UNIVERSIDADE DE COIMBRA

FACULDADE DE CIÊNCIAS E TECNOLOGIA

DEPARTAMENTO DE FÍSICA

ASSISTED ANALYSIS OF OCULAR MOVEMENTS IN THE PARAMETERIZATION OF COCHLEAR IMPLANTS

ANÁLISE ASSISTIDA DE MOVIMENTOS OCULARES NA PARAMETRIZAÇÃO DE IMPLANTES COCLEARES

DISSERTAÇÃO DE MESTRADO EM ENGENHARIA BIOMÉDICA

SÓNIA MARIA GOMES DE AMARAL FERREIRA

SETEMBRO DE 2011

Para a minha família

(o meu passado, o meu presente e o meu futuro).

To my family

(my past, my present and my future).

AKNOWLEDGEMENTS

I would like to thank all members of *Grupo de Electrónica e Instrumentação* (GEI) and give a special thanks to my supervisor, Dr. João Cardoso, for his help and dedication, to my co-worker, MSc Manuel Cabeleira, for his everyday support and fellowship, to our "summer partner" João Costa, for his precious participation, and to all the volunteers in the experimental execution of this work.

I would like to thank the collaboration of Dr. Luís Filipe Silva, for his continuous availability and contagious enthusiasm. I am also grateful for the helpful collaboration of the patients and their families and for all support received from *Serviço de Otorrinolaringologia* of *Centro Hospitalar de Coimbra* (CHC).

I want to thank Professor Eduardo Sá Marta for his important contribution in the initial phase of this work.

I would like to thank *Fundação para a Ciência e a Tecnologia* (FCT) for the financial support of this project (PTDC/SAU-BEB/100866/2008).

I am also grateful for all the "coffee breaks" that I spent with my new summer friends.

I also want to thank Miguel for his precious help in correcting this work.

I have also appreciated the help, affection and patience of all my friends during the excellent moments that I have spent with them.

I also want to thank "my Bruno Ricardo" for his patience and for always giving me hope.

Finally, I want to give a huge thanks to my parents, my brothers and my sisters because they taught me how to be strong.

RESEARCH UNIT

Centro de Instrumentação - Grupo de Electrónica e Instrumentação - GEI
(http://lei.fis.uc.pt/)

COLLABORATORS

Serviço de Otorrinolaringologia – Centro Hospitalar de Coimbra – CHC
(http://www.chc.min-saude.pt/)

SUPERVISOR

João Cardoso, Ph.D. - *Centro de Instrumentação - Grupo de Electrónica e Instrumentação* - GEI

ABSTRACT

Cochlear implants are largely applied in children in cases of sensorineural deafness to restore the auditory capacity and increase phonetic and verbal performance. The gain level of the implant electrodes must be adapted after the implantation by a rehabilitation team. Eye tracking techniques can optimize this electrode's balancing by analysing ocular movements associated with sound stimuli.

The remote eye tracking system under development in this experiment is composed by three cameras and follows two principal lines: detection and tracking of the face and the eyes. Three face detection and tracking algorithms (Red-Green, Mahalanobis Distance and Viola-Jones) and two eye region detection algorithms (Between The Eyes and Eye Gabor Filter) were constructed in MALTALB® and SIMULINK® and were tested in 6 videos of 23 subjects, each one with a different type of head movement.

Viola-Jones was the best face algorithm with a detection rate of 73.91%. Eye Gabor Filter was the best eye region detection method with a detection rate of 76.81%, when it was implemented with Viola-Jones. The frame rate of Viola-Jones associated with Eye Gabor Filter was 18 frames/second in an off-line analysis.

In future, the robustness to head movements and the frame rate of the later algorithms must be improved.

Keywords:

Cochlear Implant; Electrode Balancing; Eye Detection; Face Detection; Eye Tracking; Face Tracking; Gaze Point.

RESUMO

Os implantes cocleares são muito usados em crianças em casos de surdez neurossensorial, para restaurar a capacidade auditiva e aumentar o desempenho fonético e verbal. O ganho de cada eléctrodo do implante deve ser adaptado por uma equipa de reabilitação, após a sua implantação. As técnicas de análise assistida de movimentos oculares podem optimizar este balanceamento de eléctrodos, através do estudo da relação entre os movimentos oculares e os estímulos sonoros.

O sistema remoto de seguimento ocular em desenvolvimento neste trabalho é constituído por três câmaras e segue duas linhas principais: detecção e seguimento da face e dos olhos. Foram desenvolvidos em MALTALB® e SIMULINK® três algoritmos para a detecção e seguimento da face (*Red-Green*, *Mahalanobis Distance* e *Viola-Jones*) e dois algoritmos para detectar a região dos olhos (*Between The Eyes* e *Eye Gabor Filter*). Os algoritmos foram testados em 6 vídeos de 23 voluntários, contendo cada vídeo um tipo diferente de movimento de cabeça.

O *Viola-Jones* foi considerado o melhor método para a face com uma taxa de detecção de 73.91%. O *Eye Gabor Filter* foi considerado o melhor método para detectar a região dos olhos com uma taxa de detecção de 76.81%, quando implementado com o *Viola-Jones*. A velocidade de processamento do *Viola-Jones* associado ao *Eye Gabor Filter* foi de 18 frames/segundo.

A robustez relativa aos movimentos de cabeça e a velocidade de processamento dos algoritmos anteriores devem ser melhoradas no futuro.

Palavras-Chave:

Implante Coclear; Balanceamento de Eléctrodos; Detecção do Olho; Detecção da Face; Seguimento do Olho; Seguimento da Face; Ponto de Fixação.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

In alphabetic order

BTE Between The Eyes Algorithm.

EGF Eye Gabor Filter Algorithm.

MD Mahalanobis Distance Algorithm.

RG Red-Green Algorithm.

VJ Viola-Jones Algorithm.

CHAPTER 1

INTRODUCTION

1.1. MOTIVATION

Cochlear implants are mainly used in cases of profound or total deafness to restore hearing capacity by improving phonetic and verbal performance. This technique considerably improves the quality of life of patients.

A precocious implementation, particularly in children younger than three years old, allows exposition to sound stimuli which is essential for development of speech and articulation of language.

A cochlear implant is basically formed by a set of electrodes surgically introduced inside the cochlea (a structure existent in the inner ear). These electrodes electrically stimulate the vestibulocochlear nerve and substitute the original function of the cochlea. After the surgery, the gain level of each electrode must be adapted periodically - electrode balancing. The success of the implant depends on this process; therefore, the experience of audiologists and speech therapists who perform this task is very important. However, electrode balancing is a very slow, subjective, fallible, and laborious process because it requires many audiometric tests and periodic evaluations. It also becomes more complex when it is applied to children with less than twelve years of age.

1.2. MAIN CONTRIBUTIONS

This experiment aims to use signal and image processing technologies in order to overcome the aforementioned problems. The system under development intends to use eye tracking techniques to determine the gaze point and to relate this point with the temporal features of a sound stimulus. This system will be used as a

complement of the existing rehabilitation protocol, and will allow the analysis of the hearing perception mechanism by studying fast eye movements – saccades – which appear in the presence of a sound stimulus. The eye tracking system must be a non invasive technique, must have good reproducibility and good spatial and temporal resolutions, and must allow both on-line and off-line analysis.

This experiment will create new and innovative rehabilitation and diagnosis practices in the field of cochlear implants that will provide better quality of life to future patients.

1.3. PROJECT TEAM

This project is being developed at *Grupo de Electrónica e Instrumentação* (GEI), one of the research groups of *Departamento de Física* of *Universidade de Coimbra*, and it had the collaboration of the *Serviço de Otorrinolaringologia* of *Centro Hospitalar de Coimbra* (CHC).

The team members are João Cardoso, Ph.D. (Supervisor), Manuel Cabeleira, M.Sc., and Sónia Ferreira, M.Sc. student.

1.4. CONTENTS BY CHAPTER

This work is divided in 7 sections.

This first section or Chapter 2 contains a theoretical description of the work which is essential to understand the experimental goals. It starts with an overview about the relationship between the ear and eye anatomies and ends with the analysis of eye movements. It also contains a reference to cochlear implants.

Chapter 3 consists of a brief review of the state of the art related to face detection algorithms, eye tracking techniques, and commercial eye tracking systems.

Chapter 4 presents the complete methodology used in this experiment: the acquisition system, the face and eye region detection and tracking algorithms, and the distance calculation algorithm.

Chapter 5 describes the experimental procedure and the main results.

Chapters 6 and 7 contain a discussion of the results and the main conclusions of this work, respectively.

Chapter 8 has a description of the current work and a preview of the future directions.

CHAPTER 2

THEORETICAL BACKGROUND

2.1. RELATION BETWEEN SENSE OF HEARING AND VISUAL SENSE

2.1.1. Ear Anatomy and Physiology

The ear is the organ for hearing and equilibrium. Its hearing receptors convert sound waves into nerve impulses, and its equilibrium receptors are associated with the movements of the head. The vestibulocochlear nerve is responsible for transmitting impulses from these receptors to the brain. This organ is composed by three different parts: the outer or external, the middle, and the internal ears (Van De Graaff, 2002).

The external ear consists principally of the helix, the auricle, the external acoustic canal, and the earlobe. The auricle leads the sound waves into the external acoustic canal (Van De Graaff, 2002).

The middle ear is composed by the tympanic cavity which is filled with air. This structure communicates with the outer ear by using the tympanic membrane. This cavity has three auditory ossicles: the malleus or hammer, the incus or anvil, and the stapes or stirrup. These ossicles move when the tympanic membrane is vibrating and, in this way, the wave sounds are transmitted and amplified across the tympanic cavity to the inner ear. Besides these components, the middle ear is also constituted by the auditory tube that connects the ear and the nasopharynx, allowing the air pressure to equalize on the tympanic membrane. The vestibular or oval and cochlear or round windows make the connection between the middle and the inner ears (Van De Graaff, 2002).

The inner ear is comparable to a labyrinth with an external bony part that surrounds an internal membranous region. The fluid existent between the bony

and the membranous regions is perilymph, and the fluid inside the internal part is called endolymph. These fluids permit the conduction of hearing vibrations and the conservation of equilibrium (Van De Graaff, 2002).

The bony region is formed by the vestibule, the three semicircular canals, and the cochlea. The vestibule contains two membranous structures, the utricle and the sacule, which have receptors sensitive to gravity and linear movement of the head. The semicircular canals have membranous ducts with receptors sensitive to angular movements of the head (Van De Graaff, 2002).

The cochlea has a snail shape and is composed by three chambers: the scala vestibuli, the scala tympani, and the cochlear duct or scala media. The scala media is constituted by the basilar and the vestibular membranes and the spiral organ or organ of Corti, which have sound receptors that transform mechanical vibrations into nerve impulses. The scalas tympani and vestibuli are filled with perilymph and the scala media has endolymph (Van De Graaff, 2002).

Figure 1 illustrates the main components of the human ear.

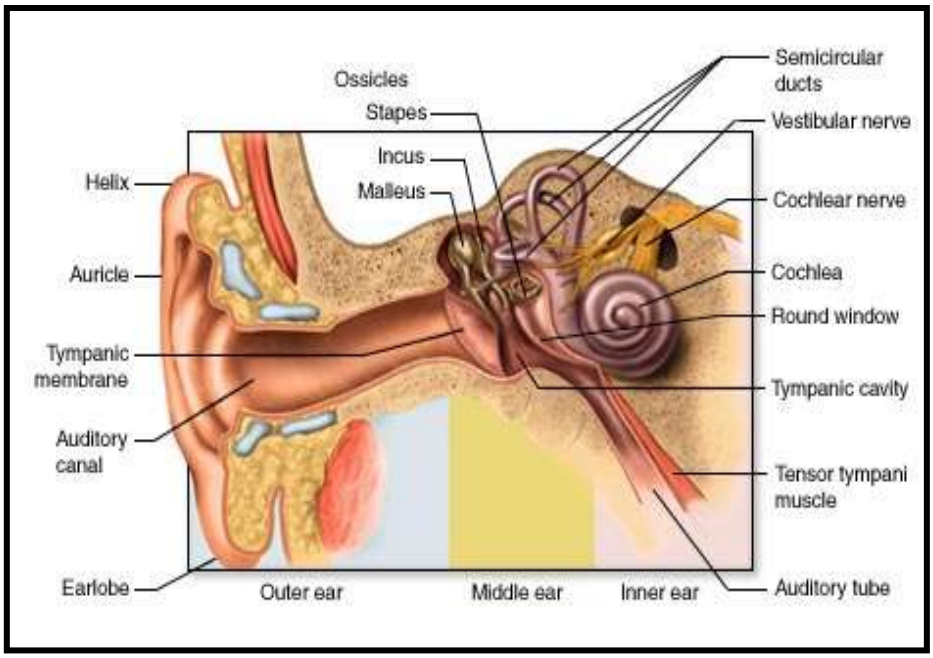

Figure 1 - Principal components of the human ear (The McGraw-Hill Companies, 2011).

2.1.2. Hearing Mechanism

The sound is propagated from its source by waves based on pressure variations in a given environment. The pressure increases and decreases alternately relative to a reference pressure. The energy of the sound wave is proportional to these variations (Pocock & Richards, 2006).

A sound is classified by its intensity (in dB), frequency (in Hz) and tone. The intensity is related to the amplitude of the sound waves. The frequency is associated with the number of waves per second. The tone or pitch is proportional to the frequency. According to these characteristics, the human auditory system is able to determine the sound origin and to analyse its information (Pocock & Richards, 2006; Van De Graaff, 2002).

The sound intensity is usually measured using the decibel scale (dB)

$$dB = 20 \times \frac{Sound\ Pressure}{Reference\ Pressure} \quad (1)$$

where the *Reference Pressure* is 20 µPa (Pocock & Richards, 2006).

An audiogram is a graphic representation of the relation between sound intensity and frequency in order to analyse the hearing threshold of a person (see Figure 2). A healthy person has a frequency range of audible sounds between 20 Hz and 20 kHz and an intensity range between 0 dB and 140 dB. The human ear can distinguish differences in sound frequency of 0.3% and differences in sound intensity of 0.1 dB (Pocock & Richards, 2006; Van De Graaff, 2002).

Sound waves are propagated through the external acoustic canal and create vibrations in the tympanic membrane. This membrane is connected to the auditory ossicles which amplify approximately 20 times the sound waves. Sounds waves pass through the vestibular window and cause displacements in the basilar membrane inside the cochlea. These displacements result on the release of a

neurotransmitter which excites the cochlear sensory neurons in the vestibulocochlear nerve. This nerve synapses with neurons in the medulla oblongata, which is connected to the midbrain. The midbrain projects neurons to the thalamus, which makes the connection with the auditory cortex in the temporal lobe of the brain. Here, these auditory impulses are recognized as sounds (Van De Graaff, 2002).

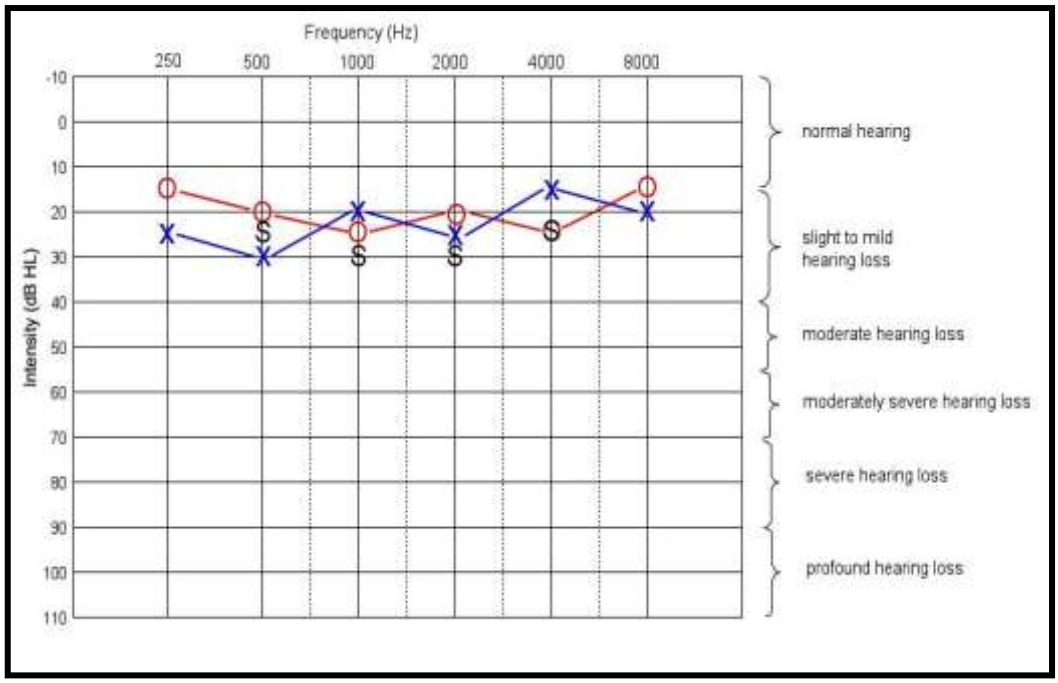

Figure 2 – Example of an audiogram. The red curve is the response of the right ear and the blue line is the response of the left ear. The different hearing levels are discriminated in the right side of the figure (BC Children's Hospital, 2010).

2.1.3. Hearing Loss

The hearing process can be disturbed leading to hearing loss. Hearing loss can result from disease, trauma or problems associated with the auditory system (cochlear nerve and the auditory pathway) or the auditory perception areas in the brain (Pocock & Richards, 2006; Van De Graaff, 2002).

Surgery or hearing aids are used in some cases of hearing loss (Van De Graaff, 2002).

The main hearing deficits are conductive deafness, sensorineural or perceptive deafness and central deafness (Pocock & Richards, 2006; Van De Graaff, 2002).

Conductive deafness results from defects of the outer or the middle ear which interfere with sound waves. This disease can be caused by wax accumulations, ruptures of the tympanic membrane, middle ear infections, otitis and otosclerosis. Medical treatment can decrease the effects of conductive deafness (Pocock & Richards, 2006; Van De Graaff, 2002).

Sensorineural deafness occurs due to disorders in the inner ear, damages of the cochlea or the vestibulocochlear nerve, or in the auditory areas of the brain. This disease can be caused by high intensity sounds or trauma, age, genetic and developmental problems, and some chemical substances like antibiotics and diuretics. This type of deafness is permanent, but the use of hearing aids can help some patients (Pocock & Richards, 2006; Van De Graaff, 2002).

Central deafness is caused by injuries in the auditory pathways. These injuries can provoke difficulty in locating sound and in recognizing speech and sound features (Pocock & Richards, 2006).

2.1.4. Cochlear Implants

Cochlear implants are electronic devices which have been applied to the treatment of severe or profound bilateral sensorineural hearing loss since the 80s. Many studies have been performed to improve these systems: the position of electrodes, the electrical stimulation, the speech processing strategies and the size of inner and outer devices (Porter et al., 2003; ASHA, 2004).

2.1.4.1. Constitution of Cochlear Implants

All cochlear implants have common characteristics. They are normally constituted by a microphone, an external speech processor, signal transfer hardware, transmitters, receptors, and electrodes. The main function of all these components is to convert sound waves into electrical stimuli (Porter et al., 2003; ASHA, 2004).

The microphone only receives and transforms electrical signals on continuous or analogical mode (Porter et al., 2003; ASHA, 2004).

The external speech processor and the signal transfer hardware shape the form of the electrical signal by using a sequence composed by amplification, compression and filtering. Amplification is important to increase the amplitude of the captured signals. Compression is an essential step on the modulation of signals, to create an intensity range useful for the patient (Porter et al., 2003; ASHA, 2004).

The transmitter or external coil is surgically placed on the mastoid, a part of the temporal bone, normally with help of magnets. It sends the processed signal to the receptor via radio waves. This component transforms the received signal into an electrical signal which is directed to the electrode array localized inside the cochlea. An electrical field is generated and causes a discharge in the auditory nerve, leading the stimulus to the central nervous system to be interpreted (Porter et al., 2003; ASHA, 2004).

The majority of cochlear implants has a technological characteristic denominated by telemetry which is used to analyse the integrity of the intracochlear electrodes, after the implantation. This characteristic is very important because patients with a few hearing experience or children can not realize the presence of abnormalities (Porter et al., 2003; ASHA, 2004).

Figure 3 is a schematic representation of a cochlear implant with its main components.

The main differences between the models of cochlear implants on the market are the techniques of sound processing and the way how the signal arrives and is transmitted to the auditory nerve (Porter et al., 2003; ASHA, 2004).

Most cochlear implants offer a range of different methods to process speech. Programming of the speech processor requires the establishment of a threshold and a stimulation maximum for each intracochlear electrode. These limits vary from patient to patient and need to be adjusted during the use of the cochlear implant. This adjustment occurs in the outside of the processor – electrode balancing or electrode mapping or electrode programming. During the electrode balancing, the electrodes are stimulated to determine the hearing range of each patient – the threshold and discomfort levels (Porter et al., 2003; ASHA, 2004; Zwolan, 2011).

There are several methods to process sound in the cochlea based on different algorithms: SPEAK-Spectral Peak, CIS-Continuous Interleaved Sampling, ACE-Advanced Combined Encoder, SAS-Simultaneous Analog Strategy and PPS-Paired Pasqueflower Sampler (Porter et al., 2003; ASHA, 2004).

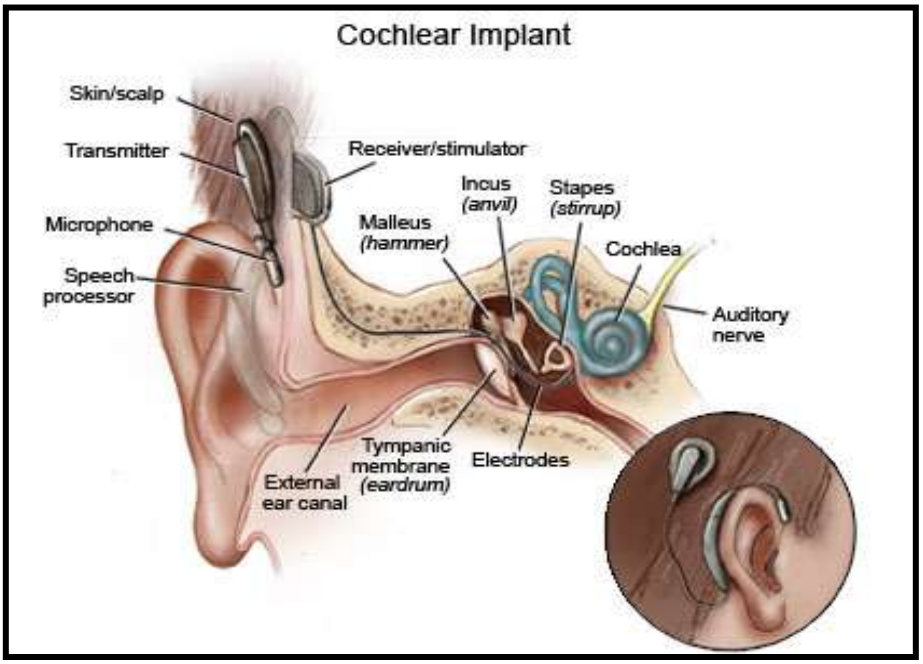

Figure 3 - Schematic representation of a cochlear implant and its main components (KidsHealth, 2011).

2.1.4.2. Advantages of Cochlear Implants

The rehabilitation of the patient depend on pre-surgery tests and communication and listening skills training (Porter et al., 2003; ASHA, 2004).

The main benefits of the use of cochlear implants are more significant in children than in adults, because the lower the period of deprivation of hearing the greater the discrimination of speech by the patients. This happens because the auditory system retains some plasticity (Porter et al., 2003; ASHA, 2004).

The cochlear implants are advantageous in improving the patient's perception of speech and these benefits remain for a long period of time (13 to 15 years), even in cases that require the reintroduction of the cochlear implant. This hearing technology enhances the patient's quality of life and is not very expensive (Porter et al., 2003; ASHA, 2004).

2.1.5. Eye Anatomy and Physiology

The surface of the eye is essentially composed by the sclera, the iris, and the pupil. The sclera is the white posterior outer layer of the eye which supports and protects the eyeball. The iris is a circular structure composed by pigment cells and muscle fibres, and is responsible for the regulation of the pupil diameter, to control the amount of light entering inside the eyeball. The pupil is a circular dark aperture located inside the iris (Van De Graaff, 2002; Pocock & Richards, 2006).

The eye is an organ complemented with protective structures and structures that allow eye movement (Van De Graaff, 2002).

The bony orbit, the eyelid, the eyebrow, the eyelashes, the lacrimal apparatus, and the conjunctiva are protective structures of the eye. The bony orbit is composed by bones that support and protect the eye. Eyelids protect the eyes from desiccation and their glands secret an oily substance to avoid these structures to become stick together. The eyebrows protect the eyes from the sun, falling

perspiration, and particles. The eyelashes protect the eye from air falling particles. The lacrimal glands secrete tears to lubricate the eye surface and the eyelids. The conjunctiva is a membrane connected with eyelids which avoids the penetration of foreign particles in the eye. (Van De Graaff, 2002).

Figure 4 represents some of the main components of the eye surface that will be referred later in this work.

