A Novel Neural Eye Gaze Tracker

Diego Torricelli, Michela Goffredo, Silvia Conforto, Maurizio Schmid and Tommaso D'Alessio

Department of Applied Electronics
University Roma Tre
Via della Vasca Navale 84, I-00146 Roma, Italy

Abstract. A gaze tracking system, based on a novel neural approach, is proposed. The work is part of a wider research project concerning the development of human computer interfaces (HCI) addressed to disabled people, that could overcome the drawbacks of most of the existing methods for gaze tracking that require either intrusive devices or expensive equipment. This work, instead, aims at developing a low cost, completely non-intrusive and self-calibrating system which combines different techniques for three blocks in Eye Gaze Tracking, i.e. blink detection, feature extraction and neural computing. The experimental results show good accuracy in eye gaze tracking (rmse < 1 degree), and adequate generalization performance (rmse < 2 degrees).

1 Introduction

The detection of human gaze has been an active research topic in the last century, in several branches of Medicine, such as Ophthalmology, Psychology and Neurology [1, 3]. In this context Bioengineering has been requested to solve the problem of Gaze Tracking, i.e. the measurement of the gaze direction. In the last two decades, with the emergence of interactive applications, several gaze tracking systems (otherwise called Eye Gaze Trackers, EGT) have been developed to provide useful tools of interaction between humans and computers.

Most of the existing gaze trackers that provide good accuracy are intrusive. These systems either use devices such as chin rests to restrict head motion, or require the user to wear cumbersome equipment, such as special contact lenses, electrodes and head mounted cameras [7, 8, 13].

Remote eye gaze trackers (REGT) represent a good non-intrusive alternative, mostly based on video analysis. In this context, REGT can be classified into two categories: feature-based and view-based. In the feature-based techniques, geometric features of the eye are extracted by image processing. Thus, by analysing the relationship between the geometric parameters, the gaze direction is computed. In this field the most relevant technique is based on infrared light [6, 9, 10, 12] to enhance the contrast between the pupil and the cornea .

Torricelli D., Goffredo M., Conforto S., Schmid M. and D'Alessio T. (2006).
A Novel Neural Eye Gaze Tracker.
In *Proceedings of the 2nd International Workshop on Biosignal Processing and Classification*, pages 86-95
DOI: 10.5220/0001224300860095

The view-based techniques do not look for geometric features. The images of the eyes are treated as points in a high-dimensional space and computed in different ways. This is what happens in the appearance-based [11] and Artificial Neural Networks (ANN) methods [2]. The ANN systems learn to associate each image with a direction of gaze, without any preliminary image processing to detect the features of interest. An accurate and extensive calibration/training procedure is required to provide the system with the right set of sample images for a good mapping between the image space and the gaze direction space. View-based techniques have the advantage of being easy to implement and generally more robust than feature-based methods. A general scheme of the classic approach for gaze tracking is shown in figure 1.

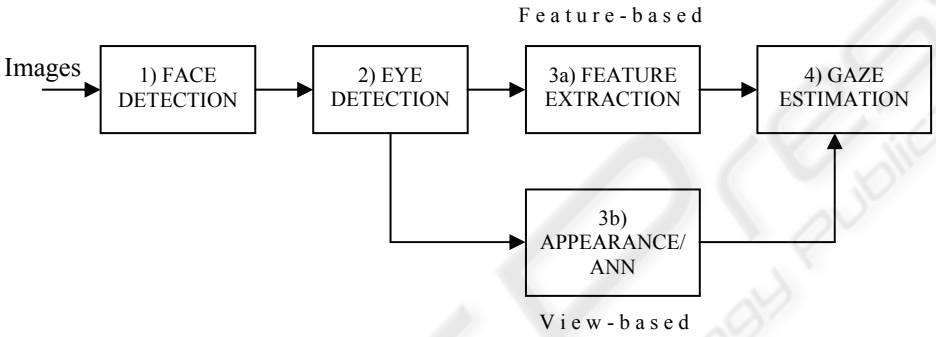

Fig. 1. Scheme of a general gaze tracking procedure.

