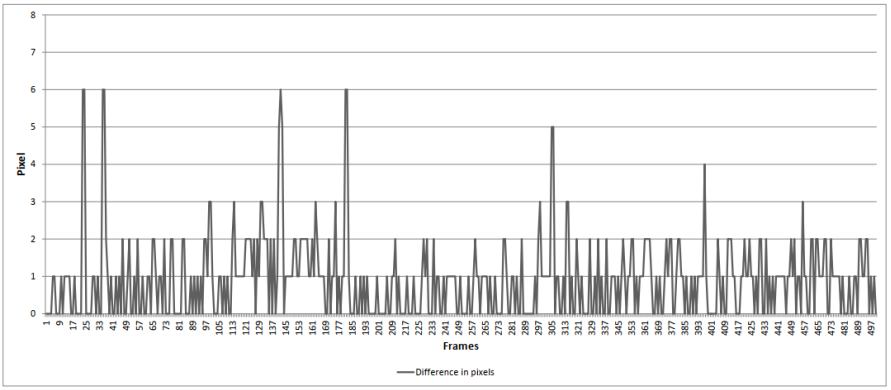

Figure 5.6: The difference Between Hand-labeled and Tracked Right Eye (Row) Positions (Experiment 1) in pixels

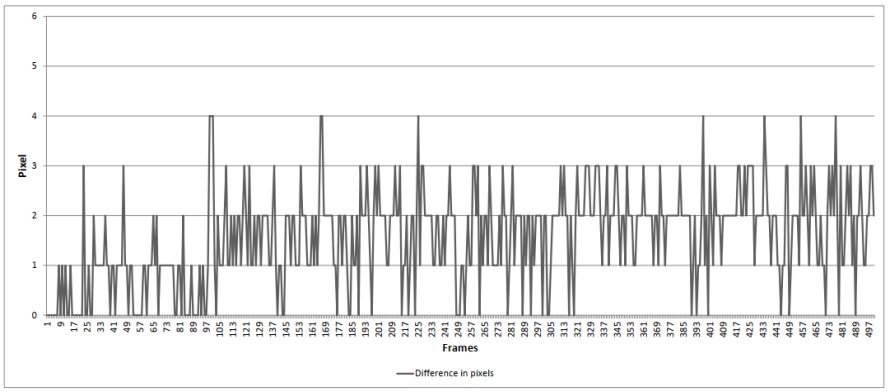

Figure 5.7: The difference Between Hand-labeled and Tracked Right Eye (column) Positions (Experiment 1) in pixels.

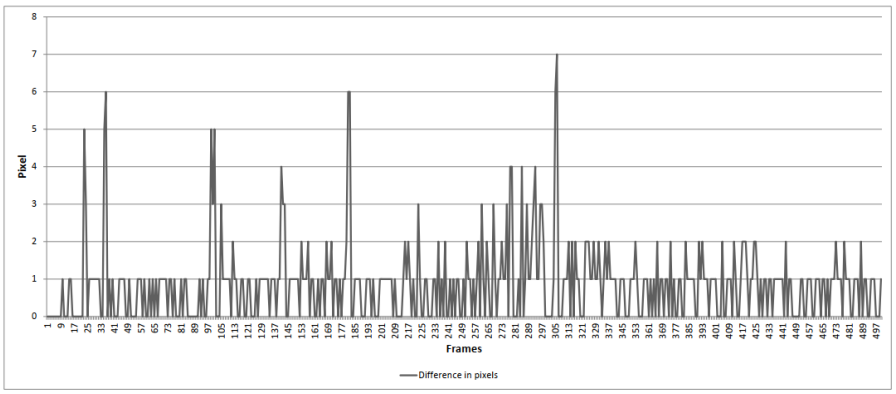

Figure 5.8: The difference Between Hand-labeled and Tracked Left Eye (Row) Positions (Experiment 1) in pixels

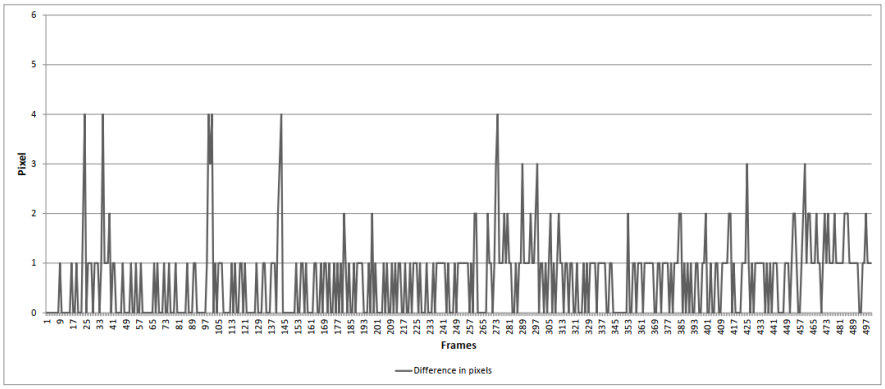

Figure 5.9: The difference Between Hand-labeled and Tracked Left Eye (column) Positions (Experiment 1) in pixels.

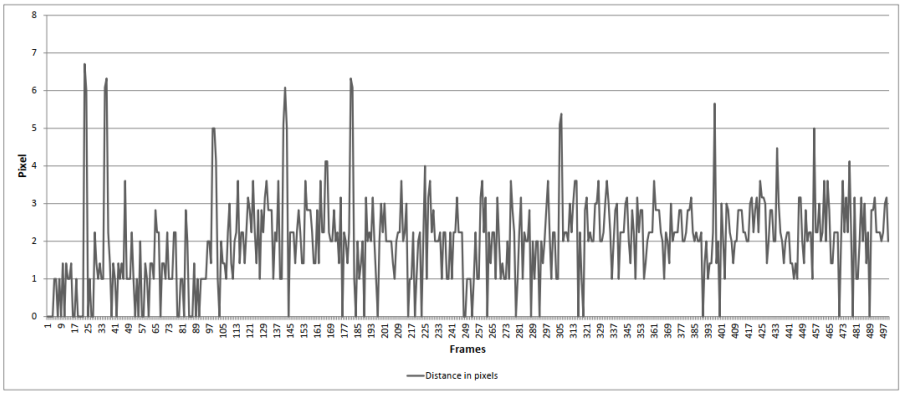

Figure 5.10: The Euclidean distance between the hand-labeled and Right Eye position (Experiment 1) in pixels.

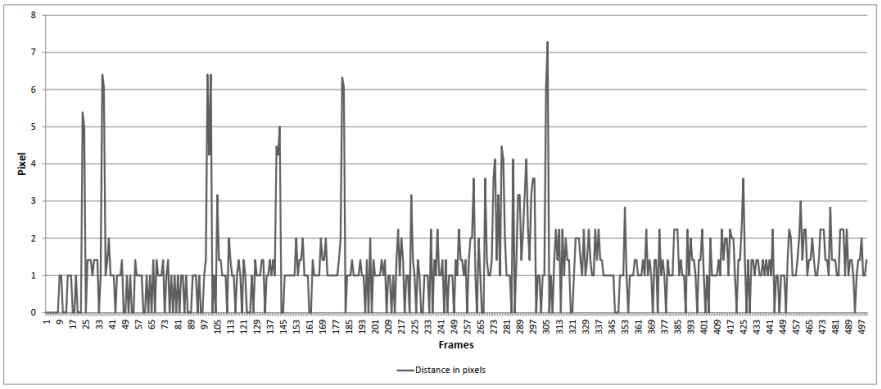

Figure 5.11: The Euclidean distance between the hand-labeled and Left Eye position (Experiment 1) in pixels.

5.4 Result Discussion

For pupil detection, from the region of the eye tracked by the particle filter, the iris is detected first. Then the center of the iris is considered as the pupil center in this eye-tracking design. To determine the parameter of the iris, a half circular mask is used which is described in Section 5.3. To approximate the shape of the irises as circles when the irises are at the corner of the eyes, a three pixels thick half circular annulus mask is used to detect the iris boundary. The pupil is determined by the center of the mask, for which the number of edge pixels located in the annulus region is maximum. As a result, it is possible that the tracked pupil center position can be three pixels apart from the original pupil center. Again, while hand-labeling the pupil center, a deviation of few pixels may occur in both row and column directions. So, if the Euclidean distance (error) between the tracked and the hand-labeled pupil position is in between zero pixel to three pixels, then the pupil is considered as tracked correctly in that frame.

