

Helping People with ICT Device Control by Eye Gaze

Fangmin Shi, Alastair Gale and Kevin Purdy

Applied Vision Research Centre, Loughborough University
Loughborough LE11 3TU, UK
{f.shi, a.g.gale, k.j.purdy}@lboro.ac.uk

Abstract. This paper presents a computer method to help people, typically having limited mobility, to be able to operate ICT devices with eye gaze in their living/work environment. The user's eye gaze is recorded and analyzed in real-time. Any ICT device in the environment that is being looked at for a certain time period is identified, located and assumed to be the object of interest that the user wants to utilise. Through a suitable interface, the user can then decide whether to operate the device. By using this state-of-the-art technology, people with impaired mobility, or able bodied people whose movements are restricted can attain a more independent life style.

1 Introduction

Living with independence is fundamentally important to people's quality of life. However, for those who have lost significant mobility due to diseases such as ALS (Amyotrophic Lateral Sclerosis), Cerebral Palsy, spinal cord injury or Multiple Sclerosis, it is very difficult to carry out even a simple operation, for example to efficiently select buttons manually from the menu of a complex remote control.

Research in eye gaze usage can enable people with limited mobility to access technology more efficiently. Communication by Gaze Interaction (COGAIN), a European research Network of Excellence, brings together researchers from both academia and industries to research and develop eye gaze driven assistive technologies with a view to achieving effective communication and environmental control using eye gaze for mobile disabled users [1]. The research 'ART - Attention Responsive Technology' [2] described here is an example of such technology.

2 Attention Responsive Technology

The ART project's aim is to design and develop a computer-based system to locate a user's eye gaze within a three dimensional room environment and use this to identify an ICT device which they may wish to operate by a suitable interface. For example, if a user wants to switch on an electric fan, with an ART system s/he stares at the fan for a certain time period when a simple 'fan ON' option would become available for

them to confirm their intention. Upon such confirmation (e.g. through a touch sensitive screen) the fan will be switched on via wireless communication.

The system currently works through two camera systems mounted on a head band worn by the user. The first camera - the eye camera- faces one of the user's eyes and tracks their eye movements, resulting in eye fixation ('gaze') data in the environment. Another compact camera – the scene camera - monitors the user's the environment in front of the user and detects any controllable devices which the user may look at. By calibrating the two cameras in advance, the position of eye gaze from the eye camera will be mapped to the image coordinate system of the scene camera. If the eye gaze is on any part of a device, the user will be offered an interface which they can then decide whether or not s/he wants to change the current status of that device. Section 3 will describe the technical details of the system. The working flow of the ART system can be seen in Figure 1.

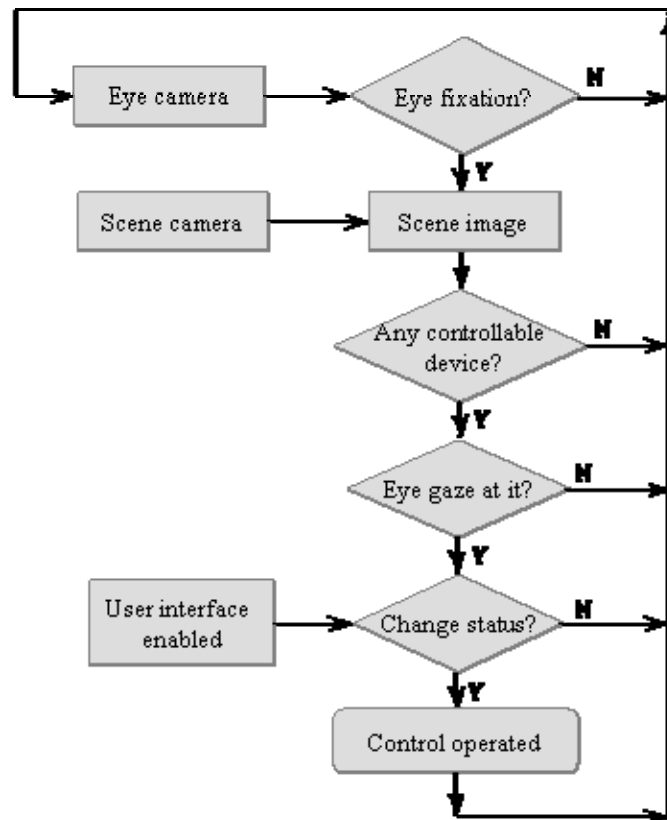


Fig. 1. ART system work flow

Compared with existing eye gaze applications, there are two main distinctive advantages of using the ART system. Firstly, its selection is relative to a real object in a three dimensional space; whereas the selections of most other systems, e.g. the Eyegaze Communication System [3], are with reference to the 2D computer screen

