Gaze-Aided Eye Detection via Appearance Learning

Lin Cao*,1,2, Chao Gou*,1,3, Kunfeng Wang1,3, Gang Xiong†,1,4, and Fei-Yue Wang1,3

[1]Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2]University of Chinese Academy of Sciences, Beijing, China
[3]Qingdao Academy of Intelligent Industries, Qingdao, China
[4]Cloud Computing Center, Chinese Academy of Sciences, Dongguan, China
{caolin2016,chao.gou,kunfeng.wang,gang.xiong,feiyue.wang}@ia.ac.cn

***Abstract*—Image based eye detection and gaze estimation have a wide range of potential applications, such as medical treatment, biometrics recognition, human-computer interaction. Though a large number of researchers have attempted to solve the two problems, they still exist some challenges due to the variation in appearance and lack of annotated images. In addition, most related work perform eye detection first, followed by gaze estimation via appearance learning. In this paper, we propose a unified framework to execute the gaze estimation and the eye detection simultaneously by learning the cascade regression models from appearance around the eye related key points. Intuitively, there is coupled relationship among location of eye center, shape of eye related key points, appearance representation and gaze information. To incorporate these information, at each cascade level, we first learn a model to map the shape and appearance around current eye related key points to the three dimension gaze update. Then, with the help of estimated gaze, we further learn a regression model to map the gaze, shape and appearance information to eye location update. By leveraging the power of cascade learning, the proposed method can alternatively optimize the two tasks of eye detection and gaze estimation. The experiments are conducted on benchmarks of GI4E and MPIIGaze. Experimental results show that our proposed method can achieve preferable results in gaze estimation and outperform the state-of-the-art methods in eye detection.**

I. Introduction

Eyes are the most salient facial components which reflect human's affective states. The accurate detection of eyes is essential for the success of wide range of applications such as iris recognition, eye state and eye gaze estimation in human-computer interaction. Image based eye detection is to estimate the pupil location in a 2D image. Gaze estimation is to obtain gaze direction or gaze point where people look at with the use of mechanical, electronic, optical and other detection means. By the estimation of eye gaze direction, the region or targets of interest can be found more accurately, on the other hand, the further study of human psychology and even physical function can be done such as the driver's fatigue state detection for safe, virtual reality, the diagnosis of cognitive impairment, etc.

With wide range of application, image based eye detection and gaze estimation have gained increasing attention. Recently, appearance based methods have achieved the state-of-the-art results [1], [2], [3]. Although much work has been done by researchers, they are still challenging tasks due to pose, glasses, illumination and facial expressions. In addition, most existing methods only focus on eye detection or gaze estimation separately, ignoring the coupled relation between the eye location and gaze direction. In this paper, we present a unified framework for simultaneous eye detection and eye gaze estimation on the basis of cascade regression. Overall framework is shown in Fig. 1. We first detect facial landmarks by [1] and coarsely extract the eye regions, followed by cascade regression for eye detection and gaze estimation. During the iteration at each cascade level, eye gaze can be updated by the current appearance and shape information. Then eye related point locations are updated simultaneously based on the estimated gaze, appearance and shape information.

The main contributions of our work are summarized into four folds: 1) Different from conventional appearance based methods for gaze estimation, we propose to incorporate local appearance and shape information of eye related landmarks for gaze estimation. 2) To capture the correspondence of gaze and eye center, we learn the cascade regression models to update the eye center based on shape, appearance and estimated gaze. With the help of gaze information, we can further improve the eye detection performance. 3) Different from the conventional methods that independently perform eye detection and eye gaze estimation, the proposed method can simultaneously detect eye center and gaze on the basis of cascade regression framework. 4) The proposed method is effective and efficient. It outperforms the state-of-the-arts in eye detection and can achieve real time applications.

The remainder of this paper is arranged as follows. In Section II, the related work about eye detection and gaze estimation are described. The details of proposed method are presented in section III. In section IV, the experimental results are discussed. Then the conclusions are drawn in Section V.

