The Exploration of Large Image Spaces by Gaze Control

Nick Adams*, Mark Witkowski and Robert Spence
Department of Electrical and Electronic Engineering
Imperial College London
Exhibition Road, London SW7 2BT, U.K.
*n8vision@gmail.com, m.witkowski@imperial.ac.uk, r.spence@imperial.ac.uk

Keywords
Image space navigation, gaze control, Stare-to-Zoom, Head-to-Zoom

INTRODUCTION

We describe methods for navigating large variable resolution images primarily using eye gaze control. In many professions a user's task involves inspecting very large image spaces at a level of granularity ranging from an overview mode to study of fine detail. Such images are becoming commonplace, for instance: maps (e.g. Google Maps), earth imaging (Google Earth, NASA World Wind[1]), surveillance footage, architectural plans, astronomical images and medical images. Traditionally the navigation – i.e. pan and zoom – of such images has been achieved by well-established means of interaction such as mouse control. We explore the use of eye-gaze – possibly in conjunction with other forms of interaction – to control the actions of panning and zooming in the context of navigating or exploring very large images.

Our exploration of these methods uses Google Earth satellite and aerial imagery to investigate how users can be enabled to traverse a complete virtual image from the broadest to finest levels of detail available. In choosing Google Earth for this purpose we note that this represents a very large image indeed[2], which is made readily and freely available on demand. The data is convenient, as it is both familiar and may be intuitively navigated by anyone with a rudimentary knowledge of geography. We investigate the comparative properties and advantages of two related methods of gaze control – Stare-to-Zoom (STZ) and Head-to-Zoom (HTZ) – and offer some preliminary findings and thereby give some pointers to designers who may wish to adopt these methods. Gaze control has been established for both disabled and able-bodied users for data input (e.g. Majaranta and Räihä, 2002), display inspection (e.g. Starker and Bolt, 1990) and spatial navigation (e.g. Bates and Istance, 2005).

Figure 1
Experimental Setup

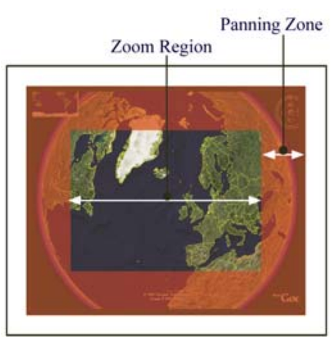

Figure 2
Pan and Zoom Regions

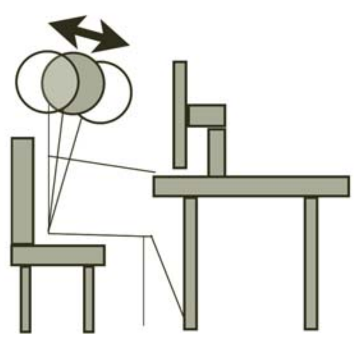

Figure 3
Head-to-Zoom Mode

[1] http://earth.google.com/; http://worldwind.arc.nasa.gov/

[2] The potential size of this "image" is staggering. Were the earth to be imaged at 1m^2 over its entire surface, the resulting image would have the equivalent of ~5x10^{14} pixels. Of course, civilian imaging does not offer this resolution yet, but the data set is still very impressive.

PAN AND ZOOM BY GAZE CONTROL

We explored two methods of pan and zoom control effected primarily by eye movements. These modes Stare-to-Zoom and Head-to-Zoom are described in this section. Both rely on the determination of a point-of-gaze on a computer screen by appropriate technology. Figure 1 shows an overview of the equipment used.

Stare-to-Zoom

This method represents a single mode user interface. All control of the image traversal uses gaze position and timing alone. Figure 2 illustrates the general strategy. The screen is divided into a central zoom region, surrounded by a pan zone. The width of the pan zone (100 pixels top and bottom, 150 pixels left and right, on a 1024x768 screen) has been established empirically. It allows the user sufficient screen space to achieve uninterrupted panning. The panning rate used (~90 pixels/sec) allows some limited visual search within the outer panning region without causing zooming. Clearly, a faster effective panning rate (i.e. across the image space) can be achieved by zooming out to a lower resolution prior to panning. No zooming takes place while gaze is in the pan zone.

Sustained gaze in the zoom central region causes the image to zoom inwards. Normal saccades and fixations in the zoom region do not cause zooming and the image may be inspected in the usual way. Extended stationary gaze (>420 ms) initiates zooming at a comfortable rate. Zooming continues while the point of gaze remains stationary, as determined by a running calculation of the standard deviation of screen position. For non-central regions zooming is accompanied by panning towards the screen centre. This is inherent in the Google Earth interface. Once the identified feature is at the centre of the screen, zooming is uninterrupted until the maximum resolution is attained, while gaze is sustained on that feature. Zooming outwards is achieved by glancing directly at the camera fixed to the base of the screen (figure 1).