where all devices are represented as buttons or menu items. When the number of devices increases, the layout of the screen may become more complex. The ART approach simply removes any need for a more complex menu. Second is that it eliminates any inadvertent actuation of controllable devices by introducing a simple user configurable interface for confirmation purpose. The user can therefore take the initiative to select operation of the object only when s/he wishes. Conventionally, the interaction with computers via eye movements such as Ward and MacKay [4] has an 'always-on' linking of the user's eye gaze direction to actions, which will include some un-intentional gazes. The ART method selection is therefore more objective.

3 Controlling ICT Devices with Eye Gaze

To implement the tasks as illustrated in Fig.1., a laboratory-based prototype system and its software control interface have been developed. To record a user's saccadic eye movements, a head-mounted ASL 501 eye tracker as shown in Fig.2 is used.

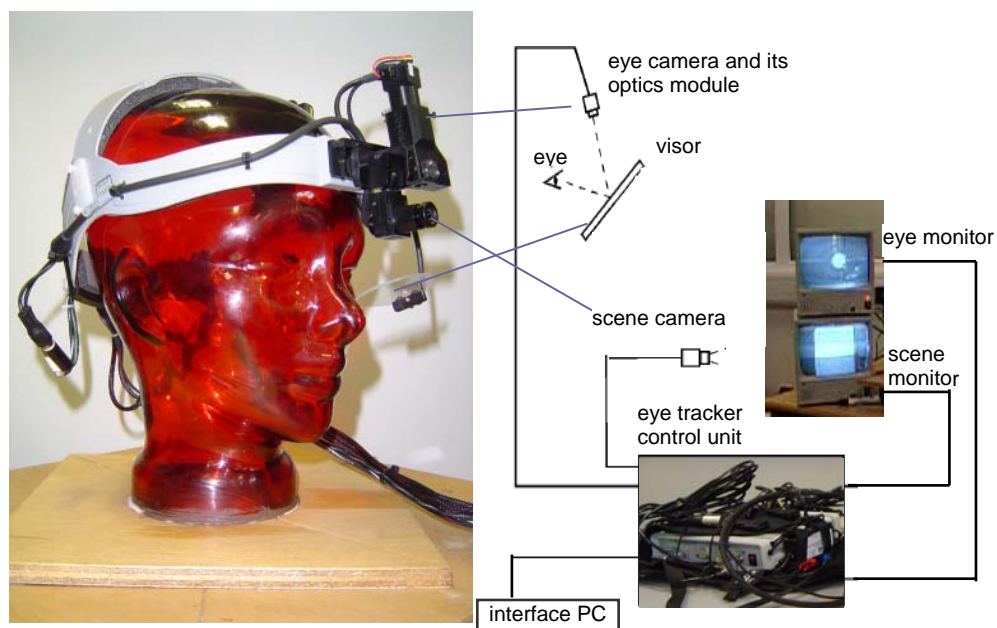


Fig. 2. ASL 501 head mounted eye tracking system

The eye, illuminated by near infra red light, is reflective from the head band visor and its image containing the eye pupil is focused onto the eye camera. The user's pupil size and the eye line of gaze with respect to his/her head, computed by the controller unit through extracting the eye pupil outline and the light reflection in the corneal, are exported as real time serial port stream data to the host PC. There is a second camera mounted on the head band which directly faces the front scene. It

forms a computer vision system, together with the connected optics and frame grabber, which videos the user's front view at a rate of 50 frames per second. Two monitors display the video signal from the eye camera and scene camera respectively.

