DEPARTMENT OF INFORMATICS ENGINEERING
SCHOOL OF ENGINEERING
TECHNOLOGICAL EDUCATIONAL INSTITUTE OF CRETE

A THESIS
Submitted in partial fulfillment of the requirements for the degree
MASTER OF SCIENCE

EXPLORING EYE ACTIVITY AS AN INDICATION OF EMOTIONAL STATE

by

CHARALAMPOS TZEDAKIS
BSc, Technological Educational Institute of Crete, 2014

Approved by

Professor Manolis Tsiknakis

HERAKLION, 2016

Statement of Originality

The work contained in this thesis has not been previously submitted for a degree or diploma at any other higher education institution or any other purpose. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except as specified in references, acknowledgments or in footnotes. I certify that the intellectual content of this thesis is the product of my own work and all the assistance received in preparing this thesis and sources have been acknowledged.

Acknowledgements

I would first like to thank my thesis advisor Prof. Manolis Tsiknakis of the Department of Informatics Engineering at Technological Educational Institute of Crete, for believing in me and supporting me. In addition, I would like to thank the Co-Supervisor Prof. Gianakakis for his help.

I would also like to thank Kostas Assariotakis. Without his past work and passionate participation and input, the survey could not have been successfully conducted. Also I would like to thank my good friend M.Sc. Yorgos Manikis for his unstinting help.

I would like to thank my parents Antonis and Stamatia for their love and support. All the support they have provided me over the years was the greatest gift anyone has ever given me.

Last but not least I would like to thank my wife Stella Alefantinou who has supported me throughout entire process, both by keeping me harmonious and helping me putting pieces together. I will be grateful forever for your love.

Abstract

The human eye allows vision but also reflects mood, emotional state, mental and physical condition. Eye activity and behavior reflected in pupil size, gaze direction, eyelid motion or eye opening are affected by states of affect. Investigating literature, it was clear that features extracted from the eye region are included in the most innovative techniques in emotion recognition giving comprehensive, reliable and objective information relating to subject's emotional state.

This thesis aims to deliver the foundational work required for investigating the role of gaze distribution as an emotional index. Towards this objective, an eye gaze detector is developed and tested analyzing videos of volunteers performing a specific gaze task. The accuracy of the method was checked using different lighting and distance conditions in order to find the best parameters to be used. 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 2 dimensional colour video is possible in real life conditions.

Περίληψη

Το ανθρώπινο μάτι, εκτός από λειτουργία της όρασης, αντικατοπτρίζει τη διάθεση, την φυσική και ψυχοσυναισθηματική κατάσταση του ατόμου. Η δραστηριότητα και συμπεριφορά του ματιού η οποία αναπαρίσταται στο μέγεθος της κόρης των ματιών, της κατεύθυνσης του βλέμματος, την κίνηση των βλεφάρων ή στο άνοιγμα των ματιών επηρεάζονται από τις διάφορες συναισθηματικές καταστάσεις. Η βιβλιογραφική επισκόπηση γύρω από αυτό το αντικείμενο κάνει σαφές ότι τα χαρακτηριστικά προερχόμενα από τα μάτια περιλαμβάνονται στις πιο καινοτόμες τεχνικές στην αναγνώριση συναισθημάτων δίνοντας ολοκληρωμένη, αξιόπιστη και τις περισσότερες φορές αντικειμενική πληροφορία σχετικά με τη συναισθηματική κατάσταση.

Η παρούσα πτυχιακή εργασία έχει ως στόχο να δώσει το αναγκαίο τεχνολογικό υπόβαθρο που θα μας επιτρέψει να διερευνήσουμε τον ρόλο και σχέση της κατεύθυνσης του βλέμματος ως έναν συναισθηματικό δείκτη. Προς αυτήν την κατεύθυνση αναπτύχθηκε και αξιολογήθηκε ένας ανιχνευτής κατεύθυνσης βλέμματος μέσω ανάλυση βίντεο καταγραφών εθελοντών που εκτελούσαν μια συγκεκριμένη διαδικασία με το βλέμμα. Η ακρίβεια της μεθόδου ελέγχθηκε χρησιμοποιώντας διαφορετικές συνθήκες φωτισμού και απόστασης για να βρεθούν οι καλύτερες παράμετροι που πρέπει να χρησιμοποιηθούν. Εκτός αυτού, η ακρίβεια της μεθόδου ελέγχθηκε χωριστά για κάθε κατεύθυνση του βλέμματος σε οκτώ διαφορετικές κατευθύνσεις. Από την ανάλυση, μπορεί να αναφερθεί ότι η μηχανική παρακολούθηση του βλέμματος σε βίντεο 2 διαστάσεων σε πραγματικές συνθήκες είναι σχετικά ακριβής και εφικτή.

Contents

Statement of Originality 2

Acknowledgements 3

Abstract 4

Περίληψη 5

Contents 6

Chapter 1 - Introduction 8

1.1 Scope and Objective 10

1.2 Thesis Overview 10

Chapter 2 – Background & Literature Review 11

2.1 Human eye – Physiology 11

2.2 Human eye - Anatomy 11

2.3 Human Eye - Functions 14

2.4 Eye tracking methods 16

2.4.1 Video - oculography (VOG) 16

2.4.2 Electrooculography (EOG) 18

2.4.3. Vision based eye tracking 19

2.5 Face detection algorithms 25

2.6 Eye detection algorithms 26

2.7 Iris tracking algorithms 28

2.8 Pupil tracking algorithms 31

2.9 Gaze Tracking 36

2.9.1 Gaze tracking methods 36

2.9.2 Extracting techniques of gaze characteristics 38

2.9.3 State of the art 40

2.10 Human eye as an emotion indicator 41

Chapter 3 – Data description 45

3.1 Available Databases – Datasets 45

3.2 Self-recorded videos 45

3.3 Semeoticons Experiment 48

Chapter 4 – Technical Implementation 50
4.1 Algorithms description 50
4.1.1 Eye region segmentation algorithm description 50
4.1.2. Iris segmentation description 50
4.1.3. Gaze tracking algorithm description 50
4.2 Gaze estimation Application 50
4.2.1 Region of Interest 50
4.2.2. Gaze estimation 51
Chapter 5 - Experimental Validation 53
5.1 Gaze Tracking Estimation Analysis 53
Chapter 6 - Discussion 55
Chapter 7 - Conclusions & Future Work 56
APPENDIX 57
References 58

Chapter 1 - Introduction

"The eyes are the most important part of an expression, and must be drawn with extreme care. Any jitter or false move … destroys both communication and believability."[1]

The study of eye movements in psychiatry and psychopathology began in 1908 [2]. The understanding of facial signals is fundamental to successfully navigate our social environment, interpreting social signals and cues, and adapting behaviors appropriately [3][4][5]. Observing direction of gaze often results in a congruent focus of attention on the part of the observer. This effect emerges quite early in life, and has been noted in infants as young as 3 months [6]. Gaze following behavior has been characterized as a window into social cognition [7] while gaze is an important signal for displaying emotion [8].

The quality and depth of human social interaction often depends upon the development of appropriate empathic responses. Eye contact has also been shown to play an important role in social interaction. Atypical eye contact and eye-gaze patterns exist as features of some psychological disorders such as autism, where empathic responses can be compromised and the capacity for effective social interaction hampered [9].

As a social signal, direction of gaze has been implicated in signaling an expressor's approach-avoidance behavioral tendencies. Stern [10] offered evidence that gaze behavior provides a way for infants to approach and withdraw from others in an effort to regulate their own emotional experience. In addition, according to Mehrabian's [11] immediacy model of social intimacy, increased eye contact is an essential cue used to enhance psychological closeness in adults. More recently, research demonstrates that both infants and adults reflexively respond to the gaze direction displayed by others [12] [13]. Infants as young as 2 days old can distinguish direct from averted gaze, suggesting that the ability to detect gaze direction is innately prepared [14]. According to Baron-Cohen [15], this innate capacity to process gaze information plays a critical role in the development of a "theory of mind", i.e., the ability to infer behavioral intentions from the nonverbal behavior of others. He defines behavioral intentions as "primitive mental states in that they are the basic ones that are needed in order to be able to make sense of the universal movements of all animals: approach and avoidance"

(pp. 33–34).

Despite wide disagreement over whether emotional displays are universal or culturally determined, cognitive or biological in origin, most recent researchers nonetheless seem able to agree that facial expressions can functionally act to forecast an organism's behavioral tendencies[16] [17].

Additionally, early animators realized that the eyes are an important aspect of the human face regarding the communication and expression of emotion. They also found that properly animating believable, emotionally expressive gaze is extremely difficult. If it's done improperly, viewers will immediately notice, destroying the illusion of life the animators were striving for. This problem becomes even more difficult with interactive virtual characters—computer-controlled characters in a virtual environment that interact with each other and human users through conversation. Unlike hand-animated characters, interactive virtual characters must function in the virtual environment and interact with human users in real time. This also entails gazing at arbitrary targets as they appear in the environment.

For example, modern video game environments, such as World of Warcraft or Oblivion, have thousands of computer-controlled non-player characters with which users can interact. Producing hand generated animations for each of these characters consumes large amounts of time and money that would often be better spent improving other game aspects.

To address this issue, researchers developed the Expressive Gaze Model (EGM), a method that enables virtual characters to generate believable gaze shifts that communicate a desired emotional expression. We define a believable, emotionally expressive gaze as an eye, head, or torso movement that places the forward vector of the virtual character's eyes on a new target while portraying behaviors that cause a human viewer to attribute a desired emotional expression to that character, in a way that doesn't damage the viewer's suspension of disbelief. The Expressive Gaze Model (EGM) is a hierarchical framework for composing simple behaviors into emotionally expressive gaze in virtual characters that encompasses eye, torso, and head movement [8].

Given the importance of gaze in both real and artificial words as a way of expressing or understanding emotion, in this thesis, we present the design and implementation of a low-cost gaze tracking system using only a laptop computer and a simple web camera. This system uses Matlab based software to estimate gaze positions. Creating a simple

and low-cost gaze tracking system can be the first step in developing a modality that can automatically identify the emotional state of a subject from its gaze.

1.1 Scope and Objective

The present study focuses on the development and validation of an automatic gaze tracking system. Attention is also payed on the optimization of the necessary methods and algorithms to estimate gaze position. Finally, it validates the measurement accuracy of the gaze tracking technique.

In addition, the thesis attempts to provide an elaborate literature review of the various circumstances in which gaze is used as an emotion indicator.

1.2 Thesis Overview

The thesis is organized in chapters, as follows

Chapter 1: The current chapter has briefly introduced the topic and highlighted the scope and the objectives of the present thesis.

Chapter 2: Provides the required background on Human Eye Anatomy, Physiology and Functions as well as an extended presentation of all methods and related studies for eye, pupil and gaze tracking.

Chapter 3: Lists all the datasets used in literature for gaze estimation and head pose tracking as well as the self-recorded videos for the validation of the current study.

Chapter 4: Analyzes all the technical implementation. In this chapter the used algorithms and the method used to estimate the gaze will be detailed.

Chapter 5: Presents all the validation results**.**

Chapter 6: In this chapter an analytic discussion over the results will be made.

Chapter 7: The last chapter is the conclusion of the present study which also identifies directions for future work.

Chapter 2 – Background & Literature Review

2.1 Human eye – Physiology

Human eye is the sense organ of vision. Vision is considered as the most important of the senses, because through vision humans and the most animals immediately perceive the outside space. Humans are highly visual creatures compared to many other animals which rely more on smell or hearing, and over our evolutionary history have developed an incredibly complex sight system.

Approximately 30% of the human brain is engaged in the processing and interpretation of visual stimuli. The human eye takes in the physical stimuli of light rays and transduces them into electrical and chemical signals that can be interpreted by the brain to construct physical images. The human visual system is capable of complex color perception, which is initiated by cones in the retina and completed by impulse integration in the brain. Depth perception is our ability to see in three dimensions and relies on both binocular (two-eye) and monocular (one-eye) cues [18].

As Greek philosophers claimed long ago, eyes are “windows of the soul”. There are many things we can read from eyes. Sparkling eyes shows strong emotions (negative or positive), narrowed eyes indicate suffering and pain while widely opened eyes serve to intimidate. Furthermore, the pupil expands or shrinks depending on human’s mental and physical condition. For example, pupil expands when a person is focusing on something or feeling stress or even is in pain. On the other hand, pupil shrinks when a person is confronted with something revolting or observing a nearby object.

2.2 Human eye - Anatomy

The eye consists of a retinal-lined fibrovascular sphere which contains the aqueous humor, the lens and the vitreous body [19]. The anatomy of human’s eye and its parts are shown in figure 1.

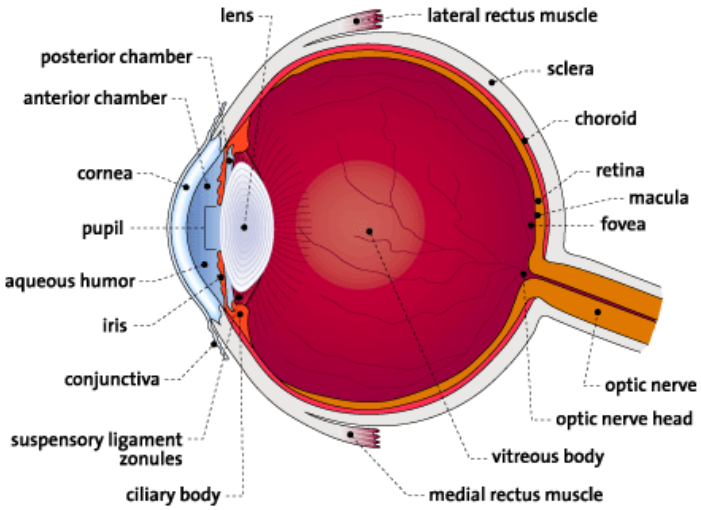

***Figure 1**: Human Eye (anatomy)*

- **Cornea:** The clear, transparent front portion of the fibrous coat of the eye; functions as an important refractive medium.
- **Sclera:** The tough white protective coat of the eye. The portion of the sclera that surrounds the cornea is covered by the conjunctiva.
- **Conjunctiva:** A mucous membrane extending from the eyelid margin to the corneal limbus, forming the posterior layer of the eyelids and the anterior layer of the eyeball.
- **Iris**: A colored, circular membrane suspended behind the cornea and immediately in front of the lens. It regulates amount of light entering the eye by adjusting size and pupil.
- **Pupil**: The opening at the center of the iris of the eye; it contracts when exposed to strong light or when the focus is on a near object and it dilates when in the dark or when the focus is on a distant object.
- **Aqueous**: Watery liquid that flows between the lens and the cornea and nourishes them.
- **Lens:** The transparent tissue behind the iris that bends light rays and focuses them on the retina.
- **Schlemm's Canal**: A passageway for the aqueous fluid to leave the eye.