II. Related Work

For the eye detection, we review the recent appearance based work. A earlier detailed review about progress and state-of-the-art methods is described in [5]. In [6], the authors proposed a discriminating feature extraction method applied to derive the discriminatory haar features (DHFs) and a new efficient support vector machine (eSVM) to improve the efficiency of the SVM for eye detection. Zhang *et al.* [7] presented an approach based on isophote and gradient features and selective oriented gradient (SOG) filter for accurate and real-time eye

*contributed equally to this paper,†corresponding author.

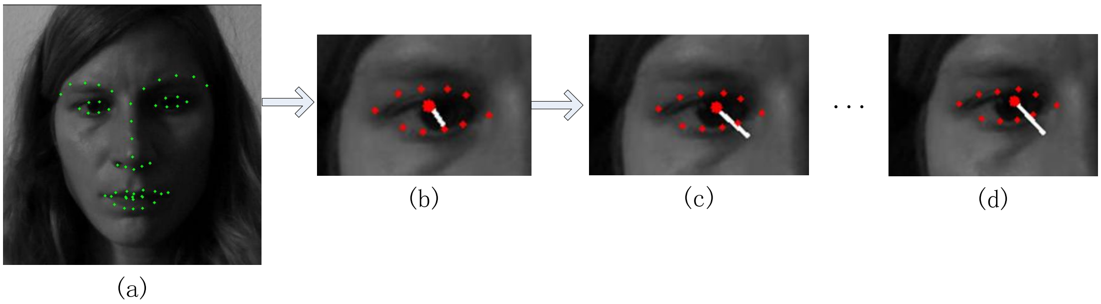

Fig. 1. Framework of our proposed approach. (a) Facial landmark detection. (b) Coarsely extract eye region, initialize key points positions and estimate the gaze. (c) Results of first iteration. (d) Final results of detection.

center detection. In [8], eye pupil center detection method is proposed using HOG features to estimate the distance between pupil center and patch center. Zhou *et al.*[9] proposed to apply SDM [10] using multi-scale nonlinear features for accurate eye detection. Florea *et al.* [11] encode the normalized images projections with zeros-crossing based method and further apply multi layer perceptron (MLP) classier for eye location. Gou *et al.* [1], [3], [4] propose coupled cascade regression for facial landmark detection and learn cascade regression eye detection models from real and synthetic images based on the local appearance, shape and structural features.

Gaze estimation methods are mainly categorized into model-based and appearance-based [5], [12]. We review the appearance based approaches in this paper. A more detailed review about the recent eye gaze research is summarized in [13]. Appearance based gaze estimation methods compute features for input eye images and try to learn the mapping functions between feature representations and low dimensional gaze direction. In [14], dually supervised manifold embedding is shown to improve the performance of gaze estimation. Lu *et al.* [15] proposed an adaptive linear regression method to map various appearance features to corresponding nine gaze points. Except for appearance features, Guo *et al.* [16] incorporate shape information to enhance the gaze estimation. Zhang *et al.* [17] introduce a large scale dataset of MPIIGaze and present a multimodal CNN for gaze estimation. In [18], an end-to-end eye tracking solution targeting mobile devices is introduced. A smartphone is used to collect eye images and a CNN model is trained for eye tracking. In [2], a large number of synthetic images of eyes generated by a dynamic eye region model are built for training eye-shape registration and gaze estimation model.

In the previous researches, most related literatures perform eye location or gaze estimation separately, or eye location first and then estimate gaze based on eye center. In this paper, we propose to jointly detect eye center and estimate gaze based on the cascade regression framework.

III. Proposed Framework

Intuitively, there is a coupled relationship between gaze direction and pupil location. The local appearance and shape of eye center and other related key points are the most notable

Algorithm 1 Proposed framework for gaze estimation and eye detection.