One of the recognized limits of EGT comes from head motion artifacts: since the image of the eyes changes with head positions, almost all the gaze trackers are sensitive to head motion. Aim of this paper is to present a novel method for gaze detection, where the elimination of the preliminary phases (face and eye detection) and the use of a specifically trained ANN allow to simplify and improve the procedure.
The paper is organised in the following way: in the next section the overall method is detailed, and a subsection is devoted to the design of the ANN. In section 3 the experimental tests and results are presented. In section 4 a brief discussion is given, with specific considerations to the future work.

2 Materials and Methods

Our method combines the use of an ANN both with an image processing algorithm for feature extraction. Such a choice represents an original approach to the gaze tracking problem. There are some reasons to justify the use of an ANN in a feature-based context. Among them, the most important is the ability of an Artificial Neural Network to generalize, once the right input set has been chosen. A bad choice can indeed invalidate the potential of the algorithm. We think that geometric relations between features of the eye could be a more effective input than that constituted by all

the pixels of an image. In fact, a set of parameters such as coordinates of iris and eye corners can be managed and generalized much better than an image. Moreover, the number of inputs to the ANN can be consistently reduced by feature extraction, with a strong benefit to the computational cost and the complexity of the ANN.

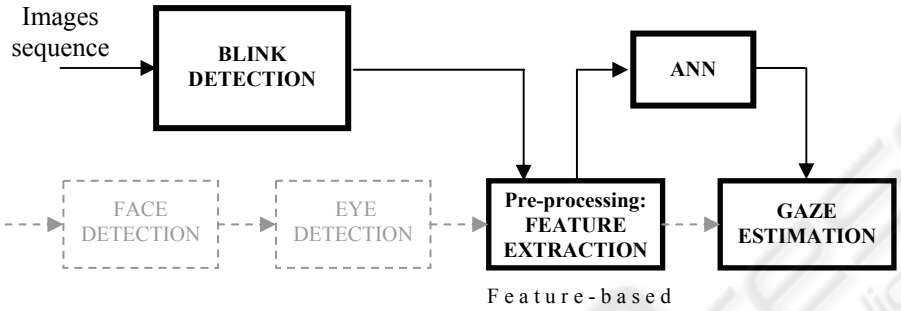

Fig. 2. The proposed method.

Figure 2 shows the scheme of the proposed method. The traditional eye tracking procedure has been modified, by replacing the Face and Eye Detection blocks (dotted line) with a Blink Detection block.

Fig. 3. Blink detection. The first two images represent two consecutive frames of a blink. A difference image is then created. An appropriate filtering enhances the zones where a high changes of intensity occur, i.e. the eye regions.

We consider the blink detection as a valid alternative to the traditional face-eye detection procedures: our approach is based on a dynamic analysis, i.e. consecutive frames of the video sequence are analysed and compared each other. With this approach we can extract information on the moving objects in the scene, such as the eyelids during a blink. Thus, the eye region (where the blink occurs) can be detected from the video sequence without passing through the face detection. The user is requested to blink three times before starting the task (figure 3), without moving the head. Since it is not possible to maintain the head perfectly still, an appropriate filtering is required to eliminate low-frequency movements and enhance the

movements with high velocity, such as the eyelids ones. The algorithm has been inspired by the works of Gorodnichy [4], Grauman et al. [5] and Ohno et al [14].
Following the scheme, after the blink detection, a feature extraction block is inserted. This pre-processing step characterizes the proposed neural-based technique, by handling images to provide the ANN with a good set of input.

2.1 Artificial Neural Network Design

In general terms, the design of an ANN comprises the assessment of the following variables:

1) typology of the net,
2) typology and number of the inputs,
3) typology and number of the outputs,
4) number of hidden layers,
5) number of neurons for each hidden layer.

A multilayer Perceptron with 2 hidden layers, 45 neurons each, has been chosen. This choice is the result of several trials in different conditions and it is strictly correlated to the typology of the input.