- **Vitreous body:** Transparent, colorless mass of soft, gelatinous material that fills the center of the eye behind the lens.
- **Retina:** Light-sensitive tissue at the back of the eye which transmits visual impulses via the optic nerve to the brain.
- **Macula**: Pigmented central area or "yellow spot" of the retina devoid of blood vessels. It is the most sensitive area of the retina and is responsible for fine or reading vision.
- **Choroid**: Blood vessel-rich tissue behind the retina that is responsible for its nourishment.
- **Optic nerve:** The nerve at the back of the eye that carries visual impulses from the retina to the brain. The area at which the optic nerve connects with the retina is known as the optic disc.

The retina is the essential component of the eye and serves the primary purpose of photoreception [20]. All other structures of the eye are subsidiary and act to focus images on the retina, to regulate the amount of light entering the eye or to provide nutrition, protection or motion. On the outside of the sphere, corresponding to the dura mater, a layer composed of dense fibrous tissue serves as a protective envelope, the fibrous tunic. The posterior part of the fibrous tunic, the sclera, is white and opaque. Although it retains its protective function, the anterior portion, the cornea, is clear and transparent. 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 posterior portion of the uvea is the choroid, a tissue composed almost entirely of blood vessels. A second portion of the uvea, the ciliary body, 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. The most anterior portion of the uveal tract, the iris, is deflected into the interior of the eye. The iris acts as a diaphragm with a central rounded opening, the pupil, which dilates to allow more light to the retina in dim lighting and constricts in bright lighting. The iris also has some degree of nutritive function, since it acts to help regulate the fluid flow in the eye. The lens, the focusing mechanism of the eye, is located immediately behind the iris and is

supported from the ciliary body by a suspensory ligament, the zonule. The space between the iris and the lens is called the posterior chamber. The anterior chamber consists of the space between the iris and the cornea. Behind the lens is the vitreous, a gel-like, transparent body which occupies the space between the lens and the retina [21].

2.3 Human Eye - Functions

The human eye functions like a camera. The inside part of the bulb could be considered as darkroom. The rays collected by the cornea and crystalline lens functioning as the zoom of the camera and passes through the aperture of the pupil, i.e. the aperture in the camera, until they reach the retina, that is the film, and wherein focus. Then the processed image is formed/decoded in the brain [22]. The picture reaches the brain inverted as seen in the image below).

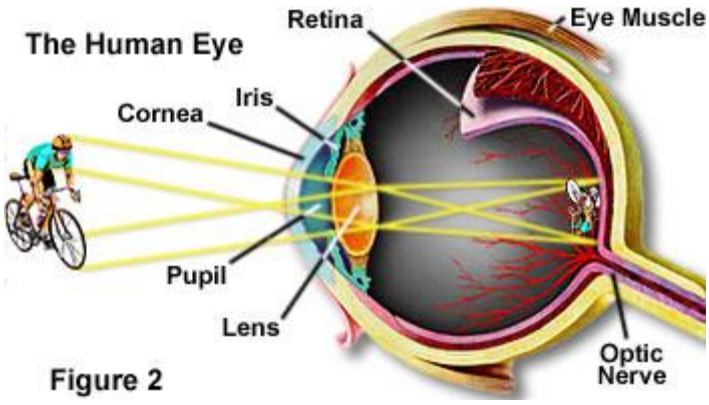

***Figure 2**: Human eye (function)*

Iris has in the middle a round opening, pupil, for 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 miosis occurs, whereas when it is dark, the pupils dilate so as to permit more light enter through them.

The cornea and the iris form an angle, the angle of the anterior chamber. The anterior chamber is filled with a clear liquid, the aqueous humor, which maintains the setting in the anterior segment of the eye. The angle has a critical role in glaucoma according to whether it is closed or open because from there the drainage of the aqueous humor takes

place. Behind the iris is the crystalline lens, approximately sized lentil. The cornea with the lens refract and concentrate the light rays on the retina, which is focused on the macula. The lens may fluctuate and shape by means of a muscle and simultaneously varying refractive power, the mechanism of adjustment. Thus, "zooming" focuses the image, depending on whether we look away or close. Behind the lens, the posterior chamber of the eye consists of the vitreous body, a gelatinous, transparent material. The sclera inside surrounded by the choroid, rich in blood vessels.

Finally, in the inner part of the eye is the retina. The light stimulus is taken up by light-sensitive cells, cones and rods, and complex mechanisms converted into nerve impulses. Nerve cells take stimulus and carrying through the optic nerve. The optic nerve consists of about 1 million axons arising from the ganglion cells of the retina. The nerve fiber layer of the retina is comprised of these axons and they converge to form the optic nerve. The orbital portion of the nerve travels within the muscle cone to enter the bony optic foramen to gain access to the cranial cavity. The optic nerve is made up of visual fibers (80%) and afferent pupillary fibers (20%) [23]. First of all, the optic nerve routes information via the thalamus to the cerebral cortex, where visual perception occurs, but the nerve also carries information required for the mechanics of vision to two sites in the brainstem. The first of these sites is a group of cells (a nucleus) called the pretectum, which controls pupillary size in response to light intensity. Information concerning moving targets and information governing scanning of the eyes travels to a second site in the brainstem, a nucleus called the superior colliculus. The superior colliculus is responsible for moving the eyes in short jumps, called saccades. Saccades allow the brain to perceive a smooth scan by stitching together a series of relatively still images. Saccadic eye movement solves the problem of extreme blurring that would result if the eyes could pan smoothly across a visual landscape; saccades can be readily observed if you watch someone's eyes as they attempt to pan their gaze across a room.

Most projections from the retina travel via the optic nerve to a part of the thalamus called the lateral geniculate nucleus (LGN), deep in the center of the brain. The LGN separates retinal inputs into parallel streams, one containing color and fine structure, and the other containing contrast and motion. Cells that process color and fine structure make up the top four of the six layers of the LGN; those four are called the parvocellular layers, because the cells are small. Cells processing contrast and motion make up the

bottom two layers of the LGN, called the magnocellular layers because the cells are large.

The cells of the magnocellular and parvocellular layers project all the way to the back of the brain to primary visual cortex (V1). Cells in V1 are arranged in several ways that allow the visual system to calculate where objects are in space. First, V1 cells are organized retinotopically, which means that a point-to-point map exists between the retina and primary visual cortex, and neighboring areas in the retina correspond to neighboring areas in V1. This allows V1 to position objects in two dimensions of the visual world, horizontal and vertical. The third dimension, depth, is mapped in V1 by comparing the signals from the two eyes. Those signals are processed in stacks of cells called ocular dominance columns, a checkerboard pattern of connections alternating between the left and right eye. A slight discrepancy in the position of an object relative to each eye allows depth to be calculated by triangulation [24]

Finally, V1 is organized into orientation columns, stacks of cells that are strongly activated by lines of a given orientation. Orientation columns allow V1 to detect the edges of objects in the visual world, and so they begin the complex task of visual recognition. The columnar organization of primary visual cortex was first described by David Hubel and Torsten Wiesel, resulting in their 1981 Nobel Prize.

2.4 Eye tracking methods

The term eye tracking refers to the process of the tracking the movement of the eye and determining where the user is looking. There are a number of principles used in measuring eye movements, including measurements of electric and photoelectric signals, tracking a number of visual features in the image of the eye, measuring relative reflection of infra-red (IR) light, and using either mechanical or optical levers or a magnetic field [25]. Some of the most-used eye tracking methods are mentioned below.

2.4.1 Video - oculography (VOG)

VOG is a non-invasive, video-based method of measuring horizontal, vertical and torsional position components of the movements of both eyes (eye tracking) using a head-mounted mask that is equipped with small cameras [26] [27]. VOG is usually employed for medical purposes. It uses electric potentials measured with electrodes

placed around the eyes. The eyes are the origin of a steady electric potential field, which can also be detected in total darkness and if the eyes are closed. It can be modelled to be generated by a dipole with its positive pole at the cornea and its negative pole at the retina.

The measurement of the horizontal and vertical components is well established technology which uses pupil tracking and/or corneal reflection tracking and has been widely applied, for example for tracking eye movements in reading. In contrast, the measurement of the torsional component (cyclorotation) is usually considered a computationally more difficult task. Approaches to solving this problem include, among others, polar cross correlation methods and iris pattern matching/tracking. In animal studies, VOG has been used in combination with fluorescent marker arrays affixed to the eye, and it has been proposed that such an array could be embedded into a scleral lens for humans.

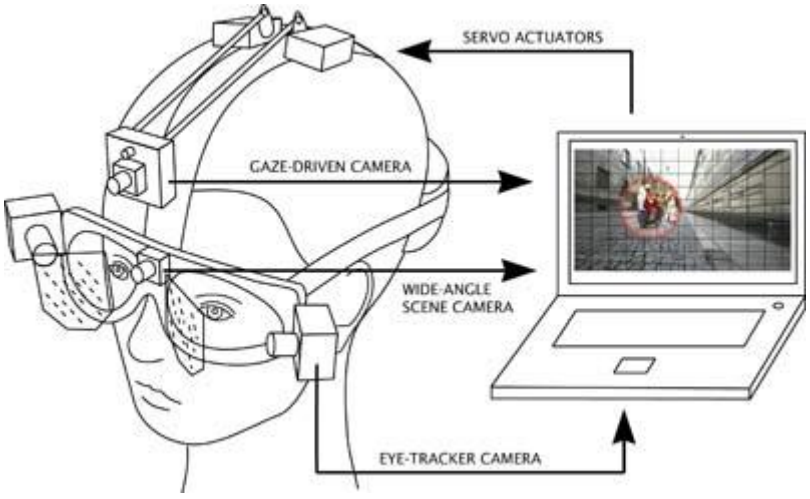

***Figure 3**: Video-Oculography (VOG) System*

VOG techniques have been put to use in a wide field of scientific research related to visual development and cognitive science as well as to pathologies of the eyes and of the visual system. For example, miniaturized ocular-videography systems are used to analyze eye movements in freely moving rodents. VOG can be used in eye examinations for quantitative assessments of ocular motility, binocular vision, vergence, cyclovergence, stereoscopy and disorders related to eye positioning such as nystagmus and strabism. It has also been proposed for assessing linear and torsional eye movements in vestibular patients and for early stroke recognition.

2.4.2 Electrooculography (EOG)

Electrooculography (EOG) is a technique for measuring the resting potential of the retina. The electric signal that can be derived using two pairs of contact electrodes placed on the skin around one eye is called electrooculogram [28]. An electrooculograph is a device that measures the voltage between two electrodes placed on the face of a subject so it can detect eye movement. Today the use of computers is extended to every field. Many sophisticated devices like touch screen, track ball, digitizers etc made interaction with computer ease from novice to professional. Assistive robotics can improve the quality of life for disable people.

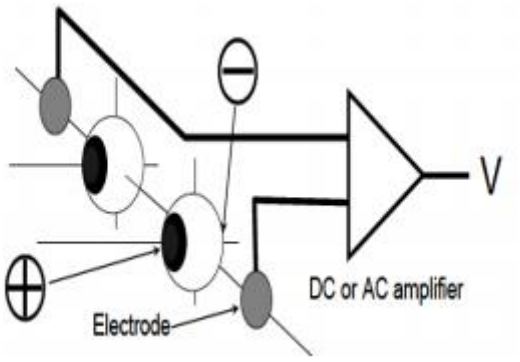

Figure 4: *Principle of Electro-oculography (EOG)*

The eyes are the origin of a steady electric potential field, which can also be detected in total darkness and if the eyes are closed. It can be modelled to be generated by a dipole with its positive pole at the cornea and its negative pole at the retina. If the eyes move from the center position towards the periphery, the retina approaches one electrode while the cornea approaches the opposing one. This change in the orientation of the dipole and consequently the electric potential field results in a change in the measured EOG signal. Inversely, by analyzing these changes in eye movement can be tracked. Due to the discretization given by the common electrode setup two separate movement components – a horizontal and a vertical – can be identified. A third EOG component is the radial EOG channel, which is the average of the EOG channels referenced to some posterior scalp electrode. This radial EOG channel is sensitive to the saccadic spike potentials stemming from the extra-ocular muscles at the onset of saccades, and allows reliable detection of even miniature saccades. Due to potential drifts and variable relations between the EOG signal amplitudes and the saccade sizes

make it challenging to use EOG for measuring slow eye movement and detecting gaze direction. EOG is, however, a very robust technique for measuring saccadic eye movement associated with gaze shifts and detecting blinks. Contrary to video-based eye-trackers, EOG allows recording of eye movements even with eyes closed, and can thus be used in sleep research.

It is a very light-weight approach that, in contrast to current video-based eye trackers, only requires very low computational power, works under different lighting conditions and can be implemented as an embedded, self-contained wearable system [29].

2.4.3. Vision based eye tracking

Vision based eye tracking methods are intrinsically non-intrusive, and does not require any overly expensive equipment. Non-obtrusive sensing technology – such as video cameras and microphones – has received special attention in this regard.

Active appearance models (AAM)

AAM and its variations is computer vision technique for fitting a statistical model of object shape and appearance to an image or sequence of images under investigation [30][31][32]. The model of shape and appearance is built during a training phase. A set of images, together with coordinates of landmarks that appear in all of the images, is provided as training set. They are used to model specific patterns of variability in shape and grey-level or colour appearance, which in turn can be used directly in image interpretation.

The model was first introduced by Edwards, Cootes and Taylor in the context of face analysis at the 3rd International Conference on Face and Gesture Recognition, 1998 [30]. Cootes, Edwards and Taylor further described the approach as a general method in computer vision at the European Conference on Computer Vision in the same year. The approach is widely used for matching and tracking faces and for medical image interpretation. The algorithm uses the difference between the current estimate of appearance and the target image to drive an optimization process based on minimizing the error of alignment between the model and the image. In relation to the active shape model (ASM) which uses only shape model (together with some information about the image structure near the landmarks) it takes advantage of all the available information including the texture across the target object.

