Far East Journal of Electronics and Communications
Volume 7, Number 2, 2011, Pages 91-114; Issue Date: April 2012
Published Online: April 2012
Available online at http://pphmj.com/journals/fjec.htm
Published by Pushpa Publishing House, Allahabad, INDIA

FAST AND EFFECTIVE SYSTEM TO LOCATE EYE AND RECOGNIZE OPEN/CLOSED STATE

Lubing Zhou and Han Wang

School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore
e-mail: zhou0145@e.ntu.edu.sg

Abstract

Automatic eye localization (EL) and eye state (open or closed) recognition (ESR) are still open challenges due to various difficulties, such as low resolution, lighting interference and changing appearance of eye. This paper proposes an automatic EL and ESR system as a solution. Given a frontal face, the face-region and two rough eye windows are successively detected using Viola Jones method. Within rough eye window, accurate position of eye center is found by a new projection function, Selective Projection Function, which selectively picks low-intensity pixels from nearby rows or columns to produce the response. To decide open/closed state, a novel descriptor is first designed based on grayscale correlogram, which expresses spatial correlations of grayscale pairs. Then pattern classification techniques, including subspace-based methods (PCA and LDA), SVM and AdaBoost algorithm are examined to train the binary eye state classifier using the descriptor. Experimental results show the proposed EL and ESR algorithms demonstrate encouraging performance on static images with large variations. Furthermore, the automatic system also displays high speed and accuracy in real time video.

Keywords and phrases: eye detection, eye state recognition, color correlogram.
Received February 24, 2012

1. Introduction

Eye is a salient feature in human face. Eye centers are often used to conduct face alignment before face recognition (FR) and facial expression analysis (FEA). Eye open/closed state is another important aspect. For one thing, it can guide real-time FR or FEA to avoid errors due to eye closure. Moreover, ESR has shown great practical values in driver fatigue monitoring system [1], human computer interface [2] and photography technology [3] in past decade. However, EL and ESR are still open topics due to eye's special characteristics. First, eye has low resolution, but it is highly deformable and has changing appearances. Another difficulty is the severe lighting interferences, such as shadows and reflective highlights, which greatly increase the within-class variations. Hence, automatic EL and ESR are attracting more and more research interests. The following two paragraphs present fast reviews on EL and ESR, respectively.

EL is often used as a preceding step before FR or FEA. According to different purposes, EL can be divided into rough eye window detection and precise localization of eye features such as eye center, eyelids and eye corners. Naturally, local eye features can be located more easily if the rough eye window is detected first. Existing EL methods can also be classified into active infrared-based and passive image-based methods. Active type is simple and efficient, but requirement of extra IR illuminators and poor performance in outdoor environment limit its popularity. On the other hand, image-based methods are conducted on visible image, and can be further divided into template-based [4, 5], feature based [6, 7, 8] and appearance-based sub-categories [9, 10]. Template-based approaches first derive a generic eye model, and then search whole face image by template matching. In [4], the iris is modeled as an ellipse, and deformable templates are introduced to track eyes in [5]. Generally, the strategy is very fast, but it is not realistic to favorably find the eye model due to high diversity of eye appearances. In feature-based methods, special characteristics of eye features like eye center, eyelids and eye corners are explored to infer eye location. Usually, methods in this type are also very efficient and able to get precise

eye position. For example, Integral Projection Function (IPF) and Variance Projection Function (VPF) [6, 7] are commonly used to get eye center. The algorithm proposed in [8] uses Circular Hough Transforms to detect iris circle. Nevertheless, feature-based methods are not stable enough as false positives (nose, eyebrow, etc.) tend to increase under large variations. Appearance-based methods attract much attention with the advances of machine learning techniques. Eye detection is transformed into an eye/non-eye classification problem. For example, a neural network classifier is trained to detect eye in [9]. Appearance-based methods can handle large variation problem as long as training set includes such conditions. However, being block-based methods, they can only detect rough eye window.

ESR does not attract so much research attention as EL. In literature, there exist many image-based methods. Likewise, they can be classified into template based [1, 2], feature-based [11, 12, 13] and appearance-based classes [3, 14]. In template-based method, first open/closed eye templates are designed, then state of a new eye is decided by similarity scores to the templates. Feature-based ESR explores the differences of open and closed eyes in some special characteristics. Such characteristics include the dark pupil, white sclera, circular iris, upper eyelid bending direction and so on. For example, contour information is used to detect whether eyes are open or closed in [11]. In [13], the grayscale characteristics, upper eyelid bending direction are employed to decide eye state. Similar to EL, appearance-based ESR methods rely on machine learning techniques. In [14], SVM is employed to verify eye state using Local Binary Patterns histograms (LBPH). Generally, appearance-based methods need a large set of training samples, but they have better performance and higher consistency to complex conditions than the other two types.

Original motivation of this paper is to provide an alternative solution for automatic ESR system. In general, such a system successively constitutes face detection (FD), EL and ESR. According to current research progress, our work emphasizes on EL and ESR, particularly on ESR. The contribution of this paper mainly consists of two parts. In EL part, a coarse-to-fine scheme

is utilized to precisely locate eye center. First, rough eye window is detected by Viola Jones method [15], which belongs to appearance-based EL method. Then precise eye center is located using proposed Selective Projection Function (SPF), which is a feature-based EL method. Our EL solution combines appearance-based and feature-based methods, which are relatively complemented to each other. Hence, eye center can be obtained more precisely and robustly. More important contribution of the paper comes from ESR part. ESR is transformed to a binary classifier problem. First, a simple and low-dimensional eye descriptor is originally introduced to represent eyes based on grayscale correlogram, which expresses spatial correlations of grayscale pairs. Then various learning algorithms, including PCA, LDA, SVM and Discrete AdaBoost (DAB) algorithms are respectively trained and compared to recognize eye state. Our ESR method displays high accuracy using only hundreds of features, and runs at a favorable speed.

The remainder of the paper is organized as follows: Section 2 introduces our EL method to accurately locate eye centers from a frontal-view face image. Next in Section 3, several ESR methods are proposed by combining the correlogram-based eye descriptor and various learning methods. Then thorough performance evaluations of the proposed EL, ESR methods, and the combined system are presented in Section 4. Finally, the conclusion is given in Section 5.

