Pattern Recognition 67 (2017) 23–31

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

A joint cascaded framework for simultaneous eye detection and eye state estimation

Chao Gou[a,c,d,*], Yue Wu[b], Kang Wang[b], Kunfeng Wang[a], Fei-Yue Wang[a,c], Qiang Ji[b]

[a] Institute of Automation, Chinese Academy of Sciences, Beijing 100190, PR China
[b] Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
[c] Qingdao Academy of Intelligent Industries, Qingdao 266109, PR China
[d] University of Chinese Academy of Sciences, Beijing 100049, PR China

ARTICLE INFO

Article history:
Received 15 September 2016
Revised 3 December 2016
Accepted 15 January 2017
Available online 1 February 2017

Keywords:
Eye detection
Eye state estimation
Learning-by-synthesis
Cascade regression framework

ABSTRACT

Eye detection and eye state (close/open) estimation are important for a wide range of applications, including iris recognition, visual interaction and driver fatigue detection. Current work typically performs eye detection first, followed by eye state estimation by a separate classifier. Such an approach fails to capture the interactions between eye location and its state. In this paper, we propose a method for simultaneous eye detection and eye state estimation. Based on a cascade regression framework, our method iteratively estimates the location of the eye and the probability of the eye being occluded by eyelid. At each iteration of cascaded regression, image features from the eye center as well as contextual image features from eyelid and eye corners are jointly used to estimate the eye position and openness probability. Using the eye openness probability, the most likely eye state can be estimated. Since it requires large number of facial images with labeled eye related landmarks, we propose to combine the real and synthetic images for training. It further improves the performance by utilizing this learning-by-synthesis method. Evaluations of our method on benchmark databases such as BioID and Gi4E database as well as on real world driving videos demonstrate its superior performance comparing to state-of-the-art methods for both eye detection and eye state estimation.

1. Introduction

Eye detection aims to estimate the pupil location in a image. Eye state prediction aims to estimate the binary state(open/close) of eye. Eye detection is becoming an increasingly important research topic due to its various applications such as iris recognition, eye gaze estimation and human-robot interaction. Eye state estimation is critical to detect the individual's affective state, and the corresponding pupil location is essential to reflect the individual's focus attention. Eye state estimation also has extensive applications in real world including diagnosing neurological disorders, sleep studies and driver drowsiness detection.

Although much work has been done for eye detection and eye state estimation, they still are challenging tasks due to variations in appearance, illumination and occlusion. In addition, most of the existing works only perform eye detection and eye state estimation separately and independently. In this paper, we propose a method for simultaneous eye localization and eye state estimation, on the basis of a joint cascaded regression framework. In cascaded regression framework, eye states and eye locations are updated simultaneously. Since it is time-consuming to collect large number of eye images with accurate eye related landmark labels for training, we propose to learn from the combination of synthetic and real images. Our main contributions are highlighted as follows:

- **Simultaneity:** On the basis of the cascade regression framework, the eye openness and eye locations are updated in each iteration simultaneously. Different from the conventional sequential eye detection and eye state estimation methods, our method is the first work that performs eye detection and eye state prediction at the same time.
- **Robustness:** The proposed framework relaxes the binary eye state to be a continuous probability, which measures the degree of openness of eyes. By setting flexible threshold, the eye states can be robustly predicted. In addition, it can estimate the location of eyes even when the eyes are closed.
- **Learning-by-synthesis:** For learning-based methods, it is time-consuming to collect various eye images and annotate them with ground truth. We propose to learn the regression mod-

* Corresponding author.
E-mail address: gouchao.cas@gmail.com (C. Gou).

http://dx.doi.org/10.1016/j.patcog.2017.01.023
0031-3203/

els from generated synthetic photorealistic eye images and it improves the result.
- **Effectiveness and efficiency:** By exploiting the cascade regression and the interactions between eye location and eye state, our method performs significantly better than other state-of-the-art methods and can achieve nearly real time.

The remainder of this paper is arranged as follows. Section 2 reviews the related work on eye detection and eye state estimation (open/close). The proposed method is described in Section 3. Experimental results are discussed in Section 4. The conclusion is drawn in Section 5.

2. Related work

2.1. Eye localization

Eye detection has been studied for decades and numerous methods have been proposed. In this section, we focus on reviewing most of the recent works. A detailed review of earlier techniques devoted to this topic can be found in [1,2]. Generally, On the basis of captured information, we summarize the representative eye localization methods into five categories: (i) shape-based, (ii) appearance-based, (iii) context-based, (iv) structure-based, and (v) others.

Shape-based models generally capture the geometric information of an iris. Yuille et al. [3] build a parameterized deformable model formulated by geometric shape with 11 parameters. Their model considers peaks, edges and valleys by using energy functions. To fit the model to a testing image, it has to optimize in a large continuous parameter space which covers shape variations. Based on the elliptical shape of an iris, Hansen and Pece [4] propose a likelihood model which incorporates neighboring information for iris detection and tracking. By using EM and RANSAC methods, the ellipse is locally fitted to the image. In [5], the authors use curvature of isophotes in the intensity image to design a voting-based method for pupil localization.

Appearance-based methods are based on the photometric appearance, which is characterized by filter responses and color distribution. In [6], the authors propose a method for eye localization based on an ensemble of randomized regression trees, which are trained by using the pixel intensity differences around pupils. Araujo et al. [7] describe an *Inner Product Detector* for eye localization based on the correlation filters. Zhang et al. [8] use local linear SVM for eye center detection, and ASEF-based filters are applied to select the candidate centers. Wu and Ji [9] propose to learn deep features to capture the appearance variations of eyes in uncontrolled conditions. In [10], the authors apply a discriminative feature extraction method to 2D Haar wavelet transformation [11] and use an efficient SVM for fast classification. Support Vector Regressor (SVR) is used to estimate the distance of patch center to the pupil center by extracted HoG features in [12].

Since eye centers have a stable relationship with other facial parts in terms of both appearance and shape, it is important to capture the contextual information to detect the eyes. Yang et al. [13] propose to detect the pupil by using different Gabor kernels to convolute with the image, which highlights the eye-and-brow regions. By employing a coarse to fine strategy for robust initialization, Zhou et al. [14] propose multi-scale nonlinear feature mapping based on the Supervised Decent Method (SDM)[15] for eye detection. They use 14 eye related key points to capture the contextual information.