To allow the ANN to accomplish the task, an appropriate set of input parameters is needed. These parameters correspond to the geometrical relations between specific features of the eyes: the irises and the eyelids. Such features are detected by a preliminary image processing. Information on the relative position of the iris in the eye socket is given by two distance vectors chosen for each eye, i.e. the distances between the centre of the iris and the corners of the eyelids (see figure 4). Each vector has an amplitude and a phase: the amplitude is important mainly for the horizontal movements of the eyes, while the phase for the vertical shifts. In order to have information on the absolute position, orientation and distance of the head, the 2D coordinates of the external corners of the eyes have been added. Considering both eyes, the total amount of the inputs is 12 (4 magnitudes, 4 phases, 2 couples of x-y coordinates).

As shown in the figure 5, the ANN provides 2 outputs corresponding to x and y coordinates of the observed point in the screen reference system.

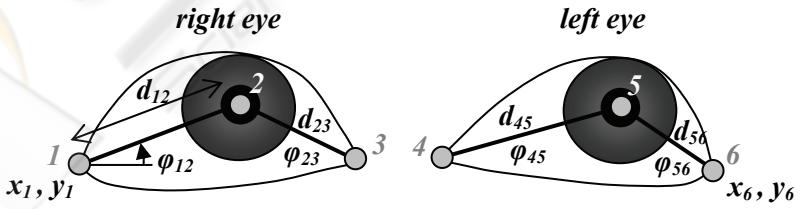

Fig. 4. The geometric features of the eyes (numbered from 1 to 6). Magnitudes and phases of the distance vectors represent 8 of the 12 inputs. The remaining 4 inputs are the x and y coordinates of the external corners of the eyelids.

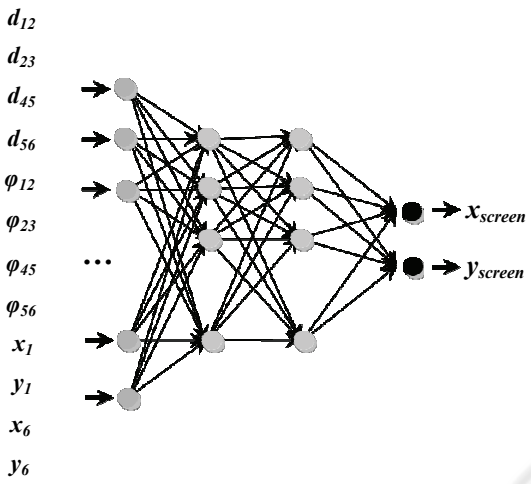

Fig. 5. The Artificial Neural Network, with 12 inputs, 2 hidden layers (45 neurons each) and 2 outputs.

To learn the right mapping between input and output, the ANN needs a preliminary training procedure. In this work a Resilient Back Propagation algorithm (RBP) has been chosen. A set of inputs has been provided together with the desired outputs. To do that, the training set has been obtained as follows: the user is requested to look at a cursor moving on the screen on known positions representing the desired outputs (figure 6). The video stream is then processed in order to extract the geometric parameters, that correspond to the inputs of the ANN. The procedure is similar to the one proposed by Baluja [2], with slight differences in the structure and length of the training set.

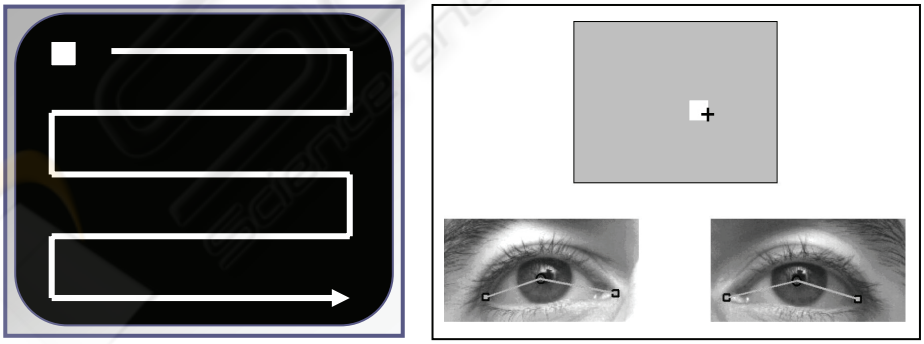

Fig. 6. On the left: cursor moving on the screen. On the right: the white square represents the cursor, while the cross shows the reconstructed position of the point observed by the eyes.