2. Eye Localization

2.1. Rough eye window detection

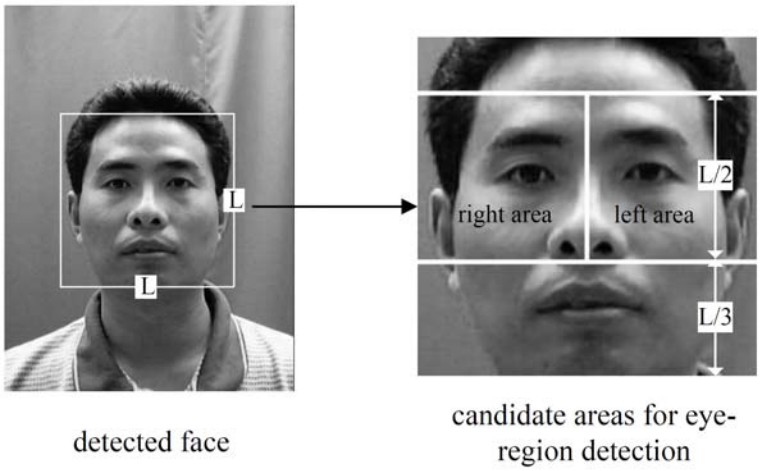
Given an image, this paper adopts well-known Viola Jones method [15] to detect the frontal face. It first trains many classifiers by using AdaBoost algorithm to select most discriminative Haar-like features, then applies a cascaded structure to discard most non-face regions quickly while maintaining almost all face regions in early stages. Thus, it is also called *boosted cascade detector*. Afterwards, the detected face is partitioned into left and right half faces, whose upper-middle parts are considered as candidate areas for eyes as shown in Figure 1(a). The detected face size is

$L \times L$. Similarly, two boosted cascade detector are then learned to detect eye-regions, which include both eye window part and eyebrow part. There are two important points during eye-region training. First one is that left eye-region and right eye-region detectors should be separately trained because of the diversity of two eyes. Another more important point is to include eyebrow in the eye-region as shown in the left of Figure 1(b). d_e is the inter-ocular distance. For one thing, eyebrows are more stable than deformable eyes, and Haar-like features have been reported to work better on rigid objects. For another thing, both open and closed eye-regions can be well detected as eyebrows are used for co-occurrence verification. Jesorsky el al. [16] have shown that eye width is approximately half of inter-ocular distance, and eye height is often adopted as half eye width. That is, true eye window should be the area of $0.5d_e \times 0.25d_e$ centered at precise eye center. In this paper, rough eye window is considered as the $0.6d_e \times 0.3d_e$ area expanding from estimated center by the eye-region detector as shown in the right of Figure 1(b). Choice of $0.6d_e \times 0.3d_e$ has taken two aspects into consideration: (1) the area must guarantee that eye center is inside, and (2) smaller area is better to reduce interference (eyebrow) during future eye center localization when condition (1) is satisfied.

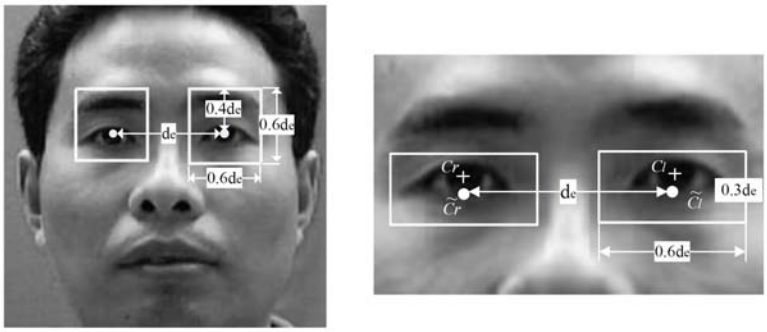
2.2. Eye center localization

Precision of EL directly affects accuracy of subsequent ESR. Given rough eye window, eye center can be precisely located more easily. Feature-based EL methods are more qualified for such purpose. Projection functions (PFs) [6, 7] can be used to locate eye center. Among them, Variance Projection Function (VPF) [7] and Integral Projection Function (IPF) [6] are widely used two. Briefly, VPF uses row (column) intensity variance to locate eye center, while IPF uses row (column) mean to find eye center. Both VPF and IPF are fast and simple to implement. Alternatively, this paper proposes the Selective Projection Function (SPF) to precisely locate eye center. To describe SPF, two observations should be mentioned. First, pixels around eye center normally have lower intensity than the rest in the rough eye window. Second, number of such lower-intensity pixels around eye centers is more

than other locations within rough eye window. Note the two observations are not just restatement to each other. First observation focuses on “lower intensity”, while the other emphasizes on “more number”.



(a) Candidate areas for rough eye window detection



(b) Left - cropping area of training samples for rough eye window detector; right - used area for eye center localization, d_e is inter-ocular distance, $\tilde{C}_r, \tilde{C}_l$ are estimated eye centers, and C_r, C_l are ground truth

Figure 1. Areas for rough eye window detection and eye center localization.

Based on the two observations, the procedure of SPF is shown in Figure 2. Let I be a $w \times h$ rough eye window, and $I(x, y)$ is intensity of pixel (x, y). I_1 is the resulting image by histogram equalization from I for contrast stretch. Next only a small portion of pixels with lowest intensity are left and the rest are masked. This paper adopts a threshold to filter I_1 as expressed as

equation (1), where $mean(I_1)$ is mean intensity of I_1 and C is optimized as 0.25 which is shown in Section 4.1,

$$I_2(x,\ y) = \begin{cases} I_1(x,\ y), & \text{if } I_1(x,\ y) < C * mean(I_1); \\ 255, & \text{otherwise.} \end{cases} \quad (1)$$