The structural locations information related to nose, mouth, *etc.* is helpful for the eye localization. Pictorial Structure [16] and enhanced Pictorial Structure [17] provide a powerful framework to model the face in terms of its appearance and geometrical relationship between parts. The pupil is part of face and this model allows for accurate eye detection by capturing the structural information.

It is a challenge to organize some other methods into a specific aforementioned types. Based on the intensity information, Chen and Liu [18] propose to extract eye regions by image enhancement, Gabor transformation, cluster analysis, and neighborhood operation with similarity measures in eye regions for final eye detection. Some researchers also propose combined models to overcome the shortcomings of separate model. Timm and Barth [19] propose to use image gradients and squared dot products to detect the pupils. The aforementioned model in [17] also combines shape and appearance features in a unified framework.

2.2. Eye state estimation

After accurate eye detection, eye state estimation can be achieved. Since pupil is frequently occluded by eyelids, hair and sunglasses, it is crucial to recognize the eye states to decide whether the pupil is occluded. Although the eye states estimation has received increasing research attention, eye state estimation is still an unsolved problem in uncontrolled scenes. Plenty of eye state estimation methods have been proposed. Generally, these methods are classified into three categories: (i) shape based, (ii) template based, and (iii) learning based.

Shape-based approaches aim to recognize the eye states based on geometric relationships or circular shape of visible iris. Kurylyak et al. [20] set some thresholds for the difference between video frames in eye region pixel level to detect the eyelid movement. Then differences of vertical and horizontal projections are used to detect the degree of eye openness. Another simple and direct way for eye state estimation is template matching. Feng et al. [21] use template matching for coarse driver eye state estimation followed by capturing the upper eyelids curvature for fine recognition. By setting flexible thresholds for the combined model, it can achieve a reasonable performance on warning system for a driver. Gonzalez et al. [22] use projection operation to produce three templates: open, nearly closed and closed eyes. Both pair of eye state classifier and individual eye state classifier measure similarities between the templates and the test image.

Since eye state estimation is a binary classification problem, machine learning techniques are widely used to tackle this problem and they significantly improve the performance compared with the aforementioned methods. Song et al. [23] propose *MultiHPOG* features to recognize the eye states and made a comparison using other features in different datasets. Extensive experimental results show that *MultiHPOG* are effective and robust. The authors also released the Closed eyes in the Wild (CEW) database which contains 2423 images with open and closed eyes. In [24], minimum intensity projection is adapted. After histogram equalization, the pixels in vertical and horizontal direction with minimum intensity value are chosen to combine the feature vector. Through their experiment, a Random Forest classifier performs better than other tree based classifiers. Another projection operation based work is presented in [25]. The authors introduce a discriminative feature by projecting the gray value distribution in x and y direction. In addition, a brightness adjustment based on the mean value of color image is proposed to overcome the variation of illumination.

2.3. Learning-by-synthesis

Learning based cascaded regression framework requires large scale training data. In our case, there is limited public available dataset with eye center location labeled under various illuminations and head pose. In addition, it is time-consuming and also

can be unreliable for accurate manual annotation of eye related landmarks. Motivated by recent work of learning-by-synthesis for appearance-based eye gaze estimation and eye detection [26–29], we propose to learn from the synthetic eye images for accurate eye detection.

In summary, we realize that utilizing multi types of features is important on eye detection. In addition, capturing appearance features and learning classifiers significantly improve eye state estimation performance. Hence, we propose to capture the shape, appearance, structural and contextual information for eye detection and eye state estimation. Moreover, we learn from the combination of real and synthetic images to boost the performance.

3. Proposed method

The conventional framework for eye state estimation is to locate the eyes first and then perform the binary classification. Since pupils are very likely to be occluded, general methods can not deal with this problem. To the best of our knowledge, there is little research focusing on eye localization and eye state estimation at the same time. In this paper, motivated by our previous work for facial landmark detection [30] which uses a robust cascaded regression method for facial landmark detection under occlusions and large head poses, we propose to simultaneously detect eyes and recognize eye state (open/close) using a joint cascaded regression framework, on the basis of eye-related shape, appearance, structural and contextual information.

Before we introduce our proposed method, we first review general cascaded regression framework which has been successfully applied to facial landmark detection [15,31,32]. The overall algorithm is shown in Algorithm 1. The facial landmark coordinates are denoted as $\mathbf{x}^t = x_1^t, x_2^t, \ldots, x_D^t$, where D denotes the number of landmarks and t denotes the iteration in cascaded regression framework. It iteratively predict the location updates $\Delta\mathbf{x}^t$ based on the extracted appearance features with regression model g_t and then adds the current estimated updates $\Delta\mathbf{x}^t$ to the previous locations $\mathbf{x}^{t-1}$ to acquire new landmark locations $\mathbf{x}^t$. It repeats until convergence.

The coarse-to-fine joint cascaded framework of our proposed method for simultaneous eye detection and eye state estimation is summarized in Fig. 1 and Algorithm 2. It should be noted that we perform eye detection and eye state estimation for left eye and right eye separately during training and testing. In the following description, we take left eye as example. To capture eye-related shape, structural, appearance and contextual information, we consider five eye-related key points (see green and red points in Fig. 1(b)), consisting of two eye corners, two eyelid points and one pupil for cascade regression. Since we focus on eye state estimation, we introduce openness probability $\mathbf{p} \in [0, 1]$ related to the landmark of eye center. Before doing cascade regression, the eye state is initialized as open with openness probability 1 as shown in Fig. 1(b). In addition, to capture the structural information for eye center detection, all eye related five key point locations $\mathbf{x} \in \Re^{2\cdot5}$ are initialized by the detected eye region and mean location from training. In cascade regression, the eye openness probability and key point locations are iteratively updated at each iteration. For updating the openness probability, a linear regression model f_t is used to predict the eye openness probability update $\Delta\mathbf{p}^t$ on the basis of current key point locations $\mathbf{x}^{t-1}$. For updating the key point locations, another regression model g_t is used to predict the key point location updates $\Delta\mathbf{x}_t$ based on the current locations $\mathbf{x}^{t-1}$ and estimated eye openness probability $\mathbf{p}^t$. In the following, we discuss the major components of the proposed method in the cascade regression framework including initialization, updating eye openness probability and key point locations.

