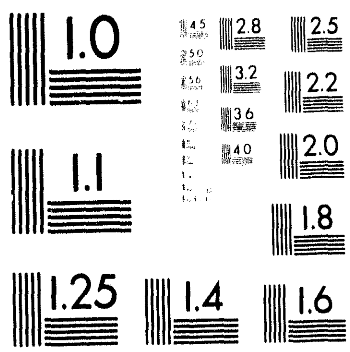
1.0
4.5
5.0
5.6
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

Conf-9310100--3

Engineering Physics and Mathematics Division

EYE GAZE CONTROL OF THE COMPUTER INTERFACE: DISCRIMINATION OF ZOOM INTENT*

J. H. Goldberg
Department of Industrial Engineering
The Pennsylvania State University
207 Hammond Bldg.
University Park, PA 16802
(814) 863-2370

J. C. Schryver
Human Factors and Cognitive Systems Group
Oak Ridge National Laboratory
P. O. Box 2008
Oak Ridge, TN 37831-6360

To be Presented at the 37th Annual Meeting of the Human Factors and Ergonomics Society in Seattle, Washington, October 11-15, 1993.

---

* Research sponsored by the Office of Technology Support Programs, U.S. Department of Energy, under Contract No. DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc., and U. S. DOE and Oak Ridge Associated Universities, under Contract No. DE-AC05-76OR00033.

EYE-GAZE CONTROL OF THE COMPUTER INTERFACE: DISCRIMINATION OF ZOOM INTENT

Joseph H. Goldberg
The Pennsylvania State University
University Park, PA 16802

Jack C. Schryver
Oak Ridge National Laboratory
Oak Ridge, TN 37831

An analysis methodology and associated experiment were developed to assess whether definable and repeatable signatures of eye-gaze characteristics are evident, preceding a decision to zoom-in, zoom-out, or not to zoom at a computer interface. This user intent discrimination procedure can have broad application in disability aids and telerobotic control. Eye-gaze was collected from 10 subjects in a controlled experiment, requiring zoom decisions. The eye-gaze data were clustered, then fed into a multiple discriminant analysis (MDA) for optimal definition of heuristics separating the zoom-in, zoom-out, and no-zoom conditions. Confusion matrix analyses showed that a number of variable combinations classified at a statistically significant level, but practical significance was more difficult to establish. Composite contour plots demonstrated the regions in parameter space consistently assigned by the MDA to unique zoom conditions. Peak classification occurred at about 1200-1600 msec. Improvements in the methodology to achieve practical real-time zoom control are considered.

INTRODUCTION

Eye-Gaze in Computer Interface Control

Eye-gaze can control aspects of the computer interface for applications such as disability aids, military weapons targeting, process control interfaces, telerobotics, and camera manipulation. Recent eye-gaze interfaces have controlled spatial cursor position and object selection. Jacob (1991; 1990) presented an algorithm and demonstration of both of these in a videogame interface. Frey, et al. (1990) developed an eye-gaze-driven word processor. Typically, objects or cursors are selected after exceeding a criterion delay, and deselected after the eye strays outside of a current fixation area (Starker and Bolt, 1990). Extensive averaging of spatial position insures that spurious blinks and/or other anomalies are not considered. Selections are reversible by fixating another area.

Though only first steps in an eye-gaze-controlled interface, the dwell time requirements prior to object or operation selection make these approaches cumbersome to use in real time. Furthermore, they cannot control more abstract operations such as object rotation, or zooming-in/ zooming-out. If controllable from eye-gaze, such operations must necessarily rely on other characteristics of eye movements.

Computer Interface Zoom Control

Zooming-in for a narrower field of view or zooming-out for a broader view are two common operations in graphics, telerobotics, and process control interfaces. A camera mounted on the end of a robotic arm or mobile platform must be controlled in addition to the arm itself (Khosla and Papanikolopoulos, 1992; NASA, 1993). Both camera zoom and position can benefit from eye-gaze control, due to already heavy use of hand controllers, and to the high compatibility of the eye controlling one's point-of-regard. Zooming under eye-gaze control may also aid the control of virtual environment presentations (Stark, et al., 1992).

OBJECTIVE

This study was conducted to determine if repeatable and definable features of eye-gaze precede a user-driven zoom-in or zoom-out decision, and whether either of these could be discriminated from a decision not to zoom. The ultimate goal of this research is to develop a demonstration system of gaze-controlled zoom, then generalize this methodology to broader user intent discrimination.

