COGAIN 2005

Proceedings of the First Conference on Communication by Gaze Interaction

May 31, 2005
Copenhagen, Denmark

Foreword

Welcome to the first COGAIN Conference.

Gaze-based communication and interaction is set to expand into new domains and new user communities thanks to the emergence of cheaper measurement systems and an increased understanding of how eye-gaze can be used effectively. As a network in its first year, COGAIN has already shown itself to be a highly effective forum for research exchanges within Europe and beyond. The annual conference is set to become a major international event in the calendar of the research, user and industrial communities associated with gaze-based communication. This first conference sets out the direction and tenor with which this will be achieved.

Howell Istance
Conference Chair
De Montfort University, United Kingdom

Talks

Date: Tuesday 31 May 2005

Place: Auditorium 2, IT University of Copenhagen, Denmark

Conference Chair: Howell Istance (De Montfort University, United Kingdom)

10:00-10:10 Opening, introduction and welcome

Kari-Jouko Räihä (University of Tampere)
Howell Istance (De Montfort University)

10:10-11:30 Session 1: Towards cheaper and less intrusive gaze measurement: technical challenges

5 Challenges in Single-Camera Remote Eye Tracking
Martin Böhme and Erhardt Barth (University of Lübeck)

7 Review of Current Camera-based Eye Trackers
Dan Witzner Hansen and John Paulin Hansen (IT University of Copenhagen)

10 Gaze Tracking with Inexpensive Cameras
Fabian Fritzer, Detlev Droege and Dietrich Paulus (University Koblenz-Landau)

12 Influence of Head Position Instability to Gaze Tracking in Remote Video-oculography
Gintautas Daunys and Nerijus Ramanauskas (Siauliai University)

11:50-13:10 Session 2: Innovations in eye-based interaction

16 EyeChess: A Tutorial for Endgames with Gaze-Controlled Pieces
Oleg Špakov and Darius Miniotas (University of Tampere)

19 Towards emotion modeling based on gaze dynamics in generic interfaces
Martin Vester-Christensen, Denis Leimberg, Bjarne Kjær Ersbøll and Lars Kai Hansen (Technical University of Denmark)

22 Learning to Type Japanese Text by Gaze Interaction in Six Hours
Hirotaka Aoki and Kenji Itoh (Tokyo Institute of Technology)
John Paulin Hansen (IT University of Copenhagen)

29 Dasher's new Gaze-tracker Mode

David MacKay and Chris Ball (University of Cambridge)

14:15-15:35 Session 3: Interaction and Inferences beyond the desktop

30 Fly Where You Look: Enhancing Gaze Based Interaction in 3D Environments

Richard Bates and Howell Istance (De Montfort University)
Mick Donegan and Lisa Oosthuizen (ACE Centre)

33 An Eye Movement Study of the Think Aloud Technique's Implications for Cognitive Processes

Kristin Due Hansen (Risø Laboratories)

34 Towards Communication of Unusual things: Attention, Consciousness and, perhaps, Feeling

Boris M. Velichkovsky, Sebastian Pannasch, Markus Joos, Jens R. Helmert and Sven-Thomas Graupner (Technische Universität, Dresden)

36 Head and Eye Tracking Inside Intelligent Houses

Fulvio Corno and Alessandro Garbo (Politecnico di Torino)

16:00-17:00 Session 4: Plenary Discussion

The COGAIN Conference: Where do we go from here?

17:00 End

Challenges in Single-Camera Remote Eye Tracking

Martin Böhme and Erhardt Barth (University of Lübeck)

Introduction

Many eye tracking systems either require the user to keep their head still or involve cameras or other equipment mounted on the user's head. While acceptable for research applications, these limitations make the systems unsatisfactory for most AAC (Augmentative and Alternative Communication) applications.

So-called “remote” eye tracking systems, which allow the user to move their head freely within certain limits, have been available for a while. Traditionally, these systems have used cameras with long focal lengths to obtain a sufficiently high-resolution image of the eye. Because of the narrow field of view, the user's head movements must be compensated for, either by panning and tilting the camera itself or by using a movable mirror. This means that the head movement speed is limited by the speed with which the mechanical system can track the eye. Furthermore, once tracking is lost, reacquiring the eye is difficult because the camera has only a narrow field of view.

With recent increases in the resolution of CCD and CMOS cameras, it has become feasible to use fixed cameras with wide field of view for eye tracking. In this approach, the camera covers the whole area within which user's head may move while still imaging the eye with sufficient resolution for eye tracking.

One important task in remote eye tracking is measuring the position of the user's eyes. The most straightforward way of doing this is to use two or more cameras so that features in the images can be triangulated to determine their position in space. However, using two cameras instead of one increases the cost and complexity of the system substantially. Cost is of particular concern for AAC applications; in the following, we will therefore investigate the single-camera remote eye tracking problem.

While a number of researchers have proposed algorithms for calibrating single-camera remote eye trackers [1, 2], the results appear to be not as accurate as those achieved using fixed or head-mounted devices (0.5 to 1 degree of accuracy). Commercial remote eye trackers with high accuracy are available [3], but no implementation details have been published.

In this talk, we will report on our on-going work on calibration algorithms for remote eye trackers that aim to achieve accuracy similar to that of fixed or head-mounted systems. Our results on simulated test data from an artificial eye model are quite promising, and we hope to achieve similar accuracy when we implement the algorithm on hardware in the near future.

Method

Most videographic eye trackers work by illuminating the eye with an infrared (IR) light source. This light source produces a glint on the cornea (the “corneal reflection” or “CR”), and the gaze angle is computed from the offset between the CR and the centre of the pupil using bilinear or biquadratic interpolation. The coefficients of the interpolation function are computed from data obtained during a calibration phase, during which the user is asked to fixate a certain number of points with known locations.

Our approach to remote eye tracking also uses infrared illumination, but instead of one light source, we use two. The distance between the CRs produced by these light sources can then be used to determine the distance of the eye from the eye tracker. This, together with the location of the eye in the camera image, allows us to deduce the three-dimensional position of the eye relative to the camera.

Note that we only determine the position and orientation of the eye; the position and orientation of the head are irrelevant for us since our approach does not use any reference points on the head.

Using an interpolation scheme to calculate gaze position from the observed pupil and CR positions, as for the fixed-head eye tracker, does not appear to be an option for the remote eye tracking scenario because it has a far greater number of degrees of freedom – covering the whole space of possible eye positions and eye orientations during calibration would not be feasible.

We therefore believe that a calibration procedure for remote eye tracking must be based on a model of all relevant physical properties of the human eye. Of course, the shape and size of the eye vary from person to person, so the model must contain a suitable set of parameters to accommodate these differences. Calibration then means estimating the values of these parameters for a specific person.

To date, our eye model contains the following parameters:

- r_{cornea}: The radius of curvature of the corneal surface (which we assume to be spherical)
- r_{pc}: The distance between the centre of corneal curvature and the pupil centre
- α_{fovea}: The angular offset between the optical axis of the eye and the direction of gaze, which caused by the fact that the fovea does not lie on the optical axis but is offset temporally and slightly upwards (at the moment, we only model the horizontal component of this offset).

The values of these parameters for a particular user are determined by taking the pupil and CR positions for a set of calibration points and then varying the parameter values to minimize the error between the observations predicted by the model and the actual observations.

Results and Outlook

We implemented our calibration algorithm in Matlab and assessed its performance on simulated test data. For the tests, the user was assumed to be seated at a distance of 50 cm from a 40x30 cm screen. To evaluate the robustness of our approach to noise, we added a certain amount of random error to the measurements of pupil centre and CR position.

The results we have obtained so far are encouraging: Assuming a maximum measurement error of 0.5 pixels, the maximum error in gaze position is 13 mm (1.5 degrees), with an average error of 4.5 mm (0.5 degrees). An error of this magnitude should be more than acceptable for most AAC applications. The assumed measurement error should be achievable if some care is taken in the image processing and camera calibration steps.

The next step, then, is to implement our algorithm on actual hardware. This will reveal whether the algorithm can live up to the potential it has demonstrated on our simulated data.

References

[1] Morimoto, C. H., Amir, A. and Flickner, M., 2002. Detecting Eye Position and Gaze from a Single Camera and 2 Light Sources. In: *16th International Conference on Pattern Recognition*, pp. 314-317.

[2] Ohno, T. and Mukawa, N., 2004. A Free-head, Simple Calibration, Gaze Tracking System That Enables Gaze-Based Interaction. In: *Eye Tracking Research and Applications 2004*, pp. 115-122.

[3] Tobii 1750 eye tracker, Tobii Technology AB, Stockholm, Sweden.

Review of Current Camera-based Eye Trackers

Dan Witzner Hansen and John Paulin Hansen (IT University of Copenhagen)

Introduction

During the last decade, tremendous effort has been made on developing robust and cheap eye tracking systems for various human computer interaction applications such as eye typing [4, 3, 6]. Robust non-intrusive eye detection and tracking is crucial for human computer interaction with attentive user interfaces, for understanding human affective states and is gaining importance outside laboratory experiments and may even be found in domestic appliances and vehicles.

Several camera-based eye trackers have been proposed in research and as commercial systems. These systems rely on different assumptions and hardware use, such as known geometry of the cameras, light sources and user, employing several light sources and cameras. This paper reviews the underlying principles of current state-of-the-art eye trackers. Approaches to eye tracking and their trade-offs with respect to accuracy and usage-potential for the general public is discussed. The aim of this paper is to focus on recent advances in eye tracking research and development. In particular, our effort is greatly placed on reviewing and analyzing different algorithms and frameworks to build the fundamental vision modules in an eye tracking system, such as automatic eye detection, eye tracking and gaze determination.

IR or not to IR

Good light conditions generally leads to greater success and less effort on algorithm research and development. Vision-based eye tracking methods greatly benefit from infrared light in practically all stages, starting from the detection, to tracking and gaze estimation. Because IR is not visible, the light does not distract the user if shone upon. The amount of emitted light can therefore be quite high as to control light conditions. Several systems have been developed under the principles of IR light and they achieve precisions on pointer control below one degree. Relying solely on the use of IR light is also limited, since IR is restricted in space and not all users produce the bright pupil (similar to red-eye) effects frequently used in current eye trackers. That is, IR has a dynamic range for which it operates and if the user moves outside this range, the eye tracker should rely on other pieces of information. The use of IR light is also challenged in outdoor scenarios and it may therefore not be sufficient for eye trackers to solely rely on the properties of the IR light if eye tracking should be used for the general public. The use of commercial off-the-shelf (COTS) products as elements in larger systems is becoming increasingly commonplace. Using COTS for camera-based eye tracking tasks has many advantages (such as price, availability and applicability). However, the use of standard components also implies giving up knowledge of exact hardware use and accurate position of the camera, as the general public do not have prior experience in calibrating and positioning cameras themselves accurately. Relying on a single camera and visible light provides less information than using IR light in the same setting. Thus several difficult problems needs to be solved: the image quality may be low due to poor light conditions and unknown light sources. Information about head pose may be difficult to obtain and thus inference on gaze direction may be difficult.