Algorithm 1 General cascaded regression framework.

Input:
Give the image $\mathbf{I}$. Facial landmark locations $\mathbf{x}^0$ are initialized by mean face.
Do cascade regression:
for t=1,2,... ,T **do**
Update the key point locations $\mathbf{x}^t$ given the current key point locations $\mathbf{x}^{t-1}$
$g_t : \mathbf{I}, \mathbf{x}^{t-1} \rightarrow \Delta\mathbf{x}^t$
$\mathbf{x}^t = \mathbf{x}^{t-1} + \Delta\mathbf{x}^t$
end for
Output:
Landmark locations $\mathbf{x}^T$.

Algorithm 2 Joint cascaded regression framework for eye detection and eye openness estimation.

Input:
Give the image $\mathbf{I}$. Left/right eye openness probability is initialized as open by $\mathbf{p}^0 = 1$. Five key point locations $\mathbf{x}^0$ are initialized by the coarse detected eye region and mean eye locations in normalized eye region.
Do cascade regression:
for t=1,2,... ,T **do**
Update the eye openness probability given the current key point locations $\mathbf{x}^{t-1}$.
$f_t : \mathbf{I}, \mathbf{x}^{t-1} \rightarrow \Delta\mathbf{p}^t$
$\mathbf{p}^t = \mathbf{p}^{t-1} + \Delta\mathbf{p}^t$
Update the key point locations given the current key point locations $\mathbf{x}^{t-1}$ and the calculated eye openness $\mathbf{p}^t$.
$g_t : \mathbf{I}, \mathbf{x}^{t-1}, \mathbf{p}^t \rightarrow \Delta\mathbf{x}^t$
$\mathbf{x}^t = \mathbf{x}^{t-1} + \Delta\mathbf{x}^t$
end for
Output:
Estimated eye states based on the eye openness probability $\mathbf{p}^T$ and the locations $\mathbf{x}^T$ of key points.

3.1. Initializaiton

In general regression based landmark detection framework, it makes sense to initialize the landmarks by mean face. In our case, directly initializing the eye related points by mean face is not adequate because it is not likely to converge to the global optimum if the initialization is far away from the ground truth. Hence, it is reasonable to focus on eye regions for accurate eye center localization and eye state estimation. We firstly extract eye regions based on 51 landmarks detection using method in [31]. In this paper, we normalize the width of two eye outer corners to 25 in pixel. As shown in Fig. 1(b), 5 key point locations are initialized by mean locations in eye region from training. The eye state is initialized as open with openness probability $\mathbf{p}^0 = 1$.

3.2. Update probability of eye openness

Even though eye state estimation is a binary classification problem, there are large variations of appearance for different individuals, especially for individual with glasses or nearly half closed eyes. Extracting features from whole eye regions is limited to represent the closed and open eyes. To use more class-specific information for robust eye state estimation, we propose to relax binary eye state to be a continuous eye openness probability, which can be inferred from pupil appearance features and the related contextual information. Since it is not easy to accurately locate the pupil, we

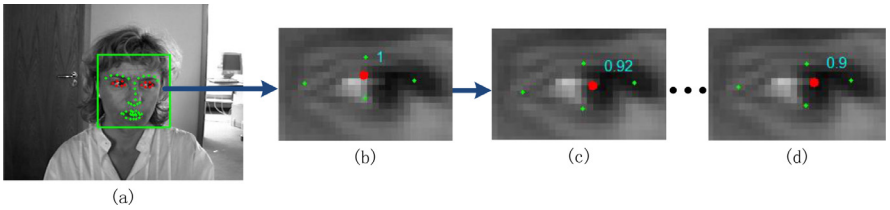

Fig. 1. Framework of our proposed approach for simultaneous eye detection and eye state estimation. (a) Face detection and 51 landmark detection. (b) Extracte the eye regions based on the detected 51 landmarks and initilaize eye state and 5 key point locations. (c) Output of first iterations. (d) Final estimated location and openness probability of eye. *Take the left eye as a example.

use the cascade regression framework and update the eye openness probability iteratively.

The eye openness probability $\mathbf{p} \in [0, 1]$ is updated at each iteration. To capture the pupil appearance and its related contextual information, SIFT features of local patches around the pupils, eye corners and eyelids are used. To capture shape information, the differences for pairwise points are calculated as shape features. Then the shape and appearance features are combined to generate a concatenated feature vector denoted as $\Psi(I, \mathbf{x}^{t-1}) \in \Re^{5\cdot128+5\cdot4}$, where I and t denote the input image and iteration index, respectively.

3.2.1. Learn the eye state prediction model

To estimate the eye openness probability, we use a linear regression model. For training at each iteration, linear model parameters β^t and bias $\mathbf{b}^t$ are estimated by the standard least-square formulation with closed form solution:

$$\beta^{t^*}, \mathbf{b}^{t^*} = arg\min_{\beta^t, \mathbf{b}^t} \sum_{i=1}^{K} \| \Delta\mathbf{p}_i^t - \beta^t \Psi(I_i, \mathbf{x}_i^{t-1}) - \mathbf{b}_i^t \|^2 \quad (1)$$

where K is the number of training samples. Given the training images with estimated key point locations $\mathbf{x}^{t-1}$, the feature $\Psi(I_i, \mathbf{x}_i^{t-1})$ of ith image can be calculated. The probability update $\Delta\mathbf{p}_i^t$ can be acquired by subtracting the current probability $\mathbf{p}_i^{t-1}$ from the ground truth. It is noted that, the eye openness probability is labeled as 1 when eye open and 0 when closed.