	Mean Absolute Error	Standard Deviation of Errors
Left Eye	1.26	1.11
Right Eye	1.98	1.18

Table 5.1: Deviation (Error) for both left and right eye in Pixels (Experiment 1)

From Table 5.1, it is found that the mean absolute error for the right eye is 1.98 pixels and the mean absolute error for the left eye is 1.26 pixels, which means the average error is less than two pixels for both eyes. This table also presents the standard deviation of errors for both eyes. Standard deviations of errors are 1.18 pixels for right pupil and 1.11 pixels for left pupil.

For Experiment 1, the Table 5.2 shows the frequency of error for both pupils. From this table it is observed that in 86.40% of frames the errors are in between zero to three pixels for the right eye. The Table 5.2 also shows that in 93.60% of frames the error is in between zero to three pixels for the left eye. The histogram of the deviations in pixels and the cumulative percent for Experiment 1 are presented in Figure 5.12 and Figure 5.13 for both eyes respectively. So, from these tables and figures, it is observed that the right pupil is tracked correctly in 86.40% frames, and left pupil is tracked correctly in 93.60% frames.

Pixels Deviation	Right Pupil		Left Pupil	
	Occurrences	Cumulative%	Occurrences	Cumulative %
0	55	11.00%	103	20.60%
1	80	27.00%	194	59.40%
2	129	52.80%	134	86.20%
3	168	86.40%	37	93.60%
4	48	96.00%	13	96.20%
5	9	97.80%	10	98.20%
6	5	98.80%	1	98.40%
7	6	100.00%	7	99.80%
8	0	100.00%	1	100.00%
9	0	100.00%	0	100.00%
10	0	100.00%	0	100.00%
More	0	100.00%	0	100.00%

Table 5.2: Frequency of Deviation (error) for right and left eye pupils (Experiment 1)

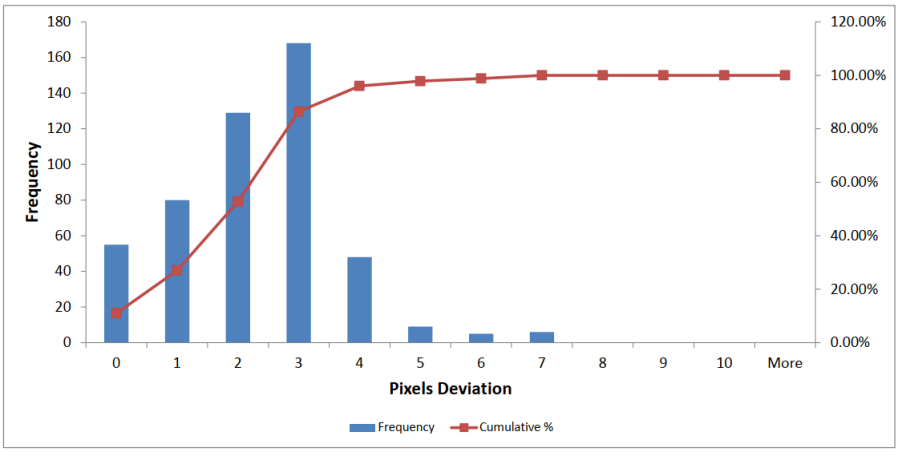

Figure 5.12: Histogram of Deviations (Error) of right pupil (Experiment 1)

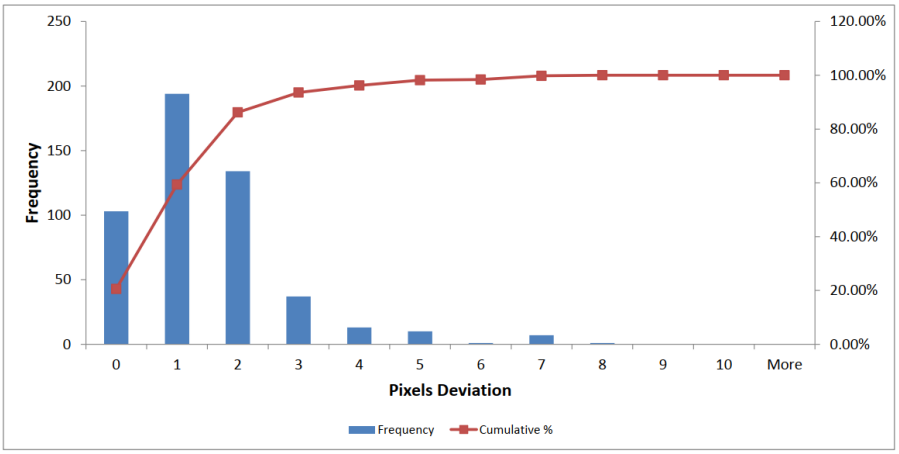

Figure 5.13: Histogram of Deviations (Error) of left pupil (Experiment 1)

5.5 Test Result of the Gaze Estimator

For gaze estimation, the first step is to make the eye-tracker to track both eyes. The gaze calibration is done next. For gaze calibration a white screen is shown to the user with some circles on it. The centers of these circles are predefined screen calibration points. The user needs to look at one circle for 30 frames, as the overall frame rate is around 10 fps, he/she needs to look at one circle for around 3 seconds. Then, the user should look at the next circle and so on. During this calibration process, the user is restricted from moving his/her head. However, the observation is that, although the user tries not to move his/her head, a movement of a few pixels may occur. Therefore it is necessary to find the amount of head movement, which is calculated using the method described in section 4.4. The amount of head movement is

subtracted from the tracked eye positions to get the actual eye movements during the calibration process.

After the head movement compensation, for each eye the 30 eye positions for each calibration point are passed to a function to get the best candidate points described in section 4.5.2. The output of this function is then used to train the KNN classifiers. Two classifiers are used, one for each eye. The label of these training data is a number associated with the predefined position on the screen. After the calibration process, a gaze estimation screen is shown to the user. In each frame, after head movement compensation, both pupil positions are passed to the classifiers. The output of the classifiers is a number associated with the corresponding screen calibration region. If the outputs of two classifiers are different, then a simple voting method is used as described in section 4.5.3. A highlighted rectangle is then presented on the screen to show the gaze region of the user. The three points calibration method for gaze estimation can be shown in Figure 5.14. A point (presented by the black circle) is shown to the user. Gaze estimation is presented by highlighting the background rectangle.

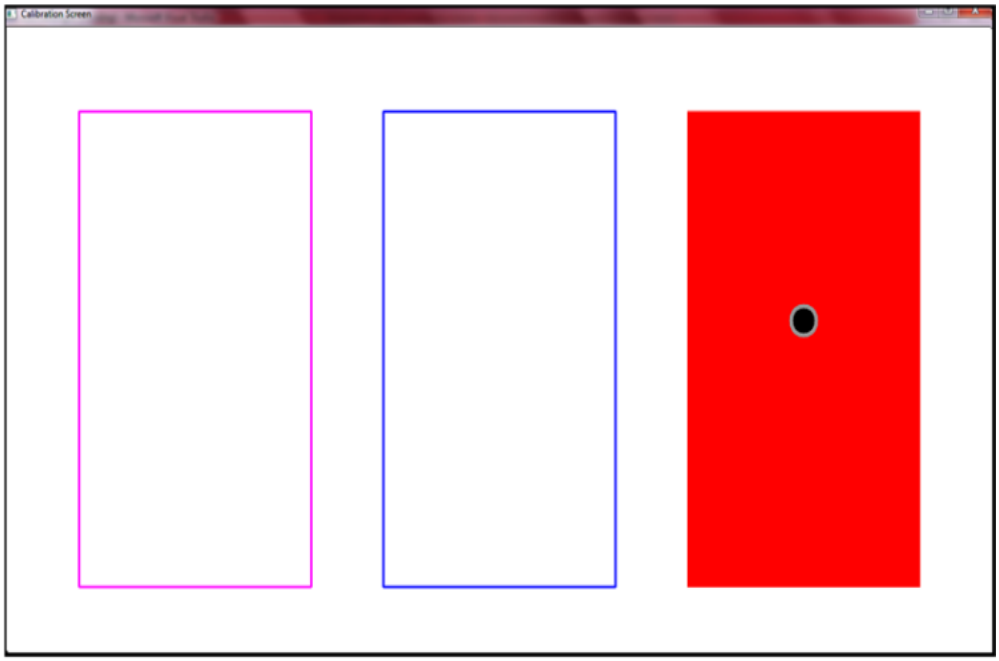

Figure 5.14: Three point gaze calibration, user need to focus on the circle in every part for 3 seconds.