The goal is to ensure that eye center will have more unmasked pixels to produce the response in image I_2. As shown in I_2 of Figure 2, the white area represents masked pixels $(I_2(x,\ y) = 255)$, and the rest remain original intensity $(I_2(x,\ y) = I_1(x,\ y))$. Meanwhile, within unmasked pixels in I_2, it is observed that pixels around eye center commonly have lower intensity than other places. However, more unmasked pixels around eye center cannot guarantee their sum (response) is still larger, as their intensities are lower. Hence, the inverse pattern of I_2 is obtained by: $I_3(x,\ y) = 255 - I_2(x,\ y)$. Be aware the black area in I_3 represents masked pixels. The horizontal selective projection $SPF_h(y)$ and vertical selective projection $SPF_v(x)$ in intervals $[y - \Delta y,\ y + \Delta y]$ and $[x - \Delta x,\ x + \Delta x]$ are, respectively, defined as:

$$SPF_h(y) = \sum_{y'=y-\Delta y}^{y+\Delta y} \sum_{x'=0}^{w-1} I_3(x',\ y'),\ I_3(x',\ y') \neq 0, \quad (2)$$

$$SPF_v(x) = \sum_{x'=x-\Delta x}^{y+\Delta x} \sum_{y'=0}^{h-1} I_3(x',\ y'),\ I_3(x',\ y') \neq 0. \quad (3)$$

In other words, $SPF_v(x)$ is intensity summation of unmasked pixels in columns from $x - \Delta x$ to $x + \Delta x$; and the response of $SPF_h(y)$ is intensity summation of unmasked pixels in rows from $y - \Delta y$ to $y + \Delta y$. In I_3, eye center still has more number of unmasked pixels than other locations. Moreover, the pixels around eye center in I_3 now have higher intensity, instead of lower intensity in I_2. Therefore, it can guarantee that intensity summation of eye center is larger than other places. That is, eye center corresponds to the row with largest $SPF_h(y)$ and the column with the largest $SPF_v(x)$.

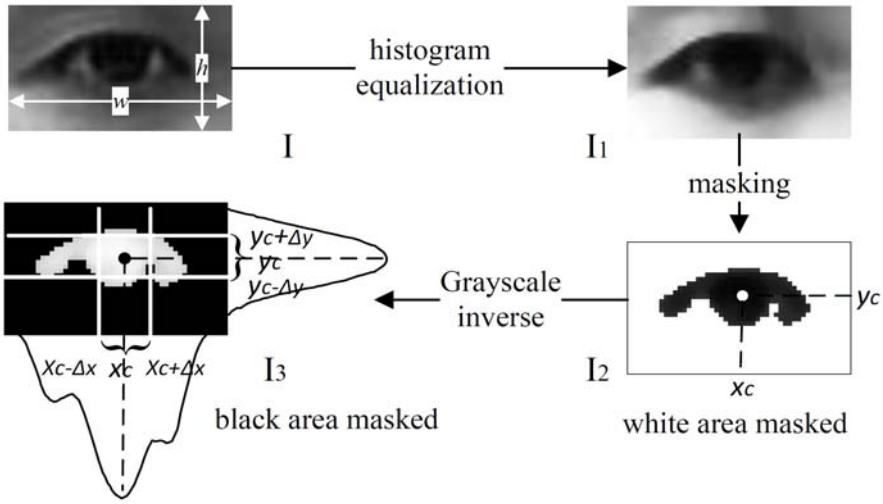


Figure 2. Illustration of accurate eye localization using SPF.

In summary, SPF selectively picks not only more number but also higher intensity pixels at eye center to get the largest projection response (summation). SPF has many good properties. First, it is efficient and simple to implement. Second, using row and column intervals enhances the consistency of PF against lighting changes. Comparatively, IPF and VPF are less capable of handling such changes. Furthermore, eye whiter and severe reflective highlights have few effects on SPF response as the corresponding pixels will be masked. While in IPF and VPF, lighting and eye white can easily degrade the performance as the responses are generated from all pixels in the row or column. For these reasons, SPF has better ability to locate eye center than conventional PFs.

3. Eye State Recognition

3.1. Preprocessing

Before ESR, preprocessing phase is performed to normalize the eye patches. Given the accurate positions of two eyes, face alignment should be implemented to ensure the two eyes are horizontally aligned and with a fixed between distance D. Roughly we choose $D = 70$ to ensure that big scaling

will not happen during future normalization of cropped eyes. After obtaining the normalized face image by interpolation operation from the original image, the eye patch is cropped from the normalized image with 40×20 dimension centered from the accurate eye center position. Often, severe lighting effects such as shadowing and highlights occur around eye area, and deteriorate the performance of eye-state recognizer. Thus, the illumination correction procedure [17] is conducted on the eye image before ESR. The method incorporates three stages, gamma correction, the Difference of Gaussian (DOG) filtering, and contrast normalization, to alleviate the effects of illumination variations, local shadowing and highlights, while preserving the essential elements of visual appearance.

3.2. Eye descriptor by grayscale correlogram

Correlogram, introduced by Huang et al. [18], describes the spatial correlation of color pairs at different distances. The correlograms were originally used for indexing and retrieval of images in RGB space. In this paper, it is used in grayscale image. Essentially, the grayscale correlogram of a $w \times h$ image I is a table indexed by grayscale pairs $\langle i, j \rangle$ and distance k. It describes the spatial correlations of grayscale pairs. Let $[m]$ denote the set of quantized grayscales 1, 2, ..., m, and $[d]$ denote a set of d fixed distances 1, 2, ..., d, which are measured by Chebyshev distance (maximum of horizontal and vertical differences). The size of a correlogram is $m^2 d$.

Consider a simple case of $m = 2$ and $d = 5$ as shown in Figure 3. Obviously, the grayscale histograms of two dashed regions in (a), (b) are identical. The spatial correlation is expressed by counting the occurrences of grayscale pairs at different distances. For example, the occurrences of pair $\langle 0, 0 \rangle$ with 1-pixel distance are marked in the figure (note the pairs are with directions). Comparing the column diagrams in (c), (d), it is clear that the spatial correlations of grayscale pairs for the two patterns are distinct. Basically, the principle of grayscale correlogram relies on the spatial correlation, and the follows present the mathematical derivation.

