Implicit calibration of a remote gaze tracker

Xavier L. C. Brolly
NASA Ames Research Center
brollyx@eos.arc.nasa.gov

Jeffrey B. Mulligan
NASA Ames Research Center
Jeffrey.B.Mulligan@nasa.gov

Abstract

We describe a system designed to monitor the gaze of a user working naturally at a computer workstation. The system consists of three cameras situated between the keyboard and the monitor. Free head movements are allowed within a three-dimensional volume approximately 40 centimeters in diameter. Two fixed, wide-field "face" cameras equipped with active-illumination systems enable rapid localization of the subject's pupils. A third steerable "eye" camera has a relatively narrow field of view, and acquires the images of the eyes which are used for gaze estimation. Unlike previous approaches [1] which construct an explicit three-dimensional representation of the subject's head and eye, we derive mappings for steering control and gaze estimation using a procedure we call **implicit calibration**. *Implicit calibration is performed by collecting a "training set" of parameters and associated measurements, and solving for a set of coefficients relating the measurements back to the parameters of interest. Preliminary data on three subjects indicate an median gaze estimation error of approximately 0.8 degree.*

1. Introduction

While some may question the truth of the English proverb "the eyes are the window to the soul," it is generally agreed that a person's gaze is directed to the current object of attention. Fixation sequences or scan-paths are therefore of interest in the study of visual cognition, and applications such as the design of human-computer interfaces. When studying eye movements that occur during natural behaviors, it is important that the measurement system be unobtrusive and not affect the behavior of the subject. In this paper we present a remote gaze tracking system designed with these goals in mind.

There are many methods available for tracking the gaze of a human observer. Unfortunately, it is difficult to achieve high accuracy in a system which is both non-invasive and unobtrusive, two qualities which are imperative for the study of natural behavior. Many high-accuracy systems require stabilizing the head, either with a chinrest or "bitebar" dental impression. While useful for many laboratory studies, restraining movement of the head is undesirable when studying natural behavior[2, 3]. The search coil technique[4, 5] provides high accuracy while allowing a large range of natural movement, but is very invasive, requiring placement of the coil on the surface of the eye.

Our system was in large part inspired by a similar system described by Beymer and Flickner [1]. Like their system, we employ a stereo pair of fixed, wide-field cameras to perform an initial localization of the head. Unlike their system (which employs a short-baseline off-the-shelf stereo rig), our wide-field cameras are located on a large baseline and are equipped with active illuminators for optical pupil-finding [6, 7]. These cameras are fitted with 8mm focal length lenses (providing a field of view of approximately 37 degrees), and provide images of the head and neck; we refer to these as the "face" cameras.

Again similar to Beymer and Flickner, we use a mirror galvonometer system to enable rapid tracking of head movements by a narrow-field eye camera. But unlike their system, we use a single narrow field eye camera instead of two. Our narrow field camera is a miniature "board camera" equipped with a 25mm microlens, providing a field of view of 14 degrees. This is a wider field of view than that obtained in the Beymer and Flickner system, which hurts system accuracy, but has other advantages. The small aperture diameter of the microlens allows the use of smaller mirrors, which have lower mechanical inertia and therefore a faster settling time. The shorter focal length results in greater depth-of-field, so dynamic focusing is not required.

The remainder of the paper is organized as follows: first, we discuss the control of the steering system, including the method of calibration and performance data; then, we present our method for estimating the gaze target from image measurement data, and data from three subjects; finally, we discuss areas where we hope to make future improvements.

2 Steering system

Before the host system can issue the steering commands, it must somehow determine where it wants the system to point. Once the eye has been acquired, small movements
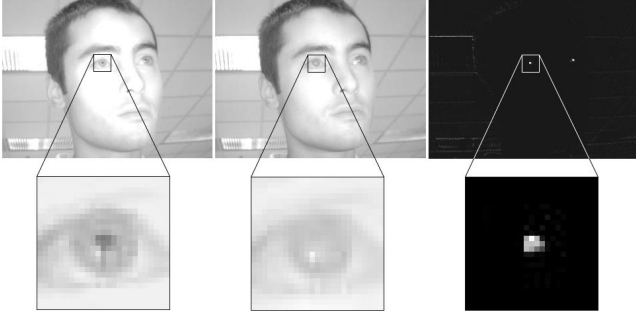
Figure 1: Montage of dark- and bright-pupil field images (left and middle images resp.), with clipped difference (right image). In addition to the pupils, artifacts resulting from spatial misregistration of the fields and subject motion can be seen.

can be detected in the image from the eye camera itself. But when the subject first sits down in front of the system, it can not be taken for granted that the eye camera will even see part of their head, let alone image one of their eyes. Therefore, two wide-field cameras are used to image the subject's head and determine the positions of the pupils.

