IST-2003-511598 (NoE)

COGAIN

Communication by Gaze Interaction

Network of Excellence

Information Society Technologies

D5.6 Report on New Approaches to Eye Tracking. Summary of new algorithms

Due date of deliverable: 31.08.2008
Actual submission date: 15.10.2008

Start date of project: 1.9.2004 Duration: 60 months

Public University of Navarra

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Villanueva, A., Cabeza, R., Porta, S., Böhme, M., Droege, D. and Mulvey, F. (2008) D5.6 Report on New Approaches to Eye Tracking. Summary of new algorithms. Communication by Gaze Interaction (COGAIN), IST-2003-511598: Deliverable 5.6. Available at http://www.cogain.org/results/reports/COGAIN-D5.6.pdf

Main Authors:

Arantxa Villanueva (UPNA, contact person)

Rafael Cabeza (UPNA)

Sonia Porta (UPNA)

Martin Böhme (UzL)

Detlev Droege (UNI KO-LD)

Fiona Mulvey (TU Dresden)

Contributors:

Luis Figueiredo (ESTG/IPG Guarda, Portugal)
Ana Isabel Gomes (Agrupamento Escolas Infante, Portugal)
Detlev Droege (University of Koblenz-Landau)
Thorsten Geier (University of Koblenz-Landau)
Dietrich Paulus (University of Koblenz-Landau)
Carola Schmidt (University of Koblenz-Landau)
Javier San Agustin (IT- University of Copenhagen)
John Paulin Hansen (IT- University of Copenhagen)
Dongheng Li (Human Computer Interaction Program Iowa State University)
Derrick Parkhurst (Human Computer Interaction Program Iowa State University)
Donatas Dervinis (Siauliai University)
Gintautas Daunys (Siauliai University)
Arantxa Villanueva (Public University of Navarra)
Dan Witzner Hansen (IT- University of Copenhagen)
Rafael Cabeza (Public University of Navarra)
Sascha Weber (Technical University Dresden, Dresden)
Dale Grover (Electrical and Computer Engineering Department, Michigan State University)
Fangmin Shi (Applied Vision Research Centre (AVRC), Loughborough University)
Alastair Gale (Applied Vision Research Centre (AVRC), Loughborough University)
Susanna Nilsson (Department of Computer and Information Science, University of Linköping)
Torbjörn Gustafsson (Division of Sensor TechnologySwedish Defence Research Agency)
Per Carleberg (Division of Sensor TechnologySwedish Defence Research Agency)
Tiago Nunes (ESTG/IPG Guarda, Portugal)
Filipe Caetano (ESTG/IPG Guarda, Portugal)
Petr Novák (Dept. of Cybernetics, Faculty of Electrical Engineering, Czech Technical Univ.)
Tomáš Krajník (Dept. of Cybernetics, Faculty of Electrical Engineering, Czech Technical Univ.)
Libor Preucil (Dept. of Cybernetics, Faculty of Electrical Engineering, Czech Technical Univ.)
Marcela Fejtová (Dept. of Cybernetics, Faculty of Electrical Engineering, Czech Technical Univ.)
Olga Štepánková (Dept. of Cybernetics, Faculty of Electrical Engineering, Czech Technical Univ.)
Martin Böhme (Institute for Neuro- and Bioinformatics, University of Lübeck)
André Meyer (Institute for Neuro- and Bioinformatics, University of Lübeck)
Thomas Martinetz (Institute for Neuro- and Bioinformatics, University of Lübeck)
Erhardt Barth (Institute for Neuro- and Bioinformatics, University of Lübeck)

Table of Contents

Executive Summary

This deliverable (D5.6) consists of a summary of the most recent studies related to eye-gaze tracking technology. The period from 2006 to 2008 has been selected since the D5.2 can be considered as covering the previous period. The most relevant work has been compiled as a first step, considering the most outstanding conferences of the field such as the Eye Tracking Research and Applications Conference (ETRA06, ETRA08) and the annual COGAIN conference (2006, 2007, 2008). In addition, journal papers with high impact factor have been reviewed. Many COGAIN members appear as authors and co-authors of recent relevant research papers. This deliverable aims to be a clarifying overview of the state of the art of this technology and most sought-after research lines for the future.

1 Introduction

There are two main working areas in eye tracking systems development. First, the image processing area is devoted to find the relevant features in the image of the eye, such as glints or pupil. Second, once the image features have been detected a mathematical procedure is needed to translate the image features into the screen or gaze coordinates. This is illustrated in Figure 1.

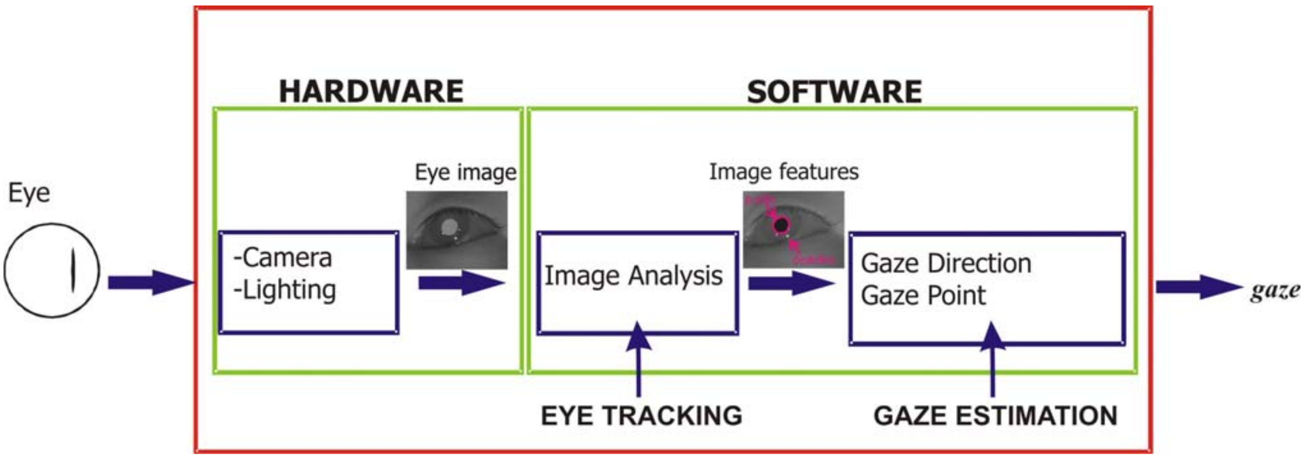

Figure 1. Schematic summary of eye-gaze tracking systems working area. Eye tracking focuses on image analysis, i.e. how to estimate working features from the image. Gaze estimation is the function that connects the image features to gaze data.