This phase can be seen as a calibration procedure for the proposed REGT, and in this work the burden time has been estimated in less than 5 minutes for the video acquisition and approximately 10 minutes for the training of the ANN. A high number of cursor positions indeed guarantees an accurate mapping, but at the same time requires an extended training procedure. Thus, a careful analysis is needed to optimize the number of training positions.
To compensate for small head movements, several images of the eyes looking at the same point have been processed. Once the training procedure is completed, the ANN is ready to receive a set of parameters from each frame of the video stream, and a gaze direction is then computed. The 2D coordinates of the observed point on the screen can be plotted on the monitor as a qualitative evaluation of the accuracy.
The quantitative aspects related to the accuracy will be presented in the next section.

3 Experimental Tests

In this section the results of the experimental testing are presented. The discussion concerns the training, the test trials and the estimation of the accuracy.
The technical equipment comprises a low cost camera (Trust 150 Spacecam Portable) with a typical spatial resolution (640x480) at a frame rate of 60 fps and a monitor at a resolution of 1024x768, together with a personal computer to manage video streams and to implement algorithms. As shown in figure 7, the distance between the user and the monitor has been set to 400 mm, and the camera has been placed below the monitor in order to reduce the influence of the eyelashes on the iris detection. The size of the monitor is 17'' (Sony Trinitron multiscan 17seII). The cursor moving on the screen has a size of 40x40 pixel, and spans over 99 positions on the screen, with a step of 100 pixels in the horizontal direction and 60 pixels in the vertical direction.

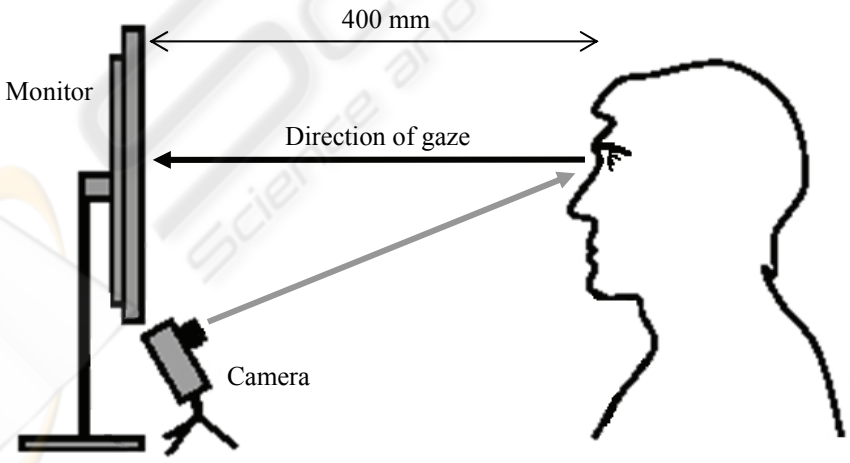

Fig. 7. Experimental setup.

The performance of the whole system has been evaluated in terms of two different quality factors:

1. The robustness with respect to small head movements, i.e. the unconscious movements that occur when the user aims at maintaining the head as still as possible. To evaluate this ability, the ANN has been trained with a set of examples that comprises all the 99 possible positions of the cursor on the screen. While the subject looks at a specific position,, her/his head moves with small fluctuations, giving rise to a change in the eyes configuration. Hence, in the training procedure the ANN needs more configuration of the eyes relative to each position. After several training trials, the optimum number of eyes configurations for each position has been chosen to 17, as figure 8 shows., that correspond to 17x99=1683 total examples.

