Combining Kalman Filtering and Mean Shift for Real Time Eye Tracking Under Active IR Illumination

Zhiwei Zhu
Department of CS, UNR
zhu_z@cs.unr.edu

Qiang Ji
Department of ECSE, RPI
qji@ecse.rpi.edu

Kikuo Fujimura
Honda R & D Americas Inc., CA
kfujimura@honda.hra.com

Kuangchih Lee
Department of CS, UIUC
klee10@uiuc.edu

Abstract

Most eye trackers based on active IR illumination require distinctive bright pupil effect to work well. However, due to a variety of factors such as eye closure, eye occlusion, and external illuminations interference, pupils are not bright enough for these methods to work well. This tends to significantly limit their scope of application. In this paper, we present a new real time eye tracking methodology that works under variable and realistic lighting conditions and various face orientations. By combining the conventional appearance based object recognition method (SVM) and object tracking method (mean shift) with Kalman filtering based on active IR illumination, our technique is able to benefit from the strengths of different techniques and overcome their respective limitations. Experimental study shows significant improvement of our technique over the existing techniques.

1 Introduction

Eye tracking based on the active remote IR illuminations is a simple and effective approach. It utilizes the special bright pupil effect under IR to detect and track the eyes. Several active IR based eye trackers [1, 2, 3, 4] were proposed, but most of them require distinctive bright pupil effect to work well because they all track the eyes by tracking the bright pupils. The success of such a system strongly depends on the brightness and size of the pupils, which are often function of eye closure, face orientations, external illumination interferences, and the distances of the subjects to the camera.

To alleviate some of these problems, Haro et al [3] proposed pupil tracking based on combining its appearance, the bright pupil effect, and motion characteristics. Ji et al [4] proposed real time subtraction and a special filter to eliminate the external light interferences. Both methods fail to track eyes when they are closed or occluded and robustness is easily affected by the external illuminations interference. In this paper, we propose a new real time eye tracking method based on a combination of the traditional appearance based tracking method and the bright pupil effect due to active IR illumination. Our method consists of two major modules. The first module is a kalman filter based bright pupil tracking, augmented with the *support vector machine* classifier [5, 6] for pupils verification. If the first module fails, then we activate the second module based on *mean shift tracking* [7, 8, 9] to continue eyes tracking. Two modules alternate during tracking to complement each other.

2 Eye Detection

To obtain the desired bright pupil effect, we build an IR illuminator that illuminates a person's face and use an IR sensitive camera to acquire image. The illuminator consists of two concentric rings of IR LEDs as shown in Fig. 1 (a). They are turned on/off alternately to produce the bright/dark pupils effect as shown in Figure 1 (b) and (c).

We have developed a circuitry to synchronize the outer ring of LEDs and inner ring of LEDs with the even and odd fields of the interlaced image respectively so that they can be turned on and off alternately. The interlaced input image is subsequently deinterlaced via a video decoder, producing the even and odd field images as shown in Figure 1 (b) and (c). While both images share the same background and external illumination, pupils in the even images look significantly brighter than in the odd images. To eliminate the background and reduce external light illumination, the odd image is subtracted from the even image, producing the difference image as shown in Figure 1 (d), with most of the background and external illumination effects removed. The difference image is subsequently thresholded. A connected component analysis is then applied to the thresholded difference image to identify each binary blob.

Typically, pupils are found among the binary blobs. However, it is usually not possible to isolate pupil blob only by picking the right threshold value, since pupils are often small and not bright enough compared with other noise blobs. Thus, we will have to make use of information other than its brightness to correctly identify them. One way to distinguish the pupils blobs from other noise blobs is to use their geometric shapes. Usually, the pupil is an ellipse-like blob and we can use an ellipse fitting method to extract the shape of each blob and use the shape and size to remove some blobs from further consideration. For example, a blob with a large size or a large major-to-minor axis ratio should not be a pupil.
From the figure 1 (e), we observe that there are still several non pupil blobs left after binary blob elimination