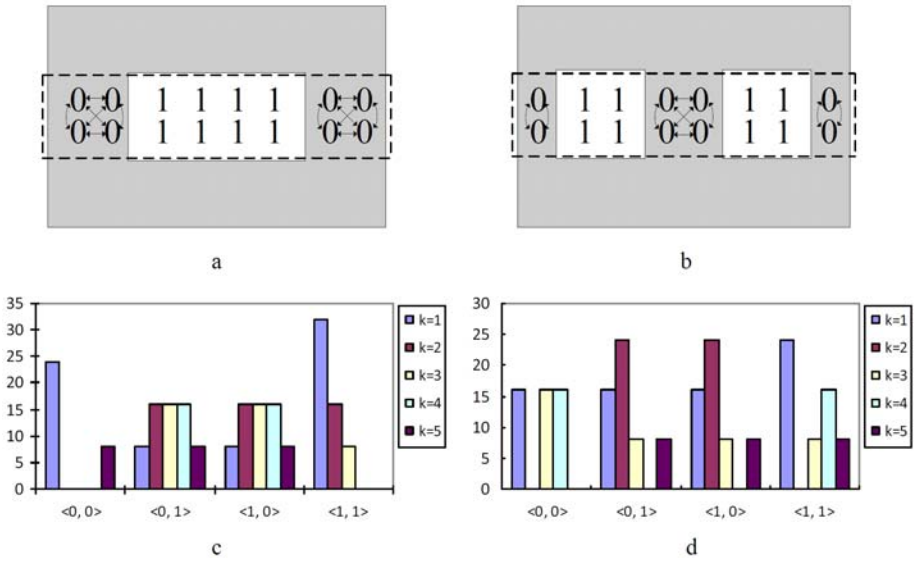


Figure 3. a, b show two binary images with identical histograms but different spatial correlations of grayscale pairs. c, d gives the occurrence statistics of different grayscale pairs at different distances corresponding to the two image patterns.

For a pixel $p(x, y) \in I$, let $I(p)$ be its grayscale, and $I_g \triangleq \{p \,|\, I(p) = g\}$. Then $p \in I_g$ is equal to $p \in I$, $I(p) = g$. The histogram of I is defined for grayscale g_i as:

$$h_{g_i} \triangleq wh \Pr_{p \in I}\{p \in I_{g_i}\}, \tag{4}$$

where h_{g_i}/wh corresponds to the probability of any pixel p in I being of grayscale g_i. Suppose $|\, p1 - p2\,|$ depicts the pixel-to-pixel distance, i.e., $|\, p1 - p2\,| = \max\{|\, x_1 - x_2\,|, |\, y_1 - y_2\,|\}$. The grayscale correlogram $\gamma^k_{g_i, g_j}$ for $i, j \in [m]$, $k \in [d]$ is defined as:

$$\gamma^{(k)}_{g_i, g_j} \triangleq \Pr_{p1 \in I_{g_i},\, p2 \in I}\{p2 \in I_{g_j} \,||\, p1 - p2\,| = k\} \tag{5}$$

which specifies the probability that given any pixel $p1$ of grayscale g_i, a

pixel $p2$ at distance k from $p1$ is of grayscale g_j. To compute $\gamma_{g_i, g_j}^{(k)}$, the first step is to count the occurrence of grayscale pair $\langle g_i, g_j \rangle$ at a distance k. Let $\Gamma_{g_i, g_j}^{(k)}$ denote the occurrences,

$$\Gamma_{g_i, g_j}^{(k)} \triangleq | \{p1 \in I_{g_i},\ p2 \in I_{g_j} \,||\, p1 - p2| = k\} |. \tag{6}$$

Regardless of boundaries, equation (5) can be rewritten as

$$\gamma_{g_i, g_j}^{(k)} = \Gamma_{g_i, g_j}^{(k)} / (h_{g_i}(I) \cdot 8k), \tag{7}$$

where the denominator counts the pixels that are at a distance k from any pixel of grayscale g_i, and the factor $8k$ is due to Chebyshev distance metric.

Correlogram has many good properties for eye representation. First, it is easy and efficient to compute. Second, it has low feature dimension $(m^2 d)$, which is only several hundreds as described in later experiment. Third, it maintains the merits of histogram and can distinguish images whose histograms are the same. More importantly, correlogram is invariant to in-plane rotation as it characterizes the global statistics of spatial correlations. This point is especially beneficial for ESR as eye appearance displays large variations. Also, note that correlograms of pair eyes are identical, a single correlogram-based eye-state classifier is adequate for both eyes.

3.3. Learning eye-state classifiers

Eye-state recognizer is learned from a set of open/closed eye samples using the eye descriptor. Different learning methods including subspaces (PCA and LDA), SVM and Discrete AdaBoost (DAB) are thoroughly studied and compared. First we give some common notations here: (x_i, y_i), $i = 1, ..., s$ denote a training set with s samples, where $x_i \in R^n$ and $y_i \in \{-1, 1\}$. The set consists of s_1 open eyes $(y = 1)$ and s_2 closed eyes $(y = -1)$. For convenience, the global and within-class means are respectively notated as μ, μ_1 and μ_2.

3.3.1. Subspace methods

Subspace methods linearly reduce feature dimensions in the form of $z_i = W^T x_i$. PCA [19] and LDA [20] are two most popular techniques to find the transformation matrix *W*. PCA searches the most informative basic vectors that can best describe the samples. Columns of *W* for PCA are the eigenvectors associated with the largest eigenvalues of covariance matrix:

$$\Sigma = \frac{1}{s}\sum_{i=1}^{s}(x_i - \mu)(x_i - \mu)^T. \tag{8}$$

For ESR, the class means μ_1, μ_2 and a new image *x* are first projected onto the eigenspace. Then the label of *x* is decided by a nearest neighbor classifier. In this paper, the distance between two feature vectors are measured by Mahalanobis distance. On the other hand, LDA searches those vectors in the original space that can best split different classes. The goal of LDA is to maximize the between class scatter

$$S_B = \frac{1}{s}\sum_{i=1}^{2} s_i(\mu_i - \mu)(\mu_i - \mu)^T \tag{9}$$

while minimizing the within-class scatter

$$S_W = \frac{1}{s}\sum_{i=1}^{2}\left[s_i \sum_{j=1}^{s_i}(x_i^j - \mu_i)(x_i^j - \mu_i)^T \right], \tag{10}$$

where x_i^j is the *j*th sample of class *i*. Euclidean distance is adopted in the nearest neighbor classifier for LDA. Relevant parameters for PCA-based and LDA-based classifier are the number of adopted basic vectors.