3.2.2. Estimate the eye openness probability

After learning the parameters β and $\mathbf{b}^t$ for each iteration, given the current key point locations, we can estimate the update probability $\Delta\mathbf{p}^t$ for next iteration by:

$$\Delta\mathbf{p}^t = \beta^t \Psi(I, \mathbf{x}^{t-1}) + \mathbf{b}^t \quad (2)$$

Then eye openness probability can be acquired though:

$$\begin{aligned} \mathbf{p}^t &= \mathbf{p}^{t-1} + \Delta\mathbf{p}^t \\ subject\ to &: \mathbf{0} \le \mathbf{p}^t \le \mathbf{1} \end{aligned} \quad (3)$$

3.3. Update point locations

After estimating the eye openness probability, the 5 key point locations can be updated. We also use a linear regression model for point location prediction. Intuitively, when the pupil is occluded with a low visibility probability, the local appearance features are less reliable for the localization of pupil. When the eyes are totally closed, the pupil related SIFT features should be discarded. Based on this intuition, we modify the regression method with the visibility probability as in Eq. (4).

3.3.1. Learn the eye location prediction model

For the training, similar to Eq. (1), we learn the weight parameters α^t and bias $\mathbf{c}^t$ by a standard least-square formulation with closed form solution:

$$\alpha^{t^*}, \mathbf{c}_i^{t^*} = arg\min_{\alpha^t, \mathbf{c}_i^t} \sum_{i=1}^{K} \| \Delta\mathbf{x}_i^t - \alpha^t[\sqrt{\mathbf{p}_i^t} \circ \Psi(I_i, \mathbf{x}_i^{t-1})] - \mathbf{c}_i^t \|^2 \quad (4)$$

where K is the number of training samples and $\circ$ denotes block-wise product. It is worth nothing that, eye center related features are weighted by the corresponding openness probability after applying block-wise multiplication. Hence, when the eye is partially occluded with a low probability, the corresponding pupil features are less reliable for the eye detection. Given the current point locations $\mathbf{x}^{t-1}$, the combined shape and appearance features can be extracted. Then the eye openness probability is acquired by Eqs. (2) and (3). During training, the update $\Delta\mathbf{x}^t$ is estimated by subtracting the current key point locations $\mathbf{x}^{t-1}$ from the ground truth locations.

3.3.2. Infer the eye location

Given the image I and corresponding key point locations $\mathbf{x}^{t-1}$, according to Eqs. (2) and (3) in inference, the current eye openness probability $\mathbf{p}^t$ can be calculated. After learning the parameters α^t and bias $\mathbf{c}^t$ for each iteration, we can estimate the update location $\Delta\mathbf{x}^t$ for iteration t by:

$$\Delta\mathbf{x}^t = \alpha^t[\sqrt{\mathbf{p}^t} \circ \Psi(I, \mathbf{x}^{t-1})] + \mathbf{c}^t \quad (5)$$

In Eq. (5), $\circ$ denotes block-wise product, and it allows for weighting the pupil related appearance features. As a result, the feature vector $\Psi(I, \mathbf{x}^{t-1})$ is weighted though $\mathbf{p}^t$ such that pupil less likely to be occluded contributes more to $\Delta\mathbf{x}^t$. Then key point locations for next iteration can be acquired through:

$$\mathbf{x}^t = \mathbf{x}^{t-1} + \Delta\mathbf{x}^t \quad (6)$$

where the eye center location can be acquired.

4. Experiments and results

In this section, we firstly describe the implementation details. Then we evaluate the proposed eye localization and eye state prediction method and compare it with the stat-of-the-art methods on two benchmark databases including BioID [33] and GI4E [34]. To verify the robustness of proposed method, we evaluate it on the extremely challenging real world driving videos [35].

4.1. Implementation details

4.1.1. Evaluation database

We train the cascade regression model using 5274 images, which consists of 1690 eye images, 2958 face images from MUCT [36], 594 images with closed eyes from CEW [23], and 32 images with one eye open and another closed collected from the Internet. In addition, the training images are augmented by perturbing the scale, rotation angle, and position of the initial eye shape for learning.

One test dataset is GI4E [34]. It contains 1236 images of 103 subjects with 12 different gaze directions. These images have a resolution of 800 × 600 in pixel and are representative for the ones that can be acquired by a normal camera. Another test set is BioID [33], which is one of the most widely used database for

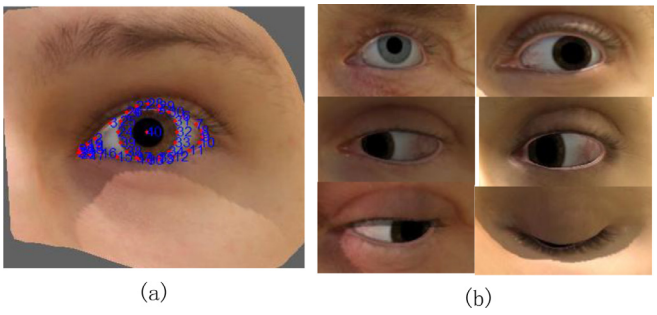

Fig. 2. Samples of synthetic eye. (a) Origninal synthetic left eye with landmark location labels. (b) Synthetic eyes with different illuminations, head pose and gaze, various eye shape and textures.

eye center localization. It is also widely used for eye state estimation. The BioID database contains 1521 gray images with a resolution of 384 × 286. This database is very challenging with complex backgrounds, various illuminations, different face sizes and head poses, and subjects with glasses or closed eyes. To verify robustness of our proposed method, 2220 extreme challenging frames from Strategic Highway Research Program (SHRP2) [35] database are chosen and manually labeled for testing. SHRP2 database consists of 44 driving videos with a resolution of 720 × 480.

4.1.2. Synthetic eye

We use *UnityEyes* from [26] to generate 1690 synthetic eye images with landmark labels for training to boost the performance. In [26], the authors adopt image-based lighting and rasterizing method to cover the various illumination conditions. By driving 3D eye region model from 3D face scans, various eyeball texture and shapes, iris width and color can be generated. In addition, different head poses can be generated by using spherical coordinates and pointing it towards the eye ball center. More details about *UnityEyes* can be found in [26] and it is public available. Some synthetic eye images are shown in Fig. 2. Since it only generates left eye images, we flip them to train models for right eye detection.

