Real-Time Eye Detection and Tracking Under Various Light Conditions

Zhiwei Zhu
Dept. of Computer Science
University of Nevada Reno
zhu_z@cs.unr.edu

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

Qiang Ji
Dept. of ECSE
Rensselaer Polytechnic Institute
qji@ecse.rpi.edu

Abstract

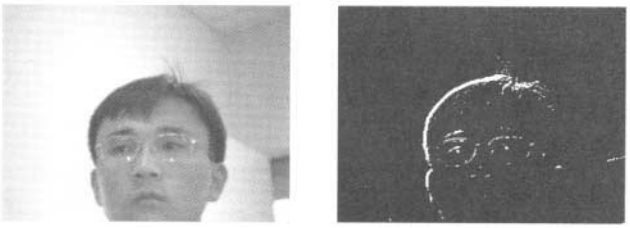
Non-intrusive methods based on active remote IR illumination for eye tracking are important for many applications of vision-based man-machine interaction. One problem that has plagued those methods is their sensitivity to lighting condition change. This tends to significantly limit their scope of application. In this paper, we present a new real-time eye detection and tracking methodology that works under variable and realistic lighting conditions. Based on combining the bright-pupil effect resulted from IR light and the conventional appearance-based object recognition technique, our method can robustly track eyes when the pupils are not very bright due to significant external illumination interferences. The appearance model is incorporated in both eyes detection and tracking via the use of support vector machine and the mean shift tracking. Additional improvement is achieved from modifying the image acquisition apparatus including the illuminator and the camera.

CR Categories: I.4.8 [IMAGE PROCESSING AND COMPUTER VISION]: Scene Analysis—Tracking;

Keywords: Eye Tracking, Support Vector Machine, Mean Shift, Kalman Filter

1 Introduction

Robust non-intrusive eye tracking is a crucial step for vision based man-machine interaction technology to be widely accepted in common environments such as homes and offices. There has been much work done in face and eye detection and tracking. The work can be classified into two categories: passive appearance-based methods [Baluja and Pomerleau 1994; Oliver et al. 1997; Smith et al. 2000] and the active IR based methods [Ebisawa and Satoh 1993; Morimoto et al. 1998; Morimoto and Flickner 2000; Haro et al. 2000; Ji and Yang 2001]. The former approaches detect eyes based on the intensity (or color) distribution of the eyes. The underlying assumption is that the eyes appear different from the rest of the face. Eyes can be detected and tracked based on exploiting the differences in appearance. This method usually needs to collect a large amount of training data representing the eyes of different subjects, under different face orientations, and different illumination conditions. These data are used to train a classifier such as a neural network and detection is achieved via classification. In fact, the well-known eigen-face approach [Turk and Pentland 1991] can simply be extended to eigen-eye for eye detection. The appearance based approach, while not requiring special illumination, requires a significant number of training data to enumerate all possible appearances of eyes since the appearance of the eyes will change dramatically under different illuminations.



(a) (b)

Figure 1: (a)original image, (b)the corresponding difference image

The eye tracking based on active IR illumination utilizes the special bright pupil effect. It's a simple and effective approach for pupil detection based on differential infrared lighting scheme. The high contrast between the pupils and the rest of the face can significantly improve the eye tracking robustness and accuracy. However, this technique is not without its shortcomings. The success of such a system strongly depends on the brightness and size of the pupils, which are often function of face orientations, external illumination interferences, and the distances of the subjects to the camera. Therefore, the most significant problems with this approach are that they require the lighting conditions to be relatively stable and the subjects close to the camera.

