Diplomarbeit

VELOC

Mouse Control Technique by Detecting and Tracking of Eye Movements in Real-Time

Cursor Steuerung durch Detektion der Augenbewegungen in Echtzeit

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Diplom-Ingenieurs am Institut für

Allgemeine Elektrotechnik und Elektronik

der Technischen Universität Wien

unter Anleitung von

ao.Univ.Prof. Dipl.-Ing. Dr.techn. Peter Pfundner

Univ.Ass. Dipl.-Ing. Dr.techn. Wolfgang Zagler

durch

Peter Skiczuk

Josef-Österreichergasse 17/6

A-1230 Wien

Wien, im Juli 1998

DECLARATION

No portion of this thesis has been submitted in support of an application for another degree or qualification at this or any other university or other institution of learning.

ACKNOWLEDGEMENTS

Kurzfassung

In unserer Gesellschaft spielen heutzutage Computer eine Schlüsselrolle und sind aus dem alltäglichen Leben nicht mehr wegzudenken. Ihre Einsatzgebiete umfassen beinahe unseren gesamten Lebensraum und dienen auch als wichtige Informationsquelle. Dies erfordert selbstverständlich auch qualifizierte Benutzer, die in der Lage sind neue Technologien effizient anzuwenden.

Im Zuge dieser rasant fortschreitenden Entwicklung und unter Betrachtnahme der vorhandenen Kommunikationsschnittstellen zwischen Mensch und Maschine mußten behinderte Menschen zusehen, wie ihre Erwartungen auf Integration in einer von Computern geprägten Welt zusehends schwanden.

Daher gilt das Hauptaugenmerk dieser Diplomarbeit der Entwicklung eines einfach zu bedienenden Eingabemediums. Die Verwendung von herkömmlichen intrusiven Eye-Tracking Techniken schränkte die Verwendung als geeignetes Eingabemedium stark ein, so daß deren Weiterentwicklung in Richtung nicht intrusiver Techniken ging.

Diese Arbeit untersucht ein nicht intrusives System, welches eine Simulation der bisher weit verbreiteten "Computermaus" darstellt und auf Basis von Beobachtungen der Augenbewegungen in Echtzeit arbeitet.

Die anschließende Realisierung zeigte, daß dieses System ein viel versprechender Schritt in Richtung einer Echtzeit-Emulation der "Computermaus" darstellt.

Abstract

Computerised systems are playing a key role in our society. They are being widely used in all areas and environments, especially in tasks regarding to access to information. The increasing complexity of this equipment and its utilisation is making the necessity of using high qualified people able to interact with them in an efficient way.

As a result of that, disabled people have seen how their expectations of integration within this high technological world are being reduced, despite having the same needs for access to information as do all other users.

The main aim of this thesis is the design of a useful tool for helping this important sector of our society to interact with computers. The physical constraints placed on a user by intrusive eye tracking techniques have given them limited applicability and emphasised the need for non-intrusive methods.

The proposal of this thesis is to undertake investigation into a non-intrusive system to simulate a pointing device such as a mouse by designing a feasible system able to recognise and determine the position of a user's eye gaze in real-time.

The final implementation of the system in a computer has pointed out its suitability for eye's location and its feasibility to user's eye gaze determination.

It is hoped that the project described in this thesis provides a contribution towards the ultimate goal of designing a real time mouse emulator.

Table of Contents

Index of figures

Index of tables

1   Introduction

1.1 Motivation for this Thesis

The effort behind this thesis is dedicated to those individuals, who can control their eye movements better than movements using any other group of muscles.

Nowadays, the most popular input device, other than the keyboard, is the „mouse", which is supported by a graphical interface of the operating system. Problems of accessibility to computer-based applications by people with disabilities are commonly attributed to the various issues related to enabling access to their computer, which are normally developed without taking into account needs, requirements and references of all potential users. Utilisation of the mouse by people with difficulties moving and co-ordinating their hands and arms is hardly possible so new input technologies have to be developed.

For these people, eye tracking holds out the possibility of providing them with an efficient and effective means of computer input. The VELOC (**Vi**su**el** Eye **Mo**vement **C**ontrol) system provides them with an opportunity to take an active part in our computer age. Getting access to computers means getting access to information and this will be more important than ever before. The VELOC system offers these people an inexpensive way so that they will not be as restricted in future.

Owing to the fact of the simplicity of using this new input device it also offers great facilities for „standard" users to accelerate their way of working by using this new technology.

1.2 Eye Gaze – Human Computer Interface for Disabled People

To explain why eye-gaze is such an interesting item to track, one can picture the cognitive processes that constitute perception as activities in working memory, which is part of the short-term memory. According to Just & Carpenter:

> „*The immediacy assumption*, which states that fixated objects are immediately processed and the *eye-mind assumption*, which states that the eye looks at an object as long as it is being processed, one can assume that if the subject's current task requires information from the display, then we can assume that the direction of eye-gaze is a good indicator of what is currently being processed."[1]

Owing to this fact you can argue that eye-gaze tracking is best used in multimodal interfaces because the user interacts with the data instead of the interface, in so-called non-command user interfaces.

The spine (Columna Vertebralis) is one of the most sensitive part of our body. If it is injured paralysis can occur. The form of paralysis due to spinal cord injury depends on the severity and site of the damage. The higher up the spine, the farther up the body the impairment. Thus tetraplegia can be caused by damage high on the spine, whereas a lower injury may cause paraplegia. Also the site of the damage is often according to the damage of the vertebra. The highest vertebrae are the cervical vertebrae, labelled C2 to C8, counting from the top down. Fractions in the lower numbers will have greater effect. Above C4 the spinal cord controls breathing, and a complete disruption above that level will cause death. Someone with severe damage at the C4 level might be able to use the chin or a mouthstick to interface with the controls of a computer, but almost certainly will be able to control the eyes. With damage at C5, the person will still need to use special devices and assistance. He/she will be able to do

all the necessary things in a wheelchair but will need help to move out of it. Someone with damage at C6 has better functions and will be able to drive an adapted car.

What is apperent is that even the most impaired people can still use their eyes to communicate with their environment. A user-friendly eye-gaze input device such as VELOC is the optimum link for them to make sure that they have still control of what is going on.

1.3 Aim and Purpose of this Project

The purpose of the VELOC project is to emulate one of the most common computer input devices, the mouse, with a hands free interaction computer interface by using the user's eye-gaze.

This system incorporates a camera for continuously monitoring the user's right eye and special software for the eye tracking process and simulating the mouse functions (movements and button clicks).

To make it as comfortable as possible for the user, a lightweight mini camera is fixed on a spectacle frame which has a good view for accquiring the required image of the eye.

Another main goal of this project has been to realise an inexpensive product which is easy to handle and available for everyone. To meet this requirement, no additional hardware, for example a frame grabber, should be used. Therefore, the camera (Quickcam product) should only need to be plugged to the parallel port of a PC. The software also should be able to run in real-time on a Windows 95 machine with an adequate processor so that an immediate interaction between the user and the graphical user interface of the operating system is possible.

1.4 Thesis Structure

This thesis investigates how the users eye gaze can be determined from a live video camera to control the cursor on a graphical user interface of an operating system.

Considering the different development stages involved in the resolution of this problem, the following chapters describe the different processes undertaken, from the early stages of the project to the final conclusions.

Chapter 2 discusses computer facilities for disabled people. An introduction to different kinds of disabilities is given. In order to enable these people to interact with computers, hands free computer interaction is introduced. At last design guidelines for disabled people to make their life easier are reviewed.

Chapter 3 gives a short introduction into the anatomy and physiology of the human eye.

Chapter 4 presents existing eye tracking techniques and discusses their advantages and drawbacks. It also contains necessarily requirements of different techniques, a comparison of those tracking techniques and a help to chose the appropriate eye tracking system as well.

Chapter 5 demonstrates existing eye tracking applications and discusses their features.

Chapter 6 offers the required image processing basics used in this thesis.

Chapter 7 introduces the VELOC system and explains its principle. Each stage of the VELOC image processing algorithm is explained in detail and underlaid with captured eye images.

Chapter 8 shows the evaluation of the VELOC system and explains the procedure to operate the system. Testing Conditions and results are illustrated as well.

The final conclusions and further improvements are considered in chapter 9. Also a preview in „Vision of the future" is given.

2 Computer Facilities for Disabled People

2.1 Introduction

Communication between human and computer has become more important over the last decades. Many input devices have been introduced but for all of them the primary target was the „normal" computer user, who had no disabilities.

However, as computer technology became more prominent in society, the need for a new form of interaction was required: a form of interaction that can be implemented for disabled people or people who are too busy with their hands. Although they share similar problems, these fields have traditionally been seen to have very little in common. In the great majority of cases, human-computer interface research seems to assume that „the user" will be an ordinary person with average abilities. Therefore, researchers often use a single model of the user, with little or even no attempt being made to specify the actual characteristics of this user. The only difference being made was to divide the user population into simple groups such as „novice" and „expert". This of course, does not necessarily include the special needs of the individual computer user. At its worst, the human interface to some software applications gives the impression that the designer's model of the user was a 26-year-old male with a doctorate in computer science who is besotted with technology.

Designing systems for those people who exhibit significantly lower than average abilities has traditionally been the remit of rehabilitation engineers. The main goal of these engineers is to communicate with the prospective users to make the most of their interface ability.

There is a view that this field addresses only small and unprofitable markets. There are over 30 Million people in the United States alone with disabilities or functional limitations and this number is increasing [2]. Anyway, new input devices are not limited to disabled people; even able-bodied computer users might find it comfortable to use these new technologies.

Clearly, all interaction techniques should be easy to handle so that they are suitable for everyone. The best technology might not launch successfully on the market if it only can be used with big effort.

This chapter first gives an overview of different types of disabilities and then presents already existing computer-input devices for disabled people for hands free interaction that is a major goal of the VELOC system.

2.2 Disabilities

2.2.1 Definitions

According to the United Nations Declaration on the Rights of Disabled Persons:

> *„The term 'disabled person' means any person unable to ensure by himself or herself, wholly or partly, the necessities of a normal individual and/or social life, as a result of a deficiency, either congenital or not, in his or her mental capabilities"* [3].

The question is to whom does this definition not apply. Everybody has his own limitations, but the point is that some people have impairments of their faculties, which severely affect their ability to take part in everyday life.

The Americans with Disabilities Act (ADA) is a little more specific:

> *A disability is „a physical or mental impairment which substantially limits one or more major life activities".*

The International Classification of Impairments, Disabilities and Handicaps (ICIDH) is an attempt to make sure that impairments do not automatically include handicaps. Impairment is „any loss or abnormality of psychological, physiological or anatomical structure or function." Disability is „any restriction or lack (resulting from an impairment) of ability to perform an activity in the manner or within the range considering normal for a human being." Finally, a handicap is „a disadvantage for a given individual, resulting from an impairment or disability, that limits or prevents the fulfilment of a role for that individual."

Thus, if there is any possibility of avoiding disadvantages caused by handicaps, then every endeavour should be made to achieve it. The VELOC project is just such an effort to prevent disabled people who are interacting with a computer, from feeling handicapped.

2.2.2 Kinds of Disabilities

2.2.2.1 Overview

The five areas of human functioning can be considered to be vision, hearing, mobility, speech and learning. Vision, hearing and mobility are necessary to handle a computer system without problems. The following pages discuss these disabilities and show some possible solution for successful interaction. The best solution might be an interplay of several techniques and depends on the needs of the individual user and his/her disabilities. For example, eye gaze tracking technique and speech input are not naturally exclusive and can be combined achieving a higher level of computer interaction.

The document „Consideration in the Design of Computers and Operating Systems to Increase Their Accessibility to Persons with Disabilities" [5] includes a set of tables which summarise very well the requirement of a range of users with different disabilities and possible approaches to meet them.

2.2.2.2 Vision

Owing to the dominance of screen based visual communication, visual impairment is the most significant sensory impairment in terms of human-computer interaction. It is necessary to distinguish between blindness and partial sights. A blind person has no useful vision and so must rely on non-visual forms of communications, for example touch and sound can be substituted. A partially sighted person has a degree of useful vision and may be able with some aids, to interact with computers. These aids can include appropriate lighting, choice of colours or an particular screen technology. The following enumeration of different input and output devices for visual disabled people should demonstrate the wide range of alternatives.

Input Access: Keyboard, Braille keycaps, large print keycaps, colour coded keys, optical character recognition (OCR)

Output Access: Braille, computer speech output, screen reviewer, screen magnification, large print, alternative documentation

Due to the fact that the VELOC system relies on eye movements, visual impaired people may not benefit.

2.2.2.3 Hearing

Users with hearing impairments are not dramatically affected when using computers. They are able to interact without problems because sounds only provide additional information for „normal" users. Instead of an acoustic signal a visual signal can be implemented very easily. Disadvantages only occur when important information is only presented in acoustic way such as in games. Tactile information could be one solution to solve this problem. The VELOC system also offers the possibility of being used by these people but it is more an additional communication device rather than a necessity.

Input Access:	Keyboard, mouse, speech recognition
Output Access:	Screen blinker, visual beep, headphones

2.2.2.4 Movement

There are a variety of different physical disabilities. People with impaired manual function may not be able to use conventional input devices such as „mouses" and keyboards but can control the movements of the rest of their body. Some other people are unable to carry out such everyday operations as walking but are in a position to co-ordinate their hands and fingers and are lucky to have no interaction problem with computers.

People with impaired manual function are one target group for the VELOC system. Due to the fact that eye movements are relative to the head and the VELOC system uses a head mounted device, the user's head must be held quite still otherwise head movements can disrupt the mechanism. The user him/herself might be able to achieve this if he/she is able to control his head movements, or his head must be fixed using some apparatus. This latter solution is not ideal but it is better than no interaction at all.

An interface between a physically disabled person and a computer can be vital. Not only does it give access to normal functions of the computer, but the computer can also be used as a gateway to many prosthetic devices. There is a wide range of new opportunities for computer interaction. Nearly every electronic device could be managed by computer-controlled remote control thus enabling disabled users. This individual remote control can than be adapted to the user, and by the user.

The following input and output devices are some examples of already existing techniques:

Input Access:	Sticky keys, variable keyboard and mouse timing, keyboard cursor control, plastic keyguard, trackball, word and phrase prediction, speech recognition, large and small keyboards
Output Access:	Speech output, physical access through raised tables

2.2.2.5 Speech

Currently, speech input is not the most used interaction facility. Thus, speech impaired people usually have no disadvantages compared to „average" users when communication with a computer takes place. The problem occurs when communication between humans is desired, so the requirement for a communication device is to replace the missing ability to speak; a means whereby the user inputs what he/she wants to say and the computer communicates it to other persons. The output can be in written or spoken form. Speech producing devices have an obvious attraction as a replacement for natural speech.

The following figure illustrates the basic concept of text-to-speech conversion for English.
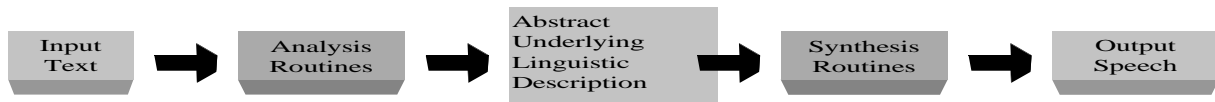
figure 1: Text must be converted to an abstract linguistic representation so as to be able to generate an accurate synthetic approximation to an English sentence [29]

First the input text has to be analysed. This is necessary to determine the underlying structure of the sentence and the phonemic composition of each word. Following tasks are involved in analysis routines:

- pronunciation of each word (phonemes and stress)

- syntactic structure of sentence

- semantic focus, ambiguity resolution

The second step is to transform this linguistic description into a speech waveform, which is managed by synthesis routines. These processes have interesting connections to linguistic theory, models of speech production and the acoustic-phonetic characterisation of language.

The synthesis routines can be separated into following main parts:

- phonetic realisation of each phoneme

- duration pattern

- fundamental frequency contour

- phonetic - to - acoustic transformation

2.2.2.6   Learning

Cognitive impairments include limitation in thinking, memory, language, learning and perception. This causes problems when using information technology such as computers. The problems faced by people with cognitive impairments are various but can be loosely categorised into four groups [6]:

- Memory problems include difficulty in retrieving information from short term, long term, or remote memory. Difficulty in recognising and retrieving information is included in this category

- Perception problems include difficulty in taking in, attending to, and discriminating sensory information

- Problem solving impairments include difficulty in recognising a problem, identifying, choosing, and implementing solutions and evaluating outcomes.

- Conceptual difficulties include problems in sequencing, generalising previously learned concepts, comprehension, and skill development

There are several design guidelines for reducing cognitive barriers to computers. These designs must address the whole of the interface. For example, many early programs for people with reading difficulties were inaccessible because the input mechanism – a keyboard – required a higher cognitive level than the task represented in the software. As summarised in the table below the Industry/Government Co-operative Initiative on Computer Accessibility recommend the following steps for reducing cognitive barriers:

- Simplify language in task instructions

- Provide on-line help

- Make display simple

- Use consistent display, menus, and selection techniques

Some additional input and output devices are listed below and should give some helpful keywords for searching:

Input Access: Word and phrase prediction, coloured keytops, speech recognition, organisational and outline tools, electronic dictionary, grammar checker

Output Access: Quiet, non- distracting environment

2.3 Hands free computer interaction

Hands free computer interaction as the name suggest is a form of interaction that allows users to communicate with a device without the need for them to use their hands. This interaction can not only be used by disabled people; it also offers common users new ways to handle their system. It can be seen as an additional input device that accelerates communication. Hands free technology is another area of input, which has plenty to offer the user who is otherwise busy with his/her hands. Nowadays, many different hands-free technologies have been developed. The most important ones are instanced below.

- Blow switch (eye and head input)

- Hand gesture

- Foot-switch

- Speech

- Facial recognition

- Eye tracking

Depending on the disability and its severity the user can choose between these technologies or even combine them.

2.3.1 Blow Switch (eye and head input)

The blow technology is used for disabled people to communicate with their environment. The main goal of this method is to enable the user to communicate with a computer by blowing into a tube to accept a choice that is presented on the screen. By controlled movements of the head it is possible to scroll through a list of predefined choices. The blow switch is used instead of the mouse button. Depending on how the user blows into the tube, the system can react in a different manner. This enables emulation of the right mouse button click.

Blow technique combined with an ultrasonic signal source mounted onto a headset enables measurement of the head position. A keyboard is displayed on the screen and the system detects slight movements of the user's head and moves the cursor accordingly. When the user blows into the tube the located key is accepted and is highlighted.

Head movement is one possibility which may be used in conjunction with blow switches. Other techniques can also be applied to the blow switch.

2.3.2 Hand Gesture

Although hand gesturing could be classed as a mode of interaction that does include the use of hands, for this special purpose, hands free interaction can involve the ability to gesture via hand signals. Sign language is a language after all, so it can be implemented as a new kind of computer interaction between users and computer.

This is an application field for pattern recognition, where the shape of the hand or fingers has to be found. To simplify image processing, the user can define a few basic hand signs, which he/she is going to use. For example the pose of the thumbs up symbol could be used to present a user's affirmative response when they are asked a question.

Another area that needs to be taken into consideration is the use of lighting to enhance the location of the hand. For example, wearing a coloured (red) glove enables the algorithm to reduce the level of green and blue and makes it easier to find the object of regard or even shows the image features more clearly.

2.3.3 Foot-switch

The foot-switch or foot mouse is a form of interaction that makes use of a block shaped pedal device to determine either direction or choice. Directions can be chosen by pressing the foot on one of the corners of the foot-switch.

Nowadays, foot-switchs are mainly used in computer games to decrease or increase the speed when playing racing games. Unfortunately it is a neglected source of input. The user has to get used to the new kind of interaction with computers, especially when extra buttons on the foot-switch are applied to increase the functionality.

2.3.4 Speech

In „21st Century Eloquence, 1996" it is written:

> *„By incorporating speech, the most natural form of communication, voice-activated products make computers „people literate" and easily accessible to a wide audience of users."*

Speech recognition is one of the most attractive research areas when it comes to hands free interaction. Many companies, such as IBM, Philips, Kolvox, Kurzweil, Dragon and Voice Connection are only some, which are involved in speech recognition. These companies have developed a variety of products enabling disabled people to communicate verbally with the computer. But of course , this technique can be used by anyone to make computer interaction easier and more comfortable.

Understanding natural speech is one of the greatest challenges, because of its large range of speaking characteristics and vocabulary. VODIS (Voice Operated Database Inquiry System) which was developed by British Telecom, Logica and Cambridge University is an example of such a system. VODIS uses a continuous speech recogniser that accepts natural spoken words and phrases and then passes them to a dialogue controller, which interprets the received speech using knowledge of the applications domain. For the system to work efficiently questions were worded so that specific responses could be given.

Speaker dependent systems are easier to implement due to the fact that it is dedicated to an individual who is trained on that program. The disadvantage of course is that no one else can use this system. A default configuration should enable the user to train the system. This is an application field for disabled people because such systems are cheaper than speaker independent systems.

Special hardware requirements are normally not necessary; users only need a microphone connected to a sound card. This technique is very cheap, only the software raises the costs.

The following example should underline the wide application fields in everyday life in the food processing industry:

The food handlers have to record information on products as they perform inspections. However using a keyboard is impractical because the inspectors would risk contaminating food products. Speech input allows the inspectors to do both tasks simultaneously without having physical contact with the computer.

2.3.5 Facial interaction

The nature of facial gesture is very similar to that of hand gesture. Both techniques require image pattern recognition. The user computer interaction has also to be modified so that questions can be answered by a simple gesture. For example, a smile could determine that a choice has been accepted or a frown could determine that a choice has been refused.

The research into pattern recognition for face identification is also a popular area.

2.3.6 Eye tracking

The general idea of eye tracking is that by analysing where a person is looking on the screen, an area of interest can be determined and thus an action can be performed. There are five areas for eye tracking application [4], which are all met by the VELOC system.

- Object selection

  Selecting an object is done by pointing at the object with a cursor and then pressing one of the buttons to select this. With the use of eye tracking, the cursor can move to the location at which the eye is fixated. After a short period, if the cursor has not moved then the object is selected. Due to the fact that you also focus on an object when you are interested in it, this method requires a special area on the screen where the computer user can rest. An advanced method of eye tracking is presented in the VELOC system where the mouse click is emulated by blinking with the eyelid. This makes it more comfortable and accelerates interaction. For further details see chapter 8.

- Continuous attribute display

  This is applied to the retrieval of further information on a selected object. For example, a text area of the screen can represent information relating to an object and by selecting the object, then looking at the text area of the screen, information will appear. This presents a new way of accessing additional information and could be used to emulate the right mouse button, particularly when working with graphical user interfaces such as Windows.

- Moving an object

  Eye tracking can also be used for moving an object. This requires the user to select an object by looking at it (VELOC offers the possibility of clicking) and then moving the cursor to its new position. This is form of multi-modal interaction, which can be used in instances where more than one interaction is required. For more information on multi-modal interaction see the next chapter.

- Eye-control scrolling text

  This is needed for working with text applications. In normal text applications it is possible to scroll up and down the text document by clicking on the directional arrows. Depending on the accuracy of the eye tracking system it might be necessary to enlarge these buttons or even define new ones.

- Menu commands

  With the use of multi-modal interaction, menu commands can be easily used. By looking at the menu bar and clicking on it, pull-down menus can be selected and by looking down

the list of options on the menu, scrolling can take place. By clicking on the menu instruction of regard it will be selected.

For details on eye tracking techniques see also chapter 4.

2.4 Design Guidelines for Disabled People

There are far too many guidelines for disabled people that they can all be discussed in this thesis. Therefore, some useful references will be summarised to give access to this information. These references also include links to other information for disabled people.

„Accessible Design of Consumer Products: Guidelines for the Design of Consumer Products to Increase Their Accessibility to Persons with Disabilities or Who are Ageing" [7]:

This reference contains the latest version of materials being developed by a task force involving researchers, industry, government, and consumers, identifying ways that standard manufactured consumer products can be made more accessible to people with disabilities.

„A Protocol for Communication Between a Device and an Auxiliary Interface"[8]:

This working paper looks at designing a universal interface to allow access to a variety of technologies by using infrared and other communication channels. Included are discussions of the need for the protocol as well as design specifications and examples.

„A Collaborative Information Model for Next Generation Information Systems"[9]:

Proposes a system for information collaboration using the convenience of the Web as well as other types of access including telephone. This system would allow professionals, consumers and providers access to useful information in accessible formats through a variety of paths.

„Access to Current and Next-Generation Information Systems by People with Disabilities"[10]:

This comprehensive document provides information and resources for those interested in learning more about accessibility issues in current and next generation information systems. The focus of the document is on the National Information Infrastructure (NII), sometimes known as the "information superhighway."

„Application Software Design Guidelines: Increasing the Accessibility of Application Software to People with Disabilities and Older Users"[11]:

The purpose of these guidelines is to document what application developers can do (or need to do) in order to make their software accessible and usable by people with disabilities, or people who have reduced abilities due to ageing. This 50-page set of design guidelines includes specific information on everything from design to support and training. It also has an extensive background and reference section.

A well-implemented *search engine* for all questions and published papers about technical helps for disabilities can be found at [12].

3   Vision - The Eye

3.1 Introduction

To understand the different kinds of image acquisition of the eye it is necessary to be acquainted with its anatomy. Therefore, the basics will be discussed in this chapter. In the next chapter, the different techniques based on reflected light for eye tracking will be demonstrated where the basic knowledge of the eye´s anatomy is needed. For more details about the eye, medical literature should be referenced.

3.2 Anatomy of Human Eye

In this chapter the anatomy of the eye will be described. Therefore, the most important and interesting parts of the eye are shown in figure 2 and an explanation is given.
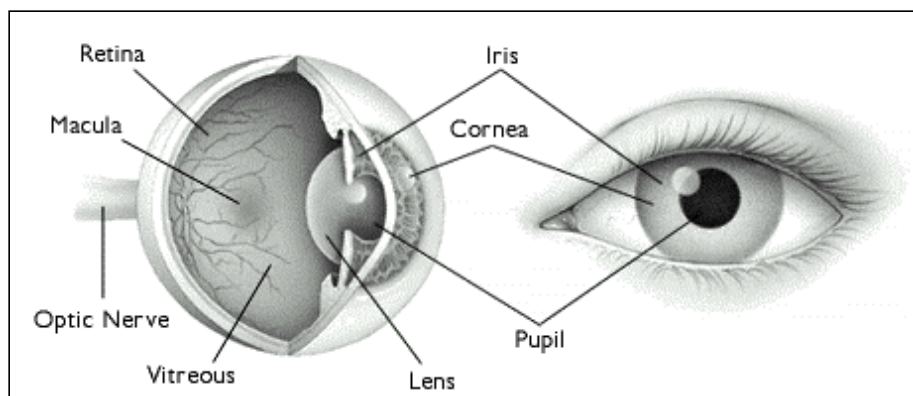
figure 2: Anatomy of the eye, picture taken from [13]

- Cornea:

  The cornea is an external transparent surface that covers both the pupil and the iris. This is the first and most powerful lens (about 70% of the total referective power) of the optical system of the eye and allows, together with the crystalline lens, the production of a sharp image at the retinal photoreceptor level. The shape of the cornea does not change, with the exception of small changes that occur over a lifetime.

- Lens:

  The lens is the transparent structure inside the eye that focuses light rays onto the retina; it finishes the focusing. It helps to „fine tune" vision, and it is able to change shape to allow focus on near objects. If it becomes cloudy it is said to have a cataract. Unlike the cornea, the lens can made to change its shape rapidly and voluntarily. Using its ability to change shape, the lens allows the eye to change its focal lenght.

- Iris:

  The iris is a coloured circular muscle, which is beautifully pigmented giving us our eye's colour. This circular muscle controls the size of the pupil so that more or less light, depending on conditions, is allowed to enter the eye.

- Pupil:

  The Pupil is the dark centre in the middle of the iris. The pupil determines how much light is let into the eye. It changes size to accommodate the amount of light that is available.

The pupil is small in bright light and large in dim light. It appears dark because of the absorbing pigments in the retina.

- Vitreous:

The vitreous is the clear, jelly like substance that fills the middle of the eye. This tends to liquefy with age, and its separation from the retina can lead to retinal tears and detachment.

- Retina:

This is a thin layer of nerve tissue that senses light at the back of the eye. Specialised cells called rods and cones convert light energy into nerve signals (impulses) that travel trough the optical nerve to the brain. The retina is analogous to the film in a camera

- Macula:

The part of the retina which is most sensitive and is responsible for the central (or reading) vision. It is located near the optic nerve directly at the back of the eye and receives the focus of the object of regard. Because of its special light-sensitive cells the macula allows us to see fine details clearly. This area is also responsible for colour vision. It is also called the Fovea.
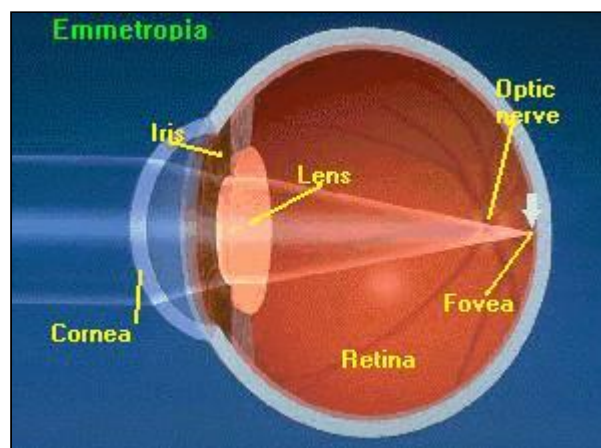
figure 3: Cross sectional view through the eye, picture taken from [14]

- Optic nerve:

The optic nerve connects the eye to the brain. The optic nerve carries the impulses formed by the retina to the brain, which interprets them as images. It consists of about 1.2 million nerve fibres. The optic disc is visible on the inside of the eye, where the nerve is viewed „end on".

- Sclera:

The Sclera is the white, tough wall of the eye. It gives a good contrast to the iris for image processing.

As the cross-sectional view of the eye shown in figure 3 illustrates, there are three different layers:

1. The external layer, formed by the sclera and cornea

2. The intermediate layer, divided into two parts: anterior (iris and ciliary body) and posterior (choroid)

3. The internal layer, or the sensory part of the eye which is called the retina

Light reflections on these different layers can be used for capturing images with special contents. This will be discussed later.

Literature structure of the retina of the human eye tells that the retina consists of approximately 127 million light-sensitive cells. Of these, 120 million (94%) are so-called rods, that have a high sensitivity, i.e. they can detect relatively small amounts of light-but without colour information. The remaining 7 million cells are the cones, that are less light-sensitive, but can capture the colour of the human visible light spectrum.

If the cones were distributed evenly among the rods across the retina, their average distance would be relatively large, and thus the acuity - the ability to detect fine spatial patterns - would be relatively poor for daytime colour vision. Therefore, the cones are concentrated in the centre of the retina, in a circular area called the macula lutea. Within this area, there is a depression called the fovea, that consists almost entirely of cones, and it is this area of high acuity, which extends over a visual angle of approximately 2°, that humans use to make their detailed observations of the surrounding world.

The remaining part of the retina offers peripheral vision, which is characterised by being of only 15-50% of the acuity of the fovea- partly because there is only one efferent optical nerve for approximately every 300 rods- and it is less colour-sensitive, but the structure of the peripheral part of the retina is more reactive to flashing objects and sudden movements.

A visual angle of 2° is slightly less than the width of one's thumb held out at arm's length, or approximately the width of an average word held at normal reading distance. Thus, detailed observations of a reasonable part of the surrounding world requires moving the eye (and eventually the head and body) to successively focus different parts of the ambient light array on the fovea, thus foveating various regions of the observed scene.

3.3 Physiology of Human Eye

The eyeball is controlled by three pairs of muscles. One pair is for horizontal eye movements, another for vertical movements and a third for rotational (rolling) movements, i.e. movements around an axis in the direction of sight. This is shown in figure 4.
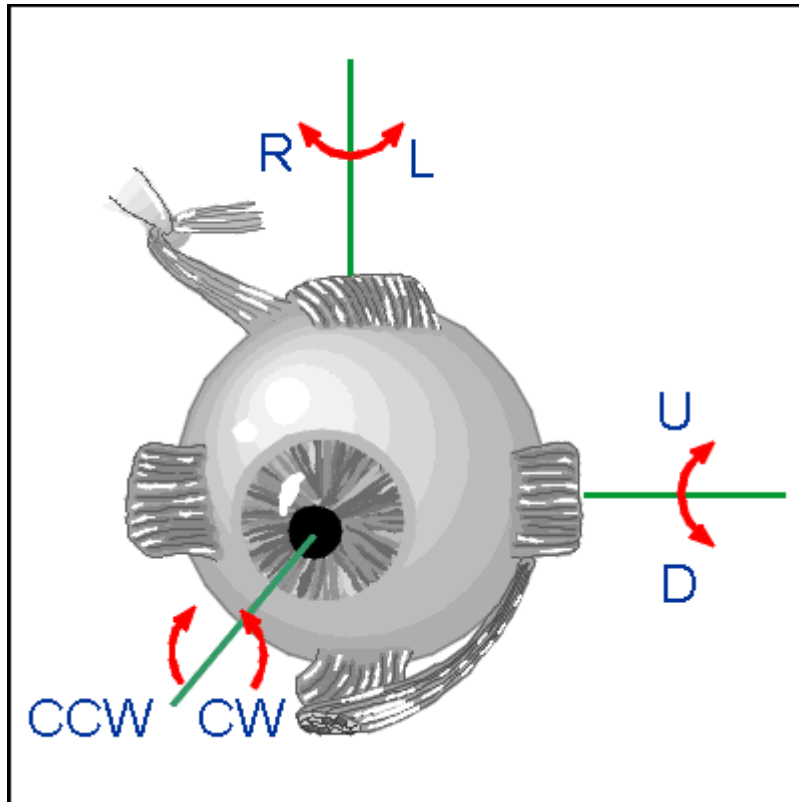
figure 4: Eye muscles of the human eye

As shown in figure 4 each eye has 3 agonist/antagonist muscle pairs.

An interesting finding is that eye-gaze occurs in cycles, creating a pattern, a scanpath. When a subject views a picture, the informative areas are rapidly sought out and scanned, and this scanpath is then repeated. It is important to notice that the scanpath is determined both by the picture compositional factors and by idiosyncratic factors.

This is accomplished mainly by fixations where the eyes are stationary separated by saccades where the eye moves in a ballistic way to view a new point in the visual scene. Research has shown that the place for the next fixation is determined during the fixation, and according to the feature integration theory proposed by Treisman, these processes can be divided into a pre-attentive, parallel phase and a subsequent attentive, serial phase. The pre-attentive phase is unlimited in capacity and operates in parallel across the entire visual field, calculating local mismatch for basic features like orientation, size, colour, movement direction. These mismatchings values are added together and in the attentive stage mismatchings are serially inspected according to their degree of mismatch. The pre-attentive phase is driven by a bottom-up process, whereas the attentive phase is at least partly under strategic control through a top-down process that - according to the zoom lens metaphor - can limit attention to a greater or smaller part of the visual scene. Finally, the different basic features from the selected objects are integrated and a stimulus is identified, before the next fixation position is determined. This process is illustrated in figure 5.

