EYE TRACKING USING ARTIFICIAL NEURAL NETWORKS FOR HUMAN COMPUTER INTERACTION

Erna Demjén[1], Viliam Aboši[2], Zoltán Tomori[1, 2]

[1]Institute of Experimental Physics, Slovak Academy of Sciences, Košice, Slovakia

[2]Department of Cybernetics and Artificial Intelligence, Technical University of Košice, Košice, Slovakia

Corresponding author:

Erna Demjén

Department of Biophysics

Institute of Experimental Physics

Slovak Academy of Sciences

Watsonova 47,SK-040 01 Košice

e-mail: demjen@saske.sk

phone: +421 55 720 41 32

Short title: Eye Tracking for Human Computer Interaction

EYE TRACKING USING ARTIFICIAL NEURAL NETWORKS FOR HUMAN COMPUTER INTERACTION

E. Demjén[1], V. Aboši[2], Z. Tomori[1, 2]

[1]Institute of Experimental Physics, Slovak Academy of Sciences, Košice, Slovakia

[2]Department of Cybernetics and Artificial Intelligence, Technical University of Košice, Košice, Slovakia

Summary

This paper describes an ongoing project that has the aim to develop a low cost application to replace a computer mouse for people with physical impairment. The application is based on an eye tracking algorithm and assumes that the camera and the head position are fixed. Colour tracking and template matching methods are used for pupil detection. Calibration is provided by neural networks as well as by parametric interpolation methods. Neural networks use back-propagation for learning and bipolar sigmoid function is chosen as the activation function. The user's eye is scanned with a simple web camera with backlight compensation which is attached to a head fixation device. Neural networks significantly outperform parametric interpolation techniques: 1) the calibration procedure is faster as they require less calibration marks and 2) cursor control is more precise. The system in its current stage of development is able to distinguish regions at least on the level of desktop icons. The main limitation of the proposed method is the lack of head-pose invariance and its relative sensitivity to illumination (especially to incidental pupil reflections).

Keywords

human-computer interaction, gaze control, neural networks, eye tracking

Temporarily or permanently disabled people often cannot handle computers using traditional interfaces, e.g. keyboard or mouse. Supplying them with appropriate devices can potentially increase their quality of life. A wide range of such devices (both invasive and non-invasive), exploiting the remained abilities of these people, has appeared in recent years to meet this need, such as simple switches controlled by sucking and blowing (Edwards, 1995), adapted joysticks, keyboards, cursor control by voice (Harada *et al.*, 2008), head movement (Bérard, 1999), eye (Duchowski, 2007), EEG (Wolpaw *et al.*, 1991, Sellers *et al.*, 2006), EOG (Wolpaw *et al.*, 2000), etc.

Object tracking is a well-researched area of computer vision (Yilmaz *et al.*, 2006) and is extensively used in practice (Petrasek *et al.*, 2009, Schutova *et al.*, 2009, etc.). Applications that track gaze have various motivations and their history goes back far beyond the occurrence of first computers. Early invasive approaches that required direct contact with the cornea were gradually replaced by less invasive or non-invasive devices. Commercial solutions exist, e.g. Tobii eye tracker. Unfortunately, their cost prevents their wider usage. On the other hand, a growing number of research groups have set a goal to design low-cost gaze tracking applications.

Despite intense research on eye tracking, it remains a very challenging task, especially due to significant pupil reflectivity and the variability of the shape and openness of the eye. The majority of current eye tracking methods are based on active IR illumination. However, their safety is questionable and is a subject of research. Other approaches utilize template based correlation maximization, simple elliptical models or complex shape descriptors of the eye, etc. (for more detailed review see Hansen *et al.*, 2010). Calibration issues and hardware setup are also critical steps for the system design. Systems using detection of glints resulted from IR illumination are often free of the calibration procedure. More sophisticated and expensive systems use more cameras and sources of illumination to eliminate the need of head pose fixing, and increase the tolerance of the system under different conditions.

This paper describes the IMouse application which should fulfil the following demands:

- cursor control without a computer mouse
- modest hardware requirements
- ability to accommodate to different users
- low cost

The aim of the tracking algorithm is to locate and track the user's eye in consecutive frames of the video stream (Fig.1). The input to our algorithm is the region of interest, where the search procedure takes place, i.e. only the eye image. Two algorithms were tested: colour tracking algorithm and template matching. Colour tracking tracks black colour, the dominant tone of the pupil. This can be easily distinguished from both the sclera and the skin. Template matching compares a predefined set of templates (pupil images) against image regions (in our case eye image) minimizing their normalized squares of differences. Template matching is more reliable, but computationally more demanding than colour tracking.

After the tracking of the pupil, conversion between its position and the position of the cursor has to be performed. For this purpose, an interaction with the user is required where a series of dots, so-called calibration marks, appears on the monitor and the user confirms that he/she is looking at them. IMouse uses neural networks to approximate a function that assigns to every single eye position a position of the cursor. IMouse uses two neural networks with a single input, a hidden layer with number of neurons dependent on the screen resolution and one output neuron. Inputs of the networks are the x and y coordinates of the eye while the outputs are the x and y coordinates of the cursor position.

