A neural system for eye detection in a driver vigilance application

T. D'Orazio, M. Leo, P. Spagnolo, C. Guaragnella*,
CNR- ISSIA Via Amendola 122/D-I 70126 BARI ITALY
[dorazio,leo,spagnolo]@ba.issia.cnr.it
*DEE - Politecnico di Bari Via E. Orabona 4 70125 BARI ITALY
guaragnella@poliba.it

Abstract

The problem of eye detection for a driver vigilance system is very important in order to monitor driver fatigue, inattention, and lack of sleep. In this paper a neural classifier has been applied to recognize the eyes in the image, selecting the couple of regions candidate to contain the eyes by using iris geometrical information and symmetry. The novelty of this work is that the algorithm works on complex images without constraints on the background, skin color segmentation and so on. Different experiments have been carried out on images of subjects with different eyes colors, some of them wearing glasses. Tests showed robustness with respect to situations such as eyes partially occluded. In particular when applied to images where people have the eyes closed the proposed algorithm correctly reveals the absence of eyes. Eyes tracking in an image sequence is applied to detect eye closure that can be dangerous if persists for a long period.

1. Introduction

Eye detection is a crucial aspect in many useful applications ranging from face recognition/detection to human computer interface, driver behavior analysis, or compression techniques like MPEG4. By locating the position of the eyes, the gaze can be determined. In this way it is possible to know where people are looking at and understand the behaviors in order to evaluate the interests (for interface purposes) and the attention levels (for safety controls).
A large number of works have been published in the last decade on this subject. Generally the detection of eyes consists of two steps: locating face to extract eye regions and then eye detection from eye window.
The face detection problem has been faced up with different approaches: neural network, principal components, independent components, skin color based methods [2,13]. Each of them imposes some constraints: frontal view, expressionless images, limited variations of light conditions, hairstyle dependence, uniform background, and so on. A very exhaustive review has been presented in [12].
On the other side many works for eye or iris detection assume either that eye windows have been extracted or rough face regions have been already located [1,3,5,6,7,8,9].
In [3], the eye detection method is performed within the possible eye region of the candidate face field, then it can be applied only after that a face detection system has extracted small candidate eye regions. Left and right eye templates have been used to detect eyes with a method that is invariant to slight rotation, scaling, and translation (up to 10%). The algorithm proposed in [1] requires the detection of face regions in order to extract intensity valleys. Only at this point the authors apply a template matching technique to extract iris candidates. The authors afford the difficult problem of face region extraction both in intensity images and in color images. Skin color models are strictly related to the considered images and cannot be so general to be applied in every light condition and with different colors of the skin. Region-growing methods or head contours methods on intensity images require strong constraints such as plain background. Also in [5] the first step is a face detection algorithm that is based on skin color segmentation on the input image with the constraints of having only one face and a simple background; then the facial feature segmentation is based on gray value reliefs. In [6] linear and non linear filters have been used for eye detection: oriented Gabor wavelets form an approximation of the eye in

gray level images; non linear filters are applied to color images to determine the color distribution of the sclera region. In both cases a face detection step is applied assuming that the face is the most prominent flash tone region in the image. The same algorithm has been used in [7] for tracking iris and eyelids in video. In [9] lip and skin color predicates are used as first step to segment lip region and skin regions in the image: the two holes above the lip region that satisfy some fixed size criteria are selected as the candidate eye regions. A hierarchical strategy is applied to track the eyes in a video sequence and evaluate the driver visual attention by finite state automata. In [4] the authors make use of multi cues extracted from a gray level image to detect the eye windows within an a priori detected face region. The precise iris and eye corner locations are then detected by variance projection function and eye variance filter.

A careful analysis of the previous works reveals some common constraints: first of all, the problem of face segmentation, distinguishing faces from a cluttered background, is usually avoided by imaging faces against a uniform background. Second, the common use of skin color information to segment the face region requires an initialization phase. Finally, the more precise is the location of the eye regions, the more reliable are the results of the eye detection algorithms. By our knowledge, no much works have been presented in literature that search directly eyes in complex images. Some works for biometrics applications detect the precise location of the iris using geometrical information [14]: however, these algorithm have been applied on eye images and it not clear if they could be applied directly on complex images.