3.3.2. Support vector machines (SVM)

SVM [21] has gained great success in pattern classification. It maps data onto a higher dimensional space, and then finds a linear separating hyperplane with maximal margin to split different classes. An input sample is classified by function: $f(x) = sign\left(\sum_{i=1}^{s}\alpha_i y_i K(x_i, x) + b\right)$, where α_i are

Lagrange multipliers of the dual optimization problem, *b* is the threshold of the hyperplane, and *K* is a kernel function. Support vectors are those training samples whose α_i are over zero. The desired hyperplane of SVM is the one with maximal distance to the support vectors. Here the eye-state recognizer uses RBF kernel: $K(x_i, x_j) = \exp(-\gamma \| x_i - x_j \|^2)$, where γ is the kernel parameter. The penalty *C* of the incorrectly classified sample is another parameter for RBF SVM. As suggested in [22], the parameter set (C, γ) is determined by a grid-search for best ESR accuracy, and $(C, \gamma) = (2.5, 0.5)$ are finally adopted.

3.3.3. Discrete AdaBoost (DAB)

Discrete AdaBoost is another powerful supervised learning algorithm [15]. Basically, the concept is to combine a sequence of better-than-chance weak classifiers to produce a strong classifier. These weak classifiers are often simple and computationally inexpensive. For example, the most popular week classifiers are to find the best threshold for each variable. A two-class DAB classifier described in Figure 4 is examined for ESR. First, each sample is initially assigned with identical weight. Then in each round, a weak classifier with least weighted error is chosen, and all sample weights are updated. The final classifier $F(x)$ is the sign of the weighted sum over the individual weak classifiers. One parameter of DAB is number of training round R_d. Finally, $R_d = 35$ is chosen from interval [1, 60] using method of exhaustion.

1. Given s examples (x_i, y_i), with $x_i \in R^n$, $y_i = \pm 1$, $i = 1, ..., s$.
2. Every sample x_i starts with the same weight: $w_{1,i} = 1/s$.
3. For round $t = 1 \cdots R_d$, do:
   - For each variable j, train a weak classifier $h_j(\,)$. Evaluate the weighted error of the classifier $\varepsilon = \sum_i w_{t,i} b_i$. $b_i = 0$ if x_i is correctly classified; otherwise, $b_i = 1$.
   - Select a weak classifier with lowest classification error ε_t.
   - Update weights: $w_{t+1,i} = w_{t,i} \beta_t^{1-b_i}$, with $\beta_t = \varepsilon_t / (1 - \varepsilon_t)$.
4. The objective function of the strong classifier is: $F(x) = sign\left(\sum_{t=1}^{R_d} \alpha_t h_t(x) \right)$, with $\alpha_t = \lg(1/\beta_t)$.

Figure 4. Discrete AdaBoost algorithm.

4. Experiment and Performance Evaluation

4.1. Performance of eye localization

To measure the precision of EL, the resolution-independent measure criterion described in [16] is adopted. Let d_l and d_r be the distances between the estimated locations of eye centers and the true positions C_l, C_r. The normalized error of EL is defined as: $d_{err} = \dfrac{\max(d_l, d_r)}{| C_l - C_r |}$. The denominator refers to inter-ocular distance. As real eye width has been shown to be roughly half of inter-ocular distance [16], hence an error of $d_{err} < 0.25$ means the localization error is no more than half of eye width. Often, $d_{err} < 0.25$ is used to verify the presence of eye window. The proposed rough eye window detector also uses this criterion.

Experiments are performed on two test sets: Normal subset of CAS-PEAL face database [23] and BioID face database [16]. The former set consists of 1040 images (360×480) from 1040 subjects. It displays large person-to-person variations in eye appearances. Lighting and background is relatively simple. BioID set includes 1520 images (381×288) from 23 subjects. It was recorded during several sessions at different places. BioID set features various illuminations, changing face scales, spectacles and complex background.

Table 1. The results of coarse eye-region detector

Dataset	CAS-PEAL	BioID
Total	1040	1521
Detected number	1013	1387
Detection rate	97.40%	91.19%

Table 1 gives the results of rough eye window detection $(d_{err} < 0.25)$. Better accuracy is gained on CAS-PEAL than BioID. The performance slightly decreases when large variations happen. Even that, the detection result (over 91%) is still encouraging and acceptable to our work. Alternative face detection and rough eye window detection algorithms may be adopted to achieve better result, but it is not the main concern in this paper.

ESR relies on localization precision of eye center. Given rough eye windows are detected, VPF, IPF, and the proposed SPF, are compared to locate eye centers. First, the relevant parameters of SPF should be optimized. Three relevant parameters are C in equation (1), and Δx, Δy in equations (2) and (3), respectively. Experiments are conducted on BioID database to decide these parameters based on harsher criterion $d_{err} < 0.10$. Δx and Δy do not have too many choices due to constraint of rough eye window size (approximately 40×20). Hence, C is first optimized by fixing Δx and Δy. Figure 5 shows the "C-Precision" curves on BioID database. The precision expresses the percentage of images whose eye centers are located within the criterion $d_{err} < 0.10$. Clearly, these curves have common trend and reach

highest precision near $C = 0.25$. Thus, $C = 0.25$ is selected as the optimal value. Next, fixing $C = 0.25$, different combinations of Δx and Δy are evaluated. As mentioned, Δx and Δy are restricted by eye window size. Table 2 shows the results of eye center localization. According to the table, $\Delta x = 1, \Delta y = 1$ is selected. That is, SPF utilizes 3 rows or columns to calculate projection response. Overall, SPF parameters are chosen as $C = 0.25, \Delta x = 1, \Delta y = 1$.