METHOD: ZOOM DISCRIMINATION

The approach used here for inferring operator intent to zoom-in, zoom-out, or do neither differs substantially from prior eye-gaze interface control methodologies. A multistep modeling procedure is used, as detailed in Figure 1. Currently computed off-line, the procedure has great promise for real-time operation during display interaction.

Cluster Characterization

While viewing a display, time-limited samples of X-Y monocular eye-gaze locations are collected at 30 Hz. Collection of the current gazepoint location, without defining fixation locations, avoids the difficulties associated with criterion dwell times for fixations. Rather than modeling via temporal scanpaths, the spatial locations are connected to form a graph. Prim's algorithm is used (Camerini, et al., 1988) to form a minimum spanning tree (MST), a minimum distance graph without circuits. The MST is separated into clusters, based upon adaptive and defined statistical tests. For each cluster, an associated mean X-Y location, mean and SD diameter, mean and SD pupil diameter, and other parameters are computed. Clusters are formed and characterized on each subsequent (and possibly overlapping) sample or data frame. Clusters are also mapped between frames on the basis of minimum distance, with each cluster in a frame mapped to its closest cluster in its preceding frame.

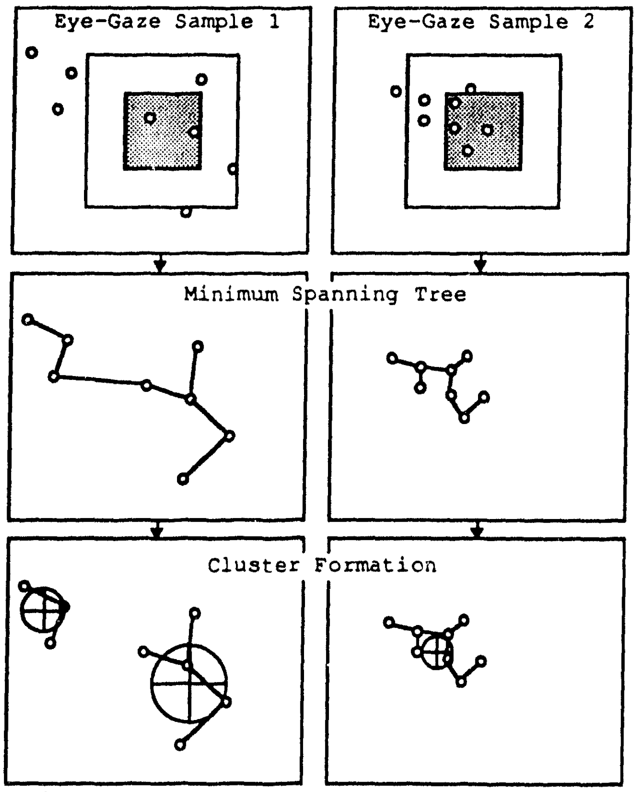

Figure 1. MST and Cluster Formation Process on Two Example Data Frames.

Multiple Discriminant Analysis

The zoom distinction is made by modeling changes in clusters within and between data frames. As a hypothesis, a user may focus his attention in gradually decreasing areas to signal an area on which to zoom-in. Focusing attention toward the outer areas of a window may signal a desire to zoom-out. Such trends become apparent via multiple discriminant analysis (MDA). Computable for any set of cluster characteristics, including changes in characteristics between frames, the MDA can rapidly define the criteria for best separating the zoom-in, zoom-out, and no-zoom conditions. These criteria are displayed by projected lines onto variable scatter plots. Code for the MDA was adapted from Murtagh and Heck (1987), and assigns each data point to the closest zoom group mean in discriminant function space, using the Mahalanobis distance. Significance of the MDA assignments are assessed using a confusion matrix, and a test statistic presented by Press (1972).

The cluster characterization and MDA provide a rapid means for describing the heuristics that optimally separate the three zoom conditions. Emergent heuristics may be user-dependent, or may be generalized to a broader population given similarities in natural eye-gaze tendencies among users. The study presented here attempted to determine if such between-user similarities exist.

METHOD: DATA COLLECTION

Using a within-subject approach, a controlled experiment was conducted to determine if repeatable and definable eye-gaze cluster characteristics precede a user-initiated decision to zoom-in, zoom-out, or not to zoom.

Subjects

Ten volunteers, recruited from the scientific staff at Oak Ridge National Laboratory, (ORNL) served in this study. Each was individually tested, in a single one-hour session. Their ages ranged from 26-60 years of age, and seven were male. All regularly worked with computers, and were familiar with zooming operations.