4.1.3. Evaluation criteria

The maximum normalized error [33] is adopted to evaluate the performance of eye center localization. It is defined as follows:

$$d_{eye} = \frac{max(d_r, d_l)}{\|C_r - C_l\|} \tag{7}$$

where d_r and d_l are the Euclidean distances between the estimated right and left eye centers and the ones in the ground truth, and C_r and C_l are the true centers of the right pupil and left pupil respectively. d_{eye} is normalized by the inter-ocular distance. It measures the error obtained by the worst of both eye estimation. In this measure, $d_{eye} \leq 0.25$ corresponds to the distance between eye center and eye corner, $d_{eye} \leq 0.1$ corresponds to the range of iris, and $d_{eye} \leq 0.05$ corresponds to the range of pupil diameter.

4.1.4. Parameters setting

The OpenCV implementation of boosted cascade face detector proposed by Viola and Jones [37] is used for face detection. The minimal face region is set to 50 × 50 and the largest detected face is chosen as the final detection result. The false negatives from the test set are discarded. As a result, the face detection rates on BioID and GI4E database are 97.5% and 99.4%, respectively. The number of iterations for the cascade regression model is set to 4. The normalized eye corner distance is 25 pixels. For binary eye state estimation, the threshold is 0.2. That means when the estimated openness probability is below 0.2, the predicted eye state is to be close.

Table 1
Eye localization results on BioID($d_{eye} \leq 0.05$) based on different training data.

Training data	Real	Synthetic	Combination
$d_{eye} \leq 0.05$	90.3%	88.6%	**91.2%**

Table 2
Eye localization comparison using the normalizaed error measurement on BioID database.

Method	$d_{eye} \leq 0.05$	$d_{eye} \leq 0.1$	$d_{eye} \leq 0.25$
Campadelli2009 [38]	80.7%	93.2%	99.3%
Timm2011 [19]	82.5%	93.4%	98.0%
Valenti2012 [5]	86.1%	91.7%	97.9%
Chen2014 [18]	87.3%	94.9%	99.2%
Araujo2014 [7]	88.3%	92.7%	98.9%
Chen2015 [10]	88.8%	95.2%	99.0%
Ours	**91.2%**	**99.4%**	**99.8%**

Table 3
Eye state estimation comparison on BioID database.

Method	Accuracy
Cheng2012 [39]	94.0%
Song2014 [23]	97.1%
Lin2015 [25]	97.5%
Ours	**98.1%**

4.2. Experimental results

To verify the effectiveness of learning-by-synthesis for eye detection, we firstly perform training on 3584 real images, 1690 synthetic images and their combination separately and test on BioID database. As show in Table 1, training on combination of real data and synthetic data improves the performance. The following experiments are based on training on combination of real and synthetic data.

4.2.1. Test on BioID

We further compare the performance of the proposed method on most widely used BioID database in Table 2 and Fig. 4 with the state-of-the-art eye localization methods. Since there is little literature focusing on both eye localization and eye state estimation, separate comparison with existing methods for eye state estimation is listed in Table 3. The best performance for evaluation criteria is highlighted in bold. As shown in Tables 2 and 3, the performance of our proposed method is significant better than other state-of-the-art methods both on eye localization and eye state estimation.

Fig. 3 shows some samples of eye localization and eye state prediction by our proposed method on BioID database, where white dot represents the manual annotation and red dot denotes the predicted eye location. Even though the eye is closed or subject with glasses, we can still predict the eye locations by the captured shape and contextual information.

Since our proposed framework for eye localization and state estimation is totally automatic given the input image, it is more reasonable for real application. Some other experiments are conducted given the face or specific eye region. In this paper, another comparison experiment with a similar work [14] is conducted. Zhou et al. [14] improves the basic SDM [15] by extracting SIFT features at first stage and LBP at the following iterations. In [14], for testing, the authors firstly generate the basic eye bounding box by annotated landmarks. To fairly compare with our method, we extract eye regions by annotated eye outer corners, instead of using the automatically detected outer eye corners as the previous experiment. The results are shown in Table 4. It shows that our

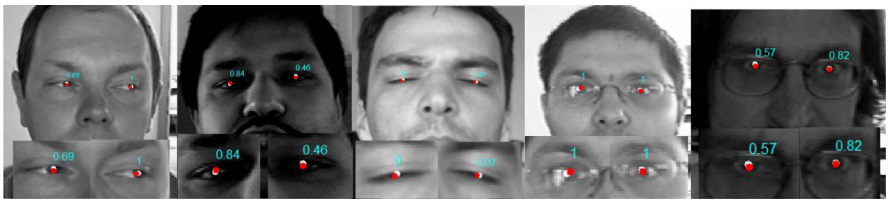

Fig. 3. Eye localization and eye openness estimation examples of successes on BioID database. The white dot represents the ground truth and red dot represents estimated eye locations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

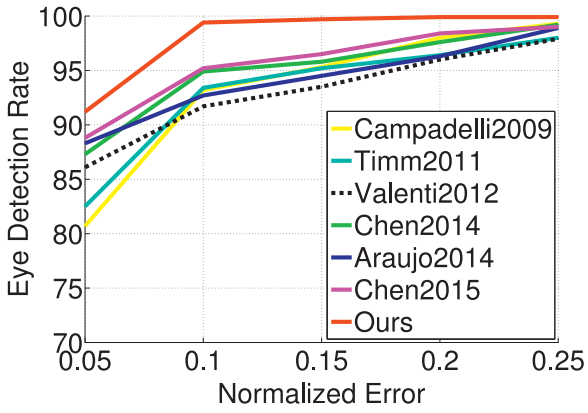

Fig. 4. Accuracy curve of eye detection rate of the our method on the BioID database, in comparison with other state-of-the-art methods.

Table 4
Eye localization comparison results with SDM-based methods.