When the pre-attentive and attentive stages have determined the position of the target, the eye
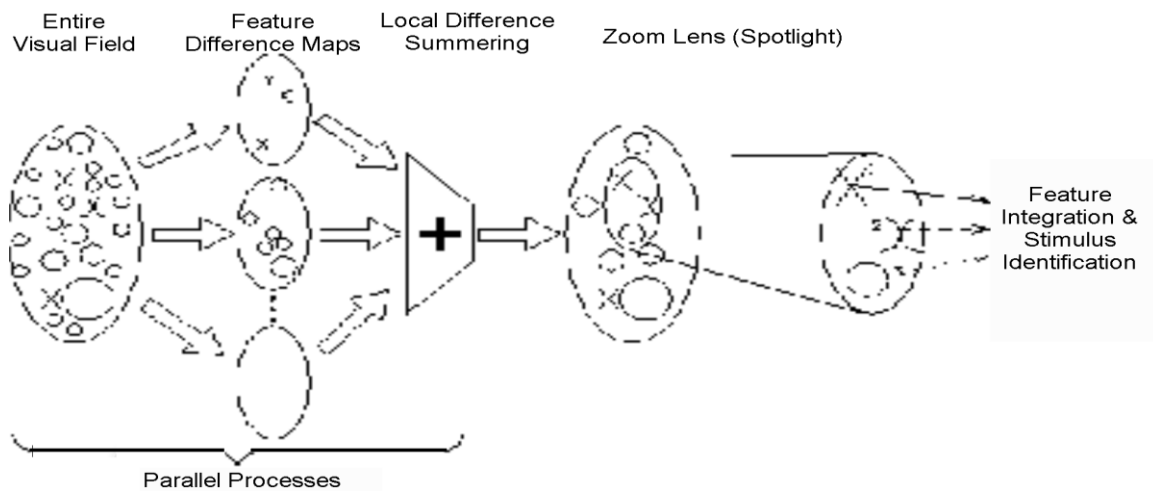
figure 5: Image processing of the human eye

must be moved in such a way that the target object can be inspected with a higher acuity, by foveating the object. The movement of the eyes to a new location is performed by executing a saccade; this is not the only type of eye movement. Research literature [31,32] classify seven different eye movements:

- Convergence

  This is a motion of both eyes relatively to each other that ensures that an object is still foveated by both eyes when its distance from the observer is changed. The closer the object is, the more the eyes point towards each other. This movement can be voluntarily controlled, but is normally the result of a moving stimulus.

- Rolling

  Rolling of the eyes is a rotational motion around an axis passing through the fovea and pupil. It is involuntary, and is influenced by among other things the angle of the neck

- Saccades

  This is the principal method for moving the eyes to a different part of the visual scene, they are sudden, rapid movements of the eyes. It takes about 100-300ms to initiate a saccade, i.e. from the time a stimulus is presented till the eye starts moving, and another 30-120ms to complete the saccade, depending on the visual angle traversed. Saccades can be initiated voluntarily, but are ballistic: that is, once they are initiated, their path of motion and destination cannot be changed- which must be taken as an indication that visual attention in the peripheral area selects the next location for the eyes to move to. During saccades- actually from about 50ms before saccades are initiated- pre-processing of the visual image is suppressed, possibly due to a detection mechanism for large-scale movements of the entire retinal mosaic. Thus, processing of the retinal image takes place mainly between the saccades, during the so-called fixations, that last for about 200-600ms.

- Pursuit motion

  This is a much smoother, slower movement than a saccade; it acts to keep a moving object foveated. It cannot be induced voluntarily but requires a moving object in the visual field.

- Nystagmus

Nystagmus is a pattern of eye movements that occur as a response to the turning of the head (acceleration detected by the inner ear) or the viewing of a moving, repetitive pattern (the train window phenomenon). It consists of smooth pursuit motion in one direction to follow a position in the scene, followed by a fast motion in the opposite direction to select a new position.

- Drift and microsaccades

These occur during fixations and consists of slow drifts followed by very small saccades (microsaccades) that apparently have a drift-correcting function. These movements are involuntary.

- Physiological nystagmus

This is a high-frequency oscillation of the eye (tremor) that serves to continuously shift the image on the retina, thus calling fresh retinal receptors into operation. If an image is artificially fixed on the retina it disappears, but physiological nystagmus causes every point of the retinal image to move approximately the distance between two adjacted foveal cones in 0.1 seconds. Physiological nystagmus actually occurs during a fixation period, is involuntary and generally moves the eye less than 1°.

3.4 Comments

This chapter described the backgrounds and basics of the anatomy and physiology of the human eye. From this it is seen that any tracking of the eyes will result in tracking data consisting of superimposed movements of different origins. Whereas the saccades and fixations originate from the attentional process, the drift, microsaccades and physiological nystagmus are physiologically determined. This implies that the attention-pertaining data somehow needs to be extracted from the „raw" eye tracking data, and that there is a lower limit on the precision with which one can measure what the person is attending to.

4 Present-day Eye-Gaze Tracking Techniques

4.1 Introduction

This chapter is based on the previous one. Different techniques of eye gaze systems will be discussed and their advantages and drawbacks will be explained. Requirements of an eye gaze system will be compared to the capacity and features of each technology. All existing products and research work are using the different techniques which will be explained in this chapter.

The electro-oculography chapter contains background information from the project Eyetrans, which was developed during my living exam at the technical school TGM in Vienna 1994.

4.2 Requirements

To date, after decades of research activities, several ways for eye gazing exist. The following requirements of an eye-gaze system are taken from Hallett [15]. Of course, almost no technique will be able to satisfy all of these requirements.

The ideal tracking device must

a) Offer an unobstructed field of view with good access to the face and head

b) Make no contact with the subject

c) Meet the practical challenge of being capable of artificially stabilising the retinal image if necessary

d) Possess an accuracy of at least one percent or a few minutes of arc; e.g. not give a 10° reading when truly 9°. Accuracy is limited by the cumulative effects of non-linearity, distortion, noise, lag and other sources of error

e) Offer a resolution of 1 minute arc sec^{-1}, and thus be capable of detecting the smallest changes in eye position; resolution is limited only by instrumental noise

f) Offer a wide dynamic range of one minute to 45° (= 3000-fold) for eye position and one minute arc · sec^{-1} to 800 ·sec^{-1} (= 50,000-fold) for eye velocity

g) Offer good temporal dynamics and speed of response (e.g. good gain and small phase shift up to 100Hz, or a good step response).

h) Possess a real-time response (to allow physiological manoeuvres).

i) Measure all three degrees of angular rotation

j) Be easily extended to binocular recording

k) Be compatible with head and body recordings

l) Be easy to use on a variety of subjects

As said before, hardly any technique can satisfy all of these, but there is also no need for it. Of course it would be nice if it were possible, but it depends on the needs of each single application. If you want to develop an eye-gaze system, you have to define your target group first. This makes the decision about which technique to use much easier. Requirement (b) might not be crucial, as long as the physical construction of the device is discreet and is tolerated by the user. But this is not a technical question; it is more a question of universal

acceptance. For example, 15 years ago, nobody could imagine that listening to music with a Walkman could be relaxing even if you go by underground during the rush hour yet. It is now very common and nobody cares about it, if it does not affect the private domain. Perhaps it is also possible that people will eventually find it natural to wear spectacle frames with a small on-board camera.

If we are to classify the current techniques by the way they make contact with the subject, there are basically three types of tracking techniques:

1. Measuring the reflection of some light that is shone onto the eye

2. Measuring the electric potential of the skin around the eyes.

3. Applying a special contact lens that facilitates tracking of its position.

Of these, type 1 includes techniques that make the least contact, and type 3 includes techniques that typically make the most contact.

These different techniques are discussed in the next chapter.

- **VELOC- requirements and their satisfaction:**

The main goal of the VELOC project is to enable motion impaired people to interact with computers. This product should also be as cheap as possible, so that everybody can afford it. The VELOC system uses a visible light CCD video camera for image acquisition. No additional hardware is used. The camera is placed in front of the right eye. This decision was made to simplify matters for the prototype and should be changed for the final product to meet requirement (a). Requirement (b) can only be satisfied if the camera is placed at the bottom of the monitor. This increases computation time dramatically and it would currently not be possible to run the eye-gaze program in the background in real-time. This does not fulfil condition (h). For the simple reason that the camera is mounted on a spectacles frame, the head and body must be held quite still or the apparatus must be fixed to the user's head (requirement (k)) This is the main disadvantage of this system. The accuracy and resolution of the VELOC system is high enough to select buttons of a size of 2x2cm. This enables most applications, which have been written for disabled people, to function satisfactorily.

For more technical details see further chapters.

4.3 Techniques based on reflected light

4.3.1 Introduction

There are five techniques using light reflected by the eye to determine the users' gaze. Most of them using infrared light that causes high contrast images. The higher the contrast of an image, the easier it is for the image-processing algorithm to detect shapes.

4.3.2 Limbus Tracking

The limbus is the boundary between the sclera and the dark iris of the eye. Due to the fact that the sclera is normally white and the iris is darker in human eyes, this boundary can easily be optically detected and tracked. This technique is based on the position and shape of the limbus relative to the head, so either the head must be held quite still or the apparatus must be fixed to the user's head. Normally, this technique also works with visible light, as used in the VELOC system.

4.3.2.1   Visible light limbus tracking

The main disadvantage of using visible light is that the ambient light causes reflections on the image of the eye, which perturb the algorithm and so the centre of the pupil can be hardly detect. The following image demonstrates the reflection problem.
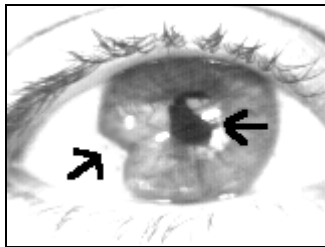
figure 6: Reflections in the eye image

Another problem with limbus tracking is that the top and the bottom of the limbus are more or less covered by the eyelids. This varies between users. Due to this fact Scott & Finlay [20] say, „it is probably fair to regard limbus tracking as suitable for precise horizontal tracking only". See also chapter 4.6 of this thesis. Figure 7 presents a geometrical model of the eye. This was used by real-time eye feature tracking with a Kalman filter [21].

Notation	Explanation
E	Whole eye region within parabolas
C	Whole iris
D	Iris circle within parabolas
W	Whit areas of the eye on the left and right sides of the iris
U	Part of the iris cut by the upper eyelid
L	Part of the iris cut by the lower eyelid
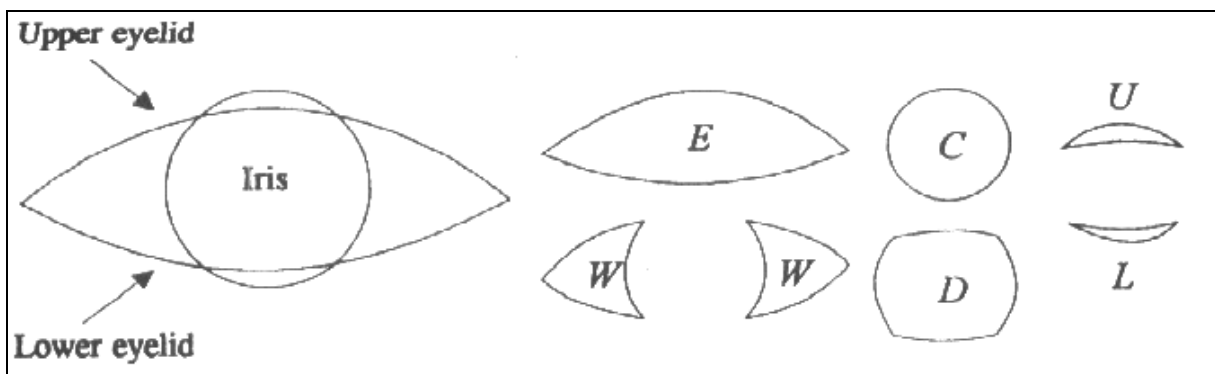
figure 7: Geometrical model and notation for the human eye

In particular both parabolas can cause shapes when the algorithm is in the state of edge detecting, because it is the most critical part of the edge detection . Owens to this fact, the subsequent steps of the algorithm will calculate with incorrect data and the correct centre cannot be determined. This also differs according where the user is looking. For example, when the user is looking down, the upper eyelid will automatically follow, so that the position of the eyelid will never be fixed.

4.3.2.2   Infrared light limbus tracking

In this method, one or two small infrared sources, together with detectors, are positioned close to the eye. They can be fixed on a spectacle frame. These sources (i.e. IR-LEDs) are situated so that they shine on either side of the iris or on the boundary of the iris and the sclera. Depending on the user's eye gaze, light is reflected back to the detector. When the eye moves to the side, more or less infrared light is reflected depending upon direction. These signals can be calibrated for eye-movement detection.

The advantage of this method is that it is relatively cheap. This results in the major disadvantage that its accuracy is not very high, especially in the vertical direction. Another problem is the set-up and alignment of the IR sources and detectors. This method might be practical for an eye gaze switch, where low accuracy is needed but for eye movement tracking it seems to be insufficiently accurate.

4.3.2.3   Photo-Electric limbus tracking

The Mackworth eye-tracker (Mackworth and Mackworth, 1958) and that described by Rashbass (1960) are examples of a photoelectric technique. In this method a spot of light images on the limbus, and the reflected light from it is detected by a nearby photocell. The amount of light reflected depends on the user's gaze. This information will be used to shift the position of the spot's source such that the limbus is continually tracked during eye-movements and the same amount of scattered light is reflected. The position of the source thus gives the relative position of the gaze. Though spatial accuracy is high, disadvantages of this method are the necessity for a moveable source and low ambient illumination.

4.3.3 Pupil tracking

Tracking the direction of gaze using the pupil tracking technique is similar to limbus tracking. The only difference is that the smaller boundary between the pupil and the iris is used instead. Once again, the apparatus must be held completely still in relation to the head.

The two main advantages of this technique over limbus tracking are that the eyelids do not cover the pupil and this enables vertical tracking, too. The second advantage is that the border of the pupil is often sharper than that of the limbus. This enables a higher resolution in both directions.

The disadvantage is that the difference in contrast is much lower between the pupil and the iris than between iris and sclera. Therefore, border detection is more difficult for image processing.

4.3.4 Cornea and pupil reflection relationship

When light, especially infrared light, is shone into the user's eye, several reflections occur on the boundaries of the lens and cornea, the so-called Purkinje images. This is demonstrated in figure 8.
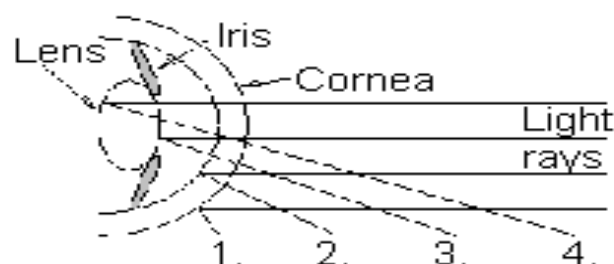
figure 8: Purkinje images

Part of the incident light beam is reflected by the front of the cornea and produces the so-called corneal reflection. It is the first Purkinje image and is also known as glint. The boundary between pupil and iris, which normally exhibits very low contrast, is enhanced with the bright pupil effect. The reflection of infrared light off the retina causes the bright eye. The reflection of infrared light from the human retina is similar to the reflection of visible light from a cat's eye at night. When using an infrared sensitive video camera this reflection can be seen as a very bright spot surrounded by a less bright disc. The corneal reflection and the bright pupil images are located in two different optical planes. When moving the eye, the relative positioning of the glint and the centre of the bright-eye change accordingly, and the direction of gaze can be calculated from these relative positions. This vector relationship between glint and bright-eye centre is demonstrated in figure 9.
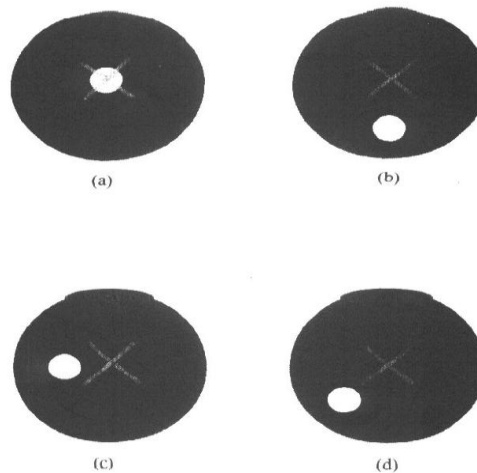
figure 9: **Vector relationship between glint and bright-eye centre, used to determine eye-gaze position (a) directly at camera (b) directly above camera (c) to right of camera (d) up and right**

As shown in figure 9, when the direction of the eye gaze coincides with the optical axis of the camera (i.e., when the user looks directly at the light source), the centre of the bright-eye coincides with the centre of the glint. The algorithm has to determine both centres of the glint and of the bright eye.

This is a very accurate method to detect a user's eye gaze. This method is also not sensitive to head movements. The disadvantage is that image processing is much more complex because of pattern recognition of the entire face to find the eyes.

The problems associated with this technique are primarily those of getting a good view of the eye. Normally the infrared sensitive camera is placed below the monitor and is tolerant of head movements in a certain range. A possible solution could be achieved by using tracking cameras that adapt to the position of the user's head. By using several, co-ordinated cameras the range over which the direction of gaze is registered, can be expanded.

4.3.5 Corneal Reflection and Eye Image using an Artificial Neural Network

This is one of the most recent developments. The raw material is still a digitised image of the user. The only image difference is that the entire head of the user is in the field of the camera. Pattern recognition is used to find the eyes. Once the eye is found, a rectangular part of the video image is feed to an Artificial Neural Network (ANN). To make it easier to find the eye

and helping decreasing computation time, an artificial visible light is placed in front of the user. Now the algorithm can search for a white reflection in a dark area, which is faster than any other pattern recognition using templates.

The disadvantage of this technique is the complexity of calibration. The ANN requires more than the simple calibration that is required by the other techniques but has the advantage of needing to be calibrated only once. It must be trained by gathering images of the user's eye and head for a couple of minutes while the user visually tracks a moving cursor on the display. No re-calibration on the next encounter is required.

A major disadvantage of this ANN-technique is the low accuracy compared to other techniques. The great advantage of this technique is the free head movement possibility of the user up to 30 cm. This could make it an attractive solution in places where a high degree of accuracy is not essential.

Results of research work can be found in [22].

4.3.6 Purkinje Image Tracking

Another possibility to determine the user's gaze by the Purkinje image tracking technique. The first and the fourth Purkinje images can be used for tracking the direction of gaze by the Dual-Purkinje image technique, which uses the relative position of these reflections to calculate the direction. This technique offers the highest accuracy and its sampling frequency is up to 40kHz. The disadvantage with this technique is that the fourth Purkinje image is rather weak, so the surrounding lighting must be heavily controlled. For more information and research results see [23] and [24].

4.4 Electro-Oculography (EOG)

4.4.1 Basics

Owing to the existence of a higher metabolic rate at the retina compared to the cornea, the eye maintains a voltage of +0.40mV to 1.0mV with respect to the retina. This potential difference causes an electric field and so surface electrodes placed on the skin around the eyes can then measure the corneoretinal potential. The electric field of the eye is similar to that of an electric dipole. The cornea is positive in vertebrates with respect to the posterior of the eye and resembles a battery situated in a variable-conducting medium. Moving your eyes results in rotation and translation of the entire eye so that the corneoretinal potential is roughly aligned with the optic axis. This enables the measurement of the user's gaze. The actual recorded potentials at the electrodes are smaller, normally in the range of 15μV to 200μV and have to be amplified before processing.
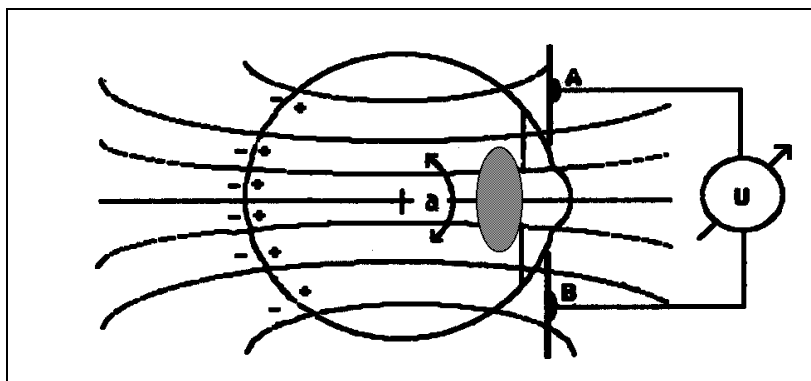
figure 10: Electric field of the human eye

Electro-Oculography has both important advantages and disadvantages over other eye tracking methods of course. Below, advantages and disadvantages of the EOG technique are listed.

Advantages over other systems:

- Head movements do not hinder recording

- Visualisation of the eye is not required

- Blinks do not produce interupt horizontal recordings

- Equipment necessary to record and store the small voltage deflections is relatively inexpensive and widely available.

- Angular deviations in eye position are easily related to the amplitude of the EOG potential

- High fidelity frequency response

- Eye movement velocity is directly obtainable from the record

- Synchronous recording from other psycho-physiological systems is easily obtained.

Drawbacks:

- Only radially accurate to within $\pm 2^{o}$, although accurate in the horizontal plane to within 1°. It is not uncommon for optical systems to be accurate to within 6 minutes of arc, with very sophisticated systems sensitive to 30 arc seconds.

- Independence from head movements makes actual input to the eyes hard to measure.

- Signal drifting

- Circumstantial procedure of electrode placing

- Sensitive to external electrical fields

On the positive side, the hardware equipment is inexpensive compared to other eye tracking techniques. A big advantage over all reflecting methods is that contact lenses or even gasses do not affect the measurements in any way. Another important fact is that any apparatus does not obstruct the visual field.

On the other hand, the main disadvantage of EOG is the drifting of the signal. There are some causes of drifting; some are listed below:

- Electrode polarisation

- Electrode slippage

- Changing skin resistance

- Variable corneoretinal potential due to light accommodation and level awareness

- Action potential of surrounding muscles

- Complex electrode placing

Because of the drifting signals the system must have an automatic gain and zeroing control to be a successful design. It might also be necessary to shield the whole system against electric fields, which can influence the small EOG-signals.

There is also a circumstantial procedure for placing the electrodes on the user's face. First areas must be found where action potentials of muscles have low importance. Then these

areas must be degreased with alcohol to achieve a good substratum. Then the area must be moistened with special electrode gel to make sure of a good contact between the skin and the electrodes. The electrode gel is also to avoid quick electrode polarisation, this depends on the electrodes and which gel can be used. So, set-up is cumbersome, and although actual discomfort is low, mental and physical awareness can be very high.

4.4.2 Eyetrans-Project

The Eyetrans project was part of my living exam at the technical school TGM in Vienna 1994. Investigations were made how eye movements can be related to cursor movements on a computer screen. The technique used in the Eyetrans project was based on EOG. The following schematic description shows the block diagram.
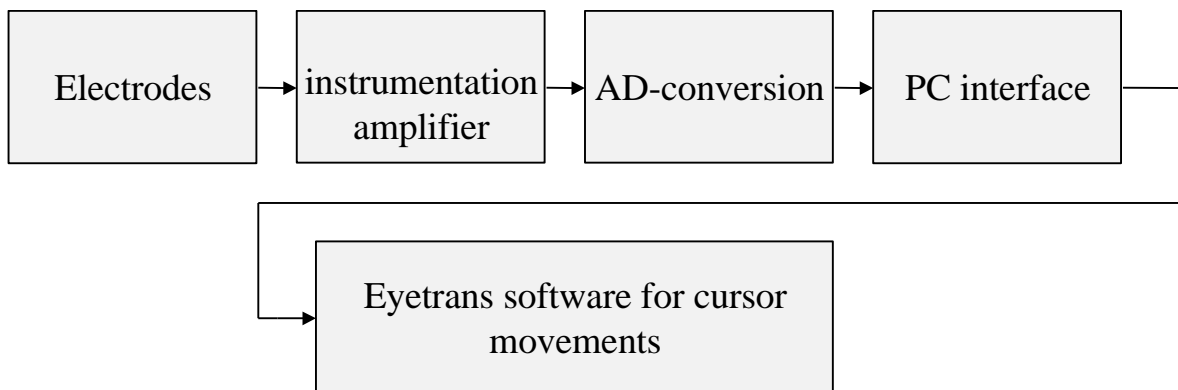
figure 11: Eyetrans schematic description

To detect the EOG signal a bipolar deduction technique was used. The electrodes were placed on the skin shown in figure 12.
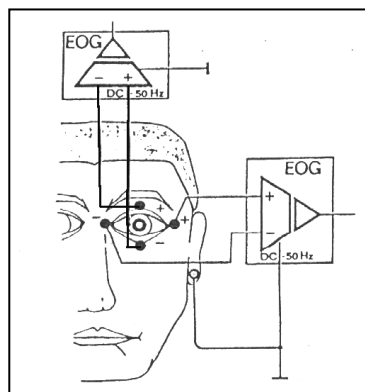
figure 12: Bipolar deduction technique

To meet the electrical - medical requirements an isolation amplifier was needed to make sure that no electric shocks can pass through the electrodes to the human skin coming from a failure of the electric circuit. Furthermore an additional security step was implemented by using a fibre optic connection between the PC-card and the AD-converter board. Using an ASIC (XILINX 3020) to transmit 4 independent input channels offered the possibility to measure other important signals (EKG,...) as well. The block diagram for the AD-conversion and transmitting stage is shown in figure 13.
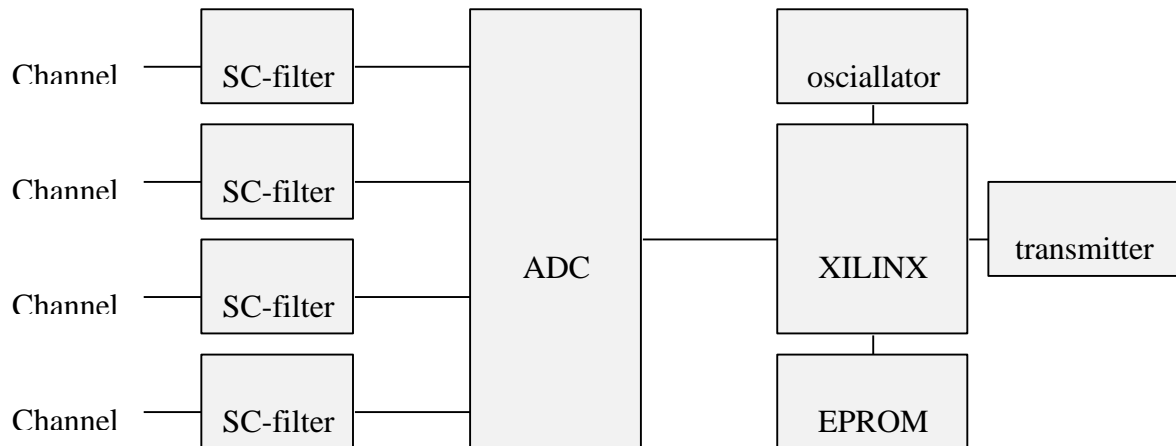
figure 13: Block diagram for AD concersion and transmitting

SC-filters (switched capacitor) were used to reduce the bandwidth for sampling of the incoming analogue signal. SC-filters are characterised by an extremely high slope of the filter curve. The ADC was an AD7874 which had the following specifications:

- 4 channels
- 12-bit conversion, $8\mu s$ conversion time per channel
- 10V input range
- 5V power supply voltage

The curve of the measured signal is illustrated in figure 14. The curve is determined by the movements of the eye. The amplitude of the resulting EOG signal is related to the deviation angle of the eye. The polarity and the amount of the amplitude correlates with moving the eyes to the left and right side or up and down.
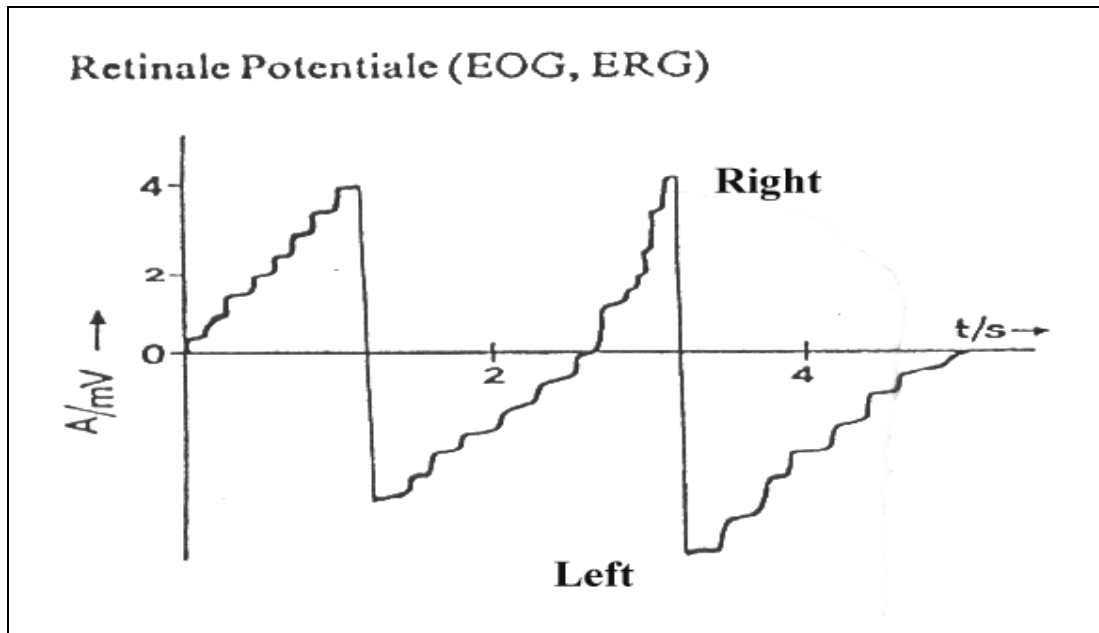
figure 14: EOG signals during reading

The EOG give useful diagnostic information about eye movements and the senso motoric integration of vision. Even when you fixate a certain point the eye moves periodically in a small area. The eye does not move continuously; there are always little steps (saccades) from one fixation point to the other. This is also shown in figure 14.

4.5 Techniques Based on Contact Lenses

4.5.1 The Sclera Coil

The subject wears a small copper coil, which is incorporated into a contact lens. The user has to sit in a high-frequency electromagnetic field. The current induced in the coil due to the angle it makes with the field gives its orientation, and hence by using a number of orthogonal fields positional information can be obtained.

This method offers highly accurate positional and temporal information, but its disadvantages include the high cost of the system. It is also inflexible to the experimental set-up and the discomfort to the subject due to its invasive nature. Experiments demonstrated that the average time for eye movement sessions were only around 30 minutes and some subjects reported eye infection. This method can hardly be seen feasible for everyday use, even the health issues concerning high-frequency electromagnetic fields are not yet resolved. Alone the requirement that the user has to wear special contact lens seems to make the technique too cumbersome for non-laboratory tasks.

4.5.2 Mirrored lens

The other possibility by wearing contact lenses is to modify the surface of the contact lens.. By engraving one or more plane mirror surfaces on the lens, the reflected light can be used for movement information. The same problem exists as with the sclera coil technique, that the user has to get used to wearing contact lenses.

4.6 Comparison of Tracking Techniques

The following overview demonstrates the advantages and drawbacks of the techniques described in the preceding chapters. The information in the table is mainly a combination of what Scott & Findlay [16], Wooding [17] and Baluja & Pomerleau [18] have written.

At the start of this section the dimension of recording compatibility has been written because any recording in principle is compatible, given enough computer power, with body localisation recordings. In some cases indicating the accuracy or range for both the horizontal and vertical directions, and also whether the technique is able to measure the pupil diameter. Finally listing the price range for the equipment.

Technique	Limbus tracking	Pupil tracking	Corneal/pupil relationship	Artificial neural network	Dual Purkinje Images	EOG	Contact lens
Face access	poor to good	good	good	good	Good	good	-
Subject contact	headmount/ chin rest	none	none	none	none	Elec- trodes	Contact lens
Image stabilising	-	-	-	-	Yes	-	-
Accuracy	h=0.5-7° v=1-7°	0.003°	0.5-2°	1.5°	0.017°	±1.5-2°	0.08°
Resolution Range	h=±15-30° v=±15-20°	h=±30-40° v=±20-40°	h=±12-40° v=±12-50°	-	±20-60°	±70°	±25°
Sampling speed	200-4000Hz	50-250Hz	25-50Hz	15Hz	4000Hz	-	1000Hz
Real-time-response	-	6-12ms delay	-	-	1ms delay	yes	-
Rotational Measurements	X/(Y)	X/Y	X/Y/Z	X/Y	X/Y	-	X/Y/Z
Pupil diameter Measurement	No	Yes	Yes	-	-	-	no
Translation Insensitivity	Poor	Good	-	Up to 30cm	Up to 25mm	-	-
Binocularly Extendable	Yes	Yes	-	-	-	-	-
Subject variety	low	Reasonable *	reasonable*	reasonable	-	-	-
Pricing	$2,700-48,000	$10,000-45,000	$22,000-120,000	-	$37,000-65,000	$2700-	$2700-

figure 15: Usability of present day eye-gaze techniques

* Some types of contact lenses can pose a problem, and bi-focal glasses are often problematic

Blanks indicate that no appropriate information was available.

4.7 Choosing an Eye Tracking System

Choosing the most suitable system for eye tracking systems is not a simple task. As shown in previous chapters every single technique has its advantages and drawbacks. This chapter should help to reduce the number of suitable systems which satisfy the requirements in best way.

Principle in reducing the selection of suitable systems are the following criteria:

- Suitability for operational conditions

- temporal and spatial accuracy

- cost

The conditions under which the system is operated will play an important part in determining which system is most suitable. Some criteria are listed below:

- *Place of operation*: Will the system be used under laboratory conditions or isolated and inaccessible places? Will it be powered by batteries and need to be therefore of low power consumption?

- *Head-free operation*: Has the equipment to be attached to the head? Is it insensitive to head movements? Is it unobtrusive?

- *Duration of recording*: How long will the subject have to wear the device? For example the discomfort of scleral coil methods limits subjects to ½ hour sessions. Other systems may limit duration due to drifting or calibration limitations.

- *Obstructions*: Is it necessary to wear spectacles, helmets or head-mounted displays? Does this affect image acquisition?

Every eye tracking technique and system have limits to the *spatial* and *temporal accuracy*. Video techniques or scleral coil systems have the edge on this area, typically with accuracy of greater than 0.5 degree in both directions. Infrared Oculography using limbus tracking is less accurate, particularly in vertical direction, but it offers high temporal accuracy. He/she difficulty with Electro Oculography is that it tends to drift which makes localisation of the point of the users gaze difficult. Another disadvantage is the sensitivity to muscle activities and electrical interference. If only saccadic latency or simply determining whether a saccade took place and its direction, then Infrared Oculography or EOG may be perfectly adequate.

Cost is not an indicator of suitability for purpose, and it is quite possible to pay a lot and get less useful piece of equipment than one which would have been cheaper and simpler. Budgets are limited to some degree so cost will have a part to play in the final decision when choosing a eye movement system.

4.8 Comments

This chapter described the present day eye tracking techniques. Different techniques to determine the eye gaze were introduced and shown in detail. The most common technique in our days is based on the cornea and pupil reflection relationship which provides best user comfort and high accuracy as well. Drawbacks and advantages of each technique were shown and summarised. Guidelines for choosing an eye tracking system were given in section 2.46. This chapter illustrates the background for existing eye tracking applications which are discussed in the next chapter.

5 Existing Eye Tracking Applications

5.1 Introduction

Today, several eye-gaze controllable applications exist. Most of them are designed for disabled people, as the use of their eyes is the fastest way to communicate with other people and manipulate their prosthetic devices.

This benefit of fast, immediate communication can also be applied to non-disabled people. Gradually eye-gaze based systems intended for everyone are being developed. This technique offers a broad range of new control mechanisms and is therefore the object of research activities and company product development. This chapter picks out some of the existing products and discusses their advantages, drawbacks and techniques used.