Realistically, however, lighting can be variable in many application domains and the view of the camera may include objects besides the subject. These pose serious challenges to this approach. Furthermore, even under normal indoors ambient lighting condition, the pupils may not look as bright under oblique face orientations. Finally, the thick eye glasses tend to disturb the infrared light so much that the pupils appear very weak. To alleviate some of these problems, Haro [Haro et al. 2000] proposed to perform pupils tracking based on combining its appearance, the bright pupil effect, and motion characteristics. Ji [Ji and Yang 2001] proposed real time subtraction and special filter to eliminate the external light interferences. However, the large and fast head movement tends to produce many very bright noises in the difference image as shown in Figure 1 and they cause difficulty for subsequent pupils detection based on intensity.

In this paper, we propose real-time robust methods for eye tracking under variable lighting conditions, based on combining the appearance-based method and the active IR illumination approach. Combining their respective strengths and overcoming their shortcomings, the proposed method uses active infrared illumination to brighten subject's faces to produce the bright pupil effect. The

bright pupil effect and appearance of eyes (statistic distribution based on eye patterns) are utilized for eyes detection and tracking. The latest technologies in pattern classification recognition (the support vector machine) and in object tracking (the mean-shift) are employed for pupil detection and tracking based on eyes appearance.

2 Image Acquisition

We use infrared LEDs that operate at a power of 32mW in a wavelength band 40nm wide at a nominal wavelength of 880nm. As in Ji [Ji and Yang 2001], we obtain a dark and a bright pupil image by illuminating the eyes with IR LEDs located off and on the optical axis, respectively. The pupil is detected from the difference of the two images (referred as the image difference method or the subtraction method). We use an optical band-pass filter which has a wavelength pass band only 10nm wide. The filter has increased the signal-to-noise ratio by a factor significantly (greater than 20) compared to the case without using the filter. The improvements in hardware apparatus include the use of IR LEDs with more powerful irradiance in a narrow optical bandwidth (40 nm), custom made IR camera with maximum spectral response around 880 nm, and the use of a band-pass filter that best match with the IR LEDs and the camera. Figure 2 illustrates the hardware setup. The new hardware

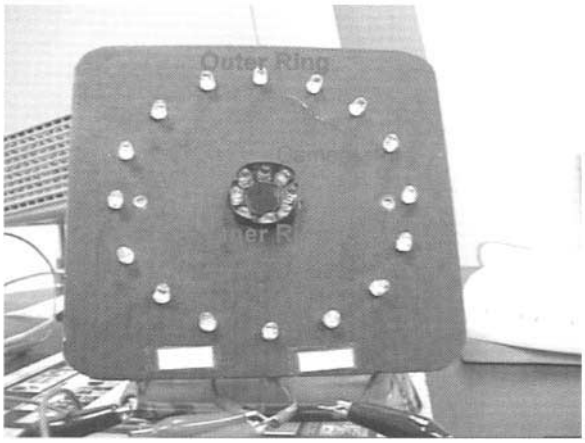
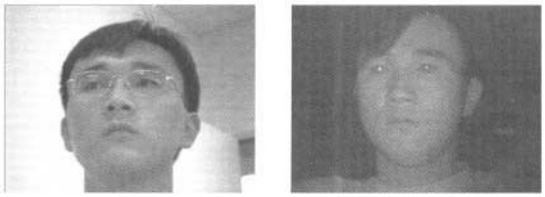


Figure 2: Hardware setup: the camera with an active IR illuminator

improvements lead to significant improvement in the quality of the images. Figure 3 illustrates the difference between two cases in which much of the ambient sources is shown to be suppressed by using the subtraction and the filter.