Using Hardware

While the use of several cameras and light sources may overcome some of these limitations, they will also add to the price. Stereo setups require calibration, which is something non-trivial for the general public. Single camera systems may on the other hand reduce costs. Eye trackers using single cameras seem not quite as accurate and robust as systems using several cameras, unless additional knowledge of geometry between the user, screen, light sources or camera is provided e.g. light sources or camera built into the monitor. In addition, the use of single camera system without pan-and-tilt faces the trade-off between large head movements and high accuracy: To get high accuracy of the eye region the camera needs to be focused on a small region around the eye, which in turn limits head movements. This problem is currently declining as the resolution of cameras is continuously improving without significantly increasing costs.

To keep costs down, an option is to use commercial-of-the-shelf components (COTS). In fact, there already exist standard video cameras that contain IR light emitters. The difference to current systems is that exact knowledge of the light emitters is unavailable and the frame rate may be too low (25-30 fps.) for particular applications. Using COTS might also imply leaving some of the popular strategies such pan-and-tilt mounts, the possibility of controlling the hardware (e.g. changing the zoom) and the choice of hardware.

Gaze Estimation

For the convenience of the users, the number of calibration points should be kept at a minimum. Methods relying on geometric information seem to require only a few calibration points. Geometric information may come at the expense of the user as knowledge of for example camera calibration or distances are needed. Obtaining this information is tedious or not a common procedure for the users, and may change between user sessions. Inference-based methods are a try to infer the underlying function from the data in the images to world coordinates. Inference-based methods require less knowledge of the geometry and thus get closer to what is required by the general public. Unfortunately, having to infer the unknown geometry seems to require a large amount of calibration points, which may have to be provided at each user session. The trade-off is therefore (not surprisingly) between the amount of prior information and the number of calibration points.

Due to the dynamics of the eye, we may probably not obtain significantly better results and therefore not obtain mouse accuracy. But do all applications really need mouse accuracy?

Appearance Changes

The problem of visual eye recognition poses a number of challenges to existing machine vision and pattern recognition algorithms. For example, it is not yet obvious how to build invariance into a commercial eye tracking technology; a system that performs equally in indoor and outdoor conditions, for any test subject, which is free of subject and sensor calibration. Eye tracking research has come far in the right direction, but a couple of issues still remain to be solved. One of the major problems for eye tracking is that the eyes appear differently under changing illumination, poses and ethnicity. The head is capable of making movements along six degrees of freedom. The eye can be subject to transitions from wide-open eye to completely closed. The appearance of the iris and pupil is therefore heavily influenced by occlusions from the eyelids and may often be totally covered. The effects of occlusion and illumination changes are also related to the ethnic origin of the user. Furthermore, the eye shape differs from one subject to another (Asian, European, African, etc). While the use of IR light generally provides sufficient information to handle most of these problems, there can also be great variance in the IR responses between subjects [5].

Several authors have suggested good solutions to the problem and tested them in quite challenging indoor conditions. However they have so far not been tested in outdoor scenes for longer periods of time. One of the problems with many current methods is the explicit assumptions on relatively stable light conditions, users close to the camera (or apparent eye size in the image), small out-of-plane face rotations, and open and un-occluded eyes. All important restrictions on both the system and the user. One reason for this lies in the frequent use of thresholds: the possible large difference in the dark and bright pupil images makes it tempting to use threshold values, connected components and other fixed conditions in the difference image. However, as the bright pupil effects may be reduced or eliminated due to light and head changes, appropriate threshold values may be difficult or impossible to set. Additionally, the use of threshold values may throw away useful information. As the bright pupil may disappear suggests that relying solely on the observations from the difference image is not necessarily sufficient. In this case information from the original bright and dark pupil images should be used. It is therefore important for future eye trackers to minimize the use of parameter settings and thresholds.

In conclusion, there are several important research challenges to be addressed before an eye tracker that will work for all under all conditions is available. This paper has pointed at some of them and argued for the use of COTS technology and self-assembly of systems as a feasible way to a large-scale use of gaze tracking.

References

[1] Amir, A., Zimet, L., Sangiovanni-Vincentelli, A., and Kao, S., 2005. An embedded system for an eye-detection sensor. *Computer Vision and Image Understanding*, 98 (1), April 2005, pp. 104-123.

[2] Andrew T. Duchowski. *Eye Tracking Methodology. Theory and Practice*. Springer, 2003.

[3] John Paulin Hansen, Anders Sewerin Johansen, Dan Witzner Hansen, Kenji Itoh, and Satoru Mashino, 2003. Language technology in a predictive, restricted on-screen keyboard with ambiguous layout for severely disabled people. In *EACL 2003 Workshop on Language Modeling for Text Entry Methods*.

[4] Päivi Majaranta and Kari-Jouko Räihä, 2002. Twenty years of eye typing: Systems and design issues. In *Symposium on ETRA 2002: Eye Tracking Research Applications Symposium*, New Orleans, Louisiana, pp. 944-950.

[5] Karlene Nguyen, Cindy Wagner, David Koons, and Myron Flickner, 2002. Differences in the infrared bright pupil response of human eyes. In *ETRA '02: Proceedings of the symposium on Eye tracking research & applications*, New York, NY, USA, ACM Press, pp. 133-138.

[6] David J. Ward, Alan F. Blackwell, and David J. C. MacKay, 2002. Dasher: A gesture driven data entry interface for mobile computing. *Human-Computer Interaction*, 17, pp. 199-228.

Gaze Tracking with Inexpensive Cameras

Fabian Fritzer, Detlev Droege and Dietrich Paulus (University Koblenz-Landau)

We present a simple gaze tracking system based on an inexpensive, yet highly sensitive camera, equipped with a cheap IR-filter and some near infrared LEDs. Using the corneal-reflection-method and common image processing algorithms we easily achieve the accuracy to control reduced keyboard gaze typing systems like UKO-II.

The system is able to track the gaze with over 20Hz and yield a mean error of 3cm of the estimated gaze point on a 30x20cm screen - equivalent to approximately 3 degrees error in viewing direction. Head movements are allowed up to 20cm horizontally and 10cm vertically. Head movement in screen direction leads to inaccuracies, but the system still remains operational. The system is not yet suited for people wearing glasses.

The hardware costs are less than 150 Euro. Two cameras were tested for this system. The current system uses a B/W CCD camera 2005XA with EXview HAD CCD SONY Chip and 0.003 Lux sensitivity (http://www.rfconcepts.co.uk/low-light_mini-cam.htm). The camera has a horizontal viewing angle of 34 degrees. The other camera tested (Philips Webcam PCVC740K) turned out not to be sensitive enough for low illumination, although being the most sensitive webcam on the market.

As infrared filter, a non-exposed, developed diapositive color film is used. The result is a high quality yet cheap infrared filter, which can be easily attached to the lens. The illumination uses eight 1.7 Volt near-infrared LEDs with integrated reflectors.

The tracking method is the corneal-reflection-method. The geometric approach is based on on-axis-illumination. We currently use a single-eye-tracking of the right eye, which is the main reason for the allowed head movement limitations.

The algorithm works as follows:

1. First, we binarize the image with a dynamic threshold. Then we select a region of interest using the binary image. As the illumination fades in the background, we easily separate person from background. In rare cases this has to be corrected by the algorithm, e.g. if the background is brightly illuminated by the sun.
2. Then we detect highlights (glints) which turn out to be the lightest points surrounded by dark pixels. In 100% of the cases when the eyes are open, this gives us the correct result. For people wearing glasses this will not be the case, as highlights on the glasses and on the frame will also be visible. For closed eyes, this step will give a false alarm, which has to be eliminated by the next step.
3. We now verify the position of the eye. We again do a binarization of the input image in a 60x60 region around the estimated glint using a dynamic threshold. The result is a binary image containing the iris and the pupil as one black filled circle. There is no difference in luminance between pupil and iris, although this is in contradiction to many publications. This may be due to the on-axis illumination. We could call this a 'half bright pupil effect'.
4. We now do a template matching to estimate the location of the eye. We reject, if the correlation of image and template is too low. If no eye is found, we assume that the eye was closed. We accept if the correlation is high enough and we use the center of the template, which corresponds to the maximum of correlation as the estimate for pupil center.
5. We now do a sub-pixel fitting again with the template to fit the iris and a heuristics for the glint.

6. Using a geometric model of the setup and using the calibration data we estimate the viewing direction, resp. the gaze point. As we use a monocular setup, we do not attempt to recover 3D or to estimate the distance of the person to the screen. As the camera has 768x576 resolution, an approach using the different reflections resulting from different illuminating diods is not feasible, as all reflections collapse to few pixels. We assume that such techniques are used in commercial systems.

The system will be tested using a handicapped person who can turn her head but she will not move it back and forth due to the headrest. This means that the restrictions above will be easily fulfilled. The gaze is subject to further computer processing and it will be used as an input device for a typewriter.

It might be interesting to use our system also for mouse interaction.

Influence of Head Position Instability to Gaze Tracking in Remote Video-oculography

Gintautas Daunys and Nerijus Ramanauskas (Siauliai University)

This study addresses the head movement influence problem to an accuracy of eye tracking. Remote video-oculography is the most suitable eye tracking method in Human Computer Interaction. Pupil centre tracking methods are sensitive to the head shifts. During this study the characteristic head movement were investigated. The corneal reflection tracking and eye corner tracking by normalized correlation coefficient were used for evaluation of the head shifts. The noise of evaluated eye gaze during fixations is less than for initial pupil centre using eye corner tracking for head shifts elimination. Head oscillations with 6.3 Hz frequencies were observed. Their amplitude is bigger in vertical direction.

Introduction

Today, video-oculography (VOG) is the most suitable eye tracking method in HCI. The physical layout of eye trackers differs in intrusive. Fixed (head is stabilized using chin rest or bite bar) and head mounted (eye tracker fixed to head) systems are more intrusive than remote systems (eye tracker typically fixed relative to display). In remote systems computer user works in the most natural conditions. However, such systems require more complex algorithms, because there is a need of head movement compensation. The characteristics of head movements must be known for better their elimination.

Method

Video frames of eye were recorded with Basler 602f video camera. Coordinates averaging method (Daunys and Ramanauskas 2004) was used for pupil centre coordinates estimation. There is a short description. It begins from pupil edge detection. The complete pupil edge is defined after two steps: (1) horizontal scanning and (2) vertical scanning. Pupil edge coordinates are extracted with subpixel accuracy.

At the second stage the average of edge points coordinates in each scanning line was calculated. A set of new points can be approximated by a vertical line. The result of approximation is equation coefficients v_0 and v_1 of vertical line:

$$y = v_0 + v_1 x; \qquad (1)$$

which are obtained by points fitting to line exploiting least squares method.