Mixture models

Mixture models are probabilistic models for representing the presence of subpopulations within an overall population, without requiring that an observed data set should identify the sub-population to which an individual observation belongs [33]. Formally a mixture model corresponds to the mixture distribution that represents the probability distribution of observations in the overall population. However, while problems associated with "mixture distributions" relate to deriving the properties of the overall population from those of the sub-populations, "mixture models" are used to make statistical inferences about the properties of the sub-populations given only observations on the pooled population, without sub-population identity information.

Some ways of implementing mixture models involve steps that attribute postulated sub-population-identities to individual observations (or weights towards such sub-populations), in which case these can be regarded as types of unsupervised learning or clustering procedures. However not all inference procedures involve such steps. Mixture models should not be confused with models for compositional data, i.e., data whose components are constrained to sum to a constant value (1, 100%, etc.). However, compositional models can be thought of as mixture models, where members of the population are sampled at random. Conversely, mixture models can be thought of as compositional models, where the total size of the population has been normalized to 1.

A typical finite-dimensional mixture model is a hierarchical model consisting of the following components:

N random variables corresponding to observations, each assumed to be distributed according to a mixture of K components, with each component belonging to the same parametric family of distributions (e.g., all normal, all Zipfian, etc.) but with different parameters N corresponding random latent variables specifying the identity of the mixture component of each observation, each distributed according to a K-dimensional categorical distribution A set of K mixture weights, each of which is a probability (a real number between 0 and 1 inclusive), all of which sum to 1 A set of K parameters, each specifying the parameter of the corresponding mixture component. In many cases, each "parameter" is actually a set of parameters. For example, observations distributed according to a mixture of one-dimensional Gaussian distributions will have a mean and variance for each component. Observations distributed according to a mixture of V-dimensional categorical distributions (e.g., when each observation is a word from a

vocabulary of size V) will have a vector of V probabilities, collectively summing to 1. In addition, in a Bayesian setting, the mixture weights and parameters will themselves be random variables, and prior distributions will be placed over the variables. In such a case, the weights are typically viewed as a K-dimensional random vector drawn from a Dirichlet distribution (the conjugate prior of the categorical distribution), and the parameters will be distributed according to their respective conjugate priors.

A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of Gaussian component densities [34]. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system, such as vocal-tract related spectral features in a speaker recognition system. GMM parameters are estimated from training data using the iterative Expectation-Maximization (EM) algorithm or Maximum A Posteriori (MAP) estimation from a well-trained prior model.

Histogram of Oriented Gradients (HOG)

HOG is a feature descriptor used in computer vision and image processing for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an image. This method is similar to that of edge orientation histograms, scale-invariant feature transform descriptors, and shape contexts, but differs in that it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy.

Navneet Dalal and Bill Triggs, researchers for the French National Institute for Research in Computer Science and Control (INRIA), first described HOG descriptors at the 2005 Conference on Computer Vision and Pattern Recognition (CVPR) [35]. In this work they focused on pedestrian detection in static images, although since then they expanded their tests to include human detection in videos, as well as to a variety of common animals and vehicles in static imagery.

The essential thought behind the histogram of oriented gradients descriptor is that local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions. The image is divided into small connected regions called cells, and for the pixels within each cell, a histogram of gradient directions is compiled. The descriptor is then the concatenation of these histograms. For improved accuracy, the local histograms can be contrast-normalized by calculating

a measure of the intensity across a larger region of the image, called a block, and then using this value to normalize all cells within the block. This normalization results in better invariance to changes in illumination and shadowing.

The HOG descriptor has a few key advantages over other descriptors. Since it operates on local cells, it is invariant to geometric and photometric transformations, except for object orientation. Such changes would only appear in larger spatial regions. Moreover, as Dalal and Triggs discovered, coarse spatial sampling, fine orientation sampling, and strong local photometric normalization permits the individual body movement of pedestrians to be ignored so long as they maintain a roughly upright position. The HOG descriptor is thus particularly suited for human detection in images.

As the eye region, iris and sclera has high contrast, HOG is very effective in eye detection. In a recent study [36], HOG was used for eye detection. The image window was divided into four sub-windows, and a histogram of intensity gradient directions, weighted by gradient magnitude was computed for each of the sub-windows separately. For each sub-window, an 8-bar histogram was calculated. So, the total gradient directions were divided into 8 parts. The gradient magnitudes with those directions were summed up to calculate the histogram value. Then, this 8-bar histogram was normalized. After the normalization, the four 8-bar histograms of the four sub-windows were concatenated into a 32-element feature vector of that window for classification. Calculation of gradient magnitude and gradient direction of a sample eye image is shown in figure 5. In the first step the horizontal and vertical gradient of the image was calculated. From these images, the gradient magnitude and gradient direction images were obtained. After calculation of the gradient magnitude and direction images, the process of calculating the HOG is shown in figure 6.

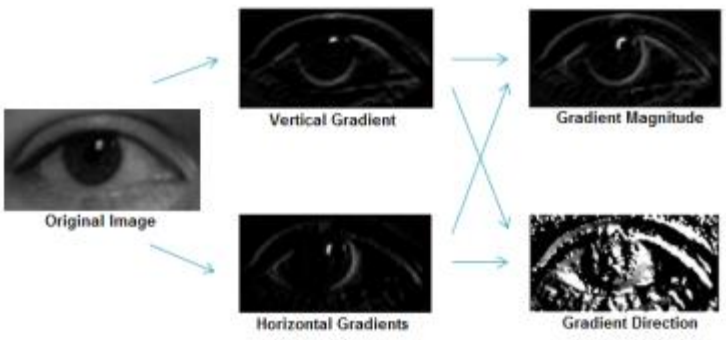

Figure 5 : *Gradient magnitude and Director Calculation*

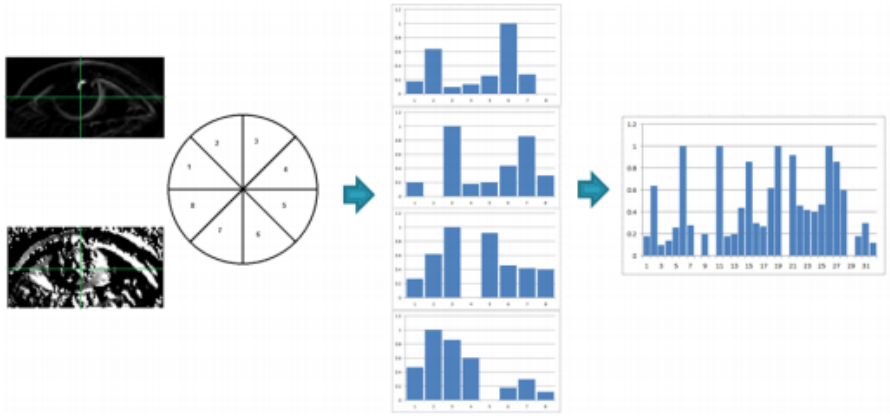

Figure 6: *Histogram of Oriented Gradients Calculation*

HOG seems to be efficient enough for eye tracking and not only for human detection in images.

Local binary patterns (LBP)

Local Binary Patterns (LBP) is a non-parametric descriptor whose aims is to efficiently summarize the local structures of images [37]. LBP is the particular case of the Texture Spectrum model proposed in 1990 [38][39]. LBP was first described in 1994 [40][41]. It has since been found to be a powerful feature for texture classification; it has further been determined that when LBP is combined with the Histogram of oriented gradients (HOG) descriptor, it improves the detection performance considerably on some datasets.

The LBP feature vector, in its simplest form, is created in the following manner: Divide the examined window into cells (e.g. 16x16 pixels for each cell). For each pixel in a cell, compare the pixel to each of its 8 neighbors (on its left-top, left-middle, left-bottom, right-top, etc.). Follow the pixels along a circle, i.e. clockwise or counter-clockwise.

Where the center pixel's value is greater than the neighbor's value, write "1". Otherwise, write "0". This gives an 8-digit binary number (which is usually converted to decimal for convenience). Compute the histogram, over the cell, of the frequency of each "number" occurring (i.e., each combination of which pixels are smaller and which are greater than the center). Optionally normalize the histogram. Concatenate (normalized) histograms of all cells. This gives the feature vector for the window.

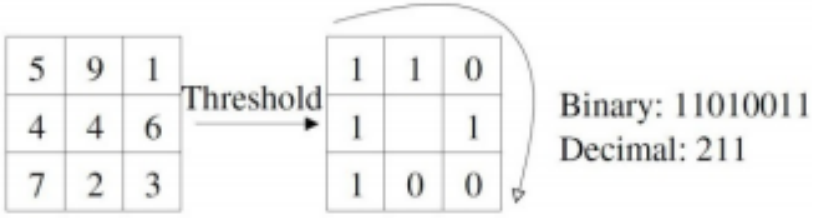

Figure 7: *An example of the basic LBP operator*

The feature vector can now be processed using the Support vector machine or some other machine-learning algorithm to classify images. Such classifiers can be used for face recognition or texture analysis. A useful extension to the original operator is the so-called uniform pattern, which can be used to reduce the length of the feature vector and implement a simple rotation invariant descriptor. This idea is motivated by the fact that some binary patterns occur more commonly in texture images than others. A local binary pattern is called uniform if the binary pattern contains at most two 0-1 or 1-0 transitions. For example, 00010000(2 transitions) is a uniform pattern, 01010100(6 transitions) is not. In the computation of the LBP histogram, the histogram has a separate bin for every uniform pattern, and all non-uniform patterns are assigned to a single bin. Using uniform patterns, the length of the feature vector for a 3x3 window reduces from 256 to 59.

In a recent study [42], a probabilistic eye localization method based on LBPs was used due to the fact that these patterns provide a simple but powerful spatial description of texture, and are robust to the noise typical to various illumination conditions and pose. LBPs were used for their higher accuracy rate and lower complexity. For a given close-up image, the center of the iris of two eyes was located. The complete system has been tested on the standard databases and web-cam pictures of people under different light conditions. The accuracy has been nearly 98 % (±1 pixel shift).

Hidden Markov model (HMM)

HMMs are widely used to model data generated from Markov processes (Barber, 2012). Specifically, it is a statistical Markov model in which the system being modelled is assumed to be a Markov process with unobserved (hidden) states. The mathematic

behind the HMM [43] [44] is closely related to an earlier work on the optimal nonlinear filtering problem [45].

In simpler Markov models (like a Markov chain), the state is directly visible to the observer, and therefore the state transition probabilities are the only parameters. In a hidden Markov model, the state is not directly visible, but output, dependent on the state, is visible. Each state has a probability distribution over the possible output tokens. Therefore the sequence of tokens generated by an HMM gives some information about the sequence of states. Note that the adjective 'hidden' refers to the state sequence through which the model passes, not to the parameters of the model; the model is still referred to as a 'hidden' Markov model even if these parameters are known exactly. Hidden Markov models are especially known for their application in temporal pattern recognition such as speech, handwriting, gesture recognition, part-of-speech tagging, musical score following, partial discharges and bioinformatics.

A hidden Markov model can be considered a generalization of a mixture model where the hidden variables (or latent variables), which control the mixture component to be selected for each observation, are related through a Markov process rather than independent of each other. Recently, hidden Markov models have been generalized to pairwise Markov models and triplet Markov models which allow consideration of more complex data structures and the modelling of nonstationary data.

HMMs is also used for eye tracking [46][47]. Eye movements can be studied and understood using HMMs. With HMMs, both spatial and the sequential aspects of eye movements can be described. Clustering HMMs can yield interesting between group differences.

2.5 Face detection algorithms

Viola-Jones

The algorithm Viola & Jones [48] is widely distributed and is based on AdaBoost algorithm which searches for features that wants to track and export it to rectangular areas. However, these features are usually not in a single rectangular but more so that it is much more complex. Initially, the algorithm locates a rectangle which runs across the image length and is capable of identifying some characteristics (features detection). The Viola - Jones used a cascade classifier (classifier) who instead detects faces reject non-faces, as soon as the process is more efficient. Furthermore, the classifier separates

another class, potential persons, which when re-running the detector subsequent times is determined whether categorized as non-persons or as individuals. This algorithm uses the complete image (integral image) which is a particular representation of the image is controlled. With the complete picture the whole detection process is accelerated significantly.

AdaBoost

The Adaptive Boosting (AdaBoost) [49] is a meta-algorithm of machine learning proposed by Yoav Freund and Robert Schapire. This algorithm uses several algorithms to optimize and reduce the time of their processes. The function of AdaBoost is as follows: initially a weak classifier performed multiple times for each execution of updated a database which contains the allocation of the basis of the algorithm which highlights the importance of data.

Kanade- Lucas- Tomasi (KLT)

The Kanade–Lucas–Tomasi (KLT) feature tracker [50][51] is an approach to feature extraction. It is proposed mainly for the purpose of dealing with the problem that traditional image registration techniques are generally costly. KLT makes use of spatial intensity information to direct the search for the position that yields the best match. It is faster than traditional techniques for examining far fewer potential matches between the images. It is widespread because of its low computational complexity and because it is based on the spread Taylor, which does not introduce errors in the calculations. Even I, J two images. The aim is, in the image I can identify a point in a known location x '= [x, y] T and find the displacement d = [dx, dy] T. Given W window, we can find the difference between the previous and the current point.

2.6 Eye detection algorithms

A lot eye tracking applications (e.g. gaze estimation) need only the detection and tracking of the iris or the pupil. Depending on the angle viewed, both the pupil and the iris are elliptical and therefore can be modeled by five shape parameters.

Simple ellipse models consist of voting-based methods [52] and model fitting methods [53][54] Voting methods select features that support a given hypothesis through a voting or accumulation process, while model fitting approaches fit the

selected features in the model (e.g. ellipse). Kim and Ramakrishna and Perez et al. [55] used thresholds of image intensities to calculate the center of the pupil ellipse. Edge detection techniques are used to extract the limbus or the pupil bounds. Many areas in the image may have the same intensity profile to the iris and pupil areas and thresholds are therefore only applicable to strained settings. The Hough transform can be used to effectively extract the iris or the pupil [56][57], but needs definite feature detection. Often a circular shape constraint is employed for efficiency reasons so the model works on near frontal faces.Kothari and Mitchell [58] suggested a different voting concept that uses spatial and temporal information to detect the area of the eyes. They use the gradient field, knowing that the gradient along the iris boundary points outward from the center of the iris. Heuristic rules and a broad temporal support are used to filter erroneous pupil nominees. An akin voting scheme is proposed by Valenti and Gevers. [59]. Their method is based on isophote curvatures in the tension image and uses edge orientation directly in the voting process. The approximation relies on a prequel face model and anthropomorphic averages to limit false positives. Since these models rely on maxima in feature space, they could mistake other features for eyes (e.g. eyebrows or eye corners) when the number of features in the eye region decreases. These methods are mostly used when a constrained search region is available.