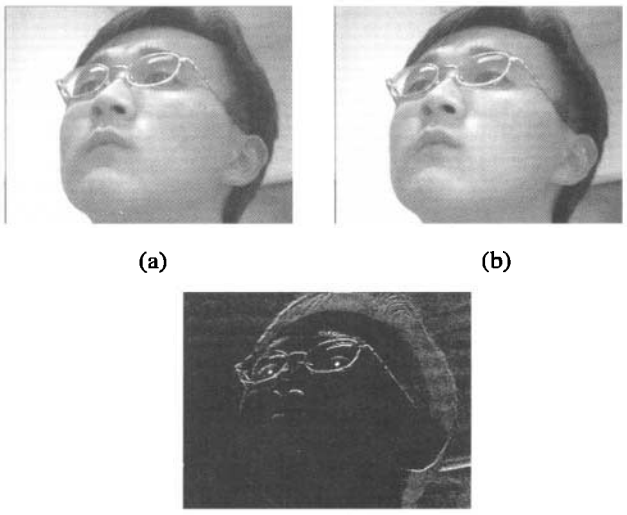


(a) (b)

Figure 3: Images (a),(b) captured before and after improvement

3 Pupil Detection

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. When the even field is being scanned, inner ring of LEDs is on and outer ring of LEDs is off and vice verse when the even filed is scanned. The interlaced input image is subsequently deinterlaced via a video decode, producing the even and odd field images as shown in Figure 4 (a) and (b). 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 4 (c), with most of the background and external illumination effects removed. The difference image is subsequently thresholded.



(a) (b)

(c)

Figure 4: (a) even field image (b) odd field image (c) the difference image

Once the subtraction (image differencing) operation and thresholding are done, we are left with a binary image which includes the pupil blobs and possibly other noise blobs. A connected component analysis is then applied to identify each binary blob. Our task is to find out which of the blobs actually contain the eyes. Initially, we mark all the blobs as potential candidates for pupils as shown in Figure 5.

Typically, pupils are found somewhere in these candidates. However, it is usually not possible to isolate eye 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 with other noise blobs is based on their geometric shapes. Usually, the pupil is an ellipse-like blob and we can use an ellipse fitting method [Fitzgibbon and Fisher 1995] 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 figure 6, we can see that there are still several non pupil blobs left because they are so similar in shape and size that we can't distinguish the real pupil from them. So we have to use other features. In the next step, we will use the Support Vector Machine clas-

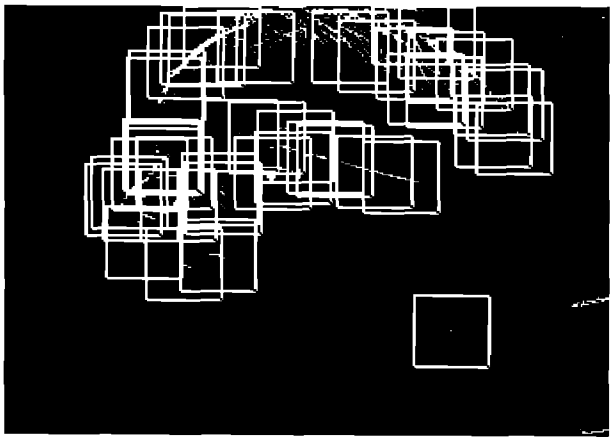


Figure 5: The thresholded difference image marked with pupil candidates

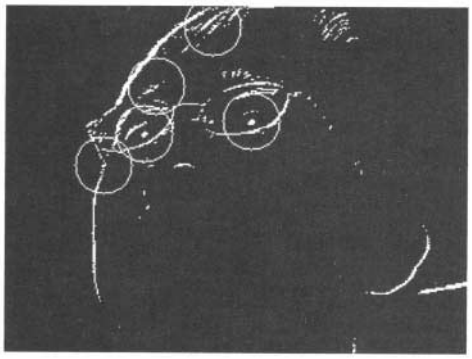


Figure 6: The thresholded difference image after removing some blobs based on their geometric properties (shape and size). The blobs marked with circles are selected for further consideration.

sifier [Cortes and Vapnik 1995; Vapnik 1995; Huang et al. 1998] to automatically identify the pupils.