Method	$d_{eye} \leq 0.05$	$d_{eye} \leq 0.1$	$d_{eye} \leq 0.25$
Basic-SDM [14]	90.3%	96.4%	100.0%
CF-MF-SDM [14]	93.8%	99.8%	99.9%
Ours	**95.1**%	**99.9**%	**100.0**%

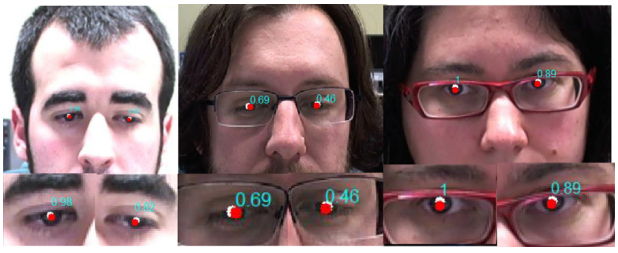

Fig. 5. Eye localization and eye openness estimation examples of successes on GI4E database. The white dot represents the ground truth and red dot represents esimated eye locations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Eye localization comparison on GI4E database.

Method	$d_{eye} \leq 0.05$	$d_{eye} \leq 0.1$	$d_{eye} \leq 0.25$
Timm2011 [19]	92.4%	96.0%[a]	97.5%[a]
Villanueva2013 [34]	93.9%	97.3%[a]	98.5%[a]
Ours	**94.2**%	**99.1**%	**99.8%**

[a] Are estimated from the accuracy curves in corresponding paper [34].

proposed method performs significantly better than the similar existing work.

4.2.2. Test on GI4E

Experimental results on GI4E database are listed in Table 5. We only conduct comparisons of eye detection on GI4E because all the testing images are with open eyes. Our proposed approach

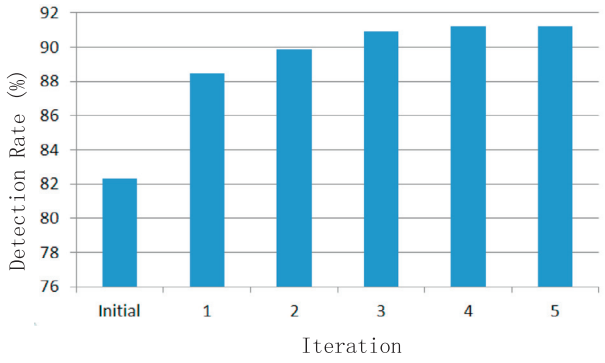

Fig. 6. Eye detection results at each cascaded iteration on BioID database. Y coordinate denotes the detection rate with the normalized error less than 0.05. It converges after fourth iteration.

Fig. 7. Samples of eye detection and eye state estimation results on extreme challenging SHRP2 driving database.The faces of the subjects are partially covered as this identity information can not be made public under the terms of a data sharing agreement.

achieves preferable location results compared with other methods. As shown in Table 5, a detection rate of 94.2% can be achieved when normalized error is $d_{eye} \leq 0.05$. Compared with results in testing on challenging database BioID, results on GI4E are better since the testing images are more clear with smaller range of head pose and illumination changes. Some qualitative results are shown in Fig. 5. It is worth nothing that purely learning-by-synthesis can perform better on GI4E with more synthetic eyes and using more landmarks. But it is not robust enough to handle the realistic images like ones from BioID.

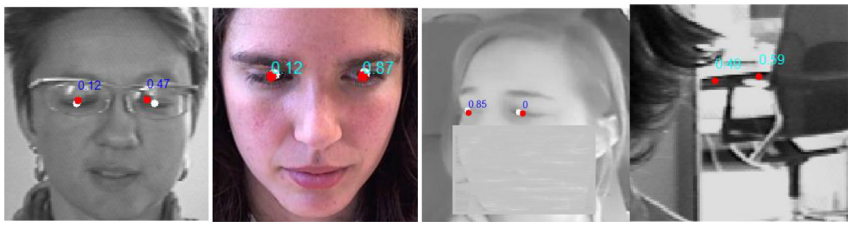

Fig. 8. Eye location and eye state estimation examples of failures on testing database. The white dot represents the ground truth and red dot represents estimated eye locations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6
Eye localization and state estimation results on SHRP2.

$d_{eye} \leq 0.1$	$d_{eye} \leq 0.25$	Eye State Accuracy
83.8%	98.2%	91.4%

4.2.3. Test on SHRP2

To verify the robustness of the proposed method, we evaluate it on extremely challenging SHRP2 database with the trained model in previous experiments. 2220 driving frames are chosen from different videos for our quantitative testing. Some examples are shown in Fig. 7. The eye detection and state estimation result are shown in Table 6. Normalized error less than 0.05 corresponding to pupil diameter is not calculated since even human can not accurately annotate the pupil location in such low resolution images. The public available code from [17] is used to test on this database which can only achieve 43.1% with normalized error 0.1. Experimental result show that the proposed method is robust enough to deal with these challenging images.

4.2.4. Further analysis

Since cascade regression converges to different eye locations at different cascade level, we further investigate the convergence of cascade regression for eye detection. We take the results of BioID database for example, as shown in Fig. 6, it converges fast at first two iterations and to optimal after 4 iterations. We initialize it by mean eye locations after coarsely extracted eye regions using our landmark detection method and achieve only detection rate of 82.3%. After several cascade regression, we can achieve a detection rate of 91.2%.

Some eye detection and state estimation results of failure are shown in Fig. 8, where white dot represents the ground truth, red dot represents the prediction and the digit is the estimated openness probability. To capture the contextual information, we combine features of 5 landmark together as the input of regression model to estimate the locations and openness probability. Hence, the feature of eye corners also have effects on final estimation. As shown in Fig. 8, for the first case, it yields inaccurate eye detection and eye state estimation under strong highlights on the glasses. For the second case, we fail to estimate the state of right eye due to the various appearance of eyelashes. Large head pose also leads to inaccurate openness estimation as shown in Fig. 8 for the third case. In addition, false positive of face detection result in false eye detection and state estimation like the last case shown in Fig. 8. It should be noted that we do not discard these images with false positives for face detection during testing.

We also use the SURF features to capture local appearance and it achieves 85.9% detection rate on Gi4E database where SIFT features can achieve 94.2% with normalized error of 0.05. All experiments are conducted with nonoptimized Matlab codes on a standard PC, which has an Intel i5 3.47 GHz CPU and 16 GB RAM. 15 frames per second can be achieved by our proposed method, which allows for near real time eye detection.