2.1 Pupil detection with active illumination

Detection of the pupils in the face camera images is simplified by the use of a special illumination system [6, 7, 8]. The illumination system consists of two banks of near-infrared LED's (Lite-On LTE-3376, 850 nm), arranged concentrically about the camera lenses. The inner ring produces a bright pupil, while the outer ring produces a dark pupil image. The two LED banks are energized alternately, with changes triggered by the video sync pulses. The difference image of the bright- and dark-pupil images is computed, resulting in an image consisting primarily of only the pupils (figure 1). The small disks are generally easy to discriminate from other artifacts, and the positions of their centroids are the primary measurements which are output from this stage.

2.2 Steering system calibration

The input data to the system are the locations of the pupils in each of the face camera images. To transform these measurements into the three-dimensional coordinates of the pupil locations is a classic stereo vision problem, for which many solutions exist (a good survey is provided by Faugeras and Luong [9]), most of which require calibration of each of the cameras [10, 11]. The key insight underlying

our method of "implicit" calibration is that it is not necessary to know the three-dimensional location of the target in order to set the steering angles, it is only necessary to know on which line-of-sight of the steering system the target lies. Of course, this is easily determined if the three-dimensional coordinates are known, but only if the steering system itself has been calibrated to the global three dimensional coordinate system. Instead of calibrating the different subsystems individually, our implicit calibration method solves for the end-to-end relation between image and galvo coordinates, without solving explicitly for either the camera calibration parameters or the three dimensional coordinates of the targets. The calibration is performed by substituting a small laser pointer in place of the steerable camera, and placing a screen at various positions within the subject volume. Both the laser and eye camera are mounted on kinematic base plates (Newport BK-3), allowing them to be interchanged rapidly while matching precise alignment. A mirror mount (Newport P100-P2) is used to align the laser beam with the camera's line of sight. Precise positioning of the screen is unnecessary; we place it by hand at several positions, so as to sample the extremes of the working volume. The steering mirrors are swept through a range of positions, and the positions of the images of the laser spot are measured in each of the face camera images. The resulting "training set" of measurements is then used to develop mappings from image coordinates to scanner settings. We obtain good results (aiming accurate to within a few pixels) by using a simple second-order polynomial mapping.

The positions of the target in each of the face camera images are measured for N positions of the steering system and projection screen. Let p_i be the pan coordinate on the ith frame, and let t_i be the tilt coordinate. (In practice, the galvo "coordinates" are the unsigned 16 bit integers which are sent to the D/A converter.) We wish to discover functions f_p and f_t such that

$$p_i = f_p(x_{L,i}, y_{L,i}, x_{R,i}, y_{R,i}) \qquad (1)$$

and

$$t_i = f_t(x_{L,i}, y_{L,i}, x_{R,i}, y_{R,i}), \qquad (2)$$

where $x_{L,i}$ is the x coordinate of the target position in the left face camera, etc. The coefficients of polynomial approximations to f_p are obtained by multiple regression. This is done by constructing a matrix $\mathbf{M}$, in which each row contains a measurement vector, a constant element with value 1, and (optionally) higher order product terms formed from the measurements. Let $\mathbf{p}$ be a column vector with elements p_i, $\mathbf{q}$ the vector of predictions f_p, and let $\mathbf{a}$ is represent the vector of unknown coefficients. By definition,

$$\mathbf{q} = \mathbf{M}\mathbf{a}. \qquad (3)$$

A vector of coefficients $\mathbf{a}$ is sought for which $\mathbf{q}$ approximates $\mathbf{p}$ as closely as possible. A least-squares solution is

obtained by

$$\mathbf{a} = \mathbf{M}^{-1}\mathbf{p}, \qquad (4)$$

where $\mathbf{M}^{-1}$ is the pseudo-inverse of $\mathbf{M}$, obtained using the singular value decomposition.