3.1 Pupil Verification Using Support Vector Machine(SVM)

Ever since its introduction, Support Vector Machine (SVM) [Cortes and Vapnik 1995] has become increasingly popular. The theory of SVM can be briefly summarized as follows. For the case of two-class pattern recognition, the task of predictive learning from examples can be formulated as shown below. Given a set of functions f_α:

$$\{f_\alpha : \alpha \in \Lambda\}, f_\alpha : R^N \longrightarrow \{-1, +1\},$$

(Λ is an index set) and a set of l examples:

$$(x_1, y_1), \ldots (x_i, y_i), \ldots, (x_l, y_l), x_i \in R^N, y_i \in \{-1, +1\},$$

where x_i is a feature vector of N dimensions and y_i represents the class, which has only two values -1 and +1. Each one is generated from an unknown probability distribution $P(x,y)$, we want to find a particular function f_α^* which provides the smallest possible value for the risk:

$$R(\alpha) = \int |f_\alpha(x) - y| dP(x,y) \quad (1)$$

The SVM implementation seeks separating hyper-planes $D(X) = (w \cdot X + w_0)$ by mapping the input data X into a higher dimensional space Z using a nonlinear function g. The data points at the maximum margin are called the support vectors since they alone define the optimal hyper-plane.

Training data are needed to obtain the optimal hyper-plane. For this project, after obtaining the positions of pupil candidates using the methods mentioned above, we cut the images from the dark image according to those positions. Usually, the eyes are included in those cropped images. The size of image we use is 20×20 pixels and the image data are preprocessed using histogram equalization and normalized to a $[0, 1]$ range before training.

The eye training images were divided into two sets: positive set and negative set. In the positive image set, we include eye images of different gazes, different degrees of opening, different subjects, and with/without glasses. The non-eye images were placed in the negative image set. Figures 7 and 8 contain some examples of eye and non-eye images in the training sets, respectively.

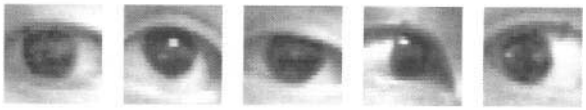


Figure 7: The eye images in the positive training set.

Figure 8: The non-eye images in the negative training set

3.1.1 Retraining Using Mis-labelled Data

Usually, supervised learning machines rely only on the limited labelled training examples and it can not reach very high learning accuracy. So we have to test on thousands of unlabelled data and pick up the mis-labelled data, then put them into the correct training sets and retrain the machine again. After doing this procedure on the unlabelled data obtained from different conditions several times, we can boost the accuracy of the learning machine at the cost of extra time needed for re-training.

In our experiment, we have eye data set from six people which are obtained using the same method. We choose the first person's data set and label the eye images and non-eye images manually, then we train the linear SVM on this training set and test linear SVM on the second person's data set. We check the second person's data one by one and pick up all the mis-labelled data by the linear SVM and label them correctly and add them into the training set.

After finishing the above step, we retrain the linear SVM on this increased training set and repeat the above step on the next person's data set. After finishing all person's eye candidate data sets, we have a very good training set and boost the accuracy of linear SVM.

3.1.2 Selection Of Learning Kernels For SVM

After finishing the above step, we get a training set which has 558 positive images and 560 negative images. In order to obtain the best accuracy, we find the best parameters for the SVM. In Figure 9, we list three different SVM kernels with various parameter settings and each SVM was tested on 1757 eye candidate images obtained from a new person.

Figure 9: Experiment results using 3 kernels with different parameters

Kernel Type	Deg	Sigma σ	# Support Vectors	Accuracy
Linear			376	0.914058
Polynomial	2		334	0.912351
Polynom.	3		358	0.936255
Polynomial	4		336	0.895845
Gaussian		1	1087	0.500285
Gaussian		2	712	0.936255
Gaussian		3	511	0.955037
Gaussian		4	432	0.9465
Gaussian		5	403	0.941377

From the above figure, we can see that the best accuracy we can get is 95.5037%, so we will use the Gaussian Kernel whose sigma term is 3 as the kernel of SVM.