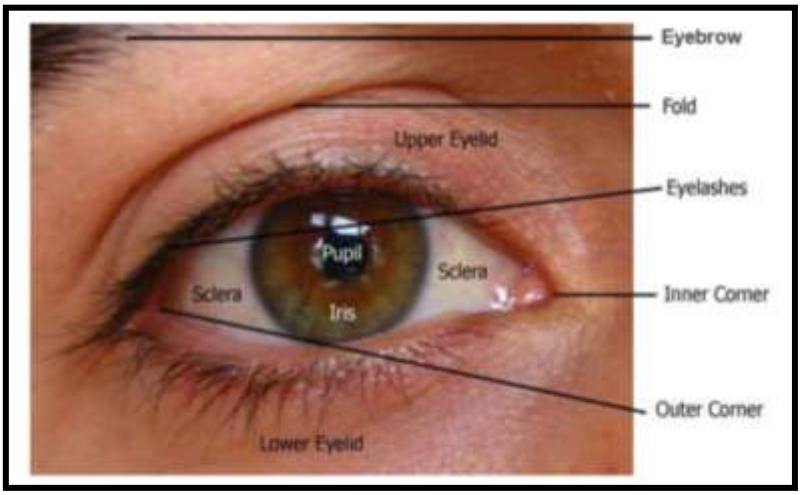

Figure 4 - Components of the eye surface that will be mentioned in this work. Image adapted from (Uzunova, 2005).

The extrinsic muscles of the eyes are skeletal muscles that help eye movements. Each muscle causes a precise eye movement, but usually an eye movement is a combination of the action of different muscles. The extrinsic muscles are composed by four rectus muscles - superior, inferior, medial and lateral - and two oblique muscles – superior and inferior (see Figure 5). The lateral and medial rectus muscles control lateral movements, the superior and inferior oblique muscles control diagonal movements and the superior and inferior rectus muscles control up-down movements (Pocock & Richards, 2006; Van De Graaff, 2002).

These muscles are innervated by three cranial nerves: oculomotor, abducens and trochlear (Van De Graaff, 2002). Figure 6 shows typical directions of eye movements due to the combined action of extrinsic muscles.

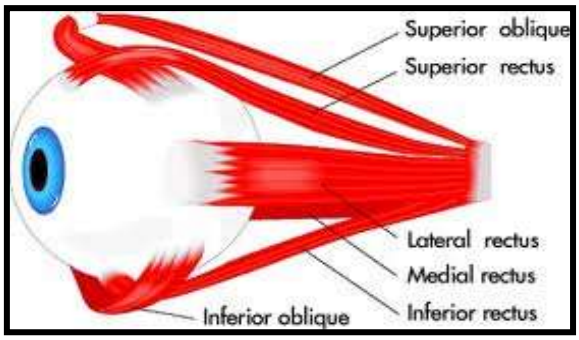

Figure 5 - Schematic representation of the extrinsic muscles of the eye (Defeat Diabetes Foundation, 2011).

Figure 6 – Typical directions of eye movements due to the combined action of the extrinsic muscles. Image adapted from (Bejjani et al., 2002).

2.1.6. Eye Movements

Many studies have been done to correlate the presence of audio stimuli with eye movements (Eberhard et al., 1995; Yee & Sedivy, 2006; Liversedge & Findlay, 2000; Harris et al., 1980; Frens et al., 1995; Zambarbieri et al., 1995; Knudsen et al., 1993).

Sound stimuli applied during the audiometric tests necessary to the electrode's balancing of the cochlear implant can be associated with the eye movements of the patient. The superior and inferior colliculi in the midbrain are the areas responsible to perform this association. The auditory information is conducted to

the inferior colliculus and after to the optic tectum or superior colliculus, in the midbrain. The optic tectum is responsible to map visual and auditory information which is used to execute ocular movements (Knudsen et al., 1993; Harris et al., 1980).

These stimuli work as an object of interest which appears in the visual field, attracting the patient's attention.

2.1.6.1. Saccade

A saccade is a sudden and rapid movement of the eye to fixate an object of interest which appears in the peripheral field of view. These movements occur when the eye appears to be still. A saccade is ballistic, so when this movement starts it is impossible to change its final destination. During a saccade, the visual system is suppressed but is not entirely shut off, so it does not take or encode visual information from objects. The visual processing is annulled in order to avoid a blurred visual image. Saccadic movements are voluntary and can involve head motion (Young & Sheena, 1975; Pocock & Richards, 2006; Van De Graaff, 2002; Jacob, 1994; Poole & Ball, 2005).

This type of movement is performed with an angular velocity between 200 and 600 deg/sec, has duration of 20 to 120 msec, covers 1 to 40 deg of the visual angle, and has initial and final accelerations up to 40.000 deg/sec^2 (Pocock & Richards, 2006; Van De Graaff, 2002; Jacob, 1994; Young & Sheena, 1975).

When the object of interest appears, there is a delay of 100 to 300 msec before the occurrence of the saccade. After the saccade, there is a refractory period of 100 to 200 msec (Jacob, 1994).

2.1.6.2. Smooth Motion/Pursuit/Slow-Tracking

A smooth pursuit movement occurs when the eye follows an object of interest. The maximum angular velocity of this type of movement is 50 deg/sec. This movement does not happen in the presence of a static scene because it requires a moving stimulus. This type of movement is not under voluntary control and is limited in acceleration and velocity (Young & Sheena, 1975; Pocock & Richards, 2006; Van De Graaff, 2002; Jacob, 1994).

2.1.6.3. Fixation

A fixation is a period of stability which normally occurs after a saccade, when the information about the object of interest is acquired by the visual system. This movement has duration between 200 and 600 msec. The frequency of fixations gives information about the complexity or salience of the target objects. During a fixation, the eye suffers small high-frequency tremor movements called microsaccades or flicks, and drifts (Poole & Ball, 2005; Jacob, 1994; Young & Sheena, 1975).

2.1.6.4. Nystagmus

An optokinetic or train nystagmus is a sequence of pursuit movements intercalated with saccades and fixations which occurs when viewing a moving pattern. A nystagmus can also be performed as a result of the motion of the head. The minimum time between each sequence cycle is about 0.2 sec, so the maximum frequency is approximately 5 Hz. The movement amplitude varies between 1 to 10 deg (Pocock & Richards, 2006; Van De Graaff, 2002; Jacob, 1994; Young & Sheena, 1975).

A vestibular nystagmus is an oscillatory eye movement which results from the stimulation of the semicircular canals during the rotation of the head. This movement has similar characteristics to the optokinetic nystagmus (Young & Sheena, 1975).

2.1.6.5. Vergence

Vergence eye movements are made to converge on an object, when the object of interest is distant or near, and are characterized by the two eyes moving in opposite directions. These movements are slower and smoother than other eye movements and are non-predictive. Vergence movements have velocities on the order of 10 deg/sec and cover a range of 15 deg (Jacob, 1994; Young & Sheena, 1975).

2.1.6.6. Torsional Rotation

Torsional rotation of the eye occurs around the horizontal axis of the eye and it is not detected by the majority of the eye trackers systems. These movements are limited to angles of less than 10 deg and result from nystagmus and compensatory movements (Young & Sheena, 1975; Jacob, 1994).

2.1.6.7. Compensatory

Compensatory eye movements are smooth movements related to pursuit and nystagmus movements. These movements compensate the motion of the head and the trunk to stabilize the acquired image (Young & Sheena, 1975).

Figure 7 shows a representation of some of the aforementioned eye movements.

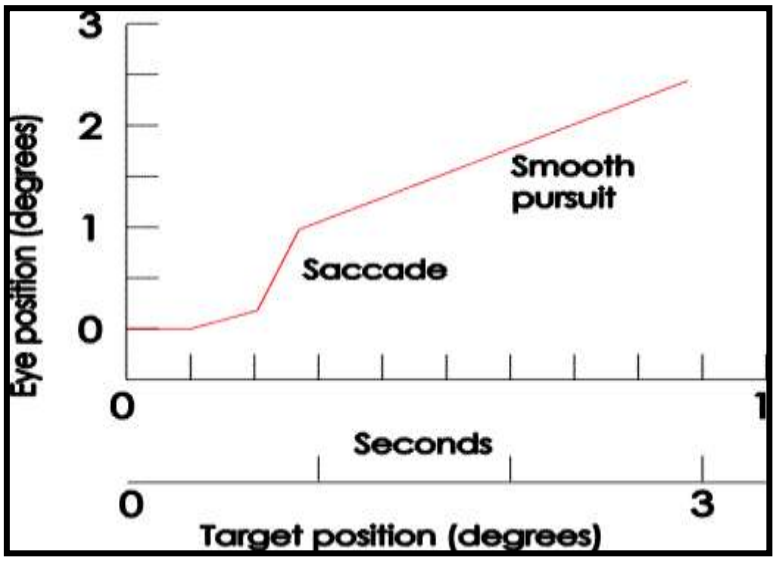

Figure 7 – Graphic representation of a saccade and a smooth pursuit movement (Laurens, 2002).

CHAPTER 3

STATE OF THE ART

3.1. EYE TRACKING TECHNOLOGY

An eye tracking system is a device constructed to measure an individual's eye movements in order to determine the gaze point or point-of-regard at any given time, and to calculate the displacement of this point over a period of time. By studying eye movements with an eye tracking system, inferences can be made about the objects or targets which "deserve" a person's attention (Poole & Ball, 2005; Jacob & Karn, 2003).

Over the years, eye tracking systems were used to study the relationship among eye movements, visual or other stimulus, and cognitive processes like learning, memory, workload, and attention. These systems are also used to interpret person's relationships, desires, needs, and emotional states (Hansen & Ji, 2010; Poole & Ball, 2005; Jacob & Karn, 2003).

The eye is the most prominent feature of the human face due to its distinctive geometry, photometry, and motion. It is characterized by the shape and the intensity of the pupil, the iris, and the cornea (Hansen & Ji, 2010).

The eye appearance is influenced by several factors: ethnicity, viewing angle, head pose, colour, texture, light conditions, positions of the iris within the eyeball, and degree of eye occlusion (Hansen & Ji, 2010). These factors complicate the eye detection.

An eye tracking system can be remote or can be head-mounted (helmets and eyeglasses). A remote eye tracking system consists on the following steps: detection of the existence of the eyes, interpretation of eye positions, tracking of the eye in video frames, estimation of head pose, and gaze estimation in 3-D (Hansen & Ji, 2010). A head-mounted eye tracking only performs the detection of the eye positions and the gaze estimation.

Eye trackers can be illuminated by artificial infra-red light source(s) – active eye trackers - or by natural light source(s) – passive eye trackers (Hansen & Ji, 2010).

Usually, the features of the eye used to calculate the gaze point are corneal reflexions named Purkinge images, the boundary between the sclera and the iris, and the pupil shape (Poole & Ball, 2005).

In order to quantify the eye movement, an eye tracking system can use different metrics. The most common metrics presented in literature are fixations, saccades, pupil size, and blink rate. Gaze and scanpath metrics, derived from fixations and saccades, are also used. A gaze or dwell or fixation cluster or fixation cycle is the sum of all duration of fixations within a certain area. A scanpath is a complete sequence saccade-fixation-saccade. Saccades have been broadly used to the detection of fatigue, vision studies, diagnostic of neurological disorders, and sleep analysis. Fixations are normally studied in vision and neuroscience and psychological areas to determine the individual's focus and attention (Hansen & Ji, 2010; Poole & Ball, 2005).

3.1.1. Eye Tracking Applications

According to Duchowski (2002), eye tracking systems can be divided in two different categories by application: diagnostic eye tracking systems and interactive eye tracking system. The first category aims to analyse the processes of vision and attention of the user in the presence of a stimulus, and is normally used to study recorded eye movements. It performs an off-line data analysis. The second type of eye tracking systems is designed to "respond to or interact with the user based on the observed eye movements"; therefore, it is suitable to real-time analysis.

The interactive eye tracking systems can be classified as selective, if they use the gaze point as a pointing or selecting device such as the computer mouse, or gaze-contingent, if they utilize the user's gaze information to alter the complexity of the display. The gaze-contingent systems can also be divided in two subtypes,

according to the display processing: screen-based systems and model-based systems. Screen-based systems are related to image manipulation and model-based systems are associated with the manipulation of graphical objects or models (Duchowski, 2002).

Eye tracking methods have been widely used in different areas in the past years. In neuroscience: in the study of attention and brain activity. In psychology: during the performance of tasks like reading, scene perception (art and films), visual search, natural tasks, auditory language processing, and other information processing tasks (problem solving, dual task situations, face perception, brain damage, sports, and human-computer interaction). In industrial engineering: in aviation, driving tasks, and visual inspection. In marketing and advertising: in copy testing, print advertising, and design of websites. In medicine: in the analysis of the performance of doctors in medical procedures (Duchowski, 2002; Poole & Ball, 2005).

3.1.2. Eye Tracking Models

According to Duchowski (2002), the eye tracking research can be separated in four distinct eras.

The first era, between 1879 and 1920, was characterized by the use of simple observation methods and invasive methods (electro-oculography and large contact lenses), and by the study of saccadic eye movements. Later, many non-invasive corneal reflection methods and motion picture techniques were combined in various ways (Duchowski, 2002; Poole & Ball, 2005; Jacob & Karn, 2003; Young & Sheena, 1975).

The second era, between 1930 and 1958, was marked by the development of head-mounted eye trackers and the application of eye tracking to engineering. This period coincided with the behaviourist movement in experimental psychology (Duchowski, 2002; Jacob & Karn, 2003).

In the third era, between 1970 and 1998, several advances in eye tracking technology (accuracy, precision and invasiveness) had been made by engineers, and had been applied in psychology and physiology to connect eye movements and cognitive processes. The first remote and real-time eye trackers were created in this time due to advances in the computer science (Duchowski, 2002; Jacob & Karn, 2003).

In the fourth era, the twenty-first century, a large variety of real-time eye trackers for diagnostic and interactive applications is being developed. These developments are associated with the proliferation of computers and the creation of new communication technologies (Duchowski, 2002; Poole & Ball, 2005; Jacob & Karn, 2003).

Due to the wide variety of eye tracking systems found in literature, a summary of these systems, their descriptions, advantages, and disadvantages is presented on Table I. Some techniques commonly associated with eye tracking are: Integral and Variance Projections, Hough Transforms, Active Appearance Models, Gabor and Kalman Filters, Hidden Markov Models, Modular Eigenspaces, Integrodifferential Operators, Neural Networks, and Morphological Operations (Hansen & Ji, 2010).

Eye tracking systems can have different gaze estimation methods: feature-based, appearance-based, natural light and dual Purkinje methods (Hansen & Ji, 2010).

The gaze estimation methods may require a camera calibration, a geometric calibration of the setup units (monitor, camera(s) and light source(s)), a personal calibration and a gaze mapping calibration (Hansen & Ji, 2010).

3.1.3. Eye Tracking Systems on the Market

Eye trackers on the market usually have high-quality cameras with variable lenses (1-2 Megapixels of resolution and 50-60 Hz of frame rate) which can have pan and tilt heads to increase the accuracy and the robustness to head movements. However, these characteristics imply much higher costs. In literature, low-quality cameras or even webcams are normally used (Hansen & Ji, 2010; Kumar, 2006).

Eye tracking systems on the market have prices between $5.000 and $40.000. The final price of an eye tracker includes hardware costs, software implementation costs and business costs (manufacturing, marketing, sales, and support). Over the last years, the commercial value of high-quality digital cameras has suffered a pronounced decrease (estimated to be in the range of $1.000-$2.000). However, the use of specialized digital processors to improve the speed of eye trackers increases the final price. The software of commercial devices includes calibration processes, APIs (Application Programming Interfaces) and SDKs (Software Development Kits) (Kumar, 2006; Winfield et al., 2005).

Specialized eye trackers can only be used by experts. Due to this limited market, vendors "are forced to charge high prices" (Kumar, 2006).

Some examples of eye tracking devices are commercialized by *Tobii*, *Eyetracking Inc.* (ETI), *Veritest*, *Eyetools Inc.*, *Applied Science Laboratories* (ASL), *Testusability*, *Erica Inc.*, *LC Technologies Inc.*, and *Fourward Technologies Inc*. (Namahn, 2001).

Nowadays, eye trackers still have disadvantages that need to be overcome: intrusiveness and obtrusiveness, the lack of robustness to face pose, light variations and the use of eyeglasses or contact lenses, the low availability, the necessity of calibration, and the high price. Another problem of eye tracking systems is to find an appropriate way to use and interpret the data (Jacob & Karn, 2003; Winfield et al., 2005; Hansen & Ji, 2010).

The remote video-based eye tracking systems allow more freedom of head movements and a good accuracy (0.5 degrees or better) (Hansen & Ji, 2010).

Table I – Type, description, advantages, and disadvantages of eye tracker models commonly used in literature (Hansen & Ji, 2010).

MODEL	DESCRIPTION	TYPE	ADVANTAGES	DISADVANTAGES	LITERATURE EXAMPLES
SHAPE-BASED	Eye or face features and their contours are used as a prior model; A similarity measure is used; Common Features: Iris, limbus, edges, eye corners, and points from filter responses.	**Fixed Shape or Feature-Based**	Robustness to illumination changes.	Difficulty to apply outdoor; Weakness to changes in image focus, eye occlusion and face rotation.	Feng & Yuen, 1998; Sirohey et al., 2002.
		Deformable Shape	Accuracy; Genericalness; Robustness to shape, scale, and illumination variability; Efficiency.	Computationally demanding; Requirement of high quality images and initialization; Weakness to face pose, eye occlusion, variations of eye features, and under IR light.	Yuille et al., 1992; Lam & Yan, 1996.
APPEARANCE-BASED OR TEMPLATE OR HOLISTIC	Directly eye detection and tracking based on photometric appearance – colour distribution or filter responses of eye and its surroundings.	**Intensity-based**	Simplicity; Efficiency; Robustness to illumination changes; Fast analysis.	Coarse-scale eye location; Necessity of further processing; Lack of size invariance and specific eye parameters; Weakness to head rotation and interpersonal variations.	Hallinan, 1991; Huang et al., 1998.
		Subspace-based	Efficiency; Accuracy; Dimensionality reduction; Robustness to illumination changes.	Lack of size invariance and specific eye parameters; Weakness to head rotation and interpersonal variations.	Pentland et al., 1994; Huang & Mariani, 2000.
		Filter Responses	Simplicity; Intuitiveness; Robustness to illumination changes.	Increasing efficiency implies decreasing runtime performance; Weakness to face rotation.	Witzner & Hansen, 2006; Huang & Wechsler, 1999.
HYBRID	Combine feature, shape and appearance	**Shape & Intensity**	Exploration of benefits of the combined	Influence of sidelights; High computational demand; Difficulty in	Ishikawa et al., 2004; Xie et al.,

MODEL	DESCRIPTION	TYPE	ADVANTAGES	DISADVANTAGES	LITERATURE EXAMPLES
	approaches to explore their respective benefits.		methods.	modelling large eye appearances.	1994.
		Colours & Shape		Influence of skin-coloured objects in the background; Requirement of prior eye location.	Witzner et al., 2003.
OTHERS		**Active IR Illumination**	Illumination does not distract subjects or cause pupil contraction; Robustness to global light changes; Simplicity; Efficiency.	Influence of variance between subjects, ethnicity, head position, background objects and reflexions on eyeglasses; Weakness in outdoor or bright environments; Influenced by eye closure, distance of the subject to the camera and eye intrinsic properties.	Ji & Yang, 2002; Zhu et al., 2002.
		Symmetry operators		Necessity of a threshold to perform feature selection; Time complexity.	Reisfeld & Yeshurun, 1992; Loy & Zelinsk, 2003.
		Blinks and Motion	Simplicity.	Weakness to fast blinking and head movements.	Grauman et al., 2001; Kawato & Testutani, 2002.

3.2. FACE DETECTION AND TRACKING

Face detection is widely used to recognition and identification of faces, model-based coding of video sequences, intelligent man-machine interfaces, security access control systems, surveillance systems, and forensic investigation (Ahlberg, 1999; Yen & Nithianandan, 2002; Shih & Chuang, 2004).

The face changes between individuals due to ethnicity, skin colour tones, facial expressions, lighting and head pose. Facial accessories like eyeglasses, hats, beard, makeup, and hair can also contribute to differences between individuals (Ahlberg, 1999; Chellappa et al., 2010).

Face detection methods can be grouped in three different categories: attribute-based or feature-based methods, template-based or example-based methods, and appearance-based methods. Table II summarizes the description, advantages, and disadvantages of these methods. These methods fail when large head movements are performed (Zhao et al., 2003; Ahlberg, 1999; Yen & Nithianandan, 2002).

Some mathematical methods associated with face detection are Principal Component Analysis, Eigenfaces, Clustering, Snakes, Neural Networks, and Active Appearance Models (Zhao et al., 2003; Shih & Chuang, 2004).

Features usually extracted from facial detection are local (lines or fiducial points) or facial features (eyes, nose, and mouth) (Zhao et al., 2003; Chellappa et al., 2010).

Nowadays, face detections techniques are still influenced by head pose and illumination effects (Zhao et al., 2003).

Table II - Type, description, advantages and disadvantages of face detection methods commonly used in literature (Ahlberg, 1999; Yen & Nithianandan, 2002; Shih & Chuang, 2004; Zhao et al., 2003).

TYPE	DESCRIPTION	ADVANTAGES	DISADVANTAGES	LITERATURE EXAMPLES
FEATURE/ATTRIBUTE-BASED	Features extracted based on geometric relationships (position and width) are search on the face image; Common Features: elliptic shape, specific colours and facial features.	Fast; Robust to image intensity, shape variations, and face rotation.	Weakness to small, multiple, and hair/beard occluded faces; Difficult to implement.	Jeng et al., 1998.
TEMPLATE/EXAMPLE-BASED	Use of single or multiple templates for features extraction based on minimum values of energy functions.	Robust to features size and shape; Simple to apply; Robust to head tilt and head scale.	Computationally demanding; Insufficient matching capability (eyeglasses and beard).	Ryu & Oh, 2001.
APPEARANCE-BASED	Extraction of facial features based on edges, lines, and curves.	Robust.	Weakness to changes in appearance – closed eyes, eyeglasses, and open mouth.	Kass et al., 1988.

CHAPTER 4

METHODOLOGY

A remote eye tracking system should follow two lines of development: one related to the movement and the position of the head/face and the other related to the detection and tracking of the eyes. These lines depend on each other because the detection of eye positions requires prior knowledge of the head position.

4.1. AQUISITION SYSTEM

4.1.1. Remote Eye Tracker

A remote eye tracking system is ideal to meet the goals of this experiment.

This system will be used with children; therefore, it is very important to construct a non-invasive method. A non-invasive method is an eye tracking technique that can be applied without interfering with hospital audiometric tests or with children's movements. The eye tracker should allow the children free movement.

In order to achieve these aforementioned goals, an eye tracker system was developed with three video cameras. The two lateral web cameras have a parallel optical axis and the central camera is situated in the middle of the two lateral cameras.

The lateral cameras are used for face detection and for face tracking, as well as eye region segmentation. These cameras are also used to compute the distance between the subject and the acquisition system. The latter is important in order to determine the gaze point in 3-D.

The central camera uses the information from the lateral cameras to acquire eye videos with a better resolution to perform the gaze point calculation. These lateral

cameras capture videos with a lower resolution (120×160 pixels) than the central camera (576×720 pixels) to decrease the processing time. The central camera will acquire eye videos with a high resolution instead of acquiring images of the entire face (in the final real-time application).

Figure 8 illustrates the disposition of the three cameras in the remote eye tracker system.

Figure 8 – Remote eye tracker system used in this experiment. This system is composed by two lateral web cameras with a low resolution and by a central camera with a high resolution.

Cameras are fixed in a metallic support to avoid undesirable movements during measurements.

Lateral cameras are connected to the computer with a USB port and the central camera is connected with a FireWire port. Videos were acquired in the RGB colour space format.

The acquired videos are processed by using a set of algorithms developed in the software MATLAB®7.11.0 and SIMULINK®7.11.0, version R2010b. These algorithms will be described in the next sections of this chapter.

Figure 9 is a schematic representation of the final eye tracking system chosen for this experiment.

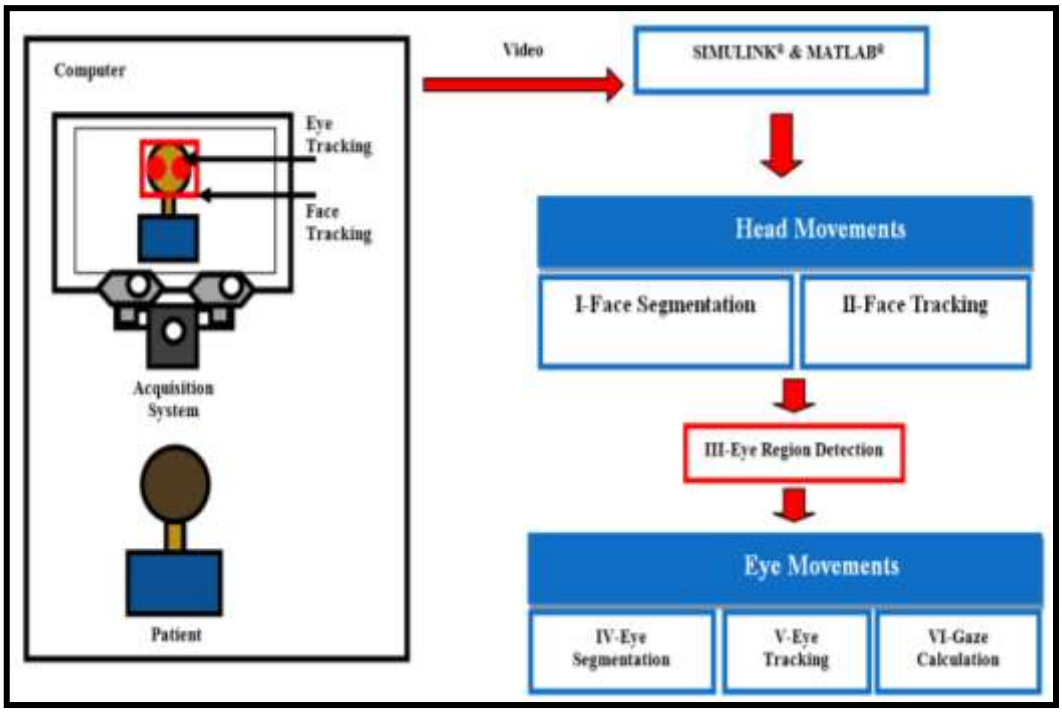

Figure 9 - Schematic representation of the final eye tracking system chosen for this experiment.