1051-4651/02 $17.00 (c) 2002 IEEE

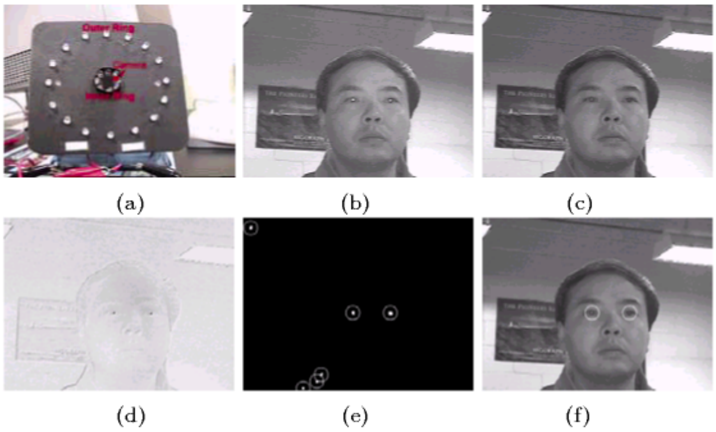


Figure 1. (a) Hardware setup: the camera with an active IR illuminator (b) bright pupils with even field image (c) dark pupils with odd field image (d)the difference image (e) The image marked with possible pupil candidates (f) The image marked with identified eyes

based on geometric properties because they are geometrically very similar to pupils. Thus we have to use other features. In the dark image, the region surrounding eyes have an unique intensity distribution. They appear different from other parts of the face. The appearance of an eye can therefore be utilized to separate it from non-eyes. We map the locations of the remaining binary blobs to the dark image and then apply the Support Vector Machine (SVM) classifier [5, 6] to automatically identify the binary blobs that correspond to eyes. We trained the SVM classifier using 558 eye image vectors and 560 non-eye image vectors and got a 95.5037% accuracy by choosing the Gaussian kernel whose σ term is 3 as the kernel of SVM. Now we crop the regions in the dark pupil image according to their locations in the difference image as shown in Figure 1 (e) and provide them to the trained SVM for classification. Figure 1 (f) shows that the SVM correctly choose the real eye regions as marked.

3 Eye Tracking Algorithm

We develop the following algorithm for the eye tracking by combining the bright pupil based Kalman filter eye tracker with the mean shift eye tracker. We call this two-stage eye tracking as outlined in Fig. 2. After locating the eyes in the initial frames, the Kalman filtering is activated to track bright pupils. If it fails in a frame due to disappearance of bright pupils, eye tracking based on mean shift will take over. We will return to bright pupil tracking as soon as bright pupil appears again since it is much more robust and reliable tracking. Pupil detection will be activated if the mean shift tracking fails. These two stage eye trackers work together and they complement each other. The robustness of the eye tracker is improved significantly. The Kalman filtering and mean shift tracking algorithm are briefly discussed below.

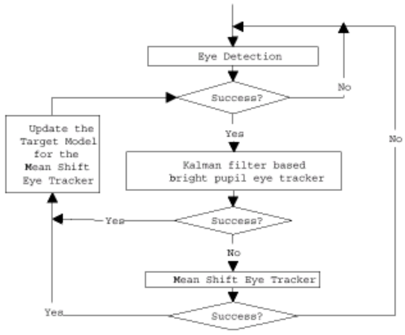


Figure 2. The Combined Eye Tracking Flowchart

3.1 Eye Tracking with Kalman Filtering

The motion of a pupil at each time instance (frame) can be characterized by its position and velocity. Let (c_t, r_t) represent the pupil pixel position (its centroid) at time t and (u_t, v_t) be its velocity at time t in c and r directions. The state vector at time t can therefore be represented as $x_t = (c_t \ r_t \ u_t \ v_t)^t$. The system can therefore be modelled as

$$\mathbf{x}_{t+1} = \phi \mathbf{x}_t + \mathbf{w_t} \tag{1}$$

where $\mathbf{w_t}$ represents system perturbation.