Pupils verification with SVM works reasonably well and can generalize for people of the same race. However, for people from a race that is significantly different from those in training images, the SVM may fail and need to be retrained. SVM can work under different illumination conditions due to the intensity normalization for the training images via histogram equalization.

4 Eye Tracking Algorithm

Given the detected pupils from the initial frames, pupils can subsequently be tracked from frame to frame. The previous method [Ji and Yang 2001] for eye tracking is based on tracking the bright pupils using Kalman filtering. This technique, however, will fail if the pupils are not as bright due to either face orientation or external illumination interferences as shown in figure 10, where pupils have been removed from the thresholded difference images due to their weak intensity.

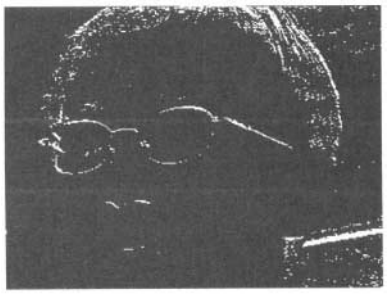


Figure 10: The thresholded difference image without pupils.

However, due to the strong illumination, the eye region in the dark pupil image exhibits strong and unique visual pattern such as the dark iris in the white part. This unique pattern due to the use of external ring of IR LEDS should be utilized to track pupils in case the bright pupils fail to appear on the difference images.

Based on this understanding, we propose a multi-stage eye tracking method based on combining the bright-pupil method with the method based on pattern of intensity distribution. Figure 11 summarizes our tracking method. Our eye tracking method consists of two stages. For the first stage, we perform eye tracking using Kalman filtering based on the bright pupil effect [Ji and Yang 2001]. If this technique fails, we activate the second stage, where we try to search the eye in the dark pupil image around the eye center

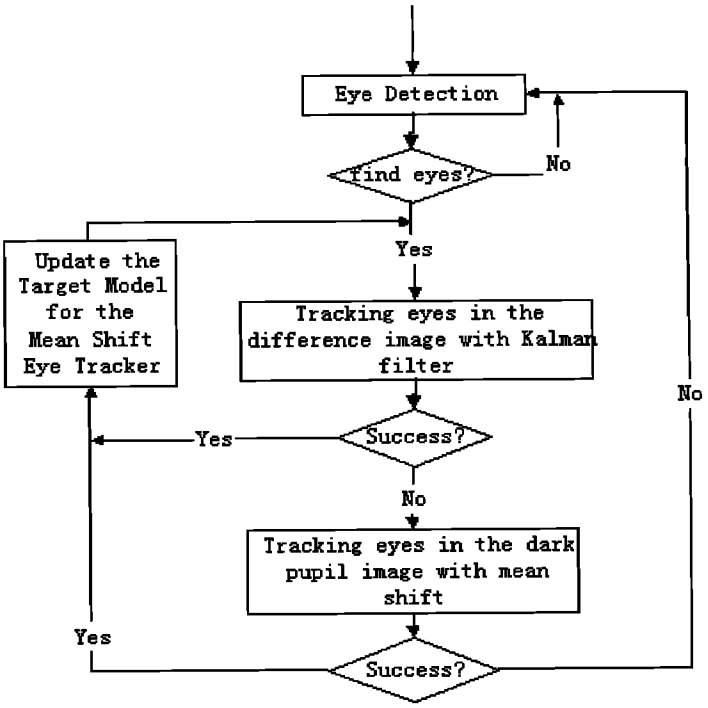


Figure 11: The Eye tracking algorithm flowchart

of the previous frame based on the intensity distribution using the mean-shift algorithm. The mean shift tracking algorithm [Comaniciu et al. 2000] is briefly summarized below.

4.1 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.

4.1.1 Tracking Model

We use a 1D histogram which is derived in the grey level dark pupil image with 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 eye target. Let us denote by $\{x_i^*\}_{i=1...n_h}$ the pixel locations of the eye target, centered at y in the previous frame. The probability of the intensity u in the eye target is given by

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

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 candidate distribution p can be built in a similar fashion.