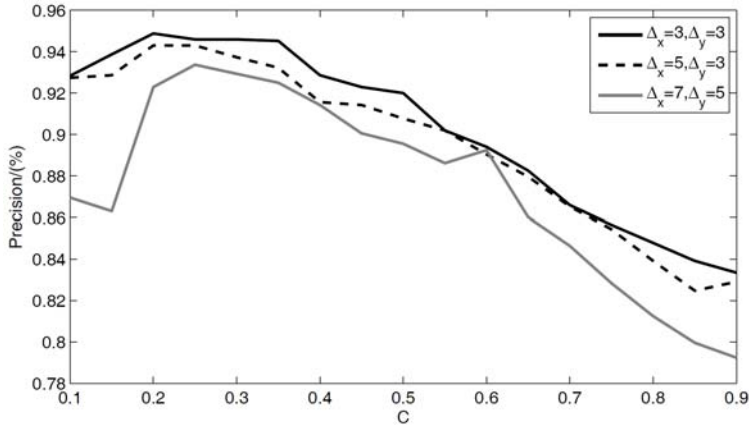


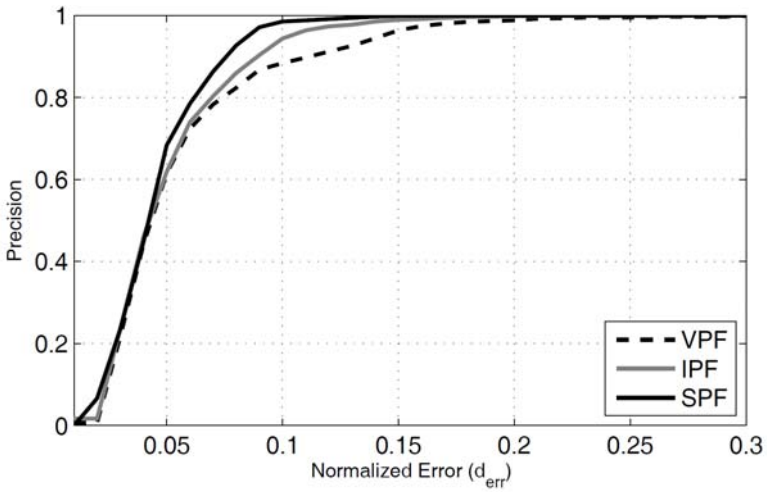
Figure 5. Selection of SPF parameter: C.

Table 2. Eye center localization rate of different Δ_x and Δ_y (%)

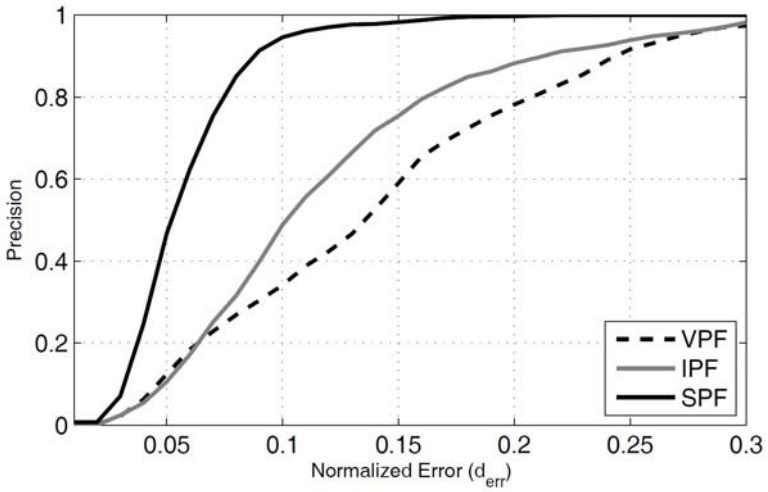
	Δ_x					
		1	2	3	4	5
Δ_y	1	94.59	94.30	93.58	89.74	79.38
	2	94.16	93.94	93.37	89.40	79.81
	3	93.58	93.44	92.14	87.53	78.95

The curves in Figure 6 depict the localization rate at different normalized error $(d_{err} = 0.2, 0.15, 0.10, 0.05)$. Obviously, the proposed method significantly outperforms IPF and VPF in both test sets. Taking criterion $d_{err} = 0.1$ for instance, all three PFs demonstrate promising accuracy in relatively simpler CAS-PEAL database: the successful rates are 88.35%,

94.37% and 98.37% for VPF, IPF and SPF respectively. SPF performs the best. When it comes to difficult BioID set, localization rates of VPF and IPF rapidly fall to only 34.10% and 48.81% under same criterion. Comparatively, accuracy of SPF slightly drops to 94.59%, which is tolerant. Therefore, SPF demonstrates higher accuracy and better consistency to large variations.



(a) CAS-PEAL



(b) BioID

Figure 6. Performance comparison of VPF, IPF and SPF on two databases.

4.2. Performance of eye state recognition

4.2.1. Experimental data for ESR

The proposed ESR methods are evaluated using the eye samples partially cropped from three databases: BioID, CAS-PEAL (normal and expression subsets), and AR [24]. The images from the three databases have different resolutions and large variations in illumination, expression and background. Meanwhile, since the closed eye samples are not enough, 736 extra closed eyes are collected from 30 persons. Table 3 shows the statistics. Note that the eye patches are cropped with a size of 40×20 from the geometric normalized face image using manually marked eye centers, and then fed into lighting correction procedure. Based on grayscale correlogram eye descriptor, four eye-state classifiers (PCA, LDA, SVM, DAB) are learned from these canonical patches. For consistency, all of them utilize the training subset of Combo dataset: 2,489 open eyes and 1,131 closed ones as described in Table 3.

Table 3. Statistics of used eye samples for eye state recognition

Dataset	Training		Testing		Total
	Open	Closed	Open	Closed	
BioID	958	27	959	27	1,917
CAS-PEAL	972	972	972	273	2,689
AR	559	96	559	96	1,310
Collected	0	636	0	100	736
Combo	2,489	1,131	2,490	596	6,706

4.2.2. Parameter optimization

Two important parameters for grayscale correlogram descriptor are: number of quantized grayscales *m* and maximal pair distance *d*. For image of size $w \times h$, we have $d \le \frac{1}{2}\min(h, w)$, i.e., $d \le 10$. On the other hand, *m* should not be two big for computational reason. Hence, range of *m* is restricted within $2 \sim 15$. This work adopts fix-one-and-optimize-another method to search optimal *m* and *d*, and DAB classifier is used for

experiment. Figure 7 gives the *m*-accuracy and *d*-accuracy curves. The accuracy is the percentage of correctly recognized eyes. The two curves labeled "mean" are the average curves of all *m*-accuracy or *d*-accuracy curves. From left sub-figure, $\{7, 8, 9\}$ may be optimal for *m*. Likewise, right sub-figure suggests $\{7, 8, 9\}$ may be optimal for *d*. Then all possible 9 combinations of (m, d) are checked. Finally $(m, d) = (8, 7)$ is the optimal parameter set. In fact, exhaustive method has also been implemented and proved that (8, 7) is optimal. Thus, the feature dimension is $8^2 \times 7 = 448$.