Li and Parkhurst [60] demonstrated an inexpensive eye tracking and propose the Starburst algorithm for locating the iris via elliptical shape model. The algorithm identifies the strongest gray level differences with rays and sparkles retrospectively new rays previously found maxima. The maximal possibility evaluation of the pupil area is found via RANSAC. While framed differently, the Starburst algorithm is an active shape model like Cootes and Taylor [61], but allowing for several features to be used along each normal. Plain shape models are in usual adequate and they can model features such as iris and pupil below lots of viewing angles. The simple models are incapable of apprehend the variations and inter-alterations of eye features such as eyelids, eye corners and eyebrows. High antithesis images and thresholds are usually used for feature extraction.

Bala et al. [62] introduced a hybrid way for eye classification by using an evolutionary algorithm to classify a group of optimal features (mean intensities, Laplacian and entropy) to qualify the eye. Feng et al [63][64] describe an eye model consisting of six points (eye corner landmarks). At first the eye points are located and used to guide the localization of iris and eye border. The methods estimate the

availability of an eye window where the eye is the only object. The gray scale face model that estimates the eye window is circumscribed. The exact eye position is resolved and checked by using the variance projection function. Variance projection functions use the variance of intensities within a given eye area to calculate the exact position and the size of iris or the positions of the eye lids. The variance projection function can be shown to be orientation and scale invariant. Experiments indicate that this method does not succeed when the eye is not open or partially covered by hair or face orientation. It is influenced by shadows and eye movements. In addition, this technique may slip up eyebrows for eyes.

Instead of detecting eye features, Kawato et al.[65] [66] in their algorithm preferred to detect the area between the eyes. This area has obscure parts on the left and right side (eyes and eyebrows) and contrastly lucent on the upper side (forehead) and the lower side (nose bridge). The whole area is argued to be the same for most of the people, discoverable for a wide range of angles and is believed to be easier for detection in comparison to the eyes themselves. They employ a circle-frequency filter to find candidate points. The spurious points are subsequently eliminated from the candidates based on studying the tension distribution pattern around the mark. To obstruct the eyebrows or other hair parts from being considered as eye regions, this method is made more robust by constructing a standard 'Between-the-eyes' template to figure the true one from within the nominees [67][68]. Experiments showed that the algorithm may fail when the forehead is covered by facial hair or when the subject is wearing black glasses.

2.7 Iris tracking algorithms

Iris is distinct for every person, even the twins have different iris pattern and it remains same for whole of the life. Thus, iris tracking technology is now considered as providing positive identification of an individual without contact and at very high confidence levels.

Since 1990s, many researchers have worked on creating iris tracking algorithms. In this section the various algorithms for iris recognition has been discussed. Human iris identification process is basically divided into four steps.

- Localization - The inner and the outer boundaries of the iris are calculated.

- Normalization - Iris of different people may be captured in different size, for the same person also size may vary because of the variation in illumination and other factors.
- Feature extraction - Iris provides abundant texture information. A feature vector is formed which consists of the ordered sequence of features extracted from the various representation of the iris images.
- Matching - The feature vectors are classified through different thresholding techniques like Hamming Distance, weight vector and winner selection, dissimilarity function, etc.

Daugman [69] [70] is the first one to give an algorithm for iris recognition. His algorithm is based on Iris Codes. For the preprocessing step i.e., inner and outer boundaries of the iris are located. Integro-differential operators are then used to detect the center and diameter of the iris, then the pupil is also detected using the differential operators, for conversion from Cartesian to polar transform, rectangular representation of the required area is made. Feature extraction algorithm uses the modified complex valued 2-D Gabor wavelets. For matching, Hamming Distance has been calculated by the use of simple Boolean Exclusive - OR operator and for the perfect match give the hamming distance equal to zero is obtained. The algorithm gives the accuracy of more than 99.9%. Also the time required for iris identification is less than one second.

Wildes has made use of an isotropic band-pass decomposition derived from application of Laplacian of Gaussian filters to the image data [71]. Like Daugman Wildes also used the first derivative of image intensity to find the location of edges corresponding to the borders of the iris. The Wildes system explicitly models the upper and lower eyelids with parabolic arcs whereas Daugman excludes the upper and the lower portions of the image. The results of this system were good enough to recognize the individuals in minimum time period.

Boashash and Boles [72] have given a new approach based on zero-crossings [73]. They first localized and normalized the iris by using edge detection and other well-known computer vision algorithms. The zero-crossings of the wavelet transform are then calculated at various resolution levels over concentric circles on the iris. The resulting one dimensional (ID) signals are then compared with the model features using different dissimilarity function. This system can handle noisy conditions as well as variations in illumination. This algorithm is also translation, rotation and scale invariant. A similar type of system has been presented in [74] which is based on zero-

crossing discrete dyadic wavelet transform representation and has shown a high level of accuracy.

In [75], an algorithm has been proposed to extract the features of iris signals by Multi-resolution Independent Component Identification (M-ICA). It provides good properties to represent signals with time frequency. This extracts the iris features which are used for matching using conventional algorithms. The accuracy obtained is low because the M-ICA does not give good performance on classseparability.

There are some other researchers who have used different algorithms for feature extraction. Dargham et. al. [76] used thresholding to detect iris from pupil and the surroundings. The detected iris is then reconstructed into a rectangular format. Self-organizing map networks are then used for recognizing the iris patterns. The accuracy obtained by the network is around 83%. In another approach by Li Ma et. al. [77][78] circular symmetry filters are used to capture local texture information of the iris, which are then used to construct a fixed length feature vector. Nearest feature line method is used for iris matching. The results obtained were 0.01% for false match and 2.17% for false non-match rate. In [79], Chen and Yuan developed the algorithm for extracting the iris features based on fractal dimension. The iris zone is partitioned into small blocks in which the local fractal dimension features are computed as the iris code. And finally the patterns are matched using the k-means and neural networks. The results obtained are 91 3% acceptance for authentic person and 100% rejection rate for fakers. Wang er. al. [80] used Gabor filters and 2-D wavelet transforms for feature extraction. For identification weighted Euclidean distance classification has been used. This method is invariant to translation and rotation and tolerant to illumination. The classification rate on using Gabor is 98.3% and the accuracy with wavelets 82.51%. Robert et. al. [81] introduced new algorithm for localization and extraction of iris. For localization a combination of the integro-differential operators with a Hough Transform is used and for feature extraction the concept of instantaneous phase or emergent frequency is used. Iris code is generated by thresholding both the models of emergent frequency and the real and imaginary parts of the instantaneous phase. Finally the matching is performed using Hamming distance. Results gave 11% of the false reject rate was obtained. Lim et. al. [82] used Haar Wavelet transform to extract features from iris images. By applying the transform four times on image of size 450x60 and combining the features 87 bit feature vector was obtained. This feature vector is the compact representation of the iris image.

Finally for classification of feature vectors, weight vector initialization and winner selection strategy has been used. The recognition rate obtained is around 98.4%. In [83] two new methods of the statistical and computer evaluations of the iris structure of a human eye in view of personal identification have been proposed which are based partly on the correlation analysis and partly on the median binary code of commensurable regions of digitized iris image. Similarly method of eye-iris structure characterization using statistical and spectral analysis of color iris images is considered in [84]. Gurianov et.al. [84] used Wiener spectra for characterization of iris patterns. In [85][86] human iris structure is explained and classified using coherent Fourier spectra of the optical transmission.

In [87], an efficient biometric security algorithm for iris recognition system with high performance and high confidence has been described. The system is based on an empirical analysis of the iris image and it is split in several steps using local image properties. The various steps are capturing iris patterns; determining the location of iris boundaries; converting the iris boundary to the stretched polar coordinate system; extracting the iris code based on texture analysis using wavelet transforms and classification of the iris code. The proposed system use the wavelet transforms for texture analysis, and it depends heavily on the knowledge of general structure of human iris. The system has been implemented and tested using a dataset of 240 samples of iris data with different contrast quality.

2.8 Pupil tracking algorithms

The determination of the direction of the gaze with the aid of the pupil is a methodology that uses the smallest boundary between the pupil and iris. Most of the techniques of this class are based on differential illumination using two infrared light sources for creating the phenomenon of dark / light pupil (dark/bright pupil effect). The eye can be traced successfully by recording the pupil in the image with differences resulting from removal of the image of the dark pupil of the image of the luminous pupil. Pupil detection techniques can find many applications. For example, pupil detection can be used for screening diabetic retinopathy, personal identification system, diagnosing of neuropsychiatric disorders, cataract surgery, cataract assessment and liveness detection [88].

In this section, a short description of the three most popular methods used for pupil tracking is presented. For all methods, human faces images must be converted to 8-bit grey scale.

Cumulative Distribution Function (CDF) algorithm.

This method principal is based on the observation that iris and pupil are much dimmer than cornea. Asadifard and Shanbezadeh [89] proposed the algorithm and its name comes from the Cumulative Distribution Function (CDF) of eye luminance used in the algorithm:

$$\text{CDF(r)} = \sum_{w=0}^{r} p(w)$$

where p(w) is the probability of finding point which have luminance equal to w.

In the first step the algorithm change the intensity l(x,y) of each pixel of the input image as :

$$l'(x,y) = \begin{cases} 255 & if\ CDF(l(x,y) < 0.05 \\ 0 & otherwise \end{cases}$$

Parameter 0.05 was chosen experimentally to provide the best possible results.

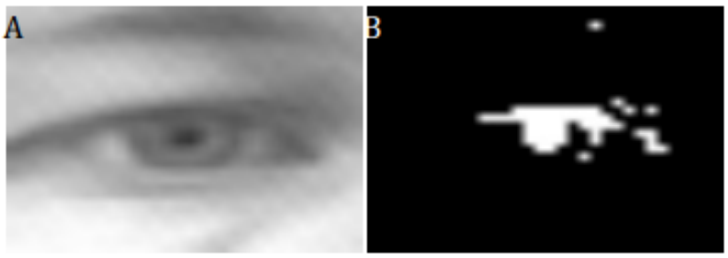

Figure 8: *Transformation results.*
CDF filter: A) input image B) filtered image

The next step is the application of the minimum filter to remove singular white points and compact white region.

Figure 9 *: Application of the minimum filter with radius 2*

Then the algorithm chooses one white pixel, which is the darkest on the original input image. This pixel is called PMI (Pixel with Minimum Intensity). As the probability that PMI belongs to iris and not pupil is significant that further processing is needed. Therefore the algorithm returns to the original image and measures average intensity (AI) in 10x10 pixel square around PMI. Then, the region is expanded to 15x15 pixels and minimum filter is applied. The eye center is assumed to be a geometrical center of points of intensity lower than AI calculated before.

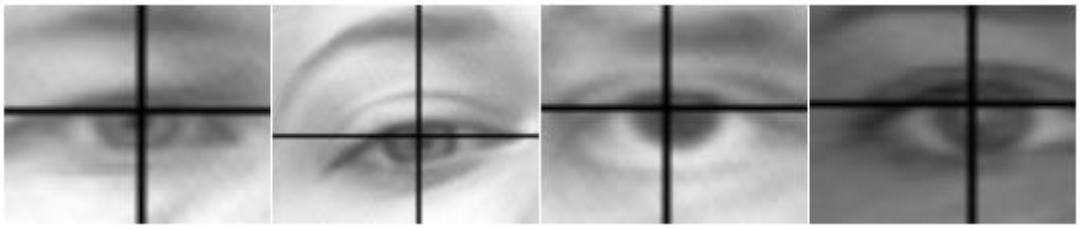

Figure 10: *Examples of eye pupil location using CDF algorithm*

Projection Functions (PF) Algorithm

Zhou and Geng [90] proposed the idea of the method. It is similar to the one used in CDF algorithm with the difference that in this case pixel intensities are projected on vertical and horizontal axes. Those projections divide the whole picture to homogenous subsets.

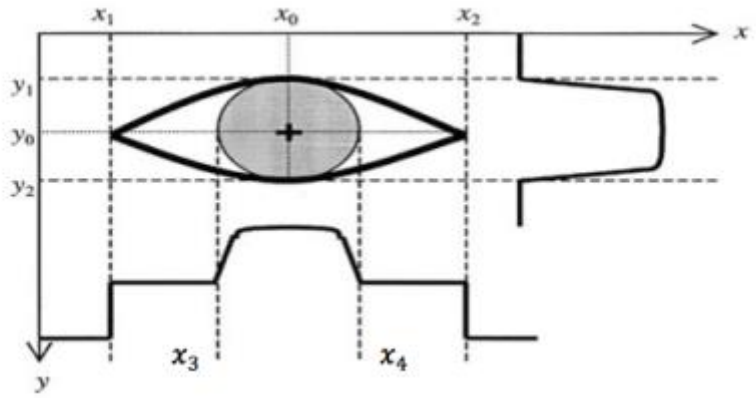

Figure 11*: Projection Function and their relation to pupil position.*

Division points are connected with rapid change of the given PF (horizontal or vertical):

$$\{x_1, x_2, x_3, x_4\} = \left\{x: \left|\frac{d\ PF_v(x)}{dx}\right| > T\right\}, \quad \{y_1, y_2\} = \left\{y: \left|\frac{d\ PF_h(y)}{dy}\right| > T\right\},$$

where T is an arbitrary chosen threshold value. Pupil position (x_0,y_0) is determined by the following function:

$$x_0 = \frac{x_3 + x_4}{2}, \quad y_0 = \frac{y_1 + y_2}{2}.$$

Values x_1,x_2 are not taken into account because they don't provide any information about changes in pupil position in relation to eye concerns. The effectiveness of presented method depends on the specific definition of a projection function. The most popular options are the Integral Projection Function (IPF) and the Variance Projection Function (VPF):

$$IPF_h(y) = \frac{1}{x_b - x_a}\sum_{x=x_a}^{x_b} I(x,y) \qquad IPF_v(x) = \frac{1}{y_b - y_a}\sum_{y=y_a}^{y_b} I(x,y)$$

$$VPF_h(y) = \frac{1}{x_b - x_a}\sum_{x=x_a}^{x_b}\left(I(x,y) - IPF_h(y)\right)^2 \quad VPF_v(x) = \frac{1}{y_b - y_a}\sum_{y=y_a}^{y_b}\left(I(x,y) - IPF_v(x)\right)^2$$

However, the best results are obtained using the General Projection Function (GPF):

$$GPF_h(y) = (1-\alpha)IPF_h(y) + \alpha VPF_h(y), \quad GPF_v(x) = (1-\alpha)IPF_v(x) + \alpha VPF_v(x),$$

with parameter $0 \leq \alpha \leq 1$. Zhou and Geng proved experimentally that the optimal value of α is 0.6, whereas in our tests the best results were obtained for $\alpha = 0$. The following figures illustrate the process of determining projection functions and its efficiency in finding the center of an eye pupil.