Input:
Give the eye image $\mathbf{I}$. Eye related key point locations $\mathbf{x}^0$ are initialized by the mean eye locations from training data and gaze position is initialized by $\mathbf{g}^0$.
Do cascade regression:
for t=1,...,T **do**
Estimate the gaze vector update $\Delta\mathbf{g}^t$ given the current eye related landmark locations $\mathbf{x}^{t-1}$
$h_t : \mathbf{I}, \mathbf{x}^{t-1} \rightarrow \Delta\mathbf{g}^t$
Update the gaze vector
$\mathbf{g}^t = \mathbf{g}^{t-1} + \Delta\mathbf{g}^t$
Estimate the key point location updates given the current key point locations $\mathbf{x}^{t-1}$ and $\mathbf{g}^t$
$f_t : \mathbf{I}, \mathbf{x}^{t-1}, \mathbf{g}^t \rightarrow \Delta\mathbf{x}^t$
Update the key point locations
$\mathbf{x}^t = \mathbf{x}^{t-1} + \Delta\mathbf{x}^t$
end for
Output:
Acquire the gaze vector $\mathbf{g}^T$ and locations $\mathbf{x}^T$ of eye landmarks including the eye center.

gaze features. Different gaze is commonly accompanied by different eye center location.

In this paper, we propose a unified framework for simultaneous gaze estimation and eye center detection on the basis of cascade regression. The overall framework of our proposed method is summarized in Algorithm 1. As shown in Fig. 1, we consider N eye related key points consisting of eye corners, eye centers, and points on eyelids to capture the shape and local appearance information. Motivated by cascade regression for facial landmark detection, we iteratively update eye related key point locations $\mathbf{x} \in \Re^{2 \cdot N}$. Given current key point locations $\mathbf{x}^{t-1}$, at cascade level t, we propose to learn a regression model h_t to map the current local appearance feature (e.g. SIFT) and shape feature to gaze update $\Delta\mathbf{g}^t \in \Re^3$ and then estimate gaze position $\mathbf{g}^t$. Motivated by SDM for eye detection [1], we learn another regression model f_t to map the estimated gaze $\mathbf{g}^t$, local appearance feature and shape feature to the updates of key points $\Delta\mathbf{x}^t$. After convergence, we can achieve gaze estimation and eye detection simultaneously. In the following, more details of the proposed method are described.

1) Estimate the gaze : As listed in Algorithm 1, we first perform gaze estimation at the cascade level t. In our case, a linear regression model h_t is applied to predict gaze updates $\Delta\mathbf{g}^t$ based on local appearance feature $\Phi(I,\mathbf{x}^{t-1})$ and shape feature $\Psi(\mathbf{x}^{t-1})$. h_t is defined as below:

$$h_t : \Delta\mathbf{g}^t = \mathbf{a}^t\Phi(I,\mathbf{x}^{t-1}) + \mathbf{b}^t\Psi(\mathbf{x}^{t-1}) + \mathbf{c}^t, \tag{1}$$

where $\Phi(I,\mathbf{x}^{t-1}) \in \Re^{N\cdot 128}$ is the calculated SIFT features representing local appearance, $\Psi(\mathbf{x}^{t-1})$ represents the shape features which are calculated as the difference of each pair of key points, $\mathbf{a}^t, \mathbf{b}^t$, and $\mathbf{c}^t$ are the parameters of regression model for gaze estimation at iteration t.

For the training at cascade level t, given K training samples with gaze annotation $\mathbf{g}_i^*$, gaze update $\Delta\mathbf{g}_i^{t,*}$ can be acquired by subtracting the current gaze $\mathbf{g}_i^{t-1}$ from the groundtruth and we can learn the parameters by a standard least square formulation with closed form solution:

$$\begin{aligned}\mathbf{a}^{t^*}, \mathbf{b}^{t^*}, \mathbf{c}^{t^*} =& arg \min_{\mathbf{a}^t,\mathbf{b}^t,\mathbf{c}^t} \sum_{i=1}^{K} \parallel \Delta\mathbf{g}_i^{t,*} - \mathbf{a}^t\Phi(I,\mathbf{x}^{t-1}) \\ &- \mathbf{b}^t\Psi(\mathbf{x}^{t-1}) - \mathbf{c}^t \parallel^2\end{aligned} \tag{2}$$