5.6 Experiment on Two Calibration Points

In this experiment, two calibration points are used on the screen. The user needs to look at two points for around three seconds. After the calibration process, the gaze estimation is tested. The mouse pointer is kept on one region on the monitor screen for around three seconds, and the user is asked to look at the mouse pointer. If the rectangle of that region is highlighted then it is considered that the gaze estimation is performed correctly. Otherwise, after three seconds, if the region is not highlighted, the estimation is considered as wrong, and the mouse pointer is moved to another region for

testing. For the gaze estimation, 7-nearest neighbor points are considered by the classifiers.

During one experiment, the mouse pointer is kept on each region to test the gaze estimation. Each region is considered randomly for 30 times during one experiment. This experiment is repeated 10 times, meaning a total of 300 tests are conducted for each region. Table 5.3 shows the result of the experiments on two calibration points.

Experiment number	Estimated Result		Accuracy
	First_part calibration	Second_part calibration	
1	287	13	95.67%
2	9	291	97%
Overall Accuracy			96.33%

Table 5.3: Gaze Estimation Result of Two Calibration Points

5.7 Experiment on Three Calibration Points

In this experiment, three calibration points are used on the screen for gaze calibration. The mouse pointer is kept on one region for around three seconds and the user is asked to look at the mouse pointer. For gaze

estimation, the classifiers consider 7-nearest neighbor points are also considered by the classifiers. If a rectangle of the region over which the mouse is moving to highlighted then it is considered that the gaze estimation is done correctly. Otherwise, after three seconds, if the region is not highlighted, the estimation is considered as wrong, and the mouse pointer is moved to another region for testing. During one experiment, here also the mouse pointer is kept for 30 times by turns on each region to test the gaze estimation. Result of 10 conducted experiments on three calibration points are presented in Table 5.4.

5.8 Experiment on Four Calibration Points

In this experiment, four calibration points are used on the screen for gaze calibration. The same procedure is used for gaze estimation described in section 5.2 and section 5.3. Here also 7-nearest neighbor is used for estimation by the classifiers. During one experiment, here also the mouse pointer is kept on each region to test the gaze estimation. Result of 10 conducted experiments on four calibration points are presented in Table 5.5.

Experiment number	Estimated Result			Accuracy
	First_part calibration	Second_part calibration	Third_part calibration	
1	286	11	3	95.33%
2	9	278	13	92.67%
3	6	10	284	94.67%
Overall Accuracy				94.22%

Table 5.4 : Gaze Estimation Result of Three Calibration Points

.

Experiment number	Estimated Result				Accuracy
	First_part calibration	Second_part calibration	Third_part calibration	Fourth_part calibration	
1	273	18	3	6	91%
2	27	254	14	5	84.67%
3	3	20	260	17	86.67%
4	2	11	9	278	92.67%
Overall Accuracy					88.75%

Table 5.5: Gaze Estimation Result of Four Calibration Points

5.9 Result Discussion

The summary of the gaze estimation results are shown in Table 5.8. From this table it is observed that, the highest accuracy of 97% is achieved for two calibration points, and lowest accuracy of 84.67% is achieved for four calibration points. In case of the gaze estimation, the screen is divided only into vertical regions. This is due to the fact that, when the iris is near the corners of the eye, the vertical movement of the pupil is very small and very hard to detect accurately. So, only the horizontal movements are considered here for gaze estimation.

Calibration Points	Lowest Accuracy	Highest Accuracy	Overall Accuracy
Two calibration method	95.67%	97%	96.33%
Three calibration method	92.67%	95.33%	94.22%
Four calibration method	84.67%	92.67%	88.75%

Table 5.6: Summary of Gaze Estimation Results.

In these experiments, the errors are found mostly when the calibration points of different regions are overlapped. In that case, the KNN classifier tries to calculate the nearest *K* points and gets erroneous neighbors. This observation is proved by the results presented in table 5.4 and table 5.5; most errors are

found in between consecutive regions. It is also found that the accuracy of gaze estimation is higher in the terminal regions. This is because the chances of overlapping gaze calibration points are less than those in the middle regions. Another cause is that if the pupil positions are beyond the terminal gaze calibration points, the closest regions are only the terminal regions.

Another cause of error is that sometimes the eye tracker provides erroneous people data. As the gaze calibration points are very close, when the iris movements are too small, for one or two pixels, then the tracker sometimes provides the pupil positions one or two pixels apart from the original pupil positions, which results in erroneous gaze estimation. In these experiments, it is also observed that, if the gaze calibration points are significantly apart, then the gaze estimations are nearly 95% correct. However, some errors still occur due to the wrong eye tracking. But, the gaze estimator faces difficulties when there are large overlaps between the gaze calibration points. It is worth noting that the extent of head movement is a crucial factor for gaze estimation - even a movement of 5 pixels might adversely affect the performance.

6 Experiments and the results for emotion detection

In our experiment, 37 volunteers participated (33 males and 4 females), their ages ranged between 19 and 45 years old. And they were chosen according to appropriate criteria, to avoid any future mistakes. Because of eye gaze data reading was slow (low capture rates (less than 59%), the data from the volunteer number 18 were excluded. All persons participating in this experiment had normal vision and did not have any pathological problems in the eye. All volunteers had at least 6 years of experience in using a computer, and watching movies and videos was also part of their daily lives. The mood was assessed by the volunteers themselves, and it was reported that 73% were in a normal mood and 20% were in a state of happiness, while the state of sadness was only about 3%.

6.1 Stimuli

Because of the reasons that mentioned in [94], [95] and [96], short videos have been chosen, which have proven effective in detecting feelings. In [97] a universal list of emotional clips are available. However, all of the videos which are used in this experiment were in English language, the translation of the Arabic language clip was not available in order to avoid any future mistakes that might happen, as the eye movement might interfere with the

translation, that leads to inaccurate readings. One of the nice and one disgusting Arabic film clips, "Heidi" and "Nobody's Boy: Remi," were chosen from the classic animations collection.

The lovely film depicts a scene of wealthy nature and scenery where the girl Heidi enjoyed the fresh air in the mountain, perceived the sun's warmth on her face, and encountered goatherd Peter. The disagreeable film featured a sequence in which Remi discovered the awful truth: Vitalis is gone. Such portions of animated films are selected because the music language can quickly be modified without any impact on lip synchronicity. The total length for each video clip was 150 seconds and the soundtracks are present. The conditions were provided for both topics.

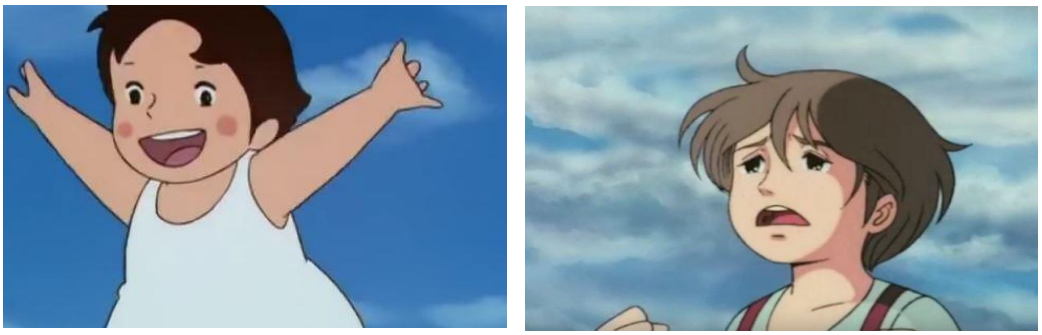

Figure 6.1: Two film clips pleasant in the left side while the unpleasant in the right side.

6.2 Procedure

Before the registration of research participants, permission was sought. Subjects on this analysis have been trained and independently checked. Before the experiment began, topics we're told to show clips (no description of the particular form of the emotional condition) and to claim that all the video clips should appear in Arabic. They were instructed to view the video because they generally went home and were advised there are concerns regarding the film clip and their thoughts to address afterward. Four-point calibration technique used for eye movement's purpose for each case, when the calibration technique has been successfully finishing, the experiment was started. Some instructions on the screen have written; these instructions asked the volunteers to clear them thoughts and minds and just focusing on the clip after this clip began. In order to allow the volunteer to take liberty when watching the video, the experimenter had to leave the room, to avoid any distraction for volunteer. The volunteer was asked after each film that if he/she watched this film before. Following this, a post-questionnaire and each subject were requested to conclude the rating of feeling emotional measures from' null' (point: 0) to ‘extremely' (point: 10).