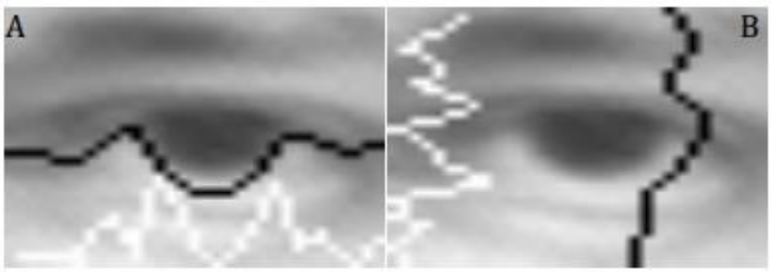

Figure 12: *The plot of A)vertical and B)horizontal General Projection Function(black) and its derivative(white) over a grey scale picture acquired with a webcam.*

Figure 13: *Edges of iris found on a picture presented in Fig. 12*

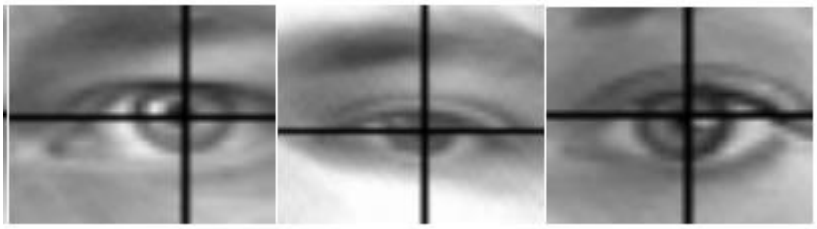

Figure 14: *Examples of pupil location using projection functions.*

Edges Analysis (EA)

This method was first developed from the work of S. Asteriadis, et. al. [91], in which the edge pixel information was used for eye location in a picture of a human face. The input frame is processed by the most popular edges detection algorithm for digital images developed by Canny [92], however before that the Gaussian blur filter is applied to eliminate the undesired noise. The next step is to find the intensity gradients of the image and then apply a non-maximum suppression to get rid of spurious response to edge detection. Afterwards, a double threshold is applied to determine potential edges. Finally, the detection edges are optimized by suppressing all the other edges that are weak and not connected to strong edges.

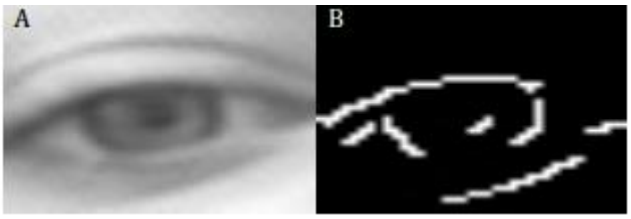

Figure 15: (A) Input image and the (B) processing result of Canny algorithm. Edges are coloured white.

The next step of the pupil detection process is to find vertical and horizontal lines sharing the next to highest number of points with the edges. The intersection of the lines indicates the pupil center [91].

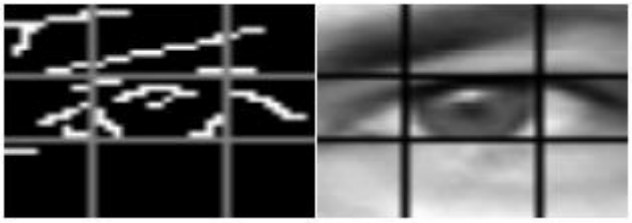

Figure 16: Example of horizontal and vertical lines calculated by the modified edge analysis algorithm.

Additionally the modified method requires the lines to be at least 7 pixels apart (for an eye region of the approximate size of 30x30 pixels) to avoid artefacts occasionally appearing on webcam frames. Having boundary lines, the center of an eye pupil is calculated in the same way as with the PF algorithm described in the previous section.

2.9 Gaze Tracking

Successful eye tracking is a prerequisite of the gaze estimation. In many applications, the head position is fixed using a forehead support, a bite bar or something that will keep the head steady. In some other cases, the head is free to move and head movement is measured with system such as magnetic or video based head trackers. This is the reason of eye and gaze tracking differentiation.

2.9.1 Gaze tracking methods

There are severe gaze tracking methods that are analyzed below.

Kalman filter

Regarding filter Kalman [93], during the decades 1960 1970 1980 used extensively missile trajectory tracking problems and navigating ships. This filter takes to make the assessment of the position of the object (object tracking) but also predicting the position

(prediction) if lost for a bit of sight of the camera. The Kalman filter is a recursive filter that estimates the state of a linear dynamic system from a series of noise measurements.

Assuming that $p(X_{K-1} | Z_{1:K-1})$ is Gaussian can be shown that the $p(X_K | Z_{1:K})$ is also Gaussian.

This applies if:

V_{H-1} and n_k are designed by the Gaussian distributions of known parameters

f_k (x_{k-1}, V_{k-1}) is known as a linear function of x_{k-1}, V_{k-1}

h_k (x_k, n_k) is a known linear function of x_k, n_k

So the initial Bayesian nonlinear functions:

$x_k = f_k(x_{k-1}, v_{k-1})$ and

$z_k = h_k(x_k, n_k)$ can respectively be written as

$x_k = F_k x_{k-1} + n_{k-1}$ and

$z_k = H_k x_k + n_k$

Where F_k, H_k known tables that define the linear functions.

Optical Flow

By the term Optical flow we mean the view in two-dimensional space motion vectors of three-dimensional objects in space, on the level of each image. So this algorithm is trying to calculate whether there are moving pixels between frames of nearby (frames), and in what direction [94]. The optical flow is represented at each point of the image by a vector (u, v) that the «u» is the velocity component along the horizontal axis and the «v» the speed component in the vertical axis.

To calculate the optical flow are three processing steps:

(1) Initially, the image is filtered with a low-pass or band-pass filter to eliminate noise and enhance the ratio of signal to noise.

(2) Then, (carried out) fundamentals such as spatial and temporal differentiation or finding correlations between local surfaces.

(3) Finally, synthesized all these results to obtain a two-dimensional plane of movement and make use of some assumptions to find the Optical Flow.

2.9.2 Extracting techniques of gaze characteristics

Based on the information obtained from the eye region and possibly head pose, the direction of gaze can be estimated. The most important parts of human eye are: the pupil – the aperture that lets light into the eye, the iris – the colored muscle group that controls the diameter of the pupil and the scelera – the white protective tissue that covers the remainder of the eye. Eye detection and tracking remains a very challenging task due to several unique issues, including illumination, viewing angle, occlusion of the eye, head pose etc. Two types of imaging processes are commonly used in video-based eye tracking: visible and infrared spectrum imaging. Gaze estimation methods are broadly grouped into the following gaze estimation categories [Hansen and Ji. (2010)]

- Feature-based
- Appearance-based

Feature-based Gaze Estimation

Feature-based methods explore the characteristics of the human eye to identify a set of distinctive features of the eyes like contours (limbus and pupil contour), eye comers and cornea reflections are the common features used for gaze estimation. The aim of feature-based methods is to identify informative local features of the eye that are generally less sensitive to variations in illumination and viewpoint [Iannizzotto and La Rosa (2011)]. These methods are performance issues in outdoors or under strong ambient light. In addition, the accuracy of gaze estimation decreases when accurate iris and pupil features are not available.

Model-based approaches

Model-based approaches use an explicit geometric model of the eye to estimate 3D gaze direction vector. Most 3D model-based (or geometric) approaches rely on metric information and thus require camera calibration and a global geometric model (external to the eye) of light sources, camera and monitor position and orientation. Most of the model-based method follows a common strategy: first the optical axis of the eye is reconstructed in 3D: the visual axis is reconstructed next: finally the point of gaze is estimated by intersecting the visual axis with the scene geometry. Reconstruction of the optical axis is done by estimation of the cornea and pupil center. By defining the gaze direction vector and integrating it with information about the objects in the scene, the

point of gaze is estimated [Hansen and Ji. (2010)]. For 3D model based approaches, gaze directions are estimated as a vector from the eyeball center to the iris center [Yamazoe et al. (2008); Taba (2012); Sigut and Sidha (2011); Yang et al. (2012); Hung and Yin (2010); Nagamatsu, et al. (2010); Model and Eizenman (2010)].

Interpolation-based approaches

These methods assume the mapping from image features to gaze co-ordinates (2D or 3D) have a particular parametric form such as a polynomial or a nonparametric form as in neural networks. Since the use of a simple linear mapping function in the first video-based eye tracker [Merchant et al. (1974)], polynomial expressions have become one of the most popular mapping techniques [Brolly and Mulligan (2004); Cerrolaza et al. (2012); Morimoto and Mimica (2005); Cerrolaza, et al. (2012)]. Interpolation-based methods avoid explicitly modeling the geometry and physiology of the human eye but instead describe the gazed point as a generic function of image features. Calibration data are used to calculate the unknown coefficients of the mapping function by means of a numerical fitting process, such as multiple linear regressions. As an alternative to parametric expressions, neural network-based eye trackers [Baluja and Pomerleau (1994); Demjen et al. (2011); Torricelli et al. (2008)] assume a nonparametric form to implement the mapping from image features to gaze coordinates. In these approaches, the gaze tracking is done by extracting the coordinates of certain facial points and sending them through a trained neural network, whose output is coordinates of the point where the user is looking at.

Appearance-based Gaze Estimation

Appearance-based methods detect and track eyes directly based on the photometric appearance. Appearance-based techniques use image content to estimate gaze direction by mapping image data to screen coordinates [Javier et al. (2009); Lu et al. (2011)]. The major appearance-based methods [Sheela and Vijaya (2011)] are based on morph able model [Rikert and Jones (1998)], gray scale unit images [Yang (2012)], appearance manifold [Kar-Han et al. (2002)], Gaussian interpolation [Sugano et al. (2012)] and cross-ratio [Flavio et al. (2012)]. Appearance-based methods typically do not require calibration of cameras and geometry data since the mapping is made directly on the image contents.

2.9.3 State of the art

Yoo et al [95] in their paper describe the technique used to achieve eye area gaze estimation. As stated the results showed that the method used was lighter, faster and easier to use in comparison to the conventional systems-methods. Five IR LED and two cameras were used in total in order to reach the final results. Four of the IR LEDs were attached to the four corners of the computer screen to produce reflections of the surface of the cornea and the fifth LED was located at the center of the of the lens of a zoom-lens camera in order to prompt a bright-eye effect. The proposed system measures a gaze point by utilizing the property of projective space. When the LEDs of the monitor were turned on and those of the camera is turned off, the light was reflected on the surface of the cornea generating glints. To estimate the exact eye gaze point, the five glints made by the light reflections and the center of the pupil should be obtained from the image sequences. The feature extraction is imperative to achieve high-performance from any eye gaze estimation system.

Several academic groups have built single-camera systems (Hennessey et al., 2006; Guestrin and Eizenman, 2006; Meyer et al., 2006). Guestrin and Eizenman's system [96] allows only restricted head movements, but it appears that their approach allowed greater head movements with a higher-resolution camera. The main additional difficulty in the single-camera setting is determining the distance of the user from the camera, since a triangulation as in the multi-camera case cannot be carried out. The advantage of a single-camera system is also the reduced cost and smaller size.

The setup that Martin Bohme was a single-camera system (Meyer et al., 2006). It consists of a high-resolution camera (1280x1024 pixels) and two infrared LEDs mounted to the sides of the camera. The LEDs provide general illumination and generate reflexes on the surface of the cornea. These reflexes are used to find the eye in the camera image and determine the location of the center of corneal curvature in space. The software consists of two main components: The image processing algorithms that determine the position of the CRs and pupils in the image, and the gaze estimation algorithm, which uses this data to compute gaze's direction. The image processing component is based on the Starburst algorithm (Li et al., 2005), which was re-implemented and modified to fit the needs of the remote eye tracking setting. The gaze estimation component is based on a physical model of the eye, which models the optical properties of the cornea both reflection and refraction, the location of the pupil

center and center of corneal curvature, and the offset of the fovea from the optical axis. The model contains three user-dependent parameters: the curvature radius of the cornea, the distance between pupil center and cornea center, and the offset of the gaze direction from the optical axis.

Arnor Amir et al [97], tried to create a pervasive eye-gaze tracking system with low complexity, low size, and price .They built a hardware-based embedded system for eye and gaze detection, implemented using simple logic gates, with no CPU and no addressable frame buffers. The image-processing algorithm was redesigned to enable highly parallel, single-pass image-processing implementation. Their system used a CMOS digital imaging sensor and an FPGA for the image processing. It processes 640 · 480 progressive scan frames at a 60 fps rate, and outputs a compact list of sub-pixel accurate (x,y) eyes coordinates through USB communication. This design, operating at the sensors pixel clock, is suitable for single-chip eye detection and eye-gaze tracking sensors.