After learning the model parameters, the gaze update $\Delta\mathbf{g}^t$ for the iteration t can be estimated by formula 1. Then gaze direction can be updated by :

$$\mathbf{g}^t = \mathbf{g}^{t-1} + \Delta\mathbf{g}^t \tag{3}$$

2) Update the location of eye center: Gaze direction is related to the location of eye center and local appearance. In this paper, we propose to update the eye location after estimating the gaze based on current location of eye related key points at each cascade iteration. In particular, we learn another regression model f_t to map the local appearance feature $\Phi(I,\mathbf{x}^{t-1})$, shape feature $\Psi(\mathbf{x}^{t-1})$ and gaze vector $\mathbf{g}^t$ to the updates of key point locations $\Delta\mathbf{x}^t$. In our case, a linear regression model for f_t is formulated as below:

$$f_t : \Delta\mathbf{x}^t = \alpha^t\mathbf{g}^t + \beta^t\Phi(I,\mathbf{x}^{t-1}) + \gamma^t\Psi(\mathbf{x}^{t-1}) + \varepsilon^t, \tag{4}$$

where $\mathbf{g}^t$ is the calculated gaze direction, $\Phi(I,\mathbf{x}^{t-1})$ is the calculated SIFT features, and $\Psi(\mathbf{x}^{t-1})$ represents the shape features as discussed before. $\alpha^t, \beta^t, \gamma^t$, and ε^t are the parameters of model at iteration t.

Similar to the optimization of model h_t, the parameters of f_t can be learned as below:

$$\begin{aligned}\alpha^{t^*}, \beta^{t^*}, \gamma^{t^*}, \varepsilon^{t^*} =& arg \min_{\alpha^t,\beta^t,\gamma^t,\varepsilon^t} \sum_{i=1}^{K} \parallel \Delta\mathbf{x}_i^{t,*} - \alpha^t\mathbf{g}^t \\ &- \beta^t\Phi(I,\mathbf{x}^{t-1}) - \gamma^t\Psi(\mathbf{x}^{t-1}) - \varepsilon^t \parallel^2\end{aligned} \tag{5}$$

where $\mathbf{K}$ is the number of input training images. The appearance feature $\Phi(I,\mathbf{x}^{t-1})$ and shape feature $\Psi(\mathbf{x}^{t-1})$ are extracted as Eq.1. The gaze vector $\mathbf{g}^t$ is calculated by Eq.3.

Fig. 2. Samples of MPIIGaze dataset

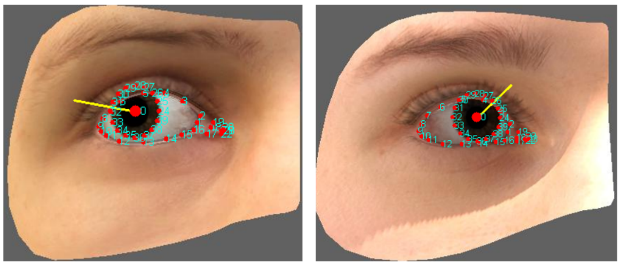

Fig. 3. Example of synthetic right eyes with variability of different head poses, illuminations, shadows

At the iteration t, given the ith image $\mathbf{I}_i$ with ground truth key points location $\mathbf{x}_i^*$, the ground truth updates of key point location can be calculated by $\Delta\mathbf{x}_i^{t,*} = \mathbf{x}_i^* - \mathbf{x}_i^{t-1}$. For the first iteration, $\mathbf{x}_i^0$ is the mean locations from training data in coarse detected eye region.

After learning the parameters of f_t, the updates of key point locations for iteration t can be acquired by Eq. 4. And then the key point locations including the eye center can be updated by $\mathbf{x}^t = \mathbf{x}^{t-1} + \Delta\mathbf{x}^t$.

IV. Experiments

To validate the performance and effectiveness of the proposed method, we conduct experiments on benchmark datasets of GI4E [19] and MPIIGaze [17]. For the proposed learning based approach, we need the annotation of key point locations. It is time-consuming to collect large-scale data with different variable appearance, pose, gaze, illuminations and it is costly and unreliable for ground truth annotation. Similar to previous work [1], we propose to learn from synthetic images generated by *UnityEyes* [20] for eye detection and gaze estimation. We can learn independent models for right and left eye. In the following, we take right eye as example for description.

