Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2016) 24: 1586 – 1603

doi:10.3906/elk-1312-150

An efficient hybrid eye detection method

Mingxin YU[1,2,*], Yingzi LIN[2], Xiangzhou WANG[1]
[1]School of Automation, Beijing Institute of Technology, Beijing, P.R. China
[2]Department of Mechanical and Industrial Engineering, Northeastern University, Boston, MA, USA

Received: 20.12.2013 • **Accepted/Published Online:** 07.04.2014 • **Final Version:** 23.03.2016

Abstract: Eye detection is the most important and critical task of diverse applications such as face detection and recognition. However, most eye detection methods do not fully consider detection robustness to people with glasses, illumination variation, head pose change, and eye occlusions. This paper proposes an efficient hybrid eye detection method based on a gray intensity variance filter (VF) and support vector machines (SVMs). Firstly, the VF is used for eliminating most of noneye region images to keep less candidate eye regions. Then accurate two eye regions are determined easily through the trained SVM classifier. Moreover, this paper provides an assessment of the sensitivity of obtained parameters in the SVM classifier on eye detection accuracy. The proposed method was evaluated on different face databases. The experimental results show that the method can improve the performance of eye detection and achieve a higher detection accuracy compared with state-of-the-art methods.

Key words: VF, SVM, eye detection

1. Introduction

Eyes are more prominent features of the human face in comparison with other features such as the nose or mouth. Hence, eye detection has been widely studied for diverse applications including drowsiness detection for intelligent vehicle systems, eye gaze tracking devices, human–machine interface for multimedia devices, human–machine collaboration, and automatic face detection and recognition systems. In recent years, many eye detection methods have been developed by researchers. According to the principles of these methods, eye detection techniques can be divided into three categories: feature-based, appearance-based, and hybrid-based methods [1].

The feature-based method takes advantage of the distinctive features from the eye image, e.g., limbus, iris intensity, and eye shape, to find eye regions. Feng and Yuen [2] used three cues for detecting eye windows on gray-scale face images. The cues were achieved through face intensity, direction of the line joining the centers of eyes, and eye variance filter (VF). Each cue indicated the candidate eye regions. The precise eye location was determined and validated by the variance projection function. The method was tested on the face database of MIT AI laboratory and achieved a 92.5% detection rate. By combining the integral projection and variance projection functions, Zhou and Geng [3] extended two projection functions to a hybrid projection function, which inherits both the robustness of integral projection and the sensitiveness of the variance projection function. The performance of eye detection is determined by the optimal parameters of the hybrid projection function. Through experiments, the detection rate is up to 95.6% tested on a Biometric Identity (BioID) face

*Correspondence: mingxinneu@gmail.com

database. Guan [4] adopted horizontal and vertical gradient projection methods on the second derivate image of the human face to extract the horizontal and vertical position of the eyes. The eye region detection rate could be up to 96.2% on the CVL face database. However, in the above projection functions the images were required to have uniform background and lighting conditions. Ando et al. [5] employed the feature-based approach. The method used a seven-segment rectangular filter over the image to find the between-the-eyes feature of the human face for locating the eye regions. A detection rate of 88% was achieved on the color FERET database.

The appearance-based method replies on models constructed directly by the appearance of eye regions. Ryu and Oh [6] developed an algorithm that used eigenfeatures derived from eigenvalues and eigenvectors of the binary edge data set from eye fields and eigenvectors of the binary edge dataset and neural networks for eye region detection. The method tested 180 face images without glasses from the Olivetti Research Laboratory (ORL) database and the detection rates of left and right eyes were 91.7% and 85.5%, respectively. Wang and Ji [7] presented a statistical learning method to extract features and construct a classifier for eye detection. Specifically, the proposed recursive nonparametric discriminant analysis (RNDA) was learned and combined with AdaBoost to form an eye detector. Because the process of classification needs to distinguish a large amount of noneye images, such as the nose, eyebrow, mouth, and teeth, the performance of the eye detector mainly depends on the diversity of eye and noneye images. The detection rate was 94.5% for the frontal face images from the Face Recognition Grand Challenge (FRGC) V1.0 database. The advantage of this method is its high computational efficiency. Wu and Trivedi [8] used a binary tree to create a model of the statistical structure of human eyes for locating accurate eye regions. With a clustering method based on the pairwise mutual information between features, the dependent features were separated into different subsets. The process of eye detection would be repeated until the features of the subset have enough mutual information. The final detection is achieved through the Bayesian criterion. The highest detection rate achieved by the method was 92.43% on 317 face images from the FERET database. Nevertheless, the approach spent a longer time on the initial position of the template.

The hybrid-based method combines the advantages of two methods or more into one system to detect the eye region. Peng et al. [9] proposed a combination approach of a feature-based method and a temple-based method to detect rough eye regions and the detection rate was 95.2% for the ORL face database without glasses. Xia and Su [10] adopted a feature extraction method for locating probable human-eye areas, finding the accurate eye regions with a modified template-matching method. The method achieved a detection rate of 94.7% on the ORL face database. Campadelli et al. [11] used a two-step technique. At first, template matching was utilized to find the sets of eye bands. Secondly, with a neural network classifier and a measure of distance, the eyes could be localized. A detection rate of 95.7% was achieved on the 750 face images without glasses from the XM2VTS face database. Hassballah et al. [12] first adopted the gray intensity variance to locate candidate eye regions; then independent components analysis (ICA) was used for eye region detection. The method was robust to eyes with glasses and obtained high detection rates of 98.4%, 97.1%, and 97.3% tested on the XM2VTS, BioID, and FERET face databases, respectively. Laxmi and Rao [13] proposed a hybrid eye detection method based on Gabor filters and support vector machines (SVMs). The Gabor filters were trained with larger data sets to reduce false positive detection. Then eye/noneye images rotated by different angles with the Gabor filter were used to train the SVM. The method was robust to rotational facial images and achieved a 96% detection rate on the GTAV and VITS databases. Kalbkhani et al. [14] used many algorithms to locate the eye. Firstly, the method created EyeMap and EyeBin used in original images for obtaining binary images, then removing border touching, small and large connected regions to leave candidate eye regions. Lastly, the

eye regions were obtained using pairs satisfying distance conditions. The method had high detection rates of 96.98%, 94.29%, 98.65%, and 95.32% on the FERET, Aberdeen, IMM, and CVL databases, respectively, and robustness to different poses, light conditions, and glasses. Phromsuthirak and Umchid [15] first adopted the EyeMap algorithm to locate possible eye regions, and then the correct eye regions were determined using the proposed geometrical test methods. The evaluation results showed a detection rate of 92.3% for frontal facial images and 85.7% for rotational facial images.