Eye-gaze can not only be used for interacting with computers; it also is suitable for other commercial products, witness the Canon company. They use the information from the user's eye gaze for auto focusing in their cameras. Eye-gazed interfaces can also be used to improve existing medical diagnostics as shown in this chapter. There is a wide range of new opportunities; some of them will be discussed later. This chapter should help to give an overview of state of the art facilities and achievements of modern eye gaze tracking.

5.2 Eye-Controlled Camera Focusing

Canon was one of the first companies to use the eye-gaze information to adjust the focus of a camera; it was introduced by the canon EOS 5 camera. This Canon technology lets the user focus on a subject anywhere within the field of view simply by looking at it. Off-centre focus, focus tracking, and focus shift techniques can all be achieved by the user moving their eye. This is achieved by using a CCD sensor in the camera's optical system. The system operates by illuminating the eye with an extremely low power, harmless infrared light. Reflections (see also chapter 4.3.2.2) of this light are then picked up and analysed by the camera, which is then able to determine where the user is looking. The gaze direction is found while the user is looking into the viewfinder and the nearest of 5 „eye-select" points shown through the lenses is selected. The camera then focuses on this area. Exposure is also linked to the focus point to ensure correct exposure of the subject. In addition to this certain functions can be operated by Eye-controlled switches.

The world's first eye-gaze controlled video camera is also manufactured by Canon and is called Canon UC-X1 Hi. The autofocus system is based on the system used in the Canon EOS 5 camera, explained above. It is an improved version where the limitation of 5 focus points no longer exists. It is simply focuses on the object gazed at. This video camera uses 4 infrared light sources. Two of them, for people wearing glasses, are placed further from the human eye than the other two. To detect if the user is wearing glasses or not, a calibration is needed when first used. The user is expected to look at two dots shown on the video display.

5.3 EyeMouse® by SensoMotorics Instruments GmbH

The EyeMouse® provides the computer users with the capability of moving the mousepointer with their eyes. A camera unit is placed in front of the user's eyes below the monitor. The user's eyes are illuminated by an area of infrared LEDs to provide independence from room lighting conditions. Normal head movements are possible and will not disturb the calibration parameters. The user has the flexibility to sit between 50cm and 1m from the computer screen. EyeMouse® uses the cornea-retina reflection relationship. All these advantages are paid for by additional expensive hardware, which calculates the centre of the pupil by using a

powerful digital signal processor. Like any other eye-gaze system, this too, must first be calibrated. To generate a mouse button the user has to fixate a location on screen for a short time that can be chosen individually.

The adaptive visual keyboard™ enables the user to enter text. Therefore, the user has to look

figure 16: EyeMouse by SensoMotorics Instruments GmbH

at key-cap icons. The design economises valuable screen real estate in order to maximise the screen area available for your application programs. To accomplish this, the choice of keys offered by the visual keyboard adapts automatically to the usage context, with special virtual keypads for text input, numerical input, dialog boxes and menu selections. Additional individual keypads can be designed by the user.

figure 17: Visual keyboard for EyeMouse

Product information is available at [25]

5.4 EyeLink ® by SensoMotorics Instruments GmbH

The EyeLink system is based, as most of its kind, on the reflection of infrared light. An Ethernet link connects the eye-tracking computer to an experimental-display computer. This guarantees real time data acquisition and control. Analyses of data into saccades, fixations and blinks are performed on-line and all data may be recorded into a file.

A lightweight headband-mounted eye-tracking system allows the subject free head motion.

figure 18: EyeLink ® by Senso-Motorics Instruments GmbH

Two high-speed cameras simultaneously take 250 images per second of both eyes to provide binocular eye tracking. A third camera is used to track 4 infrared markers mounted on the visual stimulus display for head motion compensation and true gaze position tracking. Two infrared LEDs provide the required illumination, which enables usage independent of ambient light conditions.

Further a single PC is needed to determine where the user is looking on the screen. All three images are synchronously analysed with extremely low noise and high-resolution (0.001^{o}).

More information can be found at [26].

5.5 Vision Control System's eye tracking device

Vision Control Systems (VCS) uses the latest camera technology, holographic optics and signal processing. Their system is based on the retina cornea relationship of infrared light. Figure 20 shows the principle:
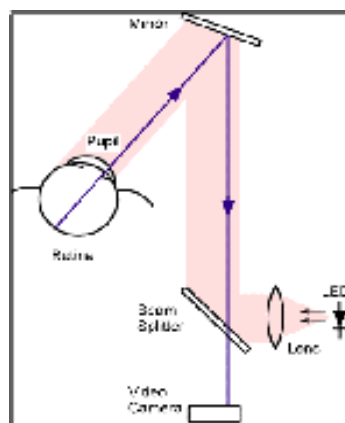
figure 19: Vision Control System's eye tracking device

The high resolution achieved with high reflectivity from the retina at infrared wavelengths against the surrounding iris can be improved by choosing the correct optical filter. VCS says, that simple uncollimated illumination does not produce optimal results unless intensity levels are undesirably high. Therefore, holographic optics is used to perform better results. Holographic optics match the infrared beam to the field of view of the camera to produce a

high-contrast ratio between the light retro-reflected from the retina and that from the surrounding eye surface. The camera is a 384 by 287 pixel CMOS camera-on-a-chip. The mayor benefit of CMOS cameras over CCD lies in the high level of product integration that can be achieved through implementing virtually all of the electronic camera functions onto the same chip. It also has only 1/3 of the power consumption of equivalent CCD cameras. The system works on an industry standard micro-controller for algorithm calculations. The eye movement is sent to the PC via an RS232 serial interface as either x, y co-ordinates or mouse commands.

Figure 21 illustrates the VCS eye-tracking device.
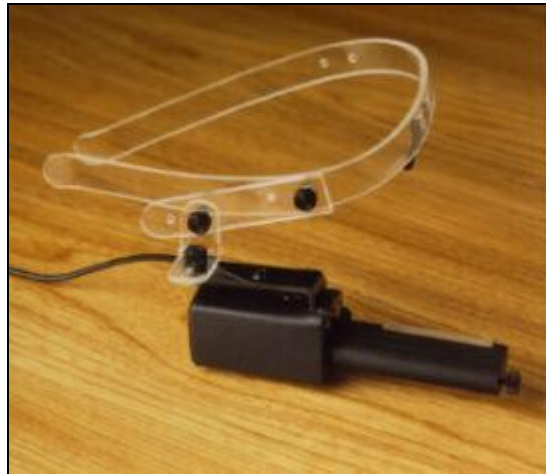
figure 20: VCS eye-tracking device

5.6 Digital chest radiograph

Results of an eye tracking device and its benefits in clinical diagnosis will be shown in this chapter.

Radiologists may overlook 30% of lung nodules even though the nodules were, in retrospect, clearly visible on previous radiographs, caused by properties of the image itself (for example low contrast,…).

The eye movement of a reader during interpretation of digital chest radiographs can provide valuable information for lesion detection. Measures of eye movements are associated with detection accuracy. During image interpretation , the reader's eye movements are tracked and recorded. This information is used to identify regions of prolonged dwell, which are then associated with false-positive detection of pulmonary nodules. Based on the location of the dwell regions, a small region of interest can be created within the image, where image processing (enhancement, zooming,…) takes place. Eye tracking combined with real-time radiology can be combined to improve image evaluation.

The next two images show the result. The digital chest radiograph shows in figure 21the overlaid gaze location data. Each observation of gaze location is indicated by a rectangle while the general direction of the gaze is indicated by an arrow. Dwells are indicated by accumulation of rectangles.
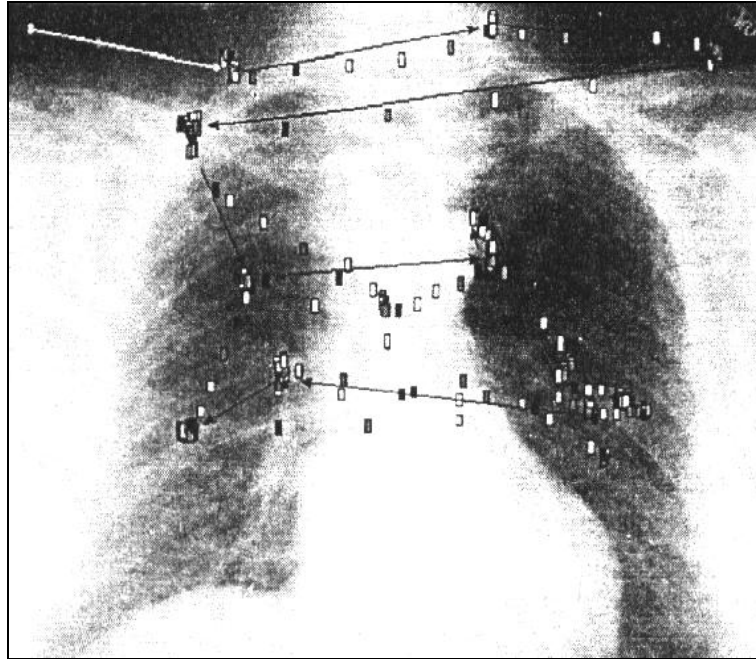
figure 21: Digital chest radiograph shows the overlaid gaze location data

figure 22 shows the three areas of longest dwell digitally enhanced. The dwell area to the left of centre has undergone simple enhancement of contrast and intensity. The area just to the right of the centre has undergone edge enhancement. The area farthest to the right of centre has undergone contrast enhancement
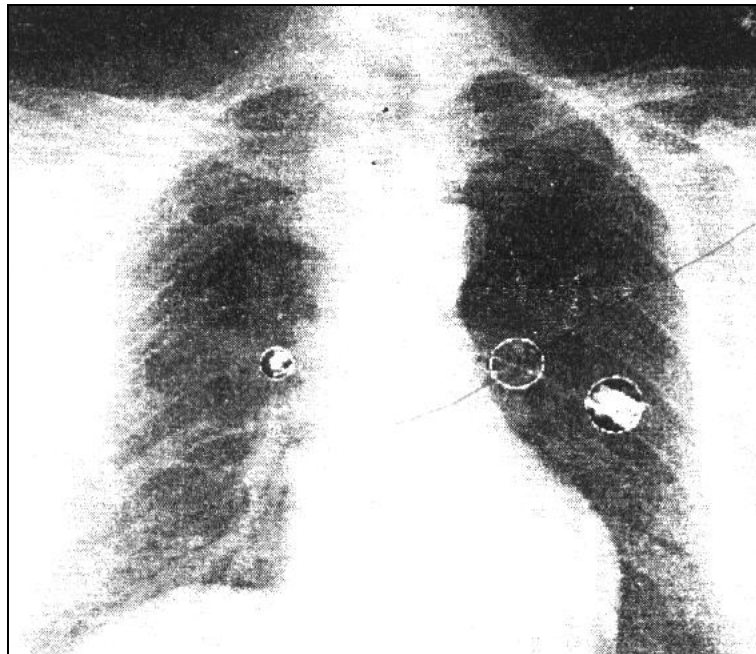
figure 22: Digital chest radiograph shows how eye movements data might be used after the reader is finished viewing the image

This is a new way of radilogic diagnosis providing powerful abilities to accelerate and improve diagnostic processes.

More information can be found at [27].

5.7 Eyegaze System, LC Technologies, Inc.

The Eyegaze System uses the traditional video tracking method, called the pupil centre/corneal reflection method, to determine the eye's gaze point on the computer display. A video camera, sensitive in the infrared range and positioned below the computer monitor continually acquires frames of the subject's eye. figure 23 represents the schematic drawing of the Eyegaze System. This system takes 30 pictures or 60 frames per second. An infrared LED is placed in the centre of the camera to illuminate the eye.

To allow the eyetracking software to run concurrently with an applications program it
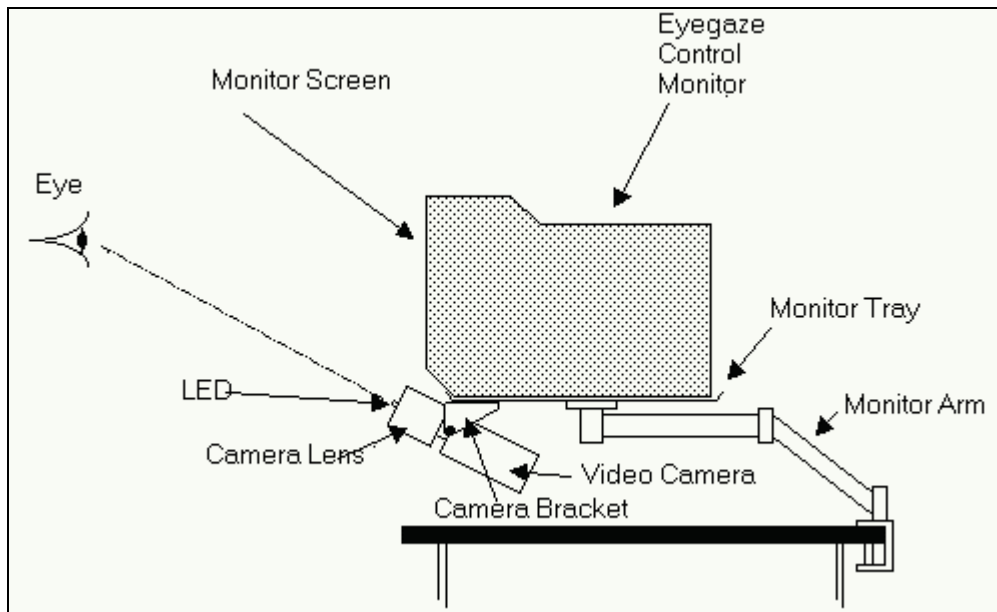
figure 23: Schematic drawing of the Eyegaze System

operates on an interrupt basis. The stay-resident interrupt service routine receives its data from the camera's frame-grabber board. This software requires between 200K and 240K of memory and takes between 30% and 50% of the available computer time. The most important specifications of the Eyegaze System will be discussed now:

The typical average bias error of the accuracy of the Eyegaze measurement with the head 50cm from the camera is 0.38cm. The tolerance to head motion with the eye within the field of view of the camera in the three ranges is 3.8cm lateral, 3cm vertical, and 1.27cm longitudinal. The LED has a wavelength of 880nm (near infrared) with a beam width of 20 degrees, between half power points. The radiated power is 20mW, radiated over the 20 degree beam with a safety factor of 5 (at a range of 38cm of the LED illumination on the eye is 20% of the HEW maximum permissible exposure).

The following image shows the Eyegaze System during operation. The left monitor shows the captured frame, the right one the calculated data and the middle one is responsible for interaction with the user who can lean his head on the seat to reduce the system's sensitivity to head movements.

figure 24: Eyegaze System during operation

More information is available at [28]

5.8 Flight Simulators

As a result of the financial cost of operating military aircraft, computer simulators are frequently used to augment tactical training. To achieve best training results, simulation requires a field of view similar to that offered by the bubble canopy of a modern fighter aircraft. Those images in the simulator are computer generated and the delay between frames is related to the size and complexity of the scene filling a large display with high-resolution image. Therefore, the computer-system behind must be very powerful. One technique for reducing the computational load is to reduce image resolution. Owing to the fact that only the fovea of the human eye has high resolution, a high-resolution scenery is only needed at the user's point of eye gaze.

The US Air Force uses eye tracking to align high-resolution **C**omputer-**G**enerated **I**magery (CGI) with a pilot's fixation point. Their system is based on the pupil cornea/retina reflection relationship and uses a fibre optic helmet mounted display. This colour display allows for a wide field of view, projecting low resolution CGI into each eye to produce a stereo view. A high-resolution insert is maintained in alignment with the fixation point.

5.9 Comments

Some latest products based on eye-gaze technique for disabled and advanced users have been evaluated in this chapter. The movements of human eyes provide a fast and precise input device (see flight simulator), especially when the user is thought to have difficulty in normal communication.

Most of the present eye gazed products use the pupil cornea/retina reflection relationship with infrared light to achieve best results. Those system only vary in the amount of needed computation power and therefore additional hardware and costs.

The state of the art products offer enough spatial and temporal accuracy to interact in a comfortable way with the (adapted) graphical user interface of operating systems. To achieve a successful launch on the market, the prices will have to drop to gain wide spread acceptance. Eye-gazed applications will become more important in future, not only for disabled people.

6   Image Processing

6.1 Introduction

Digital image processing encompasses a broad range of hardware, software and theoretical underpinnings. This chapter includes an introduction into the necessary basics required for understanding the VELOC image processing.

Figure 25 illustrates the basic concept of image processing.
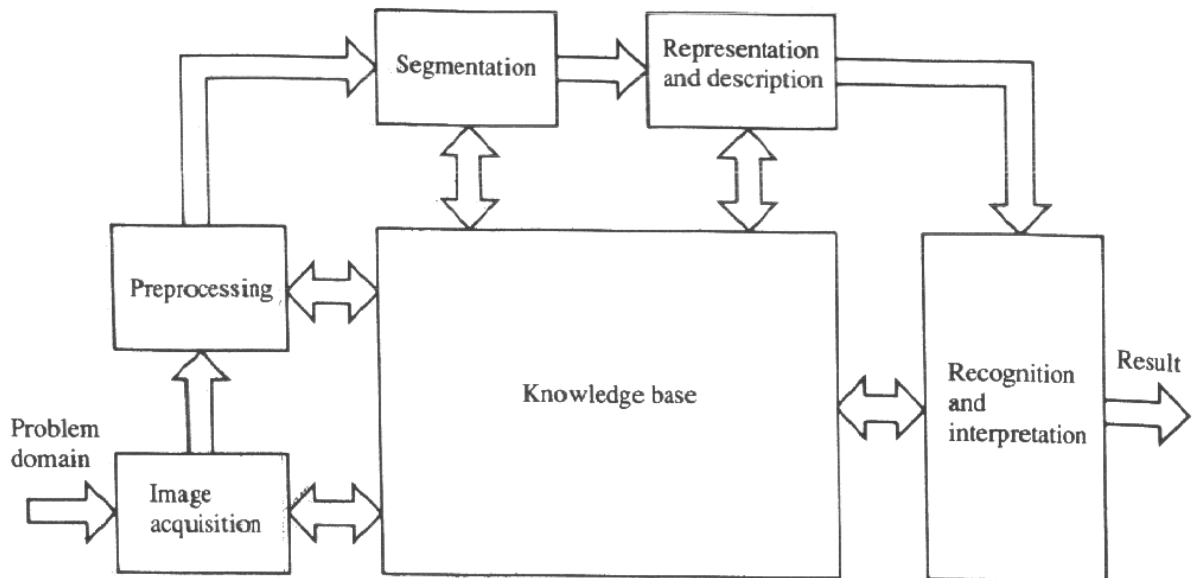
figure 25: Image processing steps

The problem domain in the VELOC system consists of acquired images of the user's eye; the objective is to determine the position of the centre of the pupil or iris. Thus the desired output in this case are the co-ordinates of the centre in the current image.

The first step in the process is *image acquisition*. This is achieved by using a CCD camera which will be discussed in detail in a further chapter.

After the digital image has been obtained, the next step deals with *pre-processing* that image. The key function of pre-processing is to improve the image in ways that increase the success of the other processes.

The next stage deals with *segmentation*. Broadly defined, segmentation partitions an image into its constituent parts or objects. In general, autonomous segmentation is one of the most difficult tasks in digital image processing. On the one hand, a robust segmentation procedure brings the process a long way toward successful solution of an imaging problem. On the other hand, weak or erratic segmentation algorithms almost guarantee eventual failure. The output of the segmentation stage is raw pixel data, constituting either the boundary of a region or all the points in the region itself. In either case, converting the data to a form suitable for computer processing is necessary.

The first decision that must be made is whether the data should be represented as a boundary or as a complete region. Boundary *representation* is appropriate when the focus is on external shape characteristics, such as corners and inflections. Regional representation is appropriate when the focus is on internal properties, such as textures or skeletal shape. However, in some applications these representations coexist; the VELOC image processing is a good example of

this. Description, also called feature selection, deals with extracting features for differentiating one class of objects from another.

The last stage involves *recognition and interpretation*. Recognition is the process that assigns a label to an object based on the information provided by its descriptors. Interpretation involves assigning meaning to an ensemble of recognised objects.

A very important part of image processing is the *knowledge base*. Knowledge about a problem domain is coded into an image processing system in the form of a knowledge database. This knowledge may be as simple as detailing regions of an image where the information of interest is known to be located. The knowledge base can also be quite complex, such as detailed information of what you are looking for. Each single stage of image processing uses the required information which is provided by the database. The more you know about the feature you wish to detect, the easier a successful search will be.

6.2 Pre-processing

6.2.1 Foundations

Thresholding is one of the most important approaches to image segmentation.

Suppose that the grey-level histogram shown in figure 26 corresponds to an image, f(x,y), composed of light objects on a dark background, in such a way that object and background
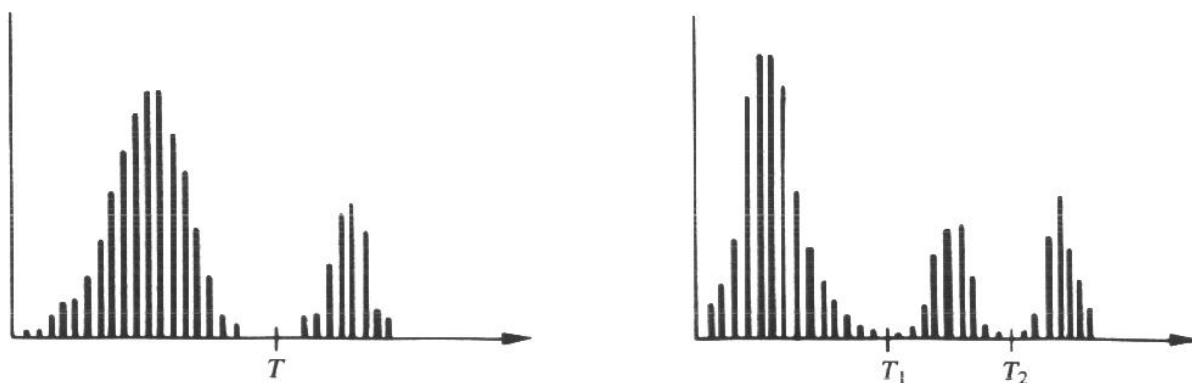
figure 26: Grey-level histograms that can be partioned by a single threshold or multiple thresholds

pixels have grey levels grouped into two dominant modes.

One obvious way to extract the objects from the background is to select a threshold T that separates these modes. Then, any point (x, y) for which $f(x, y) > T$ is called an object point; otherwise, the point is called a background point. figure 26 shows a slightly more general case of this approach. Here three dominant modes characterise the image histogram. The same basic approach classifies a point (x, y) as belonging to one object class if $T_1 < f(x, y) \le T_2$, to the other object class if $f(x, y) > T_2$, and to the background if $f(x, y) < T_1$. This type of multilevel thresholding is generally less reliable than its single threshold counterpart. The reason is the difficulty of establishing multiple thresholds that effectively isolate regions of interest, especially when the number of corresponding histogram modes is large and the illumination is changing. Typically, problems of this nature are best addressed by a single, variable threshold.

Based on the preceding discussion, thresholding may be viewed as an operation that involves tests using a function T of the form

$$T = T\left[x, y, p(x, y), f(x, y)\right] \tag{1}$$

Where $f(x, y)$ is the grey level of point (x, y), and $p(x, y)$ denotes some local property of this point – for example the average grey level of a neighbourhood centred on (x, y). A threshold image $g(x, y)$ is defined as

$$G(x, y) = \begin{cases} 1 & \text{if } f(x, y) > T \\ 0 & \text{if } f(x, y) \leq T \end{cases} \tag{2}$$

Thus pixels labelled 1 correspond to objects, whereas pixels labelled 0 correspond to the background. When T depends only on $f(x, y)$, the threshold is called global. If T depends on both $f(x, y)$ and $p(x, y)$, the threshold is called local. If, in addition, T depends on the spatial co-ordinates x and y, the threshold is called dynamic.

6.2.2 Simple Global Thresholding

With reference to the discussion in the previous chapter, the simplest of all thresholding techniques is to partition the image histogram by using a single threshold, T, as illustrated in figure 27. Segmentation is then accomplished by scanning the image pixel by pixel and labelling each pixel as object or background, depending on whether the grey level of that pixel is greater or less than the value of T. As indicated earlier, the success of this method depends entirely on how well the histogram can be partitioned.

With respect to the VELOC system, this approach will be shown in chapter 7.3.
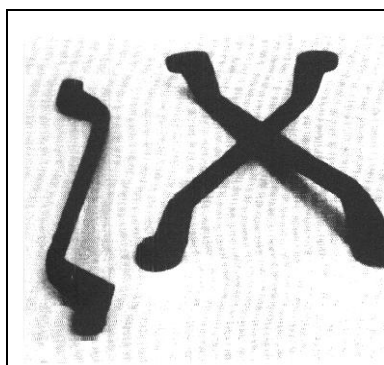
figure 27: Example of global thresholding; original image
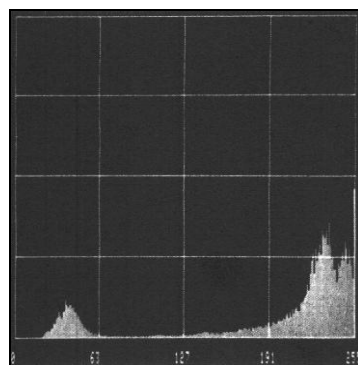
figure 28: Example of global thresholding; its histogram
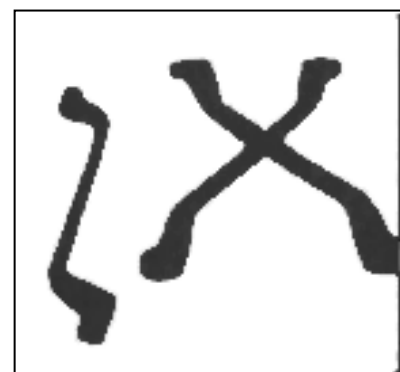
figure 29: Example of global thresholding; result of segmentation with T=90

In practice, the type of global thresholding just described can be expected to be successful in highly controlled environments. One of the areas in which this is often possible is in industrial inspection applications, where illumination control is usually feasible. Simple global thresholding within the VELOC system would result in insufficiently good results and therefore cannot be implemented.

6.2.3 Optimal Thresholding

Suppose that an image contains only two principal brightness regions. The histogram of such an image may be considered an estimate of the brightness probability density function $p(z)$. This overall density function is the sum or mixture of two unimodal densities, one for the light and one for the dark regions in the image. Furthermore, the mixture parameters are proportional to the areas of the picture of each brightness. If the form of the densities is

known or assumed, determining an optimal threshold for segmenting the image into brightness regions is possible.

Consider an image which contains two values combined with additive Gaussian noise. The mixture probability density function is

$$p(z) = P_1 p_1(z) + P_2 p_2(z) \tag{3}$$

which, for the Gaussian case, is

$$p(z) = \frac{P_1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{(z-\mu_1)^2}{2\sigma_1^2}\right] + \frac{P_2}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{(z-\mu_2)^2}{2\sigma_2^2}\right] \tag{4}$$

where μ_1 and μ_2 are the mean values of the two brightness levels, σ_1 and σ_2 are the standard deviations about the means, and P_1 and P_2 are the a priori probabilities of the two levels. The constraint

$$P_1 + P_2 = 1 \tag{5}$$

must be satisfied, so the mixture density has five unknown parameters. If all the parameters are known, the optimal threshold is easily determined. Suppose that the dark regions correspond to the background and the bright regions correspond to the objects. In this case $\mu_1 < \mu_2$, and the threshold T may be defined so that all pixels with a grey level below T are considered background points and all pixels with a level above T are considered object points. The probability of classifying an object point as a background point is

$$E_1(T) = \int_{-\infty}^{T} p_2(z)\,dz . \tag{6}$$

Similarly, the probability of classifying a background point as an object point is

$$E_2(T) = \int_{T}^{\infty} p_1(z)\,dz . \tag{7}$$

Therefore, the overall probability of error is

$$E(T) = P_2 E_1(T) + P_1 E_2(T) \tag{8}$$

To find the threshold value for which this error is minimal requires differentiating $E(T)$ with respect to T and equating the result to 0.

$$P_1 p_1(T) = P_2 p_2(T) \tag{9}$$

Applying this result to the Gaussian density, taking logarithms, and simplifying the quadric equation

$$AT^2 + BT + C = 0 \tag{10}$$

where

$$A = \sigma_1^2 - \sigma_2^2$$
$$B = 2\left(\mu_1 \sigma_2^2 - \mu_2 \sigma_1^2\right)$$
$$C = \sigma_1^2 \mu_2^2 - \sigma_2^2 \mu_1^2 + 2\sigma_1^2 \sigma_2^2 \ln\left(\sigma_2 P_1 / \sigma_1 P_2\right) \tag{11}$$

The possibility of two solutions indicated that two threshold values might be required to obtain the optimal solution. If the standard variations are equal, $\sigma^2 = \sigma_1^2 = \sigma_2^2$, a single threshold is sufficient:

$$T = \frac{\mu_1 + \mu_2}{2} + \frac{\sigma^2}{\mu_1 - \mu_2} \ln\left(\frac{P_2}{P_1}\right)$$

(12)

If the prior probabilities are equal, $P_1 = P_2$, the optimal threshold is the average of the means. The same holds for $\sigma = 0$.

A minimum mean square error approach may be used to estimate the parameters of an image from a histogram. For example, the mean square error e_{ms} between the mixture density $p(z)$ and the experimental histogram $h(z_i)$ is

$$e_{ms} = \frac{1}{n} \sum_{i=1}^{n} \left[p(z_i) - h(z_i) \right]^2$$

(13)

where an n-point histogram is assumed.

In general, analytically determining parameters that minimise this mean square error is not a simple matter. Even for the Gaussian case, the straightforward computation of equating the partial derivates to 0 leads to a set of simultaneous transcendental equations that usually can be solved only by numerical methods. Because the gradient is easily computed, a conjugate gradient or Newton's method for simultaneous non-linear equations may be used to minimise e_{ms}. Starting values must be defined. Assuming the a priori probabilities to be equal may be sufficient. Starting values may be found by dividing the histogram into two parts about its mean value, and computing means and variances of the two parts to be used as starting values.

As shown, optimal thresholding can consume a lot of computation time. Therefore it had to be adapted to the special needs of the VELOC system that will be discussed in chapter 7.3.

6.2.4 Threshold Selection Based on Boundary Characteristics

One of the most important aspects of thresholding selection is the capability of reliably identifying the mode peaks in a given histogram. This capability is particularly important for automatic threshold selection in situations where image characteristics can change over a broad range of intensity distributions.

One try for improving the shape of histograms is to consider only those pixels that lie on or near the boundary between objects and the background. An immediate and obvious improvement is that histograms would be less dependent on the relative sizes of objects and the background. For instance, the intensity histogram of an image composed of a large, nearly constant background area and one small object would be dominated by a large peak because of the high concentration of the background pixels. But, if only the pixels on or near the boundary between the objects and the background were used, the resulting histogram would have peaks of approximately equal height. In addition, the probability that one of those pixels lies on the object or on the background is nearly the same, thus improving the symmetry of the histogram peaks.

The principal problem with the approach just discussed is the implicit assumption that the boundary between objects and background is known. This information, of course, is subject to investigation on image segmentation and is not available at this stage in the algorithm procedure. However, the gradient information also includes information about edges in the image. In addition, use of the Laplacian can yield information regarding whether a given pixel

lies on the dark or light side of an edge. The average value of the Laplacian is 0 at the transition of an edge, so in practice the valleys of histograms formed from the pixels selected by a Gradient/Laplacian criterion can be expected to be sparsely populated. This value produces the highly desirable deep valleys discussed earlier.

The gradient and Laplacian information may be used to form a three-level image, as follows:

$$s(x, y) = \begin{cases} 0 & if & \nabla f < T \\ + & if & \nabla f \geq T & and & \nabla^2 f \geq 0 \\ - & if & \nabla f \geq T & and & \nabla^2 f < 0 \end{cases}$$
(14)

where the symbol 0, +, and − represent any three distinct grey levels, T is a threshold, and the gradient and Laplacian are computed at every point (x, y).

A modified version of this thresholding technique will be used by the VELOC system, which is discussed in chapter 7.3.

6.2.5 Comments

The sense of this chapter was to give an impression of different kinds of thresholding. Thresholding is one of the most critical parts in image processing. Thus, any wrong results produced in this stage can cause wrong calculations of the centre of the pupil. Owing to the fact, that thresholding can consume a lot of computation time it is also in the interest of the user, that this stage of algorithm does not take longer as it has to with regard to a real-time background-application. Therefore,, the threshold algorithm used by VELOC system is a combination of all three earlier discussed techniques to obtain best result at this stage.

6.3 Segmentation

This chapter will discuss edge detection, boundary detection and contour detection with regard to the Hough Transformation. It is not possible to cover all segmentation techniques in this thesis, so only those which are used will be explained.

6.3.1 Edge Detection

Edges are very important to any vision system (biological or machine).

- They are fairly cheap to compute

- They provide strong visual clues that can help the recognition process

- Edges are affected by noise present in an image

An edge may be regarded as a boundary between two dissimilar regions in an image. These may be different surfaces of the object, or perhaps a boundary between light and shadow falling on a single surface. In principle an edge is easy to find since differences in pixel values between regions are relatively easy to calculate by considering gradients.

Many edge extraction techniques can be broken up into two distinct phases:

- Finding pixels in the image where edges are likely to occur by looking for discontinuities in gradients. Candidate points for edges in the image are usually referred to as edge points, edge pixels, or edge lines.

- Linking these edge points in some way to produce descriptions of edges in terms of lines, curves etc.

First, this chapter will discuss the basics formulae, further mathematical backgrounds and finally several filters will be introduced.

6.3.1.1   Basic Formulations

An edge is the boundary between two regions with relatively distinct grey-level properties. Basically, the idea underlying most edge-detection techniques is the computation of a local derivative operator. Figure 30 shows an image of a light stripe on a dark background, the grey-level profile along a horizontal scan line of the image, and the first and second derivatives of this profile. Based on the fact that in digital images edges never appear abrupt (as a result of sampling), the profile is modelled as a smooth change of grey level.