4.1.2 Algorithm

After locating the eyes in the previous frame, we construct a target eye model q based on the detected eyes in the previous frame.

We then predict the locations $\hat{y}_0$ of eyes at current frame using the Kalman filter. Then we treat $\hat{y}_0$ as the initial position and use the mean shift iterations to find the most similar eye candidate with the eye target model in the current frame using the following algorithm.

1. Initialize the location of the target in the current frame with $\hat{y}_0$, then compute the distribution $\{\hat{p}_u(\hat{y}_0)\}_{u=1\ldots m}$ using equation 2 and evaluate similarity measure (Bhattacharyya coefficient) between the model density (q) and target candidate density (p)

$$\rho[\hat{p}(\hat{y}_0), \hat{q}] = \sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{y}_0)\hat{q}_u} \tag{3}$$

2. Derive the weights $\{w_i\}_{i=1\ldots n_h}$ according to

$$w_i = \sum_{u=1}^{m} \delta[b(x_i) - u] \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{y}_0)}}, \tag{4}$$

where δ is the Kronecker delta function.

3. Based on the mean shift vector, derive the new location of the eye target [Comaniciu et al. 2000]

$$\hat{y}_1 = \frac{\sum_{i=1}^{n_h} X_i w_i g(\| \frac{\hat{y}_0 - x_i}{h} \|^2)}{\sum_{i=1}^{n_h} w_i g(\| \frac{\hat{y}_0 - x_i}{h} \|^2)} \tag{5}$$

and then update $\{\hat{p}_u(\hat{y}_1)\}_{u=1\ldots m}$, and evaluate

$$\rho[\hat{p}(\hat{y}_1, \hat{q})] = \sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{y}_1)\hat{q}_u} \tag{6}$$

4. While $\rho[\hat{p}(\hat{y}_1), \hat{q}] < \rho[\hat{p}(\hat{y}_0), \hat{q}]$
Do $\hat{y}_1 \leftarrow 0.5(\hat{y}_0 + \hat{y}_1)$

5. If $\|\hat{y}_1 - \hat{y}_0\| < \varepsilon$ Stop
Otherwise, set $\hat{y}_0 \leftarrow \hat{y}_1$ and go to step 1.

The new eye locations in the current frame can be achieved in few iterations compared to the correlation based approaches which must perform an exhaustive search around previous eye locations. Due to the simplicity of the calculations, it's much faster than correlation.

5 Experiment Results

The proposed method has been experimentally evaluated under different illumination conditions, camera parameters, face orientations, and subjects. It performs reasonably well, significantly better than the one based on only bright-pupil effect. Figure 12 gives some of the tracking results. Additional Video demos are available at *http://www.cs.unr.edu/~zhu_z/Demo/demo.html*.

6 Conclusions and Future Work

In this paper, we propose a new method to improve eye tracking under various illumination conditions, face orientations, and different camera parameters. Based on combining the appearance based method with the bright pupil method that uses active IR illumination, the proposed method improves tracking robustness and accuracy under various lighting conditions. The system can work very well under strong non-infrared lighting and ordinary ambient infrared lighting. Future work will involve further improving the method so that it can work under strong Sun light as well as systematic characterization of the performance our technique against the existing systems.