4.2. FACE DETECTION AND TRACKING ALGORITHMS

For face analysis, three algorithms were used in this experiment: Red-Green (RG), Mahalanobis Distance (MD) and Viola-Jones (VJ). The first two algorithms are based on skin colour properties and the last algorithm is focused on rectangular face features.

These techniques will be detailed in the next sections.

In the next sections, the word image will be used with the same meaning of video frame.

4.2.1. Colour Based Algorithms

Colour based algorithms are applied according to the characteristics of the colour space of the acquired image. Here, two different colour spaces were used: RGB colour space and YC_BC_R colour space.

In RGB colour space, each image pixel is a combination of three components: the Red component (R), the Green component (G) and the Blue component (B). All these components have values between 0 and 255. If all the three components have a value of 0, the final pixel colour will be black and if all the components have a value of 255, the final pixel colour will be white. In this colour space, each pixel is represented by a point with three coordinates, each one corresponding to the colours red, green and blue (MathWorks®, 2011). Figure 10(a) represents the RGB colour space.

YC_BC_R colour space is formed by three distinct components: component Y which stores the luminance information, component C_B which represents the difference between the Blue component and a reference value, and component C_R which is the difference between the Red component and a reference value. C_B and C_R are the chrominance components. Component Y has values between 16 and 235 and components C_B and C_R have values between 16 and 240 (MathWorks®, 2011). Figure 10(b) shows a representation of the YC_BC_R colour space.

The conversion from RGB colour space to YC_BC_R colour space is (MathWorks®, 2011)

$$\begin{bmatrix} Y \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (2)$$

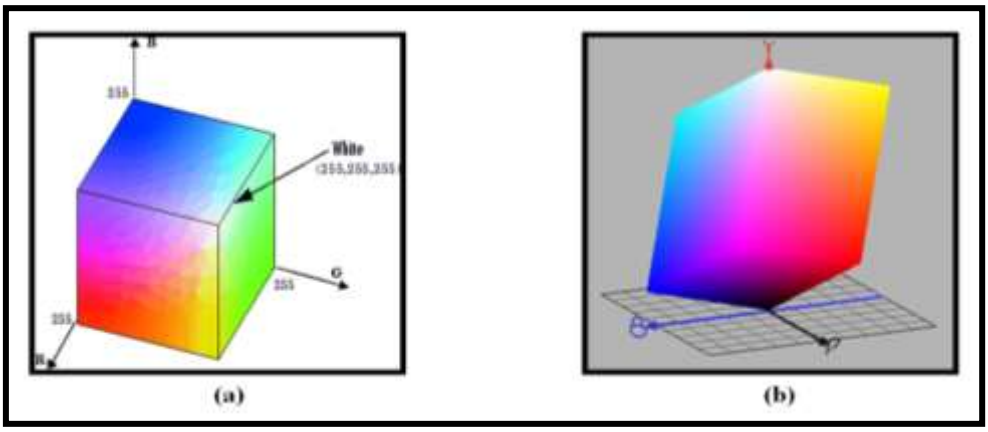

Figure 10 – Colour spaces used in this experiment: (a) Schematic representation of the RGB colour space (MathWorks®, 2011); **(b)** Schematic representation of the YC_BC_R colour space (Briggs, 2007).

4.2.1.1. RG: Red – Green

RG is an algorithm developed by Al-Shehri (2004) and it is a simple and fast method to detect faces. This method relies on human skin colour properties and requires images with sufficient quality and resolution.

Considering that the images are acquired in the RGB format, skin colour pixels are classified using two main principles: the component Red is the predominant colour in human skin; the ratio between the components Red and Green is greater than 1; therefore, the difference between these components is positive in an image containing human skin (assuming that pixels in a greyscale image have values between 0 and 255 or between 0 and 1) (Al-Shehri, 2004)

$$\frac{R}{G} > 1 \iff R > G \iff R - G > 0. \quad (3)$$

When component G is subtracted from component R, the non-skin pixels acquire low values and skin pixels have high values. Face pixels are isolated by choosing values between two given thresholds: 0 and 0.06 (Al-Shehri, 2004). The result of the RG subtraction is exemplified in Figure 11.

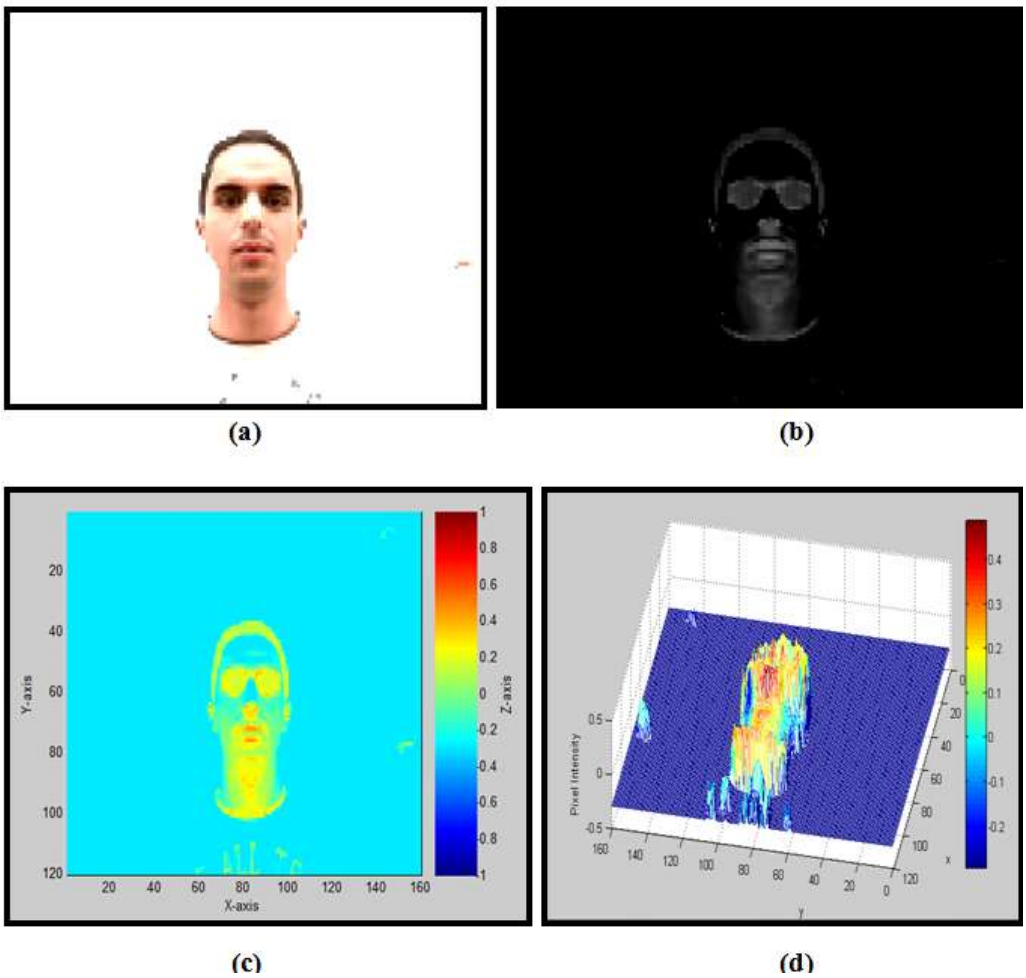

Figure 11 - Example of an image obtained with the RG method. (a) Original image; **(b)** Image after RG subtraction; **(c)** RG image displayed on a colour scale in 2-D to show the large difference between skin pixels and non-skin pixels; **(d)** RG image displayed on a colour scale in 3-D to show the large difference between skin pixels and non-skin pixels. (c) and (d) shows that almost all non-skin pixels have negative values and that skin pixels have positive values.

After the subtraction, and due to the illumination effect, it is necessary to apply a morphological operation called *Closing* to eliminate the face's holes. Many times these holes result from shadows or reflexions inside the face area.

This morphological operation is constituted by dilation followed by erosion, both with the same structuring element, in this case, a circle with a radius of 10 pixels. Dilation adds pixels to the boundaries of objects in an image, and erosion removes pixels from these boundaries by taking into account the structuring element. If any

pixel in the neighbourhood of another given pixel has a value of 1 during dilation in a binary image, the value of the pixel of interest is set to 1. With erosion the process is identical, but the pixel value is set to 0 if any of the neighbour pixels has a value of 0 (MathWorks®, 2011). Figure 12 illustrates a morphological dilation operation on a binary image.

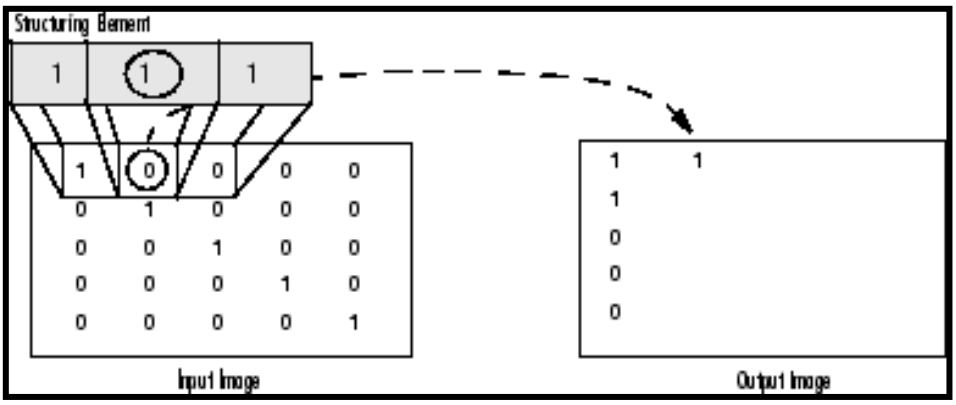

Figure 12 – Schematic representation of a morphological dilation operation. In this situation, the pixel of interest has a neighbour pixel with value of 1, so its value is set from 0 to 1 (MathWorks®, 2011).

The area of all objects in the image with connected pixels, face candidates, is calculated and the object with the largest area is chosen. By using the centroid and the area of this object, it is possible to draw a rectangular box around the face and track the face movement (Al-Shehri, 2004).

Figure 13 is a representation of the RG algorithm.

RG is a very simple and fast method to detect faces because it lies in a basic arithmetic operation. The RG algorithm is suitable for a large number of face detecting and tracking applications because it does not need a pre-processing step (Al-Shehri, 2004).

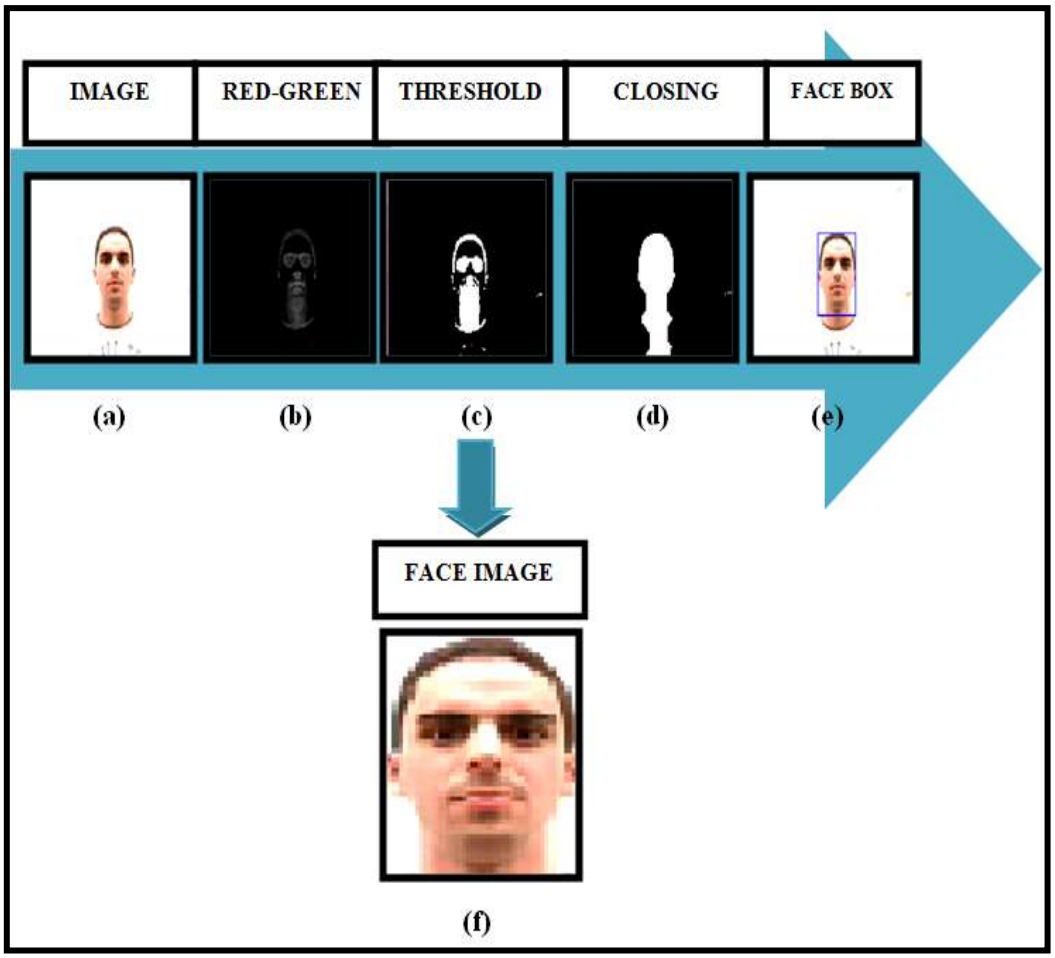

Figure 13 – Schematic representation of RG algorithm. (a) Acquisition of the RGB image; **(b)** Subtraction between images components Red and Green; **(c)** Application of the thresholds to select pixels with the highest values; **(d)** Application of the morphological operation *Closing* to connect pixels and eliminate holes; **(e)** Construction of the bounding box around the face; **(f)** Extraction of face image.

This method is also very accurate in detecting faces through various positions (frontal and profile), and to follow translational and rotational movements of the head. This happens because it is independent of face features or geometric relations that could possibly change with head movements. RG also detects all kind of skin tones (Al-Shehri, 2004).

Unfortunately, this method is influenced by strong and weak illuminations, by big skin-coloured objects in background, by clothes with colours similar to those present in the skin, and even by body parts present in the image (arms, hands, neck and neckline) (Al-Shehri, 2004).

The face is usually in the upper part of the largest object found. Therefore, the bounding box around the face can be automatically adapted when the clothes, the neckline or the neck are also detected. The algorithm used to cut the bounding box in these cases is:

$$I.\ If\ \frac{w_0}{w_i} > 0.5 \Rightarrow \begin{cases} w_0 = \frac{w_0}{2} \\ y_0 = y_0 + \frac{w_0}{4} \\ x_0 = x_0 \\ h_0 = h_0 \end{cases};\ Else \begin{cases} w_0 = w_0 \\ y_0 = y_0 \\ x_0 = x_0 \\ y_0 = y_0 \end{cases} \quad (4)$$

$$II.\ If\ \frac{h_0}{h_i} > 0.65 \Rightarrow \begin{cases} h_0 = \frac{2}{3}h_0 \\ y_0 = y_0 \\ x_0 = x_0 \\ w_0 = w_0 \end{cases};\ Else \begin{cases} w_0 = w_0 \\ y_0 = y_0 \\ x_0 = x_0 \\ h_0 = h_0 \end{cases} \quad (5)$$

where w_0 is the initial width of the bounding box, w_i is the total width of the image, h_0 is the initial height of the bounding box, h_i is the total height of the image and (x_0, y_0) are the coordinates of the upper-left corner of the bounding box (see Figure 14). The values 0.5 and 0.65 present in this algorithm were selected after the realization of tests with different face images. Figure 15 shows some examples of a cut of the bounding box.

The orientation of the face can also be calculated and be used to improve the construction of the bounding box. This benefits the experiment because if the face is very inclined relative to the body, the box could suffer some displacement. The method used to perform this adjustment is

$$If\ \theta < 0\ rad \Rightarrow \begin{cases} w_0 = w_0 + 15|\theta| \\ y_0 = y_0 - 15|\theta| \\ x_0 = x_0 \\ h_0 = h_0 \end{cases};\quad Else \begin{cases} w_0 = w_0 + 15\theta \\ y_0 = y_0 + 15\theta \\ x_0 = x_0 \\ h_0 = h_0 \end{cases} \quad (6)$$

where θ is the angle between the vertical axis of the face and the horizontal axis of the image (see Figure 14). This process is always applied to the bounding box that results from the algorithm described in (4) and (5). The value 15 on the above method was selected after testing several face images. Figure 16 illustrates this necessary adjustment of the bounding box.

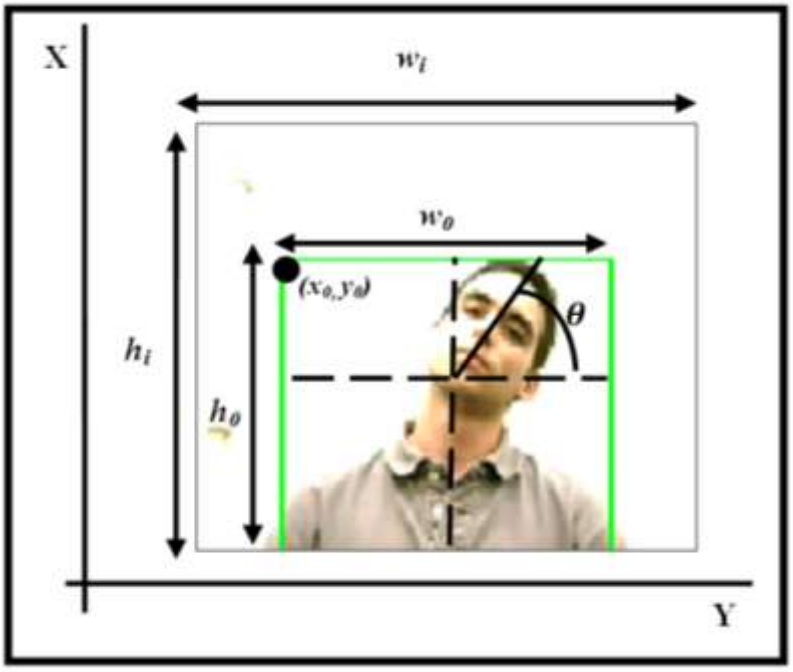

Figure 14 – Representation of the variables presented in the algorithms for the adaption of the face bounding box. θ is the angle between the vertical axis of face and the horizontal axis of the image, w_0 is the initial width of the bounding box, w_i is the total width of the image, h_0 is the initial height of the bounding box, h_i is the total height of the image and (x_0, y_0) are the coordinates of the upper-left corner of the bounding box.

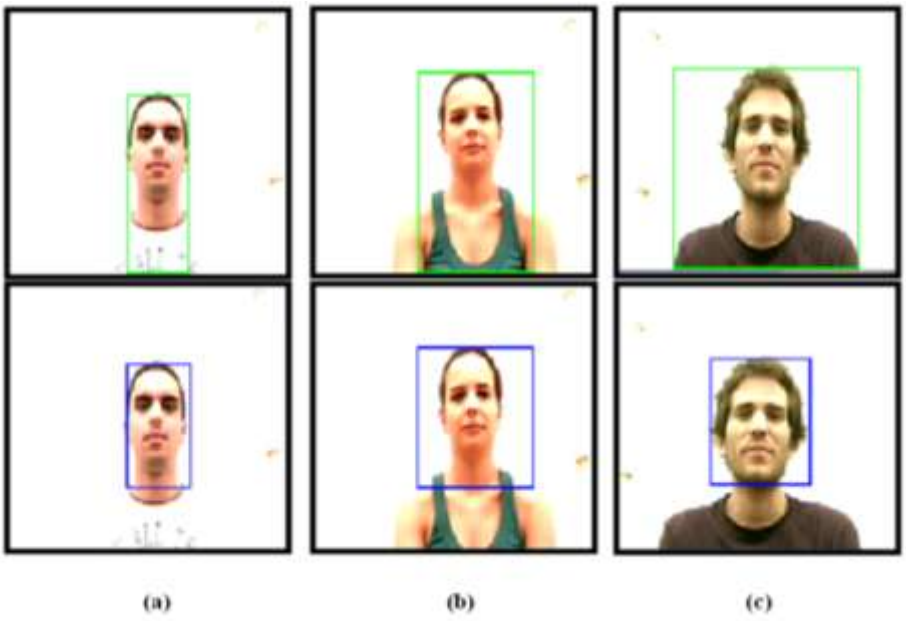

Figure 15 - Examples of automatic cuts in the face bounding box to avoid the detection of the neck (a), the neckline (b) and clothes (c). The green rectangles represent the original boxes after the application of RG and the blue rectangles represent the adapted bounding boxes.

Figure 16 - Example of adjustments in the face bounding box to avoid displacements which result from face tilt. The green box represents the original box after the application of RG, the red box represents the cut box due to the clothes effect, which is displaced relatively to the face position, and the blue box represents the adapted bounding box by using the orientation of face.

After the final bounding box is found, the face image inside it is scaled in order to always have images with the same size (50×50 pixels). This fact helps the implementation of eye region detection algorithms.

4.2.1.2. MD: Mahalanobis Distance

The colour-based segmentation algorithm described in this section is available for download in the section of SIMULINK® Demos corresponding to the Video and Image Processing Blockset in the software MATLAB®. This method was published with the version 7.11 of MATLAB® between 2004 and 2007 (MathWorks®, 2011).

The algorithm uses the square of Mahalanobis Distance to distinct skin pixels from non-skin pixels by comparing the result with a given threshold. In this experiment, the threshold has a value of 5 (MathWorks®, 2011).

Mahalanobis Distance is equal to

$$MD_i = \sqrt{(x_i - \bar{x})C_x^{-1}(x_i - \bar{x})^T} \tag{7}$$

where x_i is each element of the data matrix x, C_x^{-1} is the inverse of the covariance matrix of the data matrix x and $\bar{x}$ is the mean of the data matrix x (Maesschalck et al., 2000).

In this algorithm, x is the matrix correspondent to the image components C_B and C_R of YC_BC_R colour space. The values of $\bar{x}$ and C_x^{-1} were constructed by the author(s) of this algorithm by using a set of images with pixels corresponding to human skin. Therefore, $\bar{x} = 107.9649$ for the component C_B and $\bar{x} = 140.8913$ for the component C_R, both of the means calculated in a greyscale between 0 and 255, and (MathWorks®, 2011)

$$C_x^{-1} = \begin{bmatrix} 1.8328501 \times 10^3 & 2.2506719 \times 10^3 \\ 2.2506719 \times 10^3 & 6.8658257 \times 10^3 \end{bmatrix}. \tag{8}$$

In this method, because the original image is acquired in the RGB format, this image must be converted to the YC_BC_R colour space.

This algorithm also uses two filters, a Median Filter and a 2-D FIR Filter, and performs a morphological operation *Closing* to find the objects covered by skin in the output MD image (MathWorks®, 2011). However, these filters increase the processing time, making the implementation in real-time more difficult. For this reason, in this work, only the morphological operation *Closing* was implemented, like in the RG algorithm described in the last section.

After the application of the morphological operation *Closing*, this algorithm follows the same steps described in the RG algorithm, namely the determination of the area and the centroid of the biggest object in the image, the construction of the bounding box and the adaption of the scale of the face image.

Figure 17 is a schematic representation of the MD method.

MD algorithm applies more arithmetic operations than RG algorithm which performs only one simple subtraction, so it is expected that MD algorithm will need a longer processing time to detect a face on an image. However, MD was developed by using pre-defined values related to skin colours (the mean and the covariance), thus MD has a more restricted range of coloured pixels that can be classified as skin pixels. This method will probably be more accurate to detect human faces.

MD algorithm is influenced by illumination effects, the presence of body parts covered by skin and some objects in the background with similar colours to the skin, like the RG method.