Figure 30 also demonstrates that the first derivative of the grey-level profile is positive at the leading edge of a transition, negative at the trailing edge, and zero in areas of constant grey level. The second derivative is positive for that part of the transition associated with the dark side of the edge, negative for that part of the transition associated with the light side of the edge and of course zero in areas of constant grey level.
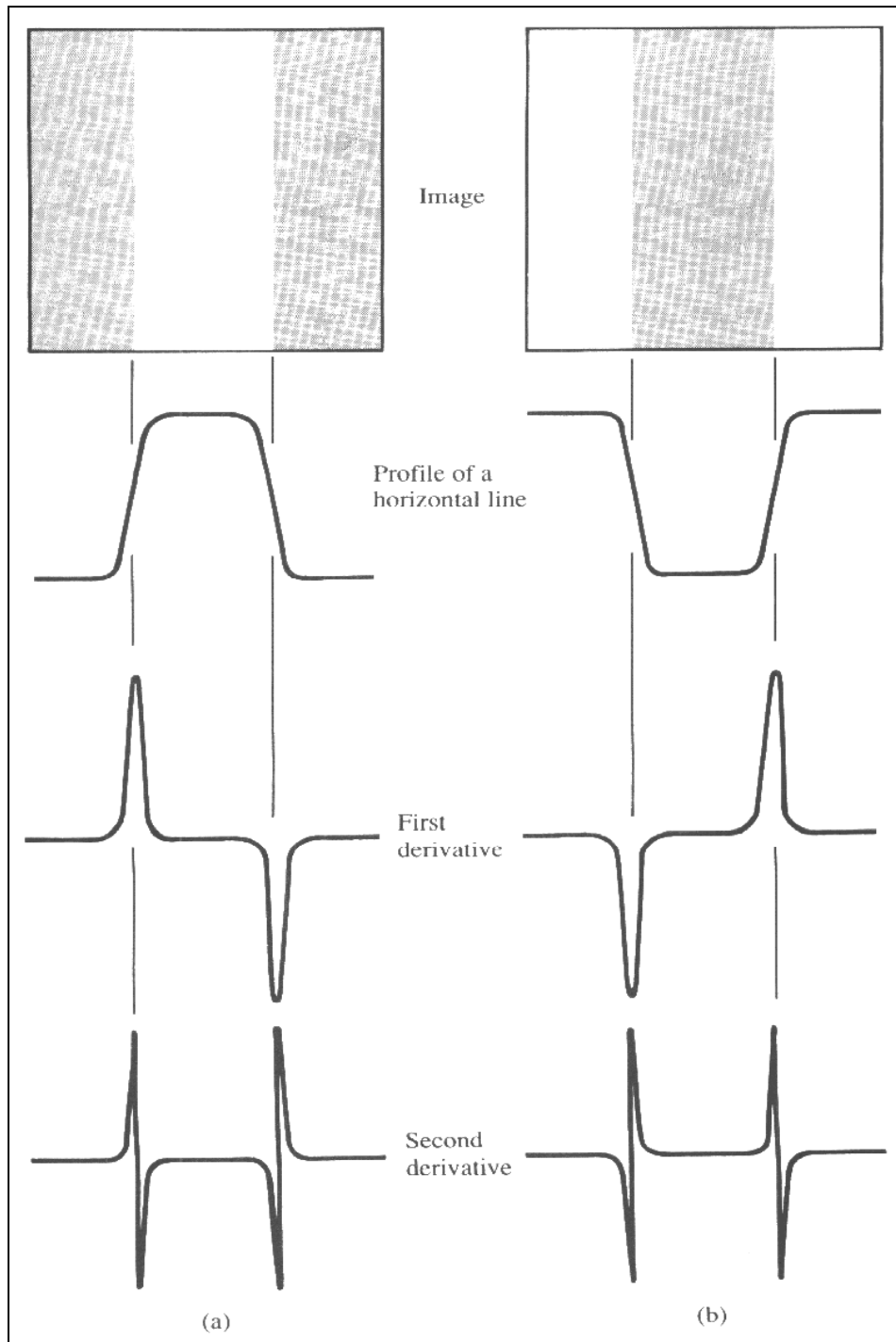
figure 30: Edge detection by derivative operators: (a) light stripe on a dark background; (b) dark stripe on a light background. Note that the second derivative has a zero crossing at the location of each edge

Hence the magnitude of the first derivative can be used to detect the presence of an edge in an image; the second derivative can be used to determine whether an edge pixel lies on the dark or light side of an edge. It is important to note that the second derivative has a zero crossing at the midpoint of a transition in grey level, which can provide a powerful approach to locating edges in an image.

6.3.1.2 Gradient Operations

The gradient is a vector, whose components measure how rapidly pixel values are changing with distance in the x and y directions. The gradient of an image $f(x, y)$ at location (x, y) is the vector

$$\nabla f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f}{\partial x} \\ \dfrac{\partial f}{\partial y} \end{bmatrix}$$

(15)

It is well known from vector analysis that the gradient vector points in the direction of maximum rate of change of f at (x, y). Thus, the components of the gradient may be found using the following approximation:

$$\frac{\partial f(x, y)}{\partial x} = \Delta x = \frac{f(x + dx, y) - f(x, y)}{dx}$$

$$\frac{\partial f(x, y)}{\partial y} = \Delta y = \frac{f(x, y + dy) - f(x, y)}{dy}$$

(16)

where dx and dy measure distance along the x and y directions respectively. In (discrete) images we can consider dx and dy in terms of numbers of pixels between two points. Thus, when $dx = dy = 1$ (pixel spacing) and we are at the point whose pixel co-ordinates are (i, j) we have

$$\Delta x = f(i + 1, j) - f(i, j)$$

$$\Delta y = f(i, j + 1) - f(i, j)$$

(17)

In edge detection an important quantity is the magnitude of the gradient vector, generally referred to simply as the gradient and denoted ∇f, where:

$$\nabla f = mag(\nabla f) = \left[G_x^2 + G_y^2 \right]$$

(18)

This quantity equals the maximum rate of increase of $f(x, y)$ per unit distance in the direction ∇f. Common practice is to approximate the gradient with absolute values:

$$\nabla f = |G_x| + |G_y|$$

(19)

which is much easier to implement, particular with dedicated hardware.

Not only the magnitude, also the direction of the gradient vector is an important quantity. Let $\alpha(x, y)$ represent the direction angle of the vector ∇f at (x, y). Then from vector analysis,

$$\alpha(x, y) = \tan^{-1}\left(\frac{G_y}{G_x} \right)$$

(20)

where the angle is measured with respect to the x-axis.

Implementation:

The difference operators in equation (17) correspond to convolving the image with the two masks Δx and Δy.

This is easy to compute:

- The top left-hand corner of the appropriate mask is superimposed over each pixel of the image in turn,

- A value is calculated for dx or dy by using the mask coefficients in a weighted sum of the value of pixel (i, j) and its neighbours.

- These masks are referred to as convolution masks or sometimes convolution kernels.

Instead of finding approximate gradient components along the x and y directions we can also approximate gradient components along directions at 45° and 135° to the axes respectively. In this case the following equations are used:

$$\Delta_1 = f(i+1, j+1) - f(i, j)$$
$$\Delta_2 = f(i, j+1) - f(i+1, j)$$

(21)

This form of operator is known as the Roberts edge operator and was one of the first operators used to detect edges in images. The corresponding convolution masks are given by:

Many edge detectors have been designed using convolution mask techniques, often using mask sizes 3x3 or even larger. An advantage of using a larger mask size is that errors due to the effects of noise are reduced by local averaging within the neighbourhood of the mask. An advantage of using a mask of odd size is that the operators are centred and can therefore provide an estimate that is biased towards a centre pixel (i, j).

6.3.1.3 Laplacian

All of the previous edge detectors have approximated the first order derivatives of pixel values in an image. It is also possible to use second order derivatives to detect edges. A very popular second order operator is the Laplacian operator. The Laplacian of a function $f(x, y)$, denoted by $\nabla^2 f(x, y)$, is defined by:

$$\nabla^2 f(x, y) = \frac{\partial^2 f(x, y)}{\partial x} + \frac{\partial^2 f(x, y)}{\partial y}$$

(22)

Once more, discrete difference approximations can be used to estimate the derivatives and represent the Laplacian operator with the convolution mask shown in figure 32:

figure 31: Laplacian operator

However there are disadvantages to the use of second order derivatives.

- As a second-order derivative, the Laplacian typically is unacceptably sensitive to noise.

- No directional information about the edge is given.

- The Laplacian produces double edges

The problems that the presence of noise causes when using edge detectors means that attempts should be made to reduce the noise in an image prior to or in conjunction with the edge detection process.

A more general use of the Laplacian is in finding the location of edges using its zero-crossings property. The zero crossings offer reliable edge detection. The price paid for these advantages is increased computational complexity and time.

6.3.1.4 Derivate Filters

Averaging of pixels over a region tends to blur detail in an image. As averaging is analogous to integration, differentiation can be expected to have the opposite effect and thus sharpen an image. In this chapter gradient filters will be discussed.

The first filter to be discussed is the *Sobel-filter*. The Sobel operator has the advantage of providing both a differencing and a smoothing effect. Because derivatives enhance noise, the smoothing effect is a particularly attractive feature of the Sobel operator. The figures 32-34 show the computation mask for a Sobel filter:

figure 32: 3x3 image region

figure 34: mask used to computate Gx at centre point of 3x3 region

figure 334: mask used to computate Gy at centre point of 3x3 region

From the figures 32-34, derivatives based on the Sobel operator mask are

$$G_y = \left(z_1 + 2z_2 + z_3 \right) - \left(z_7 + 2z_8 + z_9 \right)$$
$$G_x = \left(z_3 + 2z_6 + z_9 \right) - \left(z_1 + 2z_4 + z_7 \right)$$

(23)

Where the z's are the grey levels of the pixels overlapped by the mask at any location in an image. In the next step the magnitude and direction of the gradient vector has to be determined. To get the next value, the masks are moved to the next pixel location and the procedure is repeated. Thus, after the procedure has been completed for all possible locations, the result is a gradient image of the same size as the original image. The Sobel operator represented in figure 34 is a 3x3 mask, but there are still other mask sizes as well, which have some good characteristics. The main aim is to approximate the ideal gradient operation as closely as possible. This objective is especially important for the gradient direction, because even small errors may have a big impact of the results of successive processing steps. From this point of view, a 3x3-gradient operation should not be used. In practice, a 5x5 mask has proved to be a good choice. Larger masks yield only marginally better results, while consuming far more computation time. In figure 35 and figure 36, 5x5 and 9x9 Sobel masks are shown.

10	17	20	17	10
10	17	20	17	10
0	0	0	0	0
-10	-17	-20	-17	-10
-10	-17	-20	-17	-10

-10	-10	0	10	10
-17	-17	0	17	17
-20	-20	0	20	20
-17	-17	0	17	17
-10	-10	0	10	10

figure 35: Sobel 5x5 Gy mask　　　　　**figure 36: Sobel 5x5 Gx mask**

On the left side a mask to compute Gx and on the right side a mask to compute Gy at centre point of the 5x5 region is shown.

On the left side a mask to compute Gx and on the right side a mask to compute Gy at centre point of the 9x9 region is shown

-14	-14	-14	-14	0	14	14	14	14
-29	-29	-29	-29	0	29	29	29	29
-44	-44	-44	-44	0	44	44	44	44
-60	-60	-60	-60	0	60	60	60	60
-71	-71	-71	-71	0	71	71	71	71
-60	-60	-60	-60	0	60	7	60	60
-44	-44	-44	-44	0	44	44	44	44
-29	-29	-29	-29	0	29	29	29	29
-14	-14	-14	-14	0	14	14	14	14

14	29	44	60	71	60	44	29	14
14	29	44	60	71	60	44	29	14
14	29	44	60	71	60	44	29	14
14	29	44	60	71	60	44	29	14
0	0	0	0	0	0	0	0	0
-14	-29	-44	-60	-71	-60	-44	-29	-14
-14	-29	-44	-60	-71	-60	-44	-29	-14
-14	-29	-44	-60	-71	-60	-44	-29	-14
-14	-29	-44	-60	-71	-60	-44	-29	-14

figure 38: Sobel 9x9 Gx mask　　　　　**figure 378: Sobel 9x9 Gy mask**

The Prewitt operator is another possibility for edge detection. This filter does not offer the capability of smoothing but is used quit often. Note that all masks coefficients sum to 0, indicating a response of 0 in constant areas, as expected of a derivative operator. The figures 40, 41 illustrate a 3x3 Prewitt filter:

-1	0	1
-1	0	1
-1	0	1

1	1	1
0	0	0
-1	-1	-1
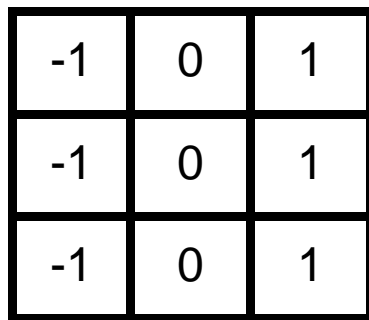
figure 40: Prewitt operator used to computate Gy at centre point of 3x3 region
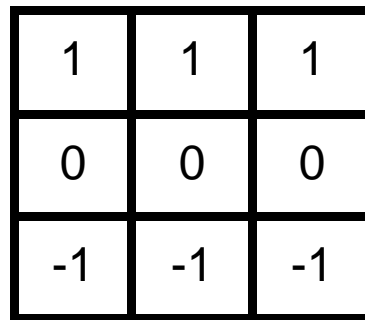
figure 39: Prewitt operator used to computate Gy at centre point of 3x3 region

6.3.2 Boundary Detection

The methods discussed in the previous chapter ideally should yield pixels lying only on the boundary between regions. In practice, this set of pixels seldom characterises a boundary completely because of noise, breaks in the boundary from nonuniform illumination (see also chapter 4.3.2.1) and many other effects causing spurious intensity discontinuities. Edge-Thinning and the 4- to 8-Transformation are two succeeding image-processing stages to determine an exact and correct shape.

6.3.2.1   Edge-Thinning

Gradient operations usually cause gradient magnitude distributions that are similar to an extended mountain edge presented in figure 41:

before   after 'thinning'
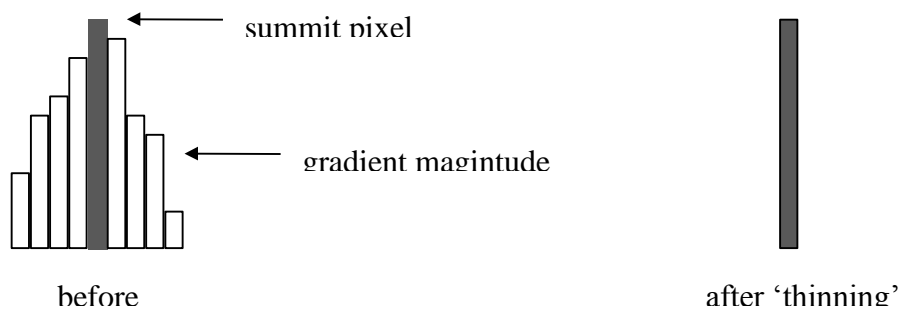
figure 41: Edge Thinning

Those pixels, which have the highest gradient magnitude are called summit pixels. Figure 41 also illustrates that using only the summit pixels would cause a sharp edge for the succeeding algorithm stages. This means that the slopes of the ridges on the left-hand and right-hand sides of the summit pixel are superfluous and therefore can be removed. This approach is also called skeletonising.

At the end of the thinning algorithm the contours are only 1 pixel wide. In order to find these local maximum magnitudes (summit pixels), it is necessary to determine the left-hand and right-hand side neighbours of every gradient pixel. The algorithm to achieve this is presented in chapter 7.3.5.

6.3.2.2   4-to 8-connected Transformation

The edge-thinning algorithm achieves a contour, which is indeed only one pixel wide. The only drawback with this algorithm is that the contour is 4-connected. This characteristic denotes that the different elements in the chain are only permitted to have vertical and horizontal orientation. This means that it is only possible to realise 90° edges. The following illustration clarifies this content.
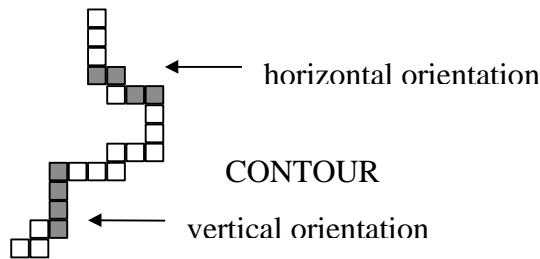
figure 42: Vertical and horizontal orientations

As you can see, redundant elements are included causing distortions during the contour extraction stage. Thus, the result of this algorithm stage should yield a contour, which is 1 pixel wide in all directions. This is the reason why a 4- connected to 8-connected transformation is needed. The reason for 8-connected is that a digital image pixel has exactly 8 neighbours. To figure 42, a 4-to 8- connected Transformation is applied and shown in figure 43.
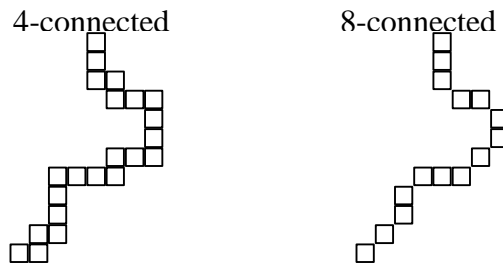
figure 43: 4-connected to 8-connected transform

The algorithm implementation for the VELOC system is presented in chapter 7.3.5.

6.3.3 Contour Detection

This algorithm stage is necessary because computers do not have any intelligence to recognise that several points or curves belong to a certain contour. The computer only sees pixel and needs an algorithm for joining diverse points to a single contour-list.

6.3.3.1   Linking contour points

This operation takes advantage of the previous step, the 4- to 8- connected transformation. Therefore, the first contour point has to be found. The next pixel within the chain will be around its 8 neighbours. If this fails for all of the eight neighbours, the pixel is the last one in the chain. The following figure should illustrate this procedure.

first pixel

neighbour

last pixel

□ chain 1

■ chain 2
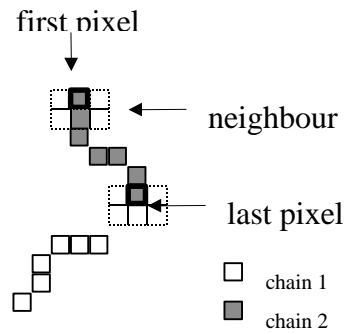
figure 44: Identification of pixels belonging to the same contour

6.3.3.2   Hough Transformation

Detecting geometric primitives in captured images is one of the basic tasks of computer vision and is not a simple exercise. The Hough Transformation is one of many possible ways to achieve this and is very popular for its conceptual simplicity.

The Hough Transformation is a main part of the eye gaze detection algorithm and so a discussion about this special transformation is proposed. First the Hough Transformation is introduced with regard to straight lines which will be adapted to circles later; these were detected in the VELOC system.

- Concept

The basic idea should be demonstrated on straight lines because of its simplicity. Suppose that, for n points in an image, we want to find subsets of these points that lie on straight lines. One possibility of course is to find all lines determined by every pair of points and then find all subsets of points that are close to particular lines. The problem with this procedure is that it involves finding $n(n-1)/2 \sim n^2$ lines and then performing $n^2(n-1)/2 \sim n^3$ comparisons of every point to all lines. This approach wastes a lot of computation time and would never allow economic development of an eye gaze system in real time, except using a high end computer.

An alternative approach was proposed by Hough [19] in 1962, commonly referred to as the Hough Transformation. Consider a point (x_i, y_i) and the general equation of a straight line. The problem is that there are infinitely many lines passing through the point (x_i, y_i), all satisfying the equation

$$y_i = ax_i + b \qquad (24)$$

for different values of a and b.

Writing this equation as $b = -x_i a + y_i$ and considering the parameter space, you get the equation of a single line for a certain point determined by (x_i, y_i). A second point (x_j, y_j) has a line in parameter space associated with it as well. This line intersects the first line with (x_i, y_i) at (a', b') where a' is the slope and b' the intercept of the line containing both (x_i, y_i) and (x_j, y_j) in the xy-plane. In fact, all points contained on this line have lines in parameter space that intersect at (a', b'). figure 45 shows this concept.
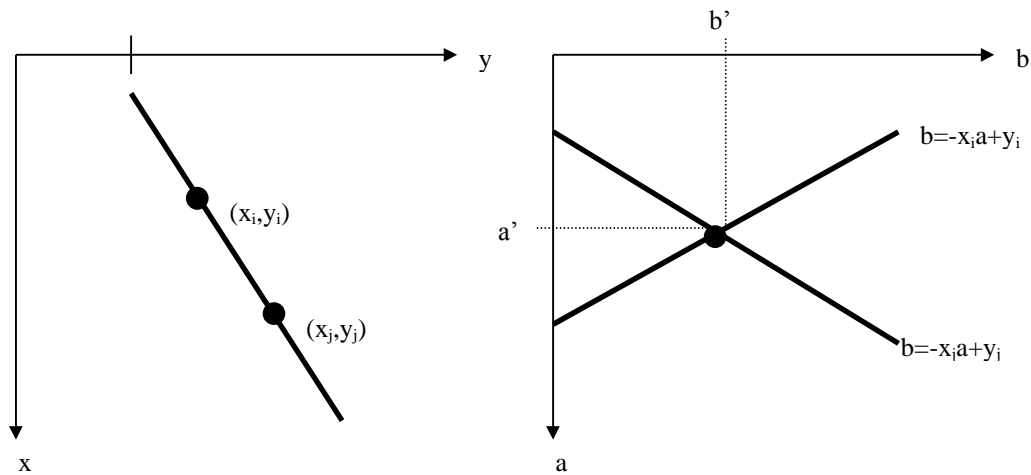
figure 45: xy- plane and parameter space

To adapt this transformation to a computer system the so-called accumulator cells are used; (a_{min}, a_{max}) and (b_{min}, b_{max}) are the expected ranges of slope and intercept values. The cell at co-ordinates (i, j), with accumulator value $A(i, j)$ corresponds to the square associated with parameter space co-ordinates (a_i, a_i) See also figure 46.
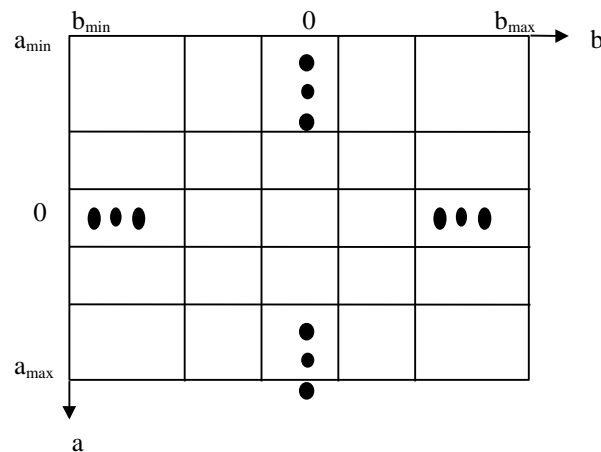
figure 46: Quantisation of the parameter plane for use in the Hough Transformation

Of course, all values of $A(i, j)$ must be set to zero when the algorithm is started. Then, for every point (x_i, y_i) in the image plane we let the parameter equal each allowed value between a_{min} and a_{max} and calculate the b parameter by using the equation

$$b = -x_i * a + y_i$$ (25)

This b value is then checked, if it is in the allowed range. Afterwards, if both parameters are in the allowed range, the accumulator cell $A(i, j)$ corresponding to a' and b' is incremented by 1. At the end of this procedure, the accumulator value of X in $A(i, j)$ represents exactly the numbers of points lying on the straight line with the values for a and b. The accuracy of the collinearity of these points is determined by the number of subdivisions in the parameter space.

The computation is much less intense than the first approach with n^3. The Hough Transformation needs for every point (x, y) exact K calculations for b when a is given. K is the number of subdivisions of the a axis. With n image points this method involves nK computations, so the Hough Transformation is linear.

This algorithm works well when the lines do not approach the vertical. With a vertical, both the slope and the intercept become infinite. One way around this difficulty is to use the normal representation of a line determined by the following equation:

$$x \cos \delta + y \sin \delta = \rho$$

(26)

Figure 47 shows the meaning of the parameters used above. It uses exactly the same procedure except δ is substituted for a and ρ for b. This causes sinusoidal curves in the parameter space instead of straight lines. Incrementing δ and solving the equation for the corresponding ρ gives M entries in accumulator $A(i, j)$ associated with the cell determined by (δ_i, ρ_i). The range of angle d is $\pm 90°$, measured with respect to the x axis.

The variable d can be calculated very easily by manipulating the original image with a derivation filter. This will be discussed in a later chapter. This means, that the variables x, y, and δ are known and consequently we can deduce the value for ρ. The following example shown in figure 47 demonstrates the procedure.

In figure 47 a section of a thinned gradient image is shown. It can be proved easily by
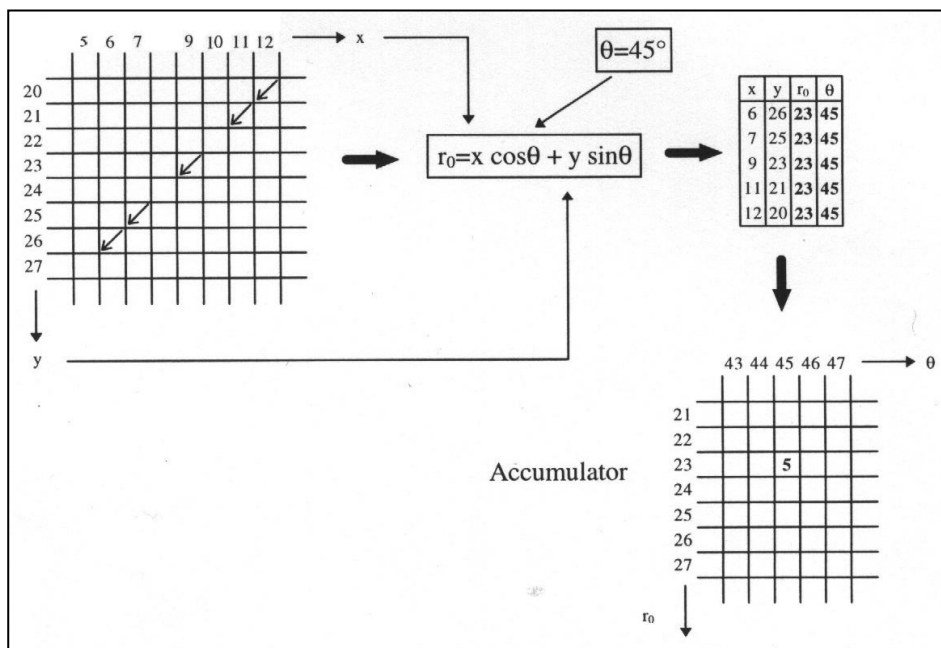
figure 47: Example of using the Hough Transformation

entering different values for x, y and δ.

The last step is to analyse the Hough accumulator. As already said, all values of the accumulator must be set to zero at the beginning of the algorithm. Thus, an accumulator value greater than 1 represents at least 2 contour points lying on the line. Of course, the greater the number in the cell, the greater the possibility of having detected a straight line in the image. Hence, a threshold value must be set in order to keep only those lines with high probability of existence in the image. This Threshold has to be found empirically.

- Circle Detection

Due to the fact that the VELOC system should determine the user's eye gaze, it was necessary to detect the centre of the pupil. Therefore, circles in the captured image had to be found, because circles can model the iris. Another benefit of the Hough transformation is its capability of being applied to any function of the form $f(v,c) = 0$, where v is a vector of co-ordinates and c is a vector of coefficients. To detect circles, a circle-to-point transformation is needed. The description of a circle in parametric form is

$$(x - x_c)^2 + (y - y_c)^2 = c^2,$$

<div align="right">(27)</div>

where x, y represent the co-ordinates of a pixel, x_c, y_c the centre of the circle and c the radius. All the points belonging to the perimeter of the circle have this characteristic. See also figure 48.
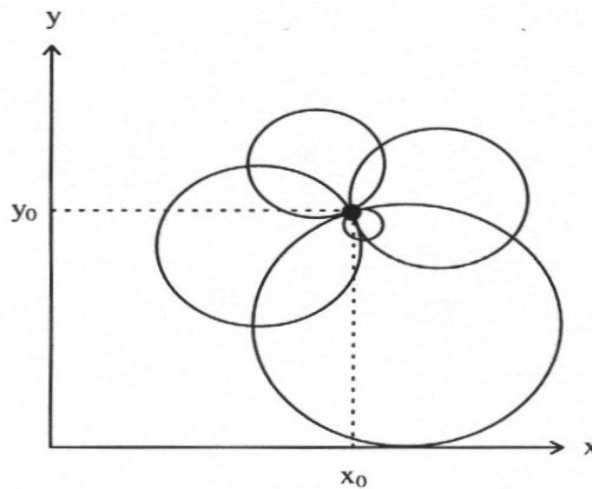
<div align="center">figure 48: Some possible circles for a given point</div>

The basic difference between detecting straight lines and circles is the presence of three parameters (x_c, y_c, and r). This results in a 3-D parametric space with cubelike cells and an accumulator of the form $A(i,j)$. The following figure demonstrates the problem of finding the centre of a circle when a point with its two co-ordinates is given.

The use of a 3-D parametric space also increases computation time, as for each single given point with the co-ordinates (x,y) every possible circle must be detected. Therefore, the procedure is to increment x_c and y_c and then check if the calculated value of the radius exceeds the allowed range. To save computation time, a knowledge of possible pupil centres and radius values is necessary. This information will be determined during calibration as described later.

- Ellipse Detection

As said in the previous chapter, the Hough Transformation can be applied to any function of the form $f(\cdot, c) = 0$. Therefore, it is also possible to detect an ellipse. The equation of an ellipse is

$$\left(\frac{x - x_1}{a}\right)^2 + \left(\frac{y - y_1}{b}\right)^2 = 1 \qquad (28)$$

With respect to.
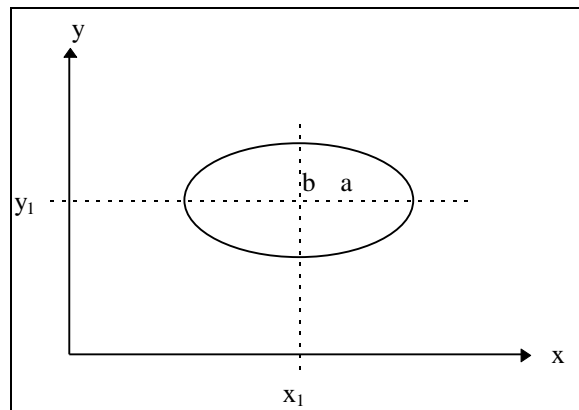
figure 49: figure of an ellipse

Solving this equation for b results in

$$b^2 = \frac{(y - y_1)^2}{1 - \left(\frac{x - x_1}{a}\right)^2} \qquad (29)$$

This equation can be used to determine b with the Hough Transformation. It is very similar to circle detection, the only difference is that one more dimension is needed because the radius r has to substituted by a and b. The Hough space will increase to 4 dimensions. This of course causes a dramatic increase of computation time.

- Comments

The objective of this chapter has been to review the abilities the Hough Transformation is offering. It is a powerful tool to determine graphical objects in a digital image. Its theoretical ideas and extension to circle and ellipse detection have been reviewed, as well as the concept of the Hough space. The Hough Transformation is the ideal solution for the VELOC system because of its simplicity and relatively short computation time compared to other detection algorithms.

The final algorithm design of the Hough Transformation will be discussed in chapter 7.3.9

6.4 Comments

The purpose of this chapter was to introduce the image processing used by the VELOC system. Image processing in general is of course much more complex and so only those areas which are implemented in the VELOC system could be established. Image processing is the most important software part used by the system and its implementation is described in chapter 7.3.

7   VELOC System

7.1 Introduction

Previous chapters described the theoretical foundations used in the VELOC system. This chapter describes the adopted practical solution and its implementation to reach a working system. This chapter will explain in detail image acquisition, the image processing algorithm and also the Windows 95 standard VfW (Video for Windows) and MCI (Multimedia Command Interface).

7.2 Hardware and Software Issues

7.2.1 PC requirements

The VELOC system uses a Connectix QuickCam product for image acquisition. QuickCam works on every Windows-based PC. The PC has to be equipped with at least a 368 processor to run QuickCam but to handle the image processing step within an adequate time (real-time) a 486 processor is preferred. The VELOC system was developed for a Windows 95 machine where a Video for Windows driver has to be installed. Further one parallel port must be available. This port may be unidirectional (compatible), bi-directional (enhanced) or an enhanced capability port. The difference between these ports is the speed of data transmission. For this purpose a „fast„ PC is one that is capable of handling 15 frames per second (fps) at the 160x120 pixel window size, used in the VELOC system. The software automatically adapts to the fastest speed the PC can handle.

7.2.2 Connectix Camera

7.2.2.1   Camera specifications

The following are QuickCam's specifications:

- CCD camera

- Up to 320x240 pixel capture.

- 4-bit grey scale (16 shades of grey) or 6-bit grey scale (64 shades of grey)

- Frame rates of up to 24 fps (varies by window size being recorded and the PC model QuickCam is used on

- Powered from keyboard port of the computer, draws less than 350mA of current

- Field of view is approximately 65° (equivalent to a 38mm lens on a 35mm camera

- Focus is fixed from 18 inches to infinity

- Lens is f 1.9

7.2.2.2   CCD-Principle

CCD technology is now about 25 years old. Using a specialised VLSI process, a very closely packed array of polysilicon electrodes is formed on the surface of a chip. These are so small and close that the individual packets of electrons can be kept intact whilst they are physically moved from the position where the light was detected, across the surface of the chip to an output amplifier. To achieve this, the array of electrodes is clocked by an off-chip source.

The succeeding figure 49a shows a typical area sensor element
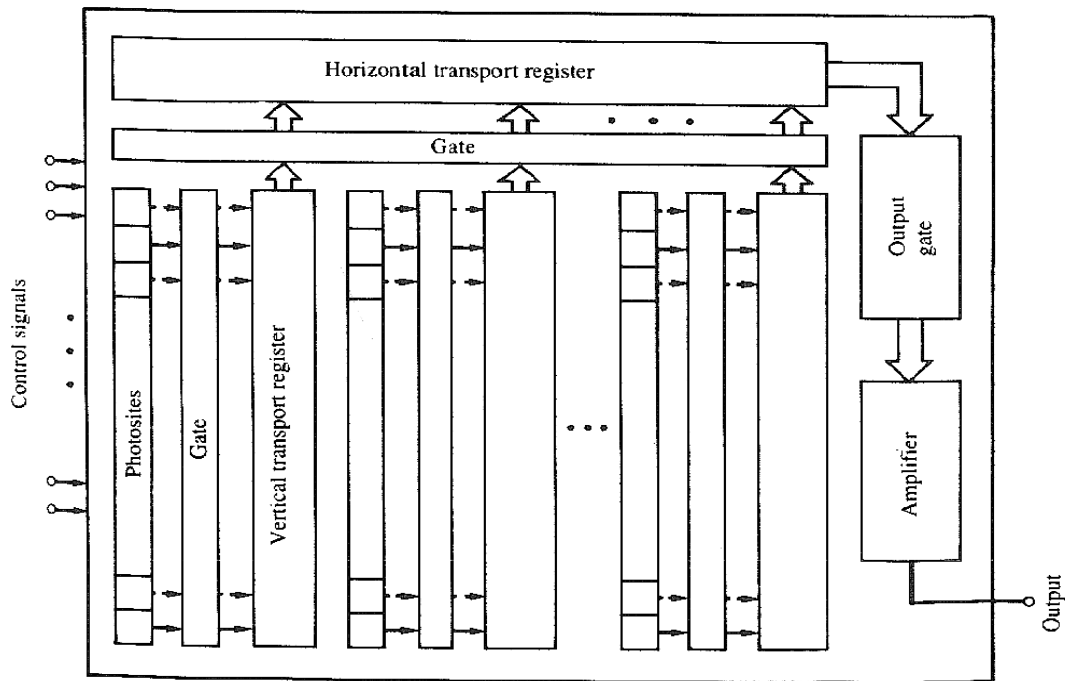
figure 49a : CCD area sensor

Charge-couple area arrays are similar to line-scan sensors, except that the photosites are arranged in a matrix format and a gate/transport-register combination separates columns of photosites. The contents of odd-numbered photosites are sequentially gated into the vertical transport registers and then into the horizontal transport register. The content of this register is fed into an amplifier whose output is a line of video. Repeating this procedure for the even-numbered lines completes the second field of an interlaced TV image frame. It is shown in figure 49a.

It is technically feasible but not economic to use the CCD process to integrate other camera functions, like the clockdrivers, timing logic, signal processing, etc. These are therefore normally implemented in secondary chips. Thus most CCD cameras comprise several chips, often as many as 8 and not fewer than 3.

Apart from the need to integrate other camera electronics in a separate chip, the Achilles heel of all CCDs is the clock requirement. The clock amplitude and rise time are critical to successful operation. Generating correctly sized and shaped clocks is normally the function of a specialised clock driver chip, and leads to two major disadvantages, multiple non-standard supply voltages and high power consumption. It is not uncommon for CCDs to require 5 or 6 different supplies at critical and obscure values. If the user is offered a simple single supply input, then several regulators will be employed internally to generate these supply requirements.