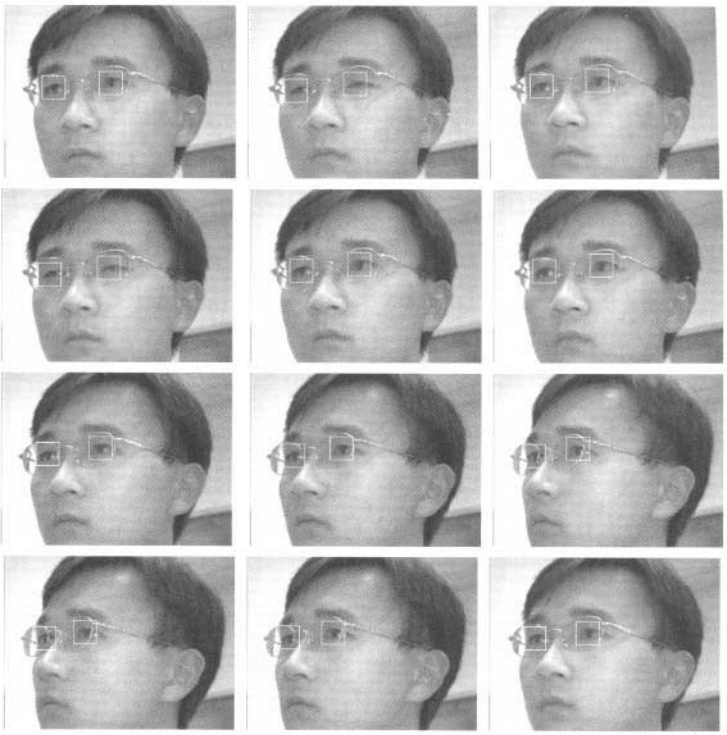


Figure 12: Detected eyes marked with white rectangles under various face orientations in a sequence of consecutive frames

Acknowledgments

This work is partially supported by Honda R&D Americas Inc., Mountain View, CA and by a grant from US Air Force office of Scientific Research.

References

BALUJA, S., AND POMERLEAU, D. 1994. Non-intrusive gaze tracking using artificial neural networks. Technical Report CMU-CS-94-102, Carnegie Mellon University.

COMANICIU, D., RAMESH, V., AND MEER, P. 2000. Real-time tracking of non-rigid objects using mean shift. In *IEEE Conf. on Comp. Vis. and Pat. Rec.*

CORTES, C., AND VAPNIK, V. 1995. Support-vector networks. *Machine Learning 20*, 273–297.

EBISAWA, Y., AND SATOH, S. 1993. 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*, 1268–1269.

FITZGIBBON, A. W., AND FISHER, R. 1995. A buyers guide to conic fitting. In *Proc.5 th British Machine Vision Conference*, 513–522.

HARO, A., FLICKNER, M., AND ESSA, I. 2000. Detecting and tracking eyes by using their physiological properties, dynamics, and appearance. In *Proceedings IEEE CVPR 2000*.

HUANG, J., II, D., SHAO, X., AND WECHSLER, H. 1998. Pose discrimination and eye detection using support vector machines (svms). In *Proceeding of NATO-ASI on Face Recognition: From Theory to Applications*, 528–536.

JI, Q., AND YANG, X. 2001. Real time visual cues extraction for monitoring driver vigilance. In *Proc. of International Workshop on Computer Vision Systems*.

MORIMOTO, C., AND FLICKNER, M. 2000. Real-time multiple face detection using active illumination. In *Proc. of the 4th IEEE International Conference on Automatic Face and Gesture Recognition 2000*.

MORIMOTO, C., KOONS, D., AMIR, A., AND FLICKNER, M. 1998. Pupil detection and tracking using multiple light sources. Technical Report RJ-10117, IBM Almaden Research Center.

OLIVER, N., PENTLAND, A., AND BERARD, F. 1997. Lips and face real time tracker. In *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition*, 123–129.

SMITH, P., SHAH, M., AND LOBO, N. D. V. 2000. Monitoring head/eye motion for driver alertness with one camera. In *Proceedings of the 2000 International Conference on Pattern Recognition*, Session P4.3A.

TURK, M., AND PENTLAND, A. 1991. Eigenfaces for recognition. *Journal of Cognitive Neuroscience 3*, 1, 71–86.

VAPNIK, V. 1995. *The nature of statistical learning theory*. Springer-Verlag, New York.