Second step is analogous to the first, but now scanning in vertical direction is processed. Fitting to line gives us an equation of horizontal line:

$$y = h_0 + h_1 x; \qquad (2)$$

To find the centre of the pupil we solve an equation system consisting of equations (1) and (2):

$$\begin{cases} y = v_0 + v_1 x, \\ y = h_0 + h_1 x; \end{cases} \tag{3}$$

Two methods were used to detect and/or eliminate head movement. They were: 1) well known corneal reflection tracking (Mulligan 1997); 2) eye corner tracking (Tian et al 2000). Both objects are shown in Figure 1. The difference between pupil centre and corneal reflection coordinates allows directly calculate line of sight direction.

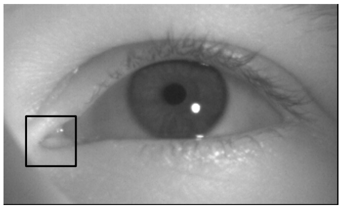

Figure 1. Video frame with 1st Purkinje image and marked eye corner.

To track the inner corner of eye, the normalized correlation coefficient was chosen (Betke 2002). The algorithm decides which 20 by 20 pixels subimage is closest to the previous selected square. It examines 400 pixels size (20 by 20) trial square subimages around the location of the previous selected square. The algorithm calculates the normalized correlation coefficient *r(s,t)* for the selected subimage *s* from the previous frame with each trial subimage *t* in the current frame:

$$r(s,t) = \left(n \cdot \sum_{i=1}^{n} s(x_i, y_i) \cdot t(x_i, y_i) - - \sum_{i=1}^{n} s(x_i, y_i) \cdot \sum_{i=1}^{n} t(x_i, y_i) \right) \cdot \frac{1}{\sigma_s \cdot \sigma_t} \tag{4}$$

where n is the number of pixels in the subimage, and

$$\begin{aligned} \sigma_s &= \sqrt{n \cdot \sum_{i=1}^{n} s(x_i, y_i)^2 - (\sum_{i=1}^{n} s(x_i, y_i))^2} \\ \sigma_t &= \sqrt{n \cdot \sum_{i=1}^{n} t(x_i, y_i)^2 - (\sum_{i=1}^{n} t(x_i, y_i))^2} \end{aligned} \tag{5}$$

From the obtained correlation coefficients the maximum location with subpixel resolution is evaluated using mass centre of contour method.

Results

Usually head movement cause a baseline variation of eye movement. Results on Figure 2 illustrate this. The subject during recording periodically fixed central target. There are plotted vertical coordinates of pupil centre, eye corner and difference between them, which proportional to gaze direction. The standard deviations are: 1.06 px (pixel) for pupil centre, 0.84 px for eye corner, 0.28 px – for gaze direction.

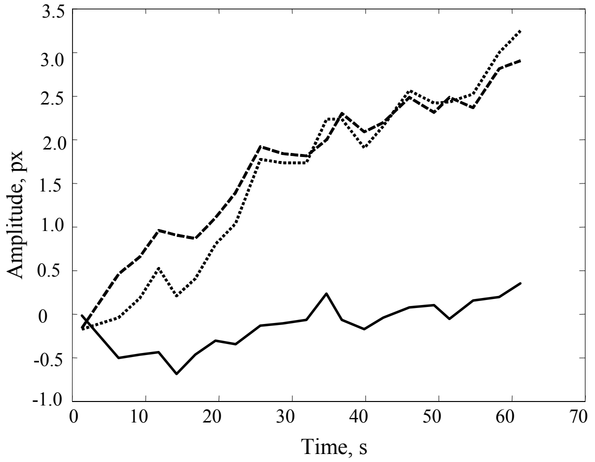

Figure 2. Baseline variation of vertical component during all recording time: solid line – resulting eye movement signal, dotted line – eye pupil centre coordinate, dashed line – head movement.

The oscillations of head were obtained during recordings. They are well noticed during eye fixation. The amplitude is bigger in vertical direction than in horizontal. Oscillations arise spontaneously. Typical example of oscillations is shown on Figure 3. The standard deviations of signals are: 0.225 px for pupil centre, 0.221 px for eye corner, 0.028 px – for gaze direction. Frequency of oscillations is about 6.3 Hz.

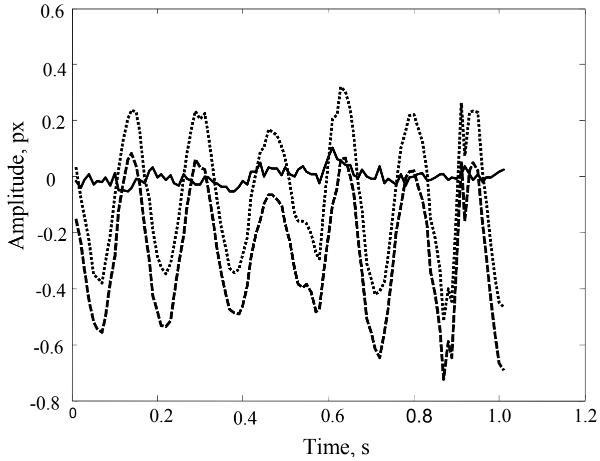

Figure 3. Signal changes during short time period: solid line – resulting eye movement signal, dotted line – eye pupil centre coordinate, dashed line – head movement.

The lowest noise level between head movement detection methods was obtained for eye corner tracking. But the imperfection of method also is evident. The eye corner form changes occur during saccades. This causes errors of saccades' amplitude detection.

The corneal reflection also has drawbacks. The signal amplitude is decreased approximately twice after difference between pupil centre and corneal reflection calculation. This causes decrease of signal to noise ratio. Also after big eye rotations corneal reflection form dramatically changes. Then the centre coordinates detection become inaccurate.

Discussion

Before the study we expected that head movement elimination methods must decrease baseline variation. Unexpected result – noise reduction during fixations. It seemed that oscillations of pupil centre coordinates reflect behaviour of eye (nystagmus). Oscillations of similar amplitude in head movement recordings suggest that they origin from head or body vibrations. Oscillations of video camera also can cause the same results. There we need more detail study.

Also the future work is design of sophisticated algorithm for head movement elimination. Eye corner tracking show good results for removing the baseline variation and head oscillations during fixations, but its performance bad during saccades. We could expect the best result combining some methods and switching between them during on line analysis.

References

Daunys G. and Ramanauskas N., 2004. The accuracy of eye tracking using image processing. In: *Proceedings of the NordiCHI 2004*, ACM Press, Finland, pp. 377-380.

Mulligan J., 1997. Image Processing for Improved Eye-Tracking Accuracy. *Behavior Research Methods, Instruments, & Computers*, 1997, vol. 29, No. 1, pp. 54-65.

Tian Y., Kanade T., and Cohn J., 2000. Dual-state parametric eye tracking. In: *Proceedings of 4th IEEE International Conference on Automatic Face and Gesture Recognition*, March 26-30, 2000, Grenoble, France, 2000, pp. 110-115.

Betke M., Gips J., and Fleming P., 2002. The Camera Mouse: Visual Tracking of Body Features to Provide Computer Access for People With Severe Disabilities. *IEEE Transactions on Rehabilitation Engineering*, 2002, vol. 10, No. 1, pp. 1-10.

EyeChess: A Tutorial for Endgames with Gaze-Controlled Pieces

Oleg Špakov and Darius Miniotas (University of Tampere)

Introduction

Advances in eye tracking have enabled the physically challenged to type, draw, and control the environment with their eyes. However, entertainment applications for this user group are still rare. We present EyeChess: a PC-based tutorial to assist novices in playing endgames.

The EyeChess software generates a virtual chessboard with the standard configuration of 8 by 8 squares (Figure 1). A square is 64 by 64 pixels in size. The chessboard thus occupies a screen area of 512 by 512 pixels. The 32 chess pieces have a dot in the center to facilitate gaze focus; otherwise, they have a common appearance. For the same purpose, the squares on the chessboard are also labeled with dots. The taken pieces appear in the frame on the right. The field above the chessboard is used for providing instructions and other information.

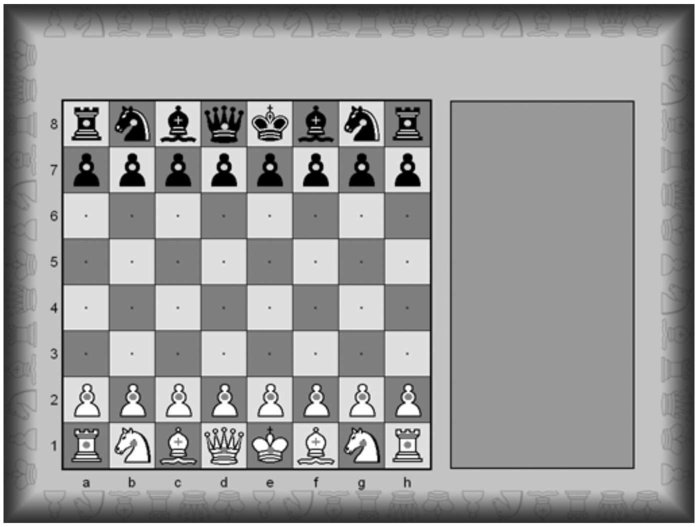

Figure 1. Virtual chessboard

To make a move, the player first selects a piece and then its destination square. After a piece has been selected, the square being looked at is highlighted according to the validity of the move for that piece. A square with a green highlight indicates a valid move, whereas red denotes invalidity. The square being looked at is highlighted with a pop-up border effect (Figure 2).

After the destination square has been chosen (as indicated by a light-yellow background), the application performs an animated piece movement from the previous position to the new one (Figure 3).

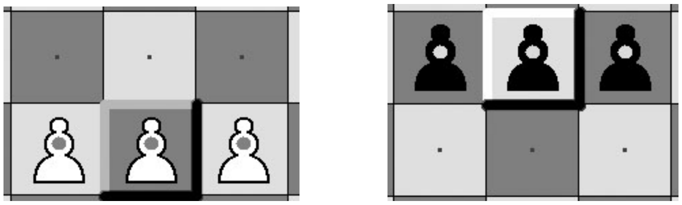

Figure 2. Highlighting with a 3D effect to indicate the active position

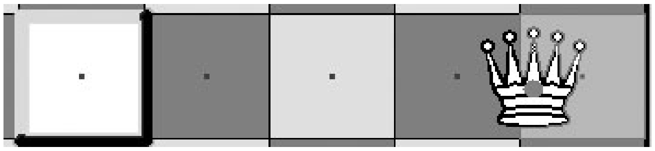

Figure 3. Animated movement of the piece

If the player does not know how to proceed, or starts making mistakes, the tutorial provides a hint. This shows up as a blinking green highlight when the gaze points at the right square.

The EyeChess software supports three methods for selection of pieces: dwell time, blink, and eye gesture (i.e., gazing at off-screen targets). Pilot experimentation revealed that the participants preferred dwell time to the other two techniques. Moreover, since playing chess involved a significant cognitive load, dwell time had to be sufficiently long. Based on the observations from the pilot study, we adopted 1.8 seconds for dwell time.