Apparatus and Calibration

The experiment was conducted in an isolated room containing computer and eye-gaze tracking apparatus. Each subject placed his chin in a chin rest in front of the workstation display. Eye-gaze was collected using an LC Technologies (Fairfax, VA) infrared system. The camera was interfaced to a host 386 PC, via a frame grabber video card. The PC sent eye-gaze sample data across an RS-232 serial port to a Sun Sparc 2 workstation. The application software, written in C, read the data from the serial port while presenting and recording the experiment.

Initial eye-gaze calibration required viewing a screen of ten points on the Sun display; the calibration was automatically repeated until a minimum distance accuracy was achieved. These calibration indices were sent over the serial port to a host PC file. The eye-screen distance was 20 in.; presented stimuli spanned a 20° visual angle. The angular error of the calibrated eye tracking system was 0.45°, or about 0.15 in. at the 20 in. viewing distance.

Procedure

A "same-different" task, using a varied stimulus comparison set, required each subject to determine whether a represented stimulus was the same (except for a possible size difference) as an earlier memorized stimulus. The procedure insured that a conscious decision and response was made to either zoom (zoom-in or zoom-out) or not to zoom. The trial procedure is illustrated in Figure 2. An initial test stimulus, presented and memorized for 3 sec., had a dark interior circle or square surrounded by a larger circle or square border. A 2 sec. mask erased any retinal afterimage traces. A comparison stimulus was displayed, and eye-gaze collection started. One-third of the trials displayed an enlarged interior of the test stimulus (zoom-out required), one-third displayed the test stimulus border with a small dot in the center (zoom-in required), and one-third displayed a stimulus with both interior and exterior (no zoom required). While viewing the comparison stimulus, the subject decided whether to zoom-in for more detail (by pressing the left mouse button), zoom-out for a broader view (right button), or immediately respond (same or different; "s" or "d"). Eye-gaze collection stopped with any of these responses. A zoom-in response was immediately followed by the interior of the comparison stimulus, and zoom-out by its exterior. For either of these, the subject now had sufficient information to make the "s" or "d" response, immediately followed by the next trial. Trials with an improper response were automatically repeated at the end of the experiment.

Subject instructions stressed accuracy over speed. Though the eye tracking camera was apparent beneath the computer display, the subject did not know when eye-gaze was actually being collected. Each subject paused when necessary; these trials were later automatically repeated.

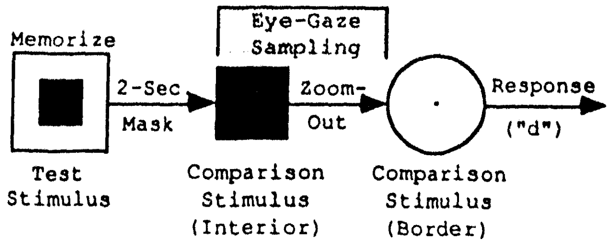

Figure 2. Screen Events During Zoom-Out.

Within each subject, 96 trials were defined as: 4 test stimuli x (2 zoom-in + 2 zoom-out + 2 no-zoom) x 4 replicates. The trial order was fully randomized between subjects. Practice trials presented the entire experiment, with chance to pause for explanation, in a different random trial order.

RESULTS

Within-Ss

Results for a typical variable pair for a single subject are shown in Figure 3. The scatterplot shows each cluster and trial surviving the 4th frame for a total of 26 points. Group means of zoom conditions are indicated by smaller solid shapes. This figure shows frame-to-frame change in cluster size as a function of mean vertical screen distance in pixels. The data are also screened for outliers to provide improved visualization.

The MDA function draws the classification boundaries in parameter space derived from the two discriminant function solution. The resulting decision space correctly classified 61%

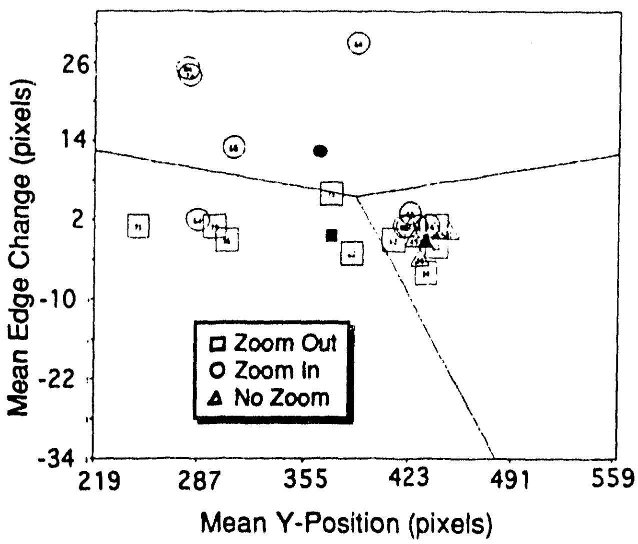

Figure 3. Single-Subject Classification.

of the data. The test statistic [$\chi^2(1)=9.31$; $p<.01$] was significant, but practical significance is more difficult to establish. The better performing combinations of two or three variables correctly classified from 50% to 65% of all trials, with occasional predictors scoring over 70% correct. We would prefer to see scores in the 80-90% range.