Issues related to image processing algorithms have been of high interest in the first periods of the technology. Eye tracking indoors can be considered as a solved issue with high resolution eye images and quasi-stable lighting conditions. However, there is still room to improve in i) outdoor scenarios in which light conditions can vary rapidly and ii) when trying to do eye tracking with off the shelf components such as web cameras in which the resolution of the eye image decreases considerably making the precise determination of features more difficult.

In recent years, the connection between the eye image (features) and the gaze coordinates (3-D/2-D) has become of great interest. The number of papers covering this topic has risen considerably in the last five years. The mathematical connection between image features and gaze direction or gaze position presents many potential benefits for gaze tracking technology; hence, it has attracted many researchers to focus on the topic.

To follow a summary of the most recent relevant work is presented below, divided into two sections: eye tracking and gaze estimation. The present document can be considered as a detailed review of the current research lines in eye-gaze tracking. The first section is devoted to the image processing part while the second one presents recent pieces of work describing gaze estimation methods, i.e. mathematical connection between the image and gaze. In the third section other work not entirely within the scope of the first two are described. Finally, some ideas for future research lines are suggested.

2 Eye Tracking Approaches

The price of eye tracking systems has always been an obstacle to overcome. The high costs of the hardware, software and distribution make this technology not available for new commercial applications. Exceptions can be found, such as Magic Eye Control (Figueiredo and Gomes, 2007). Apart from its price (it costs 1750 €), its most outstanding characteristic is that it uses a BW CMOS sensor with 1280x1024 resolution. The CMOS cameras permit to adjust image resolution and to select areas of interest from the scene. Once the pupil area has been detected in a low resolution image, this device allows for acquiring maximum resolution images of the region of interest centred in the users eye at 100 fps. This area of interest is updated as the user moves and can be reinitialized if the tracking is lost. The eye tracker uses four LEDs from which a minimum of three are used for gaze estimation. The system enables to locate the cursor in the screen with enough precision to close a Windows© window (X box) with a graphic resolution of 1024x768 (see Appendix I, paper 1: Magic Eye Control).

The University of Koblenz-Landau tries also to construct inexpensive eye trackers, such as the GoldenGaze. In their last work (Droege et al., 2007) an improved low cost eye tracker is presented. While it does not yet work entirely with commercial off the shelf (COTS) parts, most of its parts fulfil this aim. It consists of a high-sensitivity B/W camera (Sony EXView HAD CCD chip), equipped with a simple near-IR filter. In contrast to a previous setup, the IR-LEDs are now positioned below instead of besides the camera. This avoids shadowing the opposite eye by the user's nose and thus supports the usage of reflections in both eyes. In addition, this new implementation uses OpenCV Library available for Windows™ and Linux. The image processing algorithm is based on an adapted Hough transform. Due to its computational requirement the Hough transform is applied in a reduced size image and limiting Hough transform parameters variation ranges, such as the radius. Pupils are searched in both eyes, discarding those images for which no matching pairs are found. This provides an approximate shape of the pupil in the image. Taking this as a basis, a more accurate estimation of the pupil pixels is carried out and the centre of gravity calculated. The system can usually distinguish a 5x7 grid of points on the screen. Using both eyes improves notably the results comparing to their previous work. As next steps they try to find a source for highly sensitive mini cameras using a USB interface. The currently used device requires an analog video input (to be found as input in inexpensive TV tuner cards) and an external power supply. The IR illumination still is a custom built, USB powered solution (see Appendix I, paper 2: Improved Low Cost Gaze Tracker). In their recently paper Droege et al. (2008) review a considerable number of algorithms to detect the pupil in low resolution images (obtained using low cost devices). Working with low resolution images makes the detection of the pupil centre inaccurate (see Figure 2) if methods originally designed for higher resolution images are used. The paper presents a comparison of existing algorithms and proposes a new one with better results, with an accuracy of 2º of the screen resolution (see Appendix I, paper 3: A Comparison of Pupil Centre Estimation Algorithms).

Figure 2. Images of an eye using high-resolution (left) and low resolution (right) cameras. Accurate detection of the pupil centre is more difficult in the low resolution image.

Eye tracking based on off-the-shelf components has always been one of the objectives of the IT-University of Copenhagen. In Hansen and Hansen (2006) they presented a practical implementation of their method using an uncalibrated camera (Sony handy cam DCR-HC14E). By activating the option for “night-vision” the glint is created with the built in IR emitter. In this manner, the features employed by the method are obtained in the image. They propose to use the RANSAC algorithm for both estimating of the iris contour and reducing the number of outliers in the calibration procedure. The calibration is based on 9 points. The performance of this tracker is demonstrated by using the GazeTalk application developed by COGAIN. Recently, San Agustin and Hansen (2008) presented a similar implementation for a head mounted low cost prototype (see Appendix I, paper 4: Off-the-Shelf Mobile Gaze Interaction). The system uses a binocular head mounted display (glasses) to show the computer screen, making the system more portable and allowing full mobility of the user.

