USING INFRARED ILLUMINATION TO IMPROVE EYE & FACE TRACKING IN LOW QUALITY VIDEO IMAGES

R. Youmaran and A. Adler
Carleton University, Ottawa, Ontario
adler@sce.carleton.ca

ABSTRACT

We propose a novel eye and face tracking algorithm using active infrared (IR) illumination. Most eye trackers based on active IR illumination require bright pupil images to successfully detect eyes in image sequences. However, due to factors such as eye closure, head rotation, variation in illumination and occlusion, most trackers tend to fail in these situations where pupil shows weak reflections. Our proposed method overcomes these limitations by making use of the dark-bright pupil difference images as well as using adaptive thresholding techniques. The core computational module of the algorithm is based on the Kalman filter and adaptive template matching to find and update the most probable eye position in the current frame. Our eye tracker can robustly detect faces and track eyes in a sequence of images under variable lighting conditions and face orientations. Experiments show good performance in challenging image sequences with low quality and occluded images with the subject showing considerable head movements.

1. INTRODUCTION

Biometric systems identify people based on physiological or behavioral characteristics, such as voice, handprint or facial characteristics. The use of biometrics as a way to authenticate user's identities has been a topic of research for years. Eye tracking and face detection and recognition are important parts of any biometric system. There is a need for fully automated systems that are robust and efficient face and eye detection algorithms are required [4]. Given a single image or a sequence of images, the goal of face detection is to identify all image regions which contain a face regardless of its three-dimensional position, orientation and lighting conditions. Such a problem is challenging because faces and eyes are non-rigid and have a high degree of variability in size, shape, color, and texture. The ability to detect faces and eyes in a scene is critical to modern surveillance applications. While many image processing algorithms exist to detect faces in images [4], their performance is not completely reliable, especially in situations with variable lighting, and when dealing with low resolution images.

Many different techniques have been used to detect eyes in images obtained under active IR illumination. For example, appearance-based methods are developed in [6, 10] using the bright pupil effect and its motion characteristics to track the eye regions. However, these methods fail when eyes are closed or occluded and subject show rapid head movement. Another method based on the Hough transform is presented in [9] to detect the eye region. This technique is time-consuming and requires high quality eye images with a good contrast to succeed in the detection process. Other papers developed methods using Kalman filtering and the mean shift tracker [2, 12] to detect and track eyes in an image. However, the latter might fail when applied on low quality images for eye tracking.

In this paper, we explore a new robust method for eye tracking and face detection in low quality images using the" Red-Eye" effect — specular reflection from the retina of human eyes under co-axial infrared illumination ([5], [7]). By using IR illumination, it is possible to get information from which the eye positions in the image can be calculated. Our algorithm consists of three parts: face localization, eye detection and eye tracking. This is accomplished using traditional image-based passive techniques such as shape information of the eye and active based methods which exploit the spectral properties of the pupil under IR illumination. A frame differential template-based technique [3] and a feature-based principal component analysis method are used to search the image for valid eye regions. This will then follows with Kalman filtering to locate the bright pupil candidate of interest. If the processed image contains weak reflections due to occlusion or eye closure, the algorithm uses the contour and shape information of the eye to approximate the pupil location. This is achieved by using adaptive thresholding techniques to extract the eye contour [8].

2. ALGORITHM DESIGN

This section develops an algorithm to automate the detection of the face and eyes in images taken under low illumination with ON and OFF IR. The proposed algorithm is designed to detect and track the subject's eyes under challenging conditions. Using the computed eyes location, face detection in the subsequent frames can be easily accomplished.

2.1. Face detection

This section elaborates on eye and face detection using image enhancement, frame differencing and adaptive template correlation techniques. The initialization step is accomplished by using geometric and shape constraints of the eye to find the pupil regions in the enhanced difference image. The multi-stage enhancement algorithm and the initial eye detection techniques are presented in details in [11].

2.2 Eye and Face tracking

After computing the initial eye position (section 2.1), an eye and face detection algorithm is initiated for eye tracking. A Kalman filter is activated in order to track the motion bright pupils in subsequent frames. In case the images contain weak reflections or the subject exhibits rapid head movement, the Kalman filter might fail in tracking the exact location of the pupil which is then approximated using the extracted eye contour.