The number of columns of $\mathbf{M}$ increases quickly with an increase in the order of polynomial terms included. If N_m is the number of measurements per observation (4 in our example), then the total number of terms T_k in the polynomial of degree k is given by:

$$T_k = \frac{\prod_{i=1}^{k}(N_m + i)}{k!}. \qquad (5)$$

In our case, there are 5 total coefficients in the linear approximation, 15 for the quadratic, 35 for the cubic, and 70 for the quartic.

The number of coefficients sets a lower bound on the number of observations that must be made in order to guarantee a solution. While we have not derived conditions that must be met for a valid solution, it seems obvious that the entire three-dimensional volume must be sampled; we expect that a valid solution will not be obtained if only a single position of the projection screen is used, regardless of its orientation. By analogy with the 1-dimensional case, we conjecture that to obtain a valid solution of order m, it is necessary that we collect data for at least $m + 1$ different positions of the projection screen, and sample with a grid of at least $m + 1$ by $m + 1$ for each screen position.

2.3 Simulation results

To estimate the maximum attainable accuracy possible with the method, simulations were performed. The simulations differ from the real situation in several (hopefully insignificant) ways: first, the camera geometry is only a rough approximation to the actual setup; second, in our model steering system, the centers of rotation for pan and tilt are coincident (in the actual scanner, the two mirrors are separated by about a centimeter); and, finally, the face cameras are modeled as pinhole cameras, with no lens distortion. The geometry used in the simulation is illustrated in figure 2. For our model system, the origin of the coordinate system is at the center of rotation of the scanning system, sending the laser beam along the z axis for pan and tilt angles of 0. The projection screen is modeled as a plane normal to the z axis, which is placed at a series of linearly spaced depths ranging from 30 to 60 centimeters. For each screen depth, we cast a grid of rays, with azimuth and elevation sampled linearly from -15 to 15 degrees. The results presented are for an 11x11x11 sampling grid over depth, azimuth and elevation; in a series of simulations not presented here, we verified that at this sampling density the results have reached their asymptotic values.
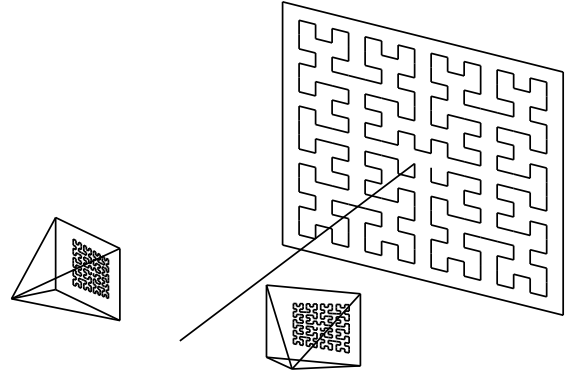

Figure 2: Schematic diagram showing the geometry used in the simulation.

The face cameras are modeled as pinhole cameras with nodal points located on the x axis. They placed 30cm apart and rotated inward 18 degrees about an axis passing through their nodal point, parallel to the y axis (see figure 2). Their positions are centered about a point 2.5cm to the left of the galvonometer.

The points of intersection of the laser beam with the screen are projected into the face camera images using perspective projection. The coordinates are converted to pixel units assuming a sensor with a 0.25 inch diagonal, a resolution of 640x480, and a 9.5 mm focal length.

Out of the original 1331 samples, 678 fall inside the field of view of both cameras. These measurements were used to construct the 678x4 matrix $\mathbf{M}$, as described in the previous section. After computing the coefficients, the predictions were generated. The prediction error was computed independently for pan and tilt, and the average prediction error was computed as the Pythagorean sum of the pan and tilt errors. Figure 3 depicts the average (RMS) prediction error, as a function of polynomial order. When no noise is added to the simulated measurements, a reduction of approximately one log unit is observed for each additional order included in the approximation. The various curves in figure 3 show the results for varying amounts of Gaussian-distributed noise added to the simulated measurements. The added noise has little effect on prediction error for the linear and quadratic cases, but eventually eliminates the advantages conveyed by the higher order polynomial approximations.