On the plus side, CCDs have matured to provide excellent image quality with low noise.

7.2.2.3   Frame Rate

There are several factors that affect the frame rate of the CCD camera which could reduce the performance of the VELOC system. The VELOC system is designed for real-time working and thus a minimum frame rate is needed to meet this requirement. The following list describes some possible conflicts caused by decreasing image acquisition speed:

- A most important fact is the amount of light in a scene. In dim lighting conditions QuickCam has to leave the „shutter„ open as long as it begins to impact frame rates. To overcome this problem one has to increase the light on the subject. For reducing reflections from the ambiguous light on the eye which affects the image processing steps it might be necessary to fix a piece of black paper on the spectacles´ frame. You have to make sure that the camera then still has enough light to operate.

- The recorded image size affects the frame rate. The VELOC uses a resolution of 160x120 pixels to obtain best results. It is a compromise between image acquisition speed and the accuracy of the system. The higher the resolution, the more accurately the algorithm can calculate the centre of the pupil. For instance a movie recorded at 160x120 will generally record at about 15 fps on most PCs, while a larger window of 320x240 may only record at 4 fps.

- The speed of the hard drive may affect the recording speed. Every single frame has to be saved to disc first before it can be used for image processing. Slow drives or machines with slower data transfer rates may not achieve 15 fps.

- The type of parallel port affects the frame rate as well. A bi-directional (enhanced) parallel port usually transfers data much faster than a uni-directional (compatible) parallel port.

7.2.3 Media Control Interface (MCI)

The Windows API provides two methods for controlling the multimedia capabilities of the system. The first is a series of low-level functions. The second is the Windows Media Control Interface (MCI). The media control interface provides a high-level, device-independent interface for controlling multimedia devices on the system and is the recommended method for controlling multimedia devices in Windows.

There are two methods for issuing MCI commands: command messages and command strings. Command messages are sent by filling a memory structure (GROUP) and passing it to the MCI system. Command strings are English-like strings that are sent to the MCI system and parsed to determine the action to be taken. Command strings provide the same functionality as command messages with a simpler format. Th*e following* examines how to use the media control interface using MCI command strings.

Using MCI, a program has the ability to play and record (as appropriate) on any supported multimedia device. This includes the usual multimedia devices (wave audio, MIDI sequencer, CD audio) as well as some more unusual devices (VCR, videodisc, digital audio tape). The configuration information for MCI is stored in the WIN.INI and SYSTEM.INI files. The [MCI] section of SYSTEM.INI maps the system device names to the device drivers controlling those devices. The [MCI extensions] section of WIN.INI maps file extensions to the appropriate multimedia device.

The MCI configuration information for the VELOC system is shown below

```
[MCI]

cdaudio=mcicda.drv

sequencer=mciseq.drv

waveaudio=mciwave.drv

avivideo=mciavi.drv

videodisc=mcipionr.drv

vcr=mcivisca.drv

QTWVideo=D:\WIN95\SYSTEM\MCIQTW.D
RV
```

figure 50: SYSTEM.INI file of VELOC system

This configuration method makes MCI extensible. New multimedia devices with the proper drivers can be integrated with the MCI system even though the devices weren't invented when MCI was designed. The system can also use multiple devices of the same type by configuring them correctly in SYSTEM.INI.

The MCI system commands are divided into four categories: system, required, basic and extended commands. The commands are described below along with some commonly used options.

- System Commands

System commands are handled by the MCI system instead of by an MCI device. These commands provide system-level information. The system commands are *break* and *sysinfo*.

Command	Description
Break	Set up key to break MCI action
Sysinfo	Gets MCI system information
	sysinfo all quantity
	sysinfo all name [index]
	sysinfo cdaudio quantity

figure 51: System Commands

- Required Commands

Required commands must be implemented for every device. Each device must also support a defined set of options for each command. The required commands are capability, close, info, open and status.

Command	Description
capability	Gets information about device capabilities
	capability cdaudio can eject
close	Closes an open device
	close all
	close cdaudio
info	Gets information about the device driver
	info cdaudio product
open	Opens and initialises a device
	open cdaudio
	open waveaudio!c:\win95\media\chimes.wav
	open c:\win95\media\chimes.wav type waveaudio
	open waveaudio!chimes.wav alias chimesound
	open sequencer!canada.mid
	open canada.mid type sequencer alias and them
status	Gets device status
	status cdaudio mode
	status sequencer mode

figure 52: Required Commands

- Basic Commands

Basic commands are device control commands that are or may not be implemented depending on the device. The commands that are implemented must support a standard set of options for that command. The basic commands are load, pause, play, record, resume, save, seek, set, status and stop.

Command	Description
load	Loads a device file from disk
pause	Pauses playing or recording
	pause cdaudio
	pause anthem
play	Starts playing the device
	play cdaudio
	play cdaudio from 3 to 4
	play waveaudio
	play chimesound
record	Starts recording data with the device
	record waveaudio
resume	Resumes playing or recording after pause
	resume cdaudio
save	Save a device file to disk
	save waveaudio c:\temp\newsound.wav
seek	Moves to a specified position for the device
	seek cdaudio to start
	seek cdaudio to 3
	seek waveaudio to 1000
set	Updates device control settings
	set cdaudio door open
	set cdaudio door closed
	set waveaudio time format milliseconds
	set cdaudio time format tmsf
status	Gets device status
	status cdaudio current track
	status anthem length
	status cdaudio length track 3
	status cdaudio number of tracks
	status waveaudio time format
stop	Stops playing or recording
	stop cdaudio
	stop anthem

figure 53: Basic Commands

- Extended Commands

Extended commands are additional commands that may be implemented for a specific device or that extended that options on a required or basic command. Examples of extended commands are delete and spin. Delete is used to delete a section of a wave audio file. Spin is used to start or stop spinning on a videodisc device. These commands are implemented only for the devices they apply to.

7.3 Functional Description

This section describes the different steps involved in the iris-centre detection, starting with image acquisition through to the final answer of the VELOC system. It illustrates the basic concept of the VELOC system. Although they have been covered from a theoretical point of view in the previous chapters, they are now going to be discussed from a practical viewpoint, reviewing their difficulties and the design considerations undertaken.
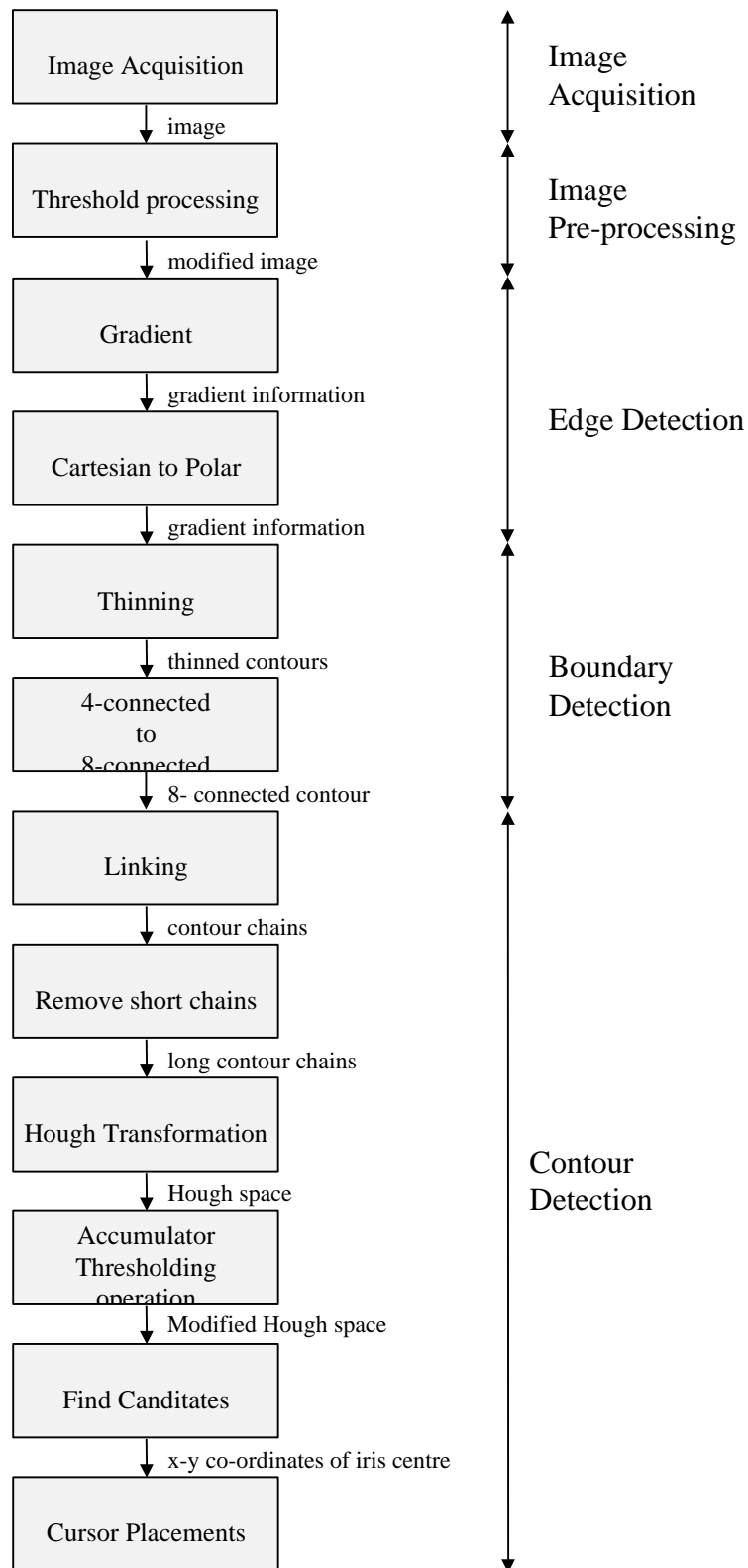
figure 54: VELOC schematic description

7.3.1 Image Acquisition

Image Acquisition is the first step to be taken. As discussed in the previous chapter acquiring an image from the camera is done by sending MCI commands. Before the program is able to handle these instructions initialisation is needed. The succeeding short program sequences (figure 55, figure 56) show the basic initialisation used in the VELOC system.

```
const WM_CAP_START                   = WM_USER;
const WM_CAP_SET_CALLBACK_YIELD      = WM_CAP_START + 4;
const WM_CAP_SET_CALLBACK_FRAME      = WM_CAP_START + 5;
const WM_CAP_DRIVER_CONNECT          = WM_CAP_START + 10;
const WM_CAP_DRIVER_DISCONNECT       = WM_CAP_START + 11;
const WM_CAP_FILE_SET_CAPTURE_FILEA  = WM_CAP_START + 20;
const WM_CAP_SAVEDIB                 = WM_CAP_START + 25;
const WM_CAP_EDIT_COPY               = WM_CAP_START + 30;
const WM_CAP_DLG_VIDEOFORMAT         = WM_CAP_START + 41;
const WM_CAP_DLG_VIDEOSOURCE         = WM_CAP_START + 42;
const WM_CAP_DLG_VIDEODISPLAY        = WM_CAP_START + 43;
const WM_CAP_GET_VIDEOFORMAT         = WM_CAP_START + 44;
const WM_CAP_SET_VIDEOFORMAT         = WM_CAP_START + 45;
const WM_CAP_DLG_VIDEOCOMPRESSION    = WM_CAP_START + 46;
const WM_CAP_SET_PREVIEW             = WM_CAP_START + 50;
const WM_CAP_SET_PREVIEWRATE         = WM_CAP_START + 52;
const WM_CAP_GRAB_FRAME              = WM_CAP_START + 60;
const WM_CAP_SEQUENCE                = WM_CAP_START + 62;
const WM_CAP_SET_SEQUENCE_SETUP      = WM_CAP_START + 64;
const WM_CAP_GET_SEQUENCE_SETUP      = WM_CAP_START + 65;
const WM_CAP_STOP                    = WM_CAP_START + 68;
```

figure 55: Definitions for message handling

Figure 55 shows the definition for the MCI-commands used by the VELOC system. Owing to the fact that all used MCI-commands call Windows messages it is recommended that WM_USER is the starting command in order to avoid any conflicts. The succeeding WM_CAP_XXX commands are used by the VELOC system and are defined in AVICAP32.DLL, which was copied during QuickCam installation.

Before communicating with the camera it is necessary to open the connection to the video

```
procedure TVELOC.CloseConnectionToDriver(Sender: TObject);
begin
    if hWndC <> 0 then begin
                          SendMessage(hWndC, WM_CAP_DRIVER_DISCONNECT, 0, 0); hWndC := 0;
                    end;
end;
procedure TVELOC.OpenConnectionToDriver(Sender: TObject);
begin
    hWndC := 0;
    hWndC := capCreateCaptureWindowA('My Own Capture Window', WS_CHILD or WS_VISIBLE ,
                               0,0,0,0,VELOC.Handle,0);
    if hWndC <> 0 then SendMessage(hWndC, WM_CAP_DRIVER_CONNECT, 0, 0);
end;
```

figure 56: Connect and disconnect driver procedures

driver. It is also recommended that the connection is closed when camera functions are no longer needed. Both procedures are shown in figure 56.

```
procedure TVELOC.CaptureEye(Sender: TObject);
begin

    if hWndC <> 0 then SendMessage(hWndC, WM_CAP_GRAB_FRAME , 0, 0);

    if hWndC <> 0 then SendMessage(hWndC, WM_CAP_SAVEDIB, 0,
                            longint(pchar(SaveDialog1.FileName)));


    ImageForLiveVideoDisplay.Picture.LoadFromFile('c:\images\temp.BMP');


    ViewCapimage.Height:=ImageForLiveVideoDisplay.Height+41;
    ViewCapimage.Width:=ImageForLiveVideoDisplay.Width+7;


ViewCapimage.view.Picture.Bitmap:=ImageForLiveVideoDisplay.Picture.bitmap;


    ViewCapimage.view.Refresh;
end;
```

figure 57: Image capturing procedure

In figure 57 is illustrated the image capturing process. First a Windows message is sent to the camera driver to grab a frame. It is much easier to save it to disc and afterwards load it back as an image component instead of copying directly to an image component. The advantage is the independence of any color palette because the file is stored in the DIB (Device Independent Bitmap) format. In practice it is no drawback because the image is normally not reloaded from hard disc, because the succeeding instruction after WM_CAP_SaveDIB is the LoadFromFile instruction so that it is only stored in the hard disc cache briefly and no unnecessary time is consumed.

The following figure illustrates the result of the image acquisition step.
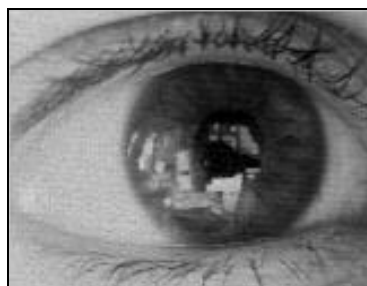
figure 58: Acquired eye image

Size, brightness, contrast and all other camera settings affect the captured image and can be adjusted (choose camera settings and adjustment menu on VELOC menu bar). Best results can be achieved when selecting the automatic brightness function. The following figures illustrate the camera adjustment capabilities. The "white balance" adjustment field in figure 59 adjusts the brightness of the video sequence.

figure 59: VELOC camera adjustments
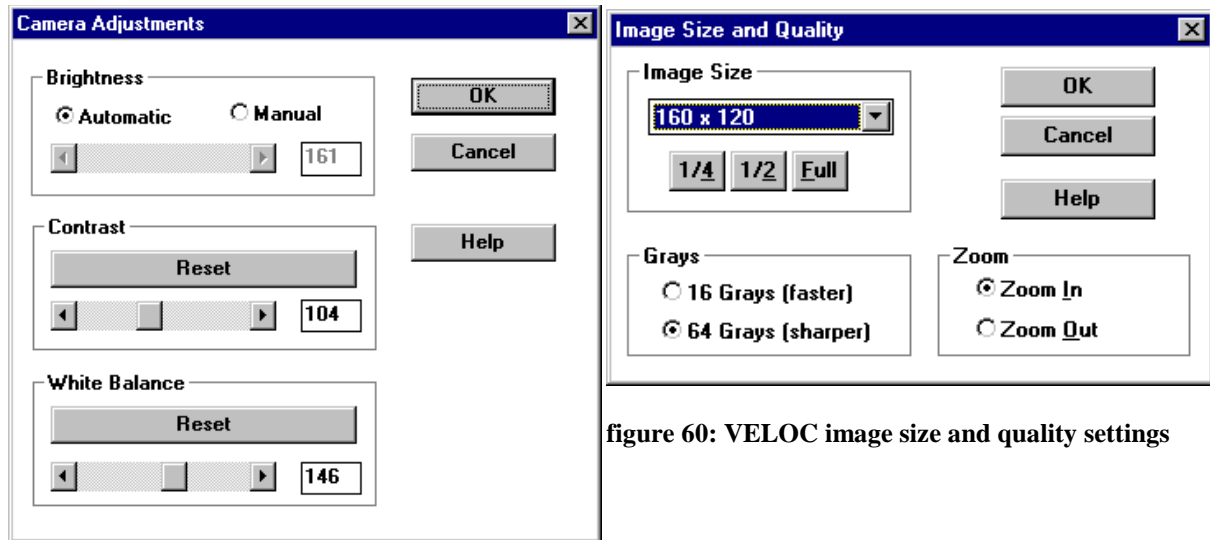
figure 60: VELOC image size and quality settings

7.3.2 Thresholding operation

To obtain better results from the gradient operation it is recommended that some thresholding work is done first. The user is able to enter 2 different values, one for the upper and one for the lower threshold. These values can be adjusted in the VELOC setting menu, shown in the figure 61.
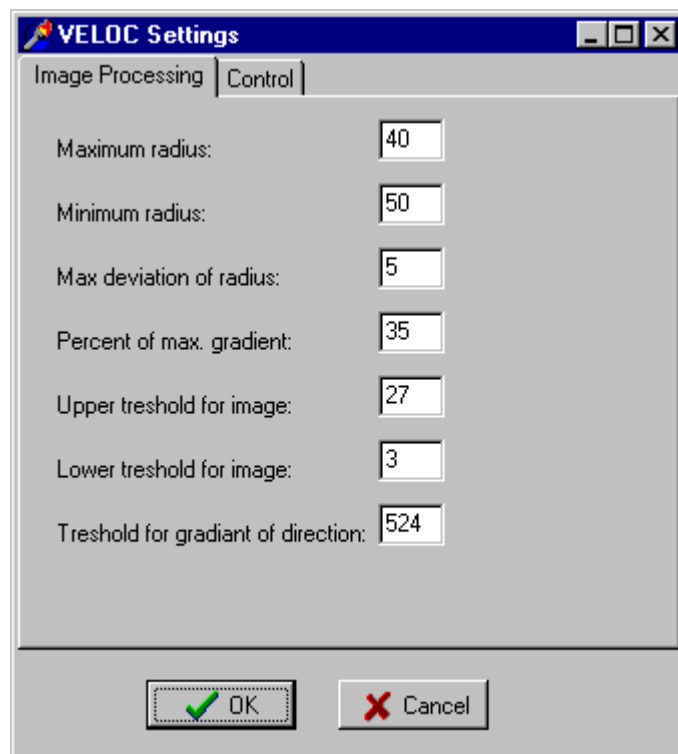
figure 61: VELOC setting menu

The total greyscale range of the acquired image is 64, where 64 stands for white and 0 for black. Therefore, with these settings all pixels which have a greyscale value lower than 3 will be replaced by black , values higher than 27 will be replaced by white pixels. Owing to the

fact that automatic brightness control in the camera adjustments menu should be selected, these settings do not have to be changed usually.
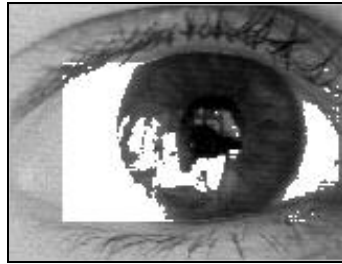
figure 62: threshold image

The following program sequence illustrates this algorithm stage:

```
FOR j:= Calc_Area[1].y TO Calc_Area[2].y DO
     BEGIN
          FOR i:=Calc_Area[1].x TO Calc_Area[2].x DO
          Begin
              if (O[i,j]>UTreshold) then O[i,j]:=colors        {colors = white}
              else
              if (O[i,j]<LTreshold) then O[i,j]:=0
          end;
     END;
```

figure 63: Part of the thresholding procedure

To save computation time in this real-time application, it is recommended that the area that is to be transformed is reduced because two For-loops are needed. The limits are determined during the calibration procedure which will be discussed in chapter 8.3.

7.3.3 Gradient Operation

As has been discussed in chapter 6.3.1.4, the Sobel operator offers the best capabilities for the VELOC system owing to the fact that it is insensitive to noise. A 5x5 mask was experimentally determined to perform best in a reasonable time. A 9x9 mask would have only increased computation time without improving the final result. The used masks are shown in figure 35 and figure 36 and are implemented in the VELOC system as shown in figure 64.

While the input is the original image captured from the camera, the outcomes are two gradient images, one corresponding to the x axis and the other one corresponding to the y axis.

The algorithm performs the following step for each pixel in the image:

- Given the co-ordinate of the pixel, the local convolution of the neighbour pixels within the window around the current point with both mask is computed. It yields two different values, one corresponding to the horizontal gradients and other corresponding to the vertical one.

```
{ Sobel Operation to the Array}

FOR j:= (Calc_Area[1].y+2) TO (Calc_Area[2].y-2) DO
BEGIN
FOR i:=(Calc_Area[1].x+2) TO (Calc_Area[2].x-2) DO
BEGIN

Dx:=(-10)*O[i-2,j-2]+(-10)*O[i-1,j-2]+10*O[i+1,j-2]+10*O[i+2,j-2]+
    (-17)*O[i-2,j-1]+(-17)*O[i-1,j-1]+17*O[i+1,j-1]+17*O[i+2,j-1]+
    (-20)*O[i-2,j]  +(-20)*O[i-1,j]  +20*O[i+1,j]  +20*O[i+2,j]  +
    (-17)*O[i-2,j+1]+(-17)*O[i-1,j+1]+17*O[i+1,j+1]+17*O[i+2,j+1]+
    (-10)*O[i-2,j+2]+(-10)*O[i-1,j+2]+10*O[i+1,j+2]+10*O[i+2,j+2];

Dy:= 10 *O[i-2,j-2]+17*O[i-1,j-2]+20*O[i,j-2]+17*O[i+1,j-2]+10*O[i+2,j-2]+
     10 *O[i-2,j-1]+17*O[i-1,j-1]+20*O[i,j-1]+17*O[i+1,j-1]+10*O[i+2,j-1]+
     (-10)*O[i-2,j+1]+(-17)*O[i-1,j+1]+(-20)*O[i,j+1]+(-17)*O[i+1,j+1]+
     (-10)*O[i+2,j+1]+(-10)*O[i-2,j+2]+(+17)*O[i-1,j+2]+(-20)*O[i,j+2]+
     17 *O[i+1,j+2]+(-10)*O[i+2,j+2];

END;
END;
```

figure 64: Sobel filter implementation

The following pictures illustrate the result of this algorithm step.
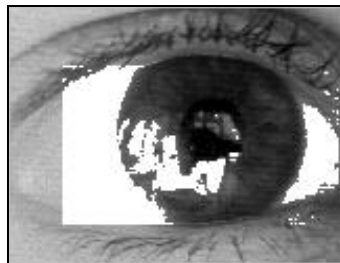
figure 65: Threshold image

figure 66: Gradient x image
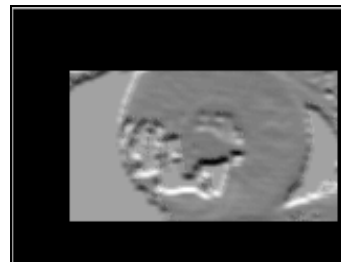
figure 67: Gradient y image

7.3.4 Cartesian to Polar

The previous algorithm step extracted the gradient information from the original image. In order to facilitate the calculations to the following stages of the algorithm, it is more

appropriate to convert this Cartesian information to Polar representation. This offers the ability of easier data handling and also improves the thresholding operations, which take place after nearly every algorithm stage, as you don't have to distinguish between G_x and G_y. Most useful information will therefore be stored in the magnitude of the gradient operation.

For each pixel in the image the following steps are employed:

- Equation (19) is applied to extract the magnitude information, which is a faster approximation to the original equation(19). It offers enough precision to give an accurate solution.

- The next step is to extract the angular information, where equation (20) is utilised.

- To obtain a good final result it is recommended to do thresholding as often as possible. Therefore, the user can specify the highest gradient magnitude permitted. This value (percent of maximum gradient) can be modified in the settings menu, shown in figure 61.
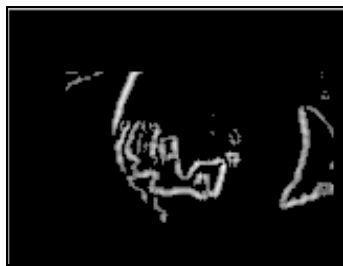
figure 69: Gradient direction image
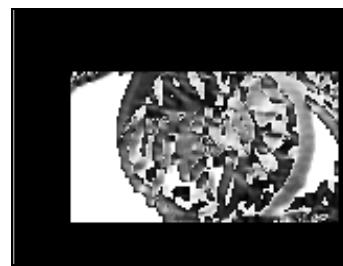
figure 689: Gradient direction image

Figure 68 and figure 69 illustrate the results of the Cartesian to Polar transformation.

7.3.5 Thinning Operation

The thinning operation is the first operation for boundary detection. The gradient magnitude and gradient direction are used. The objective of the algorithm is to find the contour point, which has the characteristic of being 1 pixel wide.

The following algorithm is repeated for each pixel in the image with no NULL value:

- The thinning procedure compares the magnitude and the direction of the current pixel with that of its neighbours on the left-hand and the right-hand sides. Their location is estimated by using figure 41.

- Once the neighbours have been located, the next question concerns the deviations between the gradient direction of the current pixel and the gradient direction of its neighbour. The highest deviation permitted is specified by the user in the settings menu (see figure 61).

- If the comparison of the gradient directions of adjacent pixels yields a deviation exceeding the maximum permitted, the operation is aborted and the next pixel is checked. Otherwise the next pixel is likely to belong to a region of homogeneous gradient directions.

- The remaining question is whether or not the current pixel is a „summit pixel". To answer the question, the gradient magnitudes are utilised. If the magnitude of the current pixel is greater than or equal to the magnitude of both the neighbours, it is classified as a „summit pixel". In this case, the magnitude and direction of the current pixel are retained in the image.

```
{********************************** Thinning **********************************}
  FOR j:=( Calc_Area[1].y+4) TO (Calc_Area[2].y-4) DO
     BEGIN  FOR i:=(Calc_Area[1].x+4) TO (Calc_Area[2].x-4) DO
            Begin

                exit1:=False; exit2:=False;
                exit3:=False; exit4:=False;
                exit5:=False; exit6:=False;
                exit7:=False; exit8:=False;

                if  (ABS(Dd[i,j]-Dd[i-1,j-1])>Radiant) then exit1:=true
                else if (ABS(Dd[i,j]-Dd[i,j-1])>Radiant)   then exit2:=true
                else if (ABS(Dd[i,j]-Dd[i+1,j-1])>Radiant) then exit3:=true
                else if (ABS(Dd[i,j]-Dd[i-1,j])>Radiant)   then exit4:=true
                else if (ABS(Dd[i,j]-Dd[i+1,j])>Radiant)   then exit5:=true
                else if (ABS(Dd[i,j]-Dd[i-1,j+1])>Radiant) then exit6:=true
                else if (ABS(Dd[i,j]-Dd[i,j+1])>Radiant)   then exit7:=true
                else if (ABS(Dd[i,j]-Dd[i+1,j+1])>Radiant) then exit8:=true;

       if (NOT(exit1 or exit2 or exit3 or exit4 or exit5 or exit6 or exit7 or exit8))
       then begin
                if (NOT(exit2 OR exit7)) then
                begin
                 if  ((Dm[i,j]>Dm[i,j-1]) and (Dm[i,j]>Dm[i,j+1]))
                 then O[i,j]:=maxm;
                END
                else
                 if (NOT(exit1 OR exit8)) then
                 begin
                  if ((Dm[i,j]>Dm[i-1,j-1]) and (Dm[i,j]>Dm[i+1,j+1]))
                  then O[i,j]:=maxm;
                 END
                else
                 if (NOT(exit4 OR exit5)) then
                 begin
                  if  ((Dm[i,j]>Dm[i-1,j]) and (Dm[i,j]>Dm[i+1,j]))
                  then O[i,j]:=maxm;
                 END
                else
                 if (NOT(exit3 OR exit6)) then
                 begin
                  if  ((Dm[i,j]>Dm[i+1,j-1]) and (Dm[i,j]>Dm[i-1,j+1]))
                  then O[i,j]:=maxm;
                 END;
                END
                else  O[i,j]:=minm;
            END;
     END;
```

figure 70: Thinning algorithm procedure

figure 71 illustrates the result obtained by this operation. The input pictures for this thinned gradient magnitude result are figure 68and figure 69.
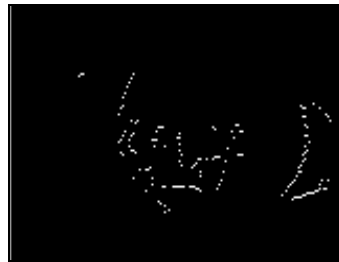
**figure 71: Thinned gradient
magnitude**

The output contours yielded by the thinning algorithm are indeed 4-connected, a feature which is not desirable for the next stage of the algorithm as explained in previous chapters. The 4-connected to 8-connected operation solves this problem.

7.3.6 4-connected to 8-connected Operation

The underlying idea is to find pixels which are unnecessary and do not contain any

	r-1 c				r-1 c									r c	r c+1
	r c	r c+1			r c-1	r c			r c-1	r c				r+1 c	
										r+1 c					
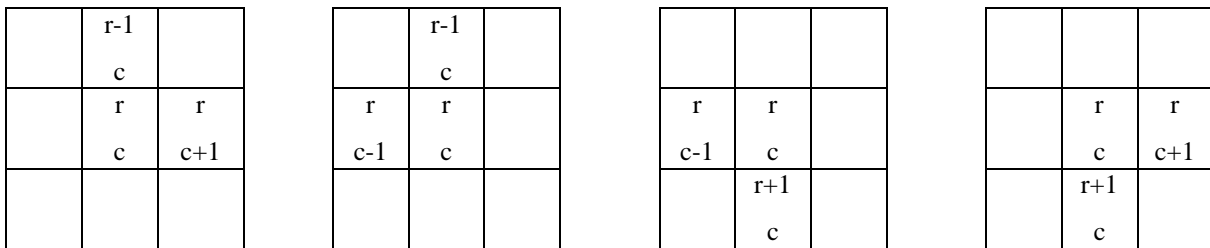
figure 72: Masks used to transform 4-connected chain into 8-connected chain

information. In this case, the pixels are in one of the following configurations:

The current pixel of each mask with co-ordinates (r,c), corresponds to the superfluous contour point. If the algorithm shown in figure 73 encounters one of the four configurations, then the current pixel is set to background color.

This algorithm step also improves the results produced by the last step, the Hough transformation, as it reduces the amount of wrong centre possibilities. The fewer contour points there are in the image, the faster the Hough transformation will be carried out. The following figure shows the result of the 4 to 8-connected contour operation applied to figure 71.
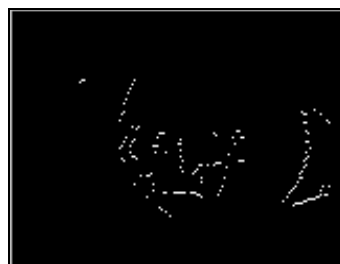
**figure 73: Result of the 4 to 8-
connected contour
operation**

```
{*************************** Connection 4 to 8 ****************************}
    FOR j:=( Calc_Area[1].y+5) TO (Calc_Area[2].y-5) DO
    BEGIN
           FOR i:=(Calc_Area[1].x+5) TO (Calc_Area[2].x-5) DO
           Begin
                If ((O[i,j]=maxm) and (O[i+1,j]=maxm) and (O[i,j-1]=maxm)) then
                O[i,j]:=minm
                else
                If ((O[i,j]=maxm) and (O[i-1,j]=maxm) and (O[i,j-1]=maxm)) then
                O[i,j]:=minm
                else
                If ((O[i,j]=maxm) and (O[i-1,j]=maxm) and (O[i,j+1]=maxm)) then
                O[i,j]:=minm
                else
                If ((O[i,j]=maxm) and (O[i,j+1]=maxm) and (O[i+1,j]=maxm)) then
                O[i,j]:=minm;
           End;
    END;
```

figure 74: Algorithm for 4- to 8-connection

7.3.7 Linking Contour Points Operation

At this algorithm stage, the system is able to begin the contour detection process. It starts by extracting the different contour chains identified by the previous step. This time, the output of the operation is not an image but a list of chains, each one containing the points belonging to the same contour. The main problem of the VELOC system image acquisition procedure is, that it is sensitive to visible light only. Therefore, reflections may cause problems for contour detection owing to the fact that reflections disturb entire contour chains of the edge between the iris and sclera.

The function operates in the following way. For each pixel in the input image,

1. That pixel is interpreted as the first element in the chain, and it is marked as „found".

2. The next step is to scan the neighbourhood around this pixel, searching clockwise for further contour points, starting with the south-eastern one.

3. If this search fails for all the neighbours, the current pixel is the last one in the chain. If it is successful, the located pixel is added to the chain and the search is continued from that pixel, repeating step 2.

The main implication of this algorithm is the way the system is aware of the information. Prior to the operation, the basic data structures were the different points within the pictures. After the linking process, the system is able to appreciate the contours as objects.

Owing to problems with reflections described above, this information is only an additional help for the Hough transformation. Experimental work showed, that it is quite useful help for further process stages, but it is not really necessary stage in the VELOC system algorithm. This algorithm stage can be switched off when it is not needed to save computation time.

7.3.8 Remove Short Chains Operation

This algorithm stage is just to remove the shortest chains to decrease computation time needed by the following Hough transformation. This is a very critical algorithm step. Owing to the fact that reflections cause distortion of entire contours it may happen that only single

pixels hold all the information needed to extract the centre of the circle. If too many single contour points are deleted (set to background colour), less accuracy is the result of this algorithm stage. Therefore, it is up to the user to set the threshold value for removing short chains in the VELOC setting menu. The default value is 1, so that a single pixel will not be deleted.

The algorithm described in words is as follows:

1. If the length of the chain is less than a given threshold, which is provided by the user, it is deleted from the contour list (or the image).

2. If it is greater than the threshold the next chain is examined.

As the default value is 1, the output image of this stage is the same as in the previous algorithm stage.

7.3.9 Hough Transformation Operation

This algorithm stage is the final step and also the most critical one. Regarding the concept introduced in chapter 6.3.3.2, the Hough transformation incorporates two consecutive processes, one involving the creation of the Hough space, and other involving the interpretation of the different values contained within that output accumulator.

The first step has been denoted as the Hough transformation because it is the core of the operation. The second processing step has been named the accumulator thresholding operation.

Referring to this early step, various things must be taken into consideration. Because a great amount of computation is needed to determine to which different circles a given pixel may belong, it would be beneficial to restrict the search to a bounded domain and also make use of some characteristic features in the picture. This involves the use of knowledge of the human eye's features.

Referring to what was already stated in previous chapters, the centre of the iris is going to be located by knowledge of its approximate diameter and its rough position. These can be added to the operation in order to improve the response time.

