

Computer Display Control and Interaction Using Eye-Gaze

M. Farid (1, 2), F. Murtagh (1) and J.L. Starck (3)

- (1) School of Computer Science, Queen's University Belfast, Belfast BT7 1NN, Northern Ireland, UK.
Email: f.murtagh@qub.ac.uk Tel: +44 28 9027-4620. Fax: +44 28 9068-3890.
- (2) Dundalk Institute of Technology, Dundalk, Ireland.
- (3) DAPNIA/SEI-SAP, CEA-Saclay, 91191 Gif-sur-Yvette Cedex, France.

Abstract

We describe implementation and initial experimentation with innovative systems for user-computer interaction based on the user's eye-gaze behavior. Two applications are described: (i) user navigation in large images, typical of astronomy or medicine, and (ii) user selection and viewing of multiple video streams. A web environment is used for these applications. System latency is negligible, and system obtrusiveness is small.

Keywords – Eye-gaze tracking, image compression and decompression, streaming video, visual interactive user interface.

1. Introduction

Eye-gaze pointing is just one possibility for vision-based interfaces (Quek, 1995). Gesture recognition is an alternative, active area of research. Istance and Howarth (1994) review user interface interaction based on eye-gaze control. Eye-gaze, when tracked and made use of for control of image displays, may suffer from computational requirements leading to latency, and obtrusiveness of the camera and positioning apparatus. In this work, we completely overcome problems related to latency, and we achieve a relatively very successful solution relating to obtrusiveness, based on the eye tracking environment used.

Goldberg and Kotval (1999) describe the analysis of eye movements for assessment of user interface designs. In work which is of immediate relevance for our objectives, Goldberg and Schryver (1995) describe the (non-realtime) clustering and discrimination of eye-gaze locations, based on multivariate clustering and discriminant analysis. Goldberg and Schryver (1995) sought to determine when the user wanted images zoomed. This is the same goal which we will discuss below in section 3. Gee and Cipolla (1994) describe an unobtrusive system based on use of a normal axis to the face. The latter work achieves 15 degree precision, compared to 1 degree in the case of Goldberg and Schryver. A system for computer control by the motor handicapped, EagleEyes, developed over many years by Gips (see Gips and Olivieri, 1996) is relatively obtrusive also, requiring electro-oculographic sensors to be placed on the subject's face. The ASL 504 system which we use (ASL, 2000) has optics which are remote mounted, 0.5 degree

visual angle accuracy, 0.1 degree visual angle precision, and allows head movement within a 1 cubic foot volume. A measured data record, 16 bits measuring pupil diameter and x- and y-coordinates, can be delivered by a serial port (RS232) connection at 115,000 bits/second, which is many times faster than the sampling rate of the system.

The longer-term goal of the work described here is to develop a new way for the human to interact with multimedia data and information streams. Eye-gaze device control is of interest to the motor handicapped, and to surgeons during medical operations when the hands are occupied. Zooming of a large radiology image, or a large astronomy image, can be the result of concentration of the observer's interest, as expressed by his or her eye-gaze behavior. A novel way to view simultaneous video streams – indeed, TV programs – is made possible, whereby video is delivered only on the basis of the viewer expressing sufficient interest by continuing to look at it. Novel approaches to player interaction with computer games are made feasible also.

Figure 1 shows the set-up. A small infrared camera under the TV set is tracking the user's eye-gaze activity. Just one eye is tracked in this set-up. A few minutes are taken initially to calibrate the particular user and their posture. To the left of the subject in Figure 1, we see from left to right (i) two monitors used for positioning and managing the cross-hairs relating to point of gaze and a graylevel image of what the subject is viewing, (ii) the Control Unit box, and (iii) a control workstation (PC).

The ASL eye-gaze tracking system Control Unit processes the information it receives from all devices connected to it, and calculates eye-gaze position – vertical and horizontal coordinates – pupil diameter and a host of other parameters. This information can be used in real time, receiving it from the RS232 port on the Control Unit which in turn is connected to the serial port of a PC.