We further assume that a fast feature extractor estimates $\mathbf{z}_t = (\hat{c}_t, \hat{r}_t)$, the pupil position at time t. Therefore, the measurement model in the form needed by the Kalman filter is

$$\mathbf{z}_t = H\mathbf{x}_t + v_t \tag{2}$$

where v_t represents measurement uncertainty. Specifically, the position of current frame t, is estimated based on a simple local thresholding in the neighborhood of the predicted position (based on the system model), assuming the existence of the bright pupil effect. Given the state model in equation 1 and measurement model in equation 2 as well as some initial conditions, the state vector x_{t+1}, along with its covariance matrix Σ_{t+1}, can be updated using the system model (for prediction) and measurement model (for updating).

3.2 Mean Shift Eye Tracking

The mean shift tracking algorithm is an appearance based tracking method and it employs the mean shift iterations to find the target candidate that is the most similar to a given model in terms of intensity distribution, with the similarity of the two distributions being expressed by a metric based on the Bhattacharyya coefficient. The derivation of the Bhattacharyya coefficient from sample data involves the estimation of the target density q and the candidate density p, for which we employ the histogram formulation. At location y, the sample estimate of the Bhattacharyya coefficient for target density q and target candidate density $p(y)$ is given by

1051-4651/02 $17.00 (c) 2002 IEEE

$$\hat{\rho}(y) \equiv \rho\left[\hat{p(y)}, \hat{q}\right] = \sum_{u=1}^{m} \sqrt{\hat{p}_u \hat{q}_u} \quad (3)$$

where m is the quantization level for histograms p and q. The distance between two distributions can be defined as

$$d(y) = \sqrt{1 - \rho\left[\hat{p}(y), \hat{q}\right]} \quad (4)$$

To reliably characterize the intensity distribution of eyes and non-eyes, the intensity distribution is characterized by two images: even and odd fields, resulted from de-interlacing the original input images. They are under different illuminations, with one producing dark pupils and the other bright pupils as shown in Fig.3.

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

Figure 3. The eye images: (a)(b) left and right bright pupil eyes; (c)(d) corresponding left and right dark pupil eyes

Thus, there are two different feature probability distributions of the eye target corresponding to dark pupil and bright pupil images. We use a 2D joint histogram which is derived in the grey level dark pupil and bright pupil image space with $m \times m$ bins to represent the feature probability distribution of the eye target. Before calculating the histogram, we employ a convex and monotonic decreasing kernel profile k to assign a smaller weight to the locations that are farther from the center of the target. Let us denote by $\{x_i{}^*\}_{i=1\ldots n_h}$ the pixel locations of the target candidate, centered at y in the current frame. The probability of the intensity vector u in the target candidate is given by

$$\hat{p_u}(y) = \frac{\sum_{i=1}^{n_h} k(\|\frac{y-x_i}{h}\|^2)\delta[b(x_i) - u]}{\sum_{i=1}^{n_h} k(\|\frac{y-x_i}{h}\|^2)} \quad (5)$$

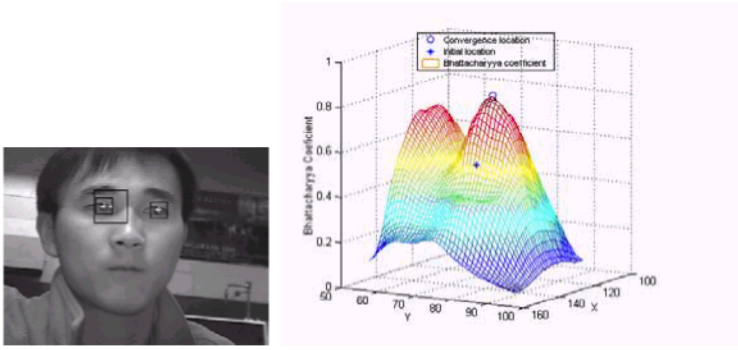
In which, the $b(x_i)$ is the index of the histogram bin, h is the radius of the kernel profile and δ is the Kronecker delta function. The target distribution q can be built in a similar fashion.