The aim of the Starburst (Li et al., 2005) algorithm is to provide a low-cost open-source eye-tracking system to integrate eye movements into interfaces (available at http://thirtysixthspan.com/openEyes/software.html). It has permitted several interface designers to use eye movements as an input to the computer using low-cost hardware. The recent work presented tries to eliminate the IR lighting from the system. Most video-oculography (VOG) systems use IR active illumination controlling the lighting and image exposure levels, thus facilitating the image processing part. However, removing the need of IR lighting would be of great interest for the technology. In addition, IR performs poorly outdoors due to existing ambient IR light. Li and Parkhurst (2006) present a new algorithm using visible-spectrum images. The most noticeable feature in these kinds of images is the limbus, i.e. the limit between the iris and the sclera. They adapt the Starburst algorithm originally designed to track the eye pupil in infrared spectrum to track the limbus. They adjust the limbus image to an ellipse and assume that the limbus is fixed with respect to the direction of gaze. The algorithm is robust against reflections from non-controlled light sources. The algorithm starts from a point calculated as the approximate limbus centre and following the Starburst method finds out the derivatives along rays extending radially from -45º to 45º and 135º to 225º. The points detected as limbus points are filtered afterwards using distance criteria to remove outliers. The remaining points are fitted to an ellipse using a RANSAC method and limiting values for the radius and pupil area. The algorithm is implemented in a low-cost eye tracker achieving an approximate 1º accuracy if the user remains still (see Appendix I, paper 5: Open-Source Software for Real-Time Visible-Spectrum Eye Tracking).

Recently, following the same philosophy, a limbus/pupil switching head mounted eye tracker has been presented by Ryan et al. (2008). They present a low-cost wearable eye tracker built from off-the-shelf components using the open source openEyes project (available at http://thirtysixthspan.com/openEyes/). The objective is to build an eye tracker that can operate in both, the visible spectrum and variable lighting conditions. The novelty of this approach is that it automatically switches between tracking the pupil/iris boundary in bright light to tracking the iris/sclera boundary (limbus) in dim light. It also introduces some changes in the Starburst algorithm. As in Starburst algorithm, feature points are detected by drawing rays from the approximate centre radially, with additional filtering process. The authors claim that although this technique is efficient it is sensitive to the threshold chosen to define the dark region. The algorithm proposed in this work iterates the process through multiple thresholds. Ellipses are fit to points generated at each threshold and only the best ellipses are kept. Regarding the ellipse fitting part they also propose an improvement by using a two step algorithm. They generate random ellipses and label each pixel that the ellipse passes through as acceptable or not depending on the magnitude and direction of the gradient at that pixel. The ellipse with the highest ratio of acceptable pixels is assumed to be the correct one. The modification makes the process more tolerant to poorly localized feature sets. The algorithm is further improved by making a previous differentiation of pixels belonging to the pupil and to the limbus based on pixel luminance by simply partitioning them of the median value. The prototype proposed is entirely based on COTS components and presents average accuracies below 2º.

Real world imaging conditions are also studied in the work by (Witzner and Hammoud, 2006). Shadows, light variations and eye partial occlusions make eye tracking systems to perform poorly in general. Problems that especially disturb eye tracking are head movement, eye blinking and light changes, all of which can cause the eyes to disappear. Normally, in such cases the solution is to restart the algorithm and try to find the image features in the whole image. This process can last several images, missing gaze information during this time. Witzner and Hammoud present an efficient and reliable method of tracking a human eye between successively produced video image frames, even in situations where the persons head turns, the eyes momentarily close and/or lighting conditions are variable. It proposes a log likelihood-ratio function of foreground and background models in a particle filtering filter-based eye tracking system. It uses bright-dark pupil images' difference as working image. Experimental validations show good performance of the proposed eye tracking method in variable lighting conditions and moderate head motion. The paper presents also an eye detector that relies on physiological infrared eye responses and a modified version of a cascaded classifier.

As mentioned before using COTS camera and performing gaze tracking in visible light eliminating the IR emitters is an interesting manner to reduce costs. As claimed in the work by Dervinis and Daunys (2007) from the Siauliai University the user's head orientation must be deduced to calculate the gaze position. In previous work (see COGAIN D5.2) they already had proposed a method for 3-D head orientation estimation using a single camera. In the present work they evaluate the influence of face expression in the estimation of head orientation and propose a method to compensate for it (see Appendix I, paper 6: 3D head orientation estimation and expression influence elimination using characteristic points of face).

Additional relevant contributions can be found in recent literature regarding image processing for eye tracking. Tan and Zhang (2005) present an algorithm to determine eye blink states by tracking iris and eyelids. This information can be highly relevant for the eye tracking system to differentiate between the alternative situations that make the eye disappear. In addition, several systems use eye blinking to produce activations (selections, mouse clicks). This method presents two relevant properties. It exploits simultaneously the intensity and edge information for detecting the eye state as well as the record of the patterns of eyelids before closing for tracking the reopened eyes. Khosravi and Safabakhsh (2008) introduce a time-adaptive self-organizing map (TASOM) based active contour models (ACM) for detecting the boundaries of the human eye sclera and tracking its movements in a sequence of images. User's face is extracted first based on a skin-colour model. The iris centre and eye corners are detected using the iris edge information. The algorithm is used to extract the inner boundary of the eye. Finally, by tracking the neighbourhood characteristics the eyes are tracked effectively. The TASOM algorithm is improved for this specific application. Additional improvements include a semi-automatic procedure for calibrating the eye and scene cameras, as well as an automatic procedure for initializing the location of the pupil in the first image frame. The reported accuracy of the system is two degrees of visual angle in both indoor and outdoor environments.

3 Gaze Estimation

As mentioned before, gaze estimation is the procedure that connects the image features to gaze coordinates (gaze position, gaze direction). The number of researchers working on gaze estimation has increased considerably in the last few years. Accordingly, the number of published papers also confirms this fact. The study of the geometry and mathematical basis of gaze tracking systems presents many potential benefits for the technology. Mathematical modelling of the system provides “a priori” knowledge about the behaviour of the system; accuracy, maximum error areas, head movement tolerance, calibration, hardware requirements, etc. The work devoted to gaze estimation based on mathematical modelling can be also classified into two groups: i) methods that estimate the 2-D position of gaze on the area of interest (e.g. screen), i.e. point of regard (PoR) and ii) methods that estimate 3-D direction of gaze, i.e. line of sight (LoS). Many authors have presented their results regarding 2-D or 3-D gaze estimation, summarized below.