Noureddin et al [98], in their paper presented a novel design for a non-contact eye detection and gaze tracking device. It uses two cameras to achieve real-time tracking of a person's eye in the during head motion. Image analysis techniques were used to gather accurate pupil locations and corneal reflections. All the computations were performed in software and the device requires simple, compact optics and electronics attached to the user's computer. Three methods of estimating the user's point of gaze on a computer monitor were evaluated. The camera motion system was capable of tracking the user's eye in real-time (9 fps) in the presence of natural head movements as fast as 100 degrees horizontally and 77 degrees vertically. Experiments using synthetic images have shown its ability to track the location of the eye in an image to within 0.758 pixels horizontally and 0.492 pixels vertically. The system had also been tested with users with different eye colours and shapes, different ambient lighting conditions and the use of eyeglasses. A gaze accuracy of 2.9 degrees was observed.

2.10 Human eye as an emotion indicator

Previous studies have demonstrated that emotions are differentially expressed on specific regions of the face. For example, negative emotions are expressed more on the upper part of the face, whereas positive emotions are expressed more on the lower part of the face. An individual may reach inaccurate conclusions about the environment if

he/she fails to look at relevant cues or to integrate what he/she has seen. Thus, it is reasonable to assume that better visual attention to specific face regions may result in better performance in emotion recognition. Information gathered from visual processing using eye tracking technology may help clarify the nature of the emotion perception impairments.

More specifically, the emotional signals are primarily based on the eye region. Pupil size, blink properties and gaze are analyzed and used to calculate an index of emotional reaction. Pupil size has been coupled with activation of the sympathetic nervous system and its size variation is known to be related to emotional reactions. The main function of the pupil is to regulate the amount of light entering the eye adjusting its diameter. However, the relationship is complex because pupil size is also related to cognitive processing load and the amount of light or hue in visual stimuli. Blink has also been related to emotional reactions, for example with defensive reactions like emotion-modulated eye blink startle. Basic emotions (anger, disgust, fear, joy, sadness, and surprise) involve a specific change in one element of the eye region (brow and furrow movement, pupil size variation, and other eye region muscle variations).

A research supporting that emotions (fearful and neutral) can influence gaze (Mathews et al., 2003; Putman et al., 2006; Tipples, 2006). These results indicate that fearful facial expressions are chiefly critical in affecting the direction of attention in anxious people, as fearful, neutral, angry, and happy expressions were all conferred to the same party in an irregular order. Thus, the present results show that it is not just negative expressions in general (e.g., fearful and angry) that are meaningful; fairly, the boost orienting appears to be distinct for fearful expressions. This discovery makes sense in that the accordance of eye gaze and emotion can give an important communal hint to the origin of menace. An angry face looking straight at a person indicates clearly where the threat is located, as does a fearful face looking at another locale. In contradiction, an angry face looking somewhere else or a fearful face looking directly at someone is a more puzzling consolidation (Adams & Kleck, 2002).

Support for this belief was also found in testings with direct gaze, in which participants with high trait anxiety were sluggish to react when the face showed anger in comparison to happy or neutral expressions. A programmed analysis showed that this lag on trials with angry straight gaze expressions diverge from fearful faces with a straight gaze. These outcomes are rational with the report of Adams and Kleck (2005),

which countenance that straight gaze boost the perception of angry expressions and has no influence on the conception of dreadful expressions.

Also Andrew J. Calder investigated the role of neutral, happy, fearful, and angry facial expressions and how much these expressions influenced the direction of eye gaze [99]. The findings of their experimentation are simple. In concession with earlier studies, the stress level characteristic boosted the orientation reply caused by deter gaze in fearful relative to the neutral facial expression (Mathews et al, 2003, Putman et al, 2006, Alcohol, 2006). An important finding was that this increased orientation response was specific for scared expressions and gaze cuing effect was not enhanced for angry or happy in comparison with neutral expressions. Indeed, the effect look kinship caused by angry expression was indeed reduced for high-trait anxious humans than other facial utterance. This conclusion contrasts with a report that angry expressions closer-look cuing results than happy and neutral expressions (Holmes et al., 2006, Experiment 3). However, this standard was only a post hoc analysis and critical Anxiety × Match × interaction term was not significant in Holmes et al. (2006) report. Andrew J. Calder's results show a very different pattern, and the interaction was considerable. There is a need for further research, Calder says, to clarify whether prevented fearful expressions strengthening of attention orientation results compared with angry facial expression on the stress people, as found in their study. A primary objective of their study was to determine whether there was general interaction between the result gaze cuing and facial emotional expression. This interaction has been underpinned, but modified to any significant extent by the level of self-reported feature anxiety.

Thus, as in previous surveys, observing a fearful facial expression with an averted gaze resulted in increased power cuing for people with high (but not low) levels trait anxiety (Mathews et al, 2003; Putman, et al, 2006; Spirits, 2006). Hietanen and Leppanen (2003) concluded after making six experiments that facial expressions are not affecting the orientation of attention caused by the direction of the gaze. Although other studies have also generally failed to find effects of facial expression to gaze cuing effects in unselected samples (eg, Mathews et al., 2003), this research stresses the weight of taking into account individual differences in stress as a decisive of the ability of emotional expressions cause increased results orientation. Other studies supporting this conclusion presented only fearful and neutral utterances in a gaze cuing task (Mathews et al., 2003; Putman et al., 2006; Tipples, 2006). Calder's results showed that fear facial expression is particularly important in affecting the orientation of attention

in distress people, such as fear, neutral, angry, and happy expressions were all presented with the same Participants in a random order. His results also show that not only negative expressions in general (e.g., scared and angry) that are important; rather, the enhanced orienting appears to be specific for fearful expressions. This conclusion makes sense in that the accordance of eye gaze and emotion can provide an important social cue to the source of menace. An angry face looking directly at someone suggests clearly where the threat is coming from, as does a fearful face looking at another location. In contrast, an angry face enhanced looking away or a fearful face looking directly at you is a more ambiguous combination.

Finally, Jakob de Lemos et al [100] introduced a state-of-the-art automated method of measuring human emotions by analyzing eye assets through an eye tracking platform. They experiment was done using still images displayed on a monitor, in a controlled test environment. The method uses eye tracking data and makes it possible to measure the actual unconscious and uncontrollable emotional responses sooner than they are cognitively perceived, interpreted, and biased by our brain. The system (iET system) they made is an automation of the methods analogous to the ones used by e.g. psychologists, witness experts and others, who have been qualified to spot the subtle diverges in facial expressions, to find emotional signals from subjects. More detailed, the emotional signals are primarily based on the eye region. Pupil size, blink properties and gaze Jakob says are analyzed and used to calculate an index of emotional reaction, using modern eye tracking equipment and specifically developed software. Pupil size is known to be related to emotional reactions. Pupil dilation has been connected with activation of the sympathetic nervous system. The relationship is complex because pupil size is also related to cognitive processing load and the amount of light or hue in visual stimuli. Blink has also been related to emotional reactions, for example with defensive reactions like emotion modulated eye blink startle. Lastly, gaze patterns have been related to emotional counteractions. The human brain is able to put fine information together in a manner that can disclose the subjects' emotions.

Chapter 3 – Data description

3.1 Available Databases – Datasets

Within the past decade, many computational approaches have been developed to estimate gaze directions. Most researchers used common face datasets with only a limited representation of different head poses to train and verify their algorithms. Moreover, in most datasets, faces have neither a defined gaze direction, nor do they incorporate different combinations of eye gaze and head pose [101].

Currently, there exist several face databases with a large number of subjects that cover at least a sparse set of head poses [102]. Moreover, some databases are available, that capture different combinations of pose, illumination [103] or with mimic expression [104] [105]. A summary of many datasets is given in Table 1 in Appendix.

3.2 Self-recorded videos

Considering the above databases requirements, we attempted to generate a number of self-recorded videos. The recordings took place in CML (Computational Biomedicine Laboratory). The experimental setup consisted of a basic web-camera (Turbo-X Webcam FHD100). All videos recorded at 24 fps, in 24-bit RGB color with a resolution of 800 × 600 pixels, lasted a total of 39 seconds and were stored in uncompressed AVI format. The experiments were conducted indoors, at different times of the day, with a varying amount of indirect sunlight and fluorescent lights as the only sources of illumination. The participants were seated at a distance of approximately 0.5 m from the cameras. Eight subjects (5 males and 3 females) participated at the recordings, none wore glasses.

They were asked to watch a video (https://www.youtube.com/watch?v=SGHwjKi4_V0) which included a grey dot moving at time intervals of 4 sec at directions: a) up, b) down, c) left, d) right, e) up-left, g) up-right, h) down-left, j) down-right k) center. The procedure is shown in figures 8 - 16.

EXPERIMENTAL VIDEO PROCEDURE

***Figure 17**: Up*

***Figure 18**: Down*

***Figure 19**: Left*

***Figure 20**: Right*

***Figure 21**: Up-Left*

***Figure 22:** Up - Right*

***Figure 23**: Down-Left*

***Figure 24**: Down-Right*

***Figure 25**: Center*

All males had a light beard. They were asked to sit naturally and look straight at the cameras during video capturing. They were free to move their head or body slightly while remaining seated. One facial video, each lasting 40 seconds, were simultaneously recorded for all participants from the camera. In total, 8 separate videos were stored for analysis (8 subjects x 1 video).

3.3 Semeoticons Experiment

The central idea in the FP7 Project SEMEOTICONS (http://www.semeoticons.eu/), which stands for SEMEiotic Oriented Technology for Individual's CardiOmetabolic risk self-assessmeNt and Self-monitoring, is to exploit the face as a major indicator of individual's well-being by tracing traits of physical and expressive status. In accordance to a semeiotics viewpoint, face signs were mapped to measures and computational descriptors, automatically assessed. SEMEOTICONS designed and constructed an innovative multi-sensory system integrated into a hardware platform having the exterior aspect of a mirror: the so-called Wize Mirror. This will easily fit into users' home or other sites of their daily life (e.g. fitness and nutritional centers, pharmacies, schools and so on). The experimental setup consisted of a Point Grey Grasshopper USB camera with lens of fixed focal length of 16 mm. All videos were recorded at 90 fps, in 24-bit RGB color with a resolution of 1600 × 1216 pixels. Semeoticons has used a high end camera, several psychological states were invested such as anxiety, stress or fatigue. These needs dictate specific requirements on the selected camera for the acquisition of raw data of such quality that could enable the computation of the selected descriptors with satisfactory accuracy. Videos of a total of 24 subjects in several emotional states such as neutral state, describing themselves, mimicking situations of stress or anxiety

or fatigue, during a Stroop test, and watching a set of relaxing or stressful images or videos, was created.

Chapter 4 – Technical Implementation

4.1 Algorithms description

4.1.1 Eye region segmentation algorithm description

The Region of Interest (ROI) of the image is then set using the EyeRegionTransform function. The function crops the ROI and returns two images consisting only of the left and right eye. This was necessary since the eye detection had to work with two different regions in order to be more accurate and have more specific results. At this point a reduced region of the face that consist of the eyes has been obtained.

4.1.2. Iris segmentation description

Iris segmentation refers to the process of automatically detecting the pupils (inner) and limbus (outer) boundaries of an iris in a given figure. This process helps to extract features from the differentiated surface of the iris, while excluding the surrounding regions. Iris segmentation has a key role in the performance of an iris recognition and detection system. This is because incorrect segmentation can lead to incorrect feature extraction from less discreet areas (e.g., facial hair, eyelashes, eyelids, pupil, etc.), therefore decreasing the recognition accuracy. After the iris estimation a blue circle is placed on the Iris region.

4.1.3. Gaze tracking algorithm description

At this point the ROI have been isolated from the rest of the face. We used CircleFinder function to identify the Pupil and the Iris. After the isolation of those two areas we proceed in Gaze tracking. We used the GazeTracking function to find the gaze based on the glare of the light inside the pupil. Depending of the position of the glimpse we were able to estimate the gaze position.

4.2 Gaze estimation Application

4.2.1 Region of Interest

In order to track the gaze of the participants we have focused on the eye region to be used for gaze estimation. For the best results in gaze tracking Cascade Object Detector was implemented in order to find the blinks and the closure of the eye. The algorithm

required as inputs the two regions of the eyes. Firstly, left and right eye regions were segmented from the rest of the face. In order to achieve that the video was divided in 24 video frames (936 video frames total) and the process continued for every frame of the video.

In all cases both eyes were in the ROI with an 85% accuracy. The other 15% loss was a result of bad lighting conditions or reduced region of interest. The most severe problems were faced when the lighting conditions were bad (e.g. low lighting, contrast lighting, very bright lighting, no lighting etc.). The algorithm could not estimate the gaze and in many cases the lighting conditions mislead the algorithm and that decreased the estimation accuracy. Therefore, room lighting and natural lighting was used for optimal results.

Also in other cases when female participants had long hair that engaged the face the algorithm was inaccurate to estimate the gaze especially when these hairs engaged the ROI. For that reason, these female participants were asked to gather their hair. When it comes to hair it is worth mentioning the facial hair (e.g. eyebrows, eyelashes). In some cases, the eyebrows confused the algorithm into estimating the ROI and the gaze position wrong.

The gaze estimation accuracy was mostly influenced by the distance between the camera and the subject. The distance was kept constant at 0.5m through the experiment and when that changed the efficiency of the algorithm decreased severely.

4.2.2. Gaze estimation

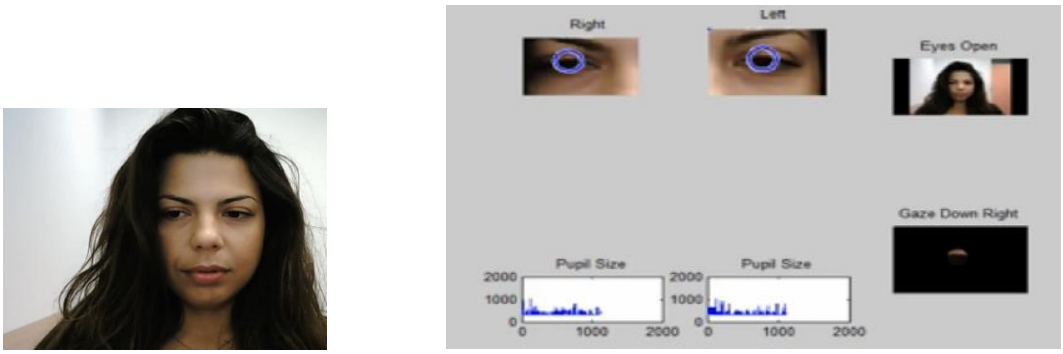

Figure 26: *The face (left) and the eye region (right)*

(published with subject's permission).

The comparative results presented in Table 3 indicate that eye region provides errors even when the tracker follows the gaze. The results were as follows.