Through analysis of the above literature, although many methods presented in their studies achieved good detection results, they did not fully consider eye detection robustness for people with glasses, different eye conditions (e.g., semiclosed eyes, closed eyes, and squints), rotation of the frontal face, and illumination changes. Furthermore, feature-based and appearance-based methods required some restricted conditions, e.g., fully open eyes, frontal face images, and uniform illumination. Compared to the two methods, the hybrid-based method achieved a better performance and higher detection rate in different conditions. In this paper, an efficient hybrid method based on VF and SVMs is introduced to detect human eyes. The method was verified by three face databases, i.e. Biometric Identity (BioID), Informatics and Mathematical Modeling (IMM), and the Face Recognition Technology (FERET), and its average eye detection rate was up to 97.5%. The robustness of the proposed method to eye with glasses, illumination, pose, and eye occlusion changes was also studied fully. The parameter sensitivity of the SVM model with eye detection was analyzed. In order to evaluate the performance of our method, the comparison results with the state-of-the-art systems for eye detection in standard and the detection accuracy in low resolution face images are given.

The rest of this paper is organized as follows. The methods used in our research are described in Section 2. Section 3 presents the process of eye detection. Experimental results are reported in Section 4. Finally, conclusions are provided in Section 5.

2. Methods

The proposed hybrid method for eye detection mainly consists of a VF and SVMs. This section describes the theory of the two methods and the processes of VF and SVM classifier construction used for our eye detection system are given. The constructed VF and trained SVM classifier are directly applied to the detection process of eye regions described in Section 3. In addition, because the two methods need a lot of eye and noneye images for VF construction and SVM training, an eye database is required to be established.

2.1. The eye database

The images used for establishment of our eye database are extracted from the Psychological Image Collection at Stirling (PICS) face database, Japanese Female Facial Expression (JAFFE) face database, Milborrow/University of Cape Town (MUCT) face database, self-face database, and face images from Internet websites. The PICS face database is a collection of images with many face sets. In this paper, the images are only extracted from 2D face sets including the Aberdeen set with 687 facial images of 90 individuals, the Pain Expressions set with 599 facial images of 23 individuals (13 women and 10 men), and the Utrecht ECVP set with 131 facial images of 69 individuals (20 women and 49 men). The JAFFE face database is constructed by Kyushu University in Japan with 213 images of 7 facial expressions posed by 10 Japanese female models. The MUCT face database is provided by the University of Cape Town in South Africa with 3755 face images. The self-face database is created by the author's Measurement & Testing Laboratory of Beijing Institute of Technology with 1000 face images from 10 people. The size of images in eye database is normalized to 48 $\times$ 24. The eye database

consists of the positive image set and the negative image set. The positive image set contains 7000 eye images with different gaze directions, different sizes, with and without glasses, facial rotation, under various lighting conditions, etc. Similarly, the negative image set contains 7000 noneye images with eyebrows, incomplete eye, skin, etc. Examples of eye and noneye images are shown in Figures 1 and 2.

Figure 1. Examples of eye training images. The first row eye images from PICS face database, the second row eye images from JAFFE face database, the third row eye images from MUST face database, the fourth row eye images from Internet websites, and the last row eye images from self-face database.

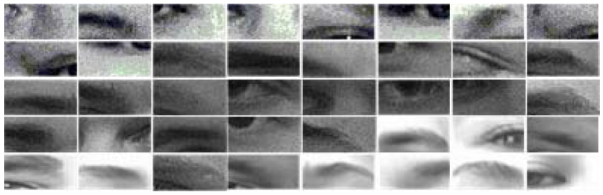

Figure 2. Examples of noneye training images.

2.2. Eye variance filter (VF)

This subsection gives the process of an eye VF construction. The goal of eye VF is to eliminate most of the noneye region images to leave fewer candidate eye regions in our proposed process of eye detection (See Section 3).

2.2.1. Eye variance images

The eye is an active model comprising an eyeball and two eyelids. Because the edge feature of an eye region is indistinct and susceptible to interference from eyebrows, a good edge of an eye region is very hard to obtain. However, the change in gray intensity in the eye region is more obvious than in other regions on the face. The second-order moment (or variance on a domain) indicates variation in gray intensity. The variance of the eye image $I(x,y)$ on the domain Ω can be defined as

$$\sigma_{\Omega} = \frac{1}{A_{\Omega}} \sum_{(x,y)\in\Omega} \left[I(x,y) - \overline{I_{\Omega}} \right]^2, \tag{1}$$

where A_{Ω} and $\overline{I_{\Omega}}$ are the area and average gray intensity on the domain Ω, respectively. The variance is a nonnegative value and has two properties [2]. Firstly, σ_{Ω} is rotation invariant on domain Ω. Secondly, σ_{Ω} reflects gray intensity variations rather than the exact shape on the domain.

The eye image of size 48 $\times$ 24 is divided into 16 $\times$ 8 nonoverlapped subblocks of size 3 $\times$ 3 pixels. For an image $I(x,y)$, the variance image is defined as

$$\begin{array}{l} V_{\sigma}(i,j) = \sigma_{\Omega}, \\ \Omega_{ij} = \{(i-1)\,l + 1 \le x \le il, (j-1)\,l + 1 \le y \le jl\} \end{array}, \tag{2}$$

where l and Ω_{ij} are the width/height (here, width = height) and the area in each subblock, respectively.

The variance image on each subblock is calculated by Eq. (2). Each subblock has different features of gray intensity. An example of an eye image from an eye database and its variance image are shown in Figure 3.

Figure 3. An eye image (a) and its variance image (b).

2.2.2. Eye VF construction

Based on the above observations about the variance image of an eye, an eye VF is constructed [2]. The 30 eye images with or without glasses are extracted from an eye database. The VF on the (i,j) subblock (called $F_e(i,j)$) is constructed by calculating the average of the variance image across all 30 eye images. $F_e(i,j)$ is defined as

$$F_e(i,j) = \frac{1}{N}\sum_{k=1}^{N}[V(i,j)]_k, \tag{3}$$

where $[V(i,j)]_k$ is the variance image $V(i,j)$ of the (i,j) subblock on the kth eye image and N is the amount of the eye images (here, $N = 30$). The eye VF is shown in Figure 4.

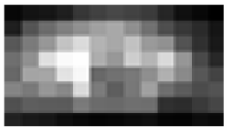

Figure 4. Eye variance filter image.

