Non-Intrusive Gaze Tracking Using Artificial Neural Networks

Dean Pomerleau & Shumeet Baluja

School of Computer Science
Carnegie Mellon University

1. Introduction

The goal of gaze tracking is to determine where a person is looking from the appearance of their eye. Numerous attempts have been made to create better gaze trackers. The interest in gaze tracking exists because of the vast number of potential applications. Two of the most common uses of a gaze tracker are as an alternative to the mouse as an input modality [Ware & Mikaelian, 1987] and as an analysis tool for human-computer interaction studies [Nodine et. al, 1992].

Viewed in the context of machine vision, successful gaze tracking requires techniques to handle imprecise data, noisy images, and a possibly infinitely large image set. The most accurate gaze tracking has come from intrusive systems which either require the subject to keep their head stable, through chin rests etc., or systems which require the user to wear cumbersome equipment, ranging from special contact lenses to a camera placed on the user's head to monitor the eye. The system described here attempts non-intrusive gaze tracking, in which the user is neither required to wear any special equipment, nor required to keep his head still.

2. Gaze Tracking

2.1. Standard Gaze Tracking

In standard gaze trackers, the image of the eye is processed in three basic steps. First, the specular reflection of a stationary light source is found in the eye's image. Second, the pupil's center is found. Finally, the relative position of the light's reflection and the pupil's center is calculated. From information about their relative positions, the gaze direction is determined [Figure 1]. In many of the current gaze tracker systems, the user is required to remain motionless, or special headgear must be worn by the user to maintain a constant offset between the position of the camera and the eye.

2.2. Neural Network Based Gaze Tracking

One of the primary benefits of the NN based gaze tracker is that it is non-intrusive; the user is allowed to move his head freely. In order to account for the shifts in the relative positions of the camera and the eye, the eye must be located in each image frame. In the current system, the eye is located by searching, in the image of the user's face, for the specular reflection of a stationary light. This can usually be distinguished by a small bright region surrounded by a very dark region. The reflection's location is used to limit the search for the eye in the next frame. A 15x30 window surrounding the reflection is extracted; the image of the eye is located within this window. [Figure 2].

To determine the coordinates of the gaze, the 15x30 window is used as the input to the neural network. The forward pass is simulated in the NN, and the coordinates of the gaze are determined by reading the output units. The output units are organized with 50 output units for specifying the X coordinate, and 50 units for the Y coordinate [Figure 3]. The gaussian output representation used is similar to that used in ALVINN [Pomerleau, 1992].

The network is trained by the user moving the cursor with the mouse around the screen, while visually tracking the cursor. The image of the eye is digitized, and paired with the (X,Y) coordinates of the cursor. A total of 1500 image/position pairs are gathered. The network is trained for approximately 200 epochs, using standard back propagation.

3. Results

The current system works at 10 hz.. The best accuracy we have achieved is 1.5 degrees with the freedom of head movement up to 30cm.. Although we have not yet matched the best gaze tracking systems, which have achieved approximately 0.75 degree accuracy, our system is non-intrusive, and does not require the expensive hardware which

many other systems require. Figure 4 shows how our system compares with a commercial gaze-tracking system, the ISCAN.

4. Future Directions

One of the largest problems in existing eye trackers is the inability to handle user motion. To address the problem of user motion, a mobile camera could be used. To control the camera, another neural network can be trained to keep the user's face in the center of the image. An extension of this idea is to use a neural network to find the eye in the image of the face. Currently, the heuristic used to find the eye in the image of the face is to locate a bright spot surrounded by dark regions. However, neural networks have been applied to facial feature tracking [Hutchinson, 1990] [Debenham, 1991] and this technology may help here. Another extension of the idea of facial tracking is using extra inputs units to represent the head position. Finally, for increased accuracy, a higher resolution image of the eye may yield improved results. The current resolution is only 15x30 gray-scale pixels. Increasing the resolution may, however, decrease the speed of the gaze tracker.

