

# A New Gaze-Based Interface for Environmental Control

Fangmin Shi, Alastair Gale, and Kevin Purdy

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

**Abstract.** This paper describes a new control system interface which utilises the user's eye gaze to enable severely disabled individuals control electronic devices easily. The system is based upon a novel human computer interface, which facilitates simple control of electronic devices by predicting and responding to the user's possible intentions, based intuitively upon their point of gaze. The interface responds by automatically pre-selecting and offering only those controls appropriate to the specific device that the user looks at, in a simple and accessible manner. It therefore affords the user conscious choice of the appropriate range of control actions required, which may be executed by simple means and without the need to navigate manually through potentially complex control menus to reach them. Two systems using the head-mounted and the remote eye tracker respectively are introduced, compared and evaluated in this paper.

## 1 Introduction

A number of diseases can progressively degenerate people's motor cells in the brain and spinal cord, greatly reducing their mobility. Statistics in 2006 [1, 2] showed that in Europe there were over three million patients suffering from ALS (Amyotrophic Lateral Sclerosis)/MND (Motor Neuron Disease), MS (Multiple Sclerosis), CP (Cerebral Palsy), SCI (Spinal Cord Injury), SMA (Spinal Muscular Atrophy), Rett Syndrome, Stroke, and TBI (Traumatic Brain Injury). All of these can gradually weaken patients' muscles and make them paralyzed, as well as impairing the ability to speak and breathe. In recent years there is evidence to suggest the incidence of these diseases is increasing. This is partly because these are more prevalent in an ageing population [3]. Due to the increasing personal and government cost on caregivers, the need for variable healthcare options attracts lots of interest from policy makers to encourage assistive technologies to be developed. As the intellect, memory or senses of such patients are unlikely to be affected by such diseases [3], a significant interest has also been generated from academics and industry in conducting research and development on assistive technologies to help them cope with their daily lives through the utilisation of their existing body functions, such as eye sight.

One of the useful assistive technologies for disabled and elderly people is the environmental control system. With it people are able to control their house lighting, open/close curtains and respond to the doorbell, for example. An environmental control system is normally designed to operate via a remote control comprising

numerous buttons or menus. Consequently, this is a serious challenge for people with severe mobility impairment to manipulate such devices due to lack of fine finger movement skills. In fact, any development without an appropriate interface to meet the capabilities that a disabled individual has would be of little or no use to them.

Because mobility-restricted people often retain their sight and the ability to control their eye movements, many eye tracking systems have been developed to aid users by means of eye control technologies. These typically utilise a person's eye gaze to form a new human computer control interface so that a user is then able to use their eyes to create documents, browse the internet and even play games for fun. Many applications can be seen on the COGAIN website [1], which is a research network concerned with developing technologies on Communication by Gaze Interaction.

The means of making environmental controls accessible to people with such specialist needs is a large research issue. Within COGAIN, a couple of applications have shown a way to operate environmental controls via selecting on-screen computer icons representing each controllable device. The user, sitting in front of the screen, uses their eye gaze to navigate around the menu. By looking at either the "on" or "off" key (for instance) of the desired menu for a sufficient 'eye dwell' time, the appropriate device control is activated. Such system copes well with those disabled people who can make accurate eye selection. However, the difficulty of such selection methods would rise with increasing numbers of devices and control options, which directly result in increasing the complexity of the screen menu layout. In addition, such a system has a generic disadvantage in that the user randomly gazing at the screen could falsely trigger an action. Such a system also intrinsically requires the user to be positioned before a computer screen; in some cases this may not be desirable or practicable, in others it may compete with other activities the user is performing with the same screen.

In this paper a user-centric control design is presented, which allows an individual user to sit comfortably in, for instance, a wheelchair with an eye tracker monitoring his/her eye movement in real time. The user's direct eye gaze at a controllable device in their three dimensional environment would result in the system responding by offering simple control options specifically tailored to only the device in question. Currently the work employs a screen to present the options to the user but this could easily be of another type, such as an audible menu, as appropriate to the situation. Similarly, control input is currently achieved by eye-gaze, but other methods such as blow/suck tubes could be employed as necessary. Although the control interface is actuated in response to direct eye gaze, the control input is therefore a separate stage, making the complete process an indirect procedure. This method reduces accidental operation of devices and allows for more detailed control options, therefore increasing conscious control and autonomy for the user in controlling their environment. It is not simply a change of state, and the responsive interface can offer much more than a simple confirmation of the control action.