***Table 2**: Estimated accuracy of the gaze tracker*

Dataset	Region	Number of Subjects	Number of Recordings	Reference	Error (percentage)	Success (percentage)
Self - recorded videos	*Eye*	*8*	*8*	*Gaze Tracker*	*26%*	*74%*

5.1 Gaze Tracking Estimation Analysis

The overall results of the experiment are presented in Table below.

***Table 3**: Estimated accuracy of the gaze tracker per person*

Subject Number	***Success (percentage)***	***Error (percentage)***
Subject 1	80%	20%
Subject 2	74%	26%
Subject 3	72%	28%
Subject 4	68%	32%
Subject 5	77%	23%
Subject 6	79%	21%
Subject 7	69%	31%
Subject 8	73%	27%
Mean	**74%**	**26%**

The overall mean success as shown above was 74% and the overall error was 26%.

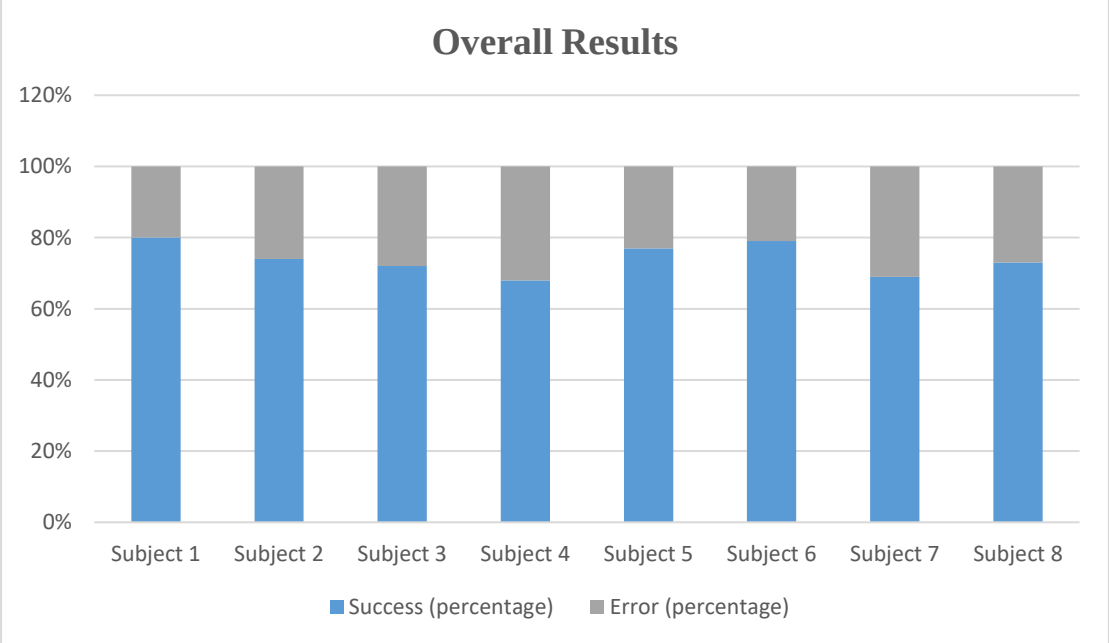

***Figure 27:** Overall results graph*

For each direction the results are given in Table 4.

***Table 4**: Estimated accuracy of the gaze tracker per grey dot direction*

Grey Dot Direction	*Success (percentage)*	*Error (percentage)*
Up	74%	26%
Down	65%	35%
Left	73%	27%
Right	71%	29%
Up-Left	72%	28%
Up-Right	71%	29%
Down-Left	68%	32%
Down-Right	67%	33%
Center	90%	10%

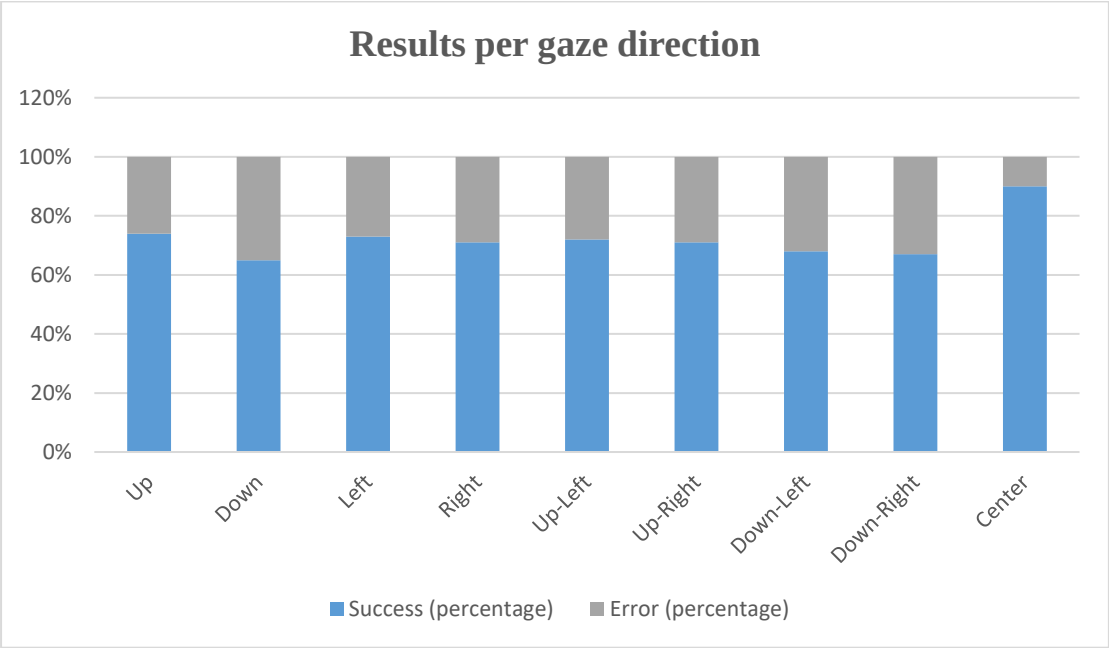

***Figure 28**: Results per gaze direction*

Chapter 6 - Discussion

.

The results of this system showed that the used gaze detection system was very promising. Since the system itself was designed to be simple, there are certain limitations that in many cases reduced the system's performance. The ROI was found in all cases successfully and the differentiation between right and left eye was done correctly. The distance between the subject and the recording camera was kept constant at 0.5m in all cases. When a higher distance was tried, the procedure stopped.

The lighting conditions played an important role in the gaze estimation. The algorithm was working properly only in good lighting conditions so the room lighting and natural lighting was used.

The mean of the overall success percentage for gaze estimation was 74% with the maximum percentage in one subject to be 80% and the minimum 68%. No significant difference was observed on males or females although it was noticed that subjects with larger eyes had better results in gaze tracking, in comparison with those who had smaller eyes. Also subjects with darker eyes had better results than those with open colored eyes due to the fact that the light glare inside pupil was easier to be tracked from the algorithm.

The algorithm also had different success rates depending on what position the grey dot was while the subject was watching the recorded video. For the 9 different locations, success rates was calculated. The best algorithm performance was recorded when the grey dot was in the center of the screen with 90% accuracy. When the grey dot was in the upper side of the screen the results were better than ones in the lower side. The success percentage for the upper grey dot positions varied from 71% to 74% while the lower grey dot positions varied from 65% to 68%. That was because, when the gaze was in the lower part of the screen the eyelids partially covered the iris and that led to decrease the efficiency of the system.

Chapter 7 - Conclusions & Future Work

In conclusion, the work reported in this thesis proves that accurate contactless measurement of gaze tracking, in RGB video, in real life conditions is possible. This confidence provides the required evidence for the commercial exploitation of such systems that are easy to use and easy to implement in everyday situations.

In all cases, there are still open issues that need to be addressed in the future. As seen in the experiments the accuracy of the estimation is significantly influenced when people have long or short hair. Also the lighting conditions are a key to the success of the gaze estimation since darker conditions provide low results. In addition, if head movement was added as a parameter in the experiment it probably would cause the deterioration imprint and this remains an issue that requires attention.

This thesis has aimed to show if it is possible to create a system that uses a regular web camera to effectively track the gaze of the user and though this system proved to be accurate enough. Future work includes the use of the system combined with videos that will create feelings to the subject (such as joy, fear, sadness etc.) in order to correlate the gaze with these emotions.

APPENDIX

Table 1 *- Datasets for gaze area estimation*

Title	Clips	Resolution	Type	Subjects	Emotions
Wallhoff et al., 2006	Video-clips	Laboratory, colour camera, 640x480@25 fps	Facial Video	18 subjects	Ekman's basic emotions + neutral
McIntyre, 2010	Movie-clips	Laboratory, colour camera, 800x600@24.94 fps	Facial Video	16 subjects+11 MDD patients	Ekman's basic emotions –Disgust
Soleymani et al., 2012	Movie-clips	Laboratory, 6 cameras (1 colour + 5 B&W), 780x580@60 fps	Facial Video, Audio, Eye Gaze, ECG, GSR, Resp, Temp	27 subjects, 50' sessions+30' setup	Three Factor Theory of Emotions*
MIT data set emotions	Skin-Surface sensing		wearable computer system	1 subject	neutral, anger, hate, grief, love, romantic love, joy, and reverence
MMI	Movie-clips	640x480 max	Facial video and sound	48 subjects	Neutral- six basic emotions circle
MAHNOB-HCI-FULL	Recording from multiple cameras	780x580pixels 60fps	Head-worn and room microphone, eye gaze and peripheral/central nervous system physiological signals	30 subjects	

** Three Factor Theory of Emotions: pleasure-displeasure, degree of arousal, and dominance-submissiveness.*

References

[1] F. Thomas and O. Johnston, Disney Animation: The Illusion of Life, Abbeville Press, 1981.

[2] Diefendorf A. R., Dodge R. (1908). An experimental study of the ocular reactions of the insane from photographic records. Brain 31, 451–489. 10.1093/brain/31.3.451

[3] Bayless, S. J., Glover, M., Taylor, M. J. & 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, 483-510

[4] Dimberg U, Peterson M (2000) Facial reactions to happy and angry facial expressions: evidence for right hemisphere dominance. Psychophysiology 37: 693–696

[5] Morrison, R. L. & Bellack, A.S (1981). The role of social perception in social skill. Behavior Therapy 12, 69-79

[6] Hood BM, Willen JD, Driver J. Adult's eyes trigger shifts of visual attention in human infants. Psychological Science. 1998;9:131–134

[7] Stephen, S.V. 2010. Following gaze: Gaze following behavior as a window into social recognition. Frontiers in Integrative Neuroscience 4:5

[8] Lance, Brent J, Marsella, Stacy C (2010). The Expressive Gaze Model: Using Gaze to Express Emotion, IEEE Comput Graph Appl. 2010 Jul-Aug; 30(4):62-73. doi: 10.1109/MCG.2010.43

[9] Cowan D.C., The Empathic Gaze and How to Find it: Eye-gaze Behaviour to Expressions of Emotion, PhD Thesis, University of Queesland, 2015.

[10] Stern, D. N. (1977). The first relationship: Infant and mother. Cambridge, MA: Harvard University Press.

[11] Mehrabian, A. (1967). Attitudes inferred from non-immediacy of verbal communications. Journal of Verbal Learning and Verbal Behavior, 6, 294–295

[12] Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & BaronCohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. Visual Cognition, 6, 509–540.

[13] Macrae, C. N., Hood, B. M., Milne, A. B., Rowe, A. C., & Mason, M. F. (2002). Are you looking at me? Gaze and person perception. Psychological Science, 13, 460–464.

[14] Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. Proceedings of the National Academy of Sciences, USA, 99, 9602–9605.

[15] Baron-Cohen, S. (1995). Theory of mind and face-processing: How do they interact in development and psychopathology? New York: Wiley.

[16] Ekman, P. (1973). Darwin and facial expression: A century of research in review. New York: Academic Press.

[17] Russell, J. A. (1997). The psychology of facial expression. New York: Cambridge University Press.

[18] Boundless. Vision: The Visual System, the Eye, and Color Vision." Boundless Psychology. 21 Jun. 2016. Retrieved 20 Sep. 2016

[19] Jüttner M., Physiological Optics, in T. G. Brown (Ed.). The Optics Encyclopedia (Wiley-VCH, Berlin, 2004), Vol. 4, p. 2511

[20] Kolb H., Fernandez E., Nelson R. Webvision: The Organization of the Retina and Visual System.

[21] McCaa C.S. The Eye and Visual Nervous System: Anatomy, Physiology and Toxicology, Environmental Health Perspectives, Vol.44, pp.1-8, 1982

[22] Gross H. Handbook of Optical Systems: Vol. 4 Survey of Optical Instruments., 2008

[23] Hayes W.A., Dixon R. L., Toxicology of the Eye, Ear and Other Special Senses. The Quarterly Review of Biology 1986 61:2, 298-299

[24] Lemos J., Sadeghnia G.R., Ólafsdóttir I., Jensen O. (2008): Measuring emotions using eye tracking, Proceedings of Measuring Behavior 2008 (Maastricht, The Netherlands, August 26-29, 2008)

[25] Singh H., Singh J. Human Eye Tracking and Related Issues: A Review. International Journal of Scientific and Research Publications, Volume 2, Issue 9, September 2012 ISSN 2250-3153

[26] Vitte E, Sémont A: Assessment of vestibular function by videonystagmoscopy. *J Vestib Res*1995;5(4):1–7

[27] Schreiber K., Haslwanter T. (April 2004). Improving calibration of 3-D video oculography systems. *IEEE Transactions on Biomedical Engineering*. 51 (4): 676–679.doi:10.1109/TBME.2003.821025

[28] Siddiqui U., Shaikh A. N. An Overview of "Electrooculography". International Journal of Advanced Research in Computer and Communication Engineering Vol. 2,

Issue 11, November 2013

[29] Yagi, T., Kuno, Y., Koga, K., Mukai. ”Drifting and blinking compensation in electro-oculography (EOG) eye-gaze interface”. Proceedings of 2006 IEEE Int. Conf

[30] Edwards G. J., Taylor, C. J., Cootes, T. F. (1998). "Interpreting face images using active appearance models". Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition. p. 300. *doi*:*10.1109/AFGR.1998.670965*. *ISBN 0-8186-8344-9*.