In order to rapidly train the neural network to new users, one method may be to use a multiple network architecture, as was used in the MANIAC autonomous road following system [Jochem, 1993]. In this system, several smaller "expert" networks were trained on different road types, ie. one lane, two lane etc.. An arbitrating network, which resided "on top" of the smaller networks, is used to select which of the expert networks is yielding the best response to the current road. In a similar manner, expert networks can be trained on several different people, with retraining only having to change the weights in the arbitrating network. Arbitration between networks could also be constructed through the use of metrics which estimate the output reliability [Pomerleau, 1993]. A benefit of the expert network approach is that each of the expert networks can be trained independently of the ôthers.

This flexibility of head position makes the connectionist gaze tracker much less intrusive than existing systems. We would like to test the viability of replacing the mouse with the connectionist gaze tracker. Other potential uses for the system, which we will be tested in the future, include helping disabled people interact with their environment, and as a tool for data collection in human-computer interaction experiments.

References

Hutchinson, R.A. (1990), "Development of an MLP feature location technique using preprocessed images". In proceedings of *International Neural Networks Conference*, 1990. Kluwer.

Jochem, T.M., D.A. Pomerleau, C.E. Thorpe (1993), "MANIAC: A Next Generation Neurally Based Autonomous Road Follower". In Proc. of International Conference on Intelligent Autonomous Systems (IAS-3).

Debenham, R.M. & S.C.J. Garth (1991), "The Detection of Eyes Using Radial Basis Functions". In Proceedings of the *1991 International Conference of Artificial Neural Networks ICANN-91*. Amsterdam, Netherlands, North-Holland.

Fahlman, S.E. & Lebiere, C. (1991) "The Cascade Correlation Learning Architecture". *Advances in Neural Information Processing Systems 2*. D.S. Touretzky (ed.) Morgan Kaufmann, pp. 524-532.

Nodine, C.F., H.L. Kundel, L.C. Toto & E.A. Krupinksi (1992) "Recording and analyzing eye-position data using a microcomputer workstation", *Behavior Research Methods, Instruments & Computers* 24 (3) 475-584.

Pomerleau, D.A. (1991) "Efficient Training of Artificial Neural Networks for Autonomous Navigation," *Neural Computation 3*:1, Terrence Sejnowski (Ed).

Pomerleau, D.A. (1992) *Neural Network Perception for Mobile Robot Guidance*. Ph.D. Thesis, Carnegie Mellon University. CMU-CS-92-115.

Pomerleau, D.A. (1993) "Input Reconstruction Reliability Estimation", *Neural Information Processing Systems 5*. Hanson, Cowan, Giles (eds.) Morgan Kaufmann, pp. 270-286.

Starker, I. & R. Bolt "A Gaze-Responsive Self Disclosing Display", In *CHI-90*. Addison Weseley, Seattle, Washington.

Ware, C. & Mikaelian, H. (1987) "An Evaluation of an Eye Tracker as a Device for Computer Input", In J. Carrol and P. Tanner (ed.) *Human Factors in Computing Systems - IV*. Elsevier.

Figures

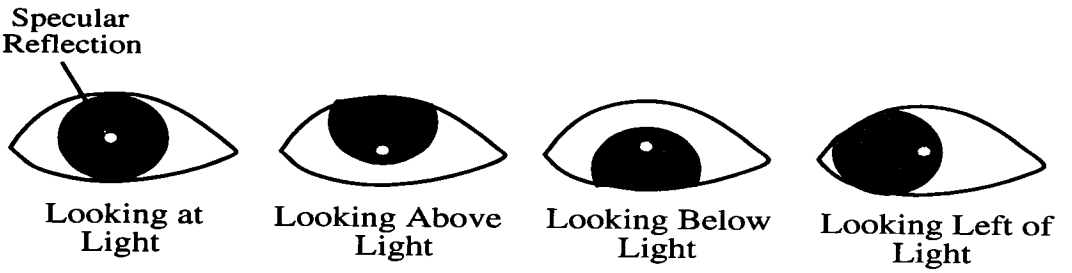

Figure 1. Relative position of specular reflection and pupil.