Then the correlation is calculated between the VF and eye/noneye variance image on the face in order to detect candidate eye regions. The correlation is defined as

$$R(V_\sigma, F_e) = \frac{E\left[\left(\xi_{V_{\sigma i}} - E\left(\xi_{V_{\sigma i}}\right)\right)\left(\xi_{F_e} - E\left(\xi_{F_e}\right)\right)\right]}{\sqrt{D\left(\xi_{V_{\sigma i}}\right)D\left(\xi_{F_e}\right)}}, \tag{4}$$

where $\xi_{V_{\sigma i}}$ and ξ_{F_e} are the concatenated vectors of the variance image of the possible eye regions and F_e. The $E(\cdot)$ and $D(\cdot)$ represent the mathematical expectation and variance of the random variable.

According to Eq. (4), the correlation value of a variance image with a VF can be obtained. Here, another 60 images including 30 eye images with or without glasses and 30 noneye images are extracted from an eye database for the calculation of correlation values. The variance images of the noneye images are constructed by the same manner as the eye variance image. The correlation values with respect to the images are plotted in Figure 5.

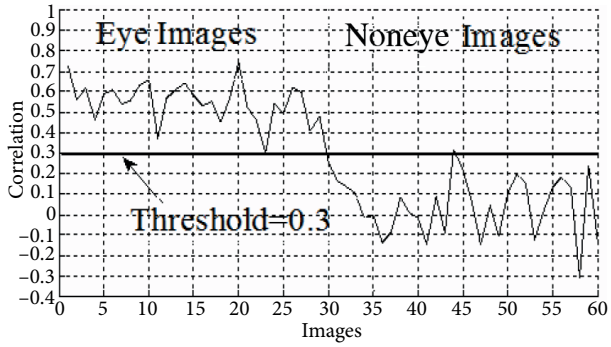

Figure 5. Correlation between eye/noneye images and VF image.

Figure 5 shows that the eye images have correlation values greater than 0.3, while the noneye region images have correlation values less than 0.3. Therefore, 0.3 can be taken as the threshold of the constructed eye VF to distinguish eye and noneye regions. However, the correlation values of some noneye images, e.g., from surrounding eyebrow above the eye region, are likely greater than the threshold value. In Figure 5, a correlation value of noneye images is a little higher than the threshold value. Hence, the role of the eye VF in the process of eye detection is used to find the most probable eye regions, i.e. candidate eyes (for details, see Section 3), and eliminate most of the noneye images.

2.3. Support vector machines (SVMs) classifier

This subsection describes the process of SVM classifier construction. Firstly, a description of the theory and selected kernel function of SVMs is given. Then principal component analysis (PCA) is used for dimension reduction and feature extraction of eye image data. In order to enhance the classification accuracy of SVMs, lastly the kernel parameters used in the SVM classifier are optimized by a genetic algorithm (GA). The constructed SVM classifier is used for determining accurate left and right eye regions (see Section 3).

2.3.1. Support vector machines (SVMs)

The theory of SVMs was proposed by Vapnik in 1995 [16]. The main idea of SVMs is to find an optimal hyperplane as a decision function in high-dimensional space.

In this paper, 3038 eye and 3164 noneye images of size 48 $\times$ 24 extracted from an eye database are used as positive and negative image sets for training the SVM classifier. The classifier distinguishes between eye images and noneye images, namely the binary classification problems. Suppose that there exists a hyperplane that could divide a sample space into two categories; one is the positive set (eye images) and the other is the negative set (noneye images). Then, in order to find the hyperplane, labeled samples (x_i, y_i) should be tested, where x_i is a vector of eye images, noneye images, or incomplete eye images (also called noneye images), and $y_i \in [-1, 1]$ is the class label. If there is a vector w and a scalar w_0, the classification can be obtained by the equation

$$y_i \left(w \cdot x_i + w_0 - 1 \right) \geq 0, \tag{5}$$

where w is the normal vector to the hyperplane and w_0 is the distance between the origin and hyperplane.

The hyperplane space can be defined as

$$f(x) = sign \left(w \cdot x + w_0 \right) \tag{6}$$

The SVM finds the separating hyperplane for maximizing the distance between the classes. The optimization problem can be solved by the equation [17]

$$\min\left(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{k}\zeta_i\right) \tag{7}$$

where ζ_i is a slack variable and C is a penalty factor for error control.

For a nonlinear decision surface, Vapnik [18] introduced the concept of using a kernel function K in the design of nonlinear SVMs. The kernel function can be defined as

$$K(x_i, x_j) = \Phi(x_i)\,\Phi(x_j). \tag{8}$$

The general kernel function has three categories: polynomial kernel function, Gaussian kernel function, and radial basis function (RBF) kernel function.

With the use of a kernel function, Eq. (6) becomes

$$f(x) = sign\left[\sum_{i=1}^{N}\lambda_i y_i K(x_i, y_i) + w_0\right], \tag{9}$$

where λ_i is a Lagrange multiplier.

The detailed theory of SVMs can be found in [16]. The RBF kernel function is used as the kernel function of the SVM in this paper. The RBF kernel function can be defined as

$$K(x_i, x_j) = \exp\left\{-\left(|x_j - x_i|^2\right)/\sigma^2\right\}, \tag{10}$$

where σ is the RBF kernel parameter.

2.3.2. Principal component analysis (PCA)

PCA is an effective method of feature extraction in pattern recognition. Its purpose is to reduce the dimension of the whole sample space and utilize fewer features of samples to describe the information of the whole sample space as much as possible [19]. Since the image data used for training and testing of SVMs are so large, PCA is introduced for data dimension reduction and feature extraction. Here, the training images are used as the example for description of the PCA used in our research.

The training images of size 48 $\times$ 24 are expanded by column for construction of 1152 $\times$ 1 vectors x_i (i = 1, 2, ..., M), where M is the number of samples (here is 6202). Then the image column data matrix $\mathrm{X} = \left(\mathrm{x}_1, \mathrm{x}_2, \cdots, \mathrm{x}_\mathrm{M}\right)$ is formed, which is the so-called observation vector.

Next, the data are centered in the observation vector. The mean vector $\mu = \mathrm{E}\{\mathrm{x}\} = (1/\mathrm{M})\sum_{\mathrm{i}=1}^{\mathrm{M}}\mathrm{x}_\mathrm{i}$. Every observation vector is subtracted by the mean vector μ, i.e. $\mathrm{x}_\mathrm{i} \leftarrow \left(\mathrm{x}_\mathrm{i} - \mu\right)$; then the centered image column data matrix $\mathrm{X}_\mu = \left(\mathrm{x}_1, \mathrm{x}_2, \cdots, \mathrm{x}_\mathrm{M}\right)$ is achieved.

