Eye-Movement Tracking Using Compressed Video Images

Jeffrey B. Mulligan and Brent R. Beutter
Mail Stop 262-2
NASA Ames Research Center
Moffett Field, CA, 94035-1000

Introduction

Infrared video cameras offer a simple noninvasive way to measure the position of the eyes using relatively inexpensive equipment. Several commercial systems are available which use special hardware to localize features in the image in real time; typically the pupil and the first Purkinje image (corneal reflex) are tracked, the difference between these two signals giving a measure of eye rotation which is relatively independent of head position. While these systems have the advantage of providing a result in real time, this speed is obtained at the cost of reduced resolution and accuracy. The limitation is not imposed by the information content of the video image, however, but rather by the complexity of the processing algorithms which can be implemented in current hardware. More accurate results can be obtained when the imagery is analyzed off-line using more complex algorithms implemented in software.

When the approach of off-line image analysis is adopted, the major technical challenge becomes real-time acquisition and storage of the video images. Although there are a number of frame-grabber products which are capable of digitizing video images at normal frame rates, the bottleneck is transfer of data from the frame buffer to disk storage. (Note the the data rate for normal broadcast quality video is approximately 7.5 megabytes per second.) There are products which store digitized video to disk in real-time by using an array of disks in parallel, but these systems are expensive, and at the low end have a capacity of only a minute or two. Analog recording to tape allows the problem to be deferred, but extraction of individual images from tape involves starting and stopping the tape for every frame to be digitized and is undesirable for a production system. Analog video disk recorders are available which overcome this problem, but these products do not satisfy our goal of low-cost.

To solve the problem of the storage bottleneck, we have pursued a strictly digital approach, exploiting the burgeoning field of hardware video compression. Between the time that we began this work and the present, a number of new products have appeared; if 2:1 spatial downsampling is acceptable, there are numerous products available for the PC which cost less than $1000. The system we are using (XVideo, Parallax Graphics, Santa Clara, CA), was one of the early offerings, and not surprisingly was more expensive. We expect to see continued rapid development in this area, as the market for such multimedia products is very large. This product implements the standard defined by the Joint Photographic Experts Group (JPEG); for an overview see Pennebaker and Mitchell (1993). The compression process reduces the amount of data needed to represent an image by between 1 and 2 orders of magnitude, allowing us to store a continuous stream of video images to a normal computer disk. When we wish to view or analyze one of the images, we must decompress the stored representation. Unfortunately, the image obtained is not identical to the original; we refer to this as "lossy" compression, indicating that some information is sacrificed.

In this paper we will describe the algorithms we have developed for tracking the movements of the eyes in video images, and present experimental results showing how the accuracy is affected by the degree of video compression. Space does not permit full specification of the algorithms, but we will attempt to present the basic ideas. We have worked with two imaging arrangements: the first was patterned after the pupil-tracking systems described above; additionally, we have constructed a simple video ophthalmoscope for acquiring images of the fundus. Imaging the fundus has several advantages for eye-tracking: first, it is relatively insensitive to head movement artifacts; although the head must be stabilized to keep the illumination and measuring beams aligned with the pupil, small movements of the head will not produce artifactual eye movement signals. Secondly, greater resolution of eye position is possible than with pupil images. When tracking the pupil, the resolution is limited by the magnification of the pupil image, which cannot be increased beyond the point at which the pupil fills the frame. The magnification of fundus images is limited only by the quality of the image and the presence of details which can be tracked in the camera’s field of view.

Pupil Tracking Algorithm

Our analysis of pupil images consists of several steps. First the pupil is crudely located, and a region of interest slightly larger than the pupil is selected for further processing. This is done to eliminate artifacts due to irrelevant portions of the image, such as the eyelids. Next the corneal reflex is localized using a matched filter for coarse localization; the position is computed with sub-pixel accuracy by computing an intensity-weighted centroid. The corneal reflex is then removed from the image of the pupil by painting over it with the mean pupil value. The pupil is then localized by low-pass filtering, thresholding, and a centroid calculation.