Active techniques exploit the spectral properties of pupil under near IR illumination. In [10] two near infrared multiplexed light sources synchronized with the camera frame rate have been used to generate bright and dark pupil images. Pupils can be detected from simple thresholding of the difference of the dark from the bright pupil images.

The main objectives of our work is to propose an eyes detection algorithm that is applicable in real time with standard cameras, in a real context such as people driving a car (then with a complex background), and skipping the first segmentation step to extract the face region as commonly done in literature.

Our eye detection algorithm works on the whole image, looking for regions that have the edges with a geometrical configuration like the expected one of the iris. Different iris radius are allowed in order to face with people having different eyes dimensions and also light variations in the distance between the camera and the person. A constraint based on the distance, orientation, and symmetry is used to select the couple of regions candidate to contain the eyes. Experimental results demonstrate that when the eyes are open, they are correctly detected; on the contrary when the eyes are closed or occluded, the algorithm provides false positives. For this reason we have added a further step for the validation of the results. A neural classifier has been trained to recognize the two classes eyes or not-eyes using a large number of examples taken from images of different people.

Tests carried out on different persons, with different eyes colors and dimensions, some of them wearing glasses, demonstrate the effectiveness of the proposed algorithm.

The rest of the paper is organized as follows: Section 2 describes the eyes detection algorithm in details. Section 3 describes the validation of the sought regions with a neural classifier. The results on different real images are reported in Section 4. Finally, in Section 5 conclusions and future works are presented.

2. Eye detection

The idea behind our work is quite simple: the eyes can be easily located in the image since the iris is always darker than the sclera no matter what color it is. In this way the edge of the iris is relatively easy to detect as the set of points that are disposed on a circle. It is not possible to know the exact diameter of the iris since people can have different iris dimensions and also the system has to manage variable distances between people and the camera. For this reason a range $[R_{min}, R_{max}]$ is set to tackle different iris radius.

In this work we have used a new circle detection operator that is based on the directional Circle Hough Transform. The gradient image is convolved with a template containing in every point of a ring with radius $[R_{min}, R_{max}]$ the directions of the radial vector. In this way the algorithm evaluates how many points in the image have the gradient direction concordant with the gradient direction of a range of circles. Then the peak in the resulting image gives the candidate center of the circle in the image.

The circle detection operator is applied on the whole image without any constraint on plain background or limitations on eye regions. The result of this step is a maximum that represents the region of the whole image that is best candidate to contain one eye. Starting from this area the search of the second eye is applied only in the two opposite regions whose distances and orientations are compatible with the range of possible eyes positions.

In this way false positives that occur in the hair regions or on other face parts (such as nose, mouth, and so on) are easily discarded out. Different similarity measures have been explored. The results obtained are quite similar; for this reason, since our main goal is to have frame rate performances, we have decided to choose the one that requires the less computational load. The similarity of the two regions has been evaluated using the Mean Absolute Error (MAE) applied on specular domains:

$$MAE_s = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}\left|a_{ij} - b_{i(M-j+1)}\right|}{N \cdot M}$$

where a and b represent the points in the two regions whose dimensions are NxM. The mirroring of the second region is necessary to evaluate the symmetry of the couple of candidate eyes. If this measure of similarity is below a fixed threshold value the two regions are considered the best match for eye candidate, otherwise a further search is activated. The whole procedure used for searching the eyes in the image is described in figure 1.

1. Search the maximum value M1 of the output convolution in the whole image
2. Search the second maximum M2 in the regions that are candidate to contain the second eye
3. Compare the two region M1 and M2 using the measure MAE_S
4. IF (MAE_S(M1,M2) < Threshold) THEN Output= M1,M2 ELSE M1=0 GO TO step 1

Figure 1. The procedure for searching the second eye.

The results of the similarity measure are of course strictly dependent on the region dimension: smaller are the regions, higher are the false positives found. Intuitively, two similar circular regions can be found in many areas of face images: among curly hair, at the sides of the mouth, on the forehead, and so on. However if the couples of regions considered for the similarity measure are larger than the circular shape searched in the first step, the probability to found false matches decreases. If misdetections happen, they have to be found in symmetric parts of the face. In this work we have used regions with dimensions of 62x62 that is quite the double of the iris dimension.