4.3. Further discussion and future work

The experimental results demonstrate that our proposed method can achieve preferable results both on eye detection and eye state estimation on benchmark databases. Based on the cascaded framework, it simultaneously updates the eye location and eye openness probability at each iteration. By further investigation, as shown in Table 4, the performance of proposed frame work is sensitive to eye region detection since the initialization of 5 key landmarks is important. It can not converge to the global optimization when the initialization of key points is far away from the ground truth. Actually, if we only train on large number of synthetic data using more eye related landmark and test on clear GI4E images, it can get improvement. But it performs much worse on BioID and SHRP2 database. In addition, it can improve the performance by using a good face detector. Further work will focus on these problems.

Due to the effectiveness and efficiency of our proposed method, it can be widely used for real application like iris detection and recognition. Moreover, Our proposed method can be applied for visual analysis for driver like gaze estimation, monitoring the driver attention and calculating PERCLOSE for driver fatigue detection.

5. Conclusions

In this paper, we propose an effective cascade regression method for simultaneous eye localization and eye state estimation. The binary eye state is mapped to a continuous variable denoted by eye openness probability. Both eye center position and eye openness probability are updated during regression iterations using captured shape, appearance, structural and contextual information. In addition, eye localization relies less on appearance information of pupil with low openness probability. Experimental results show that our proposed method is significantly better than other state-of-the-art methods for both eye localization and eye state estimation.

In the future, we will focus on applying the proposed simultaneous eye detection and eye state estimation method to other applications, such as eye tracking, gaze estimation and driver fatigue monitoring. In addition, we will further improve the method to robustly deal with large head poses in unconstrained scenarios.

Acknowledgments

This work was completed when the first author visited Rensselaer Polytechnic Institute (RPI), supported by a scholarship from University of Chinese Academy of Sciences (UCAS). The authors would like to acknowledge support from UCAS and RPI. This work was also supported in part by National Science Foundation under the grant 1145152 and by the National Natural Science Foundation of China under Grant 61304200 and 61533019.

References

[1] D.W. Hansen, Q. Ji, In the eye of the beholder: a survey of models for eyes and gaze, Pattern Anal. Mach. Intel. IEEE Trans. 32 (3) (2010) 478–500.
[2] F. Song, X. Tan, S. Chen, Z.-H. Zhou, A literature survey on robust and efficient eye localization in real-life scenarios, Pattern Recognit. 46 (12) (2013) 3157–3173.
[3] A.L. Yuille, P.W. Hallinan, D.S. Cohen, Feature extraction from faces using deformable templates, Int. J. Comput. Vis. 8 (2) (1992) 99–111.
[4] D.W. Hansen, A.E. Pece, Eye tracking in the wild, Comput. Vision Image Understanding 98 (1) (2005) 155–181.
[5] R. Valenti, T. Gevers, Accurate eye center location through invariant isocentric patterns, Pattern Anal. Mach. Intel. IEEE Trans. 34 (9) (2012) 1785–1798.
[6] N. Markuš, M. Frljak, I.S. Pandžić, J. Ahlberg, R. Forchheimer, Eye pupil localization with an ensemble of randomized trees, Pattern Recognit. 47 (2) (2014) 578–587.
[7] G. Araujo, F. Ribeiro, E. Silva, S. Goldenstein, Fast eye localization without a face model using inner product detectors, in: Image Processing (ICIP), 2014 IEEE International Conference on, 2014, pp. 1366–1370.
[8] C. Zhang, X. Sun, J. Hu, W. Deng, Precise eye localization by fast local linear SVM, in: Multimedia and Expo (ICME), 2014 IEEE International Conference on, IEEE, 2014, pp. 1–6.
[9] Y. Wu, Q. Ji, Learning the deep features for eye detection in uncontrolled conditions, in: Pattern Recognition (ICPR), 2014 22nd International Conference on, IEEE, 2014, pp. 455–459.
[10] S. Chen, C. Liu, Eye detection using discriminatory haar features and a new efficient SVM, Image Vis. Comput. 33 (2015) 68–77.
[11] B.C. Sidney, Introduction to wavelets and wavelet transforms: a primer, 1998,
[12] N.S. Karakoc, S. Karahan, Y.S. Akgul, Regressor based estimation of the eye pupil center, in: Pattern Recognition, Springer, 2015, pp. 481–491.
[13] P. Yang, B. Du, S. Shan, W. Gao, A novel pupil localization method based on gaboreye model and radial symmetry operator, in: Image Processing, 2004. ICIP'04. 2004 International Conference on, vol. 1, IEEE, 2004, pp. 67–70.
[14] M. Zhou, X. Wang, H. Wang, J. Heo, D. Nam, Precise eye localization with improved SDM, in: Image Processing (ICIP), 2015 IEEE International Conference on, IEEE, 2015, pp. 4466–4470.
[15] X. Xiong, F. De la Torre, Supervised descent method and its applications to face alignment, in: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE, 2013, pp. 532–539.
[16] P.F. Felzenszwalb, D.P. Huttenlocher, Pictorial structures for object recognition, Int. J. Comput. Vis. 61 (1) (2005) 55–79.
[17] X. Tan, F. Song, Z.-H. Zhou, S. Chen, Enhanced pictorial structures for precise eye localization under incontrolled conditions, in: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE, 2009, pp. 1621–1628.
[18] S. Chen, C. Liu, Clustering-based discriminant analysis for eye detection, Image Process. IEEE Trans. 23 (4) (2014) 1629–1638.
[19] F. Timm, E. Barth, Accurate eye centre localisation by means of gradients., in: VISAPP, 2011, pp. 125–130.
[20] Y. Kurylyak, F. Lamonaca, G. Mirabelli, Detection of the eye blinks for human's fatigue monitoring, in: Medical Measurements and Applications Proceedings (MeMeA), 2012 IEEE International Symposium on, IEEE, 2012, pp. 1–4.
[21] F. Yutian, H. Dexuan, N. Pingqiang, A combined eye states identification method for detection of driver fatigue, in: Wireless Mobile and Computing (CCWMC 2009), IET International Communication Conference on, IET, 2009, pp. 217–220.
[22] D. González-Ortega, F. Díaz-Pernas, M. Antón-Rodríguez, M. Martínez-Zarzuela, J. Díez-Higuera, Real-time vision-based eye state detection for driver alertness monitoring, Pattern Anal. Appl. 16 (3) (2013) 285–306.
[23] F. Song, X. Tan, X. Liu, S. Chen, Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients, Pattern Recognit. 47 (9) (2014) 2825–2838.
[24] A. Punitha, M.K. Geetha, Driver Eye State Detection Based on Minimum Intensity Projection Using Tree Based Classifiers, in: Intelligent Systems Technologies and Applications, Springer, 2016, pp. 103–111.
[25] X. Lin, L. Cai, R. Ji, An effective eye states detection method based on the projection of the gray interval distribution, in: Image Processing (ICIP), 2015 IEEE International Conference on, IEEE, 2015, pp. 1875–1879.
[26] E. Wood, T. Baltrušaitis, L.-P. Morency, P. Robinson, A. Bulling, Learning an appearance-based gaze estimator from one million synthesised images, in: Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications, ACM, 2016, pp. 131–138.
[27] E. Wood, T. Baltrušaitis, X. Zhang, Y. Sugano, P. Robinson, A. Bulling, Rendering of eyes for eye-shape registration and gaze estimation (2015) arXiv:1505.05916.
[28] X. Zhang, Y. Sugano, M. Fritz, A. Bulling, Appearance-based gaze estimation in the wild, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 4511–4520.
[29] C. Gou, Y. Wu, K. Wang, F.-Y. Wang, Q. Ji, Learning-by-synthesis for accurate eye detection., ICPR, 2016.
[30] Y. Wu, Q. Ji, Robust facial landmark detection under significant head poses and occlusion, in: Proceedings of the IEEE International Conference on Computer Vision, 2015a, pp. 3658–3666.
[31] Y. Wu, Q. Ji, Shape augmented regression method for face alignment, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2015b, pp. 26–32.
[32] Y. Wu, Q. Ji, Constrained joint cascade regression framework for simultaneous facial action unit recognition and facial landmark detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016.
[33] O. Jesorsky, K.J. Kirchberg, R.W. Frischholz, Robust face detection using the hausdorff distance, in: Audio-and Video-Based Biometric Person Authentication, Springer, 2001, pp. 90–95.
[34] A. Villanueva, V. Ponz, L. Sesma-Sanchez, M. Ariz, S. Porta, R. Cabeza, Hybrid method based on topography for robust detection of iris center and eye corners, ACM Trans. Multimed. Comput. Commun. Appl. (TOMM) 9 (4) (2013) 25.
[35] Transportation Research Board of the National Academies of Science, The 2nd Strategic Highway Research Program Naturalistic Driving Study Dataset, https://insight.shrp2nds.us/, 2013.
[36] S. Milborrow, J. Morkel, F. Nicolls, The muct landmarked face database, Pattern Recognit. Assoc. South Africa 201 (0) (2010).
[37] P. Viola, M.J. Jones, Robust real-time face detection, Int J Comput Vis 57 (2) (2004) 137–154.
[38] P. Campadelli, R. Lanzarotti, G. Lipori, Precise eye and mouth localization, Int. J. Pattern Recognit. Artif. Intell. 23 (03) (2009) 359–377.
[39] E. Cheng, B. Kong, R. Hu, F. Zheng, Eye state detection in facial image based on linear prediction error of wavelet coefficients, in: Robotics and Biomimetics, 2008. ROBIO 2008. IEEE International Conference on, IEEE, 2009, pp. 1388–1392.