The Visual Mouse application which we developed is to demonstrate that we can use the stream of eye-gaze data to control – or replace – the mouse of the PC which the subject is observing. Hence mouse operations on the subject's PC are emulated and are controlled by the subject's visual behaviour.

To show the feasibility of the Visual Mouse, a number of applications were prototyped with the following properties:

- The serial port is used to communicate with the eye-gaze tracking system.
- Information is continuously read from the eye-gaze tracking system.
- The information coming from the eye-gaze tracking system is decoded to get the gaze coordinates in real time.
- The coordinates of the data stream coming from the eye gaze tracker are continuously compared.
- If the subject keeps gazing at a limited area, then a mouse click is made at that position.

2. Implementation

The Visual Mouse application uses a number of parameters or settings to allow us to change the state and behavior of the program. Status of ports is indicated (“stopped”, “in progress”). The current horizontal and vertical coordinates of the point of gaze are shown. The coordinates are in the eye tracking system scale. In these coordinates the screen width is 261 and the screen height is 241. Values out of that range are also possible, including the possibility of negative values. A negative value means that the subject is looking out of the screen. A variable called index is also used. This variable is set to 0 each time the program takes new reference coordinates, i.e. every time the subject looks at a position which is not close to the previous coordinates. If the subject keeps on looking at a region around the reference coordinates, the index variable is incremented by 1 with each new coordinate pair falling in that region.

A “tolerance” variable is implemented as a slide bar going from 0 to 100, allowing user-manager interaction to change it. The user-manager sits, in Figure 1, to the left of the subject, facing the monitors, Control Unit, and PC (and beyond them, the subject). The tolerance sets the width and height of the region around the reference coordinates. The program uses this parameter to decide if the subject is looking at the same place or not. If the difference between the current value and the reference coordinate is less than the tolerance, the program considers that the subject is looking at the same place, does not change the reference coordinates, and increments the index variable. If the difference is more than the tolerance value, the program considers that the subject is looking elsewhere, sets the reference coordinates to the current coordinates, and resets the index variable. The tolerance variable uses the eye tracking system scale.

Associated with the index variable is a slide bar going from 0 to 100, with change possible by the user-manager. This index reference parameter is the value that the index variable must reach to trigger the mouse click. Changing it, we can control the time that the subject must be looking at a region before triggering the mouse click. As the index variable is updated with each new coordinate pair, and coordinate pairs are obtained at 60 Hz, we can easily control the time, in seconds, before the mouse click is triggered (i.e. divide the index reference parameter value by 60).

The Visual Mouse application was developed using Microsoft Visual Studio 6.0 Enterprise Edition. Visual Studio includes several tools. The language we used was C++, using Microsoft Visual C++ with Microsoft Foundation Class (MFC). The most important class used, EyeSerialComm, uses structures ReceiveData and ReceiveStatus to store and output information. The EyeSerialComm class does all the specific work required for the Visual Mouse application. It handles serial port communication, data decoding and mouse message posting. The Control Unit “serial out” data output port can be set to use either a demand mode or a streaming mode. The Visual Mouse applications use the streaming mode.

Two main problems were encountered during implementation. Firstly, there is flickering of eye-gaze coordinates. The coordinates given by the eye-gaze tracking system are averaged over a number of values. By default coordinates are averaged over four previous values, but this can be changed in a configuration file. In spite of this averaging, output coordinate flickering is quite appreciable. To deal with this problem, a good compromise is needed between the tolerance and the index reference parameters. Larger values of the tolerance parameter reduce the flickering effects, but also reduce the resolution of the Visual Mouse. Smaller values of the index reference parameter will generate the mouse click more quickly, which will decrease the effect of flickering. For the streaming video application described below, the tolerance was set to 10 and the index reference was set to 50.

The Visual Mouse works very well when big icon hot links are involved. Dealing with smaller clickable icon links is troublesome. An example of the latter is our attempt to use the Back button on a regular web browser window, in the context of web surfing. The Back button proved too small to deal with.