Across Ss

The practical purpose of this exploratory study was to identify common eye-gaze cluster signatures as markers for zoom conditions. Table 1 shows the top three variable pair and triplet predictors across all ten subjects. The sum of the individual test statistics is also Chi-Square-distributed with 10 degrees of freedom. The Chi-Square statistic was selected as the figure of merit because the proportion of correct classification is biased toward smaller samples sizes, and hence, later frames.

Table 1. Best Predictors of Zoom Intent

VARIABLE COMBINATION	χ^2 Sum (df=10)
Edge SD/Chg MnX	51.83
MnY/Chg Mn Edge	58.88
Chg MnX/Chg CtrDistMn	51.08
MnX/MnY/PupMn	80.46
Chg MnX/Chg MnY/Chg PupMn	87.69
Chg MnX/Chg CtrDistMn /Chg PupMn	93.82

All predictors in Table 1 were highly significant, indicating that they were consistent predictors across subjects, but this result does not guarantee that classification boundaries were reliable across the entire sample. Classification regions formed by single-subject MDAs were overlaid to form contour plots showing the density of each part of the parameter space for each zoom condition. Figure 4 shows a composite contour plot for the zoom-in, zoom-out, and no-zoom cases. The most consistent regions across subjects for each condition are marked by dense fill patterns, and second-most-consistent areas are indicated with sparse fill patterns. For example, the best region for zoom-in is located at the upper right, where increasing mean edge lengths are coincident with large mean Y values (at the bottom of the screen). In the middle of this region is a smaller area representing moderate no-zoom preference. The high-density zoom-in area not contained in the zoom-in/no-zoom intersection would then be the best basis for implementation of zoom-in, based on the predictions of this variable pair.

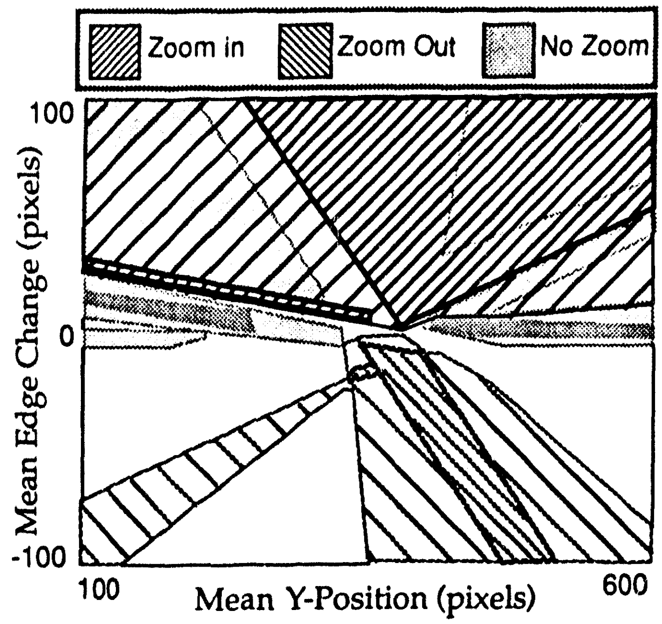

Figure 4. Composite Contour Plot

The data was collapsed over stimulus patterns and zoom condition in order to examine the goodness-of-classification as a function of temporal frame relationship. This function is plotted for the six best predictors in frames of size 12 (400 msec) in Figure 5. For nearly every predictor, classification improves monotonically, achieving peak performance at the fourth frame, and then declining at frame five. The trends noted in Figure 5 are consistent with a model in which subjects first adapt to the comparison stimulus, encode stimulus features, formulate zoom intentions, and finally engage in motor response behavior (button press). The zoom intent phase occurs at about frame four at 1200-1600 msec.

DISCUSSION

This methodology bypasses the need to determine fixations from dwell time criteria by utilizing graph-theoretical characterizations of raw eye-gaze data. Clustering techniques are becoming well-accepted in eye-gaze area (e.g., Latimer, 1988), and better capture attention locus than scanpath analysis.

Classification by MDA is initially poor, but improves in subsequent data frames. Classification was generally possible by 1.6 sec., after adaptation and encoding.