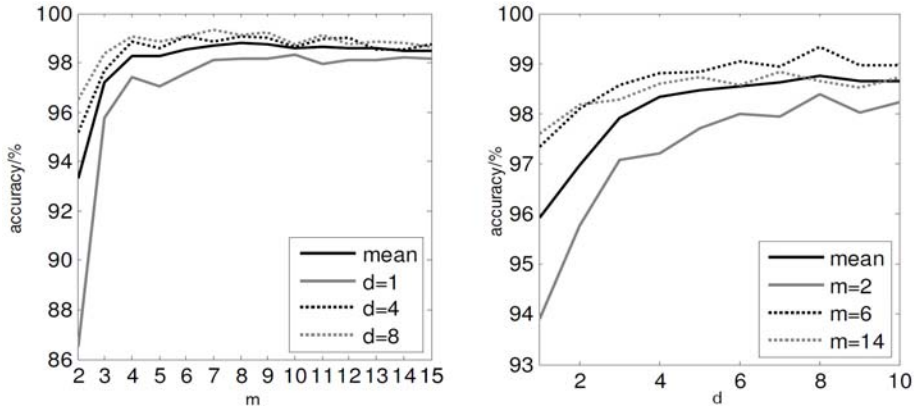


Figure 7. Parameter optimization, left: *m*-accuracy curves by fixing *d*; right: *d*-accuracy curves by fixing *m*.

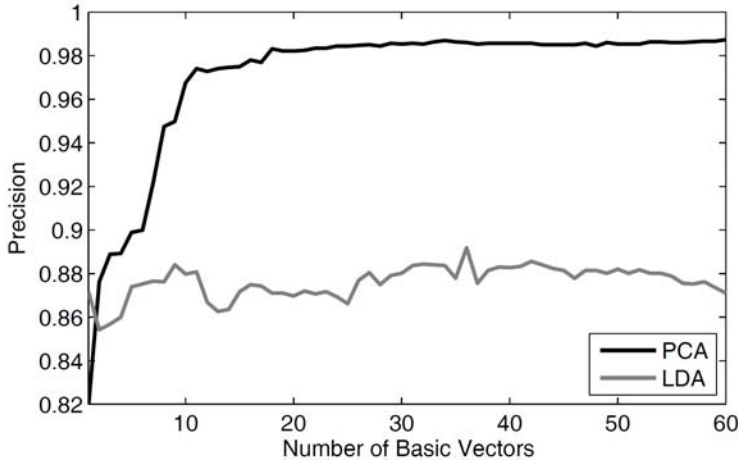


Figure 8. Selection of the number of basic vectors for PCA and LDA.

On classifiers side, numbers of basic vectors need to be optimized for PCA-based and LDA-based methods. Using optimal descriptor parameters above, the trend for ESR accuracy with the growing of number of basic vectors is given in Figure 8. Seemingly, PCA-based classifier performs much better than LDA-based classifier in ESR issue. To get best precision, 35 and 36 basic vectors are respectively adopted by PCA-based and LDA-based classifiers.

4.2.3. ESR results

The performance is measured by the *Positive Recognition Rate* (PRR) versus the *Negative Recognition Rate* (NRR). PRR is percentage of correctly recognized open eyes, while NRR is the percentage of correctly recognized closed eyes. Table 4 displays the results on the testing subset of Combo, which are encouraging. Be aware that our method uses only 448 features. Actually, PCA, SVM and DAB based classifiers can be independently utilized to recognize eye states. Among four classifiers, SVM and DAB demonstrate comparably the best performance. SVM achieves a PRR of 99.24% and NRR of 99.33%. Similarly, DAB achieves 99.18% and 99.66% respectively. In [14], Sun and Ma have reported an overall accuracy of 94.93% using SVM + LBP histograms. In [13], grayscale characteristics, upper eyelid bending direction and circular similarity have been utilized to verify eye states, and the authors declared 86.9% PRR, 90.0% NRR on BioID, and 94.3% PRR, 84.87% NRR on Normal subset of CAS-PEAL. Figure 9 presents some recognition samples. Through the misclassified samples, it is observed that these half-open eyes tend to be wrongly classified. This indicates that more half-open eyes may be used during training, or alternatively a ternary (open, half-open, closed) eye state classifier can be trained in the future.

Table 4. The accuracies for four eye-state classifiers on the combo set (%)

Classifiers	PCA	LDA	SVM	DAB
PRR	98.59	88.88	99.24	99.18
NRR	99.16	90.44	99.33	99.66

Figure 9. Some samples of the eye state recognition results: T-true, F-false, P-positive (open), N-negative (closed).

4.3. Real time analysis

Finally, we fuse presented eye localizer and eye state recognizer to an automatic ESR system. Performance evaluation constitutes two aspects: speed and accuracy. On our laptop (Visual Studio 2008, Intel Core Dual 2.53GHz, 2GB RAM), the system processes 640×480 images captured from a common webcam at a speed of 18 fps. Unfortunately, lack of standard databases and evaluation platforms limit analysis of the real time system. In order to make our evaluation results valuable to future researcher, the ESR system is evaluated on the Talking Face Video [25]. The video consists of 5000 frames taken from a video of a person engaged in conversation. It corresponds to about 200 seconds of recording which contains moderate number of closed-eye frames. The sequence was taken to model the behavior of the face in natural conversation. The system is divided into three modules for accuracy evaluation: rough eye window detection, eye center localization (SPF) and eye state recognition. Criteria $d_{err} = 0.10, 0.15$ are used to evaluate eye center localization. ESR accuracy is evaluated using SVM and DAB classifiers, because they achieved highest accuracy for static images. It is measured by the percentage of correctly recognized eyes.

Table 5 shows the results of the ESR system. With d_{err} increases from 0.10 to 0.15, overall accuracy declines slightly. The reason is that the eye localization rate increases a bit more than the decreasing amount of ESR rate. Also, it is observed that both SVM and DAB ESR classifiers still demonstrate promising accuracy. Compared to static images, ESR accuracy drops slightly. The reason is that the video includes more half-open eyes than static database, because the video recorded continuous frames. It is in

accordance with the observation in previous part that half-open eyes are often misclassified.

Table 5. Accuracy of ESR system (%)

	SVM		DAB	
Rough eye window detection	98.64			
Eye center localization	d_{err} = 0.10	d_{err} = 0.15	d_{err} = 0.10	d_{err} = 0.15
	95.40	97.41	95.40	97.41
Eye state recognition	98.24	97.64	97.92	97.37
Overall accuracy	92.45	93.82	92.15	93.56

5. Conclusions

This paper proposes a novel system to precisely detect eyes and recognize the open/closed states. For eye localization, a coarse-to-fine approach is introduced: the frontal face is first detected and divided into two halves, where two cascaded eye window detectors are trained to detect rough eye windows; afterwards, the selective projection function is novelly presented to get the accurate eye centers. Meanwhile, this paper also presents several methods for eye state recognition. The problem is considered as a two-class classification task. Originally, a discriminative eye descriptor is developed using the spatial correlations of pixel pairs. Then different classification techniques, including PCA, LDA, SVM and DAB are examined on several static databases. Experiments show that the proposed eye localizer and eye-state classifiers demonstrate promising accuracies. Furthermore, the presented eye localizer and eye state recognizer are fused to an automatic system, which displays high speed and promising accuracy on a publicly available video.

References

[1] Z. Zhang and J. Zhang, Driver fatigue detection based intelligent vehicle control, 18th Int. Conf. on Pattern Recognition, Hong Kong, China, 2006, pp. 1262-1265.

[2] K. Grauman, M. Betke, J. Lombardi and J. Gips, Communication via eye blinks and eyebrow raises: video-based human-computer interfaces, Universal Access in the Information Society 2(4) (2003), 359-373.

[3] Z. Liu and H. Ai, Automatic eye state recognition and closed-eye photo correction, 19th Int. Conf. on Pattern Recognition, Tampa, Florida, USA, 2008, pp. 1-4.

[4] D. W. Hansen and A. E. C. Pece, Eye tracking in the wild, Comput. Vis. Image Underst. 98(1) (2005), 182-210.

[5] Y. Li, S. Wang and X. Ding, Eye/eyes tracking based on a united deformable template and particle filtering, Pattern Recog. Lett. 31(11) (2010), 1377-1387.

[6] Z. Zhou and X. Geng, Projection functions for eye detection, Pattern Recog. 37(5) (2004), 1049-1056.

[7] G. C. Feng and P.c. Yuen, Variance projection function and its application to eye detection for human face recognition, Pattern Recog. Lett. 19(9) (1998), 899-906.

[8] W. M. K. W. M. Khairosfaizal and A. J. Nor'aini, Eye detection in facial images using circular hough transform, 5th Int. Colloquium on Signal Processing and Its Applications, 2009, pp. 238-242.

[9] J. Huang and H. Wechsler, Eye detection using optimal wavelet packets and radial basis functions (rbfs), Int. J. Pattern Recog. Arti. Intell. 13(7) (1999), 1009-1025.

[10] L. Jin, X. Yuan, S. Satoh, J. Li and L. Xia, A hybrid classifier for precise and robust eye detection, 18th Int. Conf. on Pattern Recognition, Hong Kong, China, 2006, pp. 731-735.

[11] Q. Wang and J. Yang, Eye location and eye state detection in facial images with unconstrained background, J. Inform. Comp. Sci. 1(5) (2006), 284-289.

[12] H. Tan and Y. J. Zhang, Detecting eye blink states by tracking iris and eyelids, Pattern Recog. Lett. 27(6) (2006), 667-675.

[13] F. Wang, M. Zhou and B. Zhu, A novel feature based rapid eye state detection method, IEEE Int. Conf. on Robotics and Biomimetics, 2009, pp. 1236-1240.

[14] R. Sun and Z. Ma, Robust and efficient eye localization and its state detection, Proc. of 4th Int. Symposium on Advances in Computation and Intelligence, 2009, pp. 318-326.

[15] P. Viola and M. Jones, Rapid object detection using a boosted cascade of simple features, IEEE Int. Conf. on Computer Vision and Pattern Recognition, USA, 2001, pp. 551-518.

[16] O. Jesorsky, K. Kirchberg and R. Frischholz, Robust face detection using the Hausdorff distance, Lect. Notes. Comput. Sci. 2091, 2001, pp. 90-95.

[17] X. Tan and B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, Int. Workshop on Analysis and Modeling of Faces and Gestures, 2007, pp. 168-182.

[18] J. Huang, S. R. Kumar, M. Mitra, W. Zhu and R. Zabih, Image indexing using color correlograms, Proc. Int. Conf. on Computer Vision and Pattern Recognition, 1997, pp. 762-768.

[19] M. Turk and A. Pentland, Eigenfaces for recognition, Int. J. Cognitive Neurosci. 3(1) (1991), 71-86.

[20] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, Eigenfaces vs. fisher-faces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19(7) (1997), 711-720.

[21] C. J. C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery 2(2) (1998), 121-167.

[22] C. W. Hsu, C. C. Chang and C. J. Lin, A practical guide to support vector classification, 2003. http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.

[23] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang and D. Zhao, The CAS-PEAL large-scale Chinese face database and baseline evaluations, IEEE Trans. Sys. Man Cyber. Part A 38(1) (2008), 149-161.

[24] A. R. Martinez and R. Benavente, The AR face database, Technical Report 24, Computer Vision Center (CVC) Technical Report, Barcelona, Spain, 1998.

[25] Talking Face Video, Face and Gesture Recognition Working Group, IST-2000-26434.
http://www.prima.inrialpes.fr/FGnet/data/01-TalkingFace/talking_face.html.