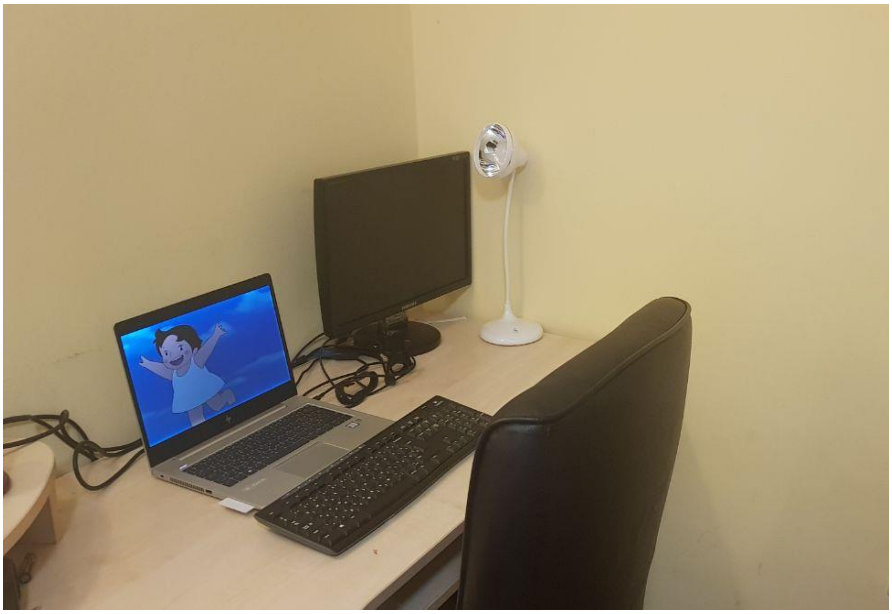

Figure 6.2: Experiment devices

6.3 Analysis

We are creating a segment by using the webcam and the method which is mentioned in Section 4.6 for analyzed the pupil size. For each case, a row of data has been generated. Nonetheless, after exporting raw data and before processing, left and right pupils, the data was filtered by eliminating incorrect reading points of the pupil's height, when the eye wasn't properly detected (enthusiasm rating validity varying from 1 to 4) and the eye tracker couldn't catch the eyes. The period for first fixation, fixation moment, fixation time and the number of fixations were derived from raw data

individually. T-tests were performed in an agreeable and uncomfortable stimulus for pupil dimensions, fixation measures.

6.4 Experiment Emotion

The rating model was used to examine Arabic dubbed syllables in this study. Where the participants rate the pleasant feelings and unpleasant feelings while he/she watched the dubbed clip, each participant inquired to give his opinion about the film clips and rate it from (0-10) scale. The findings of the self-report reveal that the friendly video clip has an effectiveness score of 7.73 out of 10 and a negative clip of 6.37 out of 10.

T-test has been added to assess whether each clip has a biased emotional impact. The good film was deemed to be substantially better than the unwelcoming video ($t=6.6$, $p<0.001$).

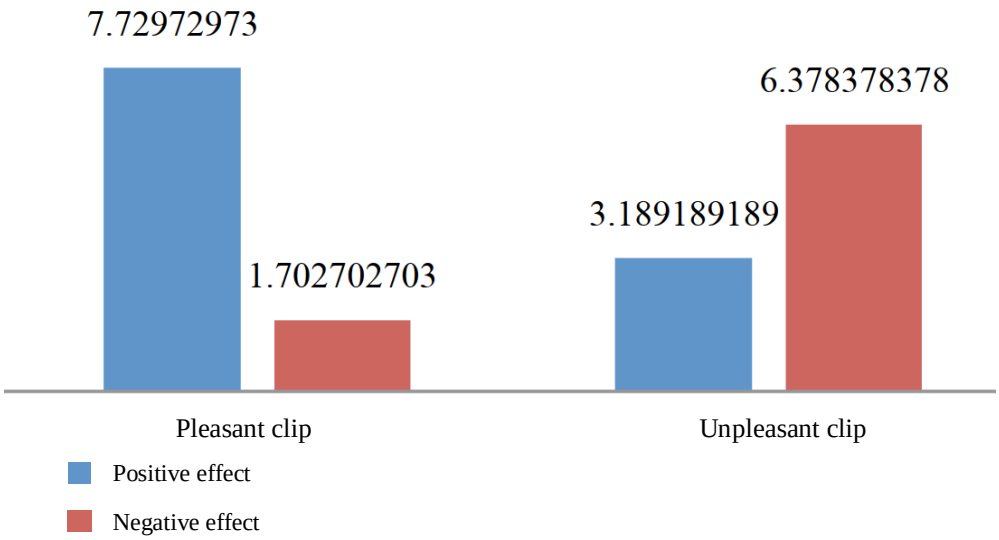

Figure 6.3: The average ranking for experienced unpleasant clip (negative and positive effect) and pleasant clip (negative and positive effect).

The findings revealed that 55% of the volunteers had seen an acceptable clip before, with 16 % having seen an awkward clip before the research, were even questioning if each clip had been seen previous to the research project.

6.5 Pupil size

The pupil measurements of the objects vary from 2.21 to 4.93 millimeters in the time during which the nice and disagreeable video clips are shown. As shown in Figure 6.4, it has been observed that the average pupil size in unpleasant excitement is approximately higher than pleasant excitement. In detail, the average pupil's size of the volunteer in unpleasant arousal was

higher than nice clip arousal by 89.6%. The average pupil size for 11% of volunteers was numerically distant from the other cases.

While this finding was fair, we also notice that 25% of the clips were scoring the satisfying clip for a good upshot 4 of 10 and 50 % of the clips ranking it as an unfavorable impact less than four in 10 to gain further comprehension when it comes to chance or for certain purposes, so we also noticed that they have an unfavorable clip.

The data from this experiment did not show a large difference in the size of the pupil for each of the pleasant and unpleasant video clips, and these results are in great agreement with the results that were proven in [98] and [99].

6.6 Fixation

In Table 6.1 we summarized all the results for both all duration for first fixation (in second), the mean for first fixation (in second), the duration for the fixation in second and count of fixation which have been investigated. In the pleasant clip, the first fixation time in seconds was between from 0 to 2.79 and an average was 0.17. While in unpleasant clip ranged from 0 and 5.13 an average was 0.49. However, the results what we received from our experiments did not show a huge difference between both unpleasant and pleasant video clips.

	Unpleasant clip Mean	**Pleasant clip Mean**
Fixation for first-time in seconds	0.47	0.19
Duration for the first fixation in second	0.24	0.28
Duration of Fixation in second	0.4	0.41
Count of fixation	342.6	318.9

Table 6.1: Using fixation measurements for measure the mean both of pleasant and unpleasant clips.

First fixation duration of effect the unpleasant was ranged from 0.11 to 1.54 and the average one was 0.24, while in the pleasant was ranging from 0.021 to 1.441 and the average was 0.29. This led us that there is no huge different between the unpleasant and unpleasant videos.

There was a slight difference in the mean of average fixation duration between the unpleasant and pleasant clips. Mean of the fixation duration in unpleasant clips was ranged from 0.17 to 0.57, while in the pleasant clips was ranged from 0.14 to 0.79, this is what shows in Figure 5.6.

T-test showing a significant difference between the unpleasant and pleasant clips, the test paired two samples for fixation duration with mean and the different was ($t=2.2$, $p=0.04$).

The count means of the fixation was ranged from 185 and 474 in the pleasant clip while was ranged from 243 and 505, as it's shown in Figure 6.6. T-test showing a difference between the unpleasant and pleasant clips, the test paired two samples for fixation count, and the difference was ($t=2.1$, $p=0.03$).

So far as we know, the standard fixation in the dynamic video for the reading process purpose have been proved in few studies and this shown in [100], with very long fixations the cognitive control can have a direct impact over saccades timing. Potentially fixation is related to cognitive control that associated with process of emotion during watching the videos.