A. Datasets

GI4E [19] consists of 1236 images collected from 103 different subjects where each subject is with 12 different gaze directions captured by standard webcam. The gaze points are uniformly distributed in screen and images are in resolution of 800*600. This dataset provides the ground truth of pupil location. It is widely used for eye detection. In some cases, it is challenging for the eyes with hidden by glasses and illumination changes.

MPIIGaze [17] contains 213659 images from 15 participants outside of controlled laboratory and covers a range of recording locations, illumination, eye appearances. The laptops are used to record images and can also come with high resolution front-facing cameras. Participants look at a random sequence of 20 on screen positions. The laptop models are different in screen size and resolution. We conduct experiments on the *Annotation Subset* of 10654 images from 15 participants where 8 landmarks consisting of four eye corners, two mouth corners and left and right eye locations. Fig. 2 shows some samples from MPIIGaze.

UnityEyes [20] is used to synthesize large numbers of variable eye region images with various appearance, shape, head pose and gaze directions. We use the same training data from [1] which consists of 2218 synthetic eye regions. Since the generated eye region images are from left eye, the right eye images can be generated by flipping. Fig. 3 shows some samples about generated images.

B. Eye detection

To evaluate the performance of eye detection, the widely used evaluation criteria of maximum normalized error introduced in [21] is used in this paper. It is formulated as below:

$$d_{eye} = \frac{max(D_{right}, D_{left})}{\|loc_{right} - loc_{left}\|} \tag{6}$$

where D_{right} and D_{left} are the Euclidean distances between the estimated right and left eye centers and the ones in the ground truth, and loc_{right} and loc_{left} are the ground truth eye center locations for right and left eye, respectively. d_{eye} is normalized by the inter-ocular distance.

Face detection is the first step for proposed framework. Eye regions are extracted based on facial landmarks detected by [1]. We compare our method with other appearance based methods. Table I shows the experimental results. The best performance for evaluation criteria is highlighted in bold. In the case of accurate location of the region of iris with d_{eye} below 0.1, the proposed method achieves a accuracy of 99.8%. In the case of the range of pupil diameter with $d_{eye} \leq 0.05$, the accuracy of the proposed method is 97.3%, which achieves the best performance. As shown in Fig. 4, the proposed method can simultaneously estimate the gaze and eye locations under various appearance, where the red dot and green dot represents the detected eye center, the ground truth and the white line represents the gaze information. It should be noted that, compared with a similar work [1], we use the same training and testing datasets. Our proposed method performs better over [1]. By further investigation, the incorporation of gaze information in cascade iteration improves the final eye detection result.

C. Gaze estimation

Since there is no ground truth gaze information in GI4E, we further conduct experiments on MPIIGaze [17] to validate the gaze estimation performance of the proposed method. Right eye is utilized for example in this paper. Since there is no full face region in MPIIGaze, similar to [9], we coarsely extract eye region by the annotated key points instead of detecting the landmarks. Mean absolute error (MAE) of gaze is adopted for the evaluation. We leave one subject out for testing, and the remaining for training. In addition, to show the effectiveness of the proposed appearance based method, we perform baselines using HOG and SIFT appearance features based on Support Vector Regression (SVR) and HOG based on Random Forests (RF). Experimental results are shown in Fig. 5, Fig. 6 and Table II. And Fig. 5 shows detection results of some samples on MPIIGaze database which contains different head pose, scales, illumination, hidden with glasses. Table II summarises mean angular errors of these methods. Angular error refers to the angle between the predicted gaze and ground truth. The average error of our method for all subjects is 6.9 degrees. The SVR with HOG features and SIFT features is 7.8 and 7.9 degrees, RF with HOG features is 7.8 degrees. Experimental results show that the proposed method outperforms the conventional gaze regression methods. In addition, the proposed method is robust to different light conditions, head poