# VIDEO PROCESSING METHODS FOR IN-FLIGHT GAZE ANALYSIS

Jeffrey B. Mulligan
NASA Ames Research Center

In-flight gaze analysis is a tool for assessing the impact of new cockpit technologies on pilots' allocation of attentional resources. In particular, gaze tracking measures allow us to determine whether external scanning is sufficient to insure the pilot's ability to see-and-avoid traffic under VFR conditions. Commercial gaze-tracking solutions, however, do not currently provide adequate performance in the presence of high levels of ambient illumination, as encountered in clear sunny weather. This report describes novel methods developed for the analysis of data collected in a series of helicopter flight tests conducted in October, 2003.

## INTRODUCTION

Pilots flying under visual flight rules (VFR) are obligated to continuously monitor the surrounding airspace for other traffic, and maneuver as necessary to eliminate conflicts ("see-and-avoid"). Thus the introduction of any new device into the cockpit raises the question of how the use of the device may impact the pilot's allocation of visual and attentional resources. Our project specifically focusses on the use of global positioning system (GPS) receivers as navigational aids. We wish to determine both how access of the information provided by the display affects performance in a precision navigation task, and how it impacts other important functions such as see-and-avoid. The use navigational aids of this sort is of particular importance for helicopter operations such as medical evacuation, in which the pilot has to fly an unfamiliar route, possibly in close proximity to obstacles and other traffic.

To this end, a series of flight tests were conducted in October, 2003, in which four video streams were recorded. Two head-mounted cameras provided images of the pilot's eye, and the forward-looking view from the pilot's perspective, while two additional fixed cameras provided a frontal view of the pilot's head and shoulders, and an over-the-shoulder view which included the control stick. A complete description of the data collection procedures has been reported previously [1].

Our initial approach to extraction of gaze estimates from the video data was to apply techniques commonly applied to similar images obtained in the laboratory under controlled illumination conditions [2]. Unfortunately, these techniques proved unsatisfactory for the conditions encountered during the flight tests. Straightforward search for key features such as the illuminator reflexes ("glints") and the pupil boundary (inner iris margin) resulted in gaze estimates for approximately 70% of the frames in the night recordings, and less that 40% of the frames of the day recordings. The primary factor contributing to the poor performance with the day recordings was the high level of ambient illumination (sunlight) which swamped the controlled illumination provided by the goggle-mounted light-emitting diodes. Additionally, the high light levels caused most of the subject pilots to maintain their eyelids in a relatively closed position, often hiding the features normally used for gaze estimation.

In order to obtain precise gaze estimates for all of the images, we therefore embarked upon a program to develop a set of new methods specifically tailored to address these problems. Our approach consisted of the following elements: 1) development of interactive tools for hand-labelling of selected images; 2) development of a geometrical model of the eye, allowing accurate gaze estimation from a minimal set of features; 3) development of a clustering procedure for selecting minimal sets of exemplar images for hand labelling, which span the space of possible appearances; 4) development of interactive tools for registration and feature-labelling of images from the head mounted scene camera, necessary for transforming head-relative gaze estimates (obtained from the eye images) to an external world-referenced gaze target. In the following

sections, we describe each of these elements in more detail.

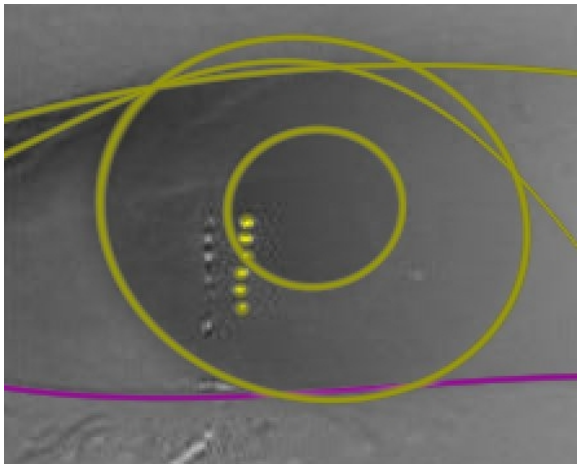

Figure 1: Typical eye image (from night flight) showing superimposed labels of eyelid and iris features.

## EYE AND LID LABELLING TOOL

The eye and lid labelling tool allows an operator to indicate the positions of the features of interest with a series of mouse clicks within a window displaying an enlarged image. The set of possible features consists of: 1) three fourth-order curves describing the lower eyelid margin, the upper eyelid margin, and the skin fold above the upper eyelid; 2) two ellipses desribing the inner and outer margins of the iris, referred to as the pupil and limbus, respectively; 3) six point locations describing the positions of the reflections of the LED illuminators. Additionally, check-boxes are provided allowing the operator to indicate the presence or absence of each feature in each image to be labelled. Figure 1 shows a typical image in which all the features are visible, along with the corresponding labels.

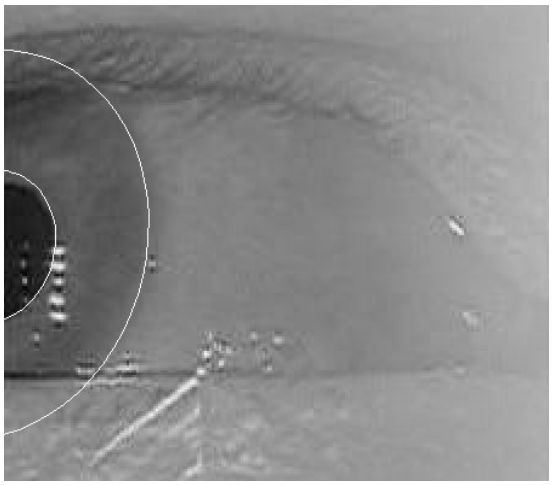

Figure 2: Eye image with superimposed labelling showing pupil/limbus model.

## GEOMETRICAL EYE MODEL

The labelling procedure described in the previous section allows the pupil and limbus to be described by ellipses which are completely independent. But because these features are part of a rigid mechanical system (the eye), they move together, and thus their projected shapes in the image are not free to vary independently, but are strongly constrained. These constraints may be exploited to obtain accurate estimates of gaze even when only a small portion of the pupil is visible in the image (as in figure 2).

We have implemented a model introduced by Ohno [3], in which the effects on the pupil image by refraction at the cornea are approximated by a change in apparent depth and size. The model has 3 structural parameters, which should be the same for all images obtained from a given subject: the limbus radius, the distance of the plane of the iris from the eye's center of rotation, and the difference in apparent depth between the pupil and the limbus. Two additional parameters are constant within a set of images obtained with a fixed position of the goggle: these are the position in the image of the center of the pupil and limbus when the eye is pointed directly at the camera, and the pupil and limbus appear as concentric circles. The corresponding viewing direction forms the origin of our gaze coordinate system.

Three additional parameters must be determined for each frame: the gaze angles, expressed as slant and tilt relative to the eye-camera axis, and the pupil radius (which varies slowly within limits). As slant increases, the pupil and limbus change in appearance from circles to ellipses; the major axis of the ellipse having a length equal to twice the relevant radius, while the minor axis is diminished by a factor equal to the cosine of the slant. If the pupil depth

diffference parameter is zero, then the ellipses will be concentric; conversely, the depth difference parameter can be adusted to account for non-concentric appearance at large gaze angles.

Several passes through the data are required to determine the fixed parameters: first we must determine the center coordinates. If the model is accurate, then all of the ellipse minor axes should interesect at the center point. In practice, the ellipses produced by the initial labelling will not have coincident minor axes, so we obtain a least-squares solution using the singular value decomposition on the matrix of line equations. Once the correct center has been found, then the shape of the limbus in the frames with large gaze deviations determines the distance of the limbus plane from the center of rotation. Finally, the offset of the pupil plane is readjusted in each frame. In each case, after labelling the individual frames, the mean is computed across frames, and this value is held fixed during subsequent iterations. Once the structural parameters have been determined, the variable parameters (gaze angles and pupil radius) can be set quickly and easily.

## IMAGE CLUSTERING

While the hand-labelling procedures described in the previous sections require only a minute or so per frame, when the number of frames is large it is impractical to hand label them all. For example, the Tullahoma flight test data set consists of 15 recordings of approximately 100,000 frames each. Fortunately, many of the frames are roughly similar; because typical gaze behaviors consist of fixational eye movements, we often encounter runs of 10 or more similar frames corresponding to a fixation. Furthermore, because gaze repetitively returns to certain targets such as the cockpit instruments, we find large sets of similar frames in the complete recordings. The purpose of the image clustering procedure is to form an efficient hierarchical representation of this strucure.

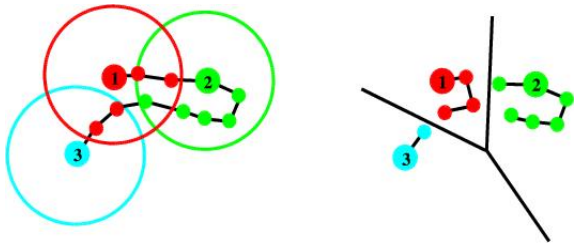


Figure 3: Two-dimensional cartoon illustrating selection of catalog exemplars and nearest neighbors. Numbered disks represent exemplar images, a new exemplar is added to the catalog when the distance of a new image from existing exemplars exceeds a threshold, indicated by the large circles.

The procedure we have adopted combines elements of vector quantization [4} and nearest-neighbor classifiers [5]. Here we present a brief overview of the procedure; a more thorough treatment is provided elsewhere [6]. We assume the existence of a metric which provides us with a measure of "distance" between two images. (We use a metric based on correlation, but the following discussion does not depend upon the choice of metric.) We treat the images as points in a high-dimensional space; the number of dimensions is potentially as large as the number of pixels, but for the restricted class of images that we are dealing with the images all lie within a manifold whose dimension is considerably lower. In figure 3, we represent the images as points in a two-dimensional plane for illustration purposes only.

We begin by choosing a threshold distance. Our goal is to come up with a minimal set of exemplar images, chosen from the input sequence, such that each exemplar differs from every other exemplar by at least the threshold distance, but every other non-exemplar image is within the threshold distance of the nearest exemplar. The catalog of exemplars is formed as follows: the first image in the sequence is

the first catalog entry. As we proceeed sequentially through the sequence, each image in the sequence is then tested against the exemplar associated with the previous frame. If the distance is below the threshold, then we proceed to the next frame. Otherwise, we test the image against the remaining catalog entries, stopping when we find one whose distance from the input is less than the threshold. If no catalog entry is found within the required distance of the input image, then the input image is added to the catalog. After the catalog has been generated, a second pass over the data is performed in which each image is associated with its nearest neighbor in the catalog. This process is illustrated in figure 3. Rather than process the entire sequence with a small threshold, we begin with a large threshold resulting in a small number of exemplars, and then apply the process recursively to the resulting neighborhoods, resulting in a tree in which the exemplars at each levels form the nodes. As we descend the tree, the images in each neighborhood become more and more similar; at some point we expect that this similarity will be high enough that an automatic labelling procedure, initialized with the values of a hand-labelled exemplar, will be able to successfully label the remaining images.

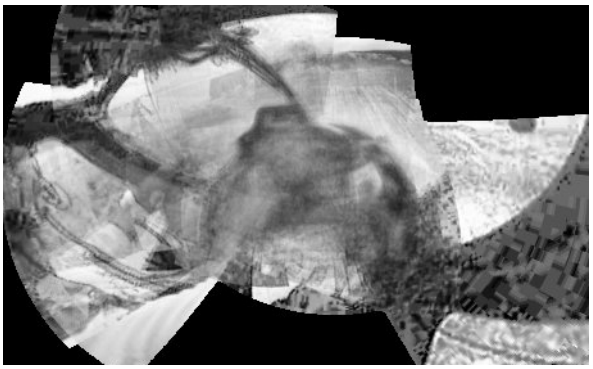

Figure 4: Cockpit mosaic image created by merging hand-aligned exemplar images.

## SCENE LABELLING TOOL

Analysis of the eye images as described above tells us the direction of gaze relative to the head; similarly, the head-mounted scene camera provides us with a head-relative view of the world; each pixel in the scene camera image corresponds to unique direction of head-relative gaze. Thus, once we have registered the scene camera image to a model of the world, we can relate the gaze computed from the eye image to an external target specified in world-coordinates.

Initially, we make the assumption that translations of the scene camera are small compared to the distance to the objects being imaged, so that we can ignore the effects of parallax, and model the appearance of the cockpit by mosaicking images from the scene camera on a sphere. We have developed a tool allowing an operator to manually register an image to another image or the complete mosaic by manipulating sliders controlling the three rotation angles (pan, tilt, and roll). This is accomplished by first computing the angles associated with each pixel in the scene camera image (which depends only on the focal length). These angles are then transformed according to the operator-selected parameters. An entire hemisphere of viewing directions is mapped into an image for viewing using stereographic projection (see figure 4).

While it is possible to obtain a reasonable looking mosaic in this way, individual features are often misaligned. This can be for two reasons: first, our assumption of zero parallax is clearly false; in addition, the focal length of the camera is initially uncalibrated. Both of these issues are ones which we ultimately hope to deal with in the correct manner, but in order to do so we need to have the coordinates of individually-labelled features. Thus our labelling tool also incorporates a feature editor. To add a new feature, the operator first clicks on its location in the mosaic image. The tool then automatically generates the list of frames which should contain that feature, based on the angles used to register each frame to the mosaic. These images are presented to the user in a second window, where (s)he indicates the precise location with another mouse click.

## SUMMARY AND CONCLUSIONS

We have described a number of new tools developed to aid the analysis of in-flight gaze recordings. Currently, a few thousand images from the Tullahoma flight tests have been hand-labelled, which allow us to directly estimate gaze, but with a low precision. Because of the fact that the GPS reciever was mounted at the top of the instrument cluster in the test vehicle (i.e., at the boundary of the windscreen), we need a high degree of precision to discriminate fixations on the receiver from out-the-window scanning. Thus our next step will be to develop procedures to use the hand-labelled data to automatically label the remaining images.

## REFERENCES


1. Mulligan, J. B. (2005). Pilot Behavior and Course Deviations during Precision Flight, in Rogowitz, B. E., Pappas, T. N., and Daly, S. J., (eds.), *Human Vision and Electronic Imaging X*, Proc SPIE vol 5666, pp. 363-373.

2. Mulligan, J. B., (1997). Image Processing for Improved Eye-tracking Accuracy. *Behavioral Research Methods, Instrumentation and Computers*, vol. 29, pp. 54-65.

3. Ohno, T., Mukawa, N., and Yoshikawa, A. (2002). FreeGaze: a Gaze Tracking System for Everyday Gaze Interaction, in Duchowski, A. T., Vertegaal, R., and Senders, J. W. (eds.), *Proc. ACM Symposium on Eye Tracking Research and Applications (ETRA)*, pp. 125-132.

4. Gersho, A. and Gray, R. M. (1992). *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Norwell MA.

5. Cover, T. M. and Hart, P. E. (1967). Nearest Neighbor Pattern Classification, *IEEE Transactions on Information Theory*, vol. IT-13, pp. 21-27.

6. Mulligan, J. B. (in press). A Tree-structured Model of Visual Appearance Applied to Gaze Tracking. Proc. 2005 IEEE International Symposium on Visual Computing.