The work by Villanueva et al. (2006) developed within the COGAIN network, presents a discussion about the gaze estimation problem from a formal geometric point of view. They do not provide final results but they present a list of objectives that should be covered in the future, such as, i) to provide a mathematical review of different methods that combine image formation and previous system data, ii) to propose new methods to estimate gaze, iii) to identify the minimum hardware requirements and the lower bound on the number of calibration points using purely geometrical criteria (see Appendix I, paper 7: Basics of Gaze Estimation).

Geometry-based methods use alternative 3-D eye models to establish a connection between the 3-D eyeball position and its image under the camera projection. In the last few years, most researchers have converged on similar 3-D eye models; however, slight differences exist in the various alternative models, particularly in terms of eye characteristics and physiology. For example, models differ in their treatment of the relationship between the optical and visual axes (horizontal offset vs. horizontal and vertical offset), corneal refraction modelling, and torsion during eye rotation (Listing’s law). Recently, Böhme et al. (2008) from the University of Lübeck presented work consisting of an open source software framework that simulates the measurement made in a video oculographic system using single or multiple cameras. In a virtual environment, it can calculate alternative image features that can be used to evaluate different gaze estimation methods.

The work by Hennessey et al. (2006) presents a model for a single camera system. Free head motion is achieved by using multiple glints and 3-D modelling techniques. The camera is modelled as a pinhole camera and Gullstrand schematic eye is used for the eyeball. Optical and visual axes of the eye are differentiated and the visual axis is considered to be the LoS. The system calculates first the optical axis of the eye as the line joining the corneal centre and the pupil centre. The cornea centre is calculated by using the information provided by the multiple glints and reflection law. The pupil centre is computed from the pupil image contour, taking into account the corneal refraction effect. The visual axis is computed in 3-D from the optical axis knowing their relative position. Once the model is constructed a four-point calibration is performed to estimate the set of unknown coefficients for gaze estimation. The work claims to have accuracy under 1º over a field of view of 14x12x20 cm. Since the system has no moving parts fast re-acquisition times are provided. This also allows the use of a one-time per user calibration.

The work by Guestrin and Eizenman (2006) provides a good review of alternative system configurations. It presents a general theory for remote estimation of the PoR or point of gaze (PoG). Its objective is to estimate the 3-D direction of the visual axis varying the number of cameras and the number of glints. It uses a similar

3-D eye model as the one proposed by Hennessey et al. (2006) with slight modifications. However, their final model uses the centre of the pupil image to estimate the pupil centre in 3-D. They use the same method to estimate the centre of the cornea based on two glints. They conclude that a system based on one camera and multiple light sources is sufficient to allow free head movements, whereas, a single light source system permits 3-D visual axis estimation in a still head position scenario. Their gaze estimation procedure presents acceptable accuracies and is based on physiological parameters of the eyeball that are due to be inferred by means of multiple point calibration. In a later work (Guestrin and Eizenman, 2007) they reduce the number of calibration marks to one but using a two-camera system.

A similar philosophy is followed in the work by Villanueva and Cabeza (2007) from the Public University of Navarra limiting the number of cameras to one; they review models based on different image features and varying number of LEDs. They also conclude that a single camera and multiple light sources are needed for gaze estimation in a free head movement scenario. However, they base their model on the shape of the pupil instead on the centre of the pupil. This conclusion is similar to the one obtained by Hennessey et al. (2006). Their model is based on physiological parameters of the eyeball that are obtained by means of calibration. The calibration procedure is also studied, and based on the obtained model, the lower bound of calibration points is calculated to be one (Villanueva and Cabeza, 2008).

The model by Ohno (2006) also suggests a one calibration point method, but additional information is used, such as the distance between the camera and the eyeball as determined by an auto-focus camera and a specific location of the two lighting sources. This model is based on a previous work of the author that used physiological information of the eye, such as the corneal radius. In the present work, the geometric eyeball model, the radius of the corneal curvature, and the distance between the camera and the eyeball are used as known parameters; thus, the distance between two Purkinje images on the eye model is derived when looking to the calibration point. If the estimated Purkinje images do not coincide with the observed ones a correction coefficient is applied and the user's gaze direction is calculated with that coefficient. The residual gaze detection error is compensated with the calibration parameters.

The work by Nagamatsu et al. (2008) is also a single calibration mark model, based on a stereo system. The centre of the pupil and multiple light sources are used as working features. The procedure to estimate the optical axis is also based on the estimation of corneal and pupil centres. Listing's Law is introduced as the main contribution of the paper to determine the 3-D orientation of the eye.

The work by Zhu and Ji (2008) also exploits eyeball 3-D anatomy to construct two models for gaze estimation. The first one is a geometry-based model using two cameras and two light sources placed around the centre of the camera lens. This configuration allows for the determination of the cornea centre. In order to estimate the pupil centre in 3-D, images of the pupil in both cameras are used. The pupil centres are employed to estimate the pupil centre position, taking into account corneal refraction effects. Afterwards, similarly to the previous models, the visual axis is estimated from the optical axis calculated as the line joining the corneal and pupil centres. In order to estimate the deviation between both axes, a 3x3 grid of points is used for calibration. The second model is not entirely a geometry-based model. It uses a specific polynomial gaze mapping function based on unknown coefficients to deduce the gazed point from the image features, using the pupil-glint vector (PCCR technique). Different equations are used for the horizontal and vertical coordinates of the gazed point on the screen. Second degree expressions are used, having the vertical coordinate of the pupil-glint vector higher influence in the expressions. The reason for that is that the gains of the vector for screen coordinates vary as a function of how high on the screen the user is looking at (these expressions are extracted from LC technologies, as referenced in the paper). The coefficients of the expressions are obtained by means of calibration. It is well-known that polynomial expressions work satisfactorily after calibration if the user keeps still, however, problems are reported when the user moves from the calibration position. This paper proposes an improvement in the mapping function by including supporting geometrical information. Having the geometrical framework allows for the theoretical measurement of head movement influence on the pupil-glint vector. After this analysis, the authors propose an elaborated dynamic head movement compensation model to improve the accuracy of the polynomial approach. A comparison between both techniques is carried out in the paper.