The field of view and resolution of the eye camera can be used to convert the angular errors plotted in figure 3 to the number of pixels we would expect the target to be off-center due to an aiming error. For order 2, the average error is between 2 and 3 pixels. Because our area of interest is on the order of 32 pixels in diameter, this amount of error
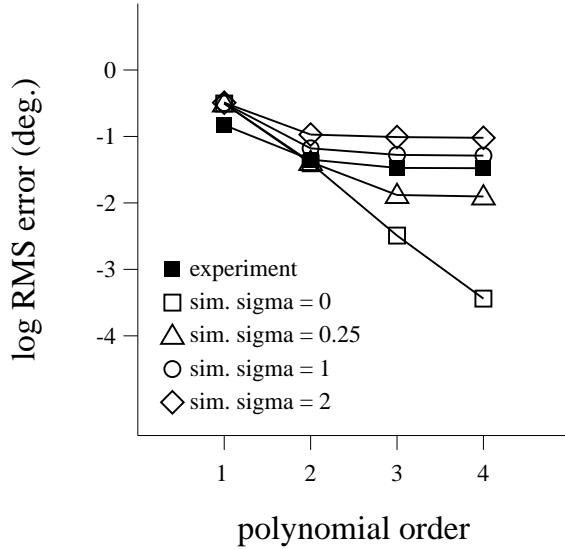
Figure 3: Average (RMS) prediction error for simulated and experimental data, as a function of the degree of the approximating polynomial. The solid squares show the values for the experimental data, while the open symbols show the results for various levels of Gaussian-distributed noise added to the pixel coordinates of the target.
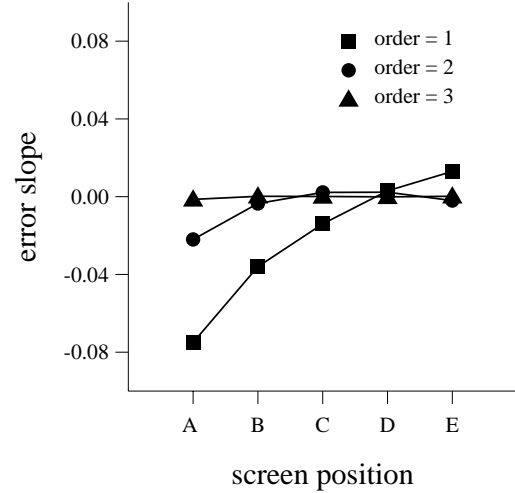

Figure 4: Reduction of systematic error with increasing accuracy of the approximation is illustrated by the "error slope." Screen position A (resp. E) corresponds to the nearest (resp. furthest) screen position.

is acceptable, and the second order approximation provides a good compromise between accuracy and computational load.

2.4 Experimental results

Training data from the actual apparatus were obtained by recording imagery from the cameras while the laser was scanned over a hand-held screen. Scanner updates are synchronized to the camera's frame rate and are performed during the vertical blanking, allowing us to acquire data at 60Hz. Small steps can be completed during the vertical blanking interval, and so to minimize the maximum step size we chose a retrace-free raster described by Koenderink [12]. We use a 32x32 pattern having 1024 samples. The laser was held in each position for two video fields, so each scan required slightly more than 30 seconds. Five scans were collected, at a variety of depths spanning to the expected range of head positions.

For each set of galvo and screen positions, and the position of the laser spot was computed for each face camera image by computing the centroid and area of all pixels exceeding a predetermined threshold. The threshold was chosen to slightly higher than the highest level of the (black) screen under the ambient illumination. The centroid locations, along with the corresponding galvo settings, were used to compute a set of prediction coefficients, as described above,

With the coefficients, prediction errors were computed for the training set. The results are shown in figure 3. We see good agreement between the empirical data and the simulated results for a 1 pixel standard deviation noise level.