To assess the performance of our algorithm, we constructed a synthetic sequence of images from a single image of an eye. Subpixel displacements of the image were performed by adding small phase shifts to the Fourier transform of the image, and then back-transforming. A linearly increasing sequence of displacements was generated, going from 0 to 2 pixels in steps of 0.1 pixel. The algorithm was then run on each of the translated copies. The estimated feature positions were plotted against actual displacement, and fit with a line of unit slope, which we used as a reference in the absence of ground-truth data as to the absolute positions of the features. The deviations of the estimates from this line provide a baseline measure of the performance of the algorithm.

The synthetic image sequence was then encoded with varying levels of compression. After compression, the images were then decoded or decompressed and passed back to the feature localization program. Errors were computed relative to the fit to the uncompressed data. For these simulations, the compression factor was controlled by a single parameter (called the "Q factor") which was passed directly to the vendor-supplied software. This number is used by the software to scale the matrix of quantization parameters for each component in the discrete cosine transform (DCT). In future work we plan to investigate the optimal matrix of coefficients for our application, but for the present study we contented ourselves with the default matrix. For a given setting of the Q factor, the actual compression rate obtained depends on the content of the image, so we computed the average compression factor over the images in the sequence. Average error magnitude is plotted as a function of compression factor in figure 1(a).

Figure 1: Average error magnitude of the tracking algorithms is shown as a function of compression factor. Compression factor of 1 corresponds to no compression and represents our best results to date from the current algorithms. Left panel: results for pupil-tracking algorithm. Squares represent errors of pupil position, circles errors of corneal reflex position. Filled symbols indicate horizontal errors, open symbols vertical errors. Right panel: results for fundus tracking algorithm. Filled symbols represent horizontal errors, open symbols vertical errors.

Fundus tracking algorithm

Although pupil-tracking is a sensible technique for many applications, we desired a more precise method for measuring small eye movements in our experimental work. To this end we constructed a simple ophthalmoscope which allows us to obtain video images of the fundus. The current configuration has a field of view of 10 degrees; given the camera resolution of 240 lines per field, this corresponds to slightly more than 2 minutes of arc of eye rotation per pixel of image displacement, a factor of 5 better than our pupil imaging set-up. The relative position at each frame is determined by cross-correlating the current image with a template of the fundus. (For small movements the template might be the first image of the sequence, to handle larger movements it is necessary to construct a larger template by compositing a number of images.) The cross-correlation image is searched for the pixel having the maximum value; the location of this pixel indicates the displacement with which the best match between the input and the template is obtained. We have obtained sub-pixel resolution by using biquadratic interpolation with the values of the 8 nearest-neighbor pixels, which improves performance but is biased if the template has an asymmetrical autocorrelation function. We have also found that the best results are obtained when the input is pre-filtered to accentuate the features to be matched (the major blood vessels), while removing high frequencies arising primarily from camera noise, and low frequencies resulting from illumination variations. The performance of this algorithm at several compression levels is shown in figure 1(b).

Conclusions

We have developed algorithms for the analysis of eye position from video images with a resolution in excess of of 1 minute of arc. Fundus tracking offers higher resolution and accuracy than pupil tracking, but requires a more complicated optical set-up, and in our implementation produces images of lower quality, particularly when infrared wavelengths are used. In this paper we have restricted our investigations to the effects of JPEG compression artifacts on the performance of the algorithms. Other factors we have not considered, such as camera noise, may degrade performance somewhat.

While not always the optimal choice, eye-movement tracking using video images is a low cost approach which is capable of high performance when off-line data analysis is acceptable. New video compression technology allows streams of video images to be acquired and stored on normal computer system disks. Our results show that at low compression rates, tracking accuracy is only slightly degraded, and is still significantly better than that of current real-time systems.

Reference

Pennebaker, W. B., and Mitchell, J. L., (1993). *JPEG Still Image Data Compression Standard.* Van Nostrand Reinhold, New York.