Regarding polynomial gaze mapping functions an interesting review was recently presented by Cerrolaza et al. (2008). Several systems base their gaze estimation procedures on general-purpose polynomial expressions using unknown coefficients. This work makes a taxonomical evaluation of thousands of mapping functions to find out the most efficient ones. It evaluates the expressions varying the number of terms, degree of the expressions and features used among others. It evaluates alternative features, such as the centre of the pupil, glint, multiple glints, pupil ellipse parameters and several combinations between features. The conclusions show that there is no single “best” expression but a set of conditions to have a good mapping equation.

The previous work present models to estimate the point of regard or point of gaze in 2-D, directly or as an intersection between the 3-D gaze direction and the screen plane. The model by Hennessey and Lawrence (2008) presents a method for interaction in the 3-D space using an eye tracker. This will greatly enhance the ability of humans to interact with 3-D displays and environments. The model presented is an improvement of the model (Hennessey et al., 2006) applied to both eyes. The PoG in 3-D is estimated as the intersection of both LoS. The Technical University of Dresden works in a similar project for virtual 3-D eye tracking. Eye based control in 3-D space allows for a more dynamic and direct interaction with alternative environments, such as, an intelligent home. LC Technologies system is used to perform a binocular tracking. The angle of parallax is calculated by using the gazed points in a plane separately for both eyes. A calibration procedure is performed first (see Appendix I, paper 8: “Virtual 3D Eye Tracking”).

A 3-D point of regard estimation method for a head mounted eye tracker is also introduced by Munn and Pelz (2008). This work is more focused on the robust feature tracking and calibration of the scene camera in order to determine the 3-D location and orientation of the scene camera in the world, rather than in the eye image. In this manner calibrated 2-D PoRs can be triangulated to 3-D positions in the world.

A head mounted eye tracker is also presented in Kohlbecher et al. (2008) using simple goggles. It is based on a stereo system to record eye images. The major novelty is that it eliminates lighting sources from the system. The underlying geometry is based on reconstructing the 3-D pupil circle by means of the ellipses in both images using projective geometry theory. However, the eye model used is simpler compared to the models used in the work mentioned before. The optical axis is used as the line of sight and refraction is obviated. These two assumptions allow for the geometrical determination of the pupil in 3-D even without glints.

The robustness of geometry-based gaze estimation methods has also been studied in recent years. One of the contributions, the work by Chen et al. (2008), is precisely the introduction of an effective constraint to reduce noise in the system. The suggested model is based on a stereo system using two IR lights located near the centres of the stereo cameras. The glints positions and the pupil centres are used to compute the corneal centre and the pupil centre in consecutive steps. As reported in this work this reconstruction method is not accurate in the estimated pupil and corneal centres. An approximate value of 1 mm can be expected for the noise as estimated in the paper what can produce non acceptable errors in the calculated gaze position. This work demonstrates that the noise affects more specially to z coordinate. In order to minimize this noise they impose a constraint in the distance between pupil and corneal centres maintaining the x and y coordinates and solving for the z coordinate. This parameter is estimated for each person during the calibration process. In this manner, they claim to reduce the noise, which indeed is demonstrated in their results. Once the optical axis is estimated, the visual axis is deduced by applying the corresponding angular offsets deduced by means of calibration.

Increasing precision is also the objective of a recent work by Hennessey et al. (2008). Fixation precision is an important issue in gaze tracking systems. The paper gives definitions for PoG fixation precision and measures it. In addition, it proposes methods for increasing fixation accuracy and evaluates the improvement when the method is applied to two PoG estimations. A main aspect of the method is the requirement of high-speed cameras (400 Hz) and image processing methods. This allows for an increment in the precision while maintaining real time performance. The method based on low-pass digital filtering is applied to two different gaze estimation methods, first a method based on a polynomial approach and second, a geometry based method. The results improved the accuracy by a factor of 5.8 (to 0.035º) and a factor of 11 (to 0.050º), respectively.

Noise in gaze estimation is also studied in (Kolakowski and Pelz 2006). The first part of the paper studies the differences in the glint pupil vector due to eye movements with respect to the head, and from camera movements with respect to the head, observed from their wearable tracker. Pupil and corneal reflection movements are observed and mathematically modelled. They claim that the noise in gaze estimation is due to the indetermination in the corneal centre estimation derived from the small size of the corneal reflection in the image. After the first analysis they achieve to differentiate two separate arrays of data, one for camera position and one for the eye position based on the pupil and corneal reflection information in the image. These arrays can be altered separately. The procedure to eliminate the noise introduced by corneal reflection is to smooth the camera position data by means of a median filter followed by a Gaussian filter. Since camera movement is slower than eye movement, the camera array can be smoothed such that the filtered eye array contains a comparable amount of noise to that in the pupil array. The PC-CR technique produces an output that can only be as low in noise as the corneal reflection data. In their proposed algorithm the corneal reflection data are important to determining the camera position but do not contribute to the noise in the final output.

4 Other

Apart from the work described above, other systems, algorithms and methods have been presented recently that cannot be entirely classified as focused on gaze estimation or eye tracking in the image.

The openEyes system (Li et al., 2006; available online at http://thirtysixthspan.com/openEyes/) consists of an open source hardware designed for a digital eye tracker that can be constructed using COTS components. In addition, it uses open source software tools for image acquisition, manipulation and processing. In fact, the origin of the openEyes system is the Human Computer Interaction Program from the Iowa State University, authors of the Starburst algorithm. In their paper they discuss mainly hardware issues and discuss different methods and technical challenges of low cost eye tracking. The openEyes has positively contributed to attract researchers to construct new inexpensive eye tracking systems, such as, the EyeSecret (Yun et al.., 2008) that presents an eye tracking system integrated in a head gear. They use two sets of autocalibration planes to automatically acquire the coordinates of calibration markers in the real scene. They claim to improve the IR source and implement laser to get qualified image and facilitate eye tracking.