3.1 Calibration between the User and the Two Image Coordinate Systems

By default, the eye camera records an eye pupil image of size 260x240 pixels. A user's line of gaze is extracted with respect to his/her head. Subject to the chosen adaptor to connect the direct scene camera to the PC, the size of the output scene image can be varied. The current system uses the existing PC graphics card with video input and gives an image size of 760x576 pixels. Through calibration the user's point of gaze is traced within the scene image to reflect where s/he is actually looking.

The performance of the head-mounted eye tracker system depends greatly on initial calibration. The user sits in a comfortable position while having one eye imaged on the eye monitor. A nine point target chart is placed at a suitable distance so that its image nearly fills the scene camera field of view and the user is instructed to look in turn at each of the calibration targets.

The target chart used in this project is shown in Fig.3. Its image is acquired by the scene camera within the 760x576 pixels. Through a set of image processing techniques, i.e. image noise reduction, image dilation and erosion, and edge detection, the image coordinates of the nine points can be obtained. The process and the corresponding images can be seen in Fig.3. The user is instructed to sequentially look at the nine points while the line of gaze to each individual point in the 260 x 240 coordinate system is recorded. Both sets of data are saved as input to the eye tracker controller interface. A calibration matrix is automatically generated linking the eye gaze coordinates in the eye camera system to those in the scene camera system.

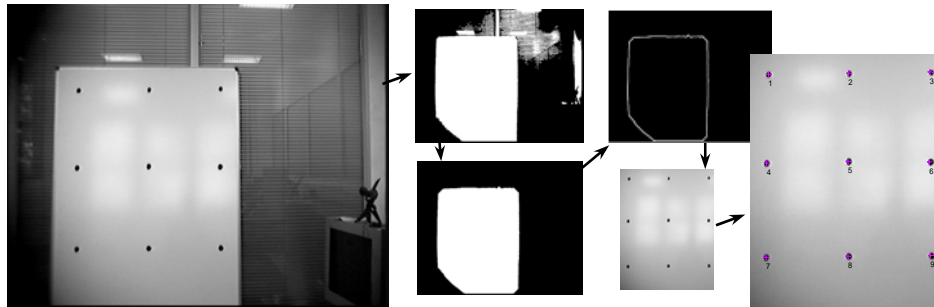


Fig. 3. Obtaining calibration target point coordinates

Typically a user calibration takes some 30 seconds. The success of such a calibration can be examined from the scene monitor where the superimposed cross hair, representing the output from the eye camera, should closely follow the real point of gaze at an object in the three dimensional room environment in real time.

3.2 Eye Data Recording and Eye Fixation Determination

In the real time application, nothing will be done until an eye fixation is determined. Simply as a starting point for development purposes, we utilise a 2s gaze at an object to indicate the user's intention to operate that object.