2.3 Template correlation

Using the initial eyes location, two eye templates are created for further matching. The templates size is chosen to be 40×40 which has shown to provide accurate tracking and correlation scores. Templates are updated after each frame using the previously computed pupil positions. The correlation scores are then calculated for every frame using the normalized correlation coefficient [1]:

$$C = \frac{\sum_{x,y}\left[f(x,y)-\bar{f}_{u,v}\right]\left[t(x-u,y-v)-\bar{t}\right]}{\sqrt{\sum_{x,y}\left[f(x,y)-\bar{f}_{u,v}\right]^2\sum_{x,y}\left[t(x-u,y-v)-\bar{t}\right]^2}} \quad (1)$$

where $f(x,y)$ is the intensity value of the video frame at point (x,y), $\bar{f}_{u,v}$ is the average value in the current search window, $t(x,y)$ is the intensity value in the template image at location (x,y), $\bar{t}$ is the mean value of the template image. In order to reduce computation time, this operation is not performed on the entire image but instead, only on the extracted face region found in the frame at time *t-1*. This computation gives correlation scores between -1 and 1 where 1 indicates a probable match between the search region of interest and the correlation template. This similarity measure has the advantage that it is insensitive to changes in lighting conditions which can then be used for images taken in low illumination.

2.4 Face detection using previous face template

In case eye regions are not detected possibly due to occlusions or head rotation, the algorithm must keep track of the face region. This is achieved by using previous face template extracted using preceding eye location. The intensity distribution of face region around the eyes is calculated in order to locate the exact contour of the face in the succeeding frames. Using the previously calculated face template at time *t-1* and the current dark pupil image, both images are downsampled prior to correlation in order to reduce the processing time. Afterwards, correlation scores are computed using equation 1. The sub-region with the highest correlation score (close to 1) is assumed to contain face information.

2.5 Pupil candidate regions computation and eye tracking

Once the eye and face regions are approximately located, the current bright and dark pupil images are preprocessed as described in section 2.1, and all possible eye candidate regions computed. The elimination of non-eye blobs is achieved by first imposing the same geometric and shape constraints in the selection process and then activating a multi-stage eye tracking module based on Kalman filtering and eye contour segmentation. The classification and tracking process is described as follows:

Step 1: Compute the centroid of all the remaining blobs in the binary image obtained using connected component analysis and mark them all as possible pupil candidates.
Step 2: Eliminate all regions with centroids located outside the face contour.
Step 3: Eliminate all regions with centroids located outside the eye regions computed through template matching. If all blobs in the binary image are eliminated, go back to step 2 and skip step 3. This might happen if the subject shows rapid head movement in which case all candidate regions will be located far from the prior eye location, outside the estimated boundary.
Step 4: Find the region that has the closest centroid to the prior eye location computed in images at instance *t-1*.
Step 5: After detecting possible pupil candidates, the kalman filter is then used to compute the exact location of the bright pupil in the difference image. In case none of the binary regions is classified as possible pupil candidate, go to step 6.
Step 6: If none of the blobs are classified as possible pupil candidates (probably because of weak reflection, eye closure or occlusion), an eye contour extractor is then activated to give a more accurate approximation of the pupil location.
Step 7: After computing the pupil coordinates, update the eye and face templates to process the successive frames.

2.6 Kalman filtering

Kalman filtering is applied on the thresholded difference image obtained using the dark and bright pupil images. The

Kalman tracker has proven to work well when the pupil shows bright reflections and when the subject's head is slowly moving. We start by characterizing the motion of the pupil in the eye image at each time instance by its position and velocity. The pixel position or centroid in the binary image is represented by (x_t, y_t) at time *t*. (u_t, v_t) represent the pixel velocity in the x and y directions, respectively. The system model (predictor) can then be written as follows:

$$x_{t+1} = \varphi x_t + w_t \qquad (2)$$

where the state vector at time *t* is $x_t = (x_t\ y_t\ u_t\ v_t)^T$, φ is the state transition matrix and w_t is the system perturbation. w_t is normally distributed as $p(w_t) \sim N(0, Q)$ where Q represents the process noise covariance. Using our feature extractor estimates z_t of the pupil position in each frame, the measurement model (updater) can be written as follows:

$$z_t = H x_t + r_t \qquad (3)$$

where H is a matrix relating the current state information to the current measurement and r_t represents the measurement uncertainty and is normally distributed as $p(r_t) \sim N(0, R)$ where R is the measurement noise covariance. The system and measurement models are used to provide an estimate of the state and its covariance matrix at time *t+1*. In this experiment $H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$ since z_t deals with the pupil position information only. Assuming a bright pupil effect, the estimate of the pupil position obtained using the geometric and shape constraints filters is combined with the Kalman estimate to provide us with the final pupil position.