During tracking, we update the target eye model whenever the bright pupil tracker tracks the eyes successfully in order to reduce the error propagation, resulted from mean shift drifting.

Figure 4 (b) presents the surface obtained by computing the Bhattacharyya coefficient for the big rectangle marked in Figure 4 (a). The mean shift algorithm exploits the gradient of the surface to climb to the closest peak which represents the maximum value of the similarity measure.

3.2.1 Mean Shift Tracking Parameters

The mean shift algorithm is very sensitive to the window size and the histogram quantization value. In order to obtain the best performance of the mean shift tracker for a specific task, we have to find the proper histogram quantization value and the proper window size.



(a) (b)

Figure 4. (a) The image frame 13; (b) Values of Bhattacharyya coefficient corresponding to the marked region(40×40 pixels) around the left eye in frames 13. Mean shift algorithm converges from the initial location($*$) to the convergence point($\circ$) which is the mode of the surface.

We choose one image sequence which contains 100 frames and manually located the left eye positions in these frames. Then we run the mean shift eye tracker under different window sizes and different histogram quantization values, then evaluate the performance of mean shift eye tracker under those conditions using the following criterion:

$$\alpha_{error} = (\sum_{i=1}^{100} \sqrt{(y_i(tracked) - y_i'(manual))^2}/100 \quad (6)$$

In which, $y_i(tracked)$ is the left eye location tracked by mean shift tracker in the image frame i; $y_i'(manual)$ is the left eye location manually located by the person in the image frame i. We treat the manually selected eye locations as the correct left eye locations.

The results are plotted in Fig. 5, from which we can determine the optimal quantization level to be 2^5 while the optimal window size is 20 pixels.

4 Experiment Results

In order to show the reliability of our proposed eye detection and tracking method, we recorded two different persons' sequences of 20 second of video at 30 fps. They all had small and large head movements with out of plane rotation and quick and long eye closure. The first person performed slow head movements while the second person did fast head movements. Two persons all sat in front of the camera under strong overhead illumination.

We calculated the number of image frames corresponding to eyes open, closure, and occlusion. We noticed when bright pupils disappear due to either eye closure or oblique face orientations as shown in Fig. 6, the Kalman filter fails because there are no bright pupil blobs

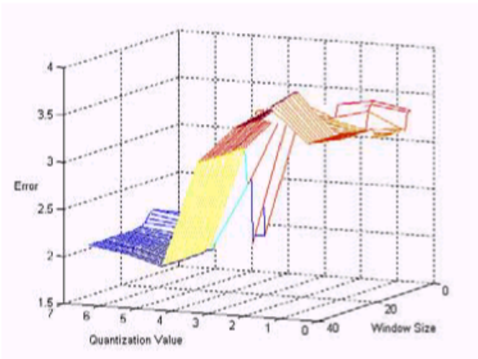


Figure 5. The error distribution of tracking results corresponding to different intensity quantization values and different window sizes

in the difference images. However the mean shift tracker compensates the failure of bright pupil tracker because it is an appearance based tracker which tracks the eye regions according to the color statistical distributions of the eye regions and doesn't need bright pupils. The black rectangles in Fig. 6 represent the tracked eye positions by the mean shift tracker.

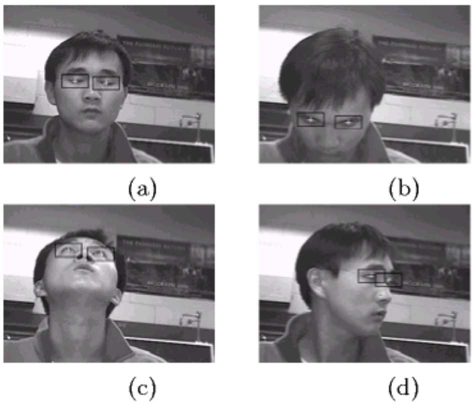


Figure 6. Bright pupil based Kalman filter tracker fails to track eyes due to absence of bright pupils caused by either eye closure or oblique face orientations. The mean shift eye tracker, however, tracks eyes successfully as indicated by the black rectangles.