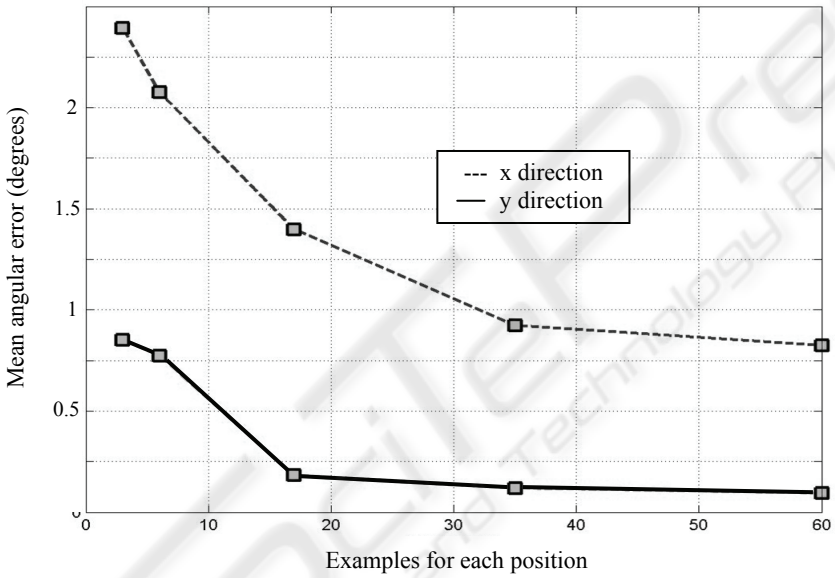

Fig. 8. Training trials. Choice of the optimum number of examples for each position of the cursor. Each example correspond to a pair of input-desired output (input: eyes configuration, output: coordinates of the cursor position).

After about 3000 epochs of training, some trials have been executed in order to evaluate the accuracy. Figure 9 shows the results in 4 different conditions: cursor moving on 24 and 35 positions in ordered sequence and randomly. In such conditions the RMSE estimation gives values less than 1 degree.

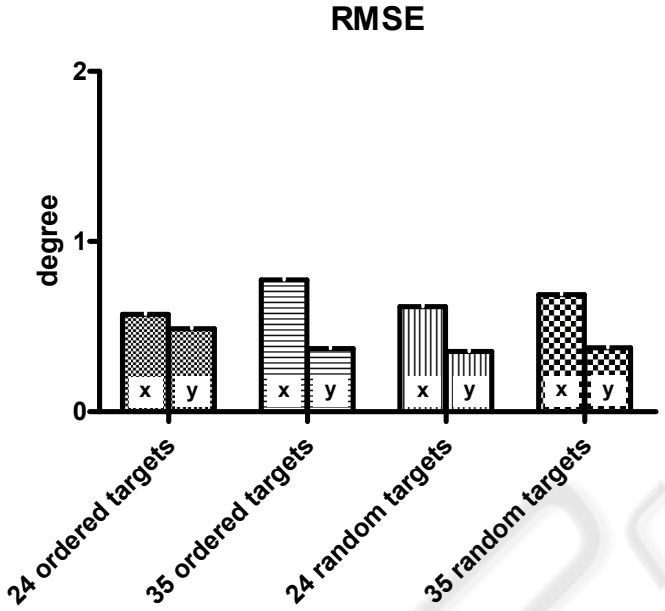

Fig. 9. Root Mean Square angular Error for x and y directions.

2. The second aspect to evaluate is the ability to generalize for position of the cursor (i.e. directions of gaze) never assumed during the training procedure. Half of the positions (50/99) have been used to train the ANN, whilst the remaining half has been utilized to test the performance. First results have shown a RMSE of about 2°. More experimental tests are expected in order to accurately quantify the variables that influence the generalization ability of the net.

4 Discussion and Conclusions

In this work, a novel eye gaze tracker based on artificial neural networks has been proposed and tested. In comparison to other neural-based methods found in literature, the proposed method allows a consistent size reduction of the input set.
Techniques discarding procedures of feature extraction are based on input sets corresponding to the whole set of pixels of the image. For example, a 15x40 pixels image of an eye provides a set of 600 parameters. In the proposed work, only 12 parameters are used for both eyes, with a strong benefit to the complexity of the net.
The results on the accuracy are encouraging, showing that the performance of the method is comparable to the most of the traditional EGT (see Table 1, [1]).

Table 1.

Technique	Accuracy	Comments
Contact lens	1’	Very intrusive
Electro Oculography (EOG)	2°	Intrusive, but simple and low cost
Infrared Oculography (IROG)	2’	Infrared-based, head mounted
Dual Purkinje Images (DPI)	1’	Infrared-based, requires bite bar
Pupil-glint	1°	Infrared-based, tolerate some head-motion
View-based	0.5°-2°	Camera-based, requires training
Proposed ANN technique	1°-2°	Video-based, low cost, requires training.