2.7 Eye contour extractor

In case where the subject's eyes show weak reflections or are occluded, the Kalman filter will not predict the correct pupil location in the difference image. For this reason, a second way of approximating the pupil location was developed to complement the first approach. The selected method processes the possible eye regions and automatically detects the eyelids using an adaptive thresholding technique described in [8]. Global thresholding methods fail in extracting the eye contour since the eye template does not necessarily have a bi-modal distribution where an optimal threshold can be easily selected to separate the desired region and the foreground. For this reason, we use a local adaptive thresholding technique that selects a unique threshold based on the local neighborhood of each sub-region in the image where illumination is assumed to be uniform. Local thresholds are chosen using the following equation:

$$T(i, j) = m(i, j) + k\,\sigma(i, j) \qquad (4)$$

where $m(i, j)$ and $\sigma(i, j)$ are the local mean and variance of the corresponding local region. In this experiment, a 10×10 window size is chosen with $k = -0.2$. This operation allows thresholding images that do not have an intensity histogram with two major peaks and that might contain strong and non-uniform illumination.

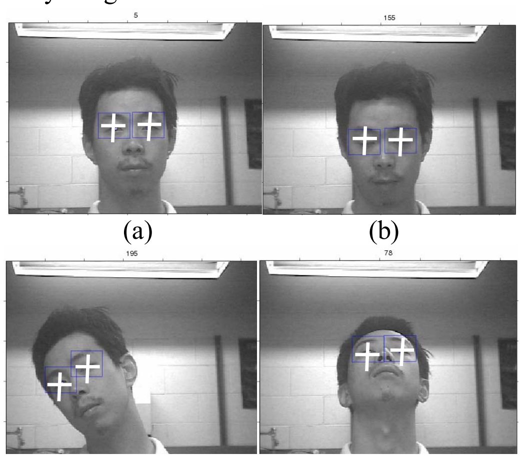
This method shows good contour extraction in complex eye images where eye closing, rotation and head movement can arise. Once the eyelids are extracted, the pupil position is then assumed to be located at the center of this region. A distance minimization method is then applied to select the candidate blob with its centroid located the closest to the estimated pupil location using the eye contour and that is nearby the prior pupil location computed in frame at time *t-1*. Once these conditions are satisfied, the binary blob is then marked as the possible pupil location in order to update the eye template model in the subsequent frame.

3. EXPERIMENTAL RESULTS

The algorithm was tested on two different eye image databases. The first set is composed of the low quality images and the other is the RPI ISL IR Eye Database [13]. The use of these images to test the proposed method demonstrates its robustness to the real-world environment. The image sequences are taken with IR cameras. Each image sequence is decomposed into dark pupil image sequence and its corresponding bright pupil image sequence. We processed 1200 image sequences taken from the ISL database and 600 images from the low quality image set. The selected frames display various facial expressions, and show small and large head movements and long eye closure.

For the initialization part, the s_i parameters were set to [0.1, 0.1, 1, 1] in section 2.1. These choices were motivated by the requirement to reduce noise and to slightly brighten the image by shifting the histogram of the image towards higher pixel intensity values.
When the images show bright pupil reflection, the Kalman filter succeeds in tracking the pupil location in the difference image. However, when the subject shows significant head movement and out-of-plane head rotations, the bright pupil effect disappears causing the Kalman filter to fail. This will then activate the eyelid extractor in order to get a better approximation of the pupil location. Fig. 1 shows eye detection results in weak and strong reflection scenarios, respectively. Table 1 shows the detection and tracking results for the two sets of images. The proposed approach detects and tracks the eyes accurately. The detection rate is 100% using the ISL database and 98.5% using the low quality images. It is worth mentioning that some of the images in low quality image data set contained

only one eye due to a 90 degree out-of-plane head rotation. For this reason, the algorithm does not detect the eye region which is not appearing on the image. However, the proposed method is able to re-locate the eye target region once it re-appears in the subsequent frames. The average and standard deviation pixel offset error in the horizontal and vertical direction are shown in table 2. This illustrates the efficiency and accuracy of our method since in general, an average pixel offset error <5 pixels means that the detected pixel position is still located within the pupil region that has a diameter of approximately 5 pixels in both the ISL and the low quality images databases.



