Vision responsive technology to assist people with limited mobility

This item was submitted to Loughborough University’s Institutional Repository by the/an author.

Citation: SHI, GALE and PURDY, 2006. Vision responsive technology to assist people with limited mobility. IN: Proceedings of the 3rd Cambridge Workshop on Universal Access and Assistive technology (CWUAAT), 2006

Additional Information:

- This is a conference paper.

Metadata Record: https://dspace.lboro.ac.uk/2134/2235

Please cite the published version.
1.1 Introduction

Automation in the home or office environment plays an increasingly important part in people’s lives. The control of such automated systems can be achieved by means as simple as pressing buttons of a remote control or as complex as using user’s speech. There are some situations where a user’s mobility is limited due to either disability, age or environmental hazards. Provided that the user can move his/her eyes properly, an approach that enables control by vision, i.e. eye selection, can be useful.

There are various commercial eye tracking systems available to record people’s saccadic eye movements, some are designed to cater for interaction with computers via eye movement such as Ward and MacKay (2002). However, the major drawback to such systems is the ‘always-on’ direct linking of the user’s gaze direction to some presumed action, as direct operation by eye selection can generate numerous false alarms. Such control strategy can therefore be unreliable and fatiguing to use over time.

The ART (Attention-Responsive Technology) research project aims to develop a system to enable mobility-limited individuals to access technology efficiently. The system monitors both the individual and the controllable devices such as lamp, TV and curtains in his/her environment. The user’s gaze direction is recorded and used to determine to which device s/he is potentially attending. This information is relayed to a user-configurable control interface, which then displays only those controls that are appropriate. The user can therefore choose to operate the device if s/he wishes. This approach eliminates inadvertent actuation of controllable devices, whilst using direction of gaze to remove any need for a more complex menu.

1.2 ART System

The ART system mainly consists of an eye gaze tracking unit employed to record the user’s eye movements and a computer vision system to identify objects in his/her field of view. By combining the output from these two camera systems, it can be identified whether the user’s eye gaze falls on any object and, if yes, what
and where the object is. Subsequently a user-configurable control interface would respond to the information and allow the user to perform the appropriate operation. Fig 1 shows a photograph of the main part of the whole system - the eye tracking and room monitoring system.

![Fig1. ART eye tracking system](image)

It has been demonstrated that saccadic eye movements can be used as an indicator of a person’s intention when interacting with his/her environment (Vertegaal, 2002). The ART project aims to build upon the natural function of the eyes in selecting environmental objects, rather than using the eyes actually as a control mechanism to operate such object. This way offers a distinct advantage as it helps to overcome the problem of potential false alarms (Gale 2005).

The eye tracking system used in the ART project at the current stage is the ASL Model 501 head mounted eye tracker. The user’s saccadic eye movements are monitored in real time and the eye gaze information is updated 50 times a second.

Being part of the ASL 501, a second mini camera is generically mounted on the helmet and ‘looks’ directly at the user’s field of view, it is therefore referred to as the scene camera. The relative position between the eye camera and the scene camera is stationary no matter how a user moves his/her head as both cameras are mounted on same solid plate connected to the head band.

Camera calibration is made for both eye and scene cameras at the same time. A pattern containing nine points at known positions is used; an image of this is taken by the scene camera whilst the user looks at each point in turn, allowing locations in the scene image to be combined with the eye gaze data. Following such calibration, it will be known where a user is staring relative to the location in the scene image.

The next question that is to be addressed is whether the eye gaze falls on any device of interest. The requirement for such an object identification and location system is that it must be stable in terms of change of viewing angle, scale, rotation and illumination, because the user can be free to move in the environment and the objects might also be moved for any purpose.

The conventional object identification idea is to generate a set of template features and perform a template matching algorithm with a view to finding the right object. However, this is limited by changes such as scale and rotation and it requires significant computational time. The feature matching method can be more
efficient, in particular, when the SIFT (Scale Invariant Feature Transform) method was developed by Lowe (2004), which has found many applications.

The SIFT algorithm provides a reliable and efficient image feature detection method. It identifies key points first that are invariant to scales and rotations by generating a pyramid of difference-of-Gaussian. The key points are then further described by their local area gradients, which allows for significant levels of local shape distortion and change in illumination. Lowe has demonstrated that the SIFT features are highly distinctive. A large number of features generated for each image can densely cover the image over the full range of scales and locations.

In the ART project, a series of reference images of each known object are firstly taken from various viewpoints, which form the original image database. SIFT features are then calculated for each of them, which then comprise the image feature database. When a new scene image is available, matching will be performed by individually comparing the scene image features with the database features. Three or more features are required in order to agree on an object. The affine transform is applied to determine the object’s pose with reference to the scene image. The results which show the combined images from both scene and reference images and their matched keypoints are presented in Fig.2. The reference image shown at the bottom can be found in the scene image with rotation and different scale, as highlighted at the top image. The matched keypoints between the two images are shown by the joined lines.

![Fig. 1. Example results of using SIFT features to identify and locate the object](image)

To improve the reliability of the SIFT approach to object recognition, the known objects in the environment are imaged at regularly spaced angles. A user could look at same objects at various distances. Their images are also taken at a number of distances in order to cover a range of scales.

It has also been found that images with more content complexity, e.g. colour contrast, richness in texts and pictures, generate more SIFT features. Whereas the objects typical in the ART environment such as; lamps, televisions and fans, are legacy items in the market, which are normally in simple black or grey colours and have repeated geometrical features in their appearance. Therefore these have fewer
SIFT features that lead to fewer opportunities to find matches. Their geometrical similarities can even result in repeated matches which are not acceptable. The SIFT algorithm is therefore modified by introducing some constraints when doing the matching and this is discussed in detail in (Shi, 2006).

Once a user’s attention to a specific object is captured, through a confirmation dialogue, the system will perform the control of it. The ART control system is implemented by use of the X10 technology (X10, 2005), a communication standard in smart home development that enables compatible products to be wirelessly controlled over the existing electrical wiring. In the ART environment, all objects are individually plugged into X10 receiver modules, which are directly plugged into the mains. A PC sends out a wireless signal via an X10 interface adaptor to enable the control of X10 connected objects. Most X10 compatible products are very affordable and no costly rewiring is necessary. Installation is simple. This also makes the adding and removing of an object much easier and thus the system can be versatile. The adoption of X10 technology just opens up a way to make the ART system usable one day in a smart home environment.

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.

1.3 Conclusions and Future Work

The ART system is developed as a solution to the problem of using gaze based control systems. It has totally avoided user’s unintentional looking. The user’s point of gaze does not actually control the object, instead it pre-selects objects for subsequent operation.

A head free eye tracking system (Smart Eye) in place of the current head mounted one is the work for the next development stage. This will release the user from the requirement of wearing a head mounted eye movement system.

1.4 References