

Using Face Position for Low Cost Input, Long Range and Oculomotor Impaired Users

G. Julian Lepinski, Roel Vertegaal

Human Media Lab

Queen's University

Kingston, Canada

{2gjl, roel}@qlink.queensu.ca

Keywords

Face tracking, long range, low cost, oculomotor impairment

Introduction

While traditional gaze-tracking techniques aim to provide the user with an input method of comparable resolution to classic hand-based input, lower-resolution input may often suffice for a number of computing tasks. Low-res input also provides benefits with respect to cost and ease of use for many users, including users with oculomotor impairments such as Parkinson's, as well as applications for very large displays.

Overview

As discussed in (Card et al., 1991), the neck and head are significantly slower than the eyes, fingers and wrists in terms of input data rate. Many computer-based tasks, such as typing, strongly depend on high data rates and alternative input methods often seek to meet or approach these rates. There exist, however, a number of computer-based tasks which do not require such a high-rate input. There are a number of motivations for using low-resolution input when available, including:

- **Cost** – head-tracking only requires an inexpensive webcam and off-the-shelf computer vision software.
- **Hands-free use** – head-tracking allows the user to provide gross control while using the hands for fine control or other tasks.
- **Eyes-free use** – eye-tracking often impedes a user's ability to scan and peruse screen real estate as they must be conscious of the input they are providing with their gaze.
- **Disabilities** – head-tracking, which is low-resolution by nature, allows users who may have difficulty with fine-motor skills, such as users with Parkinson's disease, to execute tasks they couldn't otherwise complete. A number of other diseases and injuries specifically affect control of eye movement such as those discussed in (Ciuffreda et al., 2007).

We have identified a number of software tasks which, while usually controlled with the keyboard and mouse, often use only a few common actions that do not require pixel-perfect input. These task categories include:

- File and web navigation
- Navigation in virtual worlds/cartography software
- Control of audio/video software



Figure 1. User controlling Google Earth with a standard webcam.

We chose the second category, Navigation, as our exemplary task, and have constructed a controller for Google Earth (GE), shown in use in Figure 1. Our controller takes the metaphor of user head movements in three dimensional space around a fixed, physical globe as a source of inspiration for input gestures. This mapping allows for almost instantaneous learning of controls, even for users unfamiliar with the system. As shown in Figure 2, moving to the left outside of the central zone causes a right rotation in GE, while moving to the right causes a left rotation. Moving upwards or downwards respectively causes the earth to rotate down or up, as if looking over or under the globe. In addition to the location of the face, the software also uses the size of the bounding sphere around the face as an indication of proximity to the camera. This data is used to trigger zoom thresholds, where moving towards the camera triggers a zoom-in, while moving away zooms out.

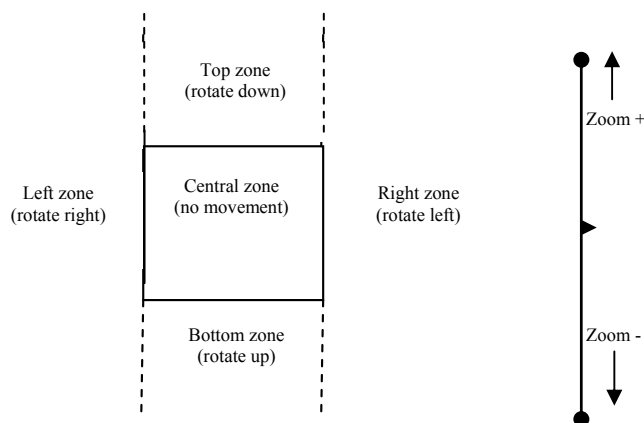


Figure 2. Movement zones and proximity (face size) threshold.

Our hardware consists of a low-cost 640x480 resolution Apple iSight webcam with a frame rate of up to 30fps. Apple has shipped all models of their personal computers with webcams built in above the screens since 2005, and it is reasonable to expect that webcam penetration into the consumer market may soon match that of the personal computer.

Our software uses a standard Haar filter included in the Open Computer Vision Library (<http://sourceforge.net/projects/opencvlibrary/>) to provide face coordinates and size data to a second Google Earth control module (Figure 3). The control module automatically calibrates to the user's starting position and assigns gesture zones to regions outside of the user's initial calibration zone. The user is able to see their position within the camera's view and with respect to their original calibration zone by using an onscreen display. Providing this feedback has shown in our tests to allow the user much more intuitive

control. The software control module will be made available for download at the Human Media Lab website: <http://hml.queensu.ca/eyeinthesky>. Users with webcams should be able to install Google Earth and the control software and be up and running within minutes.

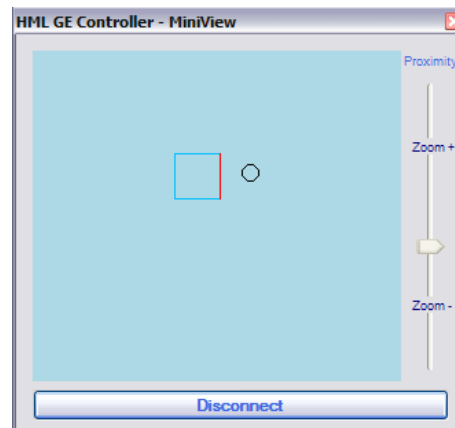


Figure 3. The GE Controller software giving the user feedback on position relative to calibrated centre.

Initial Evaluations

Informal tests indicate that users generally found the interaction techniques easy to learn, as well as use. Users accomplished various tasks, such as navigating to a specific location on the GE globe, consistently and with few errors. We further found that users were able to execute their tasks using the face-tracking controller with their hands occupied, as well as with their eyes looking elsewhere.

The face tracker is able to achieve sufficient accuracy to be used both at seated as well as standing distances, between 0.5m and 3m. The upper limit is likely a limitation imposed by the camera resolution – at higher resolutions we expect the distance to further increase. In seated operation mode the software is capable of tracking small changes in position, allowing the user to control the application without requiring gross movements.

Future Research

While we have applied our low-resolution control methods to one specific application in this paper, we see a number of tasks which may be sufficiently executed without requiring high-resolution input from the user. In the future, we would like to explore a generalized framework that would allow developers and researchers to map any number of low-resolution controlled tasks to a suitable input device; providing real, workable input applications to users and the research community.

Acknowledgements

We gratefully acknowledge the contributions of Adam Archer, Rob Retchless and Mike Rooke.

References

- Card, S. K., Mackinlay, J. D., and Robertson, G. G. A morphological analysis of the design space of input devices. *ACM Transactions on Information Systems*, 9, 2 (1991), 99-122.
- Ciuffreda, K.J., Kapoor, N., Rutner, D., Suchoff, I.B., Han, M.E., Craig, S. Occurrence of oculomotor dysfunctions in acquired brain injury: a retrospective analysis. *Optometry* (St. Louis, Mo.). 78(4):155-61, 2007 Apr.