Based on the centered image vectors $\mathrm{x}_1, \mathrm{x}_2, \cdots, \mathrm{x}_\mathrm{M}$, the covariance matrix of samples can be defined as

$$S_t = \frac{1}{M}\sum_{i=1}^{M}\overline{x_i x_i}^{\mathrm{T}} = \frac{1}{M}\Phi_t\Phi_t^{\mathrm{T}} \tag{11}$$

The orthonormal eigenvectors $\beta_1, \beta_2, \cdots, \beta_M$ and the corresponding eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_M$ are achieved through the calculation of S_t. Since the dimension of $\Phi_t \Phi_t^{\mathrm{T}}$ is too large, the singular value decomposition (SVD) technique is used. The eigenvalues are sorted in decreasing order: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M$. The correct pairings between the columns in the eigenvalues and eigenvectors matrix are maintained.

Lastly, the principal components of sample matrices will be determined in terms of the rate of contribution. Through the calculation described above, the contribution rate of the former 21 eigenvalues ($m = 21$) is above 95% in our paper, which is shown in Figure 6. In other words, the former 21 eigenvectors can describe the information of the whole samples space in this paper. The eye image data of reduction dimension are used for training the SVM classifier.

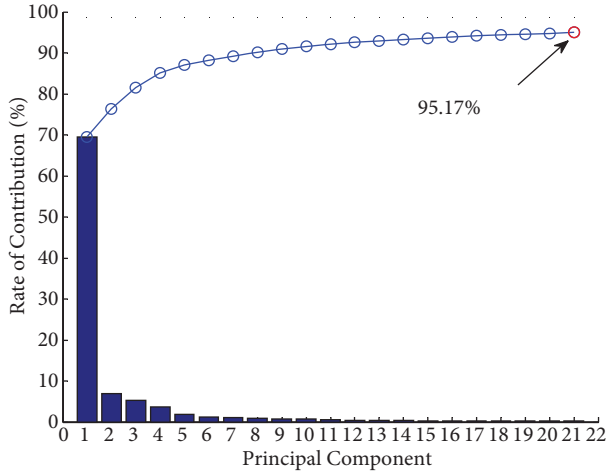

Figure 6. The distribution map of contribution rate against principal component. The blue line is the cumulative rate of contribution.

2.3.3. GA-based SVM parameter optimization

This research adopted RBF as the kernel function of the SVM for classification. The kernel function has two parameters: penalty factor C and RBF kernel parameter δ. Both of them are required to be set accurately, since they have important effects on the learning and predicting performance of the SVM. In general, the best C and δ are initially found and to be fixed first, and then the parameters can be assigned by the form of manual adjustment within certain ranges and cross-comparison is operated by the grid-search algorithm. However, the best initial values are very difficult to be found. With the increase in amount of training images, the parameters should be adjusted accordingly. Consequently, it will face low efficiency and poor improvement in SVM classification performance.

Based on the above reasons, this paper employed a GA for optimizing the SVM parameters automatically, since GAs are useful for solving global optimization problems. GA coding strategies include binary coding and real-valued coding. Here, a real-valued coding is used. At first, C and δ are initially set for the value ranges. Secondly, every chromosome in the population is decoded to obtain the C and δ. A fitness function is used for evaluating the quality of every chromosome through training the SVM classification model with samples according to the correct classification rate. The genetic operation includes selection, crossover, and mutation. The reproduction goes on until the most fitness for goal function or required number of generations (here is 100) is reached. After that, the optimized parameters of C and δ would be calculated and obtained. The detail

GA coding can be found in [20]. The SVM in training processes takes advantage of 5-fold cross-validation to reach maximum classification accuracy for the reliability of parameter estimation.

According to the GA algorithm, C and δ used in our method were obtained, i.e. $C = 17.7142$ and $\delta = 0.021453$, respectively. The fitness curve in Figure 7 shows that the best accuracy is 97.9333% after the 23rd generation. Then the 500 eye images and 500 noneye images extracted from the eye database for construction of validation set, which are different from eye and noneye image training sets of the SVM, are used to test the performance of the SVM with the optimized parameters. The accuracy of testing classification on the validation set was 98.25%. The performance of the trained SVM classifier can meet our requirement of accuracy for eye detection in this paper. The optimized parameters are used in the SVM classifier for eye detection and evaluation in the next section.

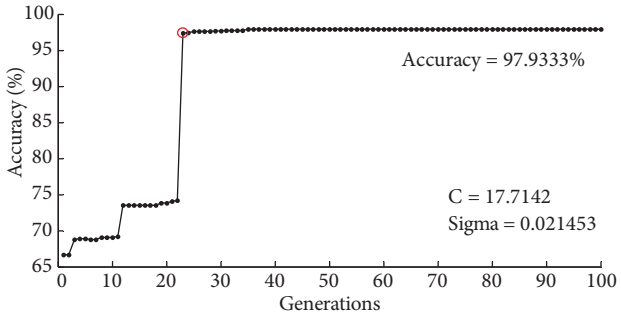

Figure 7. The fitness curve of optimized parameters for the SVM using a GA.

3. Detection of eye regions

The proposed detection process of eye regions consists of six steps. The steps from 1 to 4 are used for determining the eye search region on the face image. Then, according to the constructed VF and trained SVM classifier in Section 2, accurate left and right eye regions can be obtained through step 5 and step 6. The process of detection is shown as follows.

Firstly, eyes can be sought either on the whole image or on the face region, which is localized by the face detector. In this paper, as shown in Figure 8a, the boosted cascade face detector [21] is applied to locate the face region. Thus, eye region searching could be conducted in the detected face region instead of the whole image.

Secondly, the detected face is normalized to an image of size 200 $\times$ 200 pixels as shown in Figure 8b.

Thirdly, to improve calculation efficiency, the eye search region is limited to the top half of the detected face region, since eyes always exist in the top half of the face. According to the geometric structure of the human face, the search region is rebuilt as

$$W_{sr} \approx \frac{1}{2} W_F \times \gamma_W, H_{sr} \approx \frac{1}{2} W_F \times \gamma_H, \tag{12}$$

where W_F is the width and height of the face image (here, width = height), W_{sr} and H_{sr} are the width and height of the search region respectively, and γ_W and γ_H are the adjustment coefficient factors of the width and height of the search region, respectively. The search region of size $W_{sr} \times H_{sr}$ pixels on the human face is shown in Figure 8c.

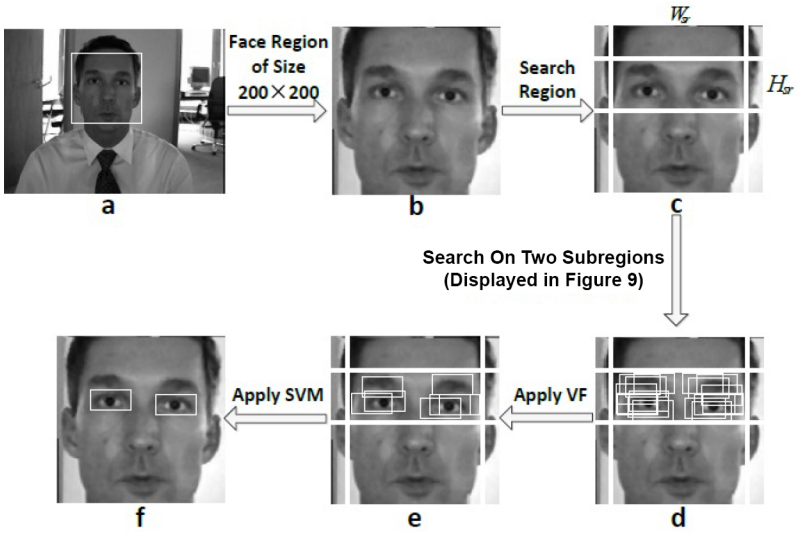

Figure 8. Flowchart of the eye detection. (Example of face image is extracted from BioID face database.)

Fourthly, the cropped search region from a human face image (Figure 9a) is displayed in Figure 9b. The search region is divided into two subregions for searching for the left eye and right eye separately. Then the overlapped windows of size 16 × 8 subblocks of size 3 × 3 pixels with interval 3 pixels are built by Eq. (2). The variance images of the two subregions are shown in Figure 9c. The searching processes of the left and right eyes operate on the two variance image regions.

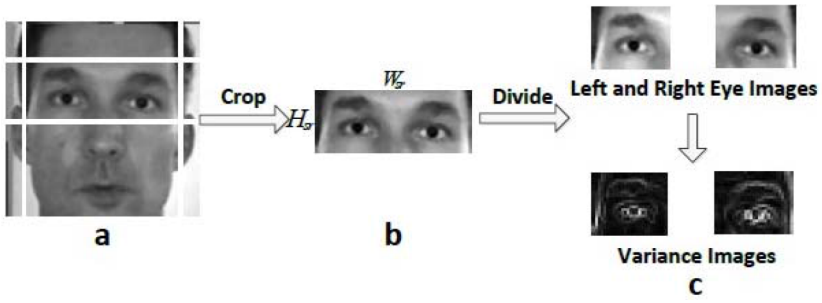

Figure 9. Eye regions search on two subregions separately. (Example of face image is extracted from BioID face database.)

Fifthly, to find the candidate eye regions, the constructed eye VF is adopted in Section 2.2. The eye VF is used as a search windows shifted with interval 3 pixels (1 subblock of size 3 × 3 pixels) on variance images of two subregions. Some examples of search windows shift are drawn on the human face image shown in Figure 8d. Then the variance vector of size 16 × 8 subblocks with the same size region as the eye VF is extracted from the variance images of the two subregions. The response value between the extracted vector block and the eye VF can be calculated using Eq. (4). According to the threshold value of the constructed eye VF, i.e. 0.3, the regions with higher correlation values (>0.3) are counted in the set of candidate eye regions. Some examples of candidate eye regions are shown in Figure 8e.

Lastly, from Figure 8e, we can see that most of noneye images in the search region have been discarded by the eye VF. Then the trained SVM classifier (see Section 2.3) is employed to select only two regions, which

are the most possible regions for the left and right eyes from the candidate eye images. The accurate left and right eye regions are shown in Figure 8f.

4. Experimental results

4.1. Evaluation of eye region detection

The proposed method was tested separately on the IMM, BioID, and FERET face databases. The IMM face database is a collection of digital color images of males and females with different illumination conditions and various facial expressions. It contains 240 images of 40 persons (six images per person) of size 480 $\times$ 640. The BioID face database contains 1521 grayscale, frontal facial images of size 384 $\times$ 286, acquired under various illumination conditions with a complex background. The database also contains titled and rotated faces, people that wear glasses, and people that have their eyes shut. The FERET face database contains 1400 images of 200 persons (seven images per person) with various color skins, rotated faces, and different lighting conditions. The size of the face images is 80 $\times$ 80. In this case, the situation of images from the three face databases is almost the same as real-world's situation, including different backgrounds, gaze directions, and illumination conditions.

The eye detection method proposed in this paper is run on the correct face regions captured by the boosted cascade face detector. The face detector was applied to the BioID and IMM face databases, not to the FERET face database due to only providing face regions. The IMM and FERET face databases have no glasses. The BioID face database consists of 1039 images without glasses and 482 images with glasses. In this paper, 406 high-quality images with glasses were extracted from the BioID face database as the testing set, namely the eye regions with glasses have no significant white spots and uniform illumination. The face detection results on the IMM and BioID face databases are given in Table 1.

Table 1. Correct images of face detection based on different databases.

Face database	Numbers of correct images
240 images from IMM	226
406 images with glasses from BioID	398
1039 images without glasses from BioID	1025

At present, there is still no criterion to evaluate the accuracy of eye region detection. Consequently, the valid correct test result considered in this paper is that the upper and lower eyelid and two corners of the eye fall into the eye regions. The detection rate for each database and average detection rate across the three databases are listed in Table 2.

Table 2. Detection rates of eye based on three face databases.

Images	Detection rate (%)
226 images from IMM	98.7%
1025 images without glasses from BioID	97.8%
398 images with glasses from BioID	92.5%
1400 images from FERET	97.6%
Total images (above)	97.5% (average detection rate)

Because the image resolution of the IMM face database is very high (640 $\times$ 480) and the illumination is uniform, the highest detection rate is expected. Without loss of generality, faces turned a certain angle left/right were also tested and a good detection result was achieved as shown in Figure 10a.