In general, the target points near the edges of the training volume are the ones where the largest prediction errors occur. We notice that for a given (fixed) position of the screen, angular error vs. angle is well-fit by a straight line, with the sign of the slope indicating expansion or contraction of the reconstructed pattern compared to the original one. Figure 4 shows the azimuthal error slope as a function of azimuth for the empirical data. Each curve in the figure represents a different regression order. For the linear predictor, we see that there is a strong dependence of the errors on screen position. The second order predictor still shows systematic errors for the nearest screen position, but has done a fairly good job of eliminating the systematic error for the other screen positions. For the third order predictor, the error slope is essentially zero at all screen positions, indicating the systematic errors are gone and the residuals are random. This observation helps us make sense of the results shown in figures 3: when there are systematic errors (as for order 1), we can generally expect to obtain a better fit by increasing the polynomial order, while no significant gains are obtained when the residual errors are random (as for orders 3 and greater).

3 Gaze estimation

The steering system described above provides us with an image from the narrow field camera that is roughly centered
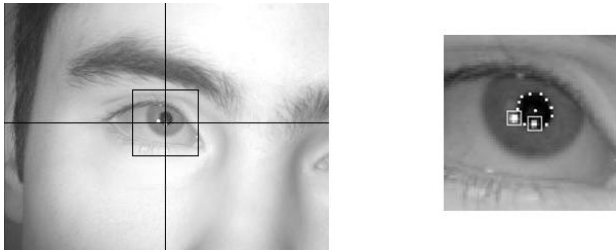
Figure 5: The left panel shows the full image from the eye camera image. A black square has been overlaid indicating the region of interest, centered on the glint found using a Gaussian curvature operator. The right panel shows an enlarged view of the region of interest, with the detected locations of the glints and pupil indicated. Two glints are produced by the active illuminators on the two face cameras.

Figure 6: Exemple of eye orientation causing multiple glint images due to reflection from the sclera.

on an eye specified by the user. The field of view is large enough that both eyes can be imaged by aiming the steering system mid-way between the two eyes, but in this paper we will describe gaze estimates based on measurements using a single eye. As in the previous section, here we will use an implicit calibration method.

3.1 Region selection

If the head were stationary, then the targeting process described above would ensure that the eye would appear in the center of the eye camera image. The purpose of the active tracking system, however, is to allow the user to move freely. When the head moves, it is no longer possible to keep the eye exactly centered in the eye camera image, because of the time lag between detection of the pupils in the face camera images and the updating of the galvo positions. Assuming we process the data at the maximum rate of 60Hz, after acquiring a pair of successive fields, the update to the galvos will occur at the end of the following field. In the absence of trajectory prediction, we expect the image of the eye to be displaced from the center of the image by the amount of motion that occurs in 17 milliseconds. Therefore, the first step is to identify the subregion of the image centered on the eye.

The bright reflections of the illuminators, or 'glints', are the most conspicuous features and are therefore used for this purpose. The glints are distinguished by two characteristics: they are usually bright, and they are small and roughly circular. With our setup, we observe two separate glints, produced by two active illuminators on each of the face cameras. We have obtained robust detection using a second
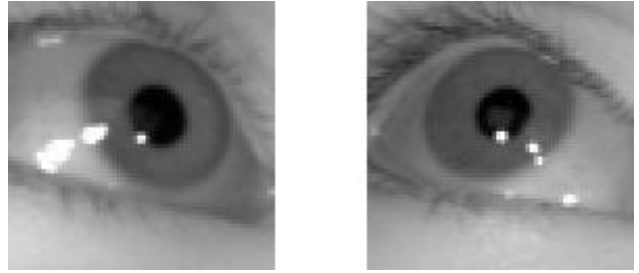
derivative operator related to Gaussian curvature [13, 14]. We compute the simplified Gaussian curvature K using the following formula:

$$K = g_{xx}g_{yy} - g_{xy}^2, \qquad (6)$$

where g_{xx}, g_{yy} and g_{xy} are image second derivatives. This operator has zero output for regions which are constant or vary in only one dimension, responds to curved edges and corners, and has a large output for small spots. The curvature is computed over a search region (which may be smaller than the entire image), and the location of the maximum is chosen as the center of a smaller subregion to be used for subsequent processing (fig. 5).