In the recent years new systems and methods have been patented as confirmed by different US Patents. As examples we have selected the Gaze Tracking System (Taylor and Rowe, Bracknell, 2007) and the Method and Apparatus for Tracking Gaze Position (Lee et al., 2008). In the first one, in order to identify images where an operator is looking in the same direction, images obtained by a camera are compared (using patches) with stored data. Thus, an initial classification based upon correspondences between areas is made. This initial classification is further processed and each image is assigned a single classification. All patches are processed to determine a best match for each of them for executing gaze conversion. The feature stored is also updated with the new information. The system is based on a single camera and no infrared illumination. The second system uses a more complex hardware including IR illumination system that reflects and captures the IR lighting at 45º. The image processing module obtains a pupil centre of the illuminated eye to find out the pupil centre by a circle detection algorithm based on a shift circle template looking for matches in the image. The pupil centre is mapped on the display plane through a predetermined transform function. The display plane position is calibrated by asking the subject to gaze at display corners using a linear interpolation transform function and the cross ratio transform function.

With the video oculography being the most popular eye tracking technique in the recent years, other radically different systems have also been proposed for eye tracking. The scleral search coil systems were largely used many years ago for gaze tracking purposes, but rapidly obviated for human computer interaction due to its intrusiveness degree. It is accepted as the standard for precise and accurate recording of eye movements in the lab and clinic. One of its drawbacks is the connecting wire that leads from the eye coil to the added electronics. Roberts et al. (2008) recently proposed an improved version of this technique, which uses a resonant scleral coil and no connecting wire. The eye coil communicates with the rest of the system in a "wireless" environment. This new approach, as claimed by the authors, retains the advantages of accuracy, precision and high sample rate, while making the system portable and more comfortable.

The OWL system was developed in the 80's for augmentative and alternative communication. It uses a small sensor close to the user's eye, combining data from multiple channels of IR light on and around user's eye to determine direction of gaze. The IR light is reflected on different parts of the eye and the reflected information is used by the OWL system to extract useful information about the movement of the eye from raw unfocused data. In its first version it used a display with 36 cells originally thought for English text typing. The later development allowed the display to be eliminated. In the work by Grover (2006), a novel implementation of the OWL system is proposed consisting of eight LEDs and eight phototransistors located uniformly around the sensor (see Appendix I, paper 9: Progress on an Eye Tracking System Using Multiple Near-Infrared Emitter/Detector Pairs with Special Application to Efficient Eye-Gaze Communication). A similar device was proposed more recently (Topal et al., 2008) named the EyeTouch system. It is based on a pair of eye glasses incorporating IrDA sensitive sensors and IrDA light sources. The light is reflected in the eye surface and captured by the sensors. This information is used to estimate eye movements. A prototype for binocular tracking is presented in the paper.

Without leaving completely aside the area of eye tracking systems, work more focused on the application field can be found. The work developed at the Loughborough University (Shi and Gale, 2007) shows an adapted eye tracking system to control devices in the user's environment by gazing the device itself instead of a computer interface (see Appendix I, paper 10: Environmental Control by Remote Eye Tracking). The work is partially based on previous work that presented a method for a head mounted device. The current paper interpolates the algorithm to a remote video oculographic system. They use a commercial eye tracker (Smart Eye) that allows for the eye camera calibration, definition of a world coordinate system (WCS), the creation of a personal profile and gaze calibration. The output of this system are the head position and the eye gaze defined with respect to the WCS. They present a pilot trial result using a prototype to control some devices, such as, a lamp from the eye tracker mounted on a wheelchair. Recently, a paper has been presented on the control of a wheelchair by means of gaze (Figueiredo et al.., 2008). The user controls the chair movement by looking to the camera in different directions as working with a joystick (see Appendix I, paper 12: Magic Environment). The control of a wheelchair is also proposed in by Novák et al. (2008) from the Czech Technical University. They use the I4Control wearable system developed in the same university to control a wheelchair. In this case, a screen based interface is presented to the user. The system incorporates additional safety systems in order to avoid potential accidents if the eye tracking system fails as the user is operating the wheelchair (see Appendix I, paper 13: AI Support for a Gaze Controlled Wheelchair). An application for hands free interaction in augmented reality (AR) is also presented in the next work by Nilsson et al. (2007). A head mounted eye tracker is developed but main focus is paid on the application (see Appendix I: paper 11: Hands Free Interaction with Virtual Information in a Real Environment). The head mounted device contains a display for AR purposes that allows for interaction by means of gaze. A low cost and lightweight head mounted equipment is also presented in (Boening et al.., 2006). The most relevant characteristic of this device is the gaze driven camera located in the top part, however it can also be used for eye movement research. The pivot able camera is rotated as the eye moves in real time.

5 Future Development

The work by Böhme et al. (2006) is an interesting summary of the state of the art of the VOG technology. However, from the point of view of this deliverable the future research directions pointed out in the paper are of great interest (see Appendix I, paper 14: Remote Eye Tracking: State of the Art and Directions for Future Development). These can be summarized as:

Regarding the image processing part the tolerance towards glasses is often a problem depending of the system and the glasses. The IR illumination can be reflected in the glasses introducing non desired glints in the image, due which it can be problematic to detect the correct glints and the pupil, especially when these reflections cover part of the pupil. One of the solutions proposed in this paper is to use more than two illuminators. In this manner, if the system detects “problems” when using a specific pair it could switch to another one to try to solve the problem. Regarding the geometry based gaze estimation, the glasses should be introduced into the 3-D eyeball model in order to account for their effect in the image. However, as claimed in the paper, preliminary tests show that the accuracy is still acceptable when current eyeball models without explicit modelling of glasses are used. Contact lenses can also produce undesirable artefacts in the image, such as glint and pupil deformation.

Geometry based gaze estimation systems are normally based on a physical model of the eye and hardware configuration. Generally, the screen and IR LEDs positions have to be known and the camera calibrated. These measurements can be obtained by hand or using laborious hardware calibration algorithms, employing additional components such as a mirror. Making the gaze estimation models independent of hardware information or easy to calibrate (easy to use) while keeping head movement tolerance and accuracy would provide more flexibility to the technology.