This research is currently being carried out under a research project named ART – Attention Responsive Technology. In the ART project, two types of eye trackers have been employed to track users' saccadic eye movements –a head mounted and a remote system. Their respective system developments have been described in previous papers [4, 5, 6]. The following sections of this paper will briefly introduce the ART architecture and the two types of ART system designs with the focus on a comparison of the two systems.

## 2 Attention Responsive Interface for Environmental Control

In order to achieve environmental control by means of eye gaze, the ART system is designed to have the following components:

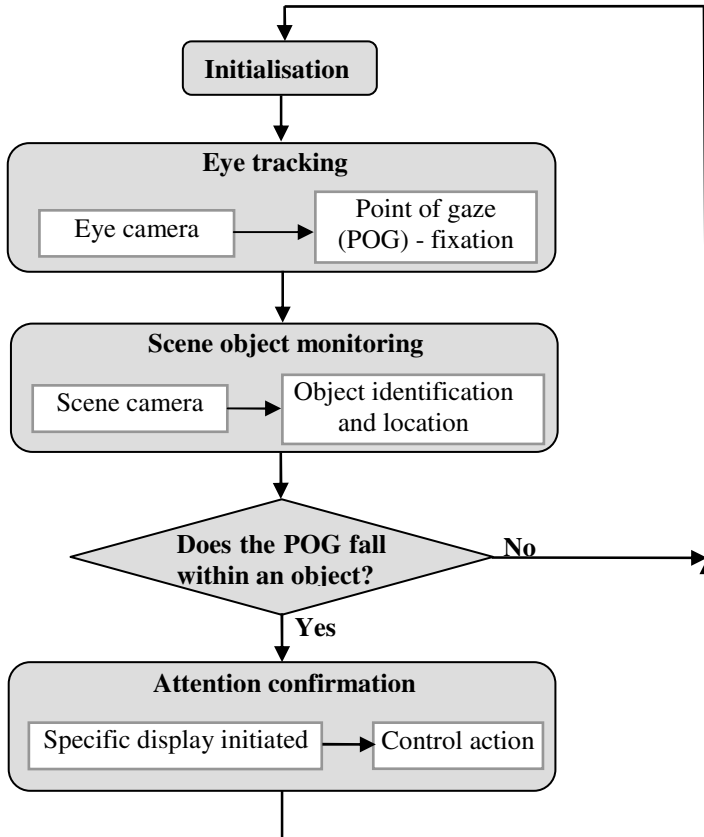
- An eye tracker, which records the user's eye movements in real time and analyses the eye fixations when available.
- A scene object monitor, which is a compact computer vision unit, used to observe the user's environment in terms of his/her field of view. Any pre-known object in their field of view will be automatically identified and located by matching it with a pre-recorded image database of available controllable objects.
- A display to present the control options specific to the identified device. For instance, a simple ON or OFF choice may be popped up for an identified electric fan or a more complex range of options may be presented for controlling a TV. This display may comprise a compact screen, a touch sensitive screen or other technologies such as verbal menus, as most appropriate to the situation.
- A user-configurable means of allowing the user to make one or more control actions. For making a selection relative to a normal display screen menu, this can be again by eye control, by a switch or any other assistive input that is within the disabled individual's capabilities. For a touch sensitive screen, the user must be able to reach and touch the choice with suitable accuracy by some part of his/her body or a control stick or wand attached to them, for example on a head-band if they have suitable neck movement. Alternatively, a user could choose to use his/her eye blinks for indicating a selection.
- A wireless controller, which executes actual device control upon user initiation.

To illustrate how the ART system works, Fig. 1 shows the work flow. The ART system constantly monitors the user's eye movements. Any gaze fixation will trigger a scene image to be analysed, i.e. to identify and locate any controllable object. It will then be determined whether the eye fixation occupies the same location as a controllable object. If it does, then the control options specifically applicable to the identified device will be displayed and appropriate control action will be implemented. For example, when a user wants to open a curtain, what s/he needs to do is to simply look at the curtain. After an analysis to match his/her stable point of gaze at the curtain, two choices of OPEN and CLOSE are offered. By making one of the selections, the user's visual attention to the curtain will become an action as s/he wishes.

## 3 System Development - Common Units

The ART system has been designed with two different types of eye trackers, a head mounted and a remote system. The former affords a user with sufficient head rotation capabilities a wider field of view over which the system can accurately work. The latter remote system has the advantage of causing less physical intrusion and no long term discomfort, as it is not worn. It would also be much easier to set up with the user for each period of use, and so is considered a more acceptable solution for those who

do not need or have a great deal of head movement. These different ways of tracking the eye gaze not only bring differences in system calibration procedures but also cause variations in setting up the scene camera. However, other functions are common in both systems.