This prior knowledge can be summarised as follows:

- Practical experience with this configuration has proved that the diameter of the iris is between 40 and 50 pixels wide, making it unnecessary to search for circles with radii outside these limits. These values can of course be adjusted in the VELOC setting menu.

- The centre of the iris is always located in a predefined area. This area has to be selected during calibration process which will be explained in chapter 8.3.

Once the search area and the appropriate radii have been predetermined, the operation itself involves the following actions for each pixel contour:

- All the cells in the Hough space are initially set to 0.

- Determine the different circles the point may belong to. Knowing the co-ordinates of the pixel, x and y, there are three free variables, x_c, y_c and c. Using two loops, one outer loop for x_c and an inner loop for y_c, the algorithm generates all the different combinations of centre points. After replacing x_c and y_c which have been determined from equation (27), a value for c is found.

- After checking if c ranges between the maximum and minimum radius, the triplet (x_c, y_c, c) indexes the Hough space, incrementing the pointed cell by one.

An example of the different circles located by the combination of this and the previous algorithm can be seen next.
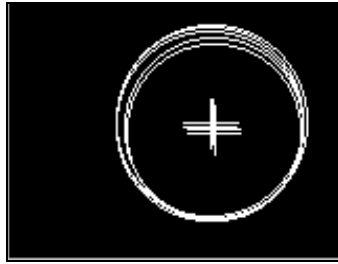
**figure 75: Circles detected
by the Hough
Transforma-tion**

This captured image of the Windows application shows the graphical representation of all the circles detected by the combination of the previous operations: *Hough transform and accumulator thresholding*.

The following program sequence shows parts of the source code written in Delphi 3.

```pascal
{*************************** HOUGH-Transformation ****************************}
for a:=0 to 160 do for b:=0 to 120 do for c:=34 to 56 do H[a,b,c]:=0;
for i:=1 to 5 do hm[i].v:=0;
Hmax:=0;
Hmax1:=0;
amax:=0;
bmax:=0;
cmax:=0;
FOR j:=( Calc_Area[1].y+5) TO (Calc_Area[2].y-5) DO
   BEGIN
        Row  := pByteArray(Image1.picture.Bitmap.Scanline[j]);
        FOR i:=(Calc_Area[1].x+5) TO (Calc_Area[2].x-5) DO
        Begin
             Hough[i,j]:=colors-Row[i];
         end;
   END;
FOR y:=( Calc_Area[1].y+5) TO (Calc_Area[2].y-5) DO
     BEGIN
       FOR x:=(Calc_Area[1].x+5) TO (Calc_Area[2].x-5) DO
       if (Hough[x,y]<>0) then
        FOR a := a_range_min TO a_range_max DO
        FOR b := b_range_min TO b_range_max DO
             Begin
                    c:=ROUND(SQRT((x-a)*(x-a)+(y-b)*(y-b)));
                    if (c>c_range_min) and (c<c_range_max) then inc(H[a,b,c]);
                    if (H[a,b,c]>Hmax) then Hmax:=H[a,b,c];
             END;
     END;
FOR a := a_range_min TO a_range_max DO
     FOR b := b_range_min TO b_range_max DO
     For c := c_range_min to c_range_max do
     begin
          if (H[a,b,c]=Hmax) then
          begin
               hm[5].v:=Hmax;
               hm[5].a:=a;
               hm[5].b:=b;
               hm[5].c:=c;
          end
          else
if ((H[a,b,c]>hm[1].v) and (H[a,b,c]>hm[2].v) and (H[a,b,c]>hm[3].v) and (H[a,b,c]>hm[4].v))
then
       begin
               hm[1]:=hm[2];
               hm[2]:=hm[3];
               hm[3]:=hm[4];
               hm[4].v:=H[a,b,c];
```

```
hm[4].a:=a;
                hm[4].b:=b;
                hm[4].c:=c;
        end

else
if ((H[a,b,c]>hm[1].v) and (H[a,b,c]>hm[2].v) and (H[a,b,c]>hm[3].v) and (H[a,b,c]<hm[4].v))
then
        begin
                hm[1]:=hm[2];
                hm[2]:=hm[3];
                hm[3].v:=H[a,b,c];
                hm[3].a:=a;
                hm[3].b:=b;
                hm[3].c:=c;
        end
else
if ((H[a,b,c]>hm[1].v) and (H[a,b,c]>hm[2].v) and (H[a,b,c]<hm[3].v) and (H[a,b,c]<hm[4].v))
then
        begin
                    hm[1]:=hm[2];
                    hm[2].v:=H[a,b,c];
                    hm[2].a:=a;
                    hm[2].b:=b;
                    hm[2].c:=c;
        end
else
if ((H[a,b,c]>hm[1].v) and (H[a,b,c]<hm[2].v) and (H[a,b,c]<hm[3].v) and (H[a,b,c]<hm[4].v))
then
        begin
                    hm[1].v:=H[a,b,c];
                    hm[1].a:=a;
                    hm[1].b:=b;
                    hm[1].c:=c;
        end;
end;
```

figure 76: Parts from the Hough transformation algorithm

7.3.10   Find Candidates operation

This is the last step of the contour detection operations sequence. Its final aim is to differenciate the circles which are likely to present the eyes from the rest.

With regard to the prior knowledge, it has been determined that the best method is to choose the centre of the circle with the highest accumulator value. It is recommended that the centre of a circle with the highest accumulator value be chosen. Consequently the 5 highest accumulator values have to be determined and then the values of the surrounding cells in the 3 dimension area are added to reduce the sensitivity to small deviations from the determined centre. Then the cell with the maximum value is determined and chosen as the correct and best centre of the iris. The following program sequence illustrates this concept in figure 76a:

```
For z:=5 DOWNTO 1 DO
     FOR ai := (hm[z].a-c_dev) TO (hm[z].a+c_dev) DO
     FOR bi := (hm[z].b-c_dev) TO (hm[z].b+c_dev) DO
     FOR ci := (hm[z].c-c_dev) TO (hm[z].c+c_dev) DO
             BEGIN
                  if ((ai>0) and (ai<160) and (bi>0) and (bi<120) and (ci>34) and (ci<56))
                  then
                  begin
                       hm[z].v:=hm[z].v + H[ai,bi,ci];
                       H[ai,bi,ci]:=0;
                  end
                  else click:=true;
             END;
For z:=1 TO 5 DO
        begin
             if (hm[z].v=hmax) then
             begin
                  x_pos:=hm[z].a;
                  y_pos:= hm[z].b;
                  r_pos:=hm[z].c;
             end;
        end;
```

figure 77: Find candidates operation program sequence

The following picture illustrates the final result of the Hough Transformation.
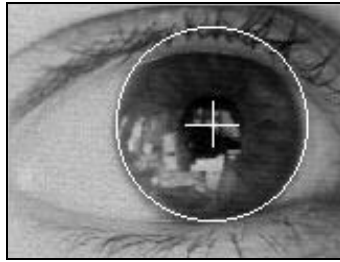
**figure 78: Final result of the
VELOC algorithm**

7.3.11   Comments

This chapter has reviewed the hardware and software issues related to the development of the VELOC system, as well as discussed the different algorithms needed to detect the centre of the iris from a user´s eye picture.

The main objective has been to review the transformation of the theoretical fundamentals detailed in chapter 6.3.3.2, into practical algorithms capable of solving real problems.

Through the previous section, the sequential steps undertaken by the VELOC system in order to successfully detect and locate the position of the iris centre have been examined. Although all the algorithms have a mathematical basis, some sections have been developed based on practical experimentation, particularly the way required parameters have been determined, most of them threshold values.

The following figure illustrates the entire algorithm processed by the VELOC system.
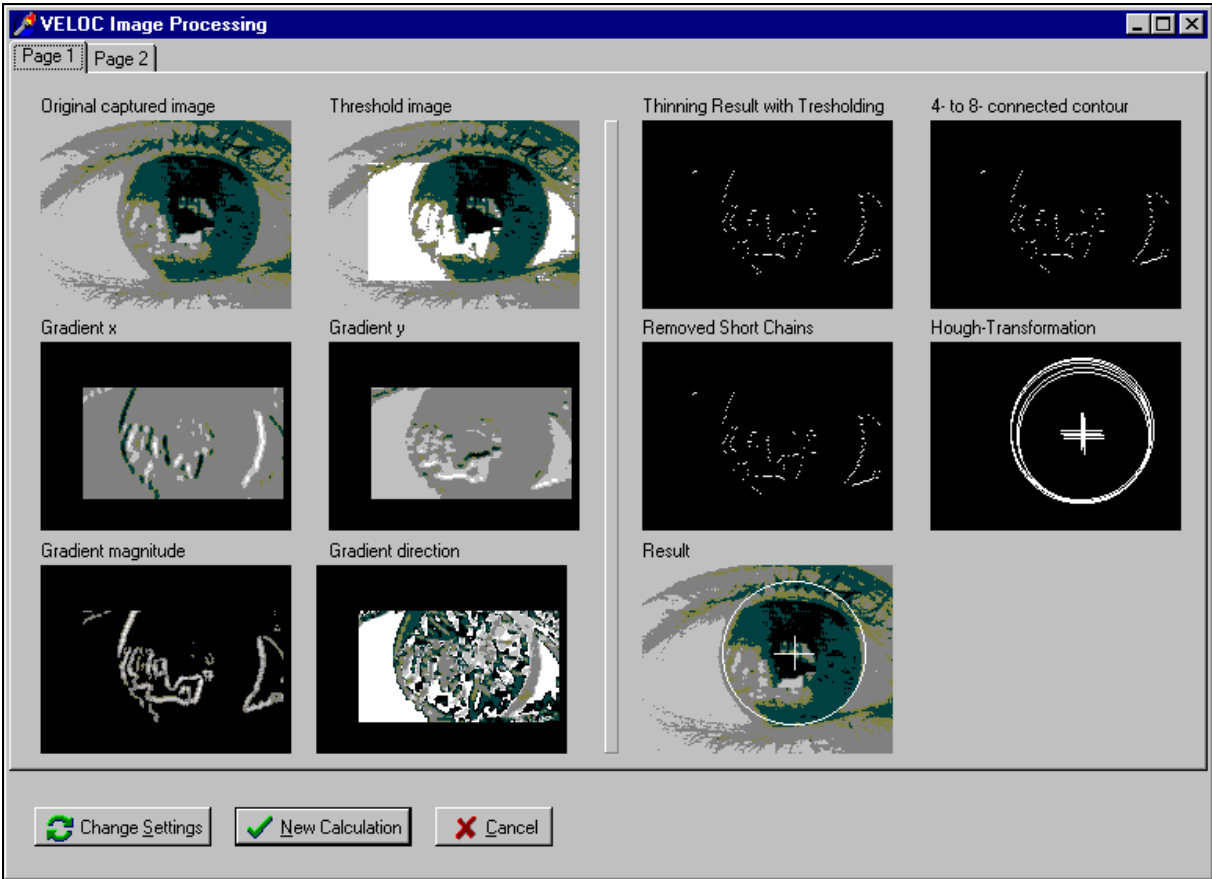
figure 79: VELOC image processing algorithm

8 Evaluation of the VELOC system

8.1 Introduction

The evaluation for the VELOC system was carried out within certain conditions, which will be discussed later. This is due to the fact that the current work is only to implement and research the suitable eye tracking system, the VELOC system will work as described as a low cost alternative option for the disabled people. An entire calibration and operating session will be described to show the benefit and simplicity of this new developed eye gaze system. The results of which will be discussed and listed to show the applicability of the VELOC system.

8.2 Operating Conditions

8.2.1 Lighting conditions and monitor placing

The operating conditions during using the VELOC system are affecting the results. First of all the light condition is the most critical part. During testing the VELOC system, the eye was shielded from ambiguous light to avoid unnecessary reflections on the eye which could disturb the clear shape between the iris and the sclera. Reflections of objects with a light surface occurred as well; so they also should be removed (for example a piece of white paper below the monitor)

The placement of the monitor plays a very important role. It is recommended that the centre of the monitor is in about the same height as the users eyes when looking straight ahead. This position of the monitor has the advantage that the eye will be covered by the eye lids as less as possible when the user is looking at the bottom of the screen.

8.2.2 System

The computer undertook the evaluation was an Intel Pentium-based CPU with clock speed of 70 MHz and 16 MB RAM. Owing to the fact that the VELOC algorithm for determining the centre of the iris during image processing needs much computation power, the speed of the processor is directly related to the execution time of the program and therefor a need to operate in real-time. Image acquisition and image processing on this machine took about 700ms depending on the shape between the iris and the sclera. Investigations were also made on a dual Pentium P2 300 MHz with 256 MB RAM. Both for image acquisition and image processing not more than 200 ms was needed.

As described in the previous chapter, a Connectix camera was attached to the computer to input the eye image into computer files.

8.2.3 Image File

Best results were achieved when using an image acquisition size of 160x120 pixels with 64 grey scales. The „zoom in" feature was also selected. Brightness control is set too automatic and therefore does not to be adjusted during operation. This provides the ability of relative independence of changing light conditions while working with the VELOC system due to the camera adapts automatically. The following illustrations show these settings.
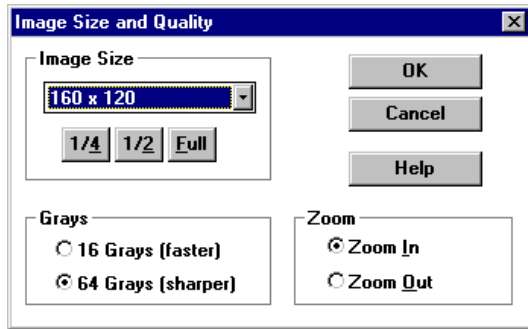
figure 81: Image size and quality
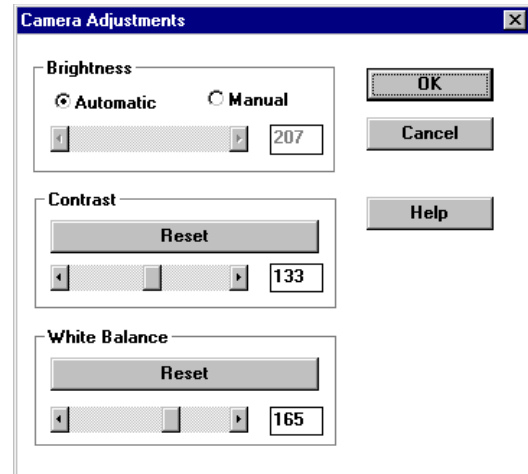
figure 801: Camera adjustments

8.3 A VELOC session

This chapter explains a complete session with the VELOC system. The handling of the program, the calibration process and of course correct using will be discussed.

After starting the VELOC program the following screen will appear shown in figure 82: VELOC welcome screen.
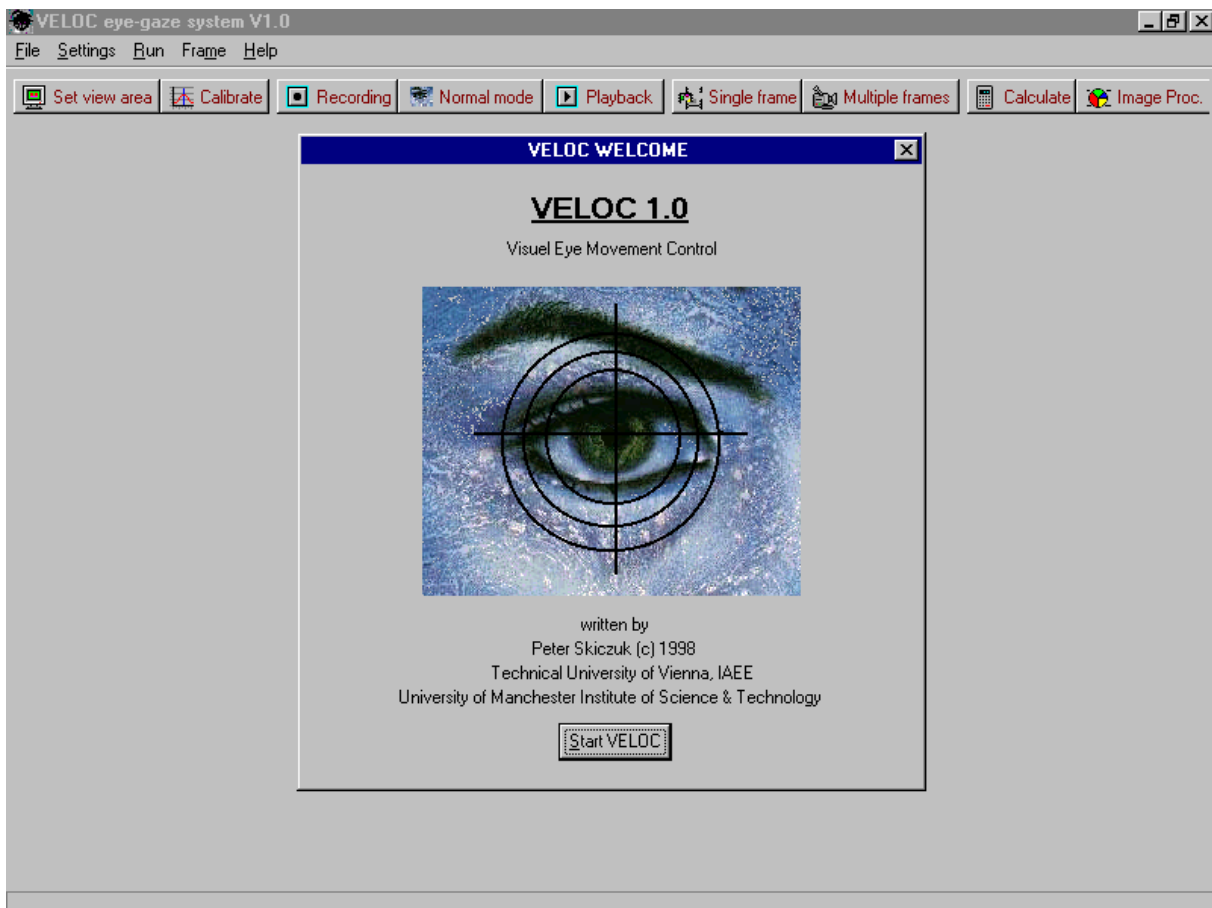
figure 82: VELOC welcome screen

After clicking the „Start VELOC" button the program is ready for use. The most important functions are displayed on the main screen and accessible through hot keys. These

instructions are listed below. An integrated voice commander will guide you through the program to make handling the program as simple as possible.

- Set view area

- Calibrate

- Recording

- Normal mode

- Playback

- Single frame

- Multiple frames

- Calculate

- Image Processing

The set view area instruction enables the user to adjust the view area in which his eye gaze will be determined. The view area is splitted into small areas with adjustable dimensions (figure 83). When the user is looking at one specific area and blinks with the right eye once, this area will change the color to indicate a successful operation. The calculated eye gaze will be displayed as well.
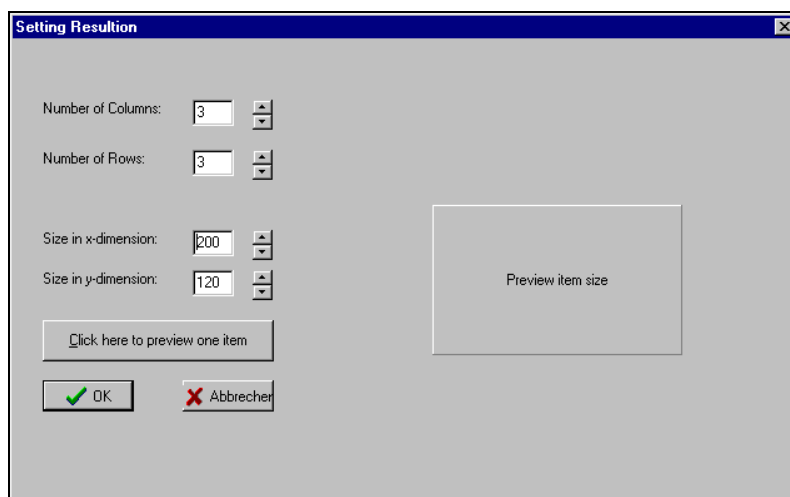
figure 83: Setting area dimensions

The next step is to calibrate the eye gaze system. The information given by this process is needed to determine the exact position of the user's eye gaze on the screen. This is the most important step due any inaccuracy during the calibration process will affect the quality of the results dramatically. To enter calibration mode, simply press the calibration button on the top of the main screen. The following screen will appear:
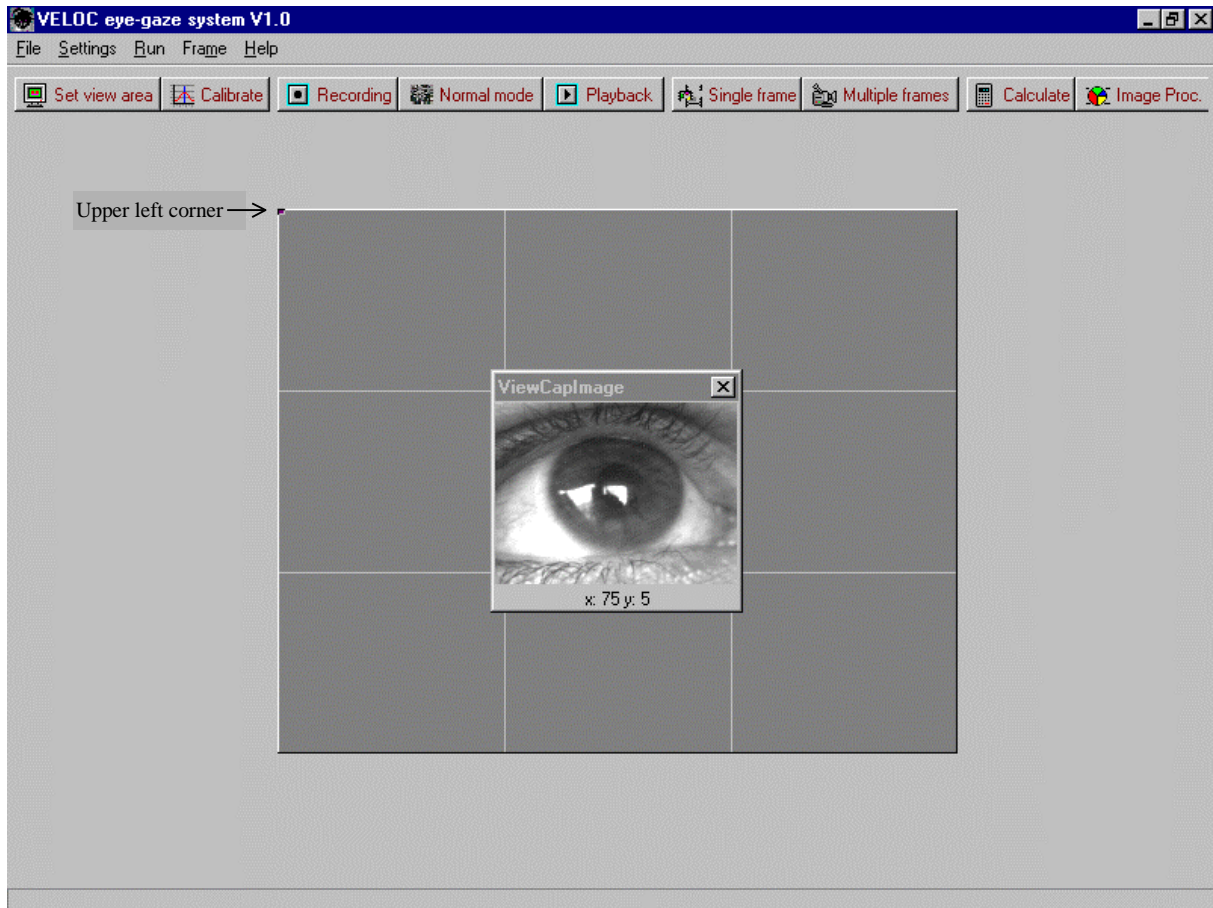
figure 84: Calibrations process stage 1

The user has to look first into the left upper corner where the black rectangle appears. When the eye fixates this rectangle the space bar is needed to be pressed to proceed to the right upper corner. The next corner will be the lower right and finally the lower left corner. At each stage the space bar has to be pressed to go to the next corner. When all 4 corners are passed the screen shown in figure 85 will appear .
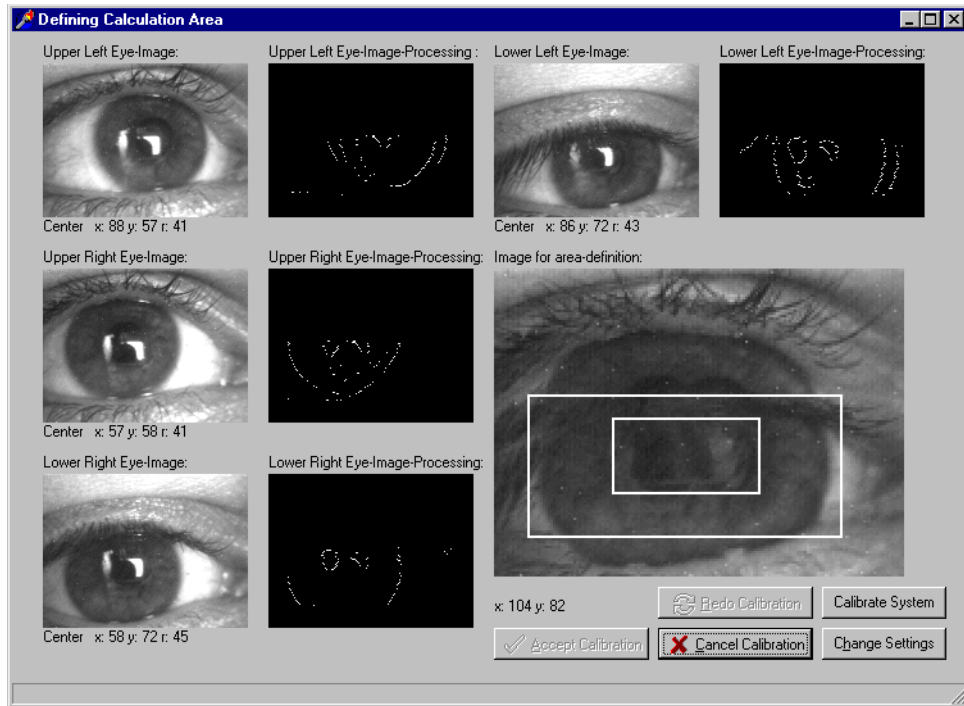
figure 85: Calibration process stage 2

It is up to the user now to determine the calculation area where image processing and eye gaze calculations take place. This is necessary to safe computation time to operate in real-time. The big right image is the result of all four captured eye images. Every pixel of all four images are compared with each other and the darkest one is displayed in the big eye image so that the pupil and the eye lids are always shown clearly.

First the big white rectangle has to be drawn to define the calculation area. Owing to the fact that the eye lids covers the iris most of the time during working it is recommended that both
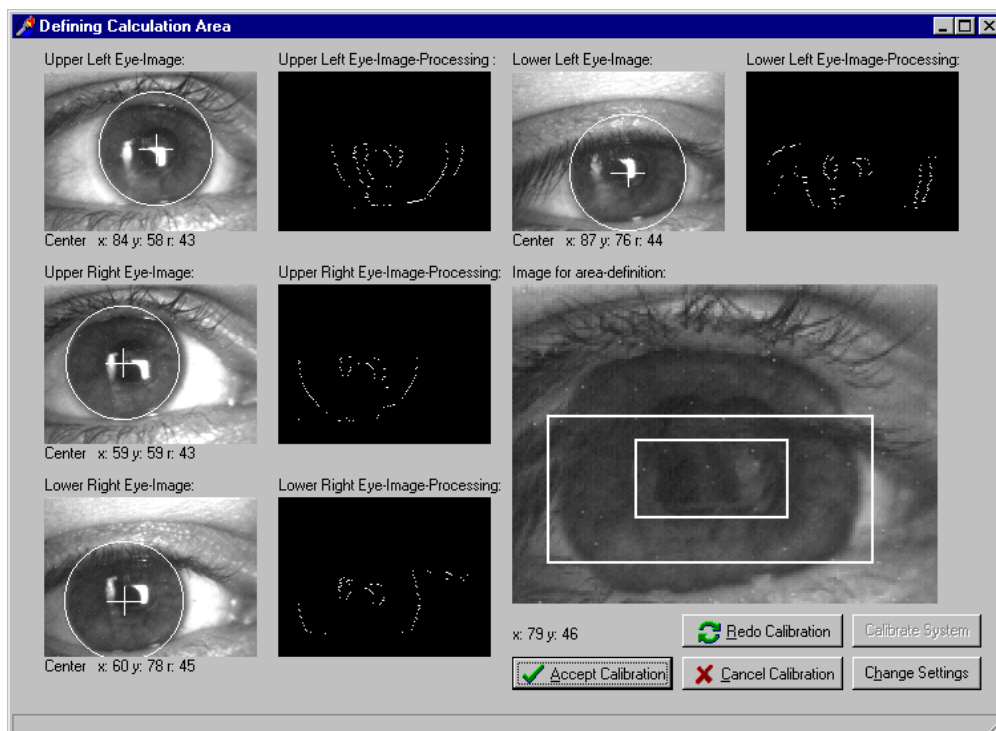
figure 86: Calibration process stage 3

the upper and the lower eye lid are not included in the white area of the large rectangle. The eye lid has a different radius compared to the iris so that they may lead to wrong results, if they are included. Drawing the rectangle is done by clicking the left mouse button at the upper left corner where the rectangle should begin and keeping it pressed until to the lower right corner where it finally should be released. The next step is to determine the possible area of the centre of the pupil. Therefore, the black area of the pupil should be surrounded with the small rectangle as shown in figure 85.

After drawing both rectangular the button „Calibrate system" should be pressed which results in the following screen:

Below each eye image the co-ordinates of the iris centre are displayed.

The user should check this co-ordinates and then decide if those values are acceptable. Here are the most important rules for checking:

- The y co-ordinate of the iris centre should not differ more than 2 pixel between the upper left and upper right image. This means that the user only shifted the eye horizontally.

- The y co-ordinate of the iris centre should not differ more than 2 pixel between the lower left and lower right image. This means that the user only shifted the eye horizontally.

- The x co-ordinate of the iris centre should not differ more than 2 pixel between the upper left and lower left image. This means that the user only shifted the eye vertically.

- The x co-ordinate of the iris centre should not differ more than 2 pixel between the upper right and lower right image. This means that the user only shifted the eye vertically.

The better the values match, the more accurate and sufficient the result will be. If all values are within the allowed range then the user should press the „Accept Calibration" button, otherwise the procedure should be done once more or the settings should be changed as described in chapter 7.3.2.

After this calibration procedure the VELOC system is ready for use. A re-calibration is only needed when the user moves his head so that the results are insufficient due the geometric relationship between the head and the monitor has changed.

Two different operation modes are now available to the user. The one is the recording the other is the normal mode. The only difference is that in recording mode all eye gazed results will be stored and are available for later presentation by clicking the „Record mode" button and selecting the previous saved file.

For example, if „Record mode" is chosen, in the centre of the boxes crosses will appear where the user has to look. The VELOC system will calculate the eye gaze position and highlight the selected button. Blinking with the right eye enables the user to select this box which will be indicated by playing a short sound-sequence.

The „Single frame" button captures a single frame from the camera and displays the image in an automatically adjusted window. The same does the „multiple frame" button, the only difference is that capturing eye images will be stopped by pressing the „Escape" button on the keyboard.

The „Calculate" button determines the centre of the iris of the currently captured eye image. The system has to be calibrated first due the calibration information is needed for the algorithm.

The „Image processing" button shows each single image processing step undertaken by the VELOC algorithm. This might be useful if the results are not sufficient enough and the user has to look for some improvements. Therefore, he/she can go through the algorithm and so adjust the specific parameters in the VELOC settings menu.

8.4 Testing Results

This chapter discribes the test procedure and discusses the results with different test users. Therefore, table 1 to table 8 are showing the testing results of the VELOC system.

The test users were split up in two groups.

One group was using the system for the very first time and had just few minutes to get used to the system. Therefore, they were using the "Normal mode" of the VELOC program allowing to see the users eye-gaze immediately. The persons were between 13 and 55 years old and were familiar with computers. They are called "normal user" in the following discussion.

The other group were those people who have been using this program for a long time. They got allready used to the program and especially the way of interacting with the new input device.

The test procedure will be described in the succeeding paragraph.

The user had to fix his/her head infront of the monitor with a distance between 30 and 50 cm. This could be met by putting the head on some books. It was not necessary to fix the head in an apparatus when the user was able to hold his head quite still. The position of the head was so adjusted that the users eye gaze was focused in the middle of the screen when he was looking straight ahead. For each user, first a table is presented which shows the testing results for 20 recorded points. The different values were yield in the "recording mode". The user had to look at the cross with the "calculated" co-ordinates and the VELOC system determined the eye gaze and stored it in "determined" co-ordinates.

These can be seen in the first three columns of the succeeding tables. On the right side of the table the deviation between the "determined" and "calculated" co-ordinates is shown. The column "area hits" indicates if the deviation of the co-ordinates were less than the half of the button size so that the button was hit successfully. A figure represents this result as well. These deviations are graphically presented in the succeeding figures. At the button of the table the mean value and the standard deviation is calculated. Screen shots for each user can be found in the chapter "Appendix".