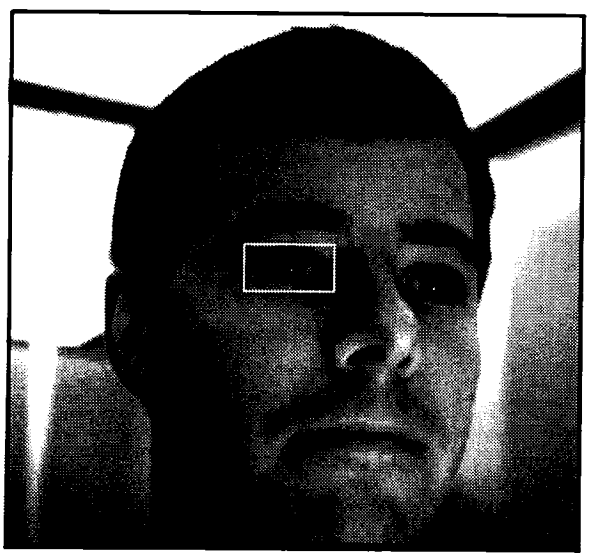

Figure 2. The 15x30 window extracted from the image of the user's face.

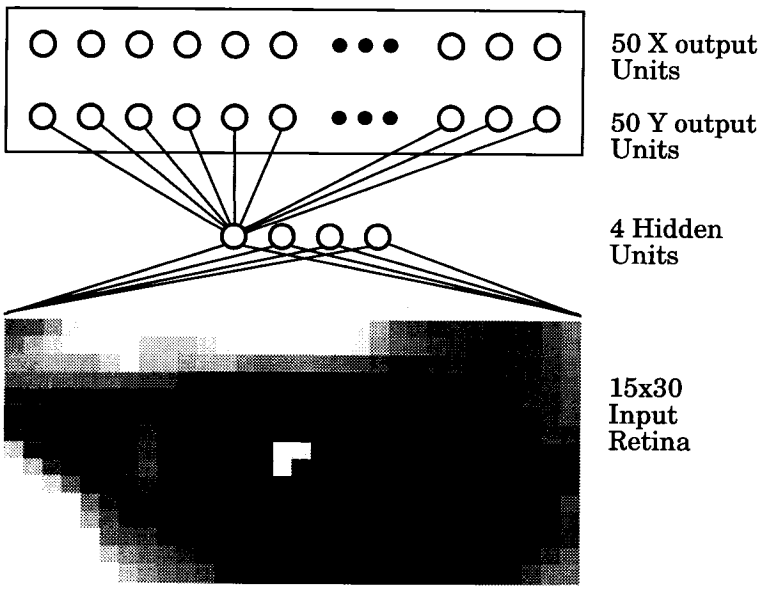

Figure 3. Network Architecture

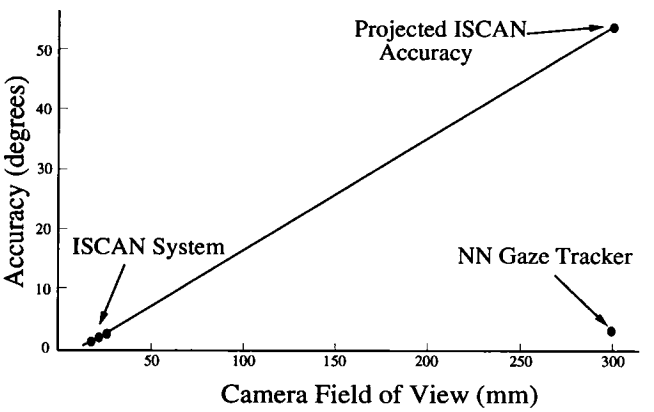

Figure 4. NN gaze tracker vs. ISCAN.