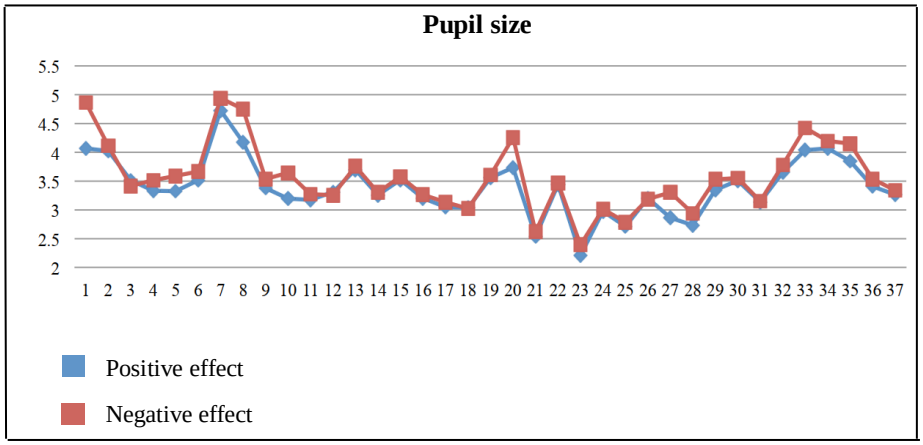

Figure 6.4: Pupil size in pleasant and unpleasant clips

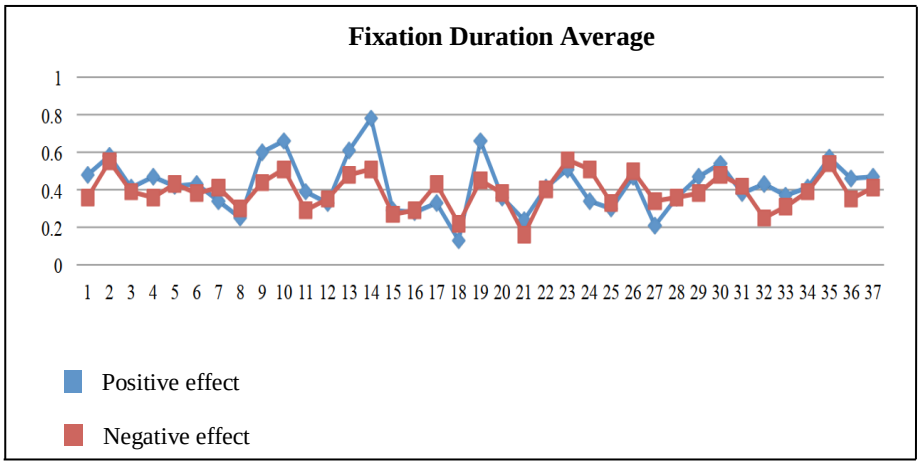

Figure 6.5: Average of fixation duration for both pleasant and unpleasant clips

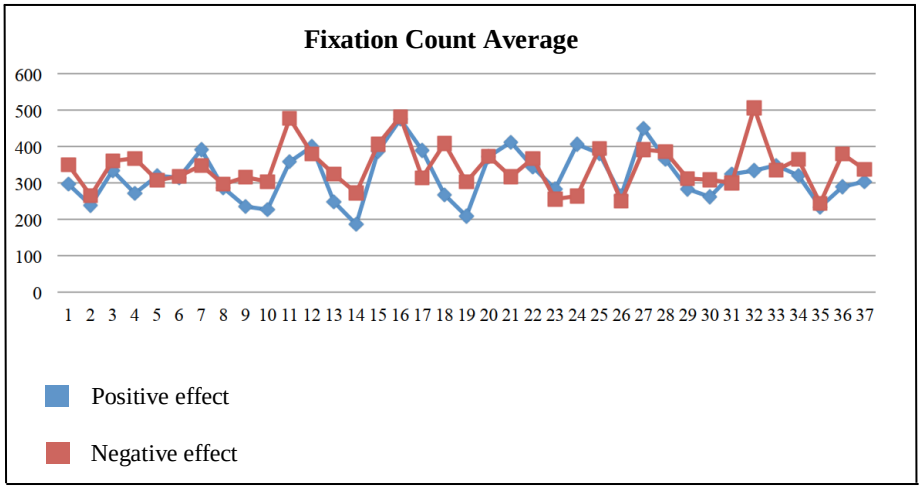

Figure 6.6: count of fixation in pleasant and unpleasant clips

Numbers of light	Numbers of camera	Head_pose estimation	Gaze_type	Calibration	Comments	Reference
0	1	Extra unit not needed	LoS/LoG	Fully	Web camera	[101],[102]
0	1	Approximate solution	LoG		Additional markets, iris radius, parallel screen	[103]
1	1 + 1 PT camera	Yes	PoR	Fully	Mirror	[104]
2(3)	2	Yes	PoR	Fully	Extra lights used during implementation	[105],[106]
4	1 + 1 PT camera	Yes	PoR		PT camera used during implementation	[107],[108]
0	2	Yes	PoR		3D face model	[109]
0	1	Yes	PoG	Fully	Built in web camera + image processing techniques	Our work

Table 6.2: Comparison of visual assessment (gaze_estimation)methods.

Measurement	Ability to Detect Distraction			Pros	Cons
	Visual	Cognitive	Visual and Cognitive		
Driving Performance	Y	N	N	Ability to indicate the effect of driving distraction	Complementary self-reporting requirement for high-resolution results
Physical Measurements	Y	Y	N	Distraction types can be distinguish	Combined type of distraction cannot be distinguish
Biological Measurements	Y	Y	Y	Ability to measure both of distraction types (cognitive and visual distraction)	Intrusiveness
Subjective Reports	N	Y	N	Ability to distinguish underlying mechanism of distraction	Requires input of an expert
Hybrid Measurements	Y	Y	Y	-Higher accuracy for discriminating types of distractions - Able to complement the blind spots of other methods	Synchronization of multiple source of data with different sampling rate

Table 6.3 Summary of measurement techniques [110]

Methods	Ages of the Volunteers	Procedure	Comment
Elaine Fox, Andrew J. Calder, Andrew Mathews method 2009 [23]	18 and 32 years old	- Selected on the basis of levels of trait anxiety rather than state anxiety because they wanted to investigate the more enduring effects of anxiety on emotion perception -Using the direction of the eye for emotional detection -Two letter T and L for watching on it and the user ask to follow them on the monitor - Distance The stimuli were presented on a 43.18 cm far from monitor	-Additional equipment was needed PsyScope software (that just work in Apple Macintosh) also It will only work under systems up to 10.14.x (Mojave). -Ask the user to press on the letter increasing the stress for the user and we will not have the correct result also using just the direction of the gaze and do not pay the attention for the pupil diameter also will affect the right result
Lassalle, A. and Itier, R.J., 2015 [111]	mean age =19.3 years	-All volunteers were females - All participants gave informed consent and were unaware of the goal of the study. - The stimuli were presented on a 15-inch monitor placed 60 cm away from the participants and controlled with E-Prime software. -A white fixation cross was presented on a black screen for 600 ms. - All volunteers respond as quickly and as accurately as possible by pressing the “c”	-Depended on one gender is not good idea -Did not inform the Volunteers from the aim of the experiments maybe will this confuse them and increasing the stress feeling - E-Prime software tools was used here and this an additional equipment - Using a black screen may be

		key on a standard keyboard with their left index finger and the “m” key with their right index finger for left and right targets	will affect the result of the experiments -Ask the volunteers to quick response and press some bottoms, this will create for the Volunteers some stress and this will increase the bad emotion
Our research	19 and 45 years old	-Using webcam -Not extra light need -Using image processing techniques -Ask the Volunteers to watch the clips in their language -No additional equipment need or third software program -Using the direction and also the diameter of the pupil for more accurate results. -Choice the volunteers randomly and in different ages. -Give free head movement for the user -Inform the Volunteers from the aim of this study	-For future work more additional status will be detection depended on this study -In our study we create a cheapest gaze tracking system , which using only usb or built in webcam

Table 6.4Comparison of emotional detection methods using gaze tracking (advantage and disadvantage)

7 Conclusions

- Three significant issues were classified: The exogenous against endogenous nature of the biasing attention to emotional motive, the deliberate capture time course of emotional visual scenes, and finally the type of the influential content (unpleasant and pleasant) that is capable of attracting attention selectively. The result has shown that both subsequent intentional participation and initial orienting are tended toward both unpleasant and pleasant emotional stimuli to attend to a neutral image displayed simultaneously.

- Throughout this experiment, we have tested whether eye movement can be a useful instrument to distinguish various adaptive states in a nice and unpleasant stimulation. Taking the advantages of the pupil size (change in the pupil diameter) and analysis the direction of the gaze to improving the performance of our methods give us an incredible result quality with no extra equipment with normal environment light.