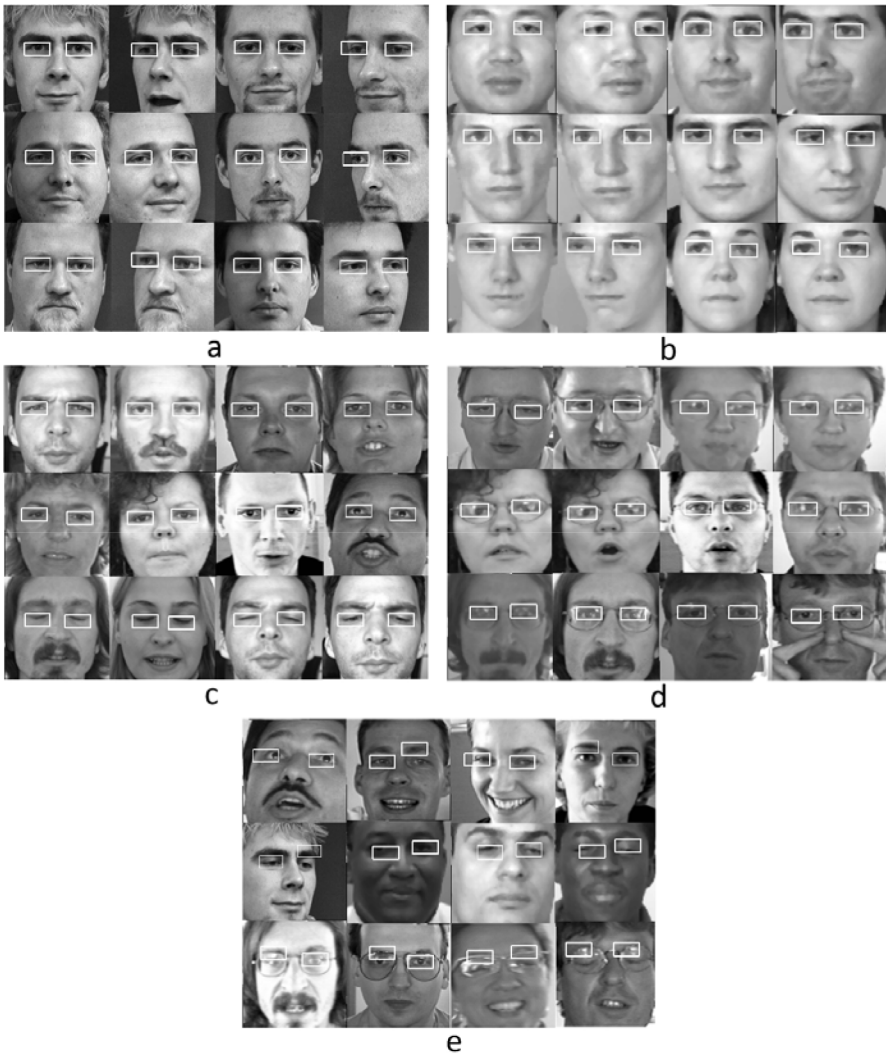

Figure 10. Results of eye detection on IMM, FERET, and BioID without glasses and with glasses face databases, respectively, and results of failure eye detection on three face databases.

The lowest detection rate was obtained on the FERET face database due to the relatively poor image quality. However, with regard to the turned face images, a good detection result was also achieved as shown in Figure 10b.

The quality of the images without glasses from the BioID face database is better than that from FERET, and so the detection result on BioID was moderate among the results on the three face databases. The proposed method was also applied to detect the closed eye region in the face image and had a good result as shown in Figure 10c.

The detection rate of images with glasses was 92.5% on average. There were two reasons for this drop in the detection rate. Firstly, large white spots emerge on the glasses. Secondly, the number of the training images with glasses of SVM is lower. Some successful examples with glasses are shown in Figure 10d.

The proposed detection method does not always achieve higher detection rates. Some examples of failure eye detection are shown in Figure 10e. The possible reasons for the failure detections including direction of large head pose, large white spot on the glasses, and low quality of face images and eye images. Some image

processing techniques, such as an image filter, can be adopted to reduce or eliminate the undesired effect of white spots in the eye region. However, other situations among those face images cannot be processed well.

Table 3 lists the comparison results between the proposed method and the existing methods. It shows that the proposed method achieves a higher detection rate than other methods excluding the method proposed by Hassaballah et al. [12] tested on the BioID face databases. However, in our paper, the statistical eye detection rate operated on the achieved correct face images, because a 100% face detection rate is hard to achieve using the boosted cascade face detector [21] on the BioID database. The proposed method is based on a baseline VF, which is proposed by Feng and Yuen. However, Feng and Yuen used a baseline VF for eye detection and achieved a detection rate of 92.5%, whereas the proposed method obtained an average detection rate of 97.5% on the three face databases.

Table 3. Comparison of the proposed method with existing methods.

Eye detection method	Face database	Detection rate
Proposed method in this paper	226 images from IMM	98.7%
	1025 images without glasses and 398 images with glasses from BioID	96.3%
	1400 images from FERET	97.6%
Hassaballah et al. [12]	XM2VTS	98.4%
	BioID	97.1%
	1500 images from FERET	97.3%
Kalbkhani et al. [14]	4517 images from color FERET	96.98%
	Aberdeen	94.29%
	IMM	98.65%
	CVL	95.32%
Guan [4]	CVL	96.2%
Laxmi and Rao. [13]	GTAV and VITS	96%
Campadelli et al. [11]	750 images of XM2VTS	95.7%
Peng et al. [9]	ORL (without glasses)	95.2%
Zhou and Geng [3]	BioID	94.81%
Xia and Su [10]	ORL	94.7%
Wang and Ji [7]	5600 frontal images of FRGC V1.0 database	94.5%
Phromsuthirak et al. [15]	473 frontal facial images from PICS	93.2%
	14 rotational facial images from PICS	85.7%
Feng and Yuen [2]	MIT AI laboratory face database	92.5%
Wu and Trivedi [8]	317 images from FERET	92.43%
Ryu and Oh [6]	180 images without glasses from ORL	91.7% left eye
		85.5% right eye
Ando et al. [5]	1227 images from color FERET	88%

4.2. Robustness to illumination and pose changes

In order to evaluate the robustness of the proposed eye detection method under the illumination changes, the Yale Face Database B were used. The face database contains 5760 grayscale images of 10 subjects under 576 viewing conditions with 9 poses and 64 illumination conditions. The size of each image is 640 $\times$ 480 pixels. The proposed eye detection method is only used on the correct face region, which depended on the performance of the boosted cascade face detector. Through the experiments, the face detection results failed when the light source direction had a high angle since the performance of the face detector was influenced by shadows. The test

results on face images from the Yale Face Database B gained 3175 correct face images by the boosted cascade face detector.

The first two rows of Figure 11 show qualitative samples of the eye detection results obtained from two subjects in the subsets of Yale Face Database B. The results show that face images under light source directions varying from different angles ($\leq \pm 40^\circ$ azimuth and elevation) with respect to the camera axis can be dealt with well. The last row in Figure 11 shows the eye detector results obtained from a subject in the Yale Face Database B. The quantitative evaluation shows the robustness of the eye detection method to pose changes. The correct rate of eye detection was 92.3% on the 3175 face images. It is possible to conclude that the eye detector can achieve a very high correct rate when there are no shadows on the eye regions.