A second problem encountered related to the accuracy of the calibration procedure. The calibration procedure for each subject and session uses nine points in order to calculate the warping matrix. It is crucial to achieve good calibration of these nine points. Notwithstanding the accuracy with which this is done, there is a decrease of accuracy as we move away from the nine calibration points. A seventeen point calibration is possible, and will give better results, but it requires appreciably more time to carry out. All results reported on below used nine point calibration.

3. Large Image Display in Astronomy and Medicine

Support of the transfer of very large images in a networked (client-server) setting requires compression, prior noise separation, and – preferably – progressive transmission. Progressive transmission consists of visualizing quickly a low resolution image, and subsequently and with greater elapsed time requirements increasing the quality. A simple form of this, using block-based regions of interest, is used in this work. Wavelet transform based methods are very attractive for support of compression and full resolution extraction of regions of interest, because they integrate a multiresolution concept in a natural way. Figure 2 illustrates the design of a system allowing for decompression by resolution scale and by region block. It is the design used in grayscale and color compression algorithms implemented in MR (2001). Systems have been prototyped which allow for decompression at full resolution in a particular block, or at given resolutions in regions around where the user points to with a cursor. Figure 3 exemplifies a standalone system on a portable PC using cultural heritage images (shown here) and digitized land-use maps.

The following two demonstrators were set up on the web.

1. The cultural heritage image example is accessible at <http://strule.cs.qub.ac.uk/zoom.html>. This image is originally a JPEG image (including compression) of size 13 MB, and with decompression it is of size 1 MB. Decompression of a block is carried out in real time. The compression method used, which supports color, is lossy and is based on the widely used biorthogonal 9/7 Daubechies-Antonini wavelet transform.

2. A medical (orthopedics) example is accessible at <http://strule.cs.qub.ac.uk/imed.html>. This image has been compressed from 8.4 MB to 568 kB, and again decompression (and format conversion) of limited area blocks is carried out effectively in real time. The compression method used here is rigorously lossless, and supports grayscale images. The multiresolution transform used is a pyramidal median transform.

Further details on the compression algorithms used can be found in Starck et al. (1996, 1998), Louys et al. (1999a, 1999b), Murtagh et al. (1998, 2001a, 2001b).

The web-based demonstrators can be used directly by the eye-gaze control system. Block sizes are sufficiently large that no precision problems are encountered. Ongoing work is now directed towards two goals: (i) better design of web-based display, through enhancing these demonstrators; and (ii) support for further Visual Mouse interaction modes. Chief among the latter is a “back” or “return” action based on lack of user interest expressed as lack of gaze concentration in a small region.

To present one argument in regard to the interest of new display and interaction environments for large scientific images, we note that various astronomical detectors are now in use or being constructed and which produce images of pixel sizes 16000 x 16000. Examples of facilities producing very large images include CFHT (Canda-France-Hawaii Telescope, Hawaii) and the UK’s Vista telescope to be built at the European Southern Observatory’s facility at Cerro Paranal, Chile. At 16 bits, and indeed 32 bits, per pixel, such images clearly necessitate compression and novel display techniques. A digitized mammogram of typical pixel sizes of 4500 x 4500, at 16 bits per pixel, gives rise to similar needs.

4. Eye-Gaze Control of Multiple Video Streams

Figure 4 exemplifies the new approach to interacting with multiple streams of multimedia data. Shown are a number of presentations from a recent workshop, each with a streaming video record of what was discussed and presented. The observer’s eye dwells just sufficiently long on one of the panels, and the video presentation is displayed in the panel. The observer’s interest remains, and the video continues to display. At any time, the observer’s interest wanders. His/her eye dwells sufficiently long on another panel. The previously playing video is replaced with a name-plate, and the new panel is now where a video stream is playing.

An HTML OBJECT element is used to insert an ActiveX component into the HTML document and all the necessary information to implement and run the object.