Test person 1: Advanced user
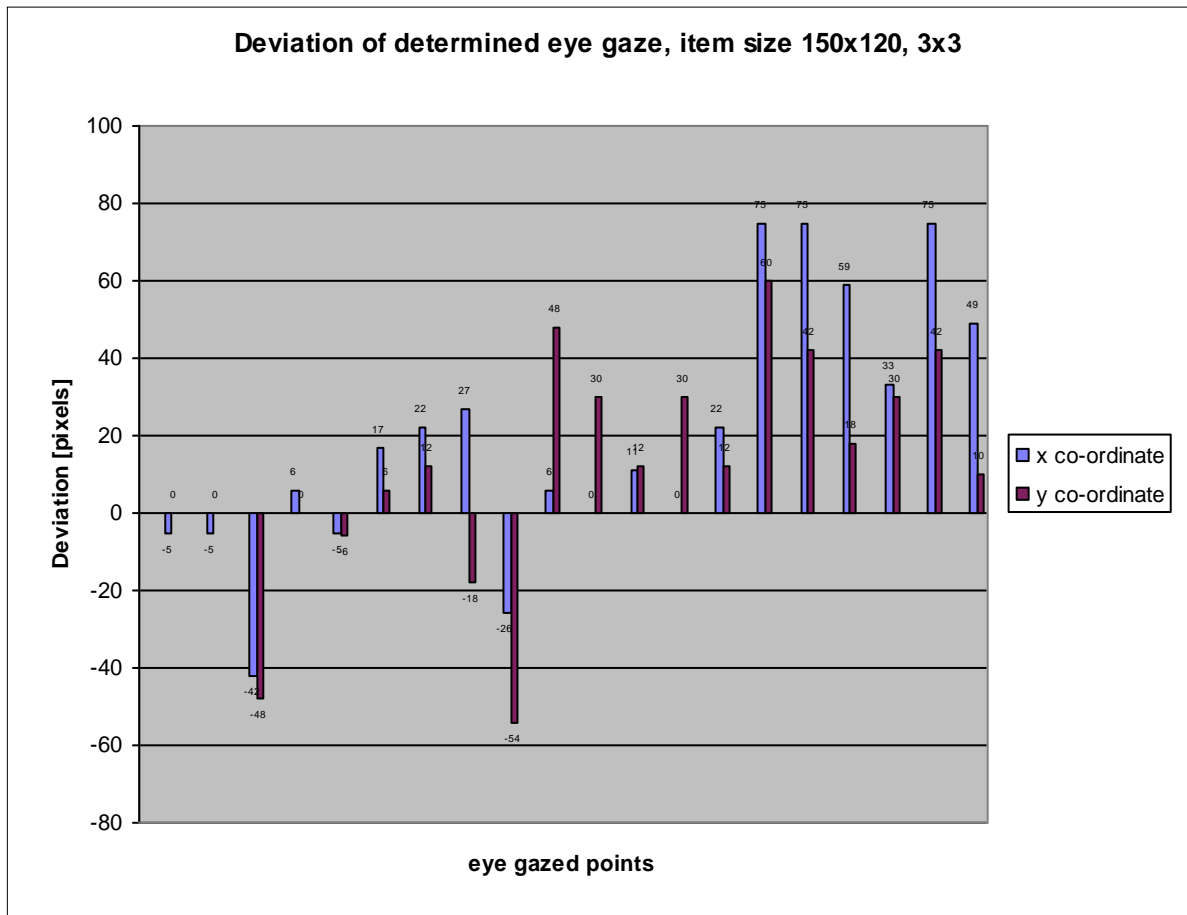
figure 87: Determined eye gaze deviation of an advanced user
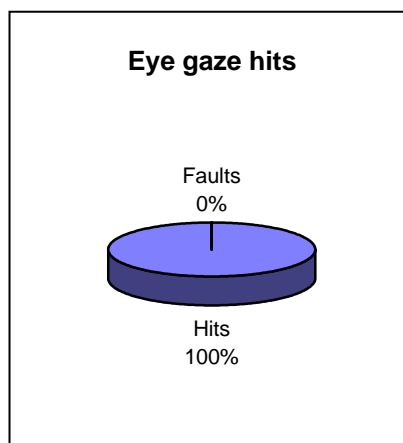
figure 88: Eye gaze hits of an advanced user

VELOC

	x	y	Deviation x	Deviation y	Area hit
Number of items:	3	3			
Dimension of an item:	150	120			
Calculated point 1	75	180			
Determined point 1	80	180	-5	0	1
Calculated point 2	75	180			
Determined point 2	80	180	-5	0	1
Calculated point 3	375	60			
Determined point 3	417	108	-42	-48	1
Calculated point 4	375	180			
Determined point 4	369	180	6	0	1
Calculated point 5	75	300			
Determined point 5	80	306	-5	-6	1
Calculated point 6	225	60			
Determined point 6	208	54	17	6	1
Calculated point 7	375	300			
Determined point 7	353	288	22	12	1
Calculated point 8	75	180			
Determined point 8	48	198	27	-18	1
Calculated point 9	375	180			
Determined point 9	401	234	-26	-54	1
Calculated point 10	375	300			
Determined point 10	369	252	6	48	1
Calculated point 11	225	300			
Determined point 11	225	270	0	30	1
Calculated point 12	75	300			
Determined point 12	64	288	11	12	1
Calculated point 13	225	300			
Determined point 13	225	270	0	30	1
Calculated point 14	375	300			
Determined point 14	353	288	22	12	1
Calculated point 15	75	60			
Determined point 15	0	0	75	60	1
Calculated point 16	75	60			
Determined point 16	0	18	75	42	1
Calculated point 17	75	180			
Determined point 17	16	162	59	18	1
Calculated point 18	225	300			
Determined point 18	192	270	33	30	1
Calculated point 19	75	60			
Determined point 19	0	18	75	42	1
Calculated point 20	75	180			
Determined point 20	26	170	49	10	1
Mean value:			19,7	11,3	
Standard deviation:			32,23	29,06	

table 1: Advanced user

As seen in the previous graphical presentation in figure 88 every eye gaze to a specific cross caused an area hit. All deviation were within the allowed limits. Due to this fact the following test persons were "normal users" to test the suitability of the system.

Test person 2: Normal user 1
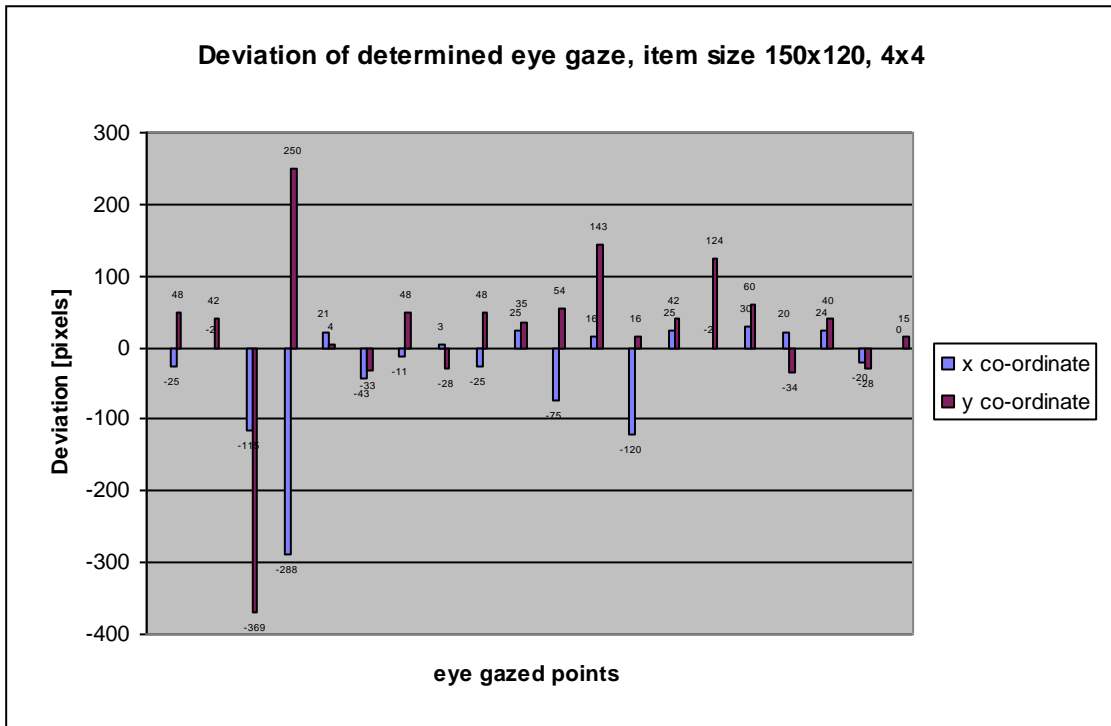
figure 89: Determined eye gaze deviation of normal user 1
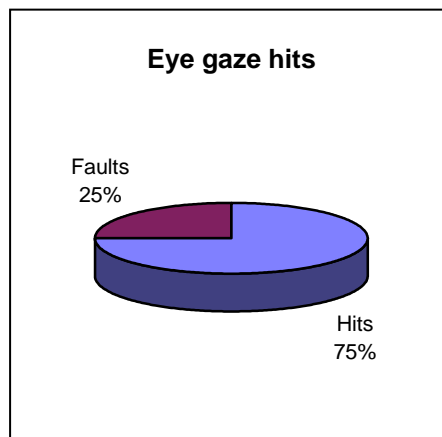
figure 90: eye gaze hits of user 1

	x	y	Deviation x	Deviation y	Area hit
Number of items:	4	4			
Dimension of an item:	150	120			
Calculated point 1	375	300			
Determined point 1	400	252	-25	48	1
Calculated point 2	525	420			
Determined point 2	527	378	-2	42	1
Calculated point 3	375	60			
Determined point 3	490	429	-115	-369	0
Calculated point 4	75	300			
Determined point 4	363	50	-288	250	0
Calculated point 5	75	180			
Determined point 5	54	176	21	4	1
Calculated point 6	375	420			
Determined point 6	418	453	-43	-33	1
Calculated point 7	225	300			
Determined point 7	236	252	-11	48	1
Calculated point 8	75	300			
Determined point 8	72	328	3	-28	1
Calculated point 9	375	300			
Determined point 9	400	252	-25	48	1
Calculated point 10	225	60			
Determined point 10	200	25	25	35	1
Calculated point 11	525	180			
Determined point 11	600	126	-75	54	1
Calculated point 12	525	420			
Determined point 12	509	277	16	143	0
Calculated point 13	225	420			
Determined point 13	345	404	-120	16	0
Calculated point 14	225	420			
Determined point 14	200	378	25	42	1
Calculated point 15	525	300			
Determined point 15	527	176	-2	124	0
Calculated point 16	375	180			
Determined point 16	345	120	30	60	1
Calculated point 17	525	180			
Determined point 17	505	214	20	-34	1
Calculated point 18	524	400			
Determined point 18	500	360	24	40	1
Calculated point 19	180	350			
Determined point 19	200	378	-20	-28	1
Calculated point 20	75	180			
Determined point 20	75	165	0	15	1
Mean value:			-28,1	23,85	
Standard deviation:			73,44	114,06	

table 2: Normal user 1

An error rate of Eye gaze hits of 25% indicates that even users with no experiences with this system can yield good results. As can seen, the deviation of the y co-ordinate are generally bigger than those of the x co-ordinate. This indicates that small errors during the calibartion stage have been made.

Test person 3: Normal user 2

figure 90: Determined eye gaze deviation of normal user 2
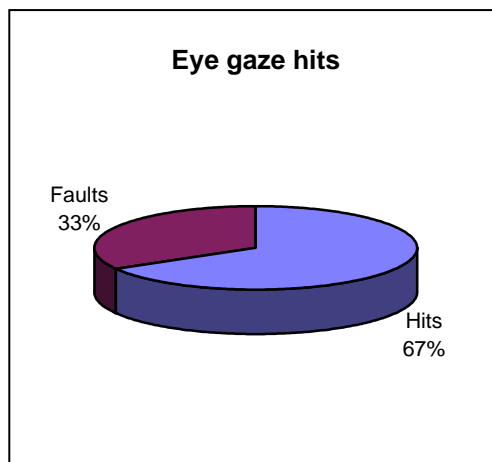
figure 91: Eye gaze hits of normal user 2

VELOC

	x	y	Deviation x	Deviation y	Area hit
Number of items:	4	4			
Dimension of an item:	66	60			
Calculated point 1	32	210			
Determined point 1	81	264	-49	-54	0
Calculated point 2	32	90			
Determined point 2	65	72	-33	18	1
Calculated point 3	162	90			
Determined point 3	195	72	-33	18	1
Calculated point 4	228	210			
Determined point 4	260	264	-32	-54	0
Calculated point 5	228	90			
Determined point 5	260	48	-32	42	0
Calculated point 6	32	90			
Determined point 6	48	72	-16	18	1
Calculated point 7	98	90			
Determined point 7	130	72	-32	18	1
Calculated point 8	228	90			
Determined point 8	276	48	-48	42	0
Calculated point 9	98	210			
Determined point 9	130	240	-32	-30	1
Calculated point 10	32	210			
Determined point 10	64	240	-32	-30	1
Calculated point 11	32	30			
Determined point 11	65	24	-33	6	1
Calculated point 12	228	90			
Determined point 12	276	48	-48	42	0
Calculated point 13	98	90			
Determined point 13	130	72	-32	18	1
Calculated point 14	228	90			
Determined point 14	276	96	-48	-6	0
Calculated point 15	228	30			
Determined point 15	260	48	-32	-18	1
Calculated point 16	162	90			
Determined point 16	195	96	-33	-6	1
Calculated point 17	32	210			
Determined point 17	81	264	-49	-54	0
Calculated point 18	32	90			
Determined point 18	97	96	-65	-6	0
Calculated point 19	228	90			
Determined point 19	292	48	-64	42	0
Calculated point 20	32	150			
Determined point 20	97	192	-65	-42	0
Mean value:			-40,4	-1,8	
Standard deviation:			13,03	33,55	

table 3: Normal user 2

The values of the deviations of table 3 allows the conclusion that an systematic error is hidden in this attempt. A reason therefore can be an unsuffiecient accurate calibration with same deviations in both dimensions. Because of the small item size (65x60) and a grid of 4x4 the system also reached the limits of accuracy. As seen below, for a grid larger than 3x3 it is recomended to use item sizes bigger than 65x60.
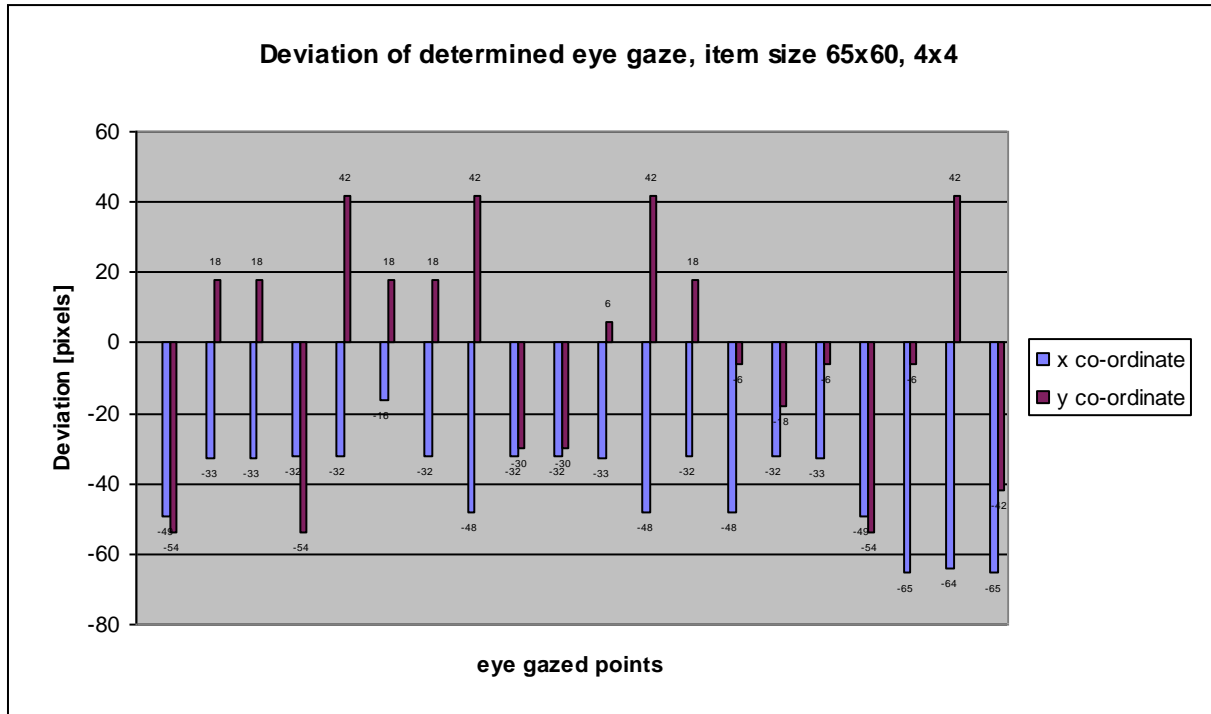
Test person 4: Normal user 3
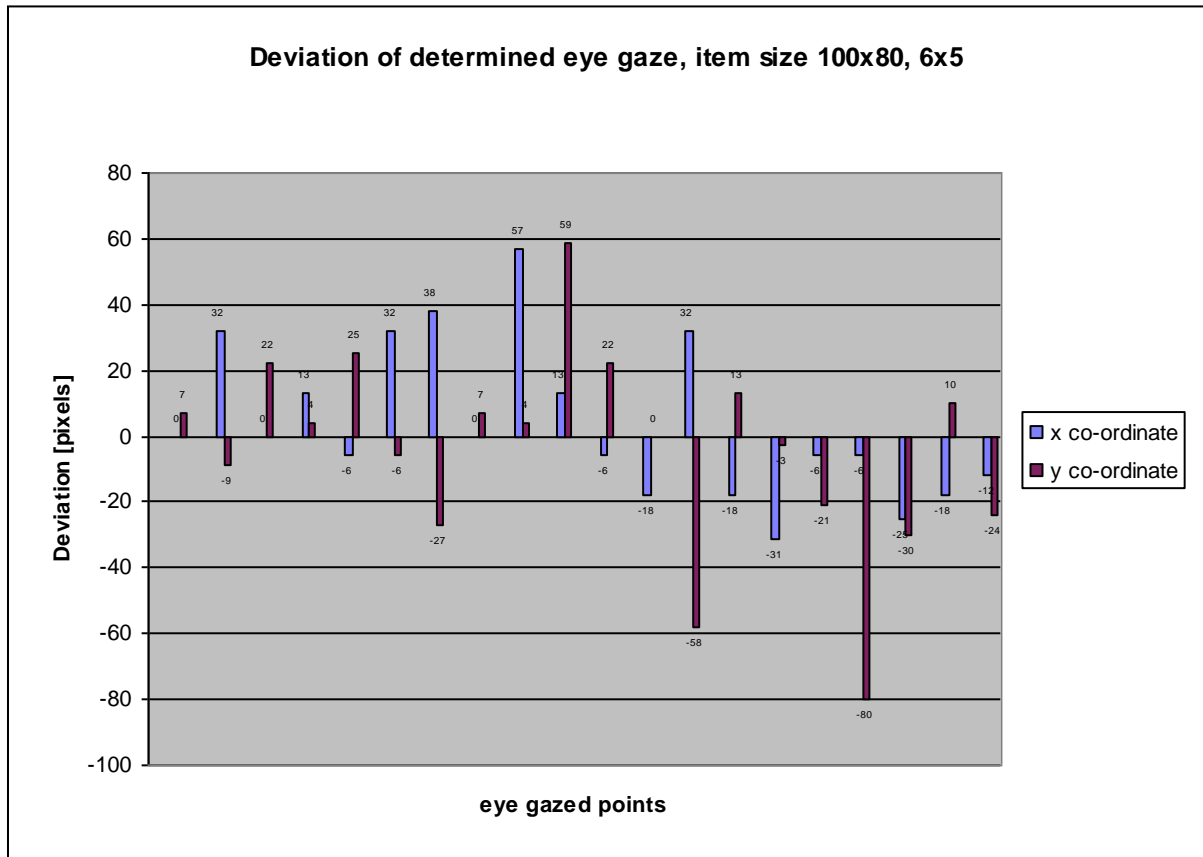
figure 92: Determined eye gaze deviation of normal user 3
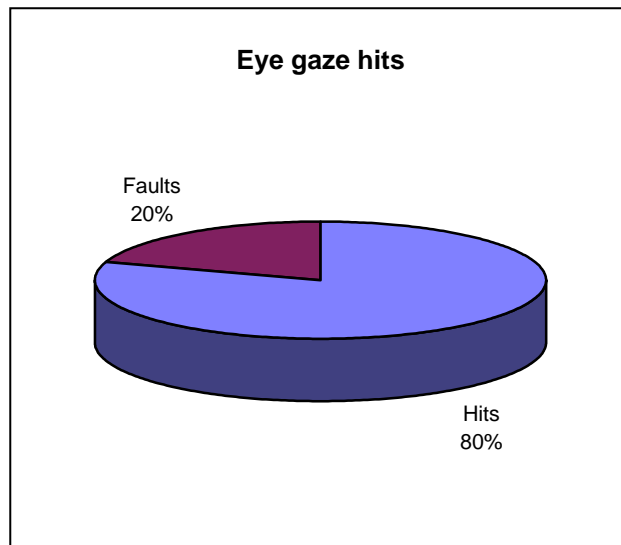
figure 93: Eye gaze hits of normal user 3

	x	y	Deviation x	Deviation y	Area hit
Number of items:	6	5			
Dimension of an item:	100	80			
Calculated point 1	450	360			
Determined point 1	450	353	0	7	1
Calculated point 2	350	360			
Determined point 2	318	369	32	-9	1
Calculated point 3	450	360			
Determined point 3	450	338	0	22	1
Calculated point 4	350	280			
Determined point 4	337	276	13	4	1
Calculated point 5	350	40			
Determined point 5	356	15	-6	25	1
Calculated point 6	350	40			
Determined point 6	318	46	32	-6	1
Calculated point 7	150	280			
Determined point 7	112	307	38	-27	1
Calculated point 8	450	360			
Determined point 8	450	353	0	7	1
Calculated point 9	150	280			
Determined point 9	93	276	57	4	0
Calculated point 10	50	120			
Determined point 10	37	61	13	59	0
Calculated point 11	350	360			
Determined point 11	356	338	-6	22	1
Calculated point 12	450	200			
Determined point 12	468	200	-18	0	1
Calculated point 13	50	280			
Determined point 13	18	338	32	-58	0
Calculated point 14	450	120			
Determined point 14	468	107	-18	13	1
Calculated point 15	550	120			
Determined point 15	581	123	-31	-3	1
Calculated point 16	350	40			
Determined point 16	356	61	-6	-21	1
Calculated point 17	50	120			
Determined point 17	56	200	-6	-80	0
Calculated point 18	350	200			
Determined point 18	375	230	-25	-30	1
Calculated point 19	450	40			
Determined point 19	468	30	-18	10	1
Calculated point 20	550	360			
Determined point 20	562	384	-12	-24	1
Mean value:			3,55	-4,25	
Standard deviation:			23,13	30,48	

table 4: Normal user 3

An error rate of 20% for an normal user with an grid size of 6x5 and item size of 100x80 indicates the high acceptance of this system. This testing results was yield with an sufficient accurate calibration stage.

Test person 5: Normal user 4
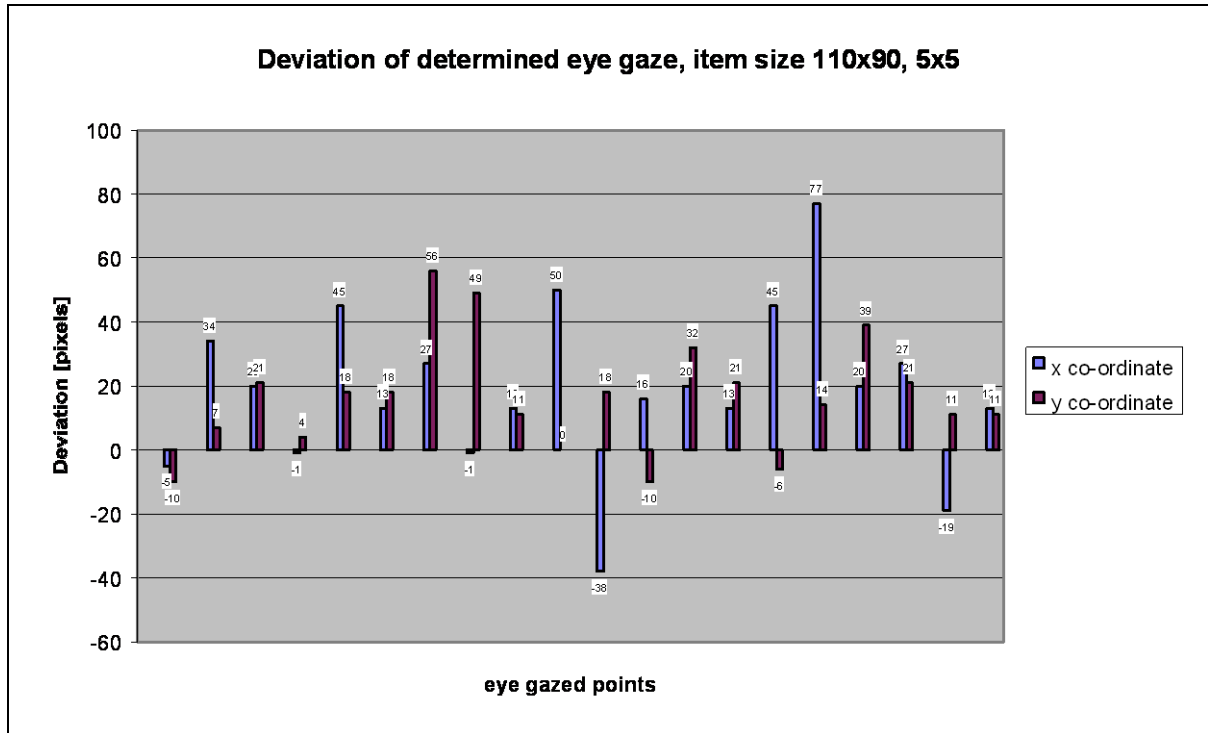
figure 94: Determined eye gaze deviation of normal user 4
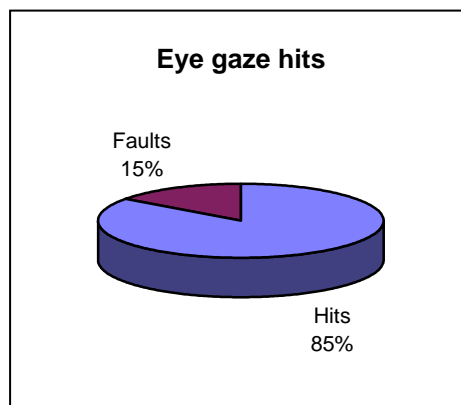
figure 95: Eye gaze hits of normal user 4

VELOC

	x	y	Deviation x	Deviation y	Area hit
Number of items:	6	5			
Dimension of an item:	100	90			
Calculated point 1	385	315			
Determined point 1	390	325	-5	-10	1
Calculated point 2	495	405			
Determined point 2	461	398	34	7	1
Calculated point 3	55	315			
Determined point 3	35	294	20	21	1
Calculated point 4	495	315			
Determined point 4	496	311	-1	4	1
Calculated point 5	275	225			
Determined point 5	230	207	45	18	1
Calculated point 6	385	225			
Determined point 6	372	207	13	18	1
Calculated point 7	275	315			
Determined point 7	248	259	27	56	0
Calculated point 8	495	135			
Determined point 8	496	86	-1	49	0
Calculated point 9	385	45			
Determined point 9	372	34	13	11	1
Calculated point 10	165	225			
Determined point 10	115	225	50	0	1
Calculated point 11	165	135			
Determined point 11	203	117	-38	18	1
Calculated point 12	495	405			
Determined point 12	479	415	16	-10	1
Calculated point 13	55	135			
Determined point 13	35	103	20	32	1
Calculated point 14	385	315			
Determined point 14	372	294	13	21	1
Calculated point 15	275	45			
Determined point 15	230	51	45	-6	1
Calculated point 16	165	135			
Determined point 16	88	121	77	14	0
Calculated point 17	55	135			
Determined point 17	35	96	20	39	1
Calculated point 18	275	315			
Determined point 18	248	294	27	21	1
Calculated point 19	495	45			
Determined point 19	514	34	-19	11	1
Calculated point 20	385	45			
Determined point 20	372	34	13	11	1
Mean value:			18,45	16,25	
Standard deviation:			24,75	17,70	

table 5: Normal user 4

This test indicates the typical improvement of area hits by increasing the item size. Both, the x-deviations and the y-deviations are regular spread which indicates a successful calibartion.

Test person 6: Normal user 5

figure 96: Determined eye gaze deviation of user normal 5

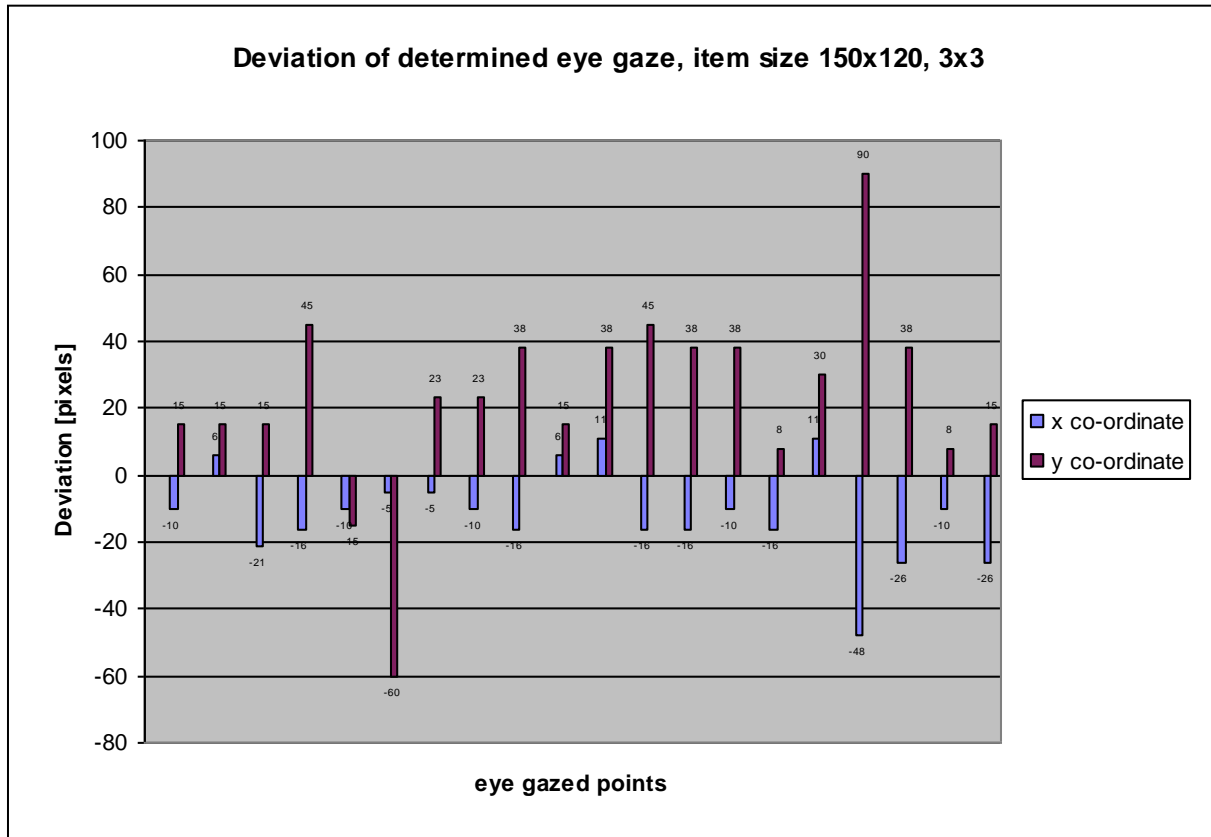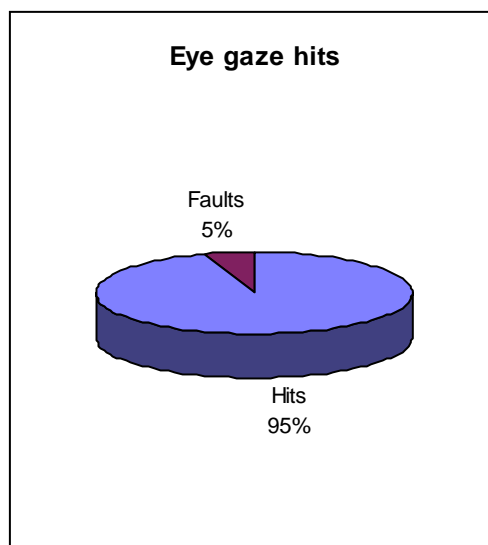**figure 97: Eye gaze hits of normal user 5**

	x	y	Deviation x	Deviation y	Area hit
Number of items:	3	3			
Dimension of an item:	150	120	x co-ordinate	y co-ordinate	
Calculated point 1	375	60			
Determined point 1	385	45	-10	15	1
Calculated point 2	375	60			
Determined point 2	369	45	6	15	1
Calculated point 3	75	60			
Determined point 3	96	45	-21	15	1
Calculated point 4	225	180			
Determined point 4	241	135	-16	45	1
Calculated point 5	375	300			
Determined point 5	385	315	-10	-15	1
Calculated point 6	75	300			
Determined point 6	80	360	-5	-60	1
Calculated point 7	75	180			
Determined point 7	80	157	-5	23	1
Calculated point 8	375	180			
Determined point 8	385	157	-10	23	1
Calculated point 9	225	60			
Determined point 9	241	22	-16	38	1
Calculated point 10	375	60			
Determined point 10	369	45	6	15	1
Calculated point 11	75	60			
Determined point 11	64	22	11	38	1
Calculated point 12	225	180			
Determined point 12	241	135	-16	45	1
Calculated point 13	225	60			
Determined point 13	241	22	-16	38	1
Calculated point 14	375	60			
Determined point 14	385	22	-10	38	1
Calculated point 15	225	300			
Determined point 15	241	292	-16	8	1
Calculated point 16	75	300			
Determined point 16	64	270	11	30	1
Calculated point 17	225	180			
Determined point 17	273	90	-48	90	0
Calculated point 18	375	60			
Determined point 18	401	22	-26	38	1
Calculated point 19	375	300			
Determined point 19	385	292	-10	8	1
Calculated point 20	375	60			
Determined point 20	401	45	-26	15	1
Mean value:			-11,35	23,1	
Standard deviation:			13,54	28,78	

table 6: Normal user 5

An grid size of 3x3 and item size of 150x120 pixels offers the user a reliable input interface. An error rate of 5% for normal users meets most reqirements for simple input interaction.
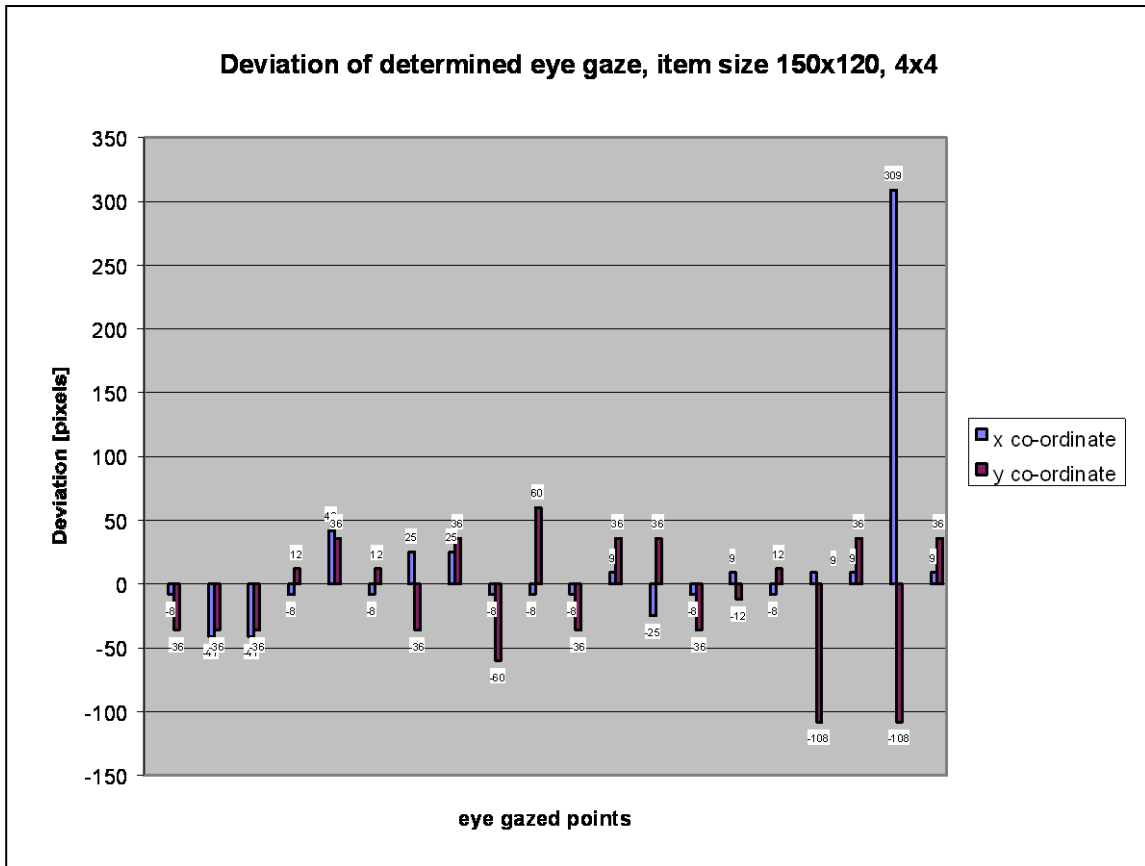
Test person 7: Normal user 6