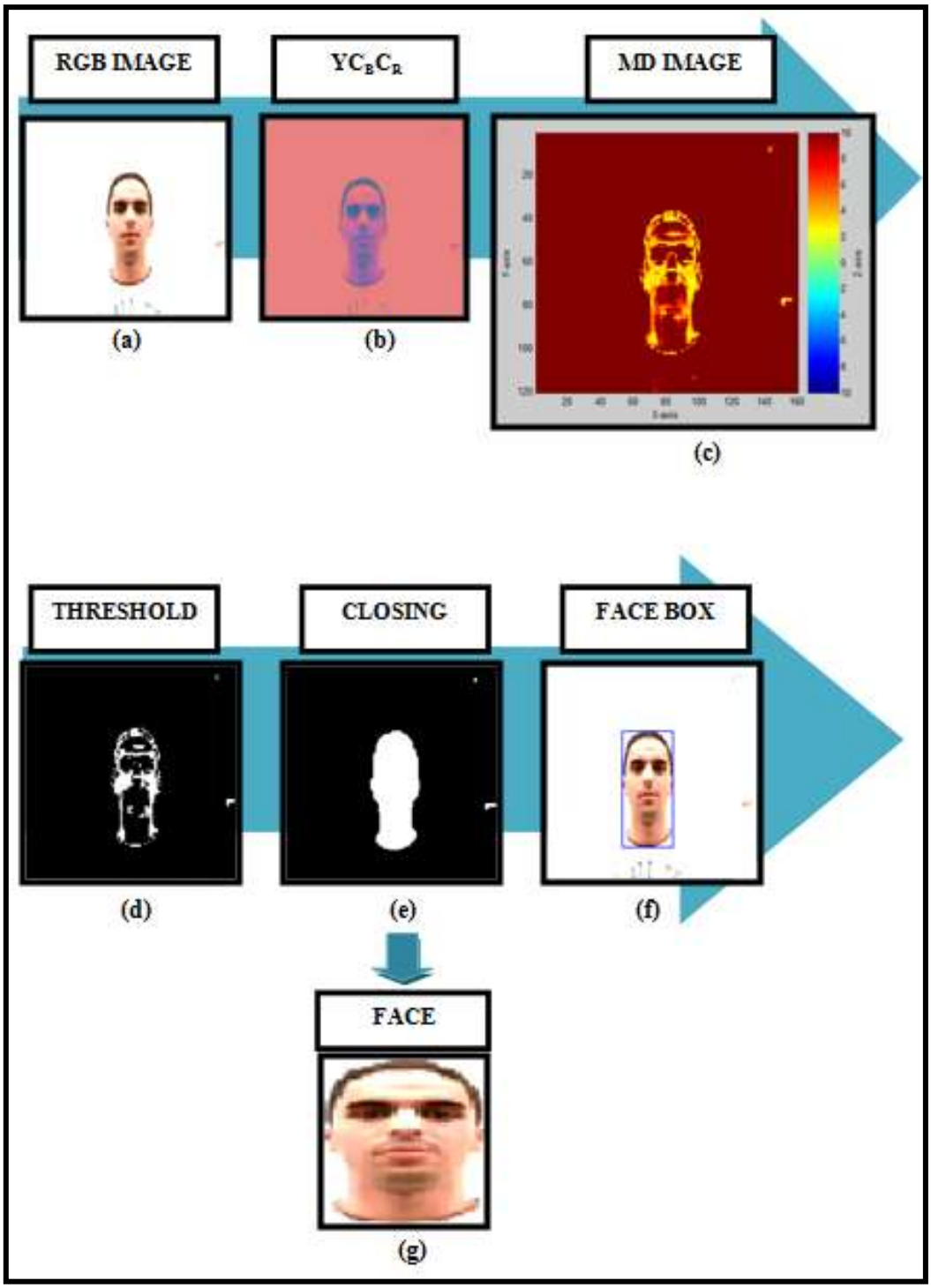

Figure 17 – Schematic representation of MD algorithm. (a) Image acquired in RGB colour space; **(b)** Conversion to YC_BC_R colour space; **(c)** Image after MD algorithm with a coloured scale to enhance contrast between the intensity of the pixels; **(d)** Application of the threshold to select skin pixels; **(e)** Application of the morphological operation *Closing* to connect pixels and eliminate holes; **(f)** Construction of the bounding box around the face; **(g)** Extracted face image.

4.2.2. Feature Based Algorithms

4.2.2.1. VJ: Viola – Jones

VJ (Viola-Jones Detector) is one of the most accurate and rapid methods applied for face detection developed by Viola and Jones (2001). Due to its low processing time, it is ideal for real-time processing. According to Viola and Jones, this method is advantageous because it can achieve "high frame rates working only with the information present in a single greyscale image", unlike other techniques that use auxiliary information like the pixel colour or the difference between two video frames. VJ is suitable for many applications as user interfaces, image databases and teleconferencing. Some face-detecting cameras, intelligent medical equipments and other "small low power devices" utilize this algorithm (Nielsen, 2010).

VJ was built by Viola and Jones following three steps related to Integral Images, Haar-like Features and Classifier Cascades. These steps will be described briefly in the next sections (Viola & Jones, 2001).

4.2.2.1.1. Integral Image

An integral image is a new representation of an image which results from a few operations per pixel in the original image. This new image will allow a fast creation of the features necessary to identify faces (Viola & Jones, 2001**;** Bucknall, 2010).

Integral images are also known as "Summed Area Tables" because they contain at a location (x,y) the sum of the pixels above and to the left of the point (x,y) of the original image, inclusive

$$ii(x, y) = \sum_{x' \leq x,\, y' \leq y} i(x', y') \tag{9}$$

where *ii(x,y)* is the integral image and *i(x,y)* is the original image. The integral image can be obtained in a single step using two recurrences

$$s(x, y) = s(x, y - 1) + i(x, y) \tag{10}$$

$$ii(x, y) = ii(x - 1, y) + s(x, y) \tag{11}$$

where *s(x,y)* is the cumulative row sum (*s(x,-1)=0* and *ii(-1,y)=0* are the initial conditions) (Nielsen, 2010; Viola & Jones, 2001).

Rectangular sums can be calculated with the integral image by using only four array references (Viola & Jones, 2001). An example of this process is illustrated in Figure 18.

The features used to analyse faces are obtained with these rectangular sums because they have information about the colour intensity of the image in given areas (Nielsen, 2010).

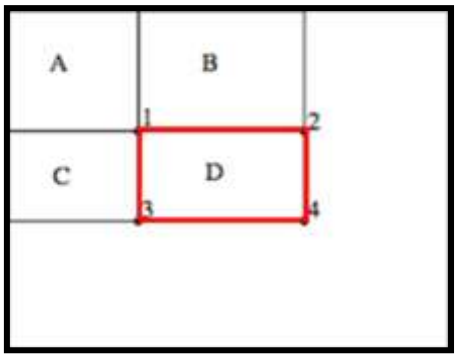

Figure 18 – Schematic representation of how to obtain an integral image and the sum of the pixels within a rectangular area, D. The value of the integral image at point 1 is the sum of pixels inside the rectangle A, at the point 2 is the sum of pixels inside rectangles A and B, at point 3 is the sum o pixels inside the rectangles A and C and at point 4 is the sum of pixels inside the rectangles A, B, C and D. The sum of pixels inside the rectangle D is calculated with these last four values of the integral image: *ii*(point 1)+*ii*(point 4)-*ii*(point 2)-*ii*(point 3)=(A)+(A+B+C+D)-(A+B)-(A+C)=D. Image adapted from (Viola & Jones, 2001).

4.2.2.1.2. Haar-Like Features

The rectangular features are also called "Haar-like features", because they work similarly to the Haar wavelets applied by Alfréd Haar to analyse complex waveforms. When the integral image is obtained, these features can be calculated at various scales and locations on the original image (Viola & Jones, 2001; Bucknall, 2010).

By using the values of the integral image, it is possible to find dark rectangles adjacent to light rectangles, because the sum of pixels correspondent to the dark areas is low and in the light areas is high. The adjacent dark and light blocks are used as features in VJ. These blocks have the same size and shape and are connected horizontally or vertically (Viola & Jones, 2001; Bucknall, 2010).

The feature value is equal to the difference between the sum of pixels in light rectangle(s) and the sum of pixels in dark rectangles(s). This value is used in a

filter to conclude if the feature is or is not present in the original image (Viola & Jones, 2001; Bucknall, 2010). Figure 19 represents some features commonly used in VJ.

The rectangular features are simple and primitive, but they allow a rapid analysis and a good representation of the image in terms of effective learning, despite being a low-resolution representation. These features can detect edges, bars and other simple structures on the image and can only be applied in three orientations: horizontal, vertical and diagonal (Viola & Jones, 2001; Bucknall, 2010). Figure 20 shows the three possible orientations of the rectangular features.

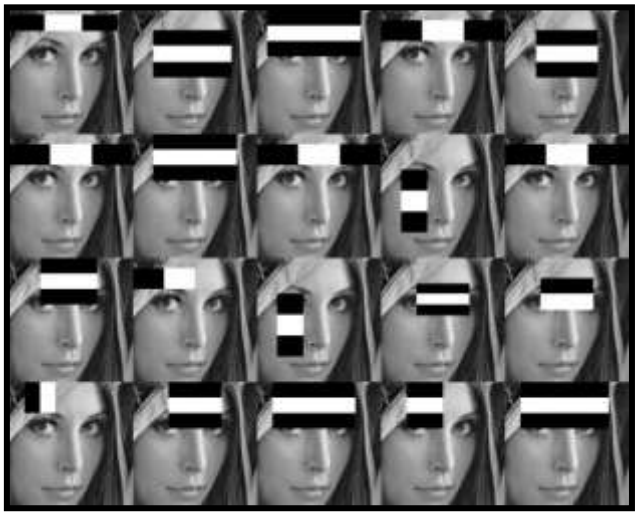

Figure 19– Example of features used in Viola-Jones (Harvey, 2010).

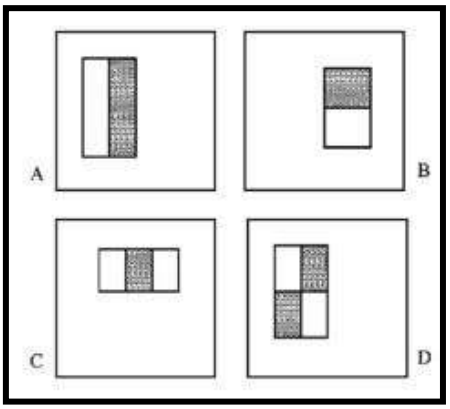

Figure 20 – Orientations of rectangular features used in Viola-Jones. A and B are two-rectangle features vertically oriented, C is a three-rectangle feature horizontally oriented and D is a four-rectangle feature diagonally oriented (Viola & Jones, 2001).

4.2.2.1.3. Classifier Cascade

The Haar-like features are called "weak classifiers" because they are very common within a face image, thus it is very difficult to decide if face is or is not present by using only one of these features (Nielsen, 2010).

The total number of rectangular features to test in an image sub-window is very large; therefore, their analysis requires expensive computation (Viola & Jones, 2001).

To overcome this problem, Viola and Jones (2001) constructed a classifier by modifying a training system called AdaBoost, an artificial intelligence method similar to Neural Networks. This method selects important features for face detection to speed up the classification process. The AdaBoost algorithm is trained with thousands of face and non-face images and, after that, it can extract critical facial features or classifiers. Critical features are the features that best separates the face images and non-face images (Bucknall, 2010; Nielsen, 2010).

VJ uses a trained classifier cascade to identify faces. A cascade is a combination of classifiers trained by the AdaBoost method. Each stage classifier is slightly more complex than the classifier in the previous stage. Each stage classifier is

trained with the selected features and the sub-windows that pass through the previous classifier (Viola & Jones, 2001; Bucknall, 2010).

The classifier cascade analyses sub-windows of the image to obtain more specific features and reject windows which fail in the classification process. The combination of selected weak classifiers during cascade stages results in a strong classifier that could detect faces with high accuracy and high frame rate. A sub-window that passes through all classifiers is considered a face (Bucknall, 2010; Nielsen, 2010).

A classifier cascade is defined by three fundamental parameters: the number of classifier stages, the number of features per stage and the threshold classification function of each stage. This function compares the sum of all previous classifiers weights attributed to a feature to a pre-defined threshold (Viola & Jones, 2001; Bucknall, 2010). Figure 21 illustrates the operation method of a classifier cascade.

The cascade train is a very extensive process but, on the Internet, it is possible to find already trained cascades for the detection of frontal and profile faces and faces including the neck and the shoulders. In this way, it is easier to use these cascades in order to save some work time. This work uses a Viola-Jones method developed by Kroon (2010) which is available for download in the website of MathWorks® (MathWorks®, 2011; Nielsen, 2010).

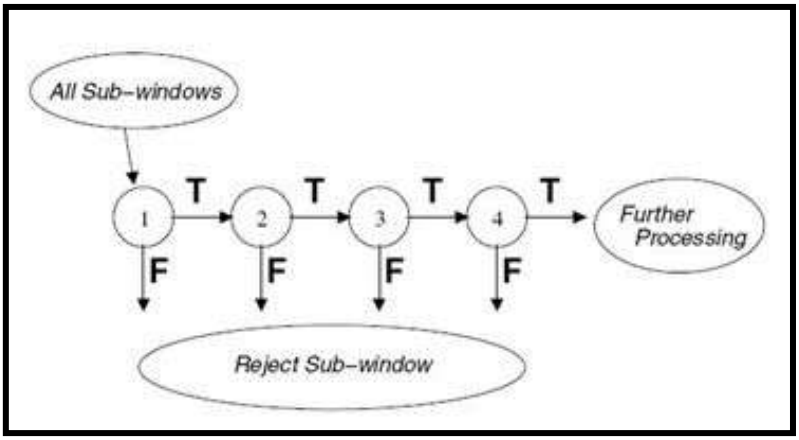

Figure 21 – Schematic representation of a classifier cascade (Viola & Jones, 2001).

In this experiment, the method developed by Kroon (2010) was changed to improve the algorithm processing time. The original algorithm was designed to process images; therefore, its application to videos has a low frame rate.

In this work, two types of VJ method above mentioned were used: VJ Heavy and VJ Light. The first one has a lower frame rate and is only used in the first two frames of the processed video. VJ Light is used in the remaining video frames and has a higher frame rate. Differences between the frame rates result from the number of stages and feature scales in the classifier cascade. VJ Light uses 5 stages and VJ Heavy uses 22 stages. VJ Light starts at the scale where VJ Heavy has stopped; therefore, it does not need to go through all the scales which had already been used by VJ Heavy, because VJ Heavy had already detected the approximated face position.

VJ Heavy is used to calculate the approximated face position in the input greyscale image. With this information, the input image is cut. The cut image is the input image of the VJ Light. To ensure that the VJ algorithm is correctly initialized, during the VJ Heavy the subject must be still.

VJ Light calculates a set of bounding boxes around the face position. To select the final bounding box, a binary image from the cut image is created (by using a threshold value between 0.45 and 0.6). The mean of the region inside each bounding box in the binary image is calculated and the box with the highest mean is chosen.

The adjustments of the bounding box described in RG are not performed in this algorithm, but the adaption of the scale of face image is applied.

Viola-Jones detector is not influenced by face illumination, face pose and face scale, and by camera variation (Viola & Jones, 2001).

Figure 22 is a schematic representation of the VJ method.

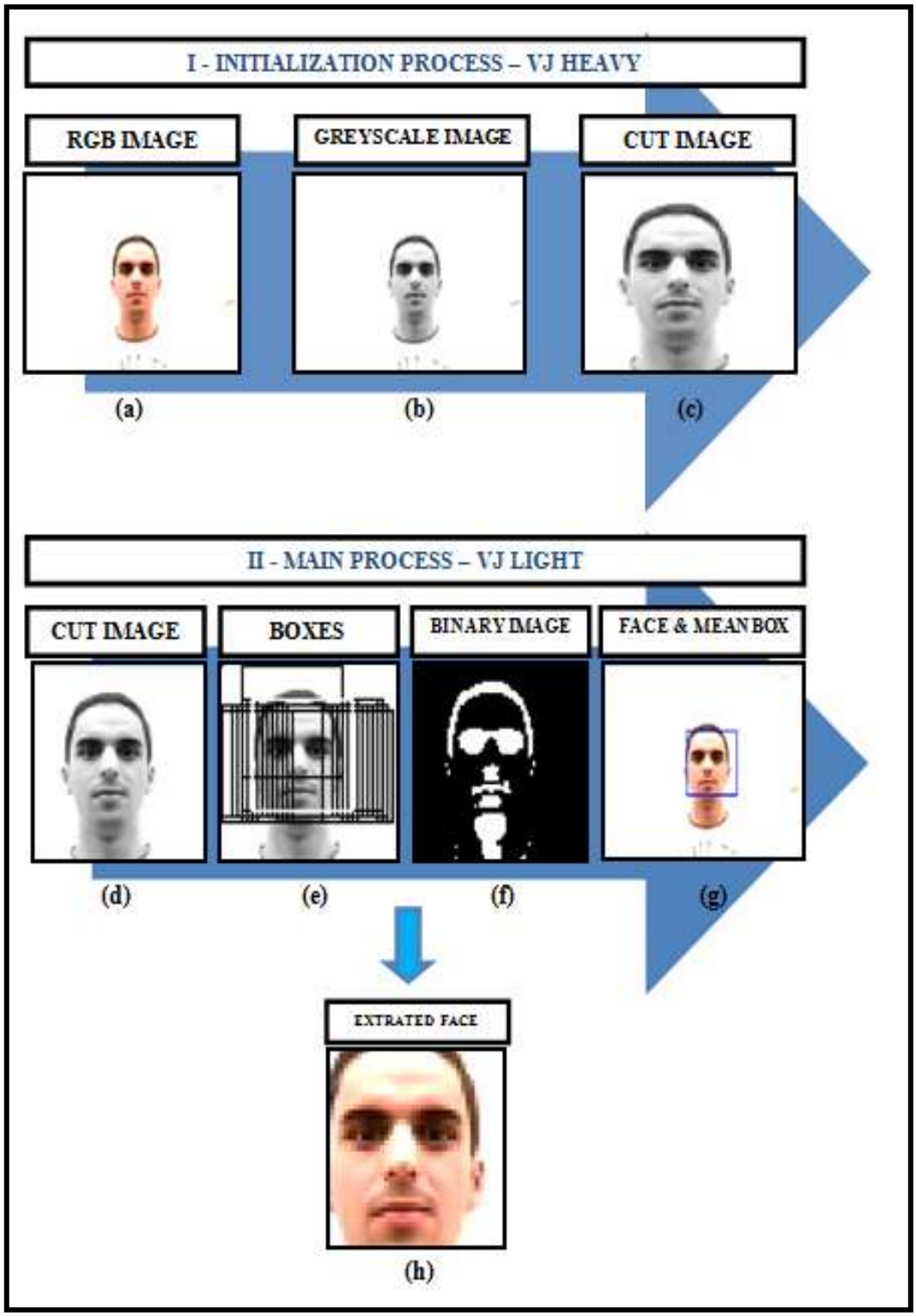

Figure 22– Schematic representation of VJ algorithm. (a) Image acquired in RGB colour space; **(b)** Conversion to greyscale; **(c)** and **(d)** Cut Image after VJ Heavy algorithm; **(e)** Boxes resultant from VJ Light algorithm; **(f)** Application of a threshold to obtain a binary image; **(g)** Construction of the bounding box around the face by using the mean of the binary image inside each box; **(h)** Extracted face image.

4.3. EYE REGION DETECTION ALGORITHMS

To locate the eye region in a previously detected face image , two methods were studied: Between The Eyes (BTE) and Eyes Gabor Filter (EGF).

These methods depend on the output of the face algorithms, thus if the face is detected with errors, the localization of the eye region will be affected.

Eye detection algorithms have two main goals in this work. In first place, these methods allow the segmentation of the region containing the eyes which is essential to the estimation of the gaze point. Second, they calculate a reference point to the estimation of the distance between the subject and the cameras, necessary to determine the gaze point on a 3-D space. The distance calculation will be discussed later in this work.

BTE and EGF techniques will be described in detail in the next sections.

4.3.1. BTE - Between The Eyes

BTE is a method applied to find the middle point between the two eyes and was developed having as reference the work of Peng et al. (2005). Other authors have created similar methods, like Ramezanpour et al. (2010) and Kawato and Tetsutani (2002).

This algorithm uses the face image resultant from one of the face detection processes described above. In the first step, the RGB face image is converted to greyscale. By using the greyscale image and a mask corresponding to the binary image of the face, also used in face segmentation algorithms, a greyscale face image with black background is obtained.

In the second phase of this method, the horizontal gradient of the greyscale image with black background is calculated by a Sobel Gradient Operator[1] with a threshold scale value of 4. Eyes have a greater variation of pixel intensity than the face, thus, in the gradient image the region containing the eyes will have pixels with higher values than the other parts of the face. To avoid the detection of other face regions that could be enhanced by the gradient application, the gradient image is divided in half, using the information that the eyes are in the upper part of the face.

After this, the horizontal projection of the half part of the gradient image is calculated, this is, the cumulative sum of all values of pixels in the image rows. The maximum value of the projection corresponds to the horizontal coordinate of the point between the eyes, because pixels in the eye region have the higher values, as was already mentioned (Peng et al., 2005).

In the last step, a morphological erosion operation (with a circular structuring element with a radius of 4 pixels) is applied to the greyscale face image with black background to enlarge the area of the dark regions. The horizontal coordinate calculated before is used to horizontally cut the greyscale image in order to segment the eyes region. After, it is performed a vertical projection (a cumulative sum of all pixels values over each column) and the maximum value of the result is found. This point corresponds to the middle part of the eyes which has pixels with higher values than its neighbour areas due to its special shine. In this way, the vertical coordinate of the point between the eyes is calculated (Peng et al., 2005).

Figure 23 represents a scheme of all steps of BTE method.

According to Peng et al. (2005), their algorithm is influenced by light reflexions present on the face and by the use of eyeglasses, since they also create unwanted reflexions inside the eye region. In addition, if the background of the input face image is not black, the algorithm will not work properly, thus it is necessary to pre-process the face image.

[1] "Sobel method finds edges using the Sobel approximation to the derivative". This method returns edges at points where the gradient of image is higher than a given threshold (MathWorks, 2011).

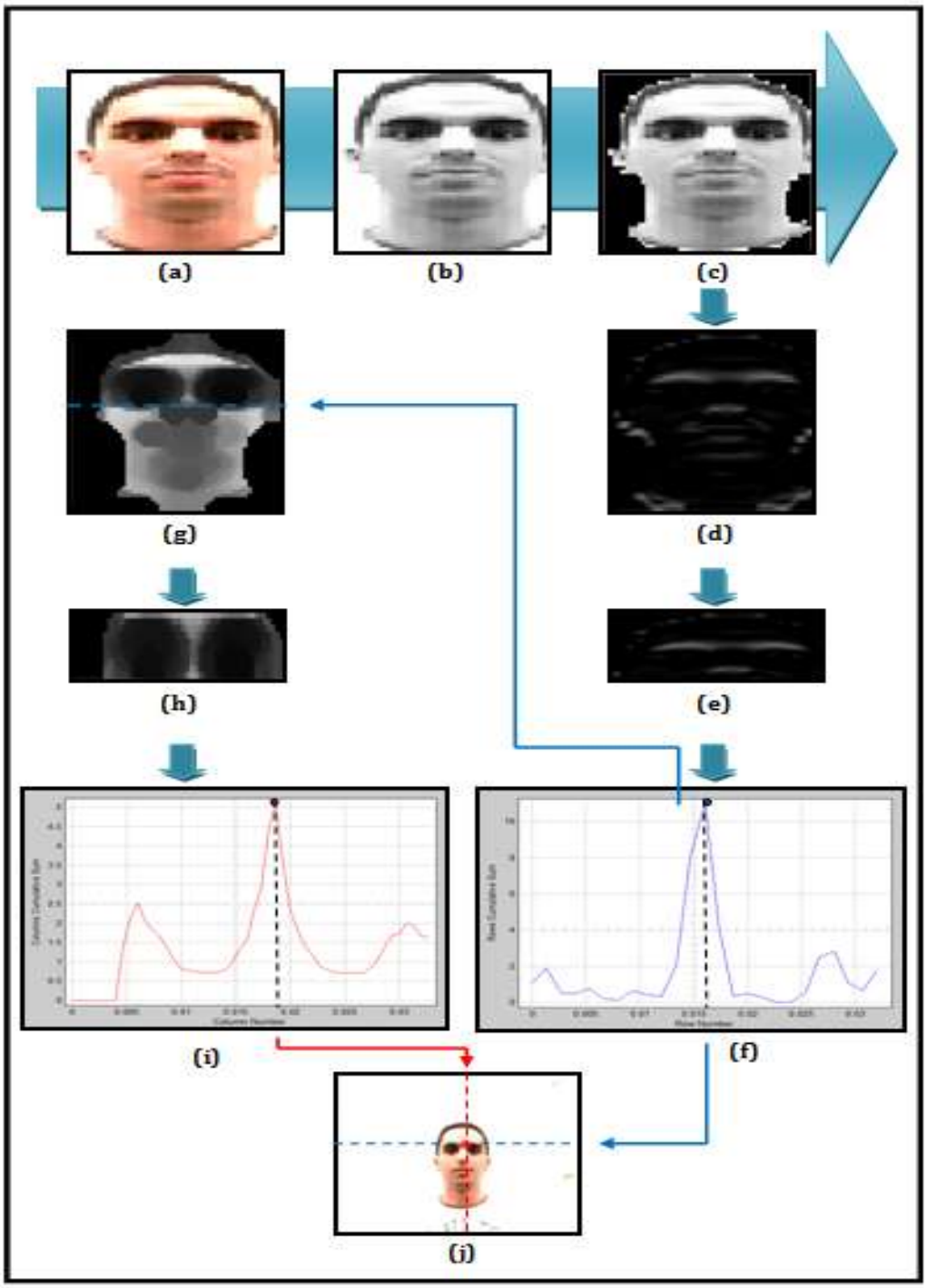

Figure 23– Schematic representation of the BTE method. (a) Face image in RGB; **(b)** Face image in greyscale; **(c)** Face image in greyscale with black background; **(d)** Horizontal gradient image; **(e)** Half of the horizontal gradient image; **(f)** Horizontal projection of the half of the horizontal gradient image and its maximum value; **(g)** Eroded greyscale face image; **(h)** Cut eroded image by using the BTE horizontal coordinate; **(i)** Vertical projection of the cut eye image and its maximum value; **(j)** Original image with the point between the eyes.