- Outcomes of our results from pupil diameter recommend that practice of mental meditation modulate the reaction of the autonomic sensory system, pupillary response reflection react to unpleasant images and quick physiological retrieval to base lines levels, recommending that pupillometry could be utilized to survey the potential medical advantages of these practices in patients.
- The first fixation, the first fixation length, the fixation duration and the number of fixations have been examined. Throughout this analysis, we observed that time for first fixation was not substantially different for both friendly and negative, surprisingly, the fixation length was important. Furthermore, the number of fixations between possible and unlikely sensations was substantially different. Fixation appears in the current study to have a racist and negative emotional impact. These findings indicate that further work is needed for high quality resolution of camera with IR diagnose various emotional conditions.
- The dissertation touched on a topic head movement positions, head movement detection has been seen as a characteristic method for interaction. It very well may be utilized as an elective control strategy and gives openness to users when utilized in human-PC interface arrangements, by applying some image processing techniques it is

possible to build software that can help users to interact with the computer without any physical connection, also this is will give a huge opportunity for a people with a permanent disability to adapt with them life.

- The result that comes from eye detection, a software can be built from these outcomes by extract the eye location and detecting the blinks that used for the safety purpose, by taking the most important features of the eye and attached the result to an alarm device that will be alarming the user, the most and famous example for this is simulation system in detecting drowsiness that which used in the cars and the trucks.
- The present dissertation was constrained in the absence of consistently producing emotions in Arabian films. Future studies will study Arabic film beyond that and offer specific emotions such as rage disgust and shock. Future research should take account of the scale and focuses of the eyes coupled with the signals of Brian Machine (BCI) as they sense emotional situations and assess the cognitive mechanism that takes place in the field of interest. This will be important to demonstrate in real-time charts of brain-behavior the results of multiple emotional conditions.
- The current era is recognized as the era of digitization, artificial

intelligence and machine learning where human-computer interaction has increased to a great extent and hence has become part of our daily life. One of the key aspects of human-computer interaction is non-verbal communication where eyes play a highly significant role. Eyes enable the user to recognize emotions and interpret the meaning of expressions. In such case, accurate detection of eye gaze plays an important role in analyzing the emotions.

- The dissertation aimed to analyze the possibility or likelihood of creating a system which used a regular web camera to follow and track the gaze along with detecting the major emotional cases. Results of the research have proved it correct and possible with cheaper equipment. Furthermore, this research has proved that accurate contactless measurement of gaze tracking, in RGB video, is possible in real life situations. This confidence provides the required evidence for the commercial exploitation of such systems that are easy to implement hence can be used in optimizing the routine activities.

- The dissertation has also identified certain outstanding issues which need to be addressed in future studies. For example, it is being noticed while conducting the experiments that length of hair of the

participants influenced accuracy level and estimation. Furthermore, it has been observed that gaze estimation is dependent of lighting conditions; given that in the case of dark environment, results are not good enough. Furthermore, the influence of head movement has not been undertaken as a key influencer; however it might have an impact and can be considered in the future studies examining if it causes imprint deterioration or not.

- The study has identified that future studies need to consider use of such a system which facilitates combined videos and aids in creating the complex versions by taking into account different feelings to the volunteer (such as embarrassment, envy, gratitude, guilt, jealousy, pride, remorse, shame, and worry etc.). This will help in correlating the gaze with these emotions. Furthermore, the new technologies and eye tracking applications that are under research can further be studied in order to understand and determine their usefulness for psychologists and emotion evaluators to analyze more of the human behavior, decision making and attention and efficiency of the user towards any specific stimuli.

References

1. Pantic, M., Pentland, A., Nijholt, A. and Huang, T.S., 2007. Human computing and machine understanding of human behavior: A survey. In Artifical Intelligence for Human Computing (pp. 47-71). Springer, Berlin, Heidelberg.
2. Wang, Q., Boccanfuso, L., Li, B., Ahn, A.Y.J., Foster, C.E., Orr, M.P., Scassellati, B. and Shic, F., 2016, March. Thermographic eye tracking. In Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications (pp. 307-310).
3. Azevedo, R., Taub, M., Mudrick, N., Farnsworth, J. and Martin, S.A., 2016. Interdisciplinary research methods used to investigate emotions with advanced learning technologies. In Methodological advances in research on emotion and education (pp. 231-243). Springer, Cham.
4. Frank, T. and Johnston, O., 1981. Disney Animation–The Illusion of Life. Abbeville Pub.
5. Diefendorf, A.R. and Dodge, R., 1908. An experimental study of the ocular reactions of the insane from photographic records. Brain, 31(3), pp.451-489.
6. Bayless, S.J., Glover, M., Taylor, M.J. and Itier, R.J., 2011. Is it in the eyes? Dissociating the role of emotion and perceptual features of emotionally expressive faces in modulating orienting to eye gaze. Visual cognition, 19(4), pp.483-510.
7. Hood, B.M., Willen, J.D. and Driver, J., 1998. Adult's eyes trigger shifts of visual attention in human infants. Psychological Science, 9(2), pp.131-134.
8. Lance, B. and Marsella, S., 2010. The expressive gaze model: Using gaze to express emotion. IEEE Computer Graphics and Applications, 30(4), pp.62-73.
9. Shepherd, S.V., 2010. Following gaze: gaze-following behavior as a window into social cognition. Frontiers in integrative neuroscience, 4, p.5.
10. Cowan, D.G., 2015. The empathic gaze and how to find it: eye-gaze behaviour to expressions of emotion.
11. Mehrabian, A., 1967. Attitudes inferred from non-immediacy of verbal communications. Journal of Verbal Learning and Verbal Behavior, 6(2), pp.294-295.
12. Farroni, T., Csibra, G., Simion, F. and Johnson, M.H., 2002. Eye contact detection in humans from birth. Proceedings of the National academy of sciences, 99(14), pp.9602-9605.

13. Baron-Cohen, S., 1995. Theory of mind and face-processing: How do they interact in development and psychopathology?.
14. Hood, B.M., Willen, J.D. and Driver, J., 1998. Adult's eyes trigger shifts of visual attention in human infants. Psychological Science, 9(2), pp.131-134.
15. Abdessalem, H.B., Chaouachi, M., Boukadida, M. and Frasson, C., 2019, June. Toward Real-Time System Adaptation Using Excitement Detection from Eye Tracking. In International Conference on Intelligent Tutoring Systems (pp. 214-223). Springer, Cham.
16. Niemic, C., 2004. Studies of Emotion: A Theoretical and Empirical Review of Psychophysiological Studies of Emotion.
17. Diefendorf, A.R. and Dodge, R., 1908. An experimental study of the ocular reactions of the insane from photographic records. Brain, 31(3), pp.451-489.
18. Oyster, C.W., 1999. The human eye. Sunderland, MA: Sinauer.
19. Hood, B.M., Willen, J.D. and Driver, J., 1998. Adult's eyes trigger shifts of visual attention in human infants. Psychological Science, 9(2), pp.131-134.
20. Shepherd, S.V., 2010. Following gaze: gaze-following behavior as a window into social cognition. Frontiers in integrative neuroscience, 4, p.5.
21. Cowan, D.G., 2015. The empathic gaze and how to find it: eye-gaze behaviour to expressions of emotion.
22. Kolb, H., Fernandez, E. and Nelson, R., 1995. Photoreceptors--Webvision: The Organization of the Retina and Visual System.
23. Fox, E., Mathews, A., Calder, A.J. and Yiend, J., 2007. Anxiety and sensitivity to gaze direction in emotionally expressive faces. Emotion, 7(3), p.478.
24. Putman, P., Hermans, E. and Van Honk, J., 2006. Anxiety meets fear in perception of dynamic expressive gaze. Emotion, 6(1), p.94
25. Tipples, J., 2006. Fear and fearfulness potentiate automatic orienting to eye gaze. Cognition & Emotion, 20(2), pp.309-320.
26. Adams Jr, R.B. and Kleck, R.E., 2002, January. Differences in perceived emotional disposition based on static facial structure. In Poster presented at the annual meeting of the Society for Personality and Social Psychology, Savannah, GA.
27. Adams Jr, R.B. and Kleck, R.E., 2005. Effects of direct and averted gaze on the perception of facially communicated emotion. Emotion, 5(1), p.3.

28. Roelofs, K., Putman, P., Schouten, S., Lange, W.G., Volman, I. and Rinck, M., 2010. Gaze direction differentially affects avoidance tendencies to happy and angry faces in socially anxious individuals. Behaviour research and therapy, 48(4), pp.290-294.

29. Hietanen, J.K. and Leppänen, J.M., 2003. Does facial expression affect attention orienting by gaze direction cues?. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(6), p.1228.

30. Tipples, J., 2006. Fear and fearfulness potentiate automatic orienting to eye gaze. *Cognition & Emotion*, *20*(2), pp.309-320.