A video record of the use of the eye-mouse can be seen at <http://strule.cs.qub.ac.uk/~fmurtagh/eye.html> It lasts 9 minutes 51 seconds. Windows Media 7 is needed to play the video. In succession, one sees the following:

1. Initial set-up.
2. Eye-gaze tracking camera.
3. Loss and regain of tracking.
4. What the subject is viewing.
5. Remote control of the eye-gaze tracking camera.
6. Calibration screen, and calibration procedure, in operation.
7. Attempt at small-size button activation.
8. Multiple video stream example in operation.
9. Orthopedics example in operation.

5. Conclusion

Performance of the systems implemented has been very good. We have shown how latency and obtrusiveness are both well catered for.

Robust use of the eye-tracking equipment in an arbitrary environment is still needed. It will be a while yet before a little camera on your TV or home PC tracks your eye and acts on your behavior, but we are moving towards that objective. A portable installation is desirable, so that the hardware is just the small infrared eye tracking camera, and the software can be hosted by a TV set-top box or handheld PC.

Leading applications of this approach to human-computer interaction are required, and we have discussed a number of proposals in this article. Following Goldberg and Schryver (1995), we have developed a system around the needs for region of interest decompression and display in the context of large image interpretation and analysis in science and medicine. Secondly we have developed a system for the viewing of multiple video streams.

In future work we will enhance these prototypes, and we will investigate how we can proceed towards mobile and robust use of this innovative technology. We also intend to study dependence on display device, e.g. screen resolution, and display transfer function (raster scanned and digitally addressed CRTs, flat panel displays, and so on).

References

1. ASL Applied Science Laboratories, Bedford, MA. <http://www.a-s-l.com> (2000).
2. A.H. Gee and R. Cipolla, "Non-intrusive gaze tracking for human-computer interaction", in *Proc. International Conference on Mechatronics and Machine Vision in Practice*, Toowoomba, Australia, IEEE Computer Society, 112-117 (1994).
3. J. Gips and P. Olivieri, "EagleEyes: an eye control system for persons with disabilities", presentation at *Eleventh International Conference on Technology and Persons with Disabilities*, Los Angeles (1996). EagleEyes homepage <http://www.cs.bc.edu/~eagleeyes>
4. J.H. Goldberg and J.C. Schryver, "Eye-gaze contingent control of the computer interface: methodology and example for zoom detection", *Behavior Research Methods, Instruments and Computers* **27**, 338-350 (1995).
5. J.H. Goldberg and X.P. Kotval, "Computer interface evaluation using eye movements: methods and constructs", *International Journal of Industrial Ergonomics* **24**, 631-645 (1999).
6. H. Istance and P. Howarth, "Keeping an eye on your interface: the potential for eye-gaze control of graphical user interfaces", *Proceedings of HCI'94* (1994). Paper at <http://www.cms.dmu.ac.uk/~hoi/hci94/hci94.html>
7. M. Louys, J.L. Starck, S. Mei, F. Bonnarel and F. Murtagh, "Astronomical image compression", *Astronomy and Astrophysics Supplement* **136**, 579-590 (1999a).
8. M. Louys, J.L. Starck and F. Murtagh, "Lossless compression of astronomical images", *Irish Astronomical Journal* **26**, 119-122 (1999b).
9. F. Murtagh, J.L. Starck and M. Louys, "Very high quality image compression based on noise modeling", *International Journal of Imaging Systems and Technology* **9**, 38-45 (1998).
10. F. Murtagh, M. Louys, J.L. Starck, F. Bonnarel and M. Farid, "On-demand delivery of large compressed images in astronomy: computational requirements", *SPIE Proceedings*, Vol. 4477, SPIE Symposium, San Diego, July/August 2001 (2001a).
11. F. Murtagh, M. Louys, J.L. Starck, F. Bonnarel and M. Farid, "Compression of grayscale scientific and medical images – principles, environments, evaluation", *Electronics & Communication Engineering Journal*, submitted (2001b).