3.2 Feature measurements

The primary features which we have used to date are the positions of the two glints, and the location of the pupil. The glints are located using the Gaussian curvature as described above. However, a problem occurs for large gaze deviations, for which the glint moves from the cornea to the sclera (figure 6). Possible approaches to this problem include detailed modeling of the reflections, or storing template images for the difficult head/gaze regions. Another possible approach would be to add additional illumination sources which could be switched on and off depending on the position of the head and the direction of gaze, such that two or more good reflections from the cornea are always obtained. This requires an alternate approach to galvo steering with active illumination, which has been successfully done by tracking the eye in the narrow-field camera image and feeding back directly to the galvos.

Fast estimation of the pupil position and shape is performed by fitting an ellipse to a set of boundary points. The direct solution method of Pilu *et al*. [15] is used. The set of boundary points is obtained as follows: a provisional center is obtained searching for the darkest pixel, the image gradient is computed along 12 radial vectors, and the location
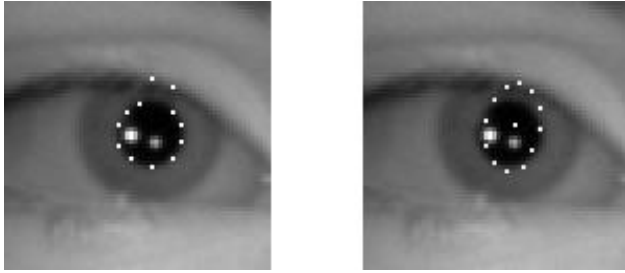
5

Figure 7: Example of pupil mislocalization due to misidentification of pupil boundary points. The image on the left shows 12 points found by searching for the maximum gradient along a radial line, including 2 which actually fall on the edge of the eyelid. The right hand panel shows the output of the ellipse fitting procedure, with substantial errors in both the center location and aspect ratio.
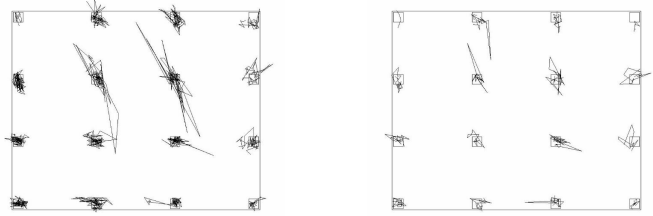

Figure 8: Graphical rendering of recovered gaze estimates for a typical sample. On the left, gaze estimates of the training data are overlaid on a representation of the display screen. The small squares (representing about 1 degree of visual angle) indicate the locations of the fixation targets. On the right is a plot of the corresponding test data. Gaze estimates were computed using galvo settings, inter-glint distance, and pupil-glint vector, and polynomial terms up to order 3. Subject WHA.

of the maximum along each direction is taken as an edge point. The main weakness with our current implementation is poor robustness with respect to outliers; for example, it sometimes occurs that the maximum change occurs at the edge of the eyelid instead of the pupil margin (figure 7). We are confident that improved outlier rejection can be achieved using existing techniques [16].

3.3 Determination of mapping coefficients

Given a set of measurements of the image and the settings of the steering system, we wish to determine the target of gaze on a display screen. One approach to this problem is exemplified by the work of Beymer and Flickner [1], who solve for the parameters of an explicit model of the eye which give the best match to the data. Here we present an alternative approach, which is analogous to our calibration of the steering system described above. We will look for a polynomial function of the image measurements that generate the screen coordinates of the fixation target. Image measurements can be composed by the following data: galvo settings, glint positions, and pupil position, size and shape (eccentricity). The estimation of the coefficients of that polynomial function requires a training period, where subject-specific data is collected. Then we choose a set of variables (measurements or derived measures), and solve for regression weights mapping the variables to the coordinates of the target location. This is analogous to the procedure described above for calibration of the steering system, but with a larger set of possible input variables, which may not all be used.

A significant difference between the gaze estimation problem and steering calibration is that for steering we knew the galvo coordinates that we were trying to predict. In the case of gaze calibration, however, we do not know the true gaze coordinates, because the subject generally makes errors of fixation. We assume that the gaze coordinates are equal to the target coordinates, and trust the random nature of the errors will not introduce biases. Before predicting the gaze coordinates, we must resolve two questions: first, which input variables should be used? Second, for a given set of inputs, which polynomial order is necessary to obtain reliable estimates?