**Fig. 1.** ART system work illustration

### 3.1 Object Recognition

The main challenges for implementing the ART system design lie in:

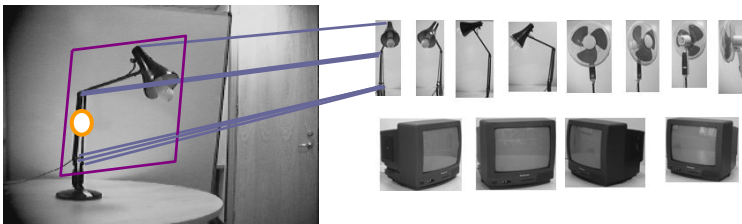
1. How to recognise an object in three dimensional space
2. For a given gaze, how to know its position in the scene image coordinate system
3. Addressing issues of setup and maintenance, such as compatibility with typical environments and developing in-built procedures to add/ delete a new device
4. Building in sufficient capacity for user customisation in order to cater for their individual capabilities and requirements
5. Catering for the wider issues of its general use, over and above its basic operation. For example, considering how the system is set up with a user on a daily basis,

designing for none or less computer-literate carers to run it, maximising its robustness in a real-world situation and considering the consequence of system failure in situations where potentially vulnerable users may rely heavily upon it.

Items 2-4 will be addressed in the later sections of the paper. The first challenge exists in both systems and is one of the most difficult of them all. This is because in the ART environment, household devices such as fans, lamps or curtains, either solid or shape-changeable, could be faced by a user at any distance and any angle of view. This may result in various forms of scene images with different scales, rotations and lighting being presented to the system for interpretation. The conventional object recognition methods such as template matching and 3D model matching are therefore impractical for the ART application.

In this research, feature matching has been extensively investigated and Lowe’s SIFT (Scale Invariant Feature Transform) features [7] were finally considered to be the most suitable method to achieve 3D object recognition by applying 2D image matching. The SIFT algorithm generates a set of key points from an image based on a pyramid of difference-of-Gaussian. Each key point is described by its location, orientation and scale as well as a series of weighted local gradient data. Such characterisation ensures that the SIFT approach is highly invariant to image scaling and rotation, and partially invariant to the change of illumination and the camera viewpoint [7].

In the ART environment all the controllable devices will need to be known to the system in advance. Images of each device are taken from a regularly spaced range of angles of view, which form the reference image database. SIFT features are then calculated for each image and are saved in the key point database. During real time system operation, each time a scene image is captured, image matching will be performed by individually comparing the scene image features with the database features. One scene image may match more than one reference image of the same indexed device. Fig. 2 shows an example of the matching result. The lines join the matched key points in the scene image and one of the reference images. The frame in the scene image outlines the affine transformed shape of the matched reference image.



**Fig. 2.** SIFT feature matching identifies and locates the controllable device

### 3.2 User Interface

When a user’s gaze at a device is detected, then an interface solely for that device is offered and the user can elect to operate the device or not. The physical user interface comprises two parts: an information display conveying the current state of the device

and range of possible control actions, and a control input allowing the user to make the necessary control actions in a way that is appropriate to their abilities. Currently the display provides the necessary information visually and comprises a flat-panel interactive display. The user-configurable control input, currently provides three methods of control input to be utilised: simple switches, a touch sensitive screen and an eye blink to the screen, however many more options are possible ranging from a second stage of eye-control to more discrete transducers such as suck/blow tubes.

### 3.3 Device Control

Actual operation of the selected electronic device is accomplished through using the X10 control system which provides wireless environmental control from the user's interface [8]. The X10 communication protocol has been widely used in the area of home automation. It builds on the existing electrical wiring in the typical household. Each electrical device is plugged into an X10 modular device before it in turn is connected to the mains. Each modular device interprets control signals from a central transmitter.

As a home automation technology X10 has several advantages. It uses the existing wiring infrastructure in the building and so additional wiring is not necessary. X10 modules are cheap, easy to use, can be retro-fitted and are readily available. There are currently no other competing protocols that work as efficiently as X10 as a starting point in simple home automation applications. However, there are well known disadvantages for X10 in terms of its speed and reliability over long wiring runs and different circuits. A neighbouring X10 action may well change another system's X10 device status. As a consequence of its generically applicable nature the system is also relatively limited in the complexity of control operations it can interface with when dealing with more complicated devices having numerous possible states or levels to control. To overcome these, it is easy to upgrade to other home automation control technologies such as Insteon and Z-Wave, although both of these increase the cost of some usage complexities and higher market price.