The experiments show that the similarity constraint is very selective: a large number of wrong candidates are discarded since their similarities overcome the fixed threshold. The threshold has been selected experimentally by evaluating the similarity of a priori known eye regions.

3. Eyes Validation by using a neural network

The couple of regions candidate to contain the eyes have to be validated to detect if the eyes are closed and false detection have been found.

The subimage containing the result of the detection process is given to a neural network trained to classify images as eye instances or no-eye ones. One of main requirements of our application is the possibility to work in real time. This requirement can be satisfied if the algorithm of features extraction is able to store all the necessary information in a small set of coefficients and the algorithm for classification is skillful to take advantage of that in order to supply right answers in reduced times. In this work we have used a Discrete Wavelet Transform that supplies a hierarchical representation of the image implemented with the iterative application of two filters: a low pass filter (approximation) filter and its complementary in frequency (detail filter). At each step the Wavelet Transform breaks the image into four subsampled or decimated images.

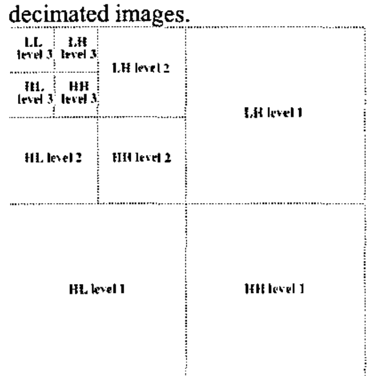

Figure 2 The decomposition of the image with a three levels wavelet transform

In figure 2 the final result of a 3-level Wavelet transform is shown. The capital letters in each subimage represent the kind of filters that have been applied on the image of the previous level; H stands for an High pass filter, L stands for a Low pass filter. The first letter is the filter that has been applied in the horizontal direction, while the second letter is the filter that has been applied in the vertical direction.

The band LL is a coarser approximation to the original image. The band LH and HL record the changes of the image along horizontal and vertical directions. The band HH shows the high frequency components of the image. Decompositions can be iterated on the LL subbands. After applying a 3-level Wavelet Transform, an image is decomposed into subbands of different frequency components. For their simplicity the Haar filters have been used for the implementation of the Wavelet transform. The coefficients of the wavelet transform enclose information about the texture and the shape of the object in the image. In this way it is possible to distinguish the eyes from other elements that could have in common one of the two aspects. We have selected for the recognition process the coefficients of the third level of decomposition of the wavelet transform because they mainly contain the information of shape and texture. The total number of features extracted is equal approximately to 1/16 of pixels of initial image.

The classifier consists in a neural network, simple to use and widely employed in pattern recognition problems, trained with a back propagation algorithm. The input layer has the number of nodes equal to the number of coefficients of the WT of the third level, the hidden layer has 80 nodes and the output layer has only one node. The neural network parameters have been found experimentally, choosing the combination that produces the best results in terms of eye detection rate. A large number of examples of eyes and no-eyes have been selected from images of different people and have been provided to the neural network to train the classifier.

3. Experimental Results

In this section the experimental results obtained on a large number of images are shown. Six different persons have been used to take images with different eyes colors, different dimensions and different shapes. In this work we assumed that the distance between the camera and the person cannot greatly change (people remain sit down on a chair in front of a camera), even if we have considered that people are allowed to slightly move on the chair as it is normally the case. The images taken in these conditions show that the iris dimensions can vary in a range of [20,30] pixels. No constraints have been imposed on the background.

Two sets of experiment have been carried out. In order to validate the ability of the first step of eye detections some tests have been done on images acquired in our laboratory selecting only the images with eyes open. Some of these images are shown in figure 3

Figure 3. The images of the six persons used in the experiments.

The results are shown in table 1. In particular in the first column of the table the results obtained on the original images are shown. In the other three columns the results obtained on the images, whose light condition have been artificially modified, are reported. The values of α represent the multiplicative constant used to modify the original images. When the value of α is less than 1 the images are uniformly darkened, while the value of α greater than 1 means that images are lightened as shown in figure 4.