12. MR, Multiresolution Software Environment, <http://www.multiresolution.com> (2001).
13. F. Quek, “Non-verbal vision-based interfaces”, keynote talk, *IWHIT’95 – International Workshop on Human Interface Technology ’95*, Aizuwakamatsu, Fukushima, Japan. Paper at <http://vislab.cs.wright.edu/Publications/nonverbal.html> (1995).
14. J.L. Starck, F. Murtagh, B. Pirenne and M. Albrecht, “Astronomical image compression based on noise suppression”, *Publications of the Astronomical Society of the Pacific* **108**, 446-459 (1996).
15. J.L. Starck, F. Murtagh and A. Bijaoui, *Image Processing and Data Analysis: The Multiscale Approach*, Cambridge University Press (1998).
16. F. Tinto Garcia-Moreno, Eye Gaze Tracking System Visual Mouse Application Development, Report, Ecole Nationale Supérieure de Physique de Strasbourg (ENSPS) and School of Computer Science, Queen’s University Belfast, 77 pp. (August 2001).



Figure 1. The subject views the monitor (here, a TV monitor). Just placed in front of – at the bottom of – the monitor is the eye-gaze tracking camera. Arrayed from left to right behind and to the left of the subject: two monitors, one on top of the other, forming part of the eye-gaze tracking system; the eye-gaze tracking system Control Unit; and a PC which hosts the control software.