As mentioned before, the high price of the majority of the systems is many times an obstacle for further applications of eye trackers. High-resolution industrial cameras with relatively high-grade lenses increase the price of the eye tracker considerably. The necessity of implementing eye tracking with lower price cameras such as webcams is clear and the number of recently published papers on the topic shows this as a promising research line.

Another straightforward step to reduce the cost and hardware complexity of an eye tracking system would be to eliminate the IR illumination; however, this complicates considerably the mathematics for gaze estimation modelling.

Moreover, using 3-D cameras in order to locate the user in the space with respect to the camera is an interesting research line. The developments of so-called 3-D time-of-flight (TOF) cameras have been applied to other computer vision problems in the last years and COGAIN tries to apply them for eye tracking purposes. The main application of this kind of cameras for gaze estimation is the possibility to determine head position and orientation robustly.

Apart from the aforementioned issues, reducing the number of calibration points while maintaining accuracy is also an interesting research line. Related to that, finding more efficient calibration patterns, one-calibration-per-user as well as auto-recalibration systems are of high interest, especially for those users for which calibration is a difficult process.

Increasing head movement tolerance is also necessary not just from the optics point of view (wide FoV systems), but also in terms of accuracy, i.e. the accuracy should be the same as the user moves. This is closely related to the gaze estimation method employed.

Finally, as mentioned before, eye tracking outdoors has to be efficiently solved. Developing new image processing algorithms and using novel hardware and sensors, more tolerant to light variations, can contribute to that.

The current advances in eye tracking and the increasing number of researchers working in the field will make widespread low-cost eye tracking to become a reality in the future.

6 References

Boening, G., Bartl, K., Dera, T., Bardins, S., Schneider, E. and Brandt, T. (2006). Mobile eye tracking as a basis for real-time control of a gaze driven head-mounted video camera. *Proceedings of the Eye Tracking Research and Applications Conference (ETRA)*, pp. 56.

Böhme, M., Meyer, A., Martinetz, T., and Barth, E., (2006). Remote eye tracking: state of the art and directions for future development. *Proceedings of the 2nd COGAIN Conference*, pp. 12-17.

Böhme, M., Dorr, M., Graw, M., Martinetz, T., and Barth, E., (2008). A software framework for simulating eye trackers. *Proceedings of the Eye Tracking Research and Applications Conference (ETRA)*, pp. 251-258

Cerrolaza, J.J., Villanueva, A. and Cabeza, R. (2008). Taxonomic study of polynomial regressions applied to the calibration of video-oculographic systems. *Proceedings of the Eye Tracking Research and Applications Conference (ETRA)*, pp.259-266.

Chen, J., Tong, Y., Gray, W. and Ji, Q. (2008). A robust 3-D eye gaze tracking system using noise reduction. *Proceedings of the Eye Tracking Research and Applications Conference (ETRA)*, pp. 189-196.

Dervinis, D. and Daunys, G. (2007). 3-D head orientation estimation and expression influence elimination using characteristic points of face. *Proceedings of the 3rd COGAIN Conference*, pp. 45-48.

Droege, D., Fritzer, F., and Paulus, D. (2006). GoldenGaze: An inexpensive real-time gaze tracking system. *Proceedings of the 2nd COGAIN Conference*, pp. 29-30.

Droege, D., Geier, T., and Paulus, D. (2007). Improved low cost gaze tracker. *Proceedings of the 3rd COGAIN Conference*, pp. 37-40.

Droege, D., Schmidt, C., and Paulus, D. (2008). A Comparison of Pupil Centre Estimation Algorithms. *Proceedings of the 4th COGAIN Conference*, pp. 23-26.

Figueiredo, L. and Gomes, A.I. (2007). Magic Eye Control. *Proceedings of the 3rd COGAIN Conference*, pp. 41-44.

Figueiredo, L., Nunes, T., Caetano, F. and Gomes, A. (2008). Magic Environment *Proceedings of the 4th COGAIN Conference*, pp. 15-18.

Grover, D. (2006). Progress on an eye tracking system using multiple near-infrared emitter/detector pairs with special application to efficient eye-gaze communication. *Proceedings of the 2nd COGAIN Conference*, pp. 21-25.

Guestrin, E.D., and Eizenman, M. (2006). General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Transactions on Biomedical Engineering* 53-6 pp. 1124-1133.

Guestrin, E.D., and Eizenman, M. (2007). Remote point-of-gaze estimation with free head movements requiring a single point calibration. *Proceedings of the Annual International Conference of the IEEE EMBS*, pp. 4456-4460.

Hansen, D.W. and Hansen, J.P. (2006). Eye typing with common cameras. *Proceedings of the Eye Tracking Research and Applications Conference (ETRA)*, pp. 55.

Hansen, D.W. and Hammoud, R.I. (2007). An improved likelihood model for eye tracking. *Computer Vision and Image Understanding*. Elsevier.

Hennessey, C., Noureddin, B. and Lawrence, P. (2006). A single camera eye-gaze tracking system with free head motion. *Proceedings of the Eye Tracking Research and Applications Conference (ETRA)*, pp. 87-94.

Hennessey, C., and Lawrence, P. (2008). 3-D point-of-gaze estimation in a volumetric display. *Proceedings of the Eye Tracking Research and Applications Conference (ETRA)*, pp. 59.

Hennessey, C., Noureddin, B., and Lawrence, P. (2008). Fixation precision in high-speed noncontact eye-gaze tracking. *IEEE Transactions on Systems, Man and Cybernetics, part B. 38-2*, pp.289-298.

Huachun, T. and Zhang, Y. (2005). Detecting eye blink states by tracking iris and eyelids. *Journal of Pattern Recognition*. Elsevier.

Kohlbecher, S. Bardins, S., Bartl, K., Schneider, E., Poitschke, T. and Ablassmeier, M. (2008). Calibration free eye tracking by reconstruction of the pupil ellipse in 3-D space. *Proceedings of the Eye Tracking Research and Applications Conference (ETRA)*, pp. 135-138.

Khosravi, M.H., and Safabakhsh, R. (2008). Human eye sclera detection and tracking using a modified time-adaptive self-organizing map. *Pattern Recognition* (41), pp 2571-2593 Elsevier.