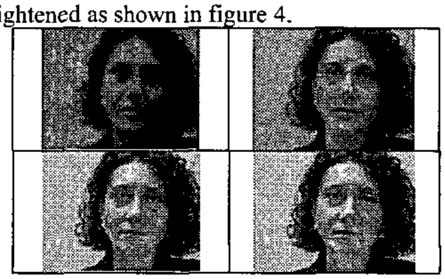

Figure 4. Four images of person 1 obtained varying the light conditions with α=0.6, 0.8, 1, 1.2

Table 1
Results obtained on the images acquired in our laboratory

	Original Images	α=0.6	A=0.8	α=1.2
Person 1 (35)	100%	100%	100%	100%
Person 2 (39)	100%	100%	100%	100%
Person 3 (39)	100%	0%	100%	100%
Person 4 (50)	100%	98%	100%	100%
Person 5 (63)	100%	95%	95%	84%
Person 6 (16)	100%	68%	100%	100%
Det. Rate	**100%**	**80%**	**98%**	**95%**

The detection rate are very high except for the case of α equals to 0.6 that correspond to very dark images. Some images with eyes correctly detected in the original images are shown in figure 5.

Figure 5. Some images with eyes detected correctly

In the second set of experiments a total of 1474 images has been considered. The images have been divided in three sets: eyes completely open, eyes closed, eyes partially visible (in the last set we have considered all the images where people are locking off the center, or the eyelids cover part of the iris). In table 2 the results obtained after the first step of our algorithm are shown. In each row of the table, the detection rates for each set of images are reported; the number of images used for each person is specified in the first column. In the last row of the tables the total detection rates have been evaluated.

Table 2

Results obtained after the first step.

	Eyes Open	Eyes Closed	Eyes partially open
Person 1 (79)	73/73	0/4	½
Person 2 (220)	212/212	0/4	¾
Person 3 (439)	390/414	0/15	6/10
Person 4 (353)	286/302	0/12	17/39
Person 5 (251)	235/235	0/2	14/14
Person 6 (132)	93/96	0/14	0/22
Det. Rate	**96%**	**0%**	**45%**

On the set of images with eyes open the proposed algorithm reaches high detection rate (96%). The results on the set of images with eyes closed are obviously all false positive (the first step of our algorithm selects in any case a couple of regions candidate to be eyes). When the eyes are partially open the detection rate is about the 45%. At this point it is very interesting to observe the results of the validation step.

The images obtained after the first step have been provided to the neural classifier that validates the obtained results. The training set contains 330 negative examples (non- eyes images) and 281 positive examples (eyes images) taken from the first three persons. It should be noted that the number of images containing eyes is the sum of the numbers of correct detection of eyes open images and eyes partially open ones, while the number of images of non eyes is the sum of wrong detection of eyes and the number of images non containing eyes.

Table 3

Results obtained after the validation with the neural classifier.on three persons of the training set.

	True Positive	False Positive	True Negative	False Negative
Person 1 (79)	74/74	0/5	5/5	0/74
Person 2 (220)	215/215	0/5	5/5	0/215
Person 3 (439)	396/396	0/43	43/43	0/396
Det. Rate	**100%**	**0%**	**100%**	**0%**

Table 4

Results obtained after the validation with the neural classifier on other three different persons.

	True Positive	False Positive	True Negative	False Negative
Person 4 (353)	277/303	0/50	50/50	26/303
Person 5 (251)	221/249	0/2	2/2	28/249
Person 6 (132)	79/93	10/39	29/39	14/93
Det. Rate	**89%**	**10%**	**89%**	**10%**

Tables 3 and 4 report the results produced by the neural classifier when applied on the candidate regions provided by the first step of our algorithm. The percentage of True Positive (eyes correctly recognized) and True Negative (non- eyes correctly recognized) are very high. False positives and False negatives have been found on images of the three persons that actually have not been considered in the training set (see table 4) These results are then very significant because demonstrate that the algorithm works very well when some images of each person are included in the training set (see table 3), but also that satisfying performances can be obtained when images of different people are used in the test phase (see table 4). In figure 6 some images of correct detection are shown.

The knowledge of the eyes position in one frame has been used in the successive frame to limit the search of the eyes and speed up the whole process. Besides counting the number of frames in which the eyes have not been found it is possible to give an alarm signal. Indeed considering the frame rate of the camera, it is possible to evaluate how many frames takes the normal behavior of eyes blinking. When the eyes are closed (or are not visible) for a number of frames greater than the estimated number of normal eye blinking, the behavior of the driver can be considered not safe and an alarm signal can be provided.