Chao Gou received his B.S. degree in automation from the University of Electronic Science and Technology of China in 2012. He is currently a Ph.D. student at the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. Since September 2015, he has been a Visiting Student at Rensselaer Polytechnic Institute, Troy, New York. His research interest covers computer vision, pattern recognition and intelligent transportation systems.

Yue Wu received the B.S and M.S degrees in electrical engineering from the Southeast University of China in 2008 and 2011, respectively. She is currently pursuing the Ph.D. degree at Rensselaer Polytechnic Institute, Troy, New York. Her areas of research include computer vision, pattern recognition, and their applications in human-computer interaction. She is a student member of the IEEE and the IEEE Computer Society.

Kang Wang received his B.S degree from Department of Electronic Engineering and Information Science, University of Science and Technology of China in 2013. He is currently pursuing the Ph.D. degree at Rensselaer Polytechnic Institute, Troy, New York. His main research interests include computer vision and machine learning.

Kunfeng Wang received his Ph.D. degree in control theory and control engineering from the Graduate University of Chinese Academy of Sciences in 2008. He is currently an associate professor at the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. Since January 2016, he has been a visiting scholar at the College of Computing, Georgia Institute of Technology, U.S.A. His research interests include intelligent transportation systems, intelligent vision computing, and machine learning.

Fei-Yue Wang received the Ph.D. degree in computer and systems engineering from Rensselaer Polytechnic Institute, Troy, NY, USA, in 1990. He was an editor-in-chief of the *International Journal of Intelligent Control and Systems*, the *World Scientific Series in Intelligent Control and Intelligent Automation*, *IEEE Intelligent Systems*, and the *IEEE Transactions on Intelligent Transportation Systems*. He has served as chairs for over 20 IEEE, ACM, INFORMS, and ASME conferences. His current research interests include social computing, complex systems, and intelligent computing and control. He is currently the Director of the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, the Vice President of the ACM China Council and the Vice President/Secretary-General of Chinese Association of Automation. He is a member of Sigma Xi and an elected fellow of IEEE, INCOSE, IFAC, ASME, and AAAS.

Qiang Ji received his Ph.D. degree in electrical engineering from the University of Washington. He is currently a professor with the Department of Electrical, Computer, and Systems Engineering at Rensselaer Polytechnic Institute (RPI). He recently served as director of the Intelligent Systems Laboratory (ISL) at RPI. Prof. Ji's research interests are in computer vision, probabilistic graphical models, information fusion, and their applications in various fields. Prof. Ji is an editor on several related IEEE and international journals and he has served as a general chair, program chair, technical area chair, and program committee member in numerous international conferences/workshops. Prof. Ji is a fellow of IEEE and IAPR.