4.3.2. EGF - Eye Gabor Filter

Gabor transform was implemented by Dennis Gabor in 1946 and is used in many applications, such as texture segmentation, fingerprint recognition, face recognition and handwritten number recognition (Zhang et al., 2008).

Gabor filters are mainly used to extract textural information from images. There is a large contrast between the eyes and the other skin regions on face, thus a Gabor filter can be applied to segment the eye region (Xiong et al., 2007; Daugman, 1993).

The EGF applied in this work is based on the method developed by Zhang et al. (2008). These authors have developed two Ring Gabor Filters, this is, 2-D band pass filters that have a pass band with a ring shape and Gaussian cross sections in the frequency domain. One of the Ring Gabor Filters has an elliptical shape and it is applied to detect the eyes region and the other one has a circular shape and it is used to segment the iris. Here, the elliptical shaped filter was used.

An Elliptical Gabor Filter is a Gaussian function modulated by a sine function with a given shape and is defined by

$$G(x,y) = g(x',y')\exp\left(2\pi jF\sqrt{\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}}\right) \quad (12)$$

where

$$(x',y') = (x\cos\theta + y\sin\theta, -x\sin\theta + y\cos\theta)\,, \quad (13)$$

F is the spatial central frequency of the filter in the frequency domain, θ is the rotation angle of the filter, (x, y) is the centre of the receptive field in the spatial domain and $g(x,y)$ is the 2-D Gaussian envelop defined by

$$g(x,y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right)\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)\right] \quad (14)$$

where σ_x and σ_y are variances or scale factors along x and y axis, respectively (Zhang et al., 2008).

After the implementation of the filter described above, only its real part is used on the image. The filter is also normalized by diving it by its maximum value.

If the Elliptical Gabor Filter is designed with the ideal frequency F, the angle of rotation θ can be set to zero because the algorithm will not be affected by it. In this way, only three parameters must be optimized for an accurate result: σ_x, σ_y and F. Pixel values in eye region change more in the horizontal direction, thus the eyes width must be considered as a reference for choosing the scale factors σ_x and σ_y (Zhang et al., 2008). After testing some images, the values of σ_x, σ_y and F were chosen: 14.5 pixels, 15 pixels and 59 Hz, respectively. In Figure 24, some examples of Elliptical Gabor Filters used by the algorithm authors can be seen.

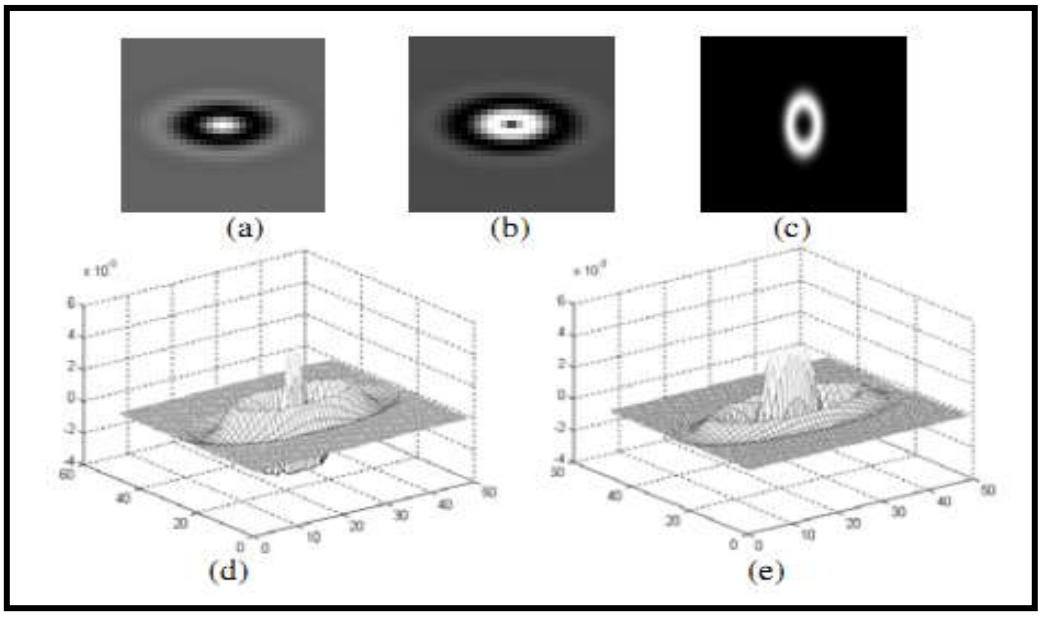

Figure 24 – Examples of Elliptical Gabor Filters used by Zhang et al. (2008). (a), **(b)** and **(c)** are filters represented in 2-D; **(d)** and **(e)** are filters represented in 3-D.

The Elliptical Gabor Filter enhances the edge of eyelids because it is selective to orientation and “matches the eye in shape”. The application of this algorithm produces accurate and rapid results, according to the authors. In addition, this method is little influenced by face pose or rotation, face accessories like eyeglasses, illumination effects, and noise effects. However, defections of the face can slightly decrease the accuracy of this method (Zhang et al., 2008; Xiong et al., 2007).

The Eye Gabor Filter or Elliptical Gabor Filter is convolved with the greyscale face image and, as result the eye region in the output image has pixels with lower values than the other face pixels. After this, the output image is binarized and a threshold equal to 85% of the minimum value of the filtered image is applied to obtain the eye candidates. The centroids of these objects are calculated and used to build rectangular boxes around both of the eyes (Zhang et al., 2008).

Figure 25 shows all the algorithm steps described above.

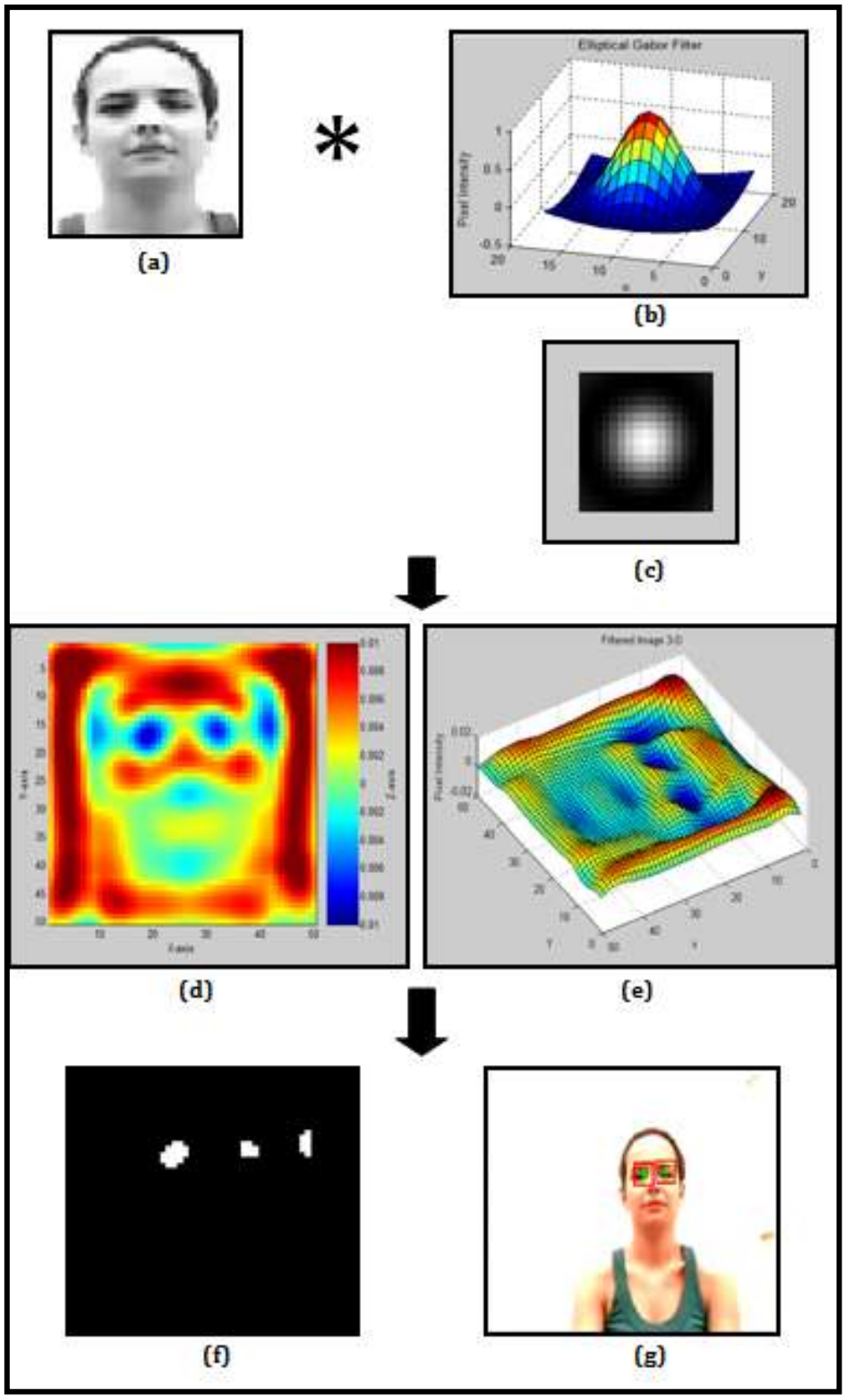

Figure 25 – Schematic representation of the EGF algorithm. (a) Face image in greyscale; **(b)** Eye Gabor filter in 3-D; **(c)** Eye Gabor filter in 2-D; **(d)** Filtered image in 2-D; **(e)** Filtered image in 3-D; **(f)** Image after the application of the threshold; **(g)** Original image with the detected eyes.

4.4. DISTANCE CALCULATION BY BINOCULAR VISION

As already mentioned in the last section, the results of BTE or the EGF algorithm will be use as reference points in the calculation of the distance between the subject and the cameras, in order to estimate the gaze point on a 3-D space.

This distance is computed by using the disparity of the images from the two lateral cameras. A binocular vision method, also known as binocular stereo triangulation, is used. This approach is a very simple and efficient method (Wu et al., 2009).

This algorithm calculates the distance using three different expressions, depending on the object position in the image plane: (a) the object is between the two cameras centres, (b) the object is at the left of the left camera centre and (c) the object is at the right of the right camera centre. Here, the term object is used as the point(s) calculated by the eye region detection algorithms, this is, the point between the eyes in the BTE method or the centres of the two eyes in the EGF method (Wu et al., 2009). Figure 26 shows the schematic representation of all the three cases.

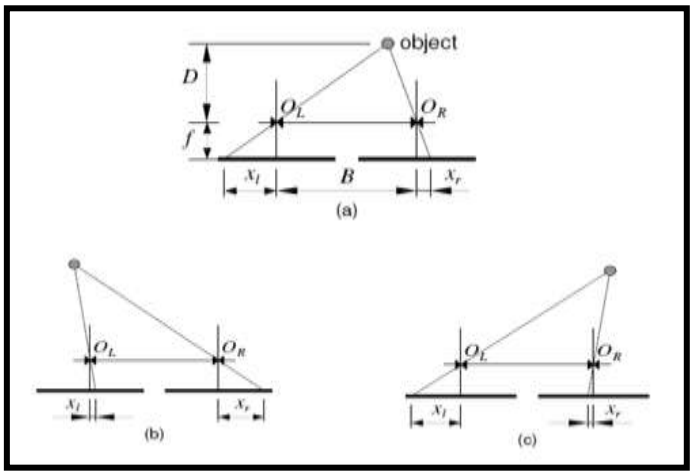

Figure 26 – Three cases of the object position in the image plane. (a) the object is between the two cameras centres; **(b)** the object is at left of the left camera centre; **(c)** the object is at right of the right camera centre (Wu et al., 2009).

The distance between the subject and the cameras, *D*, is calculated by one of the three following formulas, depending on the cases (a), (b) and (c) described in Figure 26:

If x_l<0 and x_r>0 or x_l>0 and x_r<0

$$D = \frac{fB}{|x_l| + |x_r|} = K\frac{fB}{n_{pixel}} \tag{15}$$

If x_l>0 and x_r>0

$$D = \frac{fB}{x_r - x_l} \tag{16}$$

If x_l<0 and x_r<0

$$D = \frac{fB}{|x_l| - |x_r|} \tag{17}$$

where *B* is the baseline or the distance between the cameras aperture diaphragm O_L and O_R (see Figure 26), *f* is the focal length of the two lateral cameras, x_r and x_l are the absolute horizontal distance between the image point and the left and right image centres, respectively, *K* is a constant and n_{pixel} is the displacement of pixels (Wu et al., 2009).

If cameras with different focal lengths or with non-parallel optical axes were used, the above expressions should be changed to include the angle between the optical axes, and two distinct focal distances (Wu et al., 2009). However, in this work, the optical axes of the cameras are parallel and the two cameras have the same focal

length, thus the equations (15), (16) and (17) are sufficient to calculate the distance.

According to the authors, when the object is between the two camera centres, the calculated distance will have a lower error than in the other two cases. By changing the length of the baseline *B*, the object can be maintained between the two cameras centres most of the time (Wu et al., 2009).

A lower distance between the cameras implies a lower accuracy but allows a greater viewing angle.

The algorithm developed in this work to determine the distance between the subject and the acquisition system follows three main steps. First, the reference points are detected by the eye region detection algorithms described in the last section. After this, the horizontal distance in pixels between the reference points and the centre of each of the images is calculated. In the last step, this distance in pixels is converted by using the known dimensions of CCD sensor of the cameras, in order to have a final distance in meters. This conversion is performed by using the following expression

$$d_{meter} = s_{pixel}\, z\, d_{pixel} \tag{18}$$

where s_{pixel} is the size of each CCD pixel, d_{pixel} is the distance in pixels and

$$z = \frac{h_{CCD}}{h_{image}} = \frac{w_{CCD}}{w_{image}} \tag{19}$$

where h_{CCD} and w_{CCD} are the height and the width of the CCD sensor, respectively, and h_{image} and w_{image} are the height and the width of the images from both cameras, respectively. Figure 27 illustrates an example of the process of distance calculation above described. Table III presents the values of the variables used in this experiment.

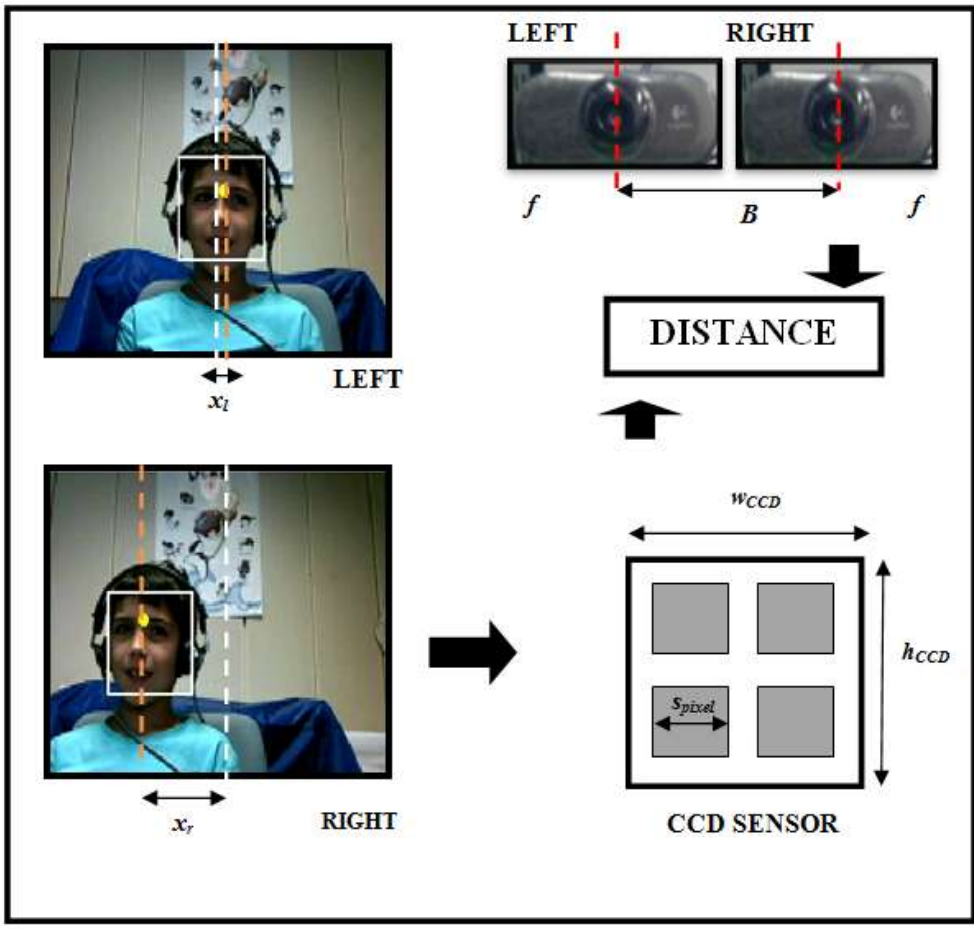

Figure 27 – Schematic representation of the calculation of the distance between the subject and the acquisition system. In this example, the face was detected using VJ algorithm and the eye region was segmented by BTE algorithm. The point between the eyes is the reference point in the left and right images. The horizontal distances in pixels x_r and x_l between these reference points and the centres of the images are converted to meters by using the dimensions of CCD sensor. By using the distance between cameras *B*, the focal distance *f* and the distances previously detected, the final distance is determined.

Table III – Variables values used in distance calculation algorithm.

VARIABLE	VALUE
f	0.0025 meters
B	0.1 meters
s_{pixel}	2.8×10^{-6} meters
w_{CCD}	640 pixels
h_{CCD}	480 pixels
w_{image}	160 pixels
h_{image}	120 pixels
z	4

CHAPTER 5

RESULTS

5.1. EXPERIMENTAL PROCEDURE

To test the face and eye algorithms described in Chapter 4, a small video database was created. The database is composed by 6 face videos of 23 subjects. The total number of videos is 138.

The 23 subjects have different haircuts, hair colours, eye colours, and skin tones. Several subjects wear eyeglasses or a beard. The majority of the subjects is Caucasian. Figure 28 shows an example of the variability among subjects from the database created in this experiment.

In each one of the 6 videos per subject, the subject was asked to do a different type of head movement. Table IV summarizes the type of head movements in the videos. Figure 29 illustrates the types of head movements described in Table IV.

Videos were acquired with a frame rate of 30 frames/second and each one has duration of 10 seconds; therefore, each video has 300 frames. Each video has a resolution of 120×160 pixels. In the first 2 seconds of all videos, subjects were asked to be still in order to initialize the algorithms, as aforementioned in Chapter 4.

All videos were obtained with the same background, the same illumination conditions, and the same initial distance between the subjects and the camera. This distance was not maintained constant throughout all tests due to the subject head movement.

To analyse and compare the results of the face and eye algorithms described in Chapter 4, all videos were studied manually.

The results were obtained with an Intel Core 2 Duo CPU E6750 @ 2.66 GHz computer (RAM 2GB and 64-bits Operating System).

All results are related to an off-line analysis because algorithms were applied to videos that had been previously recorded.

A summary of important results will be presented in the next section of this chapter. The detailed results for all subjects are in Appendix A in Table VIII.

Figure 28 – Example of the variability of the database created in this experiment.

Table IV – Type of head movements performed by each subject during experiments.

Video	Description of Head Movement
I	No movement;
II	Movement front-back;
III	Movement left-up-right;
IV	Movement up-down;
V	Movement left-front-right;
VI	Free movement of the body with the head still.

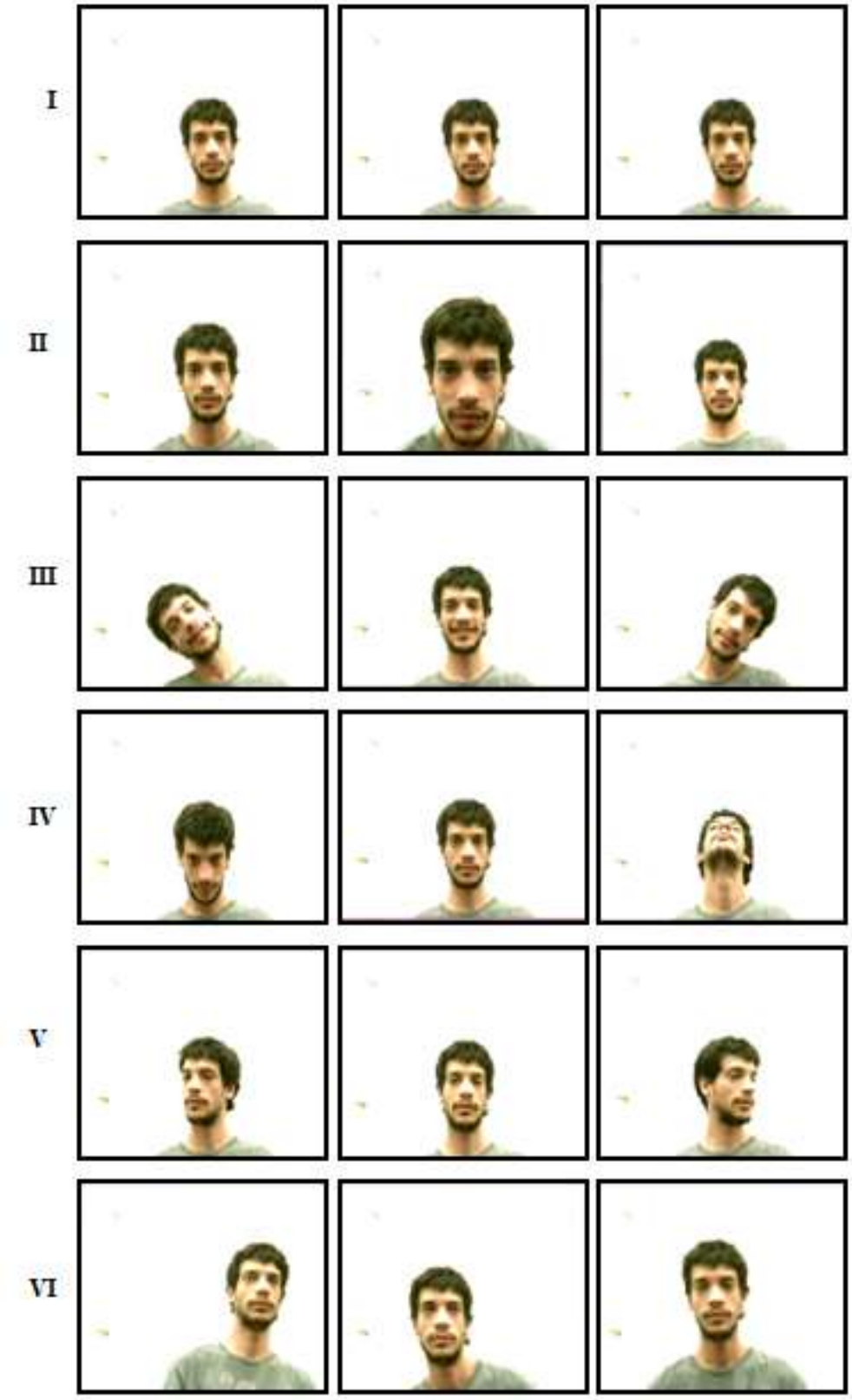

Figure 29 –Type of head movements performed by each subject during tests. I – No movement; **II** – Movement front-back; **III** - Movement left-up-right; **IV** - Movement up-down; **V** - Movement left-front-right; **VI** - Free movement of the body with the head still.

5.2. RESULTS OF FACE DETECTION AND TRACKING

In order to compare the detection rate of the three face algorithms described in Chapter 4, a classification scale was developed:

0. Face is not detected;
1. Face is detected and tracked with errors;
2. Face is correctly detected and tracked.

Figure 30 demonstrates an example of each case of the classification scale.

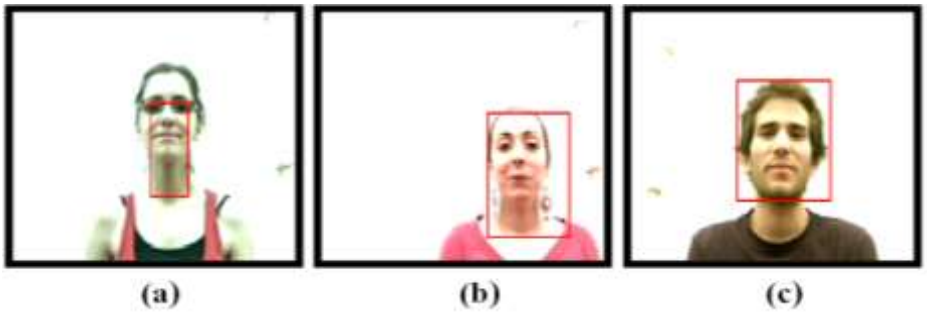

Figure 30 – Examples of each case of the classification scale obtained with MD method. (a) 0 – Face is not detected; **(b)** 1 – Face is detected and tracked with errors (in this case, the neck is detected); **(c)** 2 - Face is correctly detected.

Table V describes the results obtained with the three different face algorithms: RG, MD and VJ. In this table, the frequency of each case of the classification scale was calculated for all types of head movements previously described. A final percentage rate was determined according to the case of the classification scale and the total number of videos.

Table V – Results of face algorithms. The frequency of each case of the classification scale is represented for each type of head movement.