Head-to-Zoom

HTZ mode modifies the STZ mode just described by controlling zoom direction and rate by small movements of the head. The user zooms into the image by moving the head (or leaning) forward slightly and out by moving the head away from the screen by a small amount (~±40 mm), Figure 3. This mode allows the user to inspect any part of the image closely without initiating zoom, however the range over which the head may be moved is restricted (by the equipment properties) and the consequences of this are discussed later.

The position of the cursor remains visible in both modes. It is filtered to give the user the appearance of being centred at the point of gaze. That is, saccadic movements are preserved, but any eye movement “jitter” is suppressed during fixations.

EQUIPMENT ISSUES

The system design and investigations described here used LC Technologies (www.eyegaze.com) eyegaze position monitoring equipment. Gaze position on screen is determined by comparison of corneal and retinal reflection from an axially mounted infra-red source on the eye-imaging camera mounted beneath the screen (figure 1). The system requires a brief calibration procedure prior to use by each new user. Accuracy is quoted as 1º (about 15 pixels), readings are made 60 times a second.

Eyegaze software (supplied) and Google Earth run on a single computer. Control of Google Earth is achieved by a combination of the Google Earth COM API and emulation of mouse clicks; direct interaction through the API having been found to be too slow for this type of real time application. The effective field of view of the camera relative to eye position is a volume of 100 mm^3. If the eye position leaves this volume tracking is lost, leading to erratic zooming behaviour in HTZ mode. An eye “icon” can be displayed on screen to assist the user with their head-positioning relative to the camera, although the system, by and large, provides its own feedback in terms of pan and zoom.

PROCEDURE

We have conducted exploratory pilot investigations in a relatively informal manner. The main aims were to gain an insight into the relative merits of the two strategies at this prototype stage and to discover potential improvements for each option through user exposure prior to an extended study under controlled conditions. Seven volunteer participants were each asked to find the University site close to central London twice from a completely "zoomed out earth" manually using the (normal) mouse based interface. This was in order that the test was not influenced by the participant's ability to find the location. Then, using timed runs, participants were asked to zoom into the University site using the mouse, STZ and HTZ modes. Separately, participants were asked to give feedback on the experience, how the two methods compared and to comment on which had the better potential as a method of gaze controlled image inspection. In addition the participants were asked to rate the system on three factors, on a scale of 0 – 10: 1) How they rated their control of the system, 2) How immersed in the system they felt, and 3) How they rated their enjoyment of the system.

PRELIMINARY RESULTS

Average time to complete the zoom task were 26 seconds for STZ and 32 for HTZ on the first attempt. Unsurprisingly, practice improved performance. As a control indication, using the mouse "normally" took 17 seconds. Using the scale described, the participants rated the two systems as follows:

Average rating (0 – 10, best)	HTZ	STZ
Control	6.3	7.3
Enjoyment	6.16	7.25
Immersiveness	6.6	7

While offering no statistical significance, the users consistently rated the STZ mode more highly than HTZ. However, users tended to prefer one method over the other quite strongly, and opinion was mixed. Those participants who valued HTZ over STZ gave the following reasons: a) It was more enjoyable, b) It gave more control because they could choose when to zoom, and c) It was more responsive as there was no gap between deciding to zoom and actually zooming in. Those participants who valued STZ over HTZ cited the following reasons: a) It was more predictable, b) It gave more control because they did not have to worry about their head making erratic and unpredictable motions, c) It required less coordination and cognitive load to operate.

We noted that there was a distinct tendency for participants to drift out of the field of view of the eyegaze camera. This was a particular problem with HTZ, as users often moved their head too close or pulled to far back, resulting in suspension of screen movement while they corrected their position. This detracted from the smooth operation of the system and lead to frustration. We believe that this is partly due to equipment limitations, which might be overcome using alternative technologies.

DISCUSSION

These exploratory investigations have confirmed the value of eyegaze control in this navigation task and provided us with the confidence to proceed with a full-scale experimental investigation, which is now completed and is to be reported separately later. Several refinements of the techniques were discovered and improvements implemented as a result of this study. We have also added a further method (Dual-to-Zoom), combining gaze position input with manual zooming. Such techniques will no doubt find application for the disabled. We also believe that gaze control will be valuable as an auxiliary input mode for interface designers.

REFERENCES

Bates, R. and Istance, H. (2005) Fly Where You Look: Enhancing Gaze Based Interaction in 3D Environments, Proc. COGAIN-05, pp. 30-32.

Majaranta, P. and Räihä, K.J. (2002) Twenty Years of Eye Typing: Systems and Design Issues, Proc. ETRA-02, pp. 15-22.

Starker, I. and Bolt, R.A. (1990) A Gaze Responsive Self-Disclosing Display, Proc. CHI-90, pp. 3-9.