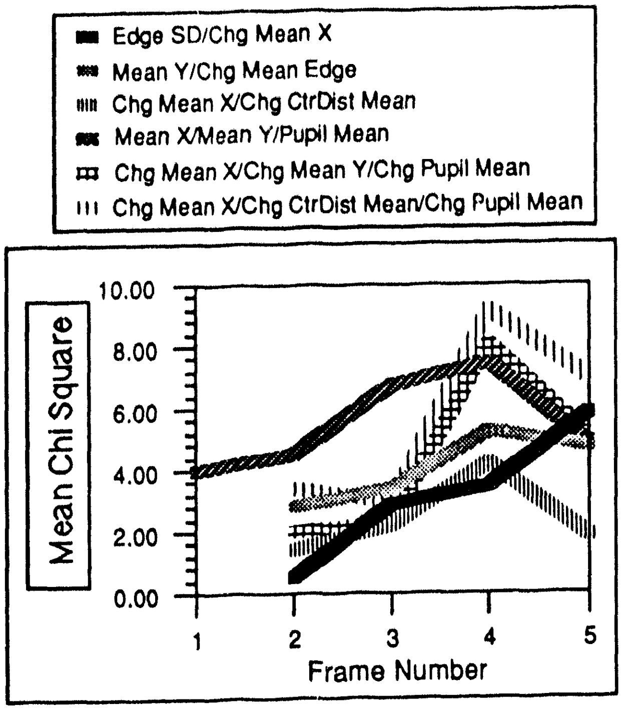

Figure 5. Frame Effect on Classification.

A number of predictor models produced consistently satisfactory classification, and may be promising for between-user applications. For example, in Figure 4 we see a strong zoom-out region at decreasing mean edge length in the bottom display-half. Heuristics for zoom intent with even better consistency can be defined by applying Boolean logic to several variable combinations. Improved classification is also likely from use of nonlinear MDA, neural nets, or similar techniques.

An important advantage of the present technique is that natural eye-gaze tendencies are utilized, and therefore the interface is transparent to the user.

Broader generalization of the present procedure to user intent discrimination is planned. Present stimuli could be replaced with those differing in rotation, other operations, or used in combination with non-eye-gaze sources such as keystrokes and additional physiological measures. Ultimately, an entire eye-gaze controlled interface may be feasible given sufficient experimental observation.

ACKNOWLEDGMENTS

The authors acknowledge the following sources for research support:

- Oak Ridge National Laboratory
- DOE/NE Robotics for Advanced Reactors Program (ORNL Engineering Physics and Mathematics Division)
- DOE Advanced Controls Program (ORNL Instrumentation and Controls Division)

REFERENCES

Camerini, P.M. Galbiati, G., and Maffioli, F. (1988). Algorithms for finding optimum trees: Description, use and evaluation. Annals of Operations Research, 13, 265-397.

Frey, L.A., White, K.P., and Hutchinson, T.E. (1990). Eye-gaze word processing. IEEE Transactions on Systems, Man, and Cybernetics, 20(4), 944-950.

Jacob, R.J.K. (1990). What you look at is what you get: Eye movement-based interaction techniques, CHI '90 Proceedings: Empowering People, Carrasco, J.C., and Whiteside (Eds.), 11-18.

Jacob, R.J.K. (1991). The use of eye movements in human-computer interaction techniques: What you look at is what you get. ACM Transactions on Information Systems, 9(3), 152-169.

Khosla, P.K., and Papanikolopoulos, N.P. (1992). Telerobotic visual servoing. Journal of Applied Intelligence, 2, 127-154.

Latimer, C. R. (1988). Eye-movement data: Cumulative fixation time and cluster analysis. Behavior Research Methods, Instruments, & Computers, 20(5), 437-470.

Murtagh, F., and Heck, A. (1987). Multivariate Data Analysis. Boston, MA: Kluwer and Dordrecht.

NASA (1993). Movable cameras and monitors for view telemanipulator. NASA Tech. Briefs, Feb., 54-55.

Press, S.J. (1972). Applied Multivariate Analysis. New York: Holt, Rinehart and Winston, Inc.

Stark, L.W., Ezumi, K., Nguyen, T., Paul, R., Tharp, G., and Yamashita, H. I. (1992). Visual search in virtual environments. Human Vision, Visual Processing, and Digital Display III, Volume 1666, Proceedings of The International Society for Optical Engineering, 577-589.

Starker, I., and Bolt, R. A. (1990). A gaze-responsive self-disclosing display. CHI '90 Proceedings: Empowering People, Carrasco, J.C., and Whiteside, J. (Eds.), 3-9.

DATE
FILMED
12/01/93
END