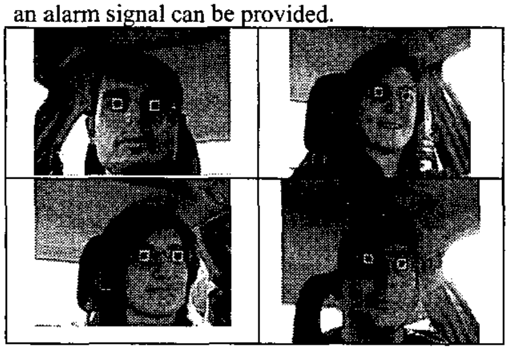

Figure 6. Some images with eyes detected correctly

5. Conclusion and Future Work

In this paper we present an effective algorithm for eyes detection in face images. It consists in two steps: first the couple of regions candidate to contain the eyes is detected in the whole image using edge direction and symmetry information. Then, a neural network is used to validate the search of eyes. Our system does not impose any constraint on the background and does not require any preprocessing step for face segmentation. Certainly, since our system searchs the eyes in the images it can works only in normal light conditions (not during nighttime driving unless a kind of artificial lightening is contemplated).

In future work an adequate attention will be given to the evaluation of anomalous behaviors such as people that are not looking straight ahead, are looking for long period to the rear-view mirror, or people that are frequently closing their eyes and so on.

6. References

[1] T. Kawaguchi, M. Rizon *Iris detection using intensity and edge information,* Pattern Recognition 36 (2003) 549-562

[2] H. A. Rowley, S. Baluja, T. Kanade *Neural Network-Based Face Detection* IEEE Transaction on Pattern Analysis and Machine Intelligence Vol. 20,No. 1, January 1998, pp 23-38

[3] Y. Li, Xiang-lin Qi, Yun-jiu Wang, *Eye detection by using fuzzy template matching and feature-parameter based judgement* Pattern Recognition Letters 22 (2001) 1111-1124

[4] G. C. Feng, P.C. Yuen *Multi cues eye detection on gray intesity image* Pattern Recognition 34 (2001) 1033-1046

[5] S. Baskan, M. Bulut, V. Atalay *Projection based method for segmentation of human face and its evaluation* Pattern Recognition Letters 23 (2002) 1623-1629

[6] S. Sirohey, A. Rosenfiled *Eye detection in a face image using linear and nonlinear filters* Pattern Recognition 34 (2001) 1367-1391

[7] S. Sirohey, A. Rosenfiled, Z. Duric *A method of detection and tracking iris and eyelids in video* Pattern Recognition 35 (2002) 1389-1401

[8] M. Rizon, T. Kawaguchi *Automatic extraction of iris with reflection using intensity information* Proceeding of the 3th Iasted Conference on Visualization Imaging and image processing September 8-10, 2003 Benalmaden Spain

[9] P. Smith, M Shah, N. da Vitoria Lobo, *Determining Driver Visual Attention with One Camera,* Accepted for IEEE Transactions on Intelligent Transportation Systems.

[10] C.H. Morimoto, D. Koons, A. Amir, M. Flickner *Pupil detection and tracking using multiple light sources* Image and Vision Computing 18, (2000) 331-335

[11] H. Kashima, H. Hongo, K. Kato, K. Yamamoto *A Robust Iris Detection Method of Facial and Eye Movement* Proceeding of VI 2001 Vision Interface Annual Conference Ottawa, Canada

[12] M. H. Yang, D. Kriegman, N. Ahuja *Detecting Faces in Images: A Survey* IEEE Transaction on Pattern Analysis and Machine Intelligence Vol. 24, No. 1 January 2002

[13] R. Hsu, M. Mottleb, A. K. Jain *Face Detection in Color Images* IEEE Transaction on Pattern Anlysis and Machine Intelligence Vol24, No. 5, May 2002

[14] J.G. Daugman *High Confidence Visual Recognition of Person by a Test of Statistical Independence* IEEE Transactions on Pattern Analysis and Machine Intelligence 15 (1993) 1148-1161