FACE METHOD	SCALE	TYPE OF HEAD MOVEMENT						TOTAL SUM	TOTAL SUM/TOTAL NUMBER OF VIDEOS (%)
		I	II	III	IV	V	VI		
RG	0	2	0	1	1	1	0	**5**	**3.62**
	1	11	9	13	11	11	10	**65**	**47.10**
	2	10	14	9	11	11	13	**68**	**49.28**
MD	0	2	0	2	2	2	0	**8**	**5.80**
	1	10	10	14	13	7	9	**63**	**45.65**
	2	11	13	7	8	14	14	**67**	**48.55**
VJ	0	3	1	1	3	3	2	**13**	**9.42**
	1	0	8	5	3	1	6	**23**	**16.67**
	2	20	14	17	17	19	15	**102**	**73.91**

5.3. RESULTS OF EYE REGION DETECTION

A similar classification scale was created to compare the efficiency of the eye region detection algorithms explained in Chapter 4:

0. Algorithm does not work;
1. Algorithm works with errors;
2. Algorithm works correctly.

Figure 31 demonstrates an example of each case of the classification scale.

Figure 31 – Examples of each case of the classification scale obtained with BTE method. (a) 0 – Algorithm does not work; **(b)** 1 – Algorithm works with errors (in this case, the point between the eyes was briefly deviated from its correct place); **(c)** 2 - Algorithm works correctly.

Table VI describes the results obtained with the two different eye region algorithms BTE and EGF conjugated with the face algorithms. In this table, the frequency of each case of the classification scale was calculated for all types of head movements previously described. A final percentage rate was determined according to the case of the classification scale and the total number of videos.

Table VI –Results of eye region detection algorithms discriminated by each case of the classification scale, the type of head movement and the face algorithm used. The frequency of each case of the classification scale is represented for each type of head movement.

Eye Method	Face Method	Scale	Type of Head Movement						Total Sum	Total Sum/Total Number of Videos (%)
			I	II	III	IV	V	VI		
BTE	RG	**0**	7	4	9	7	7	5	**39**	**28.26**
		1	4	4	10	4	8	11	**41**	**29.71**
		2	12	15	4	12	8	7	**58**	**42.03**
		0	7	5	9	7	10	7	**45**	**32.61**

Eye Method	Face Method	Scale	Type of Head Movement						Total Sum	Total Sum/Total Number of Videos (%)
			I	II	III	IV	V	VI		
	MD	**1**	6	8	8	9	5	10	**46**	**33.33**
		2	10	10	6	7	8	6	**47**	**34.06**
		0	4	3	6	4	4	4	**25**	**18.12**
	VJ	**1**	8	10	7	6	7	10	**48**	**34.78**
		2	11	10	10	13	12	9	**65**	**47.10**
		0	3	1	2	2	1	1	**10**	**7.25**
	RG	**1**	2	8	6	1	3	4	**24**	**17.39**
		2	18	14	15	20	19	18	**104**	**75.36**
		0	4	2	3	3	3	2	**17**	**12.32**
EGF	**MD**	**1**	1	8	10	1	5	9	**34**	**24.64**
		2	18	13	10	19	15	12	**87**	**63.04**
		0	3	1	1	3	3	3	**14**	**10.14**
	VJ	**1**	0	4	6	3	2	3	**18**	**13.04**
		2	20	18	16	17	18	17	**106**	**76.81**

5.4. PROCESSING TIME RESULTS

The analysis of frame rate of all algorithms is very important in order to decide if these algorithms are suitable for a real-time final application. Table VII shows the median frame rate of face methods conjugated with both of the eye region

methods for all subjects. The frame rate for each video is the mean of number of frames per second in the 300 frames. The median was selected instead of the mean because it is less sensitive to outliers.

When the face algorithm does not detect the subject face, the frame rate is considered zero. In these cases, the software operates with a higher frame rate equal to the frame rate measured when the system is only importing and displaying a video. If the real frame rate value is used, the final median value of frame rate will be misleading.

Table VII – Frame rate results of face and eye region algorithms. The presented values correspond to the median of the frame rates for all subjects.

EYE METHOD	FACE METHOD	MEDIAN OF FRAME RATE (FRAMES/SECOND)
BTE	RG	43
	MD	37
	VJ	23
EGF	RG	62
	MD	34
	VJ	18

CHAPTER 6

DISCUSSION

6.1. FACE ALGORITMS ANALYSIS

To better understand results of Table V, a graphic representation was created in Figure 32.

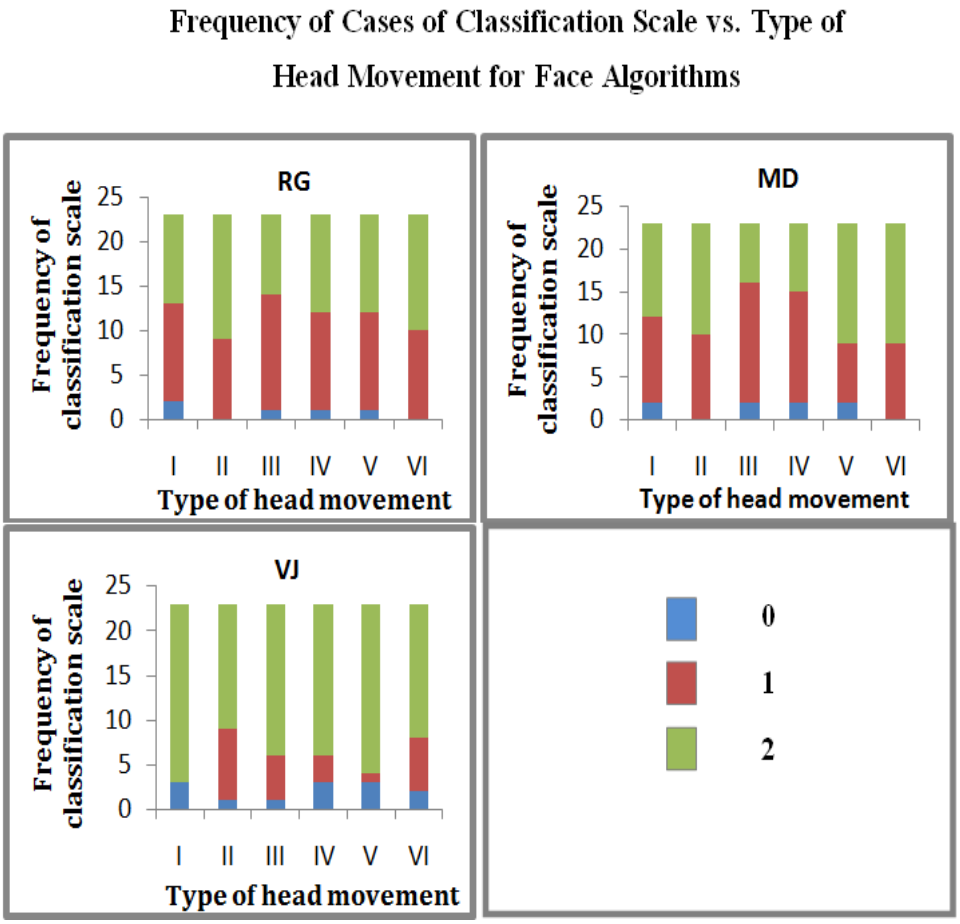

Figure 32 – Graphic representation of face algorithm results of Table V.

As shown in Table V and Figure 32, VJ method has a higher percentage of correct detections (73.91%) than the other two methods, RG (49.28%) and MD (48.55%).

RG and MD methods have a very similar percentage of correct detections because they are both based in the same face classification mechanism: skin colour properties.

When one of these methods is disturbed by illumination effects, skin-coloured objects in the background and/or subject's skin-coloured clothes, the other is also influenced. As it can be seen in Table V, the percentages of detections with errors are also very similar to both methods (RG-47.10% and MD-45.65%).

Contrary to the expectations, RG (3.62%) has a lower percentage of non-detections than MD (5.80%) and a higher detection percentage, as it was already explained. Although MD is built with pre-calculated values obtained with images of skin, as mentioned in Chapter 4, RG seems to be the best colour based algorithm for face detection and tracking.

VJ has a lower percentage of detections with errors (16.67%) than the colour based methods. However, the percentage of non detections (9.42%) is higher. This fact can result from a wrong initialization of the VJ algorithm. If the subject's face is not detected during the first two frames corresponding to the initialization process, the algorithm will not work. Fails in the initialization process could be associated with subject's movements.

This high percentage of non detections can also result from head movements. Figure 32 shows that VJ is more influenced by head movements than colour based algorithms. VJ is based on rectangular facial features; therefore, rotations of the face can affect the efficacy of this method. Colour based algorithms detect objects covered by human skin, thus when the face rotates these algorithms still work.

VJ is also influenced by illumination effects (light reflexions on face) and by hair covering the face.

Although VJ has a greater percentage of non detections, it is the best algorithm to correctly detect human faces, among those studied in this experiment. Because VJ uses a specific trained mechanism for human face detection, its results are more accurate than colour based algorithms, which can detect any skin-coloured object.

Figure 33 represents the sum of frequency of each case of the classification scale for all face algorithms, discriminated by the type of head movement mentioned in Chapter 5.

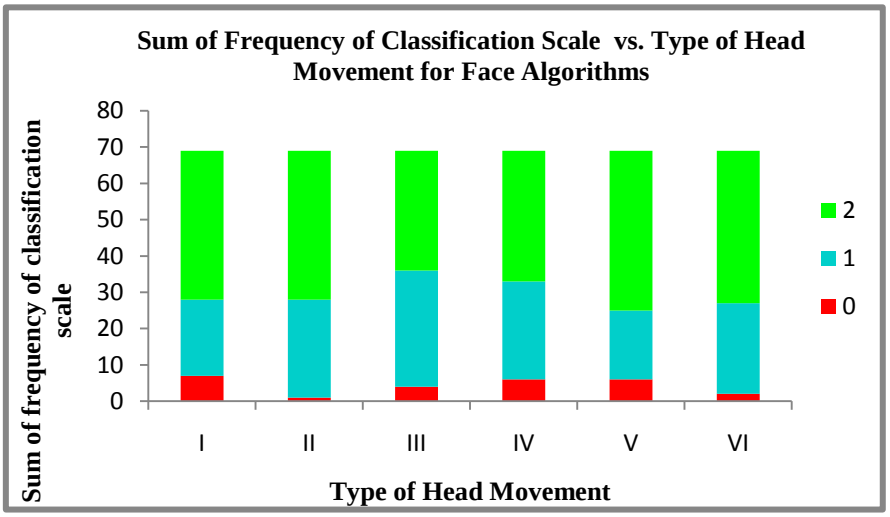

Figure 33 – Graphic representation of the relationship between efficacy of face algorithms and the type of head movement.

Figure 33 shows that the head movements I-*No movement*, III- *Movement left-up-right*, IV-*Movement up-down* and V-*Movement left-front-right* are associated with the higher percentages of non-detections.

All these face movements involve a rotation of the face, except movement I. It can be concluded that changes in face tilt seem to affect the efficacy of face algorithms.

6.2. EYE REGION ALGORITMS ANALYSIS

Figure 34 illustrates a graphic representation of results presented in Table VI.

As shown in Figure 34 and Table VI, EGF has a higher percentage of correct detections than BTE when conjugated with all face detection algorithms. The

difference of percentage of correct detections between these two methods is 33.33% with RG, 28.98% with MD and 29.71% with VJ.

In consequence, the percentage of non-detections and wrong detections is significantly reduced by using EGF instead of BTE, as it can be seen in Table VI and Figure 34.

BTE uses two principles: the special light reflexion in the middle eye region and the contrast between eyes and skin, as described in Chapter 4; therefore, illumination effects like shadows and light reflexions in other face areas decrease the efficiency of this algorithm.

For the same reason, the presence of dark regions on the face such as a beard, hair, thick eyebrows or eyeglasses can also disturb the algorithm performance.

BTE is also influenced by the head position because head movements cause changes in face illumination. BTE has an operating limit for some of the head movements described on Chapter 5. When the head reaches this limit position, BTE fails to work properly. This fact is illustrated in Figure 35.

EGF is used to isolate face structures which contrast with face itself. By choosing the most appropriate parameters, this filter can identify the eye region. Although this method is more efficient than BTE, it continues to be influenced by the existence of thick eyebrows, eyeglasses and illumination variations.

Figure 34 illustrates that EGF is less influenced by head movements than BTE; therefore, this algorithm is less sensitive to face rotations. This happens because face rotation cause illumination changes on face, and BTE is more sensitive to these changes.

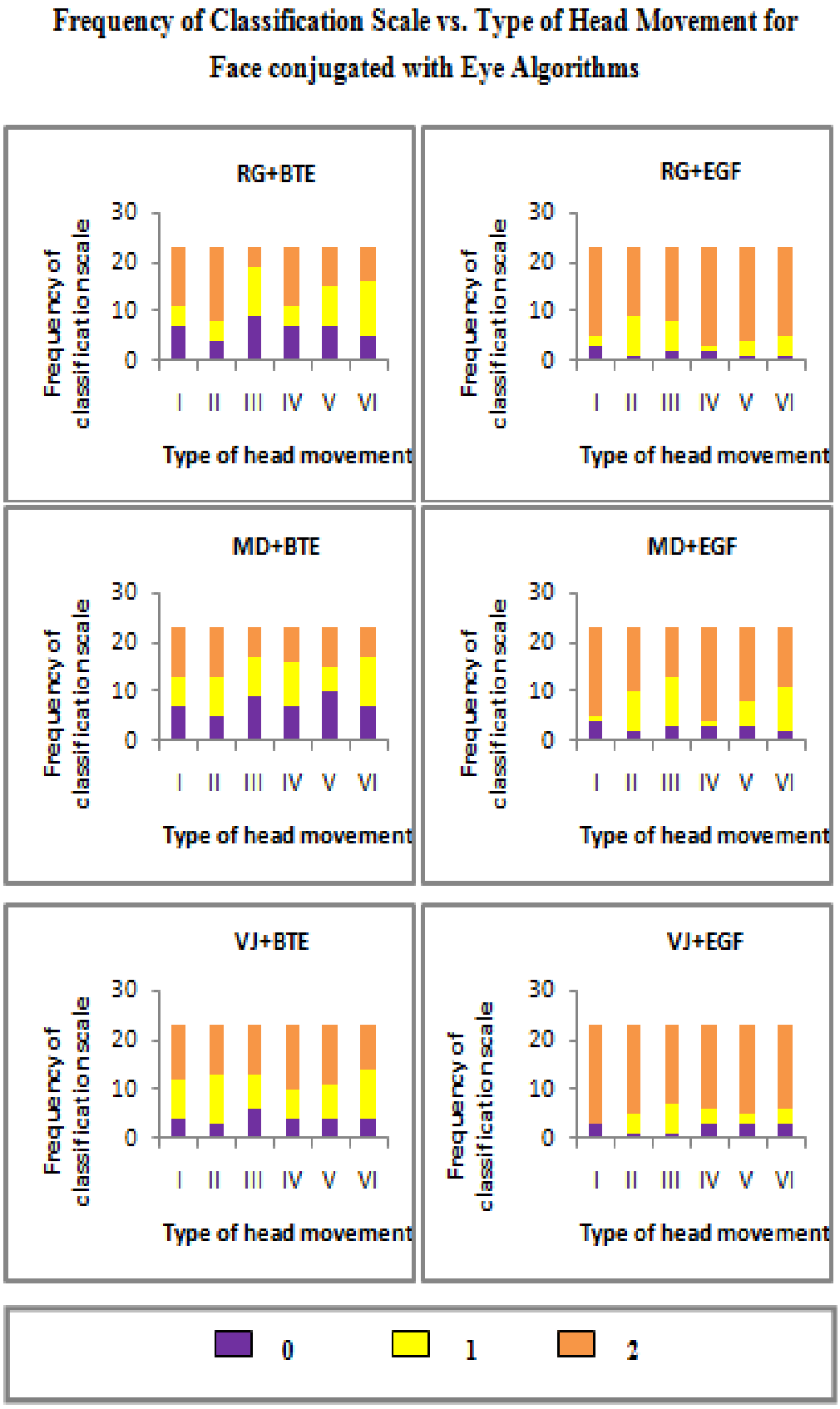

Figure 34 – Graphic representation of face and eye algorithms results of Table VI.

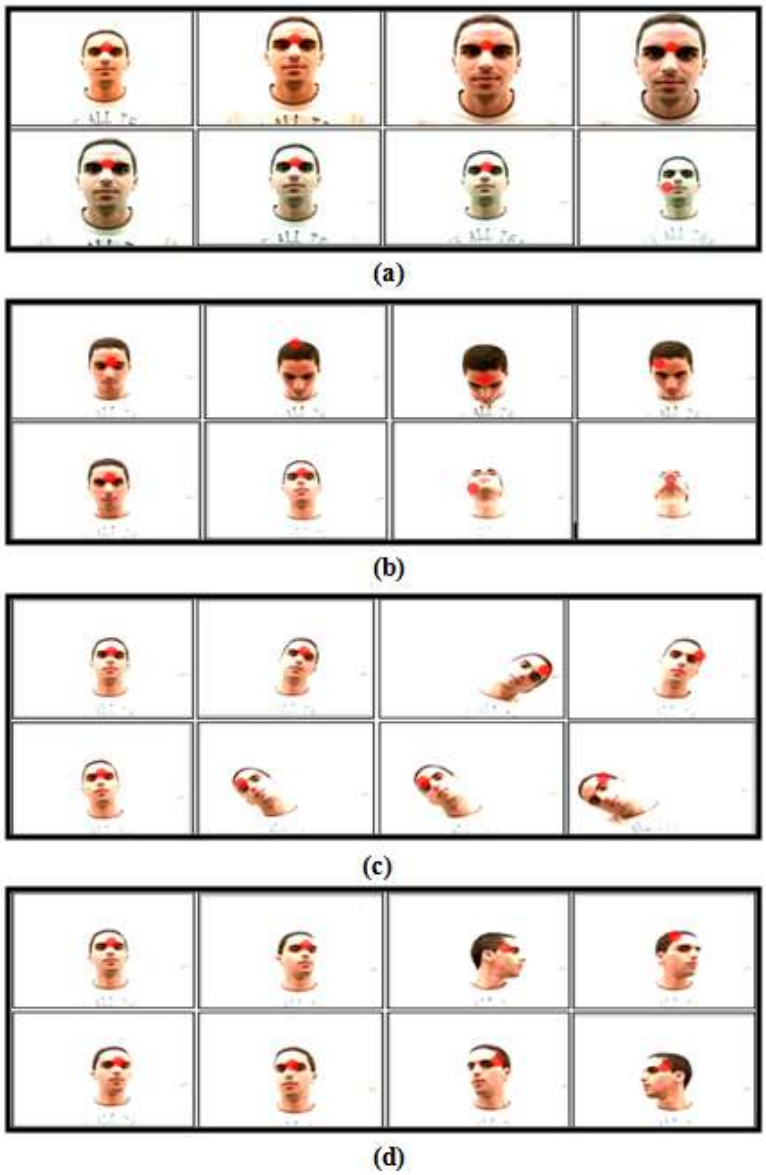

Figure 35 – Operation limits of BTE method due to some of the head movements. (a) Movement front-back; **(b)** Movement up-down; **(c)** Movement left-up-right; **(d)** Movement left-front-right.

Both of the eye region detection methods analysed above depend on the face image obtained with face algorithms. If the face is not correctly isolated, the eye region algorithms will not work properly. As it can be seen in Figure 34 and Table VI, the best results of BTE and EGF were obtained with VJ method, the best method for face detection.

Figure 36 represents the sum of frequency of each case of the classification scale for all face algorithms conjugated with eye region algorithms, discriminated by the type of head movement mentioned in Chapter 5.

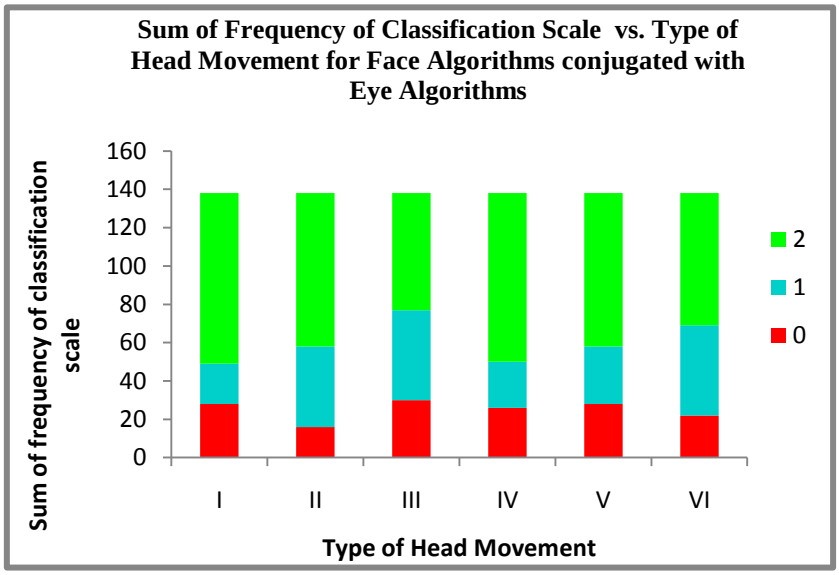

Figure 36 – Graphic representation of the relationship between the efficacy of face algorithms conjugated with eye region algorithms and the type of head movement.

Figure 36 shows that the head movements I-*No movement*, III- *Movement left-up-right*, IV-*Movement up-down* and V-*Movement left-front-right* are associated with the higher percentages of non-detections for face algorithms conjugated with eye region detection algorithms.

All these face movements involve a rotation of the face, except the movement I; therefore, changes in the face tilt seem to affect the efficacy of eye region algorithms.

6.3. PROCESSING TIME OF FACE AND EYE REGION ALGORITHMS

Results from Table VII indicate that RG is the face detection method with the highest frame rate. This method is based on a simple subtraction between the components Red and Green of the RGB colour space; therefore, it is faster than the other face detection methods which involve more complex mathematical operations and pre-processing steps.

VJ is the slowest method for face detection. The initialization phase of this algorithm during the first two frames delays all the process, because it uses a classifier cascade with a greater number of stages. After the initialization, the classifier cascade has fewer stages and feature scales; therefore, the frame rate increases.

With MD and VJ methods, the EGF has a lower frame rate than BTE. However, with RG the opposite happens. As the application of EGF requires more complex mathematical operations than BTE, it was expected that the first method would be slower, but the results with RG deny this assumption.

CHAPTER 7

CONCLUSIONS

After all the analysis performed in Chapters 5 and 6, the VJ algorithm was selected as the best face detection and tracking method. This algorithm has a percentage of correct detections of 73.91%; therefore, it is more accurate than MG and RG, the colour based algorithms studied in this experiment.

VJ has a lower frame rate and it is more influenced by head movements than colour based algorithms. Processing time is very important when implementing the system developed in a real-time application. The robustness against head movements is also essential when working with children; therefore, these limitations must be overcome in the future. This can be achieved by choosing another classifier cascade trained with a greater variety of head position images and by improving the initialization process of VJ.

EGF is the best method to segment the eye region, with a percentage of correct detections of 76.81% when implemented with VJ method.

The EGF method is less influenced by head movements than the BTE method; therefore, its implementation on children will be easier. Its frame rate (18 frames/second) is lower than in the BTE method, using the VJ algorithm. In the future, frame rate can be improved by adjusting the filter optimum parameters.

EGF is also advantageous because it gives the approximate position of the centre of the eyes, whereas the BTE method only provides the eye region. EGF facilitates the implementation of the following eye tracking steps: eye segmentation, eye tracking and gaze determination.

If EGF was only applied to straight faces this method would have better accuracy. This leads to a future in which face images can be pre-processed to achieve this purpose (Xiong et al., 2007).

To increase the frame rate of the algorithms described, a better CPU can be used in the future.

CHAPTER 8

CURRENT AND FUTURE WORK

8.1. PRELIMINARY RESULTS

8.1.1. Distance Calculation Results

Although the algorithm for the distance calculation described in Chapter 4 has already been developed, the results are still preliminary. This algorithm was tested on several different subjects and a test bench was also developed.

The test bench is composed by a rail were a skin-coloured object is inserted at a known position. During the tests, the object was moved along the rail, and the real distance marked on the rail was compared to the calculated distance. Figure 37 shows the test bench created.

The tests with subjects were performed in a similar way. The distance calculation algorithm has sensitivity in the order of centimetres. Figure 38 illustrates a graphic representation of the distance calculated during a video of 10 seconds in which the subject was still.

The distance calculated and the reference points determined with the algorithms for the eye region segmentation, BTE and EGF, are used to isolate an eye image in the video acquired by the central camera (with higher resolution). The distance in meters is again converted to pixels by using equivalent equations to (15), (16) and (17). This distance and the relationship between the resolution of the lateral cameras and the central camera are used to cut an eye image in the central camera. Figure 39 illustrates this process.

The eye image obtained by this process is used to segment the eye and to calculate the gaze point.

Some preliminary tests were executed with children participating in audiometric tests at the hospital. During the tests, three different videos were acquired with the three cameras. All the algorithms described in this work, including the distance algorithm, were applied to these videos. However, several videos could not be used because children did not collaborate with the hospital team or were all the time on incorrect positions. Figure 39 is an example of a video acquired with a child at the hospital.