To collect the subject-specific data, each subject was instructed to fixate a series of targets on the screen comprising a 4x4 grid. For each target, the subject was instructed to maintain fixation while moving their head about slowly in all three dimensions during a 2 second interval. The subject's right eye was tracked by the galvos; the central portion of the image from the eye camera was stored and for later analysis. The resulting dataset generally had some outliers arising from misidentified features. We attempted to automatically remove these from the dataset using a number of heuristics. For example, the slope of the line joining the two glints should be roughly constant, and the pupil size and aspect ratio should be within certain limits). But, without a detailed examination of the parameter values in combination with the images, it is hard to be sure that the dataset is completely free of bad data.

3.4 Results

To estimate the performance of a set of parameters, first we solve for the weights using random trials of 80% of the recorded samples (the "training sets"). Then we use these weights to make predictions for the remaining 20% of the
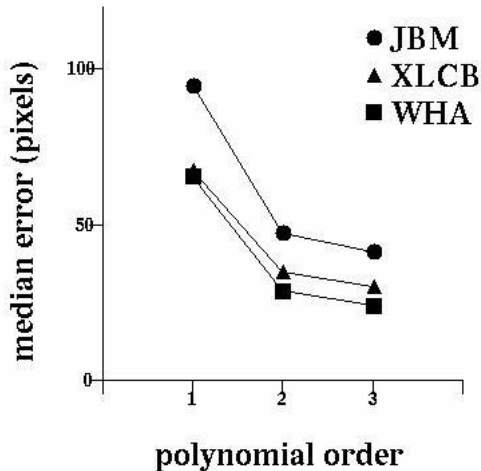
Figure 9: Median test set error as a function of polynomial order for for three subjects. The parameter set is made of galvo settings, inter-glint distance, and pupil-glint vector. At the average viewing distance, 1 pixel corresponds to about 0.025 degrees of visual angle.

recorded samples (the "testing sets"). A typical set of predictions based on a single sample trial of training data is shown in figure 8. Polynomial terms up to order 3 were included in the regression. We define the error to be the distance between the estimated gaze position and the target location, although we recognize that the actual gaze may not always be directly on the target. Because of the outliers in the testing set, we decided to examine the median error, which we believe is more representative of the potential accuracy of the system given more robust algorithms for feature localization. The RMS error is on average twice as big as the median error.

A number of different subsets of the measurements and derived measures were investigated. Using all of the measurements turned out to perform relatively poorly. Increasing the number of input variables increases the number of coefficients, especially for the higher polynomial orders (as in equation 5), so when redundant variables are included, the extra degrees of freedom end up fitting noise. Our best results to date have been obtained using a relatively small set of parameters consisting of the galvo settings, the inter-glint distance, and the pupil-glint vector. (By the pupil-glint vector, we mean a vector from the center of the pupil to the mean location of the two glints.) Figure 9 shows the median errors as a function of polynomial order for this parameter set. Those errors were computed over 20 different random trials for three subjects. The median errors obtained at order 3 are equal to 25, 30 and 40 pixels (for WHA, XLCB and JBM resp.). On a 17" monitor with a resolution of

1024x1280, viewed at a distance of 60 cm, those median errors correspond respectively to 0.63, 0.76 and 1.01 degrees. So for the moment, we estimate the precision of our gaze tracker to 0.8 degree. In order to get a more reliable estimate of that precision, we plan to test the system on a larger number of people.

4 Future work

While the mechanical aspects of our system are more-or-less complete and satisfactory, much work remains on the software side. Ultimately we hope to generate real-time binocular gaze estimates at 60Hz. Our existing software is coded in a high-level scripting language, which is convenient for prototype development, but we anticipate recoding certain key operations in C (or perhaps even assembly language) to achieve the desired processing rates.

As described above, the system uses the pupil locations from the wide-field face cameras to control the steering system. We have also experimented with using the location of the eyes within the narrow field eye camera. One advantage of this approach is that, because of the higher magnification provided by the eye camera, greater precision is possible. A feature at a given image location can be moved to the center of the image using a galvo displacement vector which is a linear transformation of the image displacement vector.

In this paper, the direct mapping from measurements to the parameters of interest was done using polynomial functions, because an efficient solution method exists for obtaining the coefficients. Neural networks can potentially provide an accurate implicit calibration capable of predicting more complex nonlinear relations between the parameters and the measurements, at the cost of more computation to determine the weights (e.g. back-propagation [17]).