## 4 ART System with Two Eye Trackers

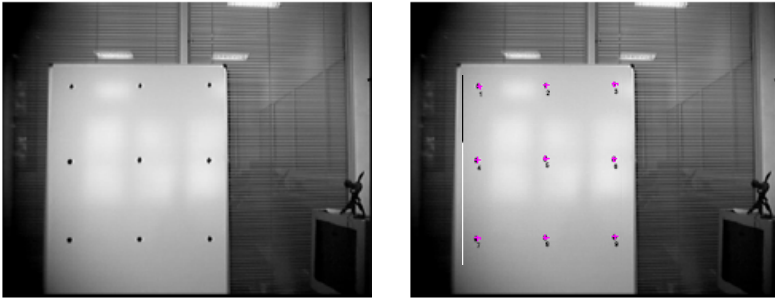
### 4.1 ASL Head-Mounted Eye Tracker

The ART system is firstly designed with the ASL 501, a light weight head-mounted eye tracker. A laboratory-based prototype system and its software control interface have been developed and described in detail [4, 5].

Basically, a user needs to wear a head band, on which is mounted a compact eye camera, monitoring the eye illuminated by near infra-red light, as well as a compact scene camera, facing the user's field of view. The user's pupil size and the eye's line of gaze are computed by extracting the eye's pupil outline, which is highlighted by IR light reflecting back from the retina, and the light reflection from the cornea.

In this configuration, the relative position of the head, the eye camera and the scene camera is constant. A user simply needs to complete a subject calibration to establish this relationship once, unless any of these subsequently moves. In practice this is typically each time the system is positioned afresh on the user's head, as it is difficult

to do so with exact duplication of previous positions. The nature of the ASL eye tracker has made the subject calibration very easy to follow. The procedure needs the user to sit in a comfortable position while wearing the head band and to look at nine calibration points consecutively on a target chart positioned in front of them. The system is nominally accurate to less than 1 degree of visual angle, which equates to an accuracy of approximately 35mm at a distance of 2m. In Fig. 3, using the calibrated parameters to verify the image points in the chart, there are statistically less than 10 pixels of error with an image size of 760x576 and at a distance of about 2m.



**Fig. 3.** ASL calibration with verified image points

Wearing both the eye and scene cameras on the head gives an easy way to set up the ART system with good accuracy, however, there are two main drawbacks with using this system. The first is that ambient lighting can significantly affect the pupil size and bright lights can also impede detection of the necessary infra-red reflections, thus leading to unstable line of gaze calculations. The second is concern about the user's comfort for prolonged periods of wearing the eye tracker on the head. The application of the ART system is expected to be one in which it is used for such extended periods of time by users, as it will facilitate autonomy throughout their daily life, so therefore comfort for the user is of vital importance.

## 4.2 Smarteye Remote Eye Tracker

To truly integrate the system into other assistive technologies, a less physically invasive system would be a distinct advantage. For this purpose, a new ART system using the Smarteye remote eye tracker has been developed [6]. There is a remote version of the ASL eye tracker available, but it was felt that in the ART project's particular application the Smarteye system would offer some advantages. Key amongst these was that it allows for greater head rotation and so an effectively wider total field of view, which could be important to users where head movement is possible and potentially also one of the few physical motor capabilities they may have available for them to utilise.

As the name suggests, the eye tracking with the Smarteye system is carried out remotely while a user sits comfortably in front of one to five small video cameras (eye cameras). There is no attachment at all to the user in this new configuration.

A three-camera Smarteye system is currently employed. The cameras are positioned completely independently from each other but are aimed to cover a range of likely user movements and head positions. The relative camera positions can then be calibrated by an automatic process through holding a calibration chart and making it visible to all three cameras. The Smarteye unit adopts a very different method from ASL to obtain the users' head position and eye line of gaze. Prior to any usage, it needs to generate a user's profile, which is a virtual 3D model of the user's head containing some prominent facial features. The features are marked in a series of snapshots to form the head and eye model and also used as a template in real time tracking of the head and the eyes. The Smarteye output is in three dimensions with reference to a World Coordinate System (WCS), which can be defined during the three-camera calibration process. This feature allows a great deal of user movement. The output can also be projected to a two dimensional plane in the World Coordinate System.

To develop the ART system, a scene camera is additionally integrated. We use a high resolution USB camera. This is positioned alongside the user's head and covers the scene which the user can reasonably view, or that part of the user's scene over which eye position can successfully be tracked, whichever is the smaller.