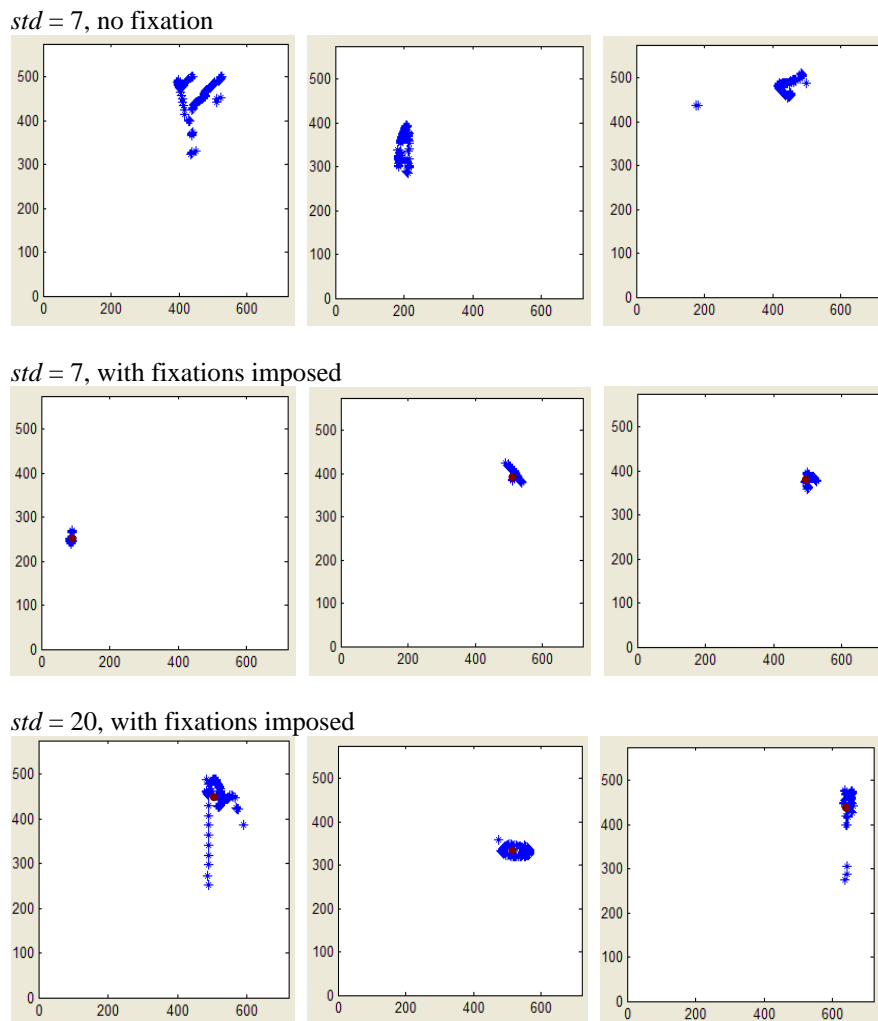


Fig. 4. Eye fixation analysis

Considering the 50Hz video frequency, 100 sets of eye data would be recorded within 2s. The eye pupil size and x, y coordinates of the points of gaze are returned, for example:

<i>pupil size</i>	<i>gaze(x)</i>	<i>gaze(y)</i>
46.0000	717.2308	410.4000
46.0000	717.2308	410.4000
45.0000	714.4615	412.8000

The criteria for determining an eye fixation are set as follows:

- 1) Remove those data where the eye pupil cannot be traced stably, either due to changes of the ambient illumination or the wrong positioning of the eye pupil image. In either case, zero pupil sizes are returned.
- 2) Remove those falling outside the boundary (720x576).
- 3) Check if the number of the remaining points is more than 60% of the original.
- 4) Continue to check if 80% of the left points are within a range of $mean \pm 2 * std$ (standard deviation). If yes, then a fixation is obtained.

Experiments have been made with various *std* and it has been found that a range of *std* from 7 to 20 work well to obtain a fixation from 'well' to 'loosely focused' gazes at an object. This parameter will be made adjustable in order to suit the real end users with different level of concentration abilities. Examples of the recorded eye data are given in Fig.4. It can be seen that the group data with larger *std* can tolerate a more widely spread distribution, (more tolerance to a user's head shaking or eye shifting).

A scene image may or may not contain an object of interest. It may also include more than one controllable object. SIFT (Scale Invariant Feature Transform) feature matching, developed by David Lowe [5], enables a stable approach to object identification and location. It works by saving a set of images of each pre-known object to form the reference image database, as shown in Fig 5. The SIFT feature of each image is calculated, and in real time these are compared with that of the scene image to perform image matching for object recognition. The method is described in [2] and matching constraints are discussed in [6].



Fig. 5. Example images from the reference image database

3.3 Controller Simulation Interface

The prototype system combines all of the features as illustrated in Fig.1. apart from the real action part once an effective eye fixation falling on an object is detected.

Alternatively, the real time system is currently controlled by a graphic user interface for the controller simulation, as shown in Fig. 6.

By pressing the toggle Start/Stop button, the scene image can be displayed continuously in the upper right window. Upon the detection of the user's attention on a controllable device, the found object and the fixation point on it will be highlighted. The name of the objects, which are listed in the left column, will flash in different colours to invite the user's confirmation by choosing either of the radio buttons ON or OFF. Each time the gaze operation is stored in the text box 'history of eye gazes'.