To evaluate our EyeChess tutorial, we conducted a small-scale user study.

Method

Four unpaid volunteers (1 male, 3 female) participated in the evaluation study. All were students at the University of Tampere with normal vision. Two participants had prior experience with eye tracking technology. All participants were novices in playing chess.

The experiment was conducted on a Pentium IV 3.06 GHz PC with a 19-inch LCD monitor with a resolution of 1024 x 768. A remote eye tracking system Tobii 1750 from Tobii Technologies served as the input device.

Participants were first provided with the guidelines in writing about how to select and move the pieces as well as the form of feedback delivered by the EyeChess tutor. Then, they had to play a series of 20 endgames. In each endgame, the player always started first and was to checkmate Black King (Blacks being played by the computer) in three moves. After calibrating the eye tracker, two endgames were given to the participants to practice before the recording session began.

Results

The average solving time was approximately 71.4 seconds. However, 78% of the time (56 s) was spent on finding the first correct move. The second and third moves took 9.6 s and 5.8 s, respectively (Figure 4).

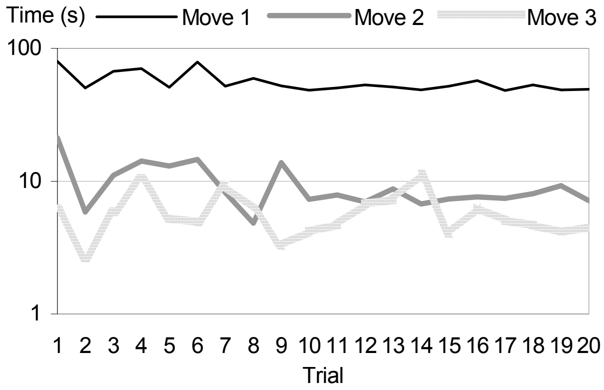

Figure 4. Task completion time vs. endgame number

Out of all attempts to make a move, 18% were wrong. As expected, the greatest portion of those (34%) was related to the first move. Meanwhile, only 6% and 1% of the attempts were wrong while making the second and the third moves, respectively (Figure 5).

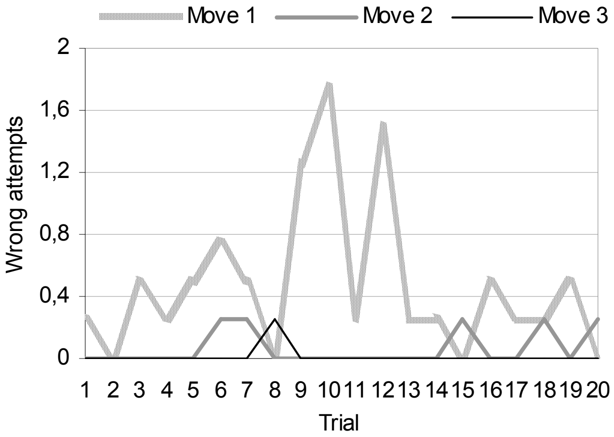

Figure 5. Percentage of wrong attempts vs. endgame number

In 9% of the trials, the participants made two or more mistakes while attempting to make the first move. They then all made use of the hint (blinking square) provided by the tutor. Nobody made two or more wrong attempts after they had found the first correct move.

Conclusion

Preliminary evaluation of the system revealed that dwell time was the preferred selection technique. Participants reported that the game was fun and easy to play using this method. Meanwhile, they found both the blinking and eye gesture methods quite fatiguing. The tutorial was rated helpful in guiding the decision-making process and training the participants.

In the future, we plan to extend the prototype to a fully-fledged game with an opponent: a computer or another player over the Internet.

Towards emotion modeling based on gaze dynamics in generic interfaces

Martin Vester-Christensen, Denis Leimberg, Bjarne Kjær Ersbøll and Lars Kai Hansen
(Technical University of Denmark)

Gaze detection can be a useful ingredient in generic human computer interfaces if current technical barriers are overcome. We discuss the feasibility of concurrent posture and eye tracking in the context of single (low cost) camera imagery. A deformable template method for eye tracking on full-face images is presented. The strengths of the method are that it is fast and retains accuracy independently of the resolution. We compare the method with a state of the art active contour approach, showing that the heuristic method is more accurate.

Detection of the human eye is a relatively complex task due to a weak contrast between the eye and the surrounding skin. As a consequence, many existing approaches use close-up cameras to obtain high-resolution images (Hansen and Pece, 2003). However, this imposes restrictions on head movements. Wang and Sung (2002) use a two camera setup to overcome the problem. We here focus on some of the image processing issues. In particular we discuss the posture estimation within the framework of active appearance models (AAM) and we discuss a recently proposed robust and swift eye tracking scheme for low-resolution video images (Leimberg, Vester-Christensen, Ersbøll and Hansen, 2005). We compare this algorithm with an existing method (Hansen and Pece, 2003) and relate the pixel-wise error to the precision of the gaze determination.

The ingredients in the approach are posture and eye region extraction based on active appearance modeling and eye tracking using a new fast and robust heuristic.

The posture is used to nail the head degrees of freedom and to locate the eye regions. In combination with eye tracking, posture can be used to infer the gaze direction. Active appearance models combine information about shape and texture. Here a face shape consists of n 2D points, landmarks, spanning a 2D mesh over the face in question. The landmarks are placed either in the images automatically (Baker, Matthews and Schneider, 2004) or by hand. Using principal component analysis a generative model of both the shape and texture of faces can be built. By minimization of the difference between the model instance and the image evidence, a face can be detected and subsequently tracked.

In many existing approaches, the shape of the iris is modeled as a circle. This assumption is well-motivated when the camera pose coincides with the optical axis of the eye. When the gaze is off the optical axis, the circular iris is rotated in 3D space, and appears as an ellipse in the image plane. Thus, the shape of the contour changes as a function of the gaze direction and the camera pose. The objective is then to fit an ellipse to the pupil contour, which is characterized by a darker color compared to the iris. To utilize this knowledge, two regions of the eye are considered. A pupil region P is the part of the image I spanned by the ellipse. The background region B is defined as the pixels inside an ellipse, surrounding but not included in P. When region P contains the darker pupil, B contains some of the brighter iris. Thus, the difference in average pixel intensity between the two regions is large. This property is to ensure equal weighting of the two regions, they have the same area.

The template model is deformed by Newton optimization of the cost function given an appropriate starting point. Due to rapid eye movements (Pelz et al., 2000}, the algorithm may break down if one uses the previous state as initial guess of the current state, since the starting point may be too far from the true state. As a consequence, we use a simple adaptive `double threshold' estimate (Sonka, M., Hlavac and Boyle, R., 1998) of the pupil region as starting point.

Although a deformable template model is capable of tracking changes in the pupil shape, there are also some major drawbacks. Corneal reflections, caused by illumination, may confuse the algorithm and cause it to deform unnaturally. In the worst case, the shape may grow or shrink until the algorithm collapses. We propose to constrain the deformation of the model in the optimization step by adding a regularization term.

Additionally a probabilistic contour based tracker is used. Hansen and Pece (2003) describe an algorithm for tracking using active contours and particle filtering. A generative model is formulated which combines a dynamic model of state propagation and an observation model relating the contours to the image data. The current state is then found recursively by taking the sample mean of the estimated posterior probability. The proposed method in this paper is based on Hansen and Pece (2003), but extended with constraints and robust statistics.

A dynamical model describes how the iris moves from frame to frame. Since the pupil movements are quite rapid at this time scale, the dynamics are modeled as Brownian motion (AR(1)).

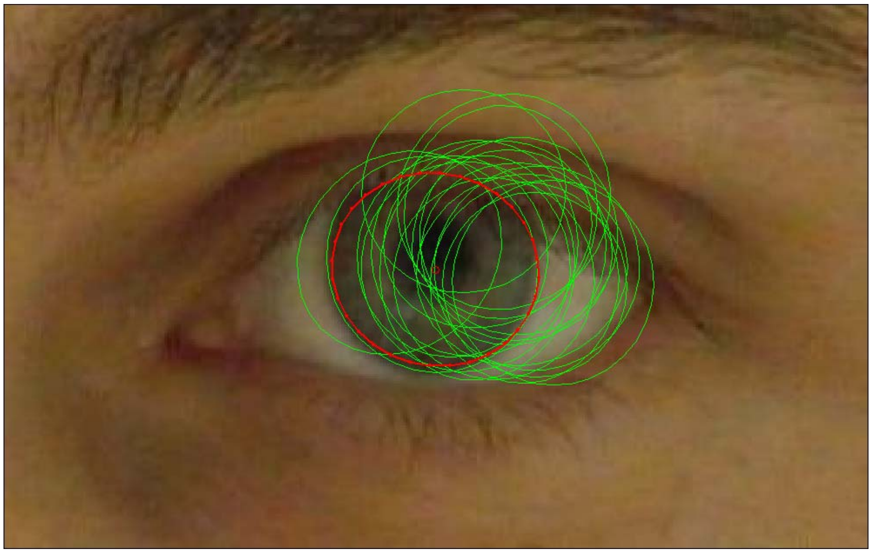

Figure 1. Hypothesized iris locations with final estimate in red

The observation model consists of two parts. A geometric component modeling the deformations of the iris by assuming a Gaussian distribution of all sample points along the contour. Secondly, a texture component defining a pdf over pixel gray level differences given a contour location. Both components are joined and marginalized to produce a test of the hypothesis that there is a true contour present. The contour maximizing the combined hypotheses is chosen.

We propose to weigh the hypotheses through a sigmoid function. This has the effect of decreasing the evidence when the inner part of the ellipse is brighter than the surroundings. In addition, this relaxes the importance of the hypotheses along the contour around the eyelids, which improves the fit.

By using robust statistics, hypotheses, which obtain unreasonably high values, compared to the others, are treated as outliers and rejected. This improves the model towards the artefact introduced by corneal reflections.

A number of experiments have been performed with the proposed methods. We wish to investigate the importance of image resolution. Therefore, the algorithms are evaluated on two datasets. One containing close up images, and one containing a down-sampled version hereof. The algorithms estimate the center of the pupil. For each frame the error is recorded as the difference between a hand-annotated ground truth and the output of the algorithms. This may lead to a biased result due to annotation error. However, this bias applies to all algorithms and a fair comparison can still be made.

References

Baker, S., Matthews, I., and Schneider, J., 2004. Automatic construction of active appearance models as an image coding problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10), October 2004.

Hansen, D.W. and Pece, A.E.C., 2003. Iris tracking with feature free contours. In *Proc. workshop on Analysis and Modelling of Faces and Gestures: AMFG 2003*, October 2003.