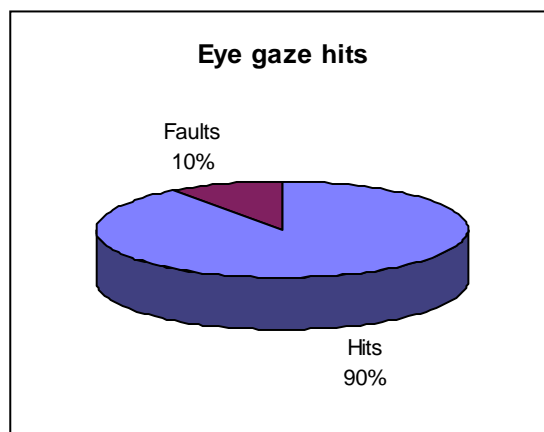**figure 98: Determined eye gaze deviation of normal user 6**

figure 99: Eye gaze hits of normal user 6

	x	y	Deviation x	Deviation y	Area hit
Number of items:	4	4			
Dimension of an item:	150	120			
Calculated point 1	75	60			
Determined point 1	83	96	-8	-36	1
Calculated point 2	375	300			
Determined point 2	416	336	-41	-36	1
Calculated point 3	75	300			
Determined point 3	116	336	-41	-36	1
Calculated point 4	375	60			
Determined point 4	383	48	-8	12	1
Calculated point 5	75	180			
Determined point 5	33	144	42	36	1
Calculated point 6	375	60			
Determined point 6	383	48	-8	12	1
Calculated point 7	525	300			
Determined point 7	500	336	25	-36	1
Calculated point 8	75	180			
Determined point 8	50	144	25	36	1
Calculated point 9	225	420			
Determined point 9	233	480	-8	-60	1
Calculated point 10	225	180			
Determined point 10	233	120	-8	60	1
Calculated point 11	375	300			
Determined point 11	383	336	-8	-36	1
Calculated point 12	75	60			
Determined point 12	66	24	9	36	1
Calculated point 13	375	60			
Determined point 13	400	24	-25	36	1
Calculated point 14	525	300			
Determined point 14	533	336	-8	-36	1
Calculated point 15	75	300			
Determined point 15	66	312	9	-12	1
Calculated point 16	225	180			
Determined point 16	233	168	-8	12	1
Calculated point 17	75	420			
Determined point 17	66	528	9	-108	0
Calculated point 18	75	60			
Determined point 18	66	24	9	36	1
Calculated point 19	375	420			
Determined point 19	66	528	309	-108	0
Calculated point 20	75	60			
Determined point 20	66	24	9	36	1
Mean value:			13,75	-9,6	
Standard deviation:			70,58	48,57	

table 7: Normal user 6

Best results can be achieved with an item size of 150x120 pixels as shown with this test person.
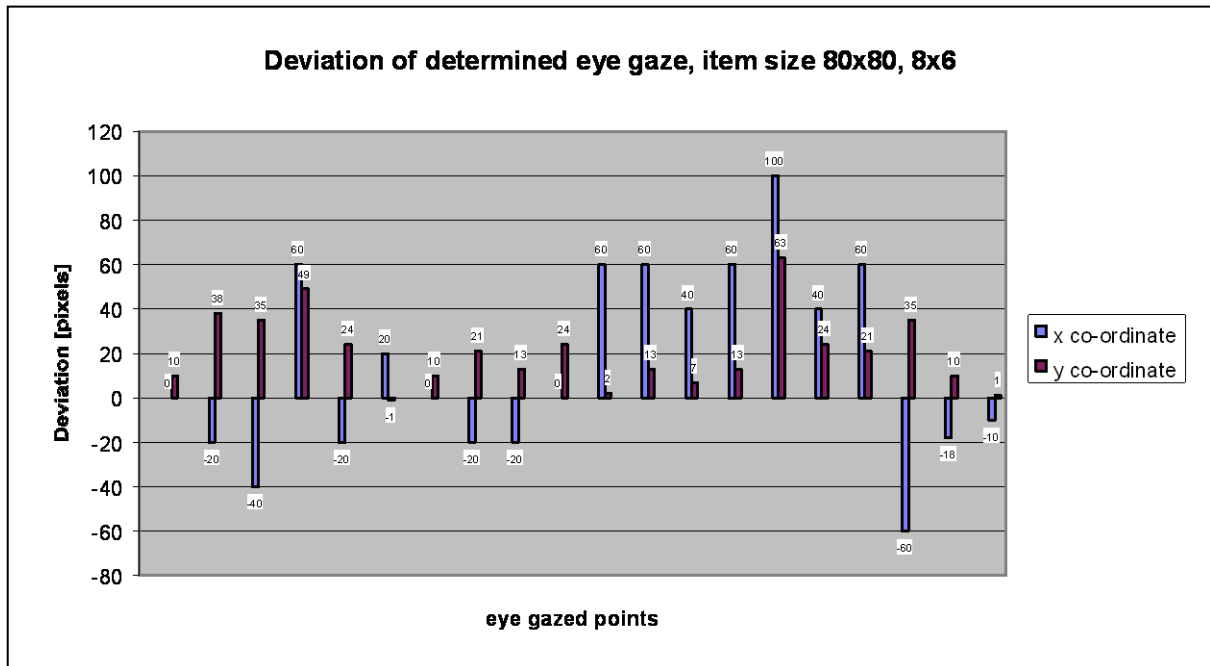
Test person 8: Normal user 7

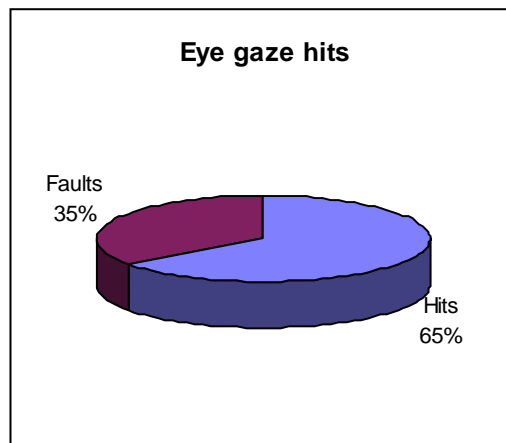**figure 100: Determined eye gaze deviation of normal user 7**

figure 101: Eye gaze hits of normal user 7

	x	y	Deviation x	Deviation y	Area hit
Number of items:	8	6			
Dimension of an item:	80	80			
Calculated point 1	520	440			
Determined point 1	520	430	0	10	1
Calculated point 2	440	120			
Determined point 2	460	82	-20	38	1
Calculated point 3	40	200			
Determined point 3	80	165	-40	35	1
Calculated point 4	200	280			
Determined point 4	140	231	60	49	0
Calculated point 5	200	40			
Determined point 5	220	16	-20	24	1
Calculated point 6	360	280			
Determined point 6	340	281	20	-1	1
Calculated point 7	600	440			
Determined point 7	600	430	0	10	1
Calculated point 8	440	120			
Determined point 8	460	99	-20	21	1
Calculated point 9	520	360			
Determined point 9	540	347	-20	13	1
Calculated point 10	360	40			
Determined point 10	360	16	0	24	1
Calculated point 11	200	200			
Determined point 11	140	198	60	2	0
Calculated point 12	200	360			
Determined point 12	140	347	60	13	0
Calculated point 13	40	40			
Determined point 13	0	33	40	7	1
Calculated point 14	200	360			
Determined point 14	140	347	60	13	0
Calculated point 15	120	360			
Determined point 15	20	297	100	63	0
Calculated point 16	280	40			
Determined point 16	240	16	40	24	1
Calculated point 17	280	120			
Determined point 17	220	99	60	21	0
Calculated point 18	40	200			
Determined point 18	100	165	-60	35	0
Calculated point 19	450	40			
Determined point 19	468	30	-18	10	1
Calculated point 20	200	200			
Determined point 20	210	199	-10	1	1
Mean value:			14,6	20,6	
Standard deviation:			41,57	16,62	

table 8: Normal user 7

The small item size and the high number of grid items is resonsible for the relatively high error rate.

During the evaluation, some statistical figures can be seen from the result. It is very important to note that no result shown before has an fault rate worse than 50%. This is sign of high acceptance by the users, because none of them have never seen and operated with the system before. Advanced user are able to yield a hit rate of 100%.

The following coherence between the difference of pixels in the camera image and the size of the eye gaze sensitive area distance in relation to the distance of the users head to the monitor can be written as

$$\frac{x}{w} = \frac{c_1}{d} \quad \text{resp.} \quad \frac{y}{h} = \frac{c_2}{d} \tag{30}$$

x...difference of the x co-ordinates of the pupil centre in the eye image of the camera [pixel].

y...difference of the y co-ordinates of the pupil centre by looking in the upper and lower corners of the eye gaze sensitive area [pixel].

w...width of this eye gaze sensitive area on screen [mm]

h...heigth of this eye gaze sensitive area on screen [mm]

c_1...constant factor for x dimension[pixels]

d...distance between the eye and the monitor[mm]

Example:

A difference of 25 pixels for x (see figure 86) by looking in the left and right corners of the eye gaze sensitive area, a width of 120mm for w and a distance of 300mm between the eye and the monitor yields c_1=62.5 .Now, the resolution of the system can be easily determined by dividing c_1 by the distance between the monitor and the eye.

During the evaluation of the VELOC system following conclusions were found:

- The system allows an advanced user to select buttons which have a size on the screen of 2cmx2cm for sure due to the maximum provided resolution.

- The calibration stage is the most critical part of the session and attention has to be paid. An unsufficient accuracy of calibration will yield dissatisfied results.

- A comfortable session without hit faults can be achieved by adjusting the grid size to 4x4 and setting the item dimension to 150x120 pixels or 4x3cm.

- Small head movements after the calibration stage are not critical if they can be corrected

- Too large x dimension of the eye sensitive area causes wide eye movements. Regarding to the geometrical position of the camera infront of the eye nonlinearities in the captured eye image can occur. This means that a circle becomes an ellipse resulting in lower accuracy of centre detection.

- The prototyp of the VELOC system provides the facility of selecting at least 12 different buttons by emulating the mouse click with an eye blink without problems. The blinking operation never caused problems and was therefore not tested. Double click can be achieved by closing the eye lid twice long compared to a single cick

9   Conclusion and Further Work

This thesis has attempted to give an overview of eye-gaze methods; how the techniques work, their backgrounds, current implementations, possible future potentials and introduced the VELOC system developed for this thesis.

The generally rising interest in eye-gaze media is a result of the development of unobstructive, video-camera-based eye-tracking techniques that make it possible to use eye-movements for real-time feed-back to the eye-tracked user. In this way eyes can also be used as output-organs, thus increasing the bandwidth of communication from the user to the computer. Using eye-gaze for interaction in multi-modal non-command user interfaces where the interface itself is as transparent as possible and the user simply interacts with his actual data, will result in a more „natural„ and „direct„ way of manipulating data.

9.1 Image Acquisition

The accuracy of the VELOC system is mainly determined by the resolution of the captured image. Owing to this fact an improvement can easily be made by using a camera which provides high resolution images. The drawback of this solution is that it costs more computation time to go through the entire algorithm. Furthermore, since the amount of data in each frame is limited, the merging of data from several frames must be considered in future. This is also a question of computation power of the processor used in the computer.

9.2 Hough Transformation

A second improvement can be achieved by performing the Hough Transformation in hardware. First of all the memory needed for this operation is very large It could probably be implemented in a stand-alone device without increasing the cost of it. Some research has been done to reduce the required memory space but the necessary memory is directly related to the size and the resolution of the image so that a compromise has to be found. As described in previous chapters, the Hough transformation can be adapted not only to circles; it can be adapted to ellipse or any other curve as well. An attempt for improving accuracy could be to detect the eyelid with an ellipse search algorithm. This would mean that the final result of the correct eye position is based on analysis of a greater part of the image. This is especially relevant for users whose eyelid cuts across the circle of their iris. This modification would certainly more satisfy their requirements.

9.3 Light Conditions

Light conditions also play a very important part in the accuracy of the VELOC system. As shown in chapter 4.3.2.1 reflections can disturb the clear shape between the iris and the sclera. Experiments with polarisation filters have been done but no improvement could be made. Even the light coming from the monitor causes reflection in the surface of the eye. The only way to avoid these reflections is to use infrared light. Infrared light produces a bright spot which can be detected easily. This in turn would mean alteration of the algorithm.

9.4 Error Precautions

While the algorithm and statistical methods for measuring and comparing the dimension to the origins set by a pre-set parameters which must be applied to the correct regions of the image if the results are to be meaningful, a series of cross-checks might be developed by measuring both the vertical and horizontal dimensions, corresponding centre estimates could then be compared. For example, if the vertical and horizontal estimate are both greater than

the mean iris diameter then it is likely that an error has occurred. Investigations for an automatic correction of typical measurement and calculation errors should also be made.

9.5 Real-time issues

The cost of real-time analysis must be kept to a minimum by improving the existing algorithms as much as possible before resorting to more processing power. An adapted search area provides one method of boosting performance. The search area used in the VELOC system is a preset area determined during calibration cycles and cannot be varied during operation time. An advanced algorithm could start searching in the area of the last centre for the new iris centre in the new captured image. Subsequent analysis would consider only those areas of the image defined as tracking regions, thus saving a considerable amount of processing time. Eventually, once the true iris centre is located, the main problem would be to keep the tracking regions centred on them. A backup algorithm would be required to interject whenever the iris was lost, in which case it would be necessary to re-start the analysis from the beginning.

9.6 Positioning of the camera

In the current development stage of the VELOC system, the LCD camera is mounted on a spectacle frame directly in front of the right eye. The right eye is therefore completely covered by the camera and has no direct view to the monitor. Owing to the fact that interacting with only one eye tires much more, the usage time for it is limited. In the next development stage of the VELOC system the camera should be mounted on the side of the head on the spectacle frame with the eye's image reflected by a mirror. Care has to be taken that the image will not be stretched so that the circle becomes an ellipse which is much more difficult to detect.

9.7 Principle problems

If eye gaze systems are to have any noticeable impact on user interfaces, some technical problems have to be addressed:

- The eye tracking equipment must be much less sensitive to user movements. The user should be allowed to move the head in a range of $\pm45°$ for comfortable working conditions. It is not acceptable for the user to sit quite still throughout the interaction.

- The eye tracking device must not constantly require user-attended re-calibration. An automatic re-calibration feature should be implemented to avoid wrong inputs and commands.

- The eye tracking equipment should be able to identify the tracked persons, most likely by iris pattern recognition. This provides the opportunity to meet the requirements and needs of each single user in the best way by storing the identification information in a database.

- Difficulties occur when taking into consideration the tracking of several persons simultaneously. A single user interface might be enough for an interaction with a computer, but multi media devices (TV,...) are mostly used by several users at the same time.

9.8 Vision of the Future

This short chapter should give a taste of what the future for eye-gaze media could look like.

Eye-gaze systems have been around for about 20 years and a big effort has been made to improve the technology to yield suitable applications for different groups of users. A great challenge will be to implement this technology into a multi modal environment to offer users the best opportunities to interact with computers. In a multi modal computer interaction the user could chose the input device which best satisfies his/her needs. This would give disabled people a better chance and more possibilities to overcome their handicaps.

As ever, innovative technologies can often take a long time to reach the general community. The typical scenario for new technologies is that for a long time they only exist in very few applications, typically owned by research centres or big industrial companies. When a sufficient number of applications exist the market starts to grow, causing the cost of the new technology to fall, finally resulting in an almost exponential growth in the number of applications utilising the new technology.

There are many current day interfaces where eye gazed systems could improve handling and accelerate the way of working. Examples can be found everywhere in our daily work. A typically application of eye-gazed tracking would be for telephone operators (for example at the directory inquiries of the telephone company) to look up information for the calling customer. Instead of typing , which is slow and can perhaps cause typing injuries, the operator could look his way through several menu levels to find the relevant information. Frequently accessed information will be quickly accessible, and the general concept of viewing the data as a hyper search space instead of a list of data (i.e. a telephone book will make for more natural perception of it.

Eye-gaze systems could also be established wherever human beings are faced with the problem of the amount of information. To keep track of the overall impression of the entire system and small details as well, eye-gazed systems could be used to provide the user with only the information that he/she is currently interested in. That might mean where he/she is looking at the moment. One possible solution might be that the user can zoom with his eyes for more information on the subject of regard on the 2 dimensional limited screen display.

Eye gaze systems might also find application in automobiles. This technology could watch the user's eye and sound an alert signal if the user gets too tired to perform the driving task. An eye gazed system might add an advantage in this situation by monitoring the eye gaze pattern, blink rate and pupil size. It also could provide useful acoustic information on an object where the user is looking without looking off the street. Another good example where eye gazed systems might be applied would be surveillance tasks.

Care must be taken though, that eye-gaze tracking data is used in a sensible way, since the nature of human eye-movements is a combination of several voluntary and involuntary cognitive processes. When being able to track using, say, computers in a network, should it be permitted by the law to track them during work, to make an assessment of how active or productive they are? Is this too close to the Orwellian 'Big brother'? On the other hand this information might be used to improve computer interaction of disabled users to aid the control of the application.

10 Appendix

This appendix lists some screen shots which were used during the evaluation of the VELOC system and represent the results yield by the users.

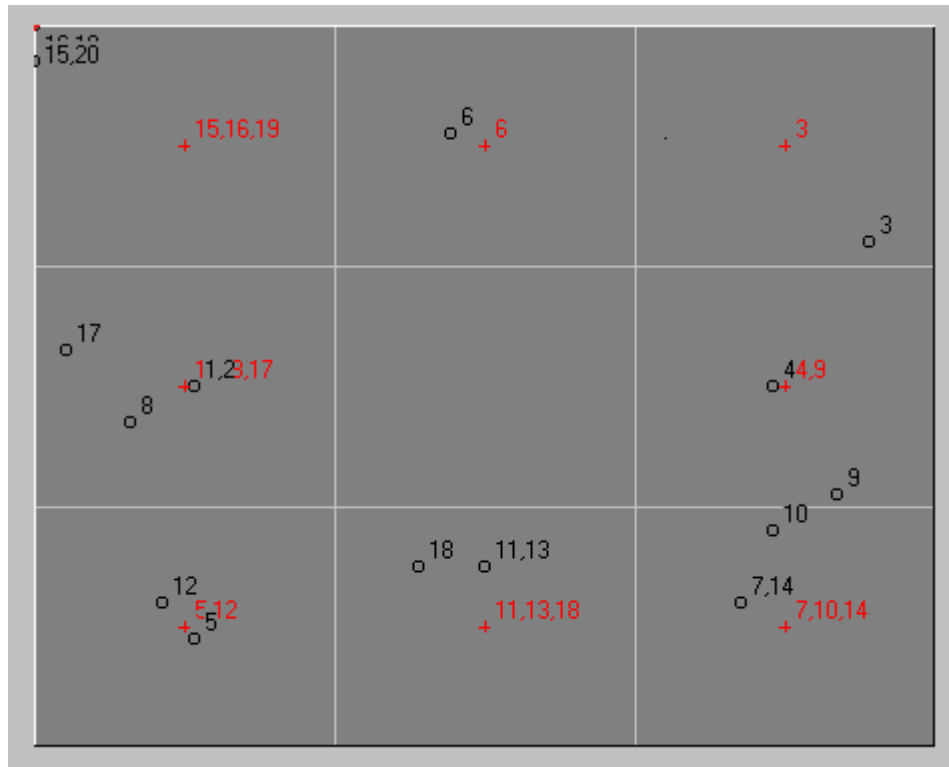In figure 107 the protype of the VELOC head apparatus is shown.

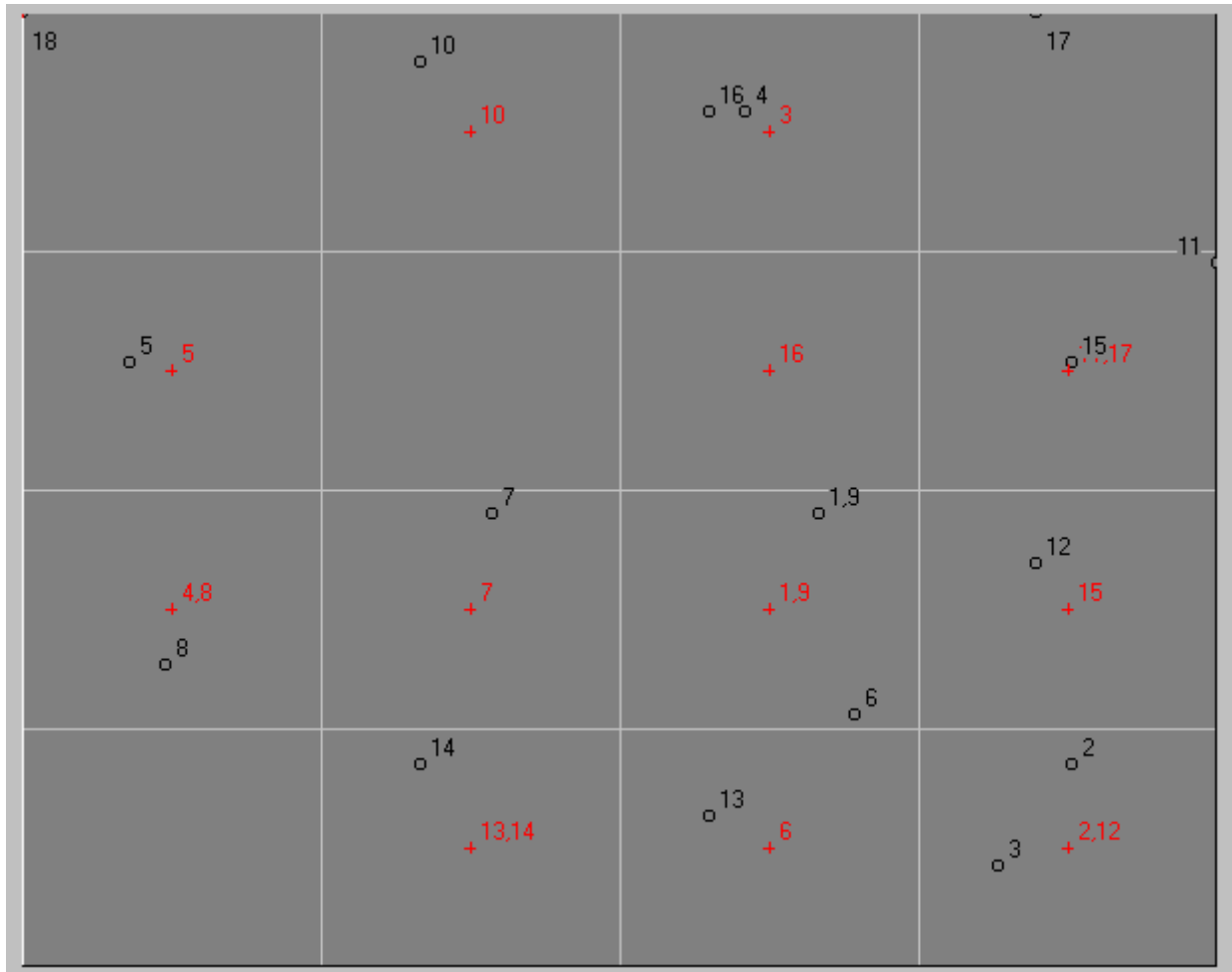**figure 102: Screen shot advanced user**
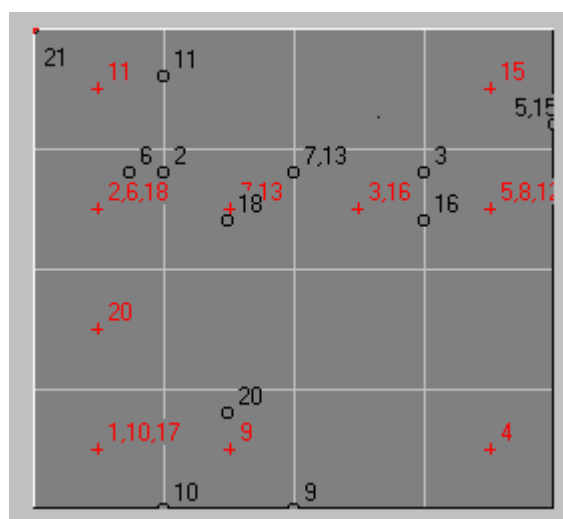
figure 103: Screen shot normal user 1

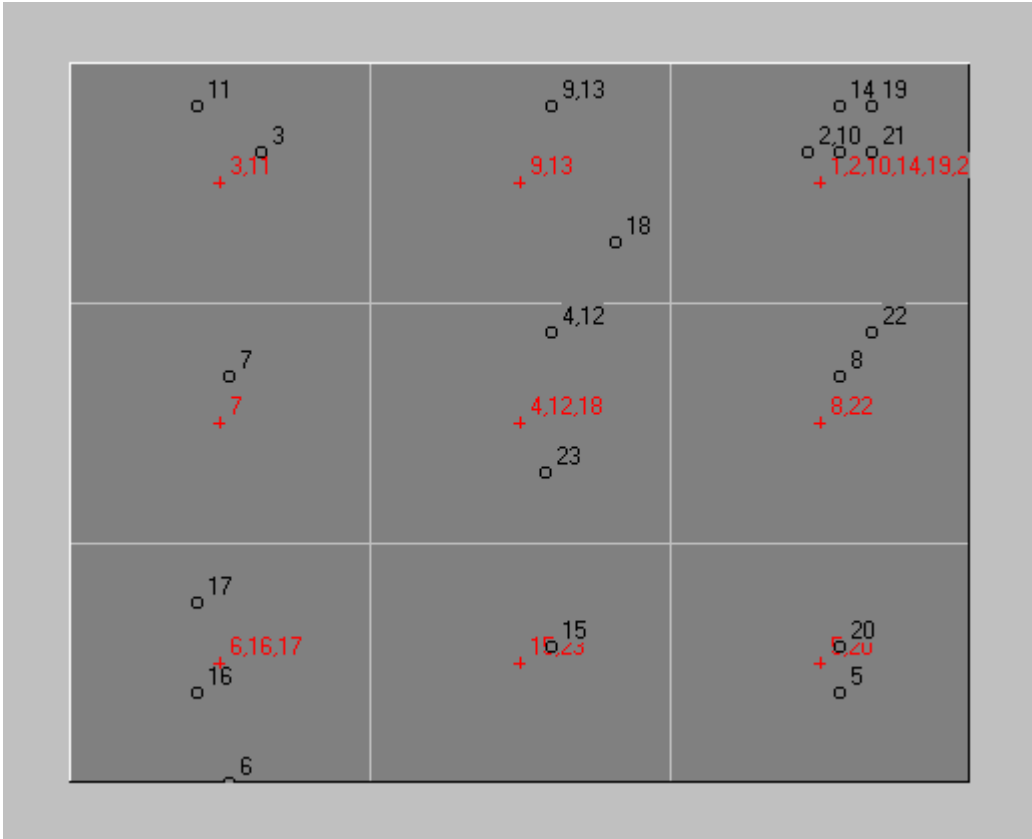**figure 104: Screen shot normal user 2**
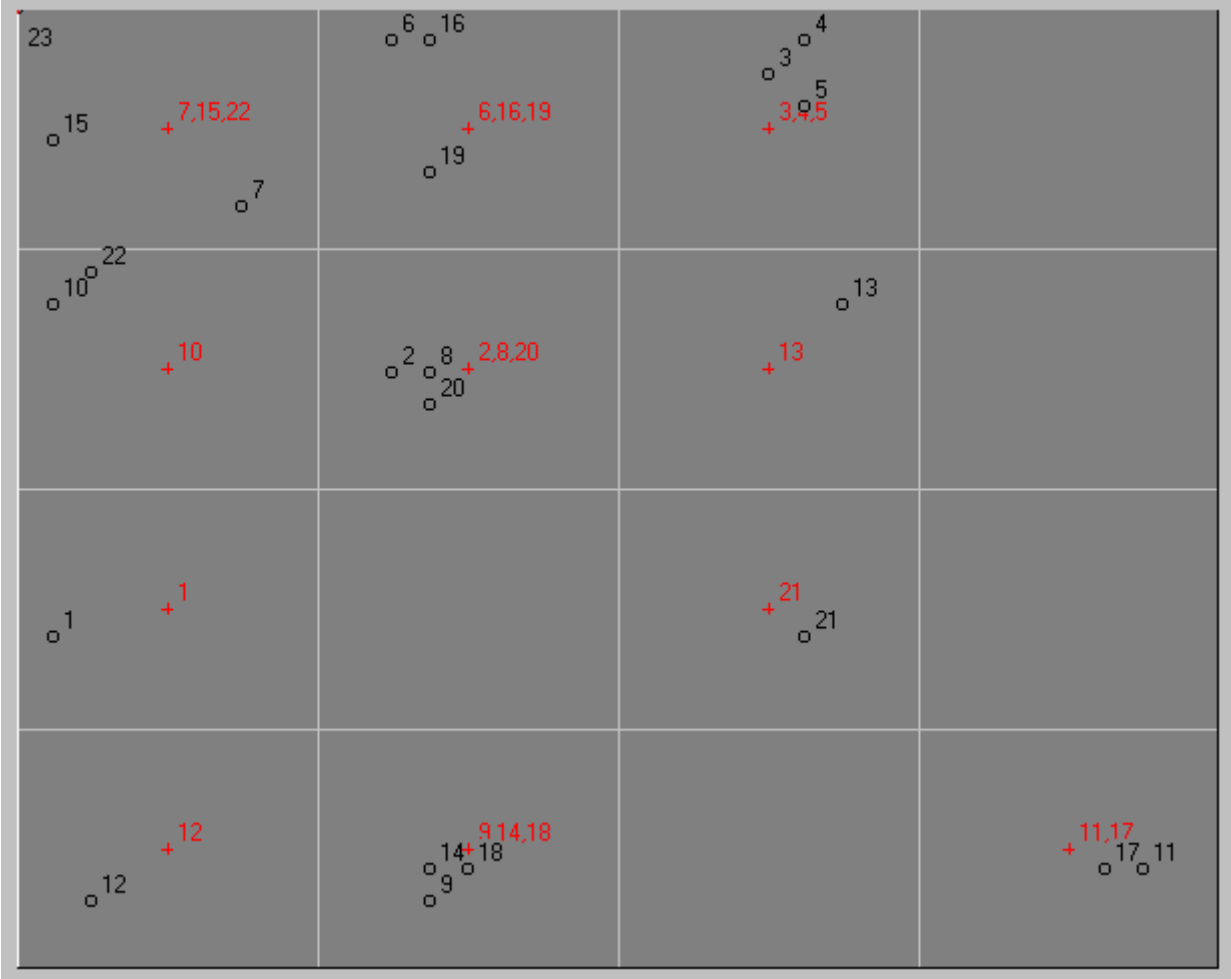
figure 105: Screen shot normal user 5

figure 106: Sscreen shot normal user 6

figure 107: VELOC prototype

11 References

[1]   Just M.A, Carpenter P.A; *A theory of reading: From eye fixations to comprehension.* Psychological Review 87(4), 1980, pp329-354

[2]   Vanderheiden Gregg C; *Thirty -Something Million: Should They Be Exceptions?* Trace Research and Development Center, Waisman Center and Department of Industrial Engineering,University of Wisconsin-Madison. http://www.trace.wisc.edu/docs/30_some/30_some.htm; 2.9.1998

[3]   The United Nations Declaration on the Rights of Disabled Persons. http://www.unhchr.ch/html/intlinst.htm; 2.9.1998

[4]   Edwards Alistair D. N; *Extra-Ordinary Human-Computer Interaction: Interfaces for Users with Disabilities.* Cambridge University Press 1995, ISBN 0-521-43413-0

[5]   Scaaden L.A, Vanderheiden G.C; *Consideration in the Design of Computers and Operating Systems to Increase Their Accessibility to Persons with Disabilities.* Working document of the Industry/Government Co-operative Initiative on Computer Accessibility. Trace Center, University of Wisconsin, 1988. http://www.trace.wisc.edu/docs/considerations/consider.txt; 2.9.1998

[6]   Foley James D. et al; *Computer graphics : principles and practice.* Addison-Wesley, Reading, Mass. 2. edition 1990

[7]   Guidelines for the design of consumer products to increase their accessibility to persons with disabilities or who are aging. http://www.trace.wisc.edu/world/consumer_products/; 2.9.1998

[8]   Vanderheiden, G.C. et al; *A Protocol for Communication Between a Device and an Auxiliary Interface.* working paper of the Trace R&D Center, 7/15/96. http://www.trace.wisc.edu/docs/uaip/uaip.htm; 2.9.1998

[9]   Vanderheiden, G.C, Bower, R; *A Collaborative Information Model for Next Generation Information Systems.* University of Wisconsin-Madison, 1996.

[10] Vanderheiden G.C, Kaine-Krolak; *Access to Current and Next-Generation Information Systems by People with Disabilities.* University of Wisconsin-Madison, 1995. http://www.trace.wisc.edu/docs/access_info_sys/full_doc.htm; 2.9.1998

[11] Vanderheiden G.C; *Application Software Design Guidelines: Increasing the Accessibility of Application Software to People with Disabilities and Older Users.* Trace R & D Center, University of Wisconsin-Madison , Dept of Industrial Engineering, 1994. http://www.trace.wisc.edu/docs/software_guidelines/toc.htm; 2.9.1998

[12] Search engine: http://trace.wisc.edu/search/; 2.9.1998

[13] Human eye anatomy: http://www.eyenet.org/public/anatomy.html; 2.9.1998

[14] Product information: http://www.visionforlife.com/understand/optics.htm; 2.9.1998

[15] Hallett, P. E; *Handbook of Perception and Human Performance I.* New York, Wiley, 1986, pp10.25-10.28

[16] Scott D, Findlay, J. M; *Visual search, eye movements and display units.* Human factors report, University of Durham, 1993

[17]   Wooding D; *Eye-Movement Monitoring Equipment.* University of Wales College of Medicine, Cardiff, UK, 1995

[18]   Baluja S, Pomerleau D; *Non-intrusive gaze tracking using artificial neural networks.* Research Paper CMU-CS-94-102, School of Computer Science, Carnegie Mellon University, Pittsburgh PA, USA, 1994

[19]   Hough P.V.C; *Methods and Means for Recognising Complex Patterns.* U.S. Patent 3,069,6541962

[20]   Scott D, Findlay J. M; E*ye movements and display units.* factors report, University of Durham,

[21]   Xie X, Sudhakar R, Zhuang H; *Real-time eye feature tracking from a video image sequence using Kalman filter.* IEEE Transactions on systems, man, and cybernetics, Vol. 25, No. 12, 1995

[22]   Baluja S, Pomerleau D; *Non-intrusive gaze tracking using artificial neural networks.* Research Paper CMU-CS-94-102, School of Computer Science, Carnegie Mellon University, Pittsburgh PA, USA, 1994

[23]   Müller P.U, Cavegn D. et al; *A comparison of a new limbus tracker, corneal reflection technique, Purkinje eye tracking and electro-oculography.* Elsevier Science Publishers, 1993, pp393-401

[24]   Cleveland D, Cleveland N; *Eyegaze eyetracking system.* Eleventh Monte-Carlo International Forum on New Images, LC Technologies, 1992

[25]   Product information: http://www.smi.de/em/embody2.htm; 2.9.1998

[26]   Product information: http://www.smi.de/el/elsys.htm; 2.9.1998

[27]   Barret J.R, Trainor F*; Eye Movements can control localised image enhancement and analysis.* RadioGraphics, 1997, 17:525-530

[28]   Product information: http://www.lctinc.com/doc/optic.htm; 2.9.1998

[29]   Allen J; *Synthesis of Speech from unrestricted Text.* Proc IEEE 64, 1976, 422-433

[30]   Eye-Movements Internet Resources: http://info.ox.ac.uk/~eye/virtua~2.htm; 2.9.1998