5 Conclusions

We have demonstrated a prototype of a remote gaze tracker which allows free head movements within a three-dimensional volume. Although the images of the eye which are obtained have fairly low resolution, we are still able to recover estimates of gaze with an average error of 0.8 degree. We hope to achieve further improvements in both accuracy and robustness through improvements to the software, but are pleased with the results obtained to date. Our accuracy is comparable to that reported by Beymer and Flickner [1], in spite of the reduced resolution of our eye camera. This suggests that the direct estimation method described here may have advantages over their model-based approach, and also that their system may be capable of higher accuracy that they have yet obtained. In future work

we hope to compare estimates obtained by both methods on a single data set.

6 Acknowledgments

References

[1] D. Beymer and M. Flickner, "Eye gaze tracking using an active stereo head," in *Proc. CVPR*, 2003, pp. xxx–xxx.

[2] H. Collewijn, R. M. Steinman, C. J. Erkelens, Z. Pizlo, and Van der Steen, "The effect of freeing the head on eye movement characteristics during 3-d shifts of gaze and tracking," in *The Head-Neck Sensory Motor System*, A. Berthoz, P. P. Vidal, and W. Graf, Eds., pp. 412–418. Oxford University Press, 1992.

[3] R. M. Steinman, "Gaze control under natural conditions," in *The Visual Neurosciences*, L. M. Chalupa and J. S. Werner, Eds., pp. xxx–xxx. MIT Press, 2003.

[4] L. Ferman, H. Collewijn, T. C. Jansen, and A. V. Van den Berg, "Human gaze stability in the horizontal, vertical and torsional directions during voluntary head movements, evaluated with a three-dimensional scleral induction coil technique," *Vision Res.*, vol. 27, pp. 811–828, 1987.

[5] M. Edwards, Z. Pizlo, C. J. Erkelens, H. Collewijn, J. Epelboim, E. Kowler, M. R. Stepanov, and R. M. Steinman, "The maryland revolving-field monitor: theory of the instrument and processing its data," Tech. Rep. CAR-TR-711, Center for Automation Research, University of Maryland, 1994.

[6] Y. Ebisawa and S Satoh, "Effectiveness of pupil area detection technique using two light sources and image difference method," in *Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA*, A. Y. J. Szeto and R. M. Rangayan, Eds., 1993, pp. 1268–1269.

[7] A. Amir C. Morimoto, D. Koons and M. Flickner, "Real-time detection of eyes and faces," in *Proceedings of 1998 Workshop on Perceptual User Interfaces*, 1998, pp. 117–120.

[8] C. Morimoto, D. Koons, A. Amir, and M. Flickner, "Pupil detection and tracking using multiple light sources," Tech. Rep. RJ-10117, IBM Almaden Research Center, 1998.

[9] O. Faugeras and Q. Luong, *The geometry of multiple images: the laws that govern the formation of multiple images of a scene and some of their applications*, MIT Press, 2001.

[10] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE Journal of Robotics and Automation*, vol. RA-3(1), pp. 323–344, 1987.

[11] A. Azarbayejani and A. Pentland, "Camera selfcalibration from one point correspondence," Tech. Rep. 341, MIT Media Laboratory, 1995.

[12] J. J. Koenderink and A. J. Van Doorn, "New type of raster scan preserves the topology of the image," *Proc. IEEE*, vol. 67, no. 10, pp. 1465–1466, 1979.

[13] C. Zetzsche and E. Barth, "Fundamental limits of linear filters in the visual processing of two-dimensional signals," *Vision Res.*, vol. 30, pp. 1111–1117, 1990.

[14] E. Barth, C. Zetzsche, and G. Krieger, "Curvature measures in visual information processing," *Open Systems and Information Dynamics*, vol. 5, pp. 25–39, 1998.

[15] M. Pilu, A. W. Fitzgibbon, and R. B. Fisher, "Ellipse-specific direct least-square fitting," in *Proc. ICIP*, 1996, pp. 001–002.

[16] D. A. Forsyth and J. Ponce, *Computer vision: a modern approach*, Pearson Education, 2003.

[17] D. E. Rumelhart and G. E. Hinton and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986.