Kolakowski, S.M. and Pelz, J.B. (2006). Compensating for eye tracker camera movement. *Proceedings of the Eye Tracking Research and Applications Conference (ETRA)*, pp. 79-85.

Lee, J, Jung, G., Han, S., Park, S., Lee, C., Park, R., Hwang, C., Lim, J.S. and Cho, Y. (2008). Method and Apparatus for tracking gaze position. United States Patent Application Publication US 2008/0137909.

Li, D., Winfield, D. and Parkhurst, D.J. (2005). Starburst:A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops - Volume 03.*

Li, D. and Parkhurst, D.J. (2006). Open-source software for real-time visible-spectrum eye tracking. *Proceedings of the 2nd COGAIN Conference*, pp. 18-20.

Li, D., Babcock, J. and Parkhurst, D.J. (2006). openEyes: a low cost head mounted eye tracking solution. *Proceedings of the Eye Tracking Research and Applications Conference (ETRA)*, pp. 95-100.

Munn, S.M. and Pelz, J.B. (2008). 3-D point of regard, position and head orientation from a portable monocular video-based eye tracker. *Proceedings of the Eye Tracking Research and Applications Conference (ETRA)*, pp. 181-188.

Nagamatsu, T., Kamahara, J., Iko, T. And Tanaka, N. (2008). One-point calibration gaze tracking based on eyeball kinematics using stereo cameras. *Proceedings of the Eye Tracking Research and Applications Conference (ETRA)*, pp. 95-98.

Nilsson, S., Gustafsson, T. and Carleberg, P. (2007). Hands free interaction with virtual information in a real environment. *Proceedings of the 3rd COGAIN Conference*, pp. 53-57.

Novak, P., Krajnik, T., Preucil, L., Fetjova, M., Stepankova, O. (2008). AI Support for a Gaze Controlled Wheelchair. *Proceedings of the 4th COGAIN Conference*, pp. 19-22.

Ohno, T. (2006). One-point calibration gaze tracking method. *Proceedings of the Eye Tracking Research and Applications Conference (ETRA)*, pp. 34.

Roberts, D., Shelhamermer, M. and Wong, A. (2008). A new "wireless" search coil system. *Proceedings of the Eye Tracking Research and Applications Conference (ETRA)*, pp. 197-204.

Ryan, W.J., Duchowski, A., and Birchfield, S.T. (2008). Limbus/pupil tracking switching for wearable eye tracking under variable lighting conditions. *Proceedings of the Eye Tracking Research and Applications Conference (ETRA)*, pp. 61-64.

San Agustin, J., and Hansen, J.P. (2008). Off-the-Shelf Mobile Gaze Interaction. *Proceedings of the 4th COGAIN Conference*, pp. 6-10.

Shi, F. and Gale, A. (2007). Environmental control by remote eye tracking. *Proceedings of the 3rd COGAIN Conference*, pp. 49-52.

Taylor, R.I. and Rowe, S.M. (2007). Gaze Tracking System. United States Patent. US 7,158,097 B2.

Topal, C., Gerek, Ö.N., and Dogan, A. (2008). A head-mounted sensor-based eye tracking device: eye touch system. *Proceedings of the Eye Tracking Research and Applications Conference (ETRA)*, pp.87-90.

Villanueva, A., Hansen, D.W., San Agustin, J. and Cabeza, R. (2006). Basics of Gaze Estimation. *Proceedings of the 2nd COGAIN Conference*, pp. 26-28.

Villanueva, A. and Rcabeza, R. (2008). A novel gaze estimation system with one calibration point. *IEEE Transactions on Systems, Man and Cybernetics, part B. 38-4*, pp.1123-1138

Villanueva, A. and Rcabeza, R. (2007). Models for gaze tracking systems. *Journal on Image and Video Processing (4)*. Hindawi Publishing Corp.

Yun, Z., Xin-Bo, Z., Rong-Chun, Z., Yuan, Z and Xiao-Chun, Z. (2008). EyeSecret: An inexpensive but high performance auto-calibration eye tracker. *Proceedings of the Eye Tracking Research and Applications Conference (ETRA)*, pp. 103-106.

Zhu, Z. and Ji, Q.(2007). Novel eye gaze tracking techniques under natural head movement. *IEEE Transactions on Biomedical Engineering* (in press).

Appendix I

1. **Magic Eye Control**. Luis Figueiredo, Ana Isabel Gomes
2. I**mproved Low Cost Gaze Tracker**. Detlev Droege, Thorsten Geier, Dietrich Paulus
3. **A Comparison of Pupil Centre Estimation Algorithms.** Detlev Droege Carola Schmidt Dietrich Paulus
4. **Off-the-Shelf Mobile Gaze Interaction**. Javier San Agustin, John Paulin Hansen
5. **Open-Source Software for Real-Time Visible-Spectrum Eye Tracking**. Dongheng Li, Derrick Parkhurst
6. **3D head orientation estimation and expression influence elimination using characteristic points of face.** Donatas Dervinis, Gintautas Daunys
7. **Basics of Gaze Estimation.** Arantxa Villanueva Dan Witzner Hansen Javier San Agustin Rafael Cabeza
8. **Virtual 3D Eye Tracking.** Sascha Weber
9. **Progress on an Eye Tracking System Using Multiple Near-Infrared Emitter/Detector Pairs with Special Application to Efficient Eye-Gaze Communication**. Dale Grover
10. **Environmental Control by Remote Eye Tracking**. Fangmin Shi, Alastair Gale
11. **Hands Free Interaction with Virtual Information in a Real Environment**. Susanna Nilsson Torbjörn Gustafsson and Per Carleberg
12. **Magic Environment**. Luis Figueiredo Tiago Nunes Filipe Caetano Ana Gomes
13. **AI Support for a Gaze Controlled Wheelchair**. Petr Novák, Tomáš Krajník, Libor Preucil Marcela Fejtová, Olga Štepánková
14. **Remote Eye Tracking: State of the Art and Directions for Future Development**. Martin Böhme, André Meyer, Thomas Martinetz, and Erhardt Barth