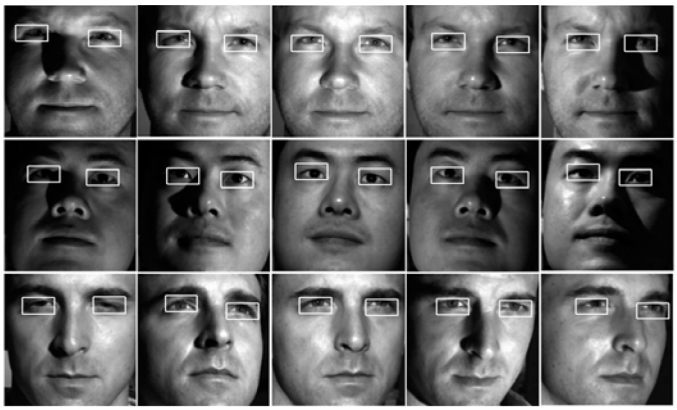

Figure 11. Results of eye detection on Yale Face Database B.

The effects of pose changes on the proposed eye detector were evaluated on the FEI Face Database, which is a Brazilian face database and contains images of 200 subjects captured under an upright frontal position with profile rotation of up to about 180 degrees for a total 2800 images (14 images of each subject). The scale of rotation angle might vary about 10% and the size of each image is 640 $\times$ 480 pixels. The database shows very challenging conditions for the proposed eye detection method, as many faces of subjects toward $\pm 45^\circ$ horizontal direction above. In that case, although the face images were obtained by the face detector, the eye detection failed due to only one whole eye being available on the face image of the subject. Consequently, the performance of the proposed eye detection method was only evaluated on face rotation angles within the range of $\pm 30^\circ$ horizontal direction. The 1000 face images with five images of each subject were captured from the FEI Face Database. The face detector obtained a 100% correct rate of face detection from the 1000 face images.

Figure 12 shows qualitative samples of eye detection obtained from three subjects on the FEI Face Database. The quantitative evaluation shows the robustness of the proposed eye detection method to pose changes. In order to analyze the correct rate of eye detection in the 1000 face images from the FEI Face Database, the experiments were conducted under the five face directions of 0°, $\pm 10^\circ$ and $\pm 30^\circ$ in the horizontal direction, respectively. The test results are shown in Table 4. The best accuracy is achieved when the subjects face toward front due to parts of the high resolution eye region image. This case is similar to the accuracy achieved when the face of subjects turns toward $\pm 10^\circ$ left or right horizontally. It is noted that when the facial rotation angle is up to $\pm 30^\circ$, the accuracy is lower than in other cases, in which part of the eye might be sheltered by the turned face region.

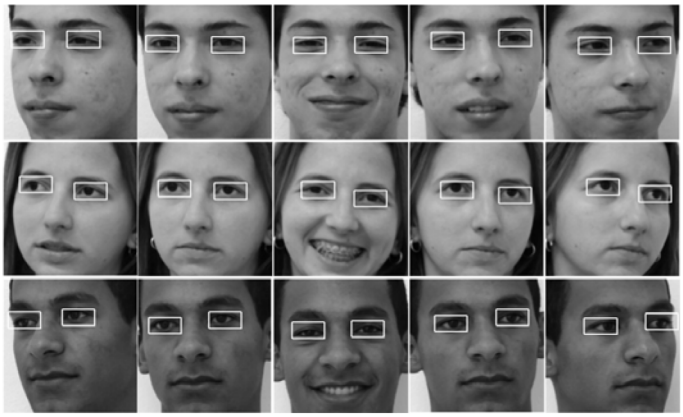

Figure 12. Results of eye detection on FEI Face Database.

Table 4. Detection rates of eye for different facial rotation angles.

Pose	Accuracy (%)
–30°	92.5%
–10°	96%
0°	99.5%
10°	97.5%
30°	90.5%
Average	95.2%

4.3. Robustness to eye occlusions

It is very important to evaluate the robustness of the proposed eye detection method to eye occlusions. Many people in the BioID face database show closed or semiclosed eyes, as shown in the last row of Figure 10c. From the accuracy listed in Table 2, it seems that the proposed method is able to handle eye occlusions. To validate the proposed method's robustness to occlusion, an experiment was designed. The three subjects were requested to sit in front of a camera and gaze at the camera. The subjects come from the author's laboratory. Then the eye was closed slowly with five conditions from open to closed. The images with the five situations of subjects' eyes were recorded. From Figure 13, eye detection images of the five conditions from open to closed are

Figure 13. Results of eye detection on subjects' faces from author's laboratory.

performed well. The evaluation results of the three subjects show the robustness of the eye detection method to eye occlusions.

4.4. Parameter sensitivity analysis of the SVM model

The eye detection accuracy of the SVM classifier mainly relies on two parameters of the SVM model, C and δ. To assess the impact of the parameters on eye detection accuracy, a training data set with 2000 eye and 2000 noneye images and testing data with 1000 eye and 1000 noneye images were taken as the analysis data set for sensitivity analysis of parameters with eye detection through learning and testing experiments. The analysis data set is a different data set to that in Section 2.2.3. Firstly, the δ was fixed at 0.021453 achieved by GA under the optimal condition, and C was given a range from 1 to 100 with interval 0.1. Figure 14 shows that the detection accuracy is stable as C ranged from 49 to 68 and from 74 to 97, respectively. The highest accuracy is obtained when C ranged from 6 to 12 and from 17 to 18, respectively. Notably, the learning and testing accuracy rates were optimal when C is rated at 17.7; the value is similar to the $C = 17.7142$ achieved in Section 2.2.3. The analysis of parameter δ sensitivity to eye detection was conducted while C was fixed at the value of 17.7 and δ was given a range from 0.001 to 0.2 with interval 0.001. Figure 15 shows that the test accuracy decreases as δ varies from 0.021 to 0.2. The best accuracy for the learning and testing is obtained when δ is rated at 0.021. Hence, the evaluation results verify the robustness of parameters $C = 17.7142$ and $\delta = 0.021453$ achieved by the GA in Section 2.2.3 as the optimal parameters for a high eye detection rate.

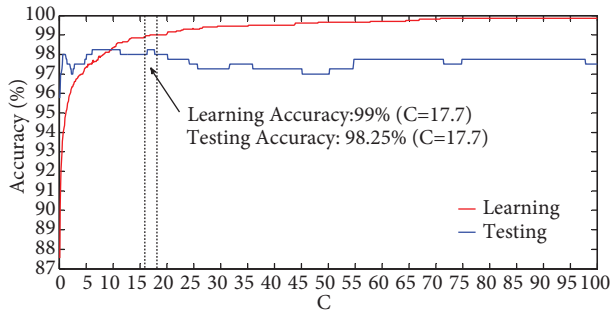

Figure 14. Parameter C impacted on sensitivity of the SVM model.