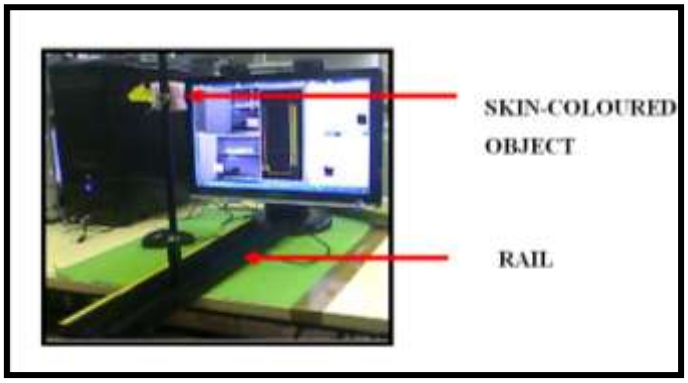

Figure 37 – Test bench created to analyse the algorithm for calculating the distance between the subject and the acquisition system.

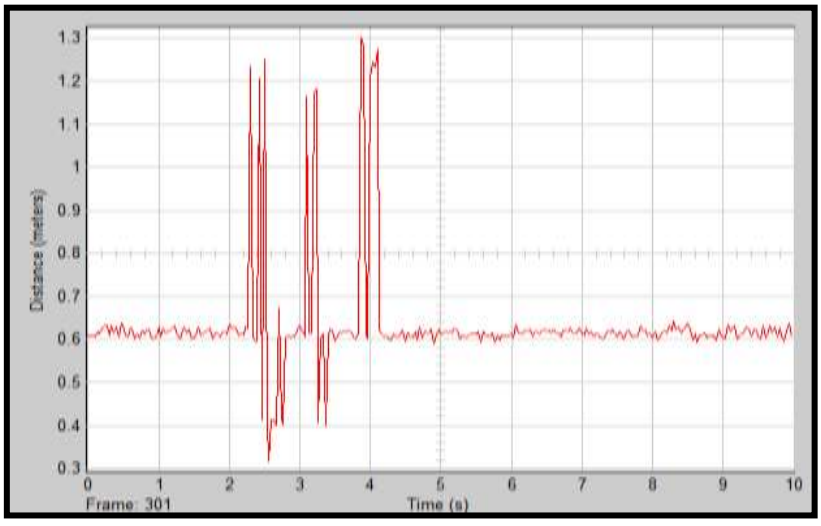

Figure 38 - Graphic representation of the calculated distance between the subject and the acquisition system in a video of 10 seconds. The subject was still during the video. In this graphic, distance errors derived from errors in the detection of the eye region by BTE can be observed.

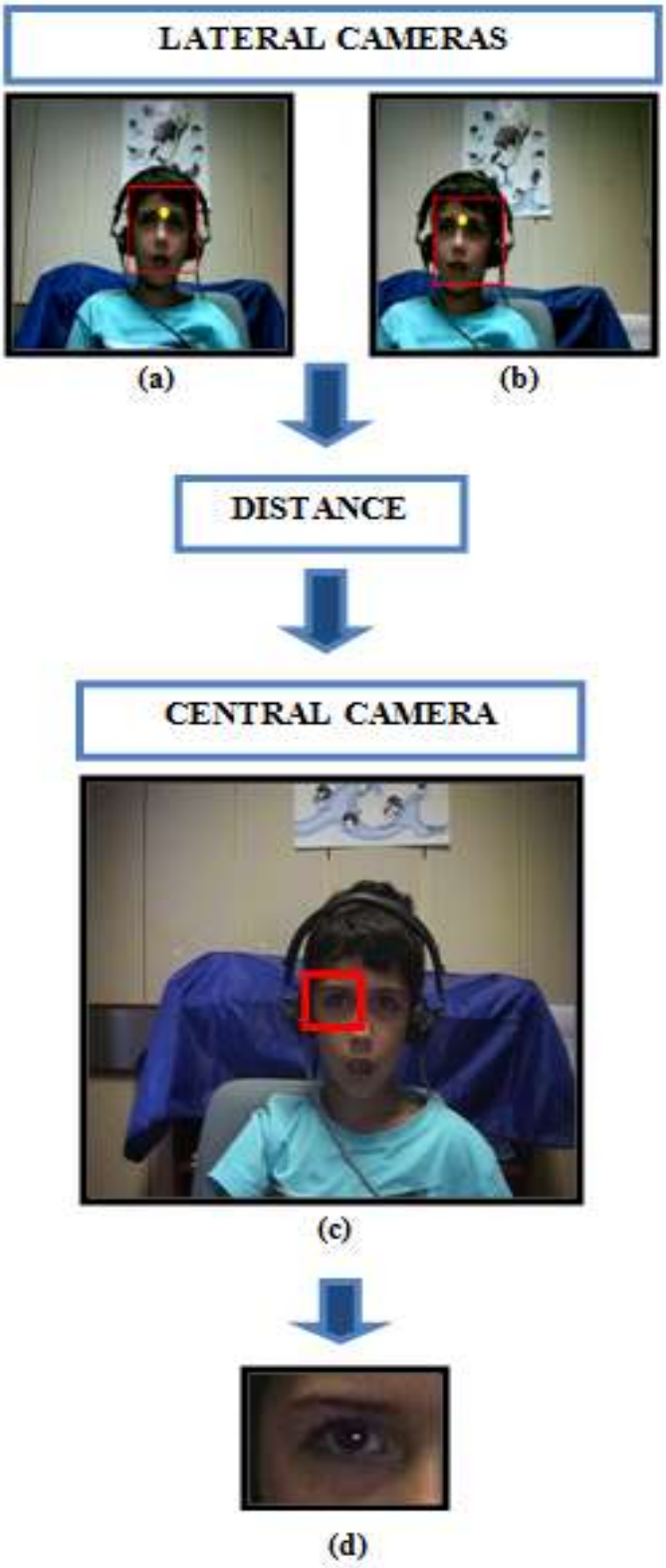

Figure 39 – Process used to obtain an eye image. (a) and **(b)** Images acquired by the lateral cameras with the face detected with VJ algorithm and the eye region segmented with BTE algorithm; **(c)** Image acquired with the central camera with the segmented eye; **(d)** Final eye image.

8.1.2. Eye Segmentation Results

This section describes a preliminary method used to segment the eye which is still being developed.

The eye image is obtained by EGF or BTE methods, as explained in the last section. The eye image has a resolution of 60×80 pixels.

To track the eye movement and to calculate the subject's gaze point, a set of eye points were defined: the reference points and the central point.

Reference points are stationary points in the eye. These points are still when the eye is moving. In this experiment the reference points are the eye corners. The central point is the centre of the iris. In turn, the gaze point can be computed as the difference between the reference and the central points.

Before the segmentation of the eye, the input eye image in the RGB format is converted to greyscale. To eliminate light reflexions, a morphological *Closing* operation with a circular structuring element with a radius equal to 3 pixels is applied. Figure 40 illustrates these changes in the original eye image.

Figure 40 – Pre-processing of eye image. (a) Original RGB eye image; **(b)** Greyscale eye image; **(c)** Eye image without reflexions obtained after the application of a morphological *Closing* operation.

8.1.2.1. Iris Segmentation

As previously mentioned in Chapter 4, Zhang et al. (2008) have developed a Circular Gabor Filter to determine the centre of the iris. This circular filter is convolved with the eye image obtained from the eye region detection.

In this experiment, an Elliptical Gabor Filter is used to detect the centre and the radius of the iris instead of the Circular Gabor Filter. This filter has different parameters than the EGF described in Chapter 4: σ_x is equal to 10 pixels, σ_y is equal to 15 pixels and F is equal to 64 Hz.

As result, the region of the iris in the output image has pixels with lower values than other eye pixels. After this, the output image is binarized and a threshold equal to 62.50% of the minimum value of the filtered image is applied to obtain the iris area. The centroid of this area is calculated and used to build a circumference around the iris. The radius of iris is also calculated.

Figure 41 is a schematic representation of images obtained with the iris segmentation process.

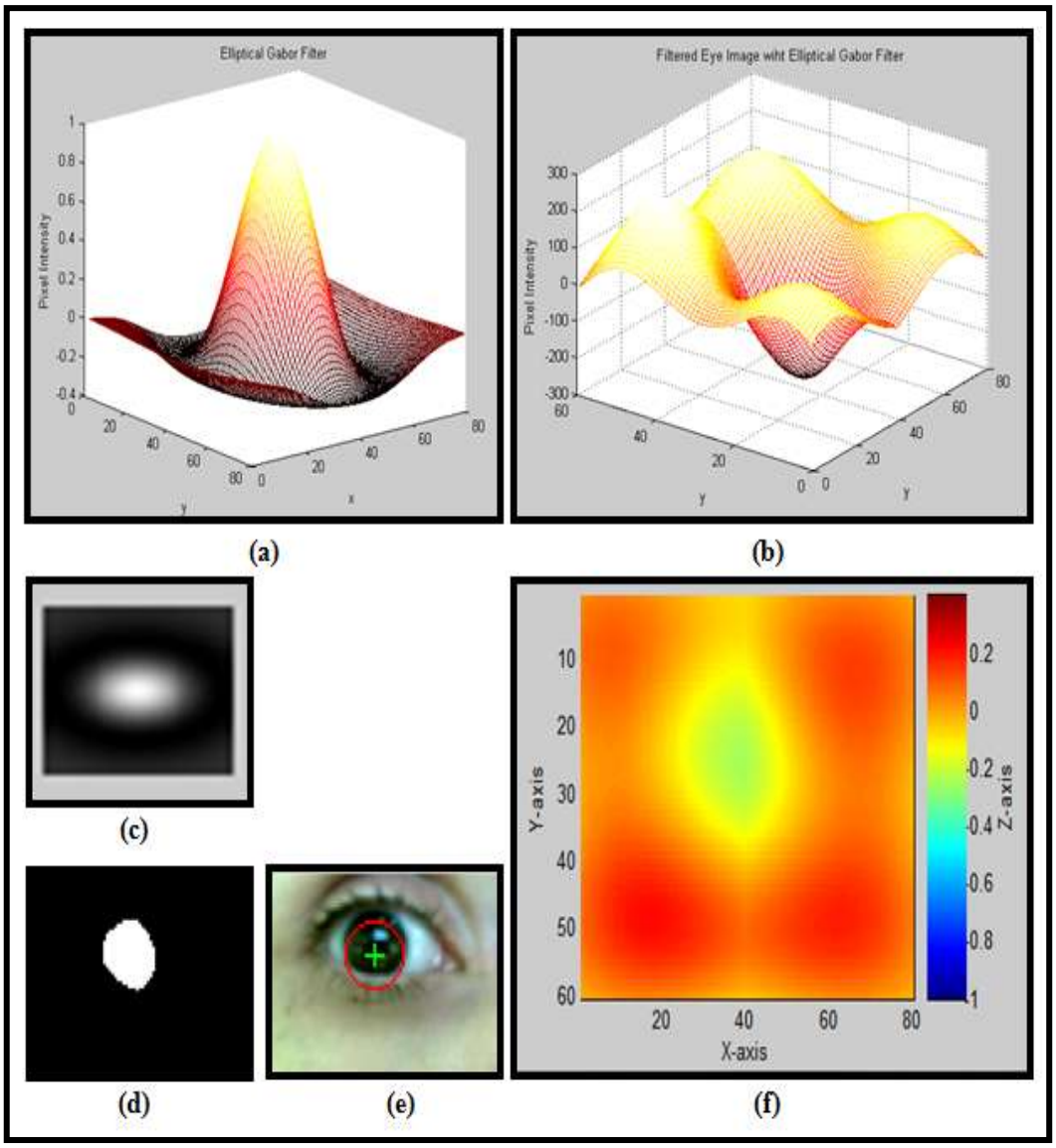

Figure 41 – Images obtained during the iris segmentation process. (a) Elliptical Gabor Filter in 3-D; **(b)** Filtered eye image in 3-D; **(c)** Elliptical Gabor Filter in 2-D; **(d)** Binary eye image obtained after the application of the threshold; **(e)** Original eye image with the iris segmented; **(f)** Filtered eye image in 2-D;

8.1.2.2. Eye Corners Segmentation

The position of eye corners is detected by using a combination of two different methods: a Circular Gabor Filter and projection functions.

Projection functions have been widely used for the segmentation of the eye (Moravčík et al., 2011; Malla, 2008; Uzunova, 2005; Zheng et al., 2005; Zhou & Geng, 2004; Feng & Yuen, 1998).

The projection functions used in this work are based on the functions created by Zhou and Geng (2004). Projection functions can give information about the boundaries of different image regions. Eye corners are located at the vertical boundary between the skin and the sclera in the region between the upper and the lower eyelids. The upper and the lower eyelids are horizontal boundaries between the skin and the iris. Projection functions are applied to detect these boundaries through the cumulative sum of pixels.

The projection functions are defined by the following equations:

1. Vertical Projection Function

$$IPF_v(x) = \frac{1}{y_2 - y_1}\left[\sum_{y_i=y_1}^{y_2} I(x, y_i)\right] \tag{20}$$

$$VPF_v(x) = \frac{1}{y_2 - y_1}\left[\sum_{y_i=y_1}^{y_2} (I(x, y_i) - IPF_v(x))^2\right] \tag{21}$$

$$GPF_v(x) = (1 - \alpha).IPF_v(x) + \alpha.VPF_v(x) \tag{22}$$

2. Horizontal Projection Function

$$IPF_h(y) = \frac{1}{x_2 - x_1}\left[\sum_{x_i=x_1}^{x_2} I(x_i, y)\right] \tag{23}$$

$$VPF_h(y) = \frac{1}{x_2 - x_1}\left[\sum_{x_i=x_1}^{x_2} (I(x_i, y) - IPF_h(y))^2\right] \quad (24)$$

$$GPF_h(y) = (1 - \alpha).IPF_h(y) + \alpha.VPF_h(y) \quad (25)$$

were *IPF* is the Integral Projection Function, *VPF* is the Variance Projection Function and *GPF* is the Generalized Projection Function, *I(x,y)* is the intensity of a pixel at the position (*x, y*) and [x_1, x_2] and [y_1, y_2] are the horizontal and vertical intervals where the projection functions are calculated, respectively. α is a constant value equal to 0.6 (Zhou & Geng, 2004).

GPF combines two different facts: the eye is darker than the skin around it (*IPF*), and the pixel intensity inside the eye area varies quickly (*VPF*) (Zhou & Geng, 2004).

After the calculation of the horizontal and vertical projection functions, GPF_h and GPF_v, the derivative is determined. The peaks of the derivative functions correspond to the location of the boundaries between the iris, the sclera and the skin. These peaks must have a minimum value equal to half of the maximum value of the derivative. To find these peaks, the values of the iris centre and the iris radius previously calculated are used as a reference. The algorithm created to find the peaks was based on the work developed by Uzunova (2005). The main principle used to find the peaks is that these peaks should be outside the iris region. Figure 42 illustrates this process.

After the vertical position of the eye corners is detected and the region between the upper and lower eyelids is determined, an Elliptical Gabor Filter is convolved with the eye image. This filter has different parameters from the filters described before: σ_x is equal to 7 pixels, σ_y is equal to 7 pixels and *F* is equal to 56 Hz.

The lower values of the filtered image are positioned at the eye fold, adjacent to the upper eyelid, and at the eye corners. The horizontal position of the eye corners is determined by choosing the minimum value of the column corresponding to the

vertical position of the eye corners previously detected, between the upper and the lower eyelids.

The application of the Elliptical Gabor Filter to detect the eye corners was based on the work of Zheng et al. (2005). They used a Gabor filter to segment eye corners because these structures, as interception points of the upper and lower eyelids, are very salient.

Figure 43 represents the application of the Elliptical Gabor Filter to detect the eye corners position.

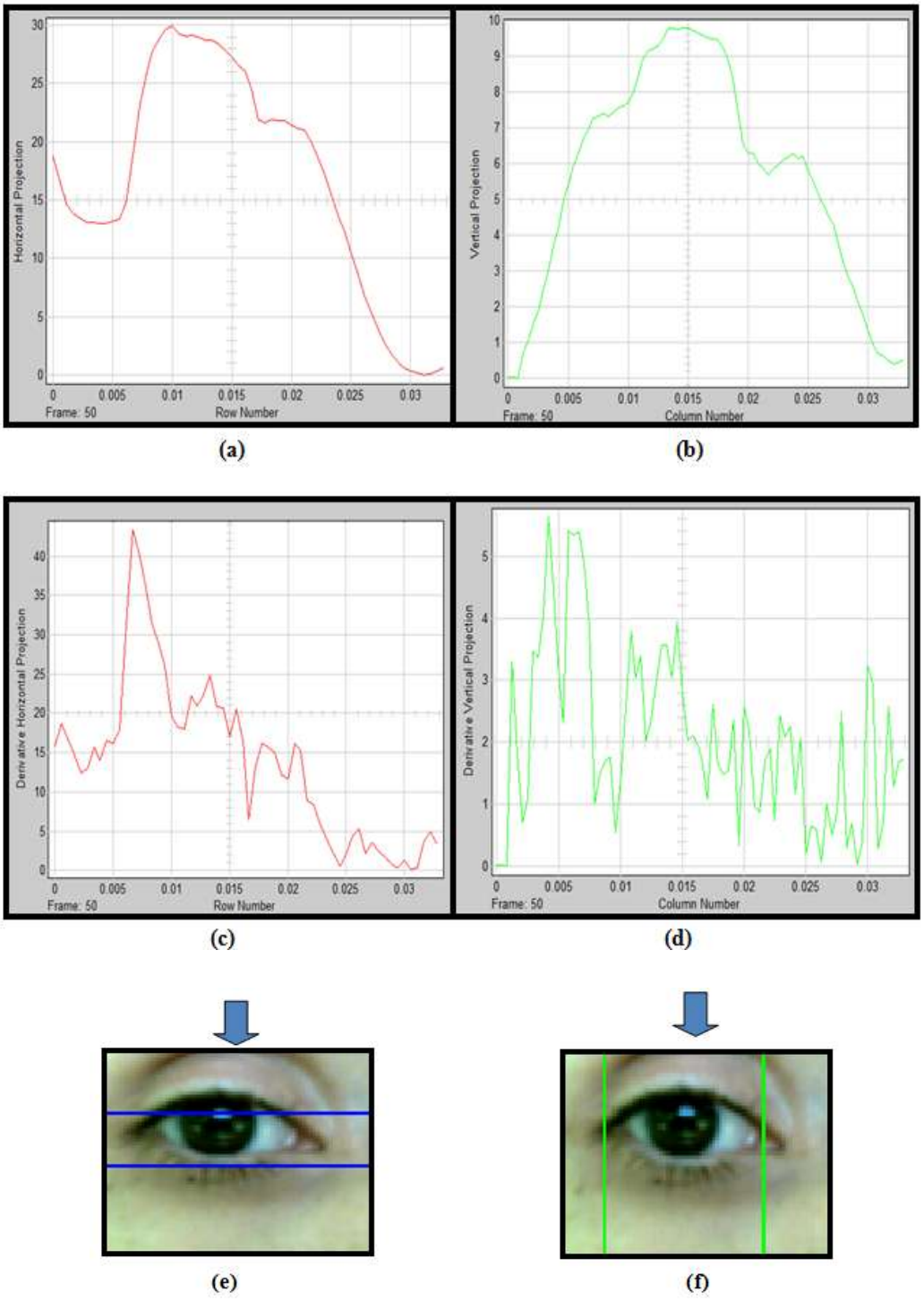

Figure 42 – First phase of eye corner segmentation – Projection Functions. (a) Horizontal projection function of eye image; **(b)** Vertical projection function of eye image; **(c)** Derivative of horizontal projection function of eye image; **(d)** Derivative of vertical projection function of eye image; **(e)** Boundaries between the skin and the iris – upper and lower eyelids; **(f)** Boundaries between the skin and the sclera – vertical position of eye corners.

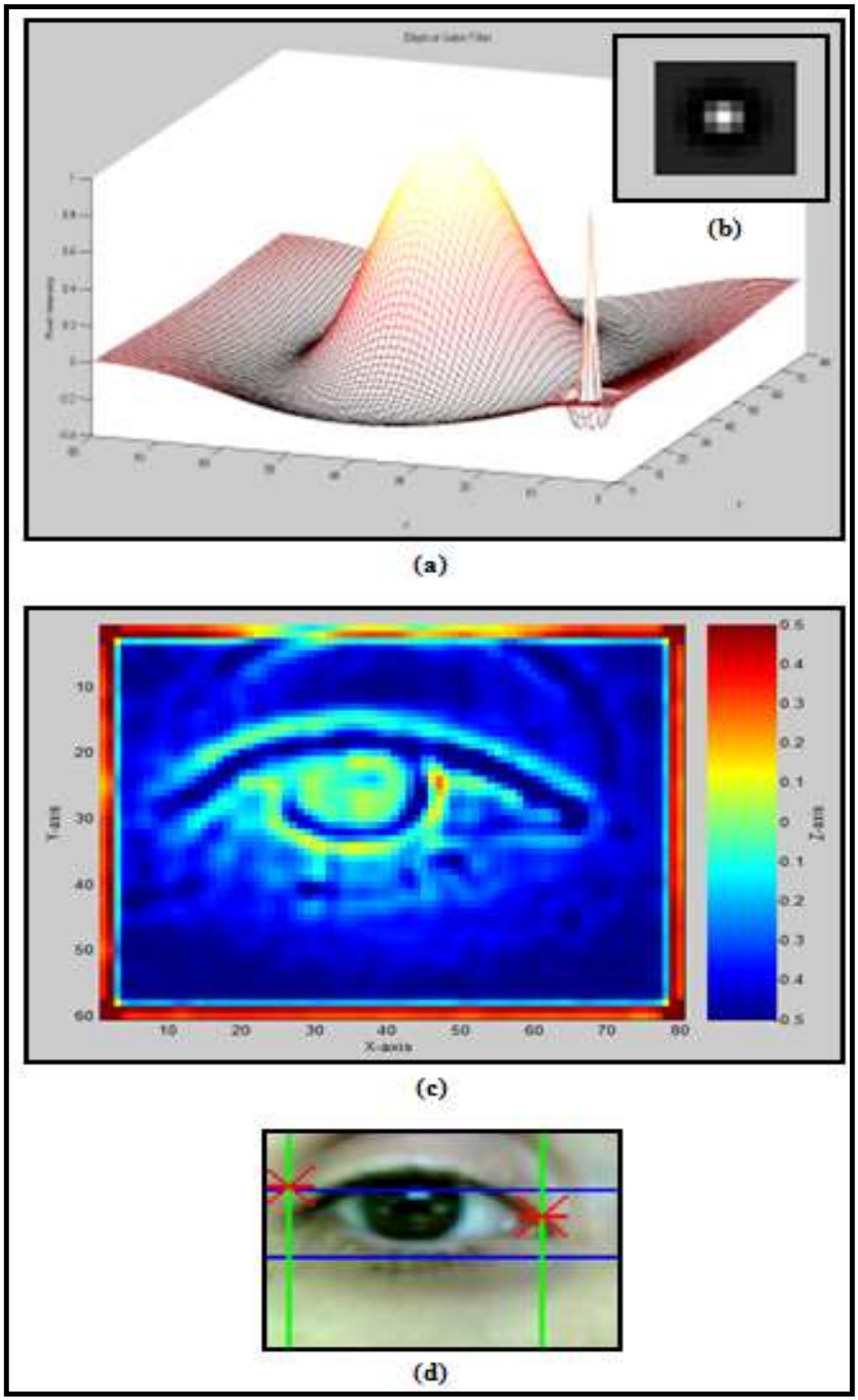

Figure 43 - Application of the Elliptical Gabor Filter to detect the eye corners position. (a) Elliptical Gabor Filter in 3-D; **(b)** Elliptical Gabor Filter in 2-D; **(c)** Filtered eye image; **(d)** Original eye image with the eye corners positions.

FINAL REMARKS

Viola-Jones was chosen as the best algorithm for face detection and tracking, and Eye Gabor Filter was considered the best method for eye region segmentation.

A distance calculation algorithm has already been developed; therefore, the next step of this work is to conclude the development of the eye segmentation method to determine the subject's gaze point.

Solutions for image processing hardware must be studied in the future to improve the performance and the temporal resolution of the system developed in this experiment.

In the future, the information about the gaze point will be related with the auditory stimuli of the audiometric tests at the hospital. The electrode's balancing of cochlear implants will be improved.

In this way, this work will provide a complete remote eye tracking system to assist the rehabilitation of patients with cochlear implants.

One advantage of the future eye tracking system will be the robustness to head movements, which is very important when working with children. This system will also have a much lower price than other similar technologies on the market.