Leimberg, D., Vester-Christensen, M., Ersbøll, B.K., and Hansen, L.K., 2005. Heuristics for speeding up gaze estimation. In: *Proc. Svenska Symposium i Bildanalys, SSBA 2005, Malmø, Sweden, SSBA, 2005*, To appear.

Pelz, J., Canosa, R., Babcock, J., Kucharczyk, D., Silver, A., and Konno, D., 2000. Portable eyetracking: A study of natural eye movements.

Sonka, M., Hlavac, V., and Boyle, R., 1998. *Image Processing, Analysis and Machine Vision*, 2nd Edition. International Thomson Publishing.

Stegmann, M.B., Ersbøll, B.K., and Larsen, R., 2003. FAME – a flexible appearance modelling environment. *IEEE Trans. on Medical Imaging*, 22(10), pp. 1319-1331.

Learning to Type Japanese Text by Gaze Interaction in Six Hours

Hirotaka Aoki and Kenji Itoh (Tokyo Institute of Technology)
John Paulin Hansen (IT University of Copenhagen)

Objective

In this paper, we examine learning effects in terms of error frequency and changes in error types as well as typing speed during learning of a gaze interaction system named Japanese version of *GazeTalk* system (Hansen *et al*, 2003).

Procedure

Six Japanese students (three female and three male subjects, ages ranged from 18 to 22 years) participated in the experiment. All of them had normal or corrected-to-normal vision. None of the subjects had previous experience with gaze typing. The task was to type Japanese sentences as fast and accurate as possible using the Japanese version of GazeTalk. Subjects were told to correct a typing error immediately when they noticed it. Each subject performed twenty-two experimental blocks in total during seven days, as successively as possible. Each block included five sentences with a total of approximately 90 characters.

Typing speed was measured in terms of *characters per minute (CPM)* for each sentence. This index was calculated by dividing the number of mixed Kana/Kanji characters in a typed sentence by the typing time in minutes. As to errors, we adopted the following two indices to measure frequencies of erroneous gaze behaviour: the *Rate of* "*Midas Touch" errors* and the *Rate of premature movement errors*. The "*Midas Touch" error* is directly relating to the "*Midas Touch Problem*" recognized by Jacob(1991). In this study, a "*Midas Touch" error* is referred to as an incorrect "gaze" activation of a not-intended-to-type key. The index *Rate of "Midas Touch" errors (RMTE)* represents the frequency of this kind of error per character, and can be calculated for each typed sentence as the total number of "gaze" activations of unintended keys divided by the number of characters included in the sentence. Because the text was very simple, everybody would know how to spell correctly. Therefore we assume that none of the incorrect activations were intended. The *Rate of premature movement errors (RPME)* is also relating to the errors recognized by Jacob (1991). He noticed that it could be difficult for some people to stare at will in order to do a dwell time selection. Naturally, the eyes are moved whenever a piece of information has been noticed and a decision to act has been taken. But if this is done before the end of the dwell time, the selection is cancelled. We counted the number of such unwanted eye movements automatically from log data by determining a threshold of fixation duration for a valid perception. In this study, the threshold was set at 170 msec. The rate of PME is calculated as the number of eye movements away from a correct key position after the threshold duration (170 msec.) but before activation (at 500 msec.) divided by the number of characters included in the sentence.

Results

Typing Speed

Results of 2-way analysis of variance (ANOVA) on the CPM with subjects and blocks as independent variables are shown in Table 1. There were significant differences both between the subjects and between the blocks, but no significant interaction of these two factors was identified. To observe a learning effect in typing speed, Figure1 depicts the transition of mean CPMs with blocks for each subject. The highest CPM of each subject obtained for a single sentence across the entire experimental session is also shown in this figure. Until blocks 7-10, the typing speed of any subject seems to increase. After that, improvement in CPM looks very modest for each subject. This result suggests that the subjects have achieved a stable performance with GazeTalk just after five or six hours of practice, i.e., about 10 blocks.

To quantify the learning effect, we formulated each subject's learning model by the "*power law of practice*", which is represented as $T_n = T_1 n^{-\alpha}$, where n = the number of blocks practiced, T_n = seconds taking to type a character in block n, and α=learning coefficient. Table 2 indicates the results of parameter estimations of a learning model for each subject by use of the least square error method to estimate CPMs after a certain number of hours of practice, e.g., block 200 (i.e., experience with 1,000 sentences). As can be seen in this table, a learning model for every subject is highly significant, and its *R-square* indicates that the estimated learning model accounts for at least 60 % of the variance.

Typing Errors

We identified a significant correlation (r=0.748, p<0.01) between RMTE (Rate of Midas Touch errors) and RPME (Rate of premature movement errors). As expected, there were negative correlations between CPM both with the RMTE (r=-0.582, p<0.001) and with the RPME (r=-0.431, p<0.01).

ANOVA was performed on the RMTE; results are as shown in Table 3. Figure 2 illustrates the transition of the RMTE with blocks for each subject. Regarding learning effect on the Midas Touch error, a rapid decrease was observed during the first three blocks. As the interaction was significant, improvements of this index from block 1 to 3 varied among the subjects. The improvement rate of the RMTE between the first three blocks, referred to as percentage improvement of the index from block 1 to 3 (base: Block 1), was very large, ranging from 58.5 to 93.0% (81.7% in average) across the subjects. In trials succeeding block 3, no significant differences were observed between any two pair of blocks, and the mean RMTE between blocks 3-22 was 0.162. This means that a user made only a few Midas Touch errors after typing 180 characters.

Like the RMTE results, there were significant differences in the rate of premature movement errors (RPME) both between subjects and blocks, as indicated in Table 4. However, no significant interaction effect of these factors was observed. Figure 3 depicts the transition of RPME with blocks for each subject. The rate of premature movement errors dropped rapidly within the first three blocks. No significant difference was observed between blocks 3-22, which is identical to the results of the other error modes mentioned above.

Table 5 summarizes estimated parameters of a learning model for both the RMTE and the RPME and their extrapolation to the 1000th typing of a sentence. In the table, A_n indicates error frequency per character in block n. As can be seen in this table, we could make significant models of practice for almost all the subjects for both of the error types. Based on the extrapolation results of the learning model at n=200, the frequency of typing error may be expected to decrease to 0.01-0.02 times per character – which corresponds to 0.0234 and 0.00515 in RMTE and RPME. This error frequency implies that a trained user selects a wrong key or moves away from a desired key at 7 times the most while he or she is typing 500 mixed Kanji/Kana characters.

To compare the frequency of the two error modes, i.e., Midas Touch errors (MTE) and premature movement errors (PME), we calculated ratios of MTE over the total number of errors for each subject based on blocks. Table 6 shows the result of ANOVA on the percentage of MTE. Each subject's transition of the percentage with blocks is depicted in Figure 4. The result shown in Table 6 indicates that there was a significant difference between subjects but no such difference was observed with the transition of trials, i.e., blocks. So, Midas Touch errors occurred far more frequently, namely 2-6 times more often than the other type of error, PME´s, regardless of blocks.

Subjective Ratings of Errors

Figure 5 indicates the percentage agreement of error estimates for each of the two error modes. The percentage of error agreement was calculated as percentage of subjects whose responses were 5 and 4 (i.e., "5: very frequently" and "4: frequently") over all the subjects. There were significant differences in the percentage agreement of error perceptions both between the two types of errors (F_0 (21, 21) =2.823, p<0.05) and between blocks (F_0 (1, 21) =7.875, p<0.05). The percentage agreement of error estimates of the Midas Touch errors is approximately 2.7 times higher than that of the premature movement errors. This seems to be in accordance with the difference between the actual rates of the two error modes (see Figures 3 and 4). As can be seen from Figure 5, however, the degree of decrease in the percentages is much smaller than the decrease in the actual rates of the two error modes. Considering this result, we infer that the subjects did perceive the actual decrease in frequency of both types of errors but remained more aware of their mistakes than the actual numbers would justify.

Conclusion

The experimental results suggest that users have a potential ability to type 23-29 characters per minute after typing 55 sentences (corresponding to approximately six hours of practice). It was found that the frequencies of both Midas Touch errors and the errors caused by premature movements were initially high (blocks 1-2), but decreased very quickly within the next 3-5 blocks, and the error rates then remained stable at a low frequency. The relative proportion of the two error types did not change with learning, though. The results are in accordance with the findings of Bates (2002): People need some practice with gaze interaction but eventually they will learn to master it.

References

Aoki, H., Itoh, K., Sumitomo, N., and Hansen, J. P., 2003. Usability of gaze interaction compared to mouse and head-tracking in typing Japanese texts on a restricted on-screen keyboard for disabled people, In: *Proceedings of the 15th Triennial Congress of the International Ergonomics Association*, 1, pp. 267-270.

Bates, R., 2002. Have patience with your eye mouse! Eye-gaze interaction with computers can work. In: *Proceedings of the 1st Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT*), Trinity Hall, University of Cambridge
Available at http://www.cse.dmu.ac.uk/~rbates/research/cambseyemouse.htm.

Hansen, J. P., Hansen, D. W., Johansen, A. S., Itoh, K., and Mashino, S., 2003. Command without a click: Dwell time typing by mouse and gaze selections, In: *Proceedings of the 9th IFIP TC 13 International Conference on Human-Computer Interaction, INTERACT 2003,* pp. 121-128.

Jacob, R. K., 1991. The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Transactions on Information Systems*, 9(3), pp. 152-169.

factor	s.s.	d.f.	V	F_o
Subject	1623.1	5	324.6	28.3**
Block	4213.0	21	200.6	17.5**
Subject × Block	1126.8	105	10.7	0.93
Error	6044.4	526	11.5	
Total	13007.3	657		

Table 1. Result of ANOVA on CPM (** *p<0.01, * p<0.05*)

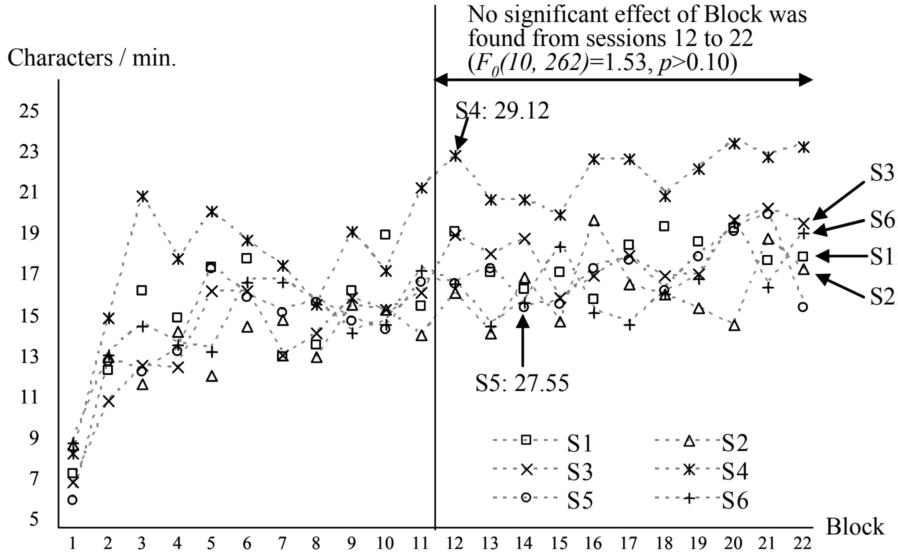

Figure 1. Mean CPM Transitions

Subjects	Learning coefficient	R-square	F_0 (1,20)	Estimated T_1 (sec.)	Estimated T_{200} (sec.)	Estimated CPM in block 200	The highest CPM observed
S1	-0.22	0.61	31.88**	6.18	1.40	30.37	23.47
S2	-0.18	0.70	47.63**	6.32	1.78	25.14	26.86
S3	-0.28	0.83	100.82**	7.45	1.05	36.22	26.12
S4	-0.24	0.67	40.16**	5.35	1.05	39.19	29.12
S5	-0.25	0.68	41.91**	6.95	1.25	32.27	27.55
S6	-0.17	0.64	35.24**	5.75	1.84	25.06	24.71