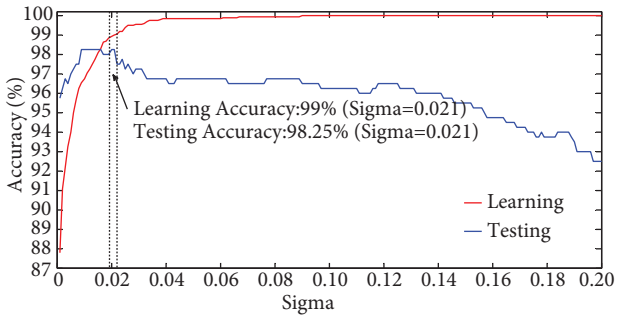

Figure 15. Parameter δ impacted on sensitivity of the SVM model.

5. Conclusions

Eye detection has been widely applied in many fields. This paper presents an efficient eye detection method based on gray intensity variance and SVMs. Firstly, the variation of gray intensity in eye regions is more obvious than that in other regions on the face. Hence, it becomes evidence to construct a gray intensity VF. The VF is used to eliminate most of the false eye regions (noneye regions) in order to leave candidate eye regions. Secondly, a trained SVM classifier is utilized to determine the precise eye location among the candidate eye regions. In order to reduce the runtime and improve the classification accuracy of SVM, PCA is used for data dimension reduction and feature extraction, and SVM parameters are optimized by a GA. The method was tested on three face databases. The experimental results show that the proposed method is sufficiently robust and accurate for a wide variety of face images. Moreover, a 92.5% detection rate was achieved based on the BioID face images with glasses. An extensive evaluation of the eye detection method was performed to verify the sensitivity of the parameters of the SVM model with eye detection and tested its robustness to different eyes conditions (e.g., semiclosed eyes, closed eyes, and squint), rotation of frontal face, and illumination changes. Some problems

remain in this method, such as the detection rate dropping when large rotation angle of face, the poor quality of face images with glasses, and insufficient diversity of training images. The comparison with the state-of-the-art methods shows the proposed method has a high accuracy and can be applied to low-resolution images of eyes on the FERET face database. From the reported accuracy of the system here, it is thought that the proposed eye detection provides a potential solution to practical applications. Based on a PC with 2.26 GHz and 1 GB RAM, the average runtime for detecting an eye region with MATLAB was about 45 ms. The method proposed in this paper is very promising in determining eye position within eye gaze tracking systems in future work.

References

[1] Hansen DW, Ji Q. In the eye of the beholder: a survey of models for eyes and gaze. IEEE T Pattern Anal 2010; 32: 478-500.

[2] Feng GC, Yuen PC. Multi-cues eye detection on gray intensity image. Pattern Recogn 2001; 34: 1033-1046.

[3] Zhou ZH, Geng X. Projection functions for eye detection. Pattern Recogn 2004; 5: 1049-1056.

[4] Guan Y. Robust eye detection from facial image based on multi-cue facial information. In: IEEE 2007 Control and Automation Conference; 30 May-1 June 2007; Guangzhou, China. New York, NY, USA: IEEE, pp. 1775-1778.

[5] Ando T, Semicond OKI, Moshnyaga VG. A low complexity algorithm for eye detection and tracking in energy-constrained applications. In: IEEE 2013 Communications, Signal Processing, and their Applications Conference; 12-14 February 2013; Sharjah, UAE. New York, NY, USA: IEEE, pp. 1-4.

[6] Ryu YS, Oh SY. Automatic extraction of eye and mouth fields from a face image using eigenfeatures and ensemble networks. Appl Intell 2002; 17: 171-185.

[7] Wang P, Ji Q. Multi-view face and eye detection using discriminant features. Computer and Vision and Image Understanding 2007; 105: 99-111.

[8] Wu J, Trivedi MM. A binary tree for probability learning in eye detection. In: IEEE 2005 Computer Vision and Pattern Recognition Conference; 25-26 June 2005; San Diego, CA, USA. New York, NY, USA: IEEE, pp. 164-171.

[9] Peng K, Chen LM, Ruan S, Kukharev G. A robust algorithm for eye detection on grey intensity face without spectacles. J Comput Sci Tech (JCS&T) 2005; 5: 127-132.

[10] Liu X, Dong YQ, Li S. Rapid human-eye detection based on an integrated method. In: IEEE 2010 Communications and Mobile Computing Conference; 12-14 April 2010; Shenzhen, China. New York, NY, USA: IEEE, pp. 2-8.

[11] Campadelli P, Lanzarotti R. Fiducial point localization in color images of face foregrounds. Image Vis Comput 2004; 22: 863-872.

[12] Hassaballah M, Kanazawa T, Ido S, Ido S. Efficient eye detection method based on grey intensity variance and independent components analysis. IET Comput Vis 2010; 4: 261-271.

[13] Laxmi V, Rao PS. Eye detection using Gabor Filter and SVM. In: IEEE 2012 Intelligent Systems Design and Applications (ISDA) Conference; 27-29 November 2012; Kochi, Japan. New York, NY, USA: IEEE, pp. 880-883.

[14] Kalbkhani H, Shayesteh MG, Moussvi SM. Efficient algorithms for detection of face, eye, and eye state. IET Comput Vis 2013; 7: 184-200.

[15] Phromsuthirak C, Umchid S. Development of a geometrical algorithm for eye detection in color images. In: IEEE 2012 Biomedical Engineering Conference; 5-7 December 2012; Ubon Ratchanthani, Thailand. New York, NY, USA: IEEE, pp. 5-7.

[16] Vapnik VN. The Nature of Statistical Theory. New York, NY, USA: Springer-Verlag, 1995.

[17] Mahesh P, Giles MF. Feature Selection for Classification of Hyperspectral Data by SVM. IEEE T Geosci Remote Sens 2010; 48: 2297-2308.

[18] Boser B, Guyon IM, Vaphik VN. A Training Algorithm for Optimal Margin Classifiers. In: ACM 1992 Computational learning theory annual workshop; 27-29 July 1992; Pittsburgh, PA, USA. New York, NY, USA: ACM. pp. 144-152.

[19] Jian Y, Zhang D, Yang JY. Constructing PCA baseline algorithms to reevaluate ICA-based face-recognition performance. IEEE T Sys Man Cybern B, Cybern. 2007; 37: 1015-1021.

[20] Haupt RL, Haupt SE. Practical Genetic Algorithms. New York, NY, USA: John Wiley and Sons, 1998.

[21] Viola P, Jones MJ. Robust real-time face detection. Int J Comput Vis 2004; 57: 137-154.