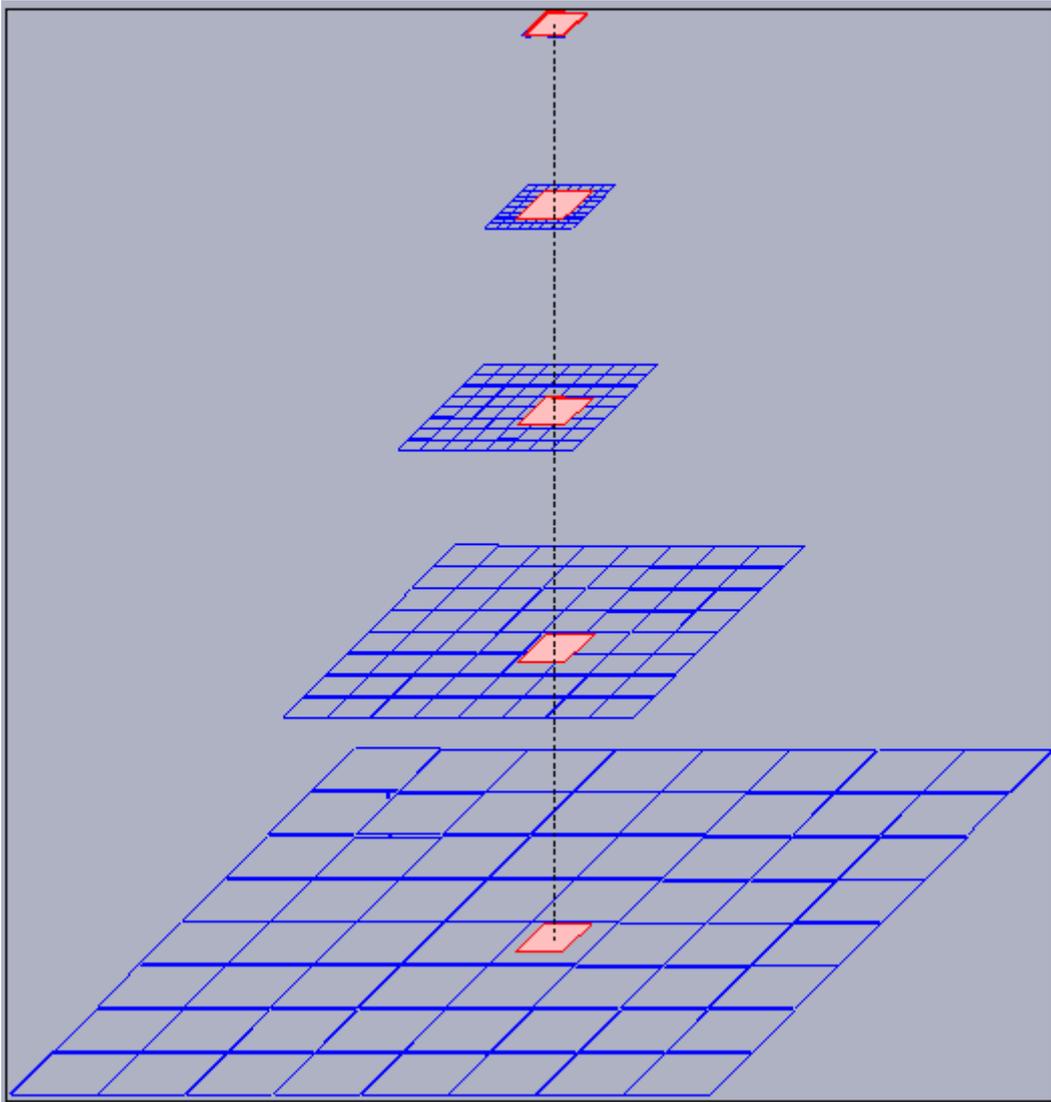


Figure 2: Design for compression of a large image, by block, and supporting five resolution levels. At each resolution level, a display window is superimposed at a given position. At low resolution, the window covers the whole image. Wavelet and other multiresolution transforms use such a pyramidal data structure.

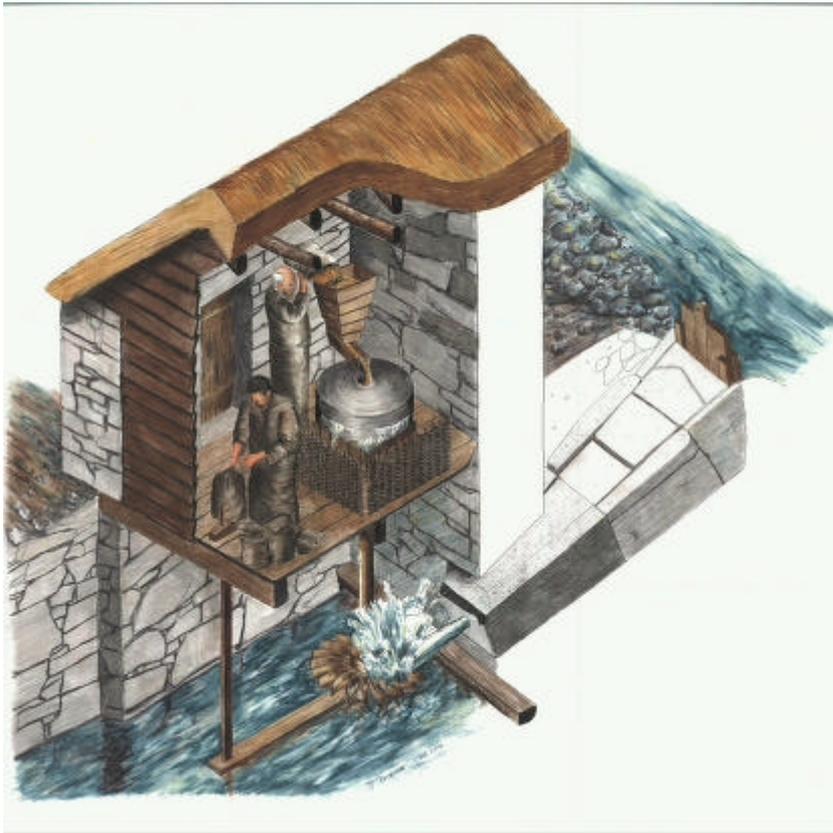


Figure 3a. A digitized image shown at full size and reduced resolution.



Figure 3b. A block from Fig. 3a shown at full resolution. This block is extracted and decompressed from the compressed image file. A mouse click on the low resolution image controls this retrieval and decompression.