Table 2. Results of Regression Analysis on CPM (** *p<0.01, * p<0.05*)

factor	s.s.	d.f.	V	F_o
Subject	4.998	5	1.00	6.88**
Block	57.73	21	2.75	18.92**
Subject × Block	23.54	105	0.224	1.54**
Error	76.72	528	0.145	
Total	163.0	659		

Table 3. Result of ANOVA on Rate of ”Midas Touch” Errors (RMTE) (** p<0.01, * p<0.05)

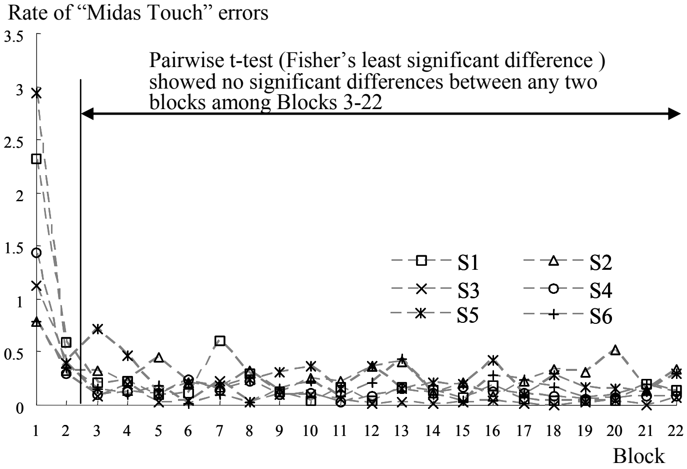

Figure 2. RMTE Transition with Blocks for Each Subject

factor	s.s.	d.f.	V	F_o
Subject	1.748	5	0.350	11.79**
Block	8.298	21	0.395	13.32**
Subject × Block	3.224	105	0.031	1.03
Error	15.66	528	0.030	
Total	28.93	659		

Table 4: Result of ANOVA on Rate of Premature Movement Errors (RPME) (** p<0.01, * p<0.05)

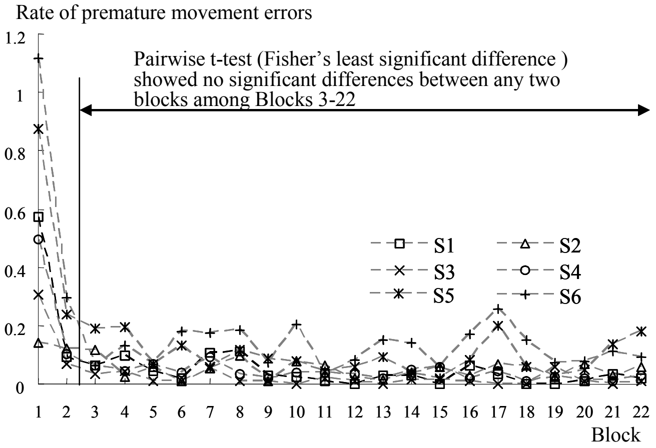

Figure 3. RPME Transition with Blocks for Each Subject

Subjects	Indices	Learning coefficient	R-square	F_0(1,20)	Estimated A_1	Estimated A_{200}
S1	RMTE	-0.91	0.53	22.28**	1.03	0.00846
	RPME	-0.83	0.54	18.81**	0.25	0.00303
S2	RMTE	-0.24	0.14	3.27	0.44	0.124
	RPME	-0.34	0.13	2.98	0.09	0.0157
S3	RMTE	-1.11	0.60	25.18**	0.62	0.00172
	RPME	-0.82	0.50	14.93**	0.11	0.00150
S4	RMTE	-0.63	0.48	18.77**	0.50	0.0175
	RPME	-0.81	0.63	34.76**	0.23	0.00311
S5	RMTE	-0.62	0.47	17.57**	1.02	0.0383
	RPME	-0.66	0.43	15.05**	0.40	0.0120
S6	RMTE	-0.29	0.07	1.52	0.25	0.0535
	RPME	-0.45	0.26	6.87*	0.34	0.0320

Table 5: Results of Regression Analysis on *RMTE* and *RPME* (** p<0.01, * p<0.05)

factor	s.s.	d.f.	V	F_o
Subjects	6.332	5	1.267	9.357**
Session	1.926	21	0.092	0.677
Error	14.21	105	0.135	
Total	22.47	131		

Table 6. ANOVA on the Percentage of Midas Touch Errors (Logit Transformed) (** p<0.01, * p<0.05)

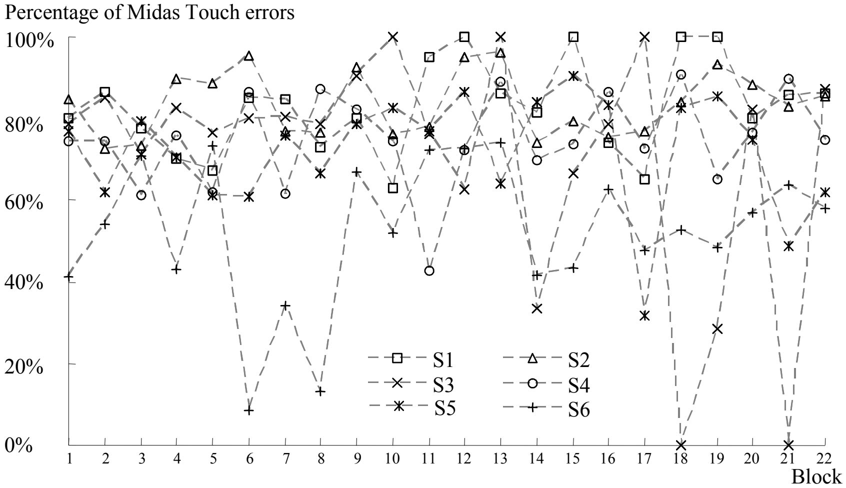

Figure 4. Mean Percentage of Midas Touch Errors Transitions

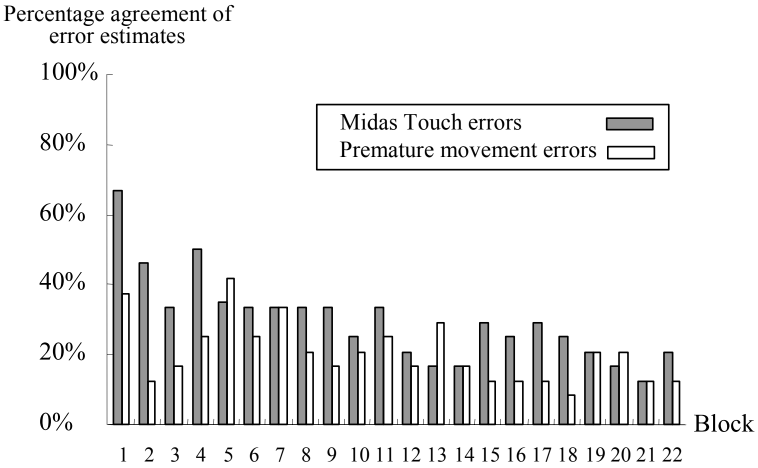

Figure 5. Percentage Agreement of Error Estimates

Dasher's new Gaze-tracker Mode

David MacKay and Chris Ball (University of Cambridge)

Dasher is a gaze-friendly and information-efficient communication interface. Whereas many gaze-based communication solutions pretend that eyes are like fingers, forcing the user to communicate by `dwelling' on on-screen buttons, Dasher offers a radical new approach to writing. We express writing as a continuous navigational task, rather like driving a car. Eyes are well suited to navigation, so Dasher works well with gaze trackers.

With version 1 of Dasher (developed from 1997 to 2002), experts achieved writing speeds of 25 words per minute by gaze direction.

In the last year, in response to user feedback, we have developed a new gaze-tracker mode for Dasher. The new mode handles navigation differently so as to give increased robustness both to gaze-tracker errors and to user errors.

Dasher is free software and can be used to write efficiently in over 100 languages. http://www.inference.phy.cam.ac.uk/dasher/

Fly Where You Look: Enhancing Gaze Based Interaction in 3D Environments

Richard Bates and Howell Istance (De Montfort University)
Mick Donegan and Lisa Oosthuizen (ACE Centre)

Introduction

Gaze based interaction in 3D virtual environments has much to offer motor impaired users such as entertainment, rehabilitation training, collaborative activities with users in remote places, and the opportunity to experience a sense of place afforded by remote locations. It is important that, for these users, interaction devices and techniques are provided that are both efficient in utilizing their residual capabilities and also do not require undue effort or impose undue workload. It has been shown (Donegan 1999) how important it is for the self-image and motivation of motor impaired users to utilize assistive technology both for work and leisure. Without assistive technology, many of these users find it difficult or even impossible to achieve success in these areas. For a significant number of these, eye control offers the potential to achieve success in the most efficient, effective and satisfying way.

Interaction in 3D environments can broadly be characterised as object manipulation, navigation and application control (Hand 1997). Zooming or flying in towards an object can be seen both as a navigation technique and as an object manipulation technique. Temporarily zooming-in on an object of interest to select it means it is easier to select objects with an inaccurate device (Bates and Istance 2002), although it is important to be able to zoom back out to the original position to prevent loss of context and orientation. Without the return to the original position from where the zoom action was initiated, zoom becomes a fly navigation technique ('fly where I look'). 'Intelligent flying' can utilise a similar technique to the 'index of interest' used to determine which object, or objects, in a virtual world a user is attending to most (Starker and Bolt 1990). Initiating an 'intelligent' fly action can assume the target to be that object with the highest index of interest. Additionally the fly can stop at a reasonable distance in front of the assumed object of interest so that it does not fill too much of the visual field, or indeed, to ensure that the user does not fly straight through the object. Moving the gaze point during the fly can either effect small corrections to the flight path or indicate the intended object to fly is not that which has been assumed by the system.