(a) (b) (c) (d)

Figure 1: Eye detection results obtained using images from the ISL IR EYE database. Frames (b,d) show eye detection under challenging conditions where the subject has his eyes closed.

Database	Subjects	Images	Detection results
ISL	4	1200	1200/1200
Under-exposed	3	600	591/600

Table 1: Table showing eye detection and tracking results. 1800 images are processed in total using two different databases.

Data	Hor. μ_{error}	Hor. σ_{error}	Ver. μ_{error}	Ver. σ_{error}
ISL	1.098	0.909	2.2352	1.9794
Under-exposed	1.440	1.329	3.0829	1.9628

Table 2: Table showing average pixel offset error and standard deviation in the horizontal and vertical direction.

4. DISCUSSION AND CONCLUSION

This paper presents a new algorithm to extract and track face and eye positions from surveillance type images with IR strobe taken under poor illumination. In the case where many reflections (blobs) occur, the algorithm will find all possible eye locations and presents the best solution using multi-stage classification techniques. In order to reduce the set of possible eye region candidates, shape and geometric constraints are imposed in the classification process. A Kalman tracker is used to approximate eye location in bright pupil images. If the image contains weak reflections, a local adaptive thresholding technique is used to extract the eye contour in order to estimate the pupil location. This improves the performance and accuracy of the system when dealing with faces at different orientation and with eye closure. The algorithm achieves a 99.5% detection rate using 1800 images taken from two different IR image databases.

5. REFERENCES

[1] Chau, M., Betke, M., "Real time eye tracking and blink detection with USB cameras", Technical Report, Boston University Computer Science, 2005.

[2] Comaniciu,D., Ramesh, V., Meer, P., "Real-time tracking of non-rigid objects using mean shift", *in: IEEE Conf. on Computer Vision and Pattern Recognition*, Hilton Head Island, South Carolina, 2000.

[3] Ebisawa, Y., Satoh, S., "Effectiveness of pupil area detection technique using two light sources and image difference method", *in: Proceedings of the 15th Annual Int. Conf. of the IEEE Eng. in Medicine and Biology Society*, San Diego, CA, 1268–1269, 1993.

[4] Fromherz, T., Stucki, P., Bichsel, M., "A Survey of Face Recognition", MML Technical Report, No 97.01,Dept. of Computer Science,Zurich, 1997.

[5] Haro, A. ,Flickner, M., Irfan, E., "Detecting and Tracking Eyes By Using Their Physiological Properties, Dynamics, and Appearance" Computer Vision and Pattern Recognition Conference, June 13-15, 2000.

[6] Huang, W., Mariani, R., "Face detection and precise eyes location", *in: Proc. Int. Conf. on Pattern Recognition,* 2000.

[7] Morimoto, C.H. ,Koons, D.,Amir, A.,Flickner,M., "Pupil Detection and Tracking Using Multiple Light Sources" Technical Report RJ-10117,IBM Almaden Research Center, San Jose, Ca, 1998. http://domino.watson.ibm.com/library/cyberdig.nsf/Home.

[8] Niblack, W., "An Introduction to Image Processing", Prentice-Hall, Englewood Cliffs, NJ, 115–116, 1986.

[9] Nixon, M., "Eye spacing measurement for facial recognition", *in: Proc. of the Society of Photo-Optical Instrument Engineers*, 1985.

[10] Pentland, A., Moghaddam, B., Starner, T., "View-based and modular eigenspaces for face recognition", *in: Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Seattle, WA, 1994.

[11] Youmaran, R., Adler, A., "Using Red-Eye to Improve Face Detection in Low Quality Video Images", CCECE, 1940-1943, 2006

[12] Zhu, Z., Ji, Q., "Robust real-time eye detection and tracking under variable lighting conditions and various face orientations", *Computer Vision and Image Understanding*, No. 1, 124-154, April 2005.

[13] ISL IR Eye Image Database, *http://www.ecse.rpi.edu/~cvrl/database/ISL_IR_Eye_Database.htm*