Back propagation was used for learning. Bipolar sigmoid function was used as the activation function, where the normalized input and output is in range from -1 to 1. Following formulas were used for normalization of the neural network inputs and outputs:

$$XIN = \left(\frac{x}{x_{screen}} - 0.5\right) \times 2 \qquad (1)$$

$$YIN = \left(\frac{y}{y_{screen}} - 0.5\right) \times 2 \qquad (2)$$

$$x = \frac{NOX+1}{2} \times x_{screen} \qquad (3)$$

$$y = \frac{NOY+1}{2} \times y_{screen} \qquad (4)$$

where *XIN* and *YIN* are the final inputs of the neural network. NOX and NOY are outputs calculated by the neural networks, while x and y are the coordinates of the eye position and x_{screen} and y_{screen} are the width and the height of the screen, respectively, each in pixels.

Three interpolation methods were implemented as an alternative to the neural network calibration procedure. The first is based on a look-up-table and enables to select only predefined squared regions of the screen by gaze. This algorithm calculates a two dimensional array where in the first column there are positions of the eye and in the second one there are coordinates of the cursor positions. The next two methods interpolate screen coordinates with linear and quadratic least square fitting, where the precision is improving with a growing number of calibration marks.

The camera and the head position were fixed during the experimental procedure. A simple web camera (MSI WebCam StarCam Clip) with backlight compensation attached to a head fixation device was used. The head position was fixed by a chin rest, while the head distance from the camera as well as from the monitor were adjustable. The quality of the camera and its sensitivity are important in poorly illuminated environments.

Various distances from the monitor (20cm - 100cm) were applied to test the outcomes of the method. Small distances naturally yielded better results since eye movements are more pronounced due to increased angular distance between close pixels.

Neural networks significantly outperformed the look-up-table based technique (both in time needed for calibration and in accuracy of the cursor control), while least square interpolation techniques gave comparable results to neural networks. On the other hand, neural networks handle situations more robustly when the camera is not placed in line with the user's eye. The system in its current stage is able to distinguish regions at least on the level of desktop icons. Moreover, increasing the number of the calibration marks increases also the accuracy of neural networks while least square methods give approximately the same results.

The comparison of the learning methods is in *Tab1*. The column "usage possibilities" evaluates the potential of the given method to adapt to various operational conditions as well as the range of possible applications. There can be seen that the look-up-table based method gave the worst results. This method can satisfy only applications where the precision is not critical, e.g. typing a text with a word prediction algorithm. The rest of the methods attained similar results as the calibration cycles took 21 points each: 20 for calibration itself and 1 as a check of the calibration process. Regression methods are quicker to calibrate and their accuracy is comparable, but future possibilities are more promising for neural networks. Neural networks could effectively eliminate a shaking cursor that was observed with each tested method, as well as they can potentially be used to predict the eye movement.

The IMouse application does not solve other functions of the computer mouse, e.g. clicking. Those can be ensured with various techniques considering the user's limitations (fixing the gaze on the desired area of the screen, eye blinking or using an external switch, just to mention a few).

The main drawback of the system is the lack of head-pose invariance and its relative sensitivity to illumination – especially to incidental reflections in the pupil. A tracking algorithm incorporating a priori information about the pupil shape and colour could compensate illumination artefacts. On the other hand, even more sophisticated algorithms cannot guarantee usability under arbitrary circumstance. Therefore, it seems to be a better option to define light sources and use polarization filters to prevent reflections from the monitor. Another possibility is to utilize the bright eye effect, which occurs when enough concentrated infrared light bounces off the back of the eye and reflects as white. For creating the possibility of head movement, a high resolution camera with remote directional and zoom control is considered to capture the whole face. In this case the tracking should also deal with face and eye localization using a classification algorithm.

We expect that our outcomes in gaze tracking methodology could be easily exploited for purposes other than cursor control and in some more specific application areas, for example in eye controlled microscopy.

Acknowledgement

This work was supported by the research grants from the Slovak Grant Agencies VEGA No. 2/0164/08, 2/0191/11, APVV 0682-07, and by the European Regional Development Fund, Project no.26220120033.

References

F. BÉRARD: The Perceptual Window: Head Motion as a New Input Stream. *Proc 7th IFIP Conf. on Human-Computer Interaction*, 238-244, 1999.

A.T. DUCHOWSKI: Eye Tracking Methodology: Theory and Practice. *Springer-Verlag*, 2007.

A. D. N. EDWARD: Extra-Ordinary Human Computer Interaction. *Cambridge University Press*, 1995.

D.W. HANSEN, W. JI: In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. *IEEE Trans. Pattern Anal. Mach. Intell.,* **32**: 478-500, 2010.