**Fig. 4.** Smart Eye calibration with verified image points

Like the ASL unit, the relationship among the user, the eye camera and the scene camera needs to be identified in order to relate the eye gaze position or direction to the scene. A similar calibration process to the ASL method is used to achieve this and Fig. 4 shows the calibration image with verified points. It can be seen that the Smarteye results have more error than those obtained from the ASL system. The average gaze accuracy is nearly 2 degrees with a larger error along the vertical axis than that along the horizontal axis. The errors are mainly due to noise in the head model generation; possibly resulting from the limited number of pixels in the eye images.

The absolute gaze error in comparison with the ground truth data with the Smarteye system is larger than with the ASL, however, it has the potential to still meet the ART system requirements well. Further investigation is currently being carried out to improve the Smarteye accuracy results in the ART system implementation.



### 4.3 System Integration and Comparison

Combining the different components described in Section 3 and either of the eye trackers above will form two complete ART systems. The systems work in different ways; however, as a result of the development carried out they will both successfully allow the detection of a purposeful gaze upon a controllable object if applicable. The circled star point imposed in Fig. 2 is an example of such a detected gaze falling within the identified object area. Both advantages and disadvantages of the two ART system approaches are evident as already discussed in the above context. Table 1 gives a comparison of the two ART systems in terms of their performance.

**Table 1.** System performance comparison with two ART systems

	ASL – based system	Smart Eye – based system
Measurement of eye line of gaze	By identifying pupil and corneal reflection and computing their separation	By computing the direction from the eye centre to the eye pupil known from the real time tracking of the head pose and the eye, based on the user's head/eye model
Calibration frequency	Once per session	Once per user
Calibration time	<30s	3 stages: 1. 3 eye camera calibration and definition of WCS: $\approx 10$ s 2. creation of subject's profile: $\approx 10$ minutes 3. system calibration with scene camera: <30s
Requirement for the user during calibration	remaining perfectly still	No special restriction
Average gaze accuracy	<1°	2° or more
Sensitivity to ambient light	Sensitive to bright lights or daylight	Little
Allowable eye movement (horizontally/vertically)	50°/35° with no head movement. With head movement it is only limited by the physical ability to rotate head/body.	170°/60° also with no head movement
Attachment to user	Yes	No

## 5 Discussion and Conclusion

ART is especially designed to help individuals whose movements are severely restricted, or potentially for cases where other forms of control and movement are otherwise undesirable or impractical. ART system users also comprise patient's carers, system installers, maintainers etc. This paper has presented two different eye tracking units for direct detection of a user's attention to real household devices with a view to providing a more usable environmental control system.

The head mounted ASL system is easier to set up and gives more accurate gaze output, however, ambient light would affect it more and the requirement for a head band attachment to the user's head seems not practical for a long term usage. The remote Smarteye system completely releases the user from donning any attachment, but it needs more time to set up a user *initially* and its generic way of computing the eye gaze currently results in a larger error in our particular application (although this may be reduced by further investigation). Both systems also have a number of other components in common, i.e. the object identification and location, the control action and the initiation interface. The system evaluation with both able-bodied people and the disabled ART system potential users is currently being undertaken. To make an assistive device fully usable, apart from technical success, issues such as cost, reliability and human factors are all important, which will comprise further work on the project.

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

## References

1. COGAIN, Communication by Gaze Interaction (2006) <http://www.cogain.org/>, accessed
2. Bates, R., Donegan, M., Istance, H.O., Hansen, J.P., Rih, K.-J.: Introducing COGAIN – Communication by Gaze Interaction. In: Clarkson, J., Langdon, P., Robinson, P. (eds.) *Designing Accessible Technology*, pp. 77–84. Springer, Heidelberg (2006)
3. ASL/MND Alliance International. (Accessed 2007) <http://www.alsmndalliance.org/>
4. Shi, F., Gale, A.G., Purdy, K.J.: Eye-centric ICT control. In: Bust, P.D., McCabe, P.T. (eds.): *Contemporary Ergonomics*, Cambridge, pp. 215–223 (2006)
5. Shi, F., Gale, A.G., Purdy, K.J.: Helping People with ICT Device Control by Eye Gaze. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.): *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, pp. 480–487 (2006)
6. Shi, F., Gale, A.G., Purdy, K.J.: Exploring Eye Responsive Control – from a head mounted to a remote system. In: Bust, P.D., McCabe, P.T. (eds.): *Contemporary Ergonomics*, Nottingham (To appear 2007)
7. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60(2), 91–110 (2004)
8. Gale, A.G., Shi, F., Purdy, K.J.: X10 - are you looking at me. In: de Waard, D. (ed.): *Development in Human Factors in Transportation, Design, and Evaluation: Human Factors in Industrial and Consumer Products and Services conference*, HFES Europe Chapter, Turin, Italy, pp. 285–294 (2006)