An Experiment in 2D

An experiment was conducted with six users in a complex 2D GUI test environment to measure the effect of providing a basic 'fly' enhancement (Bates and Istance 2002). Hand and eye based pointing devices, or hand and eye mice, with and without the fly enhancement were used to manipulate objects of four angular sizes based of the angle the objects subtended from the eye of the user. These ranged from 0.30 to 1.20 degrees visual angle at 60cm from the screen. Interaction typically lasted for 20 minutes and incorporated 150 test tasks for each user. The objective efficiency (based on time and quality of interaction metrics) and subjective user satisfaction (based on ratings of workload, comfort and ease of use) of the manipulation were measured. For this experiment, the basic fly enhancement was under the full control of each user via micro switches rather than under intelligent software control.

The results showed that the provision of a zoom or fly enhancement greatly increases the efficiency of eye-based pointing on a 2D GUI, without adversely affecting subjective ratings of comfort or workload.

Furthermore, when the user had full control over the extent of the zoom, it was used such that targets of originally different sizes all subtend approximately 1.70 degrees of visual angle at 60cm.

An Experiment in 3D

An experiment was conducted in a virtual environment to examine the extent to which hand-based and eye-based pointing would benefit from a similar fly enhancement to that previously examined with the 2D zoom enhancement. The hand-based and the two eye-based conditions (one with fly enhancement and one without) all used ray casting as the interaction technique, as did Cournia et al. (2003) previously.

Unlike previously reported work, which has used an immersive head mounted display, this experiment was conducted in a reality centre located at De Montfort University, equipped with passive stereoscopic images across an 8-metre wide 150^0 cylindrical screen. In all conditions, users were seated 6 meters away from the curved screen. A desk mounted SMI RED eye tracker was used for eye-based pointing, and a desktop mouse was used for the hand based pointing in order to enable direct comparison with the 2D environment (Figure 1 shows a user seated in the environment, with the eye tracker located on the table in front of the user).

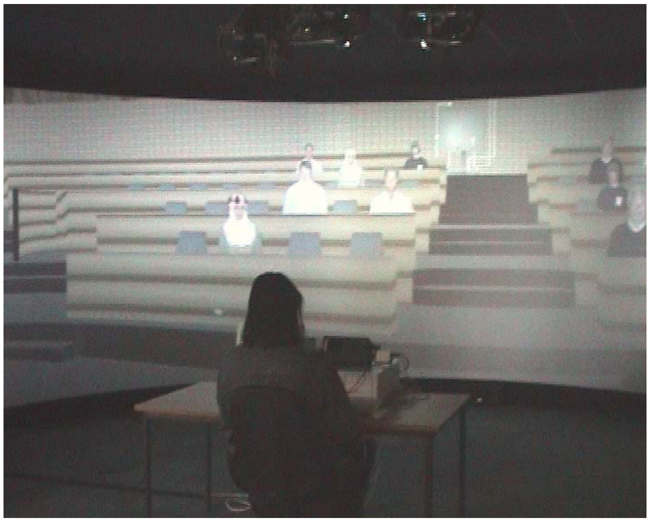

Figure 1. Eye Gaze Pointing in the Virtual Environment

The task was to select one of a group of virtual students in a virtual lecture theatre. Six users took part in the experiment, with interaction typically lasting for 20 minutes and incorporating 144 test tasks for each user. As before, the objective efficiency and subjective user satisfaction of the manipulation were measured.

A further set of trials was conducted to investigate intelligent flying. A 'smart' stopping distance was set at a distance where the visual angle of an object would subtend 2.4°; a compromise to give greatest ease of manipulation without overly enlarging objects and potentially disorienting the user, but also flying sufficiently close to give ease of selection. To do this the fly enhancement was modified such that the fly was stopped automatically when the object subtended a 2.4° visual angle. After manipulation, the user returned to the original starting point by initiating a automatic fly back. This gave the first elements of an 'intelligent' fly interaction mode.

Results

In the hand device conditions, objective task efficiency and subjective user satisfaction were highly dependent on object size, with poor performance for the smaller object sizes. The hand mouse efficiency in a 3D environment showed a marked drop in performance for all object sizes compared with the 2D environment.

Enhancing the hand mouse with the ‘fly’ soft device showed large and significant increases in efficiency for the smallest three object sizes, with no significant improvements for the largest object size, where fly was rarely used.

The efficiency results for the eye mouse without the fly enhancement showed extremely low efficiencies for the smallest objects, with efficiency increasing, as expected from the 2D results, with increasing target size. As with the 2D results, the eye mouse showed lower performance than the hand mouse, although the differences between the devices were considerably reduced in the 3D environment. With the fly enhancement, the efficiency for all target sizes was increased, with object size now having only a minor effect on efficiency, and the eye mouse achieving near parity with the hand mouse.

The subjective hand mouse satisfaction ratings showed increased workload and lower ease of use in the 3D environment compared to the 2D environment. The fly enhancement reduced this workload and increased the ease of use in the 3D environment. The fly enhancement also resulted in an improvement for all ratings in conditions where the eye-mouse was tested.

Considering the trials conducted with the ‘intelligent’ fly mode, device efficiency was essentially unchanged from the basic ‘fly’ mode. Although test subjects tended to fly closer to objects with the eye mouse than the estimated ideal distance, there were no performance benefits from doing so. The results also suggested that reduced workload is possible in comparison to the basic fly mode.

Conclusions

The work has demonstrated that the benefits of enhancing eye pointing by zoom previously demonstrated in 2D interaction are also apparent in 3D interaction. Our results comparing eye and hand pointing in virtual environments accord with those reported by Cournia et al, and show a performance advantage for hand based pointing. Our work shows that this benefit is only apparent for larger target sizes, however, when a ‘fly’ enhancement is provided the performance levels of eye based pointing increase to a similar level to that of hand based pointing. Our initial attempts to go further and introduce a degree of ‘intelligence’ have indicated some success. The limited but promising results suggest that more effort is required to add further intelligence to interaction to gain performance benefits. The addition of an intelligent control based on optimal object subtended angles is currently under investigation and is expected to further enhance performance with eye based pointing in 3D environments, to enable the naturalness and efficiency benefits offered by this modality.

References

Bates, R. and Istance, H. O., 2002. Zooming interfaces! Enhancing the performance of eye controlled pointing devices, *Proceedings of ASSETS 2002, The Fifth International ACM SIGCAPH Conference on Assistive Technologies*, July 8 - 10, 2002, Edinburgh, Scotland.

Cournia, N., Smith, J. D., and Duchowski, A. T., 2003. Gaze- vs. Hand-Based Pointing in Virtual Environment. In: *Proceedings of SIGCHI 2003* (Short Talks & Interactive Posters), April 5-10, 2003, Ft. Lauderdale, FL.

Donegan, M., 1999. Computers and Inclusion - Factors for Success. ACE Centre Advisory Trust.

Hand, C., 1997. A Survey of 3D Interaction Techniques, *Computer Graphics Forum*, 16 (5), pp. 269-281.

Starker, I. and Bolt, R. A., 1990. A Gaze-Responsive Self-Disclosing Display, *Proceedings of CHI ‘90*, ACM, pp. 3-9

An Eye Movement Study of the Think Aloud Technique's Implications for Cognitive Processes

Kristin Due Hansen (Risø Laboratories)

The think-aloud technique is a valuable tool because it gives access to the thoughts of the users, which would otherwise be considered a ‘black-box’ to usability experts and systems developers. Think-aloud tests are generally considered to be well suited for identifying and explaining the causes of usability problems. However, there seems to be a contradiction between the cognitive psychological theory that forms the basis for the technique in usability testing, and practice in usability evaluation studies.

The question under investigation is what kind of eye movement patterns evolve when the think-aloud procedures are as fundamentally different as they seem to be in cognitive psychology compared with usability. From the eye movements an attempt is made to draw conclusions regarding differences in cognitive information processing. Given that eye movements are the primary measurement in the experiment, an introduction to the method is presented together with possible ways of analyzing eye movement data.

Two assumptions are made on the basis of the literature: 1) Current procedures in the usage of the think-aloud technique potentially change the cognitive processes of the users. Consequently, the data from the test does not reflect the cognitive processes of the users outside the test situation and 2) think-aloud testing done according to the theory does not change the cognitive processes. Thus two kinds of think-aloud tasks are created: tasks consistent with guidelines from theory, and tasks consistent with guidelines from usability evaluation studies. These are compared with silent control tasks, which represent the situation outside the test setting.

The assumptions are not supported by the results. The results indicate that information processing is more challenging to the participants in connection with the silent tasks. One interpretation could be that both of the think-aloud tasks may have improved the visual and cognitive performance.

Towards Communication of Unusual Things: Attention, Consciousness and, perhaps, Feeling

Boris M. Velichkovsky, Sebastian Pannasch, Markus Joos, Jens R. Helmert and Sven-Thomas Graupner
(Technische Universität, Dresden)

The implementation of gaze-based-interaction systems will only be successful if underlying mechanisms are taken into account. The registration and analysis of eye movements can be pursue two different purposes. Although this is a more artificial separation it should be stated here: eye movements can be used to analyse the ongoing information processing or they serve to control external devices (e.g. gaze typing). While in the first case, the eyes are more concerned with the representation of internal processing, in the latter they are more treated as an input device. However, the development of gaze-based applications needs to consider the fact that our eyes are *input* and *output* systems at the same time.

A promising approach to this task is based on the distinction between two routes of visual processing. The classification first came to prominence with a special issue of *Psychologische Forschung* in 1967 (Held, Ingle, Schneider, and Trevarthen, 1967) and has survived decades of critical analysis under different guises (e.g. dorsal vs. ventral or ambient vs. focal). Surprisingly, the idea of two visual systems has never been related to the major output of visual processing, i.e. eye movements. Is it possible that both visual pathways can selectively influence oculomotor mechanisms and that the balance of these influences can change flexibly? Assuming there are indeed different processes accompanied by distinct eye movement behaviour it would be interesting to find a method for their reliable distinction, for instance, by a combined consideration of parameters of both, fixations and saccades. Neurophysiology tells us that dorsal stream areas can mediate large saccades throughout much of the visual field on the basis of simple visual properties such as contrast and location. In contrast to this, ventral stream areas receive inputs chiefly from central regions of the retina, but construct a richer, memory-based representation of the stimulus, including its semantic properties. We present evidence that analysis of eye movements can provide an on-line identification of the extent to which these modes are involved. This opens a perspective on explicating perception, as only ventral and not dorsal pathway seems to be related to conscious representation. Three examples of our work will be given to demonstrate the feasibility of this approach.

In a recent experiment, we tried to validate these eye movement parameters by testing assumptions about memory representations related to these two modes (Velichkovsky, Joos, Helmert, and Pannasch, in press). After a short presentation of various real world scenes, subjects had to recognize cut-outs from them, which were selected according to their fixation parameters. Random cut-outs from not seen pictures (catch trials) were also presented. The results confirmed our hypothesis: cut-outs corresponding to presumably focal mode of processing were better recognized than cut-outs similarly fixated in the course of ambient exploration.