The tracking statistics from the first person are summarized in table 7, from which we can conclude that our algorithm is more robust than the conventional kalman filter based bright pupil tracker, especially for the closed eyes and partially occluded eyes due to the face orientations. Even for open eyes under strong external illuminations, we have achieved good results. Video demos are available at *http://www.cs.unr.edu/~zhu_z/Demo/demo.html*.

5 Summary

In this paper, we have proposed a real-time eye tracker which can track eyes robustly under various illuminations and face orientations. Our method performs well regardless of whether the pupils are directly visible or not. This has been achieved by combining an appearance based pattern recognition method (SVM) and object tracking (mean shift) together with a bright pupil eye tracker based on Kalman filtering. The use of two channels (dark and bright images) for mean shift tracking as well as the experimental determination of the optimal window size and quantization level for mean shift tracking further enhance the performance of our technique. By experimental results, we have demonstrated that the proposed method dramatically improves the robustness and accuracy of eye tracking.

Image 600 frames	Bright pupil tracker	Combined tracker
Left eye(open) 452 frames	400/452	452/452
Left eye(closed) 66 frames	0/66	66/66
Left eye(occluded) 82 frames	0/82	82/82
Right eye(open) 425 frames	389/425	425/425
Right eye(closed) 66 frames	0/66	66/66
Right eye(occluded) 109 frames	0/109	109/109

Figure 7. Tracking statistics comparison for both trackers under different eyes conditions (open, closed, occluded)

6 Acknowledgments

This work is partially supported by a grant from AFOSR and a grant from Honda R&D Americas Inc., Mountain View, CA.

References

[1] Y. Ebisawa. Unconstrained pupil detection technique using two light sources and the image difference method. Visualization and Intelligent Design in engineering and architecture, pages 79-89, 1995.

[2] C.H. Morimoto and M. Flickner. Real-time multiple face detection using active illumination. Proc. of the 4th IEEE International Conference on Automatic Face and Gesture Recognition 2000, Grenoble, France, March 2000.

[3] A. Haro, M. Flickner, and I. Essa. Detecting and Tracking Eyes By Using Their Physiological Properties, Dynamics, and Appearance. Proceedings IEEE CVPR 2000, Hilton Head Island, South Carolina, June 2000.

[4] Q. Ji and X. Yang, Real time visual cues extraction for monitoring driver vigilance, In Proc. of International Workshop on Computer Vision Systems, July 7-8, 2001, Vancouver, Canada.

[5] C. Cortes and V. Vapnik. Support-vector networks. Machine Learning,20:273-297, 1995.

[6] Huang, J., D. Ii, X. Shao, and H. Wechsler. Pose Discrimination and Eye Detection Using Support Vector Machines (SVMs), Proceeding of NATO-ASI on Face Recognition: From Theory to Applications, pp.528-536, 1998.

[7] D. Comaniciu, V. Ramesh, P. Meer, Real-Time Tracking of Non-Rigid Objects using Mean Shift, IEEE Conf. on Comp. Vis. and Pat. Rec., Hilton Head Island, South Carolina, 2000.

[8] D. Comaniciu, V. Ramesh, Mean Shift and Optimal Prediction for Efficient Object Tracking, IEEE Int'l Conf. Image Processing, Vancouver, Canada, Vol. 3, 70-73, 2000.

[9] D. Comaniciu, V. Ramesh, Robust Detection and Tracking of Human Faces with an Active Camera, IEEE Int'l Workshop on Visual Surveillance, Dublin, Ireland, 11-18, 2000.

1051-4651/02 $17.00 (c) 2002 IEEE