Furthermore, a good robustness to unconscious movements of the head has been observed. Nevertheless, the main potentials of the system haven’t been wholly explored yet, i.e. the capacity to measure gaze directions in presence of big changes in head position and/or orientation.
Thus, the future work comprises the following topics:

- optimization of the blink detection procedure in order to have a totally automatic eye detection for different persons and in presence of eye glasses;
- generalization for larger head movements by a more extensive training procedure that takes into account user’s position;
- feature extraction without the use of templates. New solutions for the identification of the irises and eyelids will be designed to shorten the computational time. In this context, a preliminary calibration procedure could be helpful to recognize such specific features;
- real-time implementation, to use the system in a realistic interactive environment.

Human vision has the peculiarity to be very effective in analysing the visual information. With a glance, a human is capable of determining the line of sight of another person. With a neural approach we want to mimic the ability of a neural system, such as human being, to accomplish the task of gaze tracking. This represents a deeper motivation that goes beyond the aim of an accurate and robust measurement.

5 Acknowledgments

This research activity has been partially funded by the Italian Ministry of Education, University and Research (MIUR).

References

1. C.H. Morimoto, M.R.M. Mimica, Eye gaze tracking techniques for interactive applications, Computer Vision and Image Understanding 98 (2005) 4–24

2. S. Baluja, D. Pomerleau, Non-intrusive gaze tracking using artificial neural networks, Tech. Rep. CMU-CS-94-102, School of Computer Science, CMU, CMU Pittsburgh, Pennsylvania 15213 (January 1994).
3. A.T. Duchowski, A breadth-first survey of eye tracking applications, Behav. Res. Methods Instrum. Comput. (2002) 1–16.
4. D.O. Gorodnichy, Towards automatic retrieval of blink-based lexicon for persons suffered from brain-stem injury using video cameras, Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'04)
5. Kristen Grauman, Margrit Betke, James Gips, Gary R. Bradski, Communication Via Eye Blinks-Detection and Duration Analysis in Real Time, Proceedings IEEE Conf. on Computer Vision and Pattern Recognition, 2001.
6. T. Hutchinson, K. White Jr., K. Reichert, L. Frey, Human–computer interaction using eye-gaze input, IEEE Trans. Syst. Man Cybernetics 19 (1989) 1527–1533.
7. D.A. Robinson, A method of measuring eye movements using a scleral search coil in a magnetic field, IEEE Trans. Biomed. Eng. 10 (1963) 137–145.
8. A. Kaufman, A. Bandopadhay, B. Shaviv, An eye tracking computer user interface, in: Proc. of the Research Frontier in Virtual Reality Workshop, IEEE Computer Society Press, 1993, pp. 78–84.
9. J. Reulen, j.T. Marcus, D. Koops, F. de Vries, G. Tiesinga, K. Boshuizen, J. Bos, Precise recording of eye movement: the iris technique, part 1, Med. Biol. Eng. Comput. 26 (1) (1988) 20–26.
10. T. Cornsweet, H. Crane, Accurate two-dimensional eye tracker using first and fourth purkinje images, J. Opt. Soc. Am. 63 (8) (1973) 921–928.
11. K. Tan, D. Kriegman, H. Ahuja, Appearance based eye gaze estimation, in: Proc. of the IEEE Workshop on Applications of Computer Vision—WACV02, 2002, pp. 191–195.
12. Y.Ebisawa, Realtime 3D Position Detection of Human Pupil, VECIMS 2004- IEEE Intenational Conference on Vitual Environments, Human-Computer Interfaces, and Measurement Systems
13. Soo Chan Kim, Ki Chang Nam, Won Sang Lee, Deok Won Kim, A new method for accurate and fast measurement of 3D eye movements, Medical Engineering & Physics 28 (2006) 82–89
14. Ohno, Mukava, Kawato, Just blink your eyes: a head-free gaze tracking system, CHI 2003, April 5-10, 2003, Ft. Lauderdale, Florida, USA.