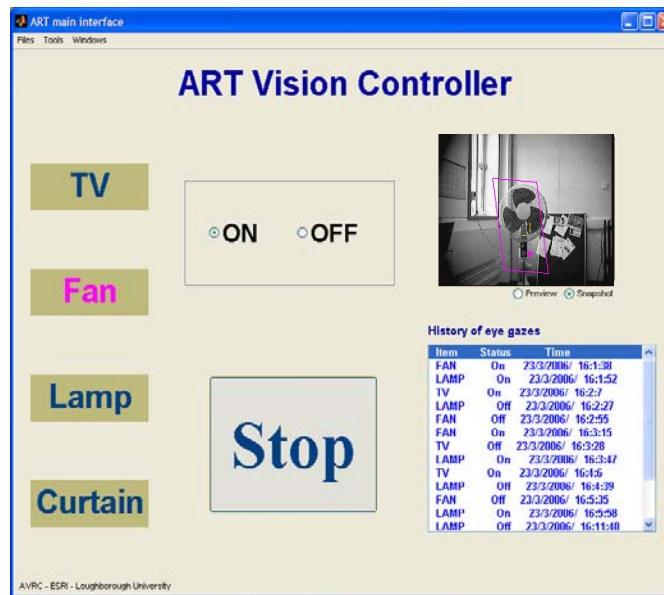


Fig. 6. A GUI to run the prototype system in real time

4 User Group

The initial research work is targeting potential end user groups and determining which objects, and object controls are most relevant to them. Additionally this work will ascertain potential user demands from such a system and the outputs from this research phase will then help shape the subsequent technical solutions formulated.

A questionnaire has been sent out to potential end users - http://www.lboro.ac.uk/research/esri/applied-vision/projects/art/art-survey/art_q.htm. The survey questions are composed of five main sections, including the users' general feeling of the ART concept, their requirement for the speed of using the system, the opinion on the number and the type of controllable devices, where users want to place them and finally about the users themselves. The design of the online survey takes special account of users' disability. Most of the questions are accessible by repeatedly using the Tab key and Y/N keys.

The survey is still ongoing. The initial feedback has been used as a guide for the development of the system. For instance, setting up a living room environment, the selection of controllable devices such as; TV, fan, lamp and electric curtain, and the two second item selection speed are examples reflected by the participants' answers.

4 Discussion and Future Work

The ART system currently runs on a Matlab programmed interface. The system run time can be broken into the following details: 100 eye data - 2~3s; a snap shot - 2~3s; and SIFT matching - 3~5s (40 reference images). In total it takes 7~11s. This initial slow speed can be significantly improved by either programming in C instead of Matlab or reducing the number of eye data points to be recorded.

This paper has presented a solution to the problem of using eye gaze based control systems. It offers a practical way for people with restricted mobility to make use of his/her vision ability to control devices with a view to achieving a more independent lifestyle. A laboratory-based prototype system is available with a head mounted eye tracker. A head free eye tracking system (Smart Eye) in place of the current head mounted one is under development. This will release the user from the requirement of wearing a head mounted eye movement system.

5 Acknowledgements

This research is funded by the ESRC PACCIT Programme. The work is also part of the COGAIN network.

6 References

1. Bates R., Donegan M., Istance H.O., Hansen J.P. and Rih K.-J.: Introducing COGAIN – Communication by Gaze Interaction. In: Clarkson J., Langdon P. and Robinson P. (Eds.): *Designing Accessible Technology*. Springer-Verlag, London (2006) 77-84
2. Shi, F., Gale, A.G. & Purdy, K.J.: Eye-centric ICT control. In: Bust P.D. & McCabe P.T. (Eds.): *Contemporary Ergonomics* (2006)
3. The Eyegaze Communication System:
<http://www.eyegaze.com/2Products/Disability/Disabilitymain.htm>, accessed April 2006
4. Ward D.J. and MacKay D.J.C. (2002) Fast Hands Free Writing by Gaze Direction. *Nature*. August
5. Lowe D.G. (2004) Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*. 60, 2, 91-110
6. Shi, F., Gale, A. and Purdy, K. (2006) Sift Approach Matching Constraints for Real-Time Attention Responsive System. *Proceedings of APIS5 2006, China*, 31-34