[31] Cootes, T. F.; Edwards, G. J.; Taylor, C. J. (1998). "Active appearance models". *Computer Vision — ECCV'98*. Lecture Notes in Computer Science. 1407. p. 484. doi:10.1007/BFb0054760. ISBN 3-540-64613-2.

[32] Cootes, T. F.; Edwards, G. J.; Taylor, C. J. (2001). "Active appearance models". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 23 (6): 681. doi:10.1109/34.927467

[33] Dinov, ID. "Expectation Maximization and Mixture Modeling Tutorial". *California Digital Library*, Statistics Online Computational Resource, Paper EM_MM,http://repositories.cdlib.org/socr/EM_MM, December 9, 2008

[34] Reynolds, D.A.; Rose, R.C. (January 1995). "Robust text-independent speaker identification using Gaussian mixture speaker models". IEEE Transactions on Speech and Audio Processing. 3 (1): 72–83. doi:10.1109/89.365379

[35] Dalal N. Triggs B. Histograms of Oriented Gradients for Human Detection. In CVPR, pages 886-893, 2005

[36] Iqbal T., A robust real time eye tracking and gaze estimation system using particle filters. Department of Computer Science, University of Texas at El Paso. 2012

[37] Huang D., Shan C., Ardebilian M., Wang Y., Chen L., Local Binary Patterns and Its Application to Facial Image Analysis: A Survey

[38] He D.C., Wang L.(1990), "Texture Unit, Texture Spectrum, And Texture Analysis", Geoscience and Remote Sensing, IEEE Transactions on, vol. 28, pp. 509 – 512

[39] Wang L., He D.C. (1990), "Texture Classification Using Texture Spectrum", Pattern Recognition, Vol. 23, No. 8, pp. 905 - 910

[40] Ojala T., Pietikäinen M., Harwood D. (1994), "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions", Proceedings of the 12th IAPR International Conference on Pattern Recognition (ICPR 1994), vol. 1, pp. 582 - 585

[41] Ojala T., Pietikäinen M., Harwood D. (1996), "A Comparative Study of Texture

Measures with Classification Based on Feature Distributions", Pattern Recognition, vol. 29, pp. 51-59.

[42] Shylaja S. S., Murthy K. N. B., Natarajan S., Kumar N. Agarwal R.. Effective Eye Localization using Local Binary Patterns. IJCA Proceedings on International Conference and workshop on Emerging Trends in Technology (ICWET 2012) icwet(2):40-47, March 2012.

[43] Baum L. E., Petrie T. (1966). "Statistical Inference for Probabilistic Functions of Finite State Markov Chains". The Annals of Mathematical Statistics. 37 (6): 1554–1563. doi:10.1214/aoms/1177699147. Retrieved 28 November 2011.

[44] Baum L. E., Eagon, J. A. (1967). "An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology". Bulletin of the American Mathematical Society. 73 (3): 360. doi:10.1090/S0002-9904-1967-11751-8. Zbl 0157.11101

[45] Stratonovich, R.L. (1960). "Conditional Markov Processes". Theory of Probability and its Applications. 5 (2): 156–178. doi:10.1137/1105015

[46] Sodergren M.H., Orihuela-Espina F., Clark J. et al. A hidden Markov model-based analysis framework using eye-tracking data to characterize re-orientation strategies in minimally invasive surgery. Cogn Process (2010) 11: 275. doi:10.1007/s10339-009-0350-3

[47] Viola P., Jones M. J., Robust Real-Time Face Detection, International Journal of Computer Vision 57(2), 137-154, 2004.

[48] Chuk T., Ng A. C. W., Coviello E., Chan A. B., Hsiao J. H.. Understanding eye movements in face recognition with hidden Markov model. Journal of Vision September 2014, Vol.14, 8. doi:10.1167/14.11.8

[49] Freund Y., Schapire R. E., A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1):119–139, August 1997

[50] Lucas B. D., Kanade T. An iterative Image Registration Technique with an Application to Stereo Vision. International Joint Conference on Artificial Intelligence, pages 674-679, 1981.

[51] Tomasi C., Kanade T. Detection and Tracking of Point Features. Carnegie Mellon University Technical Report CMU-CS-91-132, April 1991

[52] Kim KN., Ramakrishna R. S. Vision-based eye-gaze tracking for human computer interface. IEEE International Conf. on Systems, Man, and Cybernetics, 1999

[53] Daugman J. The importance of being random: statistical principles of iris recognition. Pattern Recognition, 36(2):279–291, 2003

[54] Hansen D.W., Pece A.E.C. Eye tracking in the wild. Computer Vision and Image Understanding, 98(1):182–210, April 2005.

[55] Perez A., Cordoba M.L., Garcia A., Mendez R., Munoz M.L., Pedraza J.L., Sanchez F. A precise eye-gaze detection and tracking system. Journal of WSCG, pages 105–8, 2003.)

[56] Nixon M. Eye spacing measurement for facial recognition. In Proceedings of the Society of Photo-Optical Instrument Engineers, 1985

[57] Young D., Tunley H., Samuels R. Specialised hough transform and active contour methods for real-time eye tracking. Technical Report 386, School of Cognitive and Computing Sciences, University of Sussex, 1995

[58] Kothari R., Mitchell J.L. Detection of eye locations in unconstrained visual images. Image Processing, 1996. Proceedings. International Conference on, 3:519–522, 1996.

[59] Valenti R. Gevers T. Accurate eye center location and tracking using isophote curvature. In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2008.

[60] Li D., Winfield D., Parkhurst D.J. Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. In Proceedings of the Vision for Human-Computer Interaction Workshop, IEEE Computer Vision and Pattern Recognition Conference, 2005.

[61] Cootes T.F., Taylor. Active shape models – 'smart snakes'. In Proc. British Machine Vision Conf., BMVC92, pages 266–275, 1992

[62] Bala J., DeJong K., Huang J., Vafaie H., Wechsler H. Visual routine for eye detection using hybrid genetic architectures. In Proc. of International Conf. on Pattern Recognition, Vienna, Austria, 1996

[63] Feng G.C., Yuen P.C. Variance projection function and its application to eye detection for human face recognition. International Journal of Computer Vision, 19:899–906, 1998

[64] Feng C.G., Yuen P.C. Multi-cues eye detection on gray intensity image. Pattern recognition, 34:1033–1046, 2001.

[65] Kawato S., Ohya J. Real-time detection of nodding and head-shaking by directly detecting and tracking the between-eyes. In Proc. IEEE 4th Int. Conf. on Automatic Face and Gesture Recognition, pages 40–45, 2000

[66] Kawato S., Ohya J. Two-step approach for real-time eye tracking with a new

filtering technique. In Proc. Int. Conf. on System, Man & Cybernetics, pages 1366–1371, 2000

[67] Kawato S., Tetsutani N. Detection and tracking of eyes for gaze-camera control. In Proc. of 15th International Conference on Vision Interface, 2002.

[68] Kawato S., Tetsutani N. Real-time detection of between-the-eyes with a circle frequencey filter. In Proc. of Asian Conference of Computer Vision 2002, volume II, pages 442–447, 2002

[69] Daugman J. G., High confidence visual recognition of persons by a test of statistical independence, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume: 15, No. 1 I, (l993), pp. 1148-1 161

[70] Daugman J. G;, Statistical Richness of Visual Phase Information: Update on Recognizing Persons by Iris Patterns, International Journal of Computer Vision, Vol 45, No.1, (2001), pp. 25-38

[71] Wildes R. P. Iris Recognition: An Emerging Biometric Technology, Proceedings of the IEEE. Vol 85, No. 9, (1999), pp.1348-1363

[72] Boles W.W, Boashash B., A human identification technique using inzages of the iris and wavelet transform, IEEE Transactions on Signal Processing, Vol.46, No. 4, (1998), pp. 1185-1188.

[73] Mallat S. G., Zero-crossing of a Wavelet Transform, IEEE Transactions, on Information Theory, Vol. 37. No. 14, (1991), pp. 1019 - 1033

[74] Sanchez-Avila C., Sanchez-Rei R., de Martin-Roche D., Iris recognition for biometric identification using dyadic wavelet transform zero-crossing, Proceedings of the IEEE 35th International. Camahan Conference on Security Technology, (2001). pp. 272 -277.

[75] Noh S., Pae K., Lee C., Kim J. Multiresolution Independent Component Identification, Proceedings of the 2002 International Technical Conference on Circuit Systems, Computers and Communications, Phuket, Thailand, (2002).

[76] Dargham, J. A., Chekima, A., Liau Chung Fan and Lye Wil Liam, Iris recognition using sev- organizing neural network, Student Conference on Research and Development, (2002), pp. 169 -172

[77] Ma L.., Tan T., Wang Y. Iris recognition based on multichannel Gabor filtering, Proceedings of the International Conference on Asian Conference on Computer Vision, (2002), pp. 1-5.

[78] Ma L., Tan T., Wang Y., Iris recognition using circular symmetric filters,

Proceedings of the 16th International Conference on Pattern Recognition, .Vol. 2, (2002), pp. 414 -417.

[79] Chen W. S., Yuan S. Y. A Novel Personal Biometric Authentication Technique Using Human Iris Based on Fractal Dimension Features, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, (2003).

[80] Zhu Y., Tan T., Wang Y. Biometric Personal Identification Based on Iris Patterns, Proceedings of the IEEE International Conference on Pattern Recognition, (2000), pp. 2801-2804

[81] Tisse CL., Martin L., Torres L., Robert M. Person Identification Technique Using Human Iris Recognition, Proceedings of the 15th International Conference on Vision Interface (2002).

[82] Lim S., Lee K., Byeon O., Kim T., Efficient Iris Recognition through Improvement of Feature Vector and Classifier, Journal of Electronics and Telecommunication Research Institute, Vol. 23, No. 2, (2001), pp. 61 - 70.

[83] Machala, Libor, Pospisil, Jaroslav, Alternatives of the statistical evaluation of the human iris structure, Proceedings of the SPIE Vol. 4356, (2001) pp. 385-393.

[84] Gurianov E. V., Zimnyakov. D. A., Galanzha, V. A., Iris patterns characterization by use of Wiener spectra analysis: potentialities and restrictions, Proceedings of the SPIE Vol. 4242, (2001), pp. 286- 290.

[85] Kois P., Muron A., Pospisil J. Human iris structure by the method of coherent optical Fourier transform, Proceedings of the SPIE Vol. 4356, (2001), pp.394-400.

[86] Ales M., Petr K., Pospisil J. Identification of persons by means of the Fourier spectra of the optical transmission binary models of the human iris, Optics Communications, Vol. 192, (2001)

[87] Ali J.M.H., Hassanien A.E. An Iris Recognition System to Enhance E-security Environment Based on Wavelet Theory, AM0 - Advanced Modeling and Optimization, Volume 5, No. 2, (2003), pp. 93-104.

[88] Ciesla M. Eye Pupil Location Using Webcam. Online. Accessed on 31 October 2012. Available from: http://arxiv.org/ftp/arxiv/papers/1202/1202. 6517.pdf

[89] Asadifard M., Shanbezadeh J, Automatic Adaptive Center of Pupil Detection Using Face Detection and CDF Analysis, Available from: http://www.iaeng.org/publication/IMECS2010/IMECS2010_pp130F133.pdf

[90] Zhou Z., Geng X. Projection functions for eye detection, Available from: ftp://ftp.dii.unisi.it/pub/users/sarti/eyedetection/Zhou_2004_PR.pdf

[91] .Asteriadis S., Nikolaidis N., Hajdu A., Pitas I., An Eye Detection Algorithm Using Pixel to Edge Information, Int. Symp. on Control, Commun. and Sign. Proc, 2006

[92] Canny J. A Computational Approach to Edge Detection, IEEE Trans. Pattern Analysis and Machine Intelligence, 8(6):679–698, 1986.

[93] Kalman, R. E. (1960). "A New Approach to Linear Filtering and Prediction Problems". Journal of Basic Engineering. 82: 35.doi:10.1115/1.3662552

[94] Burton A., Radford J. (1978) Thinking in Perspective: Critical Essays in the Study of Thought Processes. Routledge.ISBN 0-416-85840-6.

[95] Yoo D.H., Chung M.J. A novel non-instrusive eye gaze estimation using cross-ratio under large head motion. Computer Vision and Image Understanding 98 (2005) 25–51

[96] Guestrin E.D., Eizenman M. “Remote point-of-gaze estimation with single-point personal calibration based on the pupil boundary and corneal reflections,” in Proc. of the IEEE Canadian Conference on Electrical and Computer Engineer-ing 2011 (CCECE 2011), Niagara Falls, ON, Canada, May 2011, pp. 971-976.

[97] Amir A., Zimet L., Sangiovanni-Vincentelli A., Kao S. An embedded system for an eye-detection sensor. CVIU, 98(1): 104-123, April 2005.

[98] Noureddin B., Lawrence P.D., Man, C.F. A non-contact device for tracking gaze in a human computer interface. Comput. Vis. Image Underst. 2005, 98, 52–82

[99] Fox E., Mathews A., Calder A.J., Yiend J. Anxiety and Sensitivity to Gaze Direction in Emotionally Expressive Faces, Emotion. 2007 Aug; 7(3):478-86.

[100] De Lemos, J. Reza Sadeghnia, G. Olafsdottir, I. Jensen, Measuring emotions using eye tracking. In: Proceedings of Measuring Behavior, vol.2008, p. 226

[101] Weidenbacher U., Layher G., Strauss P.M., Neumann H. A comprehensive head pose and gaze database. Intelligent Environments, 2007. IE07, 3rd IET International Conference on

[102] Philips P.J., Moon H., Rauss P., Rizvi S.A. The FERET evaluation methodology for face-recognition algorithms. In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 22(10), 1090

[103] Georghiades A.S., Belhumeur P.N., Kriegman D.J. From few to many: Generative models for recognition under variable pose and illumination. In Proc. 4th IEEE Int’l. Conf. on Automatic Face and Gesture Recognition, 2000.

[104] Kanade T., Cohn J., Tian Y-L. Comprehensive database for facial expression analysis. In Proc. 4th IEEE Int’l. Conf. on Automatic Face and Gesture Recognition,

2000.

[105] Sim T., Baker S., Bsat M. The CMU Pose, Illumination, and Expression Database, IEEE Trans. PAMI, 25(12), 1615 – 1618, 2003.