31. Putman, P., Hermans, E. and Van Honk, J., 2006. Anxiety meets fear in perception of dynamic expressive gaze. *Emotion*, *6*(1), p.94.

32. De Lemos, J., Jensen, O.B. and Sadeghnia, G.R., De Lemos Jakob, 2010. *System and method for calibrating and normalizing eye data in emotional testing*. U.S. Patent Application 12/170,059.

33. *iView X. System Manual. Version 2.4. December 2009*

34. Hansen, D.W. and Ji, Q., 2009. In the eye of the beholder: A survey of models for eyes and gaze. IEEE transactions on pattern analysis and machine intelligence, 32(3), pp.478-500.

35. Soltany, M., Zadeh, S.T. and Pourreza, H.R., 2011. Fast and accurate pupil positioning algorithm using circular Hough transform and gray projection. In *International Conference on Computer Communication and Management* (pp. 556-561).

36. George, A. and Routray, A., 2016. Fast and accurate algorithm for eye localisation for gaze tracking in low-resolution images. *IET Computer Vision*, *10*(7), pp.660-669.

37. Hansen, D.W. and Ji, Q., 2009. In the eye of the beholder: A survey of models for eyes and gaze. IEEE transactions on pattern analysis and machine intelligence, 32(3), pp.478-500.

38. Periketi, P.R., 2011. Gaze estimation using sclera and iris extraction.

39. Gordon, G.M., Austin, J.S., Sklar, A.L., Feuer, W.J., LaGier, A.J. and Fini, M.E., 2011. Comprehensive gene expression profiling and functional analysis of matrix metalloproteinases and TIMPs, and identification of ADAM-10 gene expression, in a corneal model of epithelial resurfacing. *Journal of cellular physiology*, *226*(6),

pp.1461-1470.

40. Park, K.R., Lee, J.J. and Kim, J., 2002, November. Facial and eye gaze detection. In *International Workshop on Biologically Motivated Computer Vision* (pp. 368-376). Springer, Berlin, Heidelberg.
41. Ohno, T. and Mukawa, N., 2004, March. A free-head, simple calibration, gaze tracking system that enables gaze-based interaction. In Proceedings of the 2004 symposium on Eye-tracking research & applications (pp. 115-122).
42. Zhao, C., Fan, B., Hu, J., Tian, L., Zhang, Z., Li, S. and Pan, Q., 2017, October. Pose estimation for multi-camera systems. In *2017 IEEE International Conference on Unmanned Systems (ICUS)* (pp. 533-538). IEEE.
43. Chang, S.A., Tyler, R.S., Dunn, C.C., Ji, H., Witt, S.A., Gantz, B. and Hansen, M., 2010. Performance over time on adults with simultaneous bilateral cochlear implants. *Journal of the American Academy of Audiology*, *21*(1), pp.35-43.
44. Xiong, C., Huang, L. and Liu, C., 2014, August. Gaze estimation based on 3D face structure and pupil centers. In *2014 22nd International Conference on Pattern Recognition* (pp. 1156-1161). IEEE.
45. Nagamatsu, T., Sugano, R., Iwamoto, Y., Kamahara, J. and Tanaka, N., 2010, March. User-calibration-free gaze tracking with estimation of the horizontal angles between the visual and the optical axes of both eyes. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (pp. 251-254).
46. Jones, M., and Viola, P., 2003. Fast multi-view face detection. Mitsubishi Electric Research Lab TR-20003-96, 3(14), p.2.
47. Sirohey, S., Rosenfeld, A. and Duric, Z., 2002. A method of detecting and tracking irises and eyelids in video. Pattern Recognition, 35(6), pp.1389-1401.
48. https://imotions.com/blog/eye-tracker-prices/ Accessed 20 March 2020.
49. https://www.oculus.com/?locale=en_US Accessed 20 March 2020
50. Eyegaze (formerly LC Technologies) (https://eyegaze.com/) .Accessed 20 March 2020.
51. Pupil Labs (https://pupil-labs.com/). Accessed 20 March 2020
52. FOVE 0 Eye tracking vr devkit (https://www.getfove.com/). Accessed 20 March 2020
53. Patney, A., Kim, J., Salvi, M., Kaplanyan, A., Wyman, C., Benty, N., Lefohn, A.

and Luebke, D., 2016. Perceptually-based foveated virtual reality. In ACM SIGGRAPH 2016 Emerging Technologies (pp. 1-2).

54. Ergoneers (https://www.ergoneers.com/en/). Accessed 20 March 2020.
55. smivision (http://www.smivision.com/). Accessed 20 March 2020.
56. Interactive-minds (https://interactive-minds.com/). Accessed 20 March 2020.
57. Imotions (https://imotions.com/). Accessed 20 March 2020.
58. institut-fw (http://institut-fw.de/) Accessed 20 March 2020.
59. https://eyezag.com/ Accessed 20 March 2020
60. xLabs: EyesDecide (https://www.eyesdecide.com/) . Accessed 20 March 2020
61. Höffner, S., 2018. Gaze Tracking Using Common Webcams.
62. Eyelike (https://github.com/trishume/eyeLike). Accessed 20 March 2020.
63. Timm, F. and Barth, E., 2011. Accurate eye centre localisation by means of gradients. *Visapp*, *11*, pp.125-130.
64. Opengazer (http://inference.org.uk/opengazer/) Accessed 20 March 2020.
65. CVC eyetracker (https://github.com/tiendan/OpenGazer). Accessed 20 March 2020.
66. Pygaze (http://www.pygaze.org/). Accessed 20 March 2020
67. Dalmaijer, E.S., Mathôt, S. and Van der Stigchel, S., 2014. PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior research methods*, *46*(4), pp.913-921.
68. Gazr (https://github.com/severin-lemaignan/gazr) Accessed 20 March 2020.
69. ROS (https://www.ros.org/) Accessed 20 March 2020.
70. Lemaignan, S., Garcia, F., Jacq, A. and Dillenbourg, P., 2016, March. From real-time attention assessment to "with-me-ness" in human-robot interaction. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 157-164). Ieee.
71. Krafka, K., Khosla, A., Kellnhofer, P., Kannan, H., Bhandarkar, S., Matusik, W. and Torralba, A., 2016. Eye tracking for everyone. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2176-2184).
72. Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J. and Hays, J., 2016, January. Webgazer: Scalable webcam eye tracking using user interactions. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence-IJCAI 2016*.

73. Deepgaze (https://github.com/mpatacchiola/deepgaze) Accessed 20 March 2020.
74. Patacchiola, M. and Cangelosi, A., 2017. Head pose estimation in the wild using convolutional neural networks and adaptive gradient methods. *Pattern Recognition*, *71*, pp.132-143.
75. Vezhnevets, V. and Degtiareva, A., 2003, September. Robust and accurate eye contour extraction. In *Proc. Graphicon* (pp. 81-84).
76. Tambe, S.B., Kulhare, D., Nirmal, M.D. and Prajapati, G., 2013. Image processing (IP) through erosion and dilation methods.
77. Test, B.I., CASIA iris image database, version 1.0," 2010.
78. Nummiaro, K., Koller-Meier, E. and Van Gool, L., 2003. An adaptive color-based particle filter. *Image and vision computing*, *21*(1), pp.99-110.
79. Sadri, M., Kangarloo, K. and Farokhi, F., 2012. Particle Filtering in the Design of an Accurate Pupil Tracking System. *International Journal of Computer Applications*, *51*(8).
80. BioID Face Database. http://www.bioid.com. Accessed 20 March 2020
81. Dalal, N. and Triggs, B., 2005, June. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (Vol. 1, pp. 886-893). IEEE.
82. Dalal, N., Triggs, B. and Schmid, C., 2006, May. Human detection using oriented histograms of flow and appearance. In *European conference on computer vision* (pp. 428-441). Springer, Berlin, Heidelberg.
83. Viola, P. and Jones, M., 2001, December. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001* (Vol. 1, pp. I-I). IEEE.
84. Paul Viola and Michael J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57:137{154, 2004.
85. Hsu, C.W., Chang, C.C. and Lin, C.J., 2003. A practical guide to support vector classification.
86. Kawato, S. and Tetsutani, N., 2004. Detection and tracking of eyes for gaze-camera control. *Image and Vision Computing*, *22*(12), pp.1031-1038.
87. Asuncion, A. and Newman, D., 2007. UCI machine learning repository.