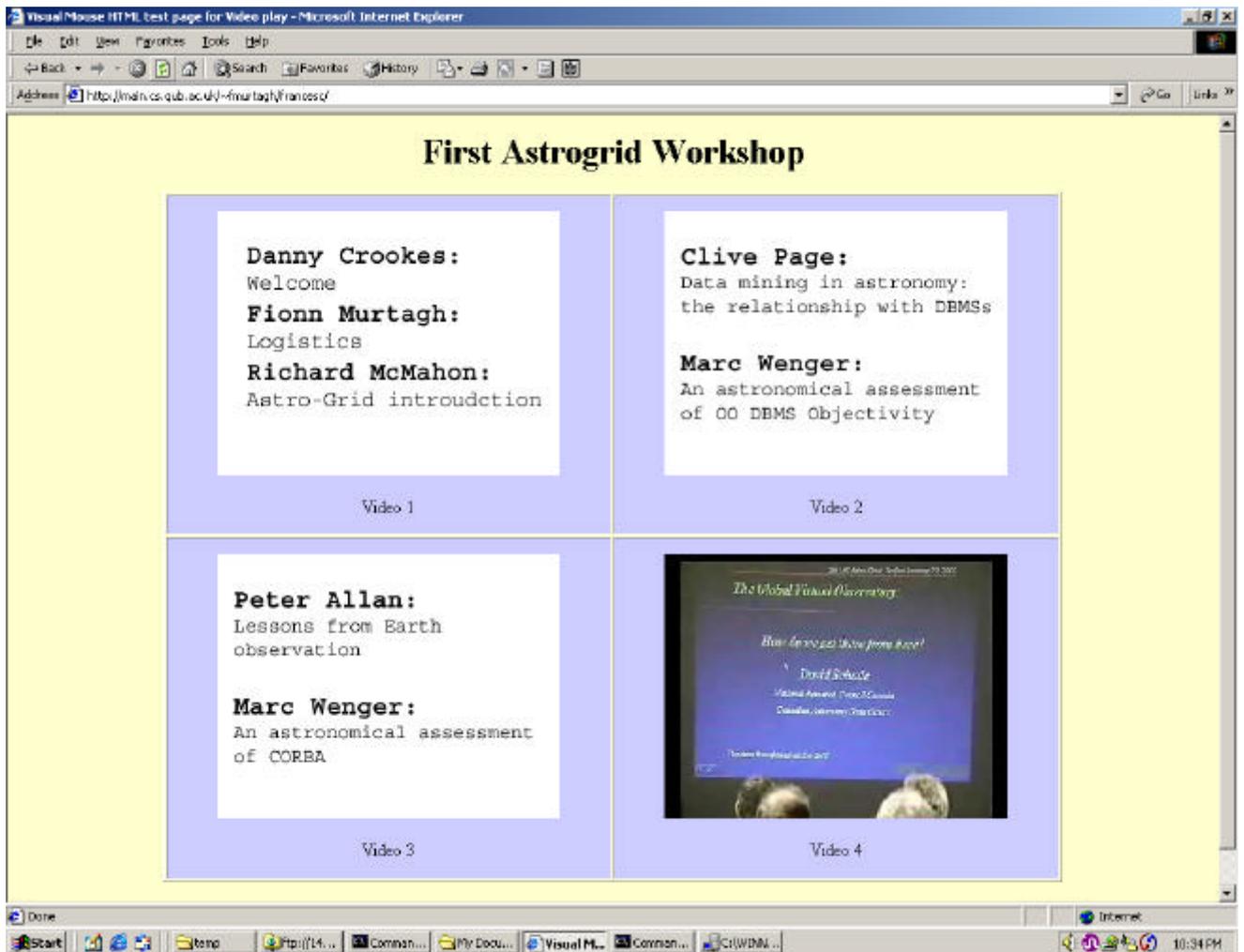


Figure 4: Eye-gaze control of a streaming video record of a workshop held in Computer Science, Queen’s University Belfast, on 29-30 January 2001. User interest, demonstrated by eye-gaze of sufficient duration on a description panel, causes the streaming video to be activated. If the user continues to watch that panel, the streaming video continues. If the user switches his/her attention to another panel, the previous one is turned off, and the panel of interest starts to show video.