S. HARADA, J.A. LANDAY, J. MALKIN, X. LI, AND J.A. BILMES: The Vocal Joystick: Evaluation of Voice-based Cursor Control Techniques for Assistive Technology. *Disability and Rehabilitation: Assistative Technology, Special Issue ACM Conf. on Comp. and Access.*, **3**: 22-34, 2008.

T. PETRÁSEK, A. STUCHLÍK: Serotonin-Depleted Rats are Capable of Learning in Active Place Avoidance, a Spatial Task Requiring Cognitiv Coordination. *Phys. Res.* **58**: 299-303, 2009.

B. SCHUTOVÁ, L. HRUBÁ, M. POMETLOVÁ, K. DEYKUN, R. ŠLOAMBEROVÁ: Cognitive Functions and Drug Sensitivity in Adult Male Rats Prenatally Exposed to Methampetamine. *Phys. Res.* **58**: 741-750, 2009.

E. W. SELLERS, E. DONCHIN: A P-300-Based Brain-Computer Interface: Initial Tests by ALS Patients. *Clin. Neurophysiol.*, **117:** 538-548, 2006.

J. R. WOLPAW, D. J. MCFARLAND, G. W. NEAT, C. A. FORNERIS: An EEG-based brain-computer interface for cursor control. *Electroencephalogr. Clin. Neurophysiol.,* **78**: 252-259, 1991.

J. R. WOLPAW, N. BIRBAUMER, W. J. HEETDERKS, D. J. MCFARLAND, P. H. PECKHAM, G. SCHALK, E. DONCHIN, L. A. QUATRANO, CH. J. ROBINSON, T. M. VAUGHAM: Brain-Computer Interface: A Review of the First International Meeting. *IEEE Trans. Rehabil. Eng.*, **8:** 164-173, 2000.

A. YILMAZ, O. JAVED, M. SHAH: Object Tracking. A Survey. *ACM Comput. Surv.* **38**: 13, 2006.

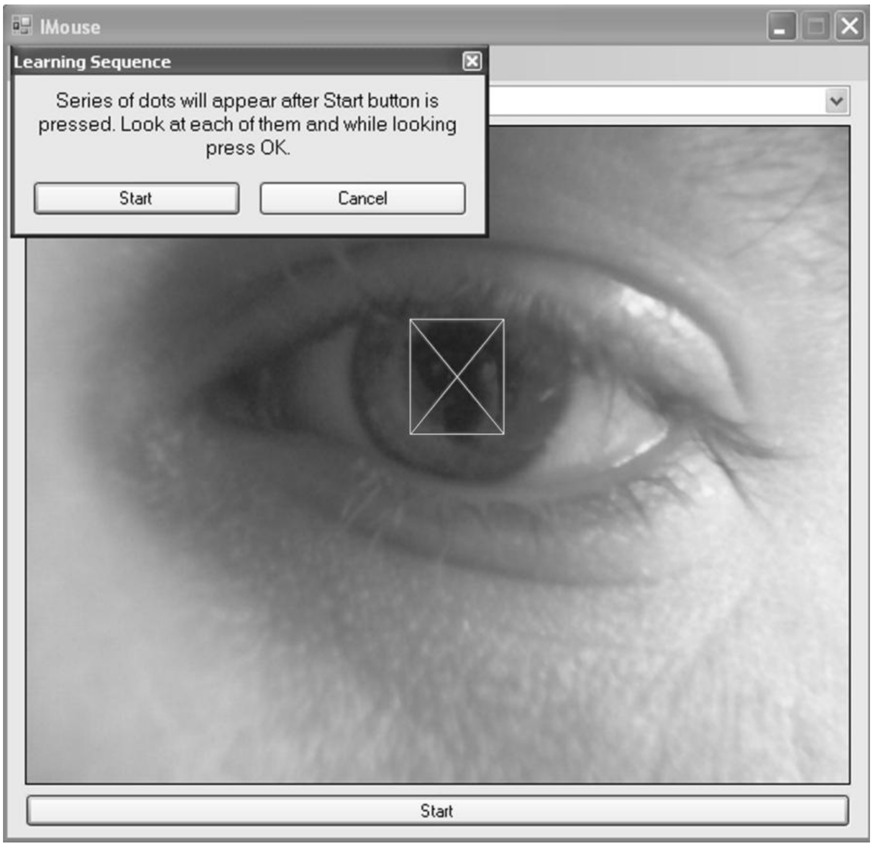

Figure 1. Eye pupil tracking. The centre of the circumscribed rectangle is considered as the pupil's coordinates. The figure is a screenshot of the application before the calibration procedure.

Method	Calibration points	Accuracy	Usage possibilities
LUT Table	(resX/64) x (resY/64)	low	low
Linear Regression	21-42	high	good
Quad. Regression	21-42	high	good
Neural Networks	21-63	high	high

Tab.1 Comparison of methods