88. Mokhayeri, F., Toosizadeh, S. and Akbarzadeh-T, M.R., 2011, November. A Novel Approach for Pupil Diameter Measurement Based on Soft Computing Techniques. In *2011 7th Iranian Conference on Machine Vision and Image Processing* (pp. 1-5). IEEE.

89. Jomier, J., Rault, E. and Aylward, S.R., 2004, April. Automatic quantification of pupil dilation under stress. In *2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821)* (pp. 249-252). IEEE.

90. Ahn, S.J., Rauh, W. and Warnecke, H.J., 2001. Least-squares orthogonal distances fitting of circle, sphere, ellipse, hyperbola, and parabola. *Pattern Recognition*, *34*(12), pp.2283-2303.

91. Ahn, S.J., Rauh, W. and Warnecke, H.J., 2001. Least-squares orthogonal distances fitting of circle, sphere, ellipse, hyperbola, and parabola. *Pattern Recognition*, *34*(12), pp.2283-2303.

92. Mokhayeri, F., Toosizadeh, S. and Akbarzadeh-T, M.R., 2011, November. A Novel Approach for Pupil Diameter Measurement Based on Soft Computing Techniques. In *2011 7th Iranian Conference on Machine Vision and Image Processing* (pp. 1-5). IEEE.

93. Long, X., Tonguz, O.K. and Kiderman, A., 2007, August. A high speed eye tracking system with robust pupil center estimation algorithm. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 3331-3334). IEEE.

94. Ray, R.D. and Gross, J.J., 2007. Emotion elicitation using films. *Handbook of emotion elicitation and assessment*, *9*.

95. Zeng, Z., Pantic, M., Roisman, G.I. and Huang, T.S., 2008. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE transactions on pattern analysis and machine intelligence*, *31*(1), pp.39-58.

96. Jerritta, S., Murugappan, M., Nagarajan, R. and Wan, K., 2011, March. Physiological signals based human emotion recognition: a review. In *2011 IEEE 7th International Colloquium on Signal Processing and its Applications* (pp. 410-415). IEEE.

97. Gross, J.J. and Levenson, R.W., 1995. Emotion elicitation using films. *Cognition & emotion*, *9*(1), pp.87-108.

98. Partala, T., Jokiniemi, M. and Surakka, V., 2000, November. Pupillary responses to emotionally provocative stimuli. In *Proceedings of the 2000 symposium on Eye tracking research & applications* (pp. 123-129).

99. Partala, T., Jokiniemi, M. and Surakka, V., 2000, November. Pupillary responses to emotionally provocative stimuli. In *Proceedings of the 2000 symposium on Eye tracking research & applications* (pp. 123-129).

100. Yang SN, McConkie GW. Eye movements during reading: A theory of saccade initiation times. Vision research. 2001 Nov 1;41(25-26):3567-85.

101. Wang, J.G., Sung, E. and Venkateswarlu, R., 2005. Estimating the eye gaze from one eye. *Computer Vision and Image Understanding*, *98*(1), pp.83-103.

102. Villanueva, A., Cabeza, R. and Porta, S., 2007. Gaze tracking system model based on physical parameters. *International Journal of Pattern Recognition and Artificial Intelligence*, *21*(05), pp.855-877.

103. Kim, K.N. and Ramakrishna, R.S., 1999, October. Vision-based eye-gaze tracking for human computer interface. In *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)* (Vol. 2, pp. 324-329). IEEE.

104. Noureddin, B., Lawrence, P.D. and Man, C.F., 2005. A non-contact device for tracking gaze in a human computer interface. *Computer Vision and Image Understanding*, *98*(1), pp.52-82.

105. Shih, S.W., Wu, Y.T. and Liu, J., 2000, September. A calibration-free gaze tracking technique. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000* (Vol. 4, pp. 201-204). IEEE.

106. Shih, S.W. and Liu, J., 2004. A novel approach to 3-D gaze tracking using stereo cameras. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *34*(1), pp.234-245.

107. Coutinho, F.L. and Morimoto, C.H., 2006, October. Free head motion eye gaze tracking using a single camera and multiple light sources. In *2006 19th Brazilian Symposium on Computer Graphics and Image Processing* (pp. 171-178). IEEE.

108. Yoo, D.H. and Chung, M.J., 2005. A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *Computer Vision and Image Understanding*, *98*(1), pp.25-51.

109. Newman, R., Matsumoto, Y., Rougeaux, S. and Zelinsky, A., 2000, March. Real-time stereo tracking for head pose and gaze estimation. In *Proceedings Fourth IEEE*

International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580) (pp. 122-128). IEEE.

110. Khan, M.Q. and Lee, S., 2019. Gaze and Eye Tracking: Techniques and Applications in ADAS. *Sensors*, *19*(24), p.5540.

111. Bayliss, A.P., Schuch, S. and Tipper, S.P., 2010. Gaze cueing elicited by emotional faces is influenced by affective context. *Visual Cognition*, *18*(8), pp.1214-1232..

PUBLICATIONS:

- Alhamzawi, H., Fazekas, A.: Assessment of Patients Emotional Status According To iris Movement.
  In: Contemporary Computational Science : 3rd Conference on Information Technology, Systems Research and Computational Physics and 6th International Symposium CompIMAGE'18 : Computational Modeling of Objects Presented in Images: Fundamentals, Methods, and Applications. Proceedings of the International Multi-Conference on Computational Science (CS 2018), 2-5 July 2018, Kraków, Poland. Eds.: Piotr Kulczycki, Piotr A. Kowalski, Szymon ükasik, AGH University of Science and Technology, Kraków, 160-176, 2018. ISBN: 9788366016224

- Alhamzawi, H.: Control Mouse Cursor by Head Movement: Development and Implementation. Applied Medical Informatics. 40 (3-4), 39-44, 2018.
- Alhamzawi, H.: Faces and eyes Detection in Digital Images Using Cascade Classifiers.ComEngApp. 7 (1), 57-66, 2018.
- Alhamzawi, H.: Location of eyes in images of human faces through analysis variance shine intensity, Indonesian Journal of Electrical Engineering and Computer Science. 11 (3), 949-953, 2018.
- Alhamzawi, H.: Automatics vehicle license plate recognition using MATLAB.Int. J. Comput. Sci. Netw. 6 (6), 677-681, 2017.
- Alhamzawi, H.: Real time gaze estimation for medical field using normally webcam with OpenCV. Int. J. Sci. Res. Sci. Eng. Techn. 3 (8), 136-141, 2017.

Alhamzawi Hussein Ali Mezher (image processing)

1.

Alhamzawi, Hussein ; Attila, Fazekas
Assessment of Patients Emotional Status According To iris Movement
In: Piotr, Kulczycki; Piotr, A Kowalski; Szymon, Łukasik (szerk.) Proceedings of the International Multi-Conference on Computational Science
Cracow, Lengyelország : AGH, (2018) pp. 160-176. , 17 p.
Közlemény:3410998 Érvényesített Forrás Könyvrészlet (Konferenciaközlemény)

2.

Alhamzawi, Hussein Ali Mezher
Faces and eyes Detection in Digital Images Using Cascade Classifiers
Computer Engineering and Applications Journal 7 : 1 pp. 57-66. , 10 p. (2018)
DOI DEA Egyéb URL
Közlemény:30359003 Admin láttamozott Forrás Folyóiratcikk (Szakcikk)

3.

Hussein, Ali Alhamzawi
Control Mouse Cursor by Head Movement: Development and Implementation
APPLIED MEDICAL INFORMATICS 40 : 3-4 pp. 39-44. , 6 p. (2018)
Teljes dokumentum
Közlemény:3410986 Érvényesített Forrás Folyóiratcikk (Szakcikk)

4.

Hussein, Ali Alhamzawi
Location of Eyes in Images of Human Faces through Analysis Variance Shine Intensity
Indonesian Journal of Electrical Engineering and Computer Science 11 : 3 pp. 949-953. , 5 p. (2018)
DOI
Közlemény:3381896 Admin láttamozott Forrás Folyóiratcikk (Szakcikk)

5.

Alhamzawi, Hussein Ali mezher
Automatics vehicle license plate recognition using MATLAB
INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND NETWORK 6 : 6 pp. 677-681. (2017)
DEA Teljes dokumentum
Közlemény:3298665 Érvényesített Forrás Folyóiratcikk (Szakcikk)

6.

Alhamzawi, Hussein Ali Mezher
Real time gaze estimation for medical field using normally webcam with OpenCV
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY 3 : 8 pp. 136-141. (2017)
DEA
Közlemény:3298338 Érvényesített Forrás Folyóiratcikk (Szakcikk)