In a study of a simulated driving activity, we received evidence that these assumptions may be of importance for the perception of hazardous situations (Velichkovsky Rothert, Kopf, Dornhoefer, and Joos, 2002). 12 healthy and well-trained subjects had to drive in a dynamic virtual environment fulfilling all the common rules and in particular preventing accidents. The hazardous events were sudden changes of traffic lights from green to red, pedestrians' appearance on the road and the behaviour of other drivers. The experiment, which lasted for 5 consecutive weeks, has allowed collecting a large database on parameters of eye movements in this dynamic situation and on their correlation with correct or erroneous reactions to dangerous events. First of all, we found a systematic combination of the visual fixation duration with amplitude of the following saccades. There have been two distinctive segments on the scale of fixation durations. The first segment, with fixations from 90 to about 260 ms, was related to larger saccades of more than 5 deg. In other words, these saccades aimed at targets seen as blobs not as individualized objects – a strong case for the ambient mode of processing. Fixations longer than 260–280 ms rather seemed to be related to focal processing: they initiated saccades mainly within the parafoveal region where objects are relatively easily seen and continuously attended. The next major result of the study was a strong relationship between parameters of 2 to 3 visual fixations that immediately preceded a hazardous event and subject performance: if such an event hit them in their ambient processing mode there was a significantly higher chance for an error than otherwise.

In two recent studies, we explored the biological basis of social interaction with virtual characters. Anthropomorphic virtual characters were presented which appeared moving on-screen and turned either towards the participant or towards someone beside them. In dynamic animations, virtual characters then exhibited FACS coded facial expressions, which were judged as socially relevant, i.e. indicative of the intention to establish interpersonal contact. Otherwise facial movements were shown which were judged as arbitrary. These four conditions thus established a two-by-two factorial design. This paradigm was developed for the purpose of an fMRI study and a study recording eye movements and facial muscle activity (EMG). Functional neuroimaging revealed that medial prefrontal activation is observed not only during one's own personal involvement in social interaction- as indicated by adequate facial expressions - but also during the experience of an interaction between the virtual character and a third other. Similarly, differential EMG activity was observed when the virtual characters smiled towards the human observer, but also when the smiles were directed towards someone else. Conversely, eye movements of human participants showed that the intensity of visual attention – as manifested in visual fixation duration – is specifically related to having eye-to-eye contact with a virtual other. The data from these two studies demonstrate a clear-cut difference between visual attention and neuro- and electrophysiological correlates dependent upon the observer's personal involvement, i.e. adopting a second-person perspective, versus being a passive bystander, i.e. adopting a third-person perspective.

References

Held, R., Ingle, D., Schneider, G., and Trevarthen, C., 1967. Locating and identifying: Two modes of visual processing. A symposium. *Psychologische Forschung*, 31, pp. 42-43.

Velichkovsky, B. M., Rothert, A., Kopf, M., Dornhoefer, S. M., and Joos, M. (2002). Towards an express diagnostics for level of processing and hazard perception. *Transportation Research, Part F, 5*(2), pp. 145-156.

Velichkovsky, B.M., Joos, M., Helmert, J.R. and Pannasch, S. (in press). Two visual systems and their eye movements: Evidence from static and dynamic scene perception. *Proceedings of the XXVII Annual Conference of the Cognitive Science Society*.

Head and Eye Tracking Inside Intelligent Houses

Fulvio Corno and Alessandro Garbo (Politecnico di Torino)

Introduction

Recently, many devices, usually present inside the house, have been improved to meet the needs of elderly or disabled people and to obtain more self-sufficiency. New central-lighting plants let the users control several electronic devices directly from a central control point. Open the doors, open the shutters, turn on the lights and so forth need increasingly smaller efforts. The word *domotic* concerns the ability of the house to become, as it were, more intelligent about the requests of the people who live inside it. However, some of these devices are still precluded to some people, such as disabled ones with severe motor diseases, that can’t use the hands to handle them.

For this reason an eye tracking system can make it possible for disabled people with severe motor diseases, to become more independent about the world that surround them by accessing domotic technology. Eye tracking system could be used not just for communication or for using a Personal Computer, but also for environmental control: eye tracking would act as an interface between the disabled person and the devices designed for people with less critical diseases.

Objectives

The main purpose of our research work is the task to combine several systems, designed and developed for many different purposes, into an application that is able to satisfy or at least meet the requests of a small number of users.

On the market there are already many devices that are able to attend to exhausting tasks as opening or closing shutters, lifting a wheel chair over stairs or help people to rise from the bed or from the bathtub. Other devices have been planned for lighter tasks that are easy for people in good health but could be uncomfortable for disabled people. For example, there are devices that can open and close doors, turn on and off lights in every room from a central control point. Moreover, electronic household appliances may be able to execute indirect commands coming, for example, from the electrical equipment. And yet these devices or mechanisms have an isolated behavior because they haven’t been imagined to work together.

Another system that will be part of our application is new electrical equipment. Until few years ago, conventional equipment was certainly characterized by high complexity and every control device needed a separate and distinct circuit. That obviously takes a remarkable increase of set up time, as well as restrictions, modifications or additions upon pre-existing equipments.

The solution to this problem is suggested by new digital technologies that permit to replace conventional equipment with *intelligent devices* that are able to communicate to each other. Every device, using a digital bus for the communication, takes care of data processing and sends the processed data to other nodes. This system lets to control any home device (lighting, automatism, alarms and so on) in two different ways: from any point of the house or remotely from a cell phone, telephone or Personal Computer.

The union of simple and intelligent devices with new electrical equipment forms what is called a domotic house. This term involves all the sciences and techniques correlated with the data processing utilizable inside the house. It is possible to define it as an intelligent and automatic house equipped with mechanisms that can carry out certain operations. An automatic house can be intelligent if it places itself at user’s disposal, if it can really improve the quality of life and if the cost for its realization and management can be justified, on the other hand, by some useful solutions for the user.

An interesting component to be integrated into an intelligent house is the Eye and Head tracking system. Our system [1], currently a research prototype, makes use of low-cost cameras to capture the head and eye's images. The most important characteristic of the applications developed for this system is the ability to change their layout according to the accuracy that the system can give. Additionally, unlike normal applications presented by other Eye and Head tracking systems that are able to help the communication of the users, our application is also capable of environmental control. This application helps the user to check and change the state of the devices that surround him.

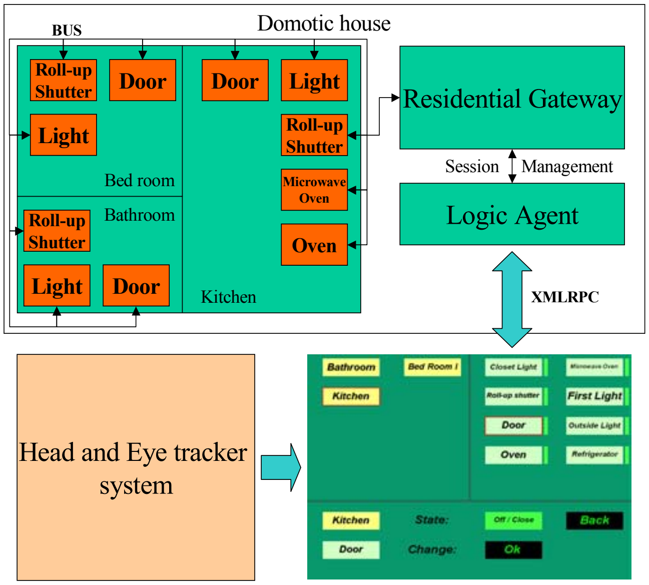

Figure 1. System architecture

System architecture

Inside the domotic house all the devices are connected with the same domotic bus (Figure 1). There are only two possibilities to connect the devices to the bus: a control connection or a power connection. The control connection carries signals to and from endpoints such as switches or relays. A power connection is useful to connect standard devices that need electrical power to work. Every connection has a univocal network address. For example, when users operate the light switch in the kitchen (control connection) the switch produces a short message that is forwarded on the bus. This message includes the power connection address and the command relevant to the kitchen light. The message reaches all the power connections, in the house, but only the power connection with the same address collects and processes it, giving power to the corresponding bulb.

Besides control connections available in the rooms there is another device, called Residential Gateway that can send messages on a TCP/IP connection to a remote user and forward the user's messages over the bus of the electrical equipment. For security reasons the remote user must identify himself/herself with a password; then the exchanged messages are encrypted and the TCP/IP connection is interrupted if unused for long time. For the complexity of the connection rules with the Residential Gateway we have designed another module called Logic Agent. This module (written in Java) opens the connection, sends the password, checks the connection, keeps it active and encrypts and sends the user's commands. Additionally, the Logic Agent shows a friendly interface for the applications that interact with the domotic house. Through XMLRPC messages the Logic Agent can show to the applications the available devices inside the house so that the application can get control over the devices through simple XMLRPC calls.

Once the devices configuration inside the rooms has been received, the application (written in Visual C++) permits the user to navigate through the various devices and to select them and check or change their state. The user, with the Head and Eye tracker system, has only to move the face or the gaze towards the items (commands) shown by the application, selecting them with a gaze fixation and change their state if it is needed (Figure 1, the application works in Head tracking mode; Figure 2, it works in Eye tracking mode).

Now the whole system can only process the commands composed by the user, but in the near future both the Logic Agent and the application, will be extended to test and learn the user's behavior and to show him/her more complex commands that involve multiple devices at the same time. For example, the system can understand that every day the user gets up at eight o'clock and before breakfast he has a shower. So it can show to the user a single command to turn on the oven in the kitchen for the breakfast and the water heater for the shower-bath.

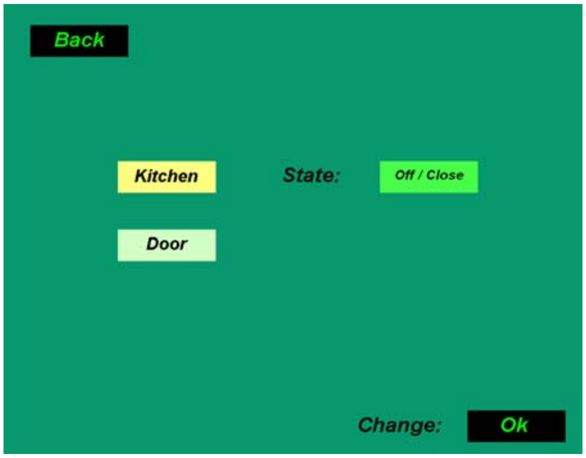

Figure 2. Application layout with the system in Eye tracking mode (less accurate)

References

[1] Corno, F. and Garbo, A., 2005. Multiple Low-cost Cameras for Effective Head and Gaze Tracking. *HCI International 2005,* 22-27 July 2005, Las Vegas, Nevada USA.