REFERENCES

All the non-referenced figures in this work were created by the author;

Ahlberg, J. (1999). A system for face localization and facial feature extraction. *CiteSeerX, Tech. Rep. LiTH-ISY-R-2172*;

Al-Shehri, S. A. (2004). A simple and novel method for skin detection and face locating and tracking. *Springer, Computer Human Interaction, Lecture Notes in Computer Science, 3101*, 1-8;

ASHA. (2004). Technical report: Cochlear implants. Working group on cochlear implants. *ASHA Supplement 24, in press*;

BC Children's Hospital. (2010). *How to read an audiogram*. Consulted in 23 of August of 2011, in BC Children's Hospital: http://www.bcchildrens.ca/Services/ClinicalDiagnosticFamilyServices/Audiology/Forfamilies/readaudiogram.htm;

Bejjanie, B.-P. et al. (2002). Concurrent excitatory and inhibitory effects of high frequency stimulation: an oculomotor study. *J Neurol Neurosurg Psychiatry, 72*, 517-522;

Briggs, D. (2007). *The dimensions of colour*. Consulted in 15 of August of 2011, in Ibiblio: http://www.huevaluechroma.com/093.php;

Bucknall, J. M. (2010). *How face detection works: In Depth: You can do it, and your camera can do it - but how?*. Consulted in 10 of August of 2011, in TechRadar: http://www.techradar.com/news/software/applications/how-face-detection-works-703173;

Chellappa, R., Sinha, P. & Phillips, P. J. (2010). Face recognition by computers and humans. *IEEE Computer Society*;

Daugman, J. G. (1993). High confidence visual recognition of persons by a test of statistical independence. *IEEE, Transactions on Pattern Analysis and Machine Intelligence, 15,* No. 11;

De Maesschalck, R., Jouan-Rimbaud, D. & Massart, D. L. (2000). The Mahalanobis distance. *Elsevier Science B.V., Chemometrics and Intelligent Laboratory Systems*, *50*, 1-18;

Defeat Diabetes Foundation. (2011). *Anatomy of an eye*. Consulted in 23 of August of 2011, in Defeat Diabetes Foundation, inc.: http://www.defeatdiabetes.org/about_diabetes/text.asp?id=Anatomy_of_the_eye;

Duchowski, A. T. (2002). A breadth-first survey of eye tracking applications. *Behavior Research, Methods, Instruments, and Computers*;

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C. & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, *4*, No. 6;

Feng, G. & Yuen, P. (1998). Variance projection function and its application to eye detection for human face recognition. *Elsevier Science B.V., Pattern Recognition Letters*, *19*, 899-906;

Frens, M. A., Van Opstal, A. J. & Van der Willigen, R., F. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Psychonomic Society, Inc*., *Perception & Psychophysics*, *57*, 6, 802-816;

Grauman, K., Betke, M., Gips, J. & Bradski, G. R. (2001). Communication via eye blinks: detection and duration analysis in real time. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, *1*, pp.1010-1017;

Hallinan, P. W. (1991). Recognizing human eyes. *Geometric Methods in Computer Vision*, pp. 212-226, SPIE;

Hansen, D. W. & Hansen, J. P. (2006). Robustifying eye interaction. *Proc. Conf. Vision for Human Computer Interaction*, pp. 152-158;

Hansen, D. W. & Ji, Q. (2010). In the eye of the beholder: a survey of models for eyes and gaze. *IEEE, Transactions on Pattern Analysis and Machine Intelligence*, *32*, No. 3;

Harris, L. R., Blakemore, C & Donaghy, M. (1980). Integration of visual and auditory space in the mammalian superior colliculus. *Macmillan Journals Ltd.*, *Nature*, *288*, No. 5786, pp. 56-59;

Harvey, A. (2010). *Visualizing Haar cascade files*. Consulted in 12 of August of 2011, in AH Projects: http://ahprojects.com/blog/c/itp;

Huang, J. & Wechsler, H. (1999). Eye detection using optimal Wavelet Packets and Radial Basis Functions (RBFs). *Int'l J. Pattern Recognition and Artificial Intelligence*, *13*, No.7;

Huang, J., Ji, D., Shao, X. & Wechsler, H. (1998). Pose discrimination and eye detection using Support Vector Machines (SVMs). *Proc. Conf. NATO-ASI on Face Recognition: From Theory to Applications*, pp. 528-536;

Huang, W. M. & Mariani, R. (2000). Face detection and precise eyes location. *Proc. Int'l Conf. Pattern Recognition*;

Ishikawa, T., Baker, S., Matthews, I. & Kanade, T. (2004). Passive driver gaze tracking with Active Appearance Models. *Proc. 11th World Congress Intelligent Transportation Systems*;

Jacob, R. J. & Karn, K. S. (2003). Eye tracking in Human-Computer Interaction and usability research: ready to deliver the promises. *Elsevier Science BV, The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research, Hyona, Radach & Deubel (eds.)*;

Jacob, R. J. K. (1994). Eye tracking in advanced interface design. *Oxford University Press, Advanced Interface Design and Virtual Environments, ed. W. Bartfield and T. Fumess*;

Jeng, S.-H., MarkLiao, H. Y., Han, C. C., Chern, M. Y. & Liu, Y. (1998). Facial feature detection using geometrical face model: an efficient approach. *Pattern Recognition*, *31*, 273-282;

Ji, Q. & Yang, X. (2002). Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. *Real-Time Imaging, 8*, No. 5, pp. 357-377;

Kass, M., Witkin, A. & Terzopoulos, D. (1988). Snakes: active contour models. *International Journal Comuter Vision*, 321-331;

Kawato, S. & Tetsutani, N. (2002). Detection and tracking of eyes for gaze-camera control. *Elsiever B.V., Image and Vision Computing, Proceedings from the 15th International Conference on Vision Interface, 22*, Issue 12, pp. 1031-1038;

KidsHealth. (2011). *Cochlear Implants*. Consulted in 10 of February of 2011, in The Nemours Foundation: http://kidshealth.org/parent/general/eyes/cochlear.html;

Knudsen, E. I., Knudsen, P. F. & Masino, T. (1993). Parallel pathways mediating both sound localization and gaze control in the forebrain and midbrain of the barn owl. *The Journal of Neuroscience, 13*, 7, 2837-2852;

Kumar, M. (2006). Reducing the Cost of Eye Tracking Systems. *Building, 4. Citeseer*;

Lam, K. & Yan, H. (1996). Locating and extracting the eye in human face images. *Pattern Recognition, 29*, pp. 771-779;

Laurens, A. D. (2002). *Eye Movements. A Discourse of the Preservation of Sight: of Melancholike Diseases; of Rheumes, and of Old Age***.** Consulted in 28 of August of 2011: http://www.psy.vanderbilt.edu/courses/psy216/eyemove_physio.htm;

Liversedge, S. P. & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Elsevier Science Ltd, Trends in Cognitive Sciences, 4*, No.1;

Loy, G. & Zelinsky, A. (2003). Fast radial symmetry for detecting points of interest. *IEEE Trans. Pattern Analysis and Machine Intelligence, 25*, No. 8, pp. 959-973;

Malla, A. M. (2008). Automated video-based measurement of eye closure using a remote camera for detecting drowsiness and behavioural microsleeps.

University of Canterbury, Electrical and Computer Engineering, Theses and Dissertations;

MathWorks®. (2011). *Video and Image Processing Toolbox*. Consulted in July-August of 2011, in The MathWorks®, Inc.: http://www.mathworks.com/help/toolbox/images/;

Moravčík, T., Bubeníková E. & Muzikářová, L'. (2011). Detection of determined eye features in digital image. *Annals of Faculty Engineering Hunedoara – International Journal of Engineering*, Fascicule 1;

Namahn. (2001). *Using eye tracking for usability testing*. Consulted in 28 of August of 2011, in Namahn bvba: www.namahn.com;

Nielsen, J. B. (2010). Face detection and recognition in video-streams. *IMM-B.Sc*;

Peng, K., Chen, L., Ruan, S. & Kukharev, G. (2005). A robust algorithm for eye detection on gray intensity face without spectacles. *JCS&T, 5*, No.3;

Pentland, A., Moghaddam, B. & Starner, T. (1994). View-based and modular eigenspaces for face recognition. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*;

Pocock, G. & Richards, C. D. (2006). *Fisiologia Humana - A Base da Medicina* (2ª ed.). Guanabara Koogan (Grupo GEN);

Poole, A. & Ball, L. J. (2005). Eye tracking in Human-Computer Interaction and usability research: Current status and future prospects. *Idea Group, Inc, Prospects, Chapter in C. Ghaoui (Ed.), Encyclopedia of Human-Computer Interaction*;

Porter, G. T., Gadre, A. K., Quinn, F. B. & Ryan, M. W. (2003). Cochlear Implants. *Grand Rounds Presentation, UTMB, Dept. of Otolaryngology*;

Ramezanpour, M., Azimi, M. A. & Rahmati, M. (2010). A new method for eye detection in color images. *Journal of Advances in Computer Research, 2*, 55-61;

Reisfeld, D. & Yeshurun, Y. (1992). Robust detection of facial features by generalized symmetry. *Proc. Int'l Conf. Pattern Recognition*, *I*, pp. 117-120;

Ryu, Y.-S. & Oh, S.-Y. (2001). Automatic extraction of eye and mouth field from a face image using eigenfeatures and multilayer perceptrons. *Pattern Recognition, 34*, 2459-2466;

Shih, F. Y. & Chuang, C.-F. (2004). Automatic extraction of head and face boundaries and facial features. *Elsevier, Information Sciences, An International Journal, 158*, 117-130;

Sirohey, S., Rosenfeld, A. & Duric, Z. (2002). A method of detecting and tracking irises and eyelids in video. *Pattern Recognition*, *35*, No.6, pp. 1389-1401;

The McGraw-Hill Companies, inc. (2011). *Sense of Hearing*. *External, Middle, and Inner Ear.* Consulted in 23 of August of 2011, in McGraw-Hill Higher Education: http://www.mhhe.com/biosci/esp/2001_saladin/folder_structure/in/m4/s5/inm4s5_1.htm;

Uzunova, V. I. (2005). An eyelids and eye corners detection and tracking method for rapid iris tracking. *Technical University of Denmark (DTU);*

Van De Graaff, K. M. (2002). *Human anatomy* (6th ed.). New York: McGraw-Hill Higher Education;

Viola, P. & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *IEEE, Computer Society Conference on Computer Vision and Pattern Recognition, 1*, pp.511;

Winfield, D., Li, D., Babcock, J. & Parkhurst, D. J. (2005). Towards an open-hardware open-software toolkit for robust low-cost eye tracking in HCI applications. *IEEE, Computer Society Conference on Computer Vision and Pattern Recognition*;

Witzner Hansen, D., Hansen, J. P., Nielsen, M., Johansen, A. S. & Stegmann, M. B. (2003). Eye typing using Markov and Active Appearance Models. *Proc. IEEE Workshop Applications on Computer Vision*, pp. 132-136;

Wu, J., Abdulla, H. M. D. & Snášel, V. (2009). Application of binocular vision and diamond search technology in computer vision navigation. *IEEE, International Conference on Intelligent Networking and Collaborative Systems*, 87 – 92;

Xie, X., Sudhakar, R. & Zhuang, H. (1994). On improving eye feature-extraction using deformable templates. *Pattern Recognition*, *27*, No. 6, pp. 791-799;

Xiong, F., Zhang, Y. & Zhang, G. (2007). A pupil localization algorithm based on adaptive Gabor filtering and negative radial symmetry. *Springer, LNCS 4679*, pp. 87-96;

Yee, E. & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology*: *Leaning, Memory, and Cognition*, *32*, No.1, 1-14;

Yen, G. G. & Nithianandan, N. (2002). Facial feature extraction using genetic algorithm. *IEEE*;

Young, L. R. & Sheena, D. (1975). Methods & Designs: Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation*, *7*, 5, 397-429;

Yuille, A., Hallinan, P. & Cohen, D. (1992). Feature extraction from faces using deformable templates. *Int'l J. Computer Vision*, *8*, No.2, pp. 99-111;

Zambarbieri, D. Beltrami, G. & Versino, M. (1995). Saccade latency toward auditory targets depends on the relative position of sound source with respect to the eyes. *Elsevier Science Ltd., Vision Res.*, 35, No. 23/24, pp. 3305-3312;

Zhang, Y., Sun, N., Gao, Y. & Cao, M. (2008). A new eye location method based on Ring Gabor Filter. *Proceedings of the IEEE, International Conference on Automatic and Logistics, 978*, 1, 4244-2503;

Zhao, W., Chellappa, R., Phillips, P. J. & Rosenfeld, A. (2003). Face Recognition: A Literature Survey. *ACM Computing Surveys, 35*, No. 4, pp. 399-458;

Zheng, Z., Yang, J. & Yang, L. (2005). A robust method for eye features extraction on color image. *Elsiever B.V., Pattern Recognition Letters, 26*, 2252-2261;

Zhou, Z.-H., Geng, X. (2004). Projection functions for eye detection. *Elsevier Science B.V., Pattern Recognition, 37*, Issue 5, 1049-1056;

Zhu, Z., Ji, Q. & Fujimura, K. (2002). Combinig Kalman filtering and mean shift for real time eye tracking. *Proc. Int'l Conf. Pattern Recognition, IV*, pp. 318-321;

Zwolan, T. (2011). Cochlear Implants. *American Speech-Language-Hearing Association, Audiology Information Series,* 7976-Y5.

APPENDIX A

Table VIII – Results obtained with face and eye algorithms for all subjects. FR (frames/second) is the mean of frame rate for a video with 300 frames.

SUBJECT	HEAD MOVEMENT	RG					MD					VJ				
		RG	BTE	FR	EGF	FR	MD	BTE	FR	EGF	FR	VJ	BTE	FR	EGF	FR
1	I	2	1	43	2	58	2	1	35	2	35	2	2	24	2	20
	II	2	2	44	1	57	2	0	36	1	35	1	1	24	2	17
	III	2	1	42	2	61	1	0	35	1	35	2	0	29	2	19
	IV	2	2	42	2	59	1	0	35	2	33	2	1	26	2	21
	V	2	1	44	2	58	2	1	36	2	34	2	1	23	2	18
	VI	2	1	40	2	55	2	1	37	2	32	2	1	26	2	20
2	I	1	1	44	2	56	1	1	38	1	34	2	2	24	2	18
	II	1	0	43	2	56	1	0	37	2	33	1	2	24	2	18
	III	1	0	42	1	59	1	0	37	1	33	2	1	24	2	21

SUBJECT	HEAD MOVEMENT	RG					MD					VJ				
		RG	BTE	FR	EGF	FR	MD	BTE	FR	EGF	FR	VJ	BTE	FR	EGF	FR
	IV	1	0	43	2	59	1	0	38	2	32	1	2	29	2	20
	V	1	0	44	2	57	1	0	35	1	33	2	0	24	1	17
	VI	1	0	43	2	64	1	0	35	1	33	1	1	23	2	19
3	I	2	2	44	2	58	2	2	38	2	34	2	2	23	2	17
	II	2	2	43	1	62	1	2	37	2	32	1	2	25	2	18
	III	2	2	42	2	62	1	1	39	1	34	2	2	23	2	18
	IV	2	2	45	2	64	2	1	37	2	34	2	2	23	2	17
	V	2	1	44	2	65	2	1	37	2	33	2	2	28	2	17
	VI	2	2	45	2	64	2	2	37	1	33	2	2	23	2	20
4	I	1	0	43	2	58	1	0	37	2	34	2	1	27	2	19
	II	2	1	40	2	65	2	0	37	2	36	1	1	26	2	18
	III	2	0	43	1	60	2	0	37	1	36	2	1	25	2	19
	IV	2	0	43	2	61	2	0	37	2	36	2	2	26	2	19

SUBJECT	HEAD MOVEMENT	RG					MD					VJ				
		RG	BTE	FR	EGF	FR	MD	BTE	FR	EGF	FR	VJ	BTE	FR	EGF	FR
	V	2	0	41	1	66	2	0	38	1	34	0	0	0	0	0
	VI	2	0	41	2	64	1	0	39	1	35	2	2	29	2	18
5	I	2	2	43	1	60	1	1	37	0	34	2	1	26	2	18
	II	1	2	45	1	60	2	2	38	0	33	2	2	26	2	18
	III	2	1	44	2	59	1	2	38	1	35	2	2	26	2	16
	IV	1	2	44	2	60	1	2	37	2	35	2	2	26	2	18
	V	2	2	42	2	64	2	2	37	2	35	1	2	28	2	18
	VI	2	2	40	2	66	2	1	37	2	35	2	1	25	2	19
6	I	2	2	43	2	66	2	2	36	2	35	2	1	23	2	19
	II	2	2	42	2	64	2	2	37	2	34	2	2	24	2	18
	III	2	1	42	2	59	2	0	38	1	35	2	1	28	2	20
	IV	2	2	41	2	60	2	1	37	2	34	2	2	26	2	18
	V	2	1	40	2	61	2	0	37	2	33	2	1	25	2	19

SUBJECT	HEAD MOVEMENT	RG					MD					VJ				
		RG	BTE	FR	EGF	FR	MD	BTE	FR	EGF	FR	VJ	BTE	FR	EGF	FR
	VI	2	1	41	2	65	2	1	38	2	34	2	1	25	2	18
7	I	2	2	42	2	61	2	2	37	2	34	2	2	24	2	17
	II	2	2	44	2	61	2	2	37	2	35	2	1	22	2	16
	III	1	2	43	2	66	2	2	39	2	35	2	1	23	2	18
	IV	1	2	42	2	61	2	2	36	2	35	2	2	22	2	18
	V	2	2	41	2	60	2	2	38	2	35	2	2	24	2	17
	VI	2	2	40	2	56	2	2	38	1	34	2	2	24	2	18
8	I	1	1	42	2	67	1	2	37	2	35	2	2	27	2	18
	II	2	2	42	2	63	2	2	37	2	34	1	2	23	2	17
	III	2	2	42	2	64	1	2	35	2	34	1	2	20	1	18
	IV	2	2	42	2	67	1	2	37	2	31	2	2	24	2	17
	V	2	2	44	2	59	1	2	37	1	33	2	2	27	2	16
	VI	1	2	44	1	64	1	1	37	1	35	2	2	22	2	18

SUBJECT	HEAD MOVEMENT	RG					MD					VJ				
		RG	BTE	FR	EGF	FR	MD	BTE	FR	EGF	FR	VJ	BTE	FR	EGF	FR
9	I	2	2	42	2	65	2	2	37	2	34	2	2	26	2	16
	II	2	2	43	2	58	2	2	38	2	33	1	2	20	1	17
	III	2	2	42	2	60	2	2	37	2	34	2	2	27	2	17
	IV	2	2	41	2	63	2	2	37	2	34	2	1	24	2	19
	V	2	2	42	2	58	2	2	37	2	33	2	2	21	2	18
	VI	1	1	42	1	66	1	1	36	1	34	1	2	24	2	18
10	I	1	1	41	2	57	1	1	37	2	33	2	1	22	2	18
	II	2	2	42	1	65	2	2	37	1	35	1	0	20	1	18
	III	1	1	43	1	59	1	2	34	2	35	2	2	28	2	17
	IV	1	2	43	2	59	1	2	38	2	35	2	2	24	2	19
	V	1	2	42	2	62	2	2	39	2	34	2	1	22	2	18
	VI	2	2	42	2	60	1	2	38	2	35	2	1	22	2	17
11	I	2	2	42	2	62	2	2	37	2	34	2	2	29	2	19

SUBJECT	HEAD MOVEMENT	RG					MD					VJ				
		RG	BTE	FR	EGF	FR	MD	BTE	FR	EGF	FR	VJ	BTE	FR	EGF	FR
	II	2	1	42	1	61	2	1	36	1	34	2	1	23	2	20
	III	2	1	42	2	68	2	1	38	2	34	2	2	22	2	20
	IV	1	1	43	2	63	2	1	39	2	34	1	1	22	2	21
	V	2	2	43	2	62	2	2	37	2	34	2	2	25	2	18
	VI	2	1	41	2	63	2	1	37	2	35	1	1	25	1	19
12	I	1	2	43	2	57	1	2	37	2	35	2	2	21	2	19
	II	1	2	43	1	62	1	1	37	1	35	2	1	21	2	19
	III	1	1	42	2	65	1	1	38	2	32	2	2	18	2	17
	IV	2	2	45	2	67	1	1	38	2	35	2	1	22	2	20
	V	1	1	42	2	63	1	1	36	2	35	2	2	24	2	18
	VI	1	1	43	2	65	1	1	37	2	35	2	1	23	2	18
13	I	1	0	43	0	61	2	1	38	0	34	2	0	27	2	16
	II	1	0	48	0	59	1	1	37	0	34	1	0	20	1	17

SUBJECT	HEAD MOVEMENT	RG					MD					VJ				
		RG	BTE	FR	EGF	FR	MD	BTE	FR	EGF	FR	VJ	BTE	FR	EGF	FR
	III	1	0	42	0	65	1	0	37	0	34	2	0	21	1	17
	IV	1	0	43	0	61	1	0	38	0	34	2	1	21	1	18
	V	1	0	45	1	58	1	0	37	0	35	2	1	27	2	17
	VI	2	1	41	0	62	2	1	37	0	34	2	0	20	1	18
14	I	1	2	45	2	69	1	1	38	2	34	2	2	22	2	18
	II	2	2	44	2	66	2	1	37	2	35	2	2	21	2	16
	III	2	0	43	1	61	2	1	38	1	35	1	1	21	2	16
	IV	2	1	42	2	66	1	1	37	2	35	2	2	28	2	18
	V	1	1	43	2	59	1	0	38	2	35	2	2	27	2	18
	VI	1	1	46	2	64	2	1	37	2	34	2	2	20	2	17
15	I	2	0	59	2	61	2	0	37	2	34	2	1	22	2	17
	II	1	0	42	2	62	1	0	37	1	35	2	1	24	2	18
	III	1	0	48	2	66	1	1	38	2	35	1	0	21	2	19

SUBJECT	HEAD MOVEMENT	RG					MD					VJ				
		RG	BTE	FR	EGF	FR	MD	BTE	FR	EGF	FR	VJ	BTE	FR	EGF	FR
	IV	1	0	45	2	61	1	0	37	2	35	0	0	0	0	0
	V	1	0	45	2	60	1	0	36	2	36	2	1	24	2	17
	VI	1	1	47	2	61	2	0	38	1	34	0	0	0	0	0
16	I	1	2	42	2	60	1	2	36	2	35	2	1	23	2	18
	II	2	2	42	2	56	1	2	37	2	35	2	2	24	2	18
	III	1	1	42	2	66	1	2	38	2	34	2	2	22	2	18
	IV	2	2	42	2	60	1	2	38	2	34	2	2	24	2	17
	V	2	1	42	2	65	2	2	36	1	35	2	2	20	2	16
	VI	2	2	45	2	58	2	2	38	2	34	2	2	20	2	16
17	I	0	0	46	0	62	0	0	36	0	34	0	0	0	0	0
	II	1	1	42	1	60	1	1	38	1	38	0	0	0	0	0
	III	0	0	44	0	61	0	0	38	0	34	0	0	0	0	0
	IV	0	0	42	0	60	0	0	38	0	34	0	0	0	0	0

SUBJECT	HEAD MOVEMENT	RG					MD					VJ				
		RG	BTE	FR	EGF	FR	MD	BTE	FR	EGF	FR	VJ	BTE	FR	EGF	FR
	V	0	0	44	0	64	0	0	38	0	34	0	0	0	0	0
	VI	2	0	42	1	63	1	0	37	0	35	0	0	0	0	0
18	I	2	2	43	2	61	2	2	36	2	34	2	2	23	2	16
	II	2	2	44	2	62	2	2	37	2	35	2	1	20	2	16
	III	1	1	42	2	68	1	1	38	2	35	1	1	22	1	19
	IV	2	1	43	2	64	2	1	37	2	34	2	2	23	2	18
	V	2	2	43	2	65	2	2	37	2	34	2	2	20	2	18
	VI	2	1	44	2	66	2	1	38	2	34	1	1	23	2	17
19	I	1	2	69	2	67	1	0	39	2	34	2	2	22	2	17
	II	2	2	46	2	61	1	1	37	2	34	2	2	24	2	18
	III	1	0	46	2	60	1	1	37	2	35	2	2	23	1	18
	IV	1	1	62	2	66	1	1	38	2	35	2	2	22	2	17
	V	1	1	48	2	65	1	1	37	2	34	2	2	23	2	19

SUBJECT	HEAD MOVEMENT	RG					MD					VJ				
		RG	BTE	FR	EGF	FR	MD	BTE	FR	EGF	FR	VJ	BTE	FR	EGF	FR
	VI	1	2	47	2	62	2	2	39	2	35	2	2	27	2	20
20	I	0	0	42	0	60	0	0	39	0	34	2	1	27	2	17
	II	1	2	42	2	66	1	1	37	1	35	2	1	23	2	18
	III	1	1	46	1	62	0	0	37	0	34	2	0	23	2	17
	IV	1	2	43	2	63	0	1	38	0	35	1	1	28	1	18
	V	1	2	42	2	66	0	0	38	0	35	2	1	24	2	18
	VI	1	1	43	2	66	1	0	30	1	34	1	0	21	0	20
21	I	2	2	45	2	62	1	2	38	2	34	2	1	23	2	18
	II	2	2	41	2	63	1	2	38	2	34	2	2	22	2	16
	III	1	1	43	2	61	1	1	37	1	35	2	2	25	2	18
	IV	2	2	47	2	57	1	2	38	2	34	2	2	26	2	18
	V	1	1	41	2	65	2	1	37	2	34	2	2	25	2	18
	VI	1	1	42	2	63	2	2	38	2	33	2	2	24	2	17

SUBJECT	HEAD MOVEMENT	RG					MD					VJ				
		RG	BTE	FR	EGF	FR	MD	BTE	FR	EGF	FR	VJ	BTE	FR	EGF	FR
22	I	1	0	41	1	68	2	0	38	2	34	0	0	0	0	0
	II	1	0	43	2	59	2	0	38	2	33	2	1	23	1	19
	III	1	0	42	2	65	2	0	38	1	34	2	1	24	1	19
	IV	1	0	42	1	68	2	0	38	1	35	2	0	27	1	16
	V	1	0	42	1	61	2	0	38	2	33	2	1	29	1	17
	VI	2	0	42	1	66	2	0	38	2	34	1	1	27	2	15
23	I	1	0	43	2	63	2	0	37	2	34	0	0	0	0	0
	II	1	1	42	1	54	2	1	38	1	35	2	1	26	2	18
	III	1	0	42	1	58	1	0	39	1	35	1	0	24	1	18
	IV	1	0	45	2	62	1	1	39	2	35	0	0	0	0	0
	V	1	0	42	2	64	2	0	34	1	34	0	0	0	0	0
	VI	1	0	43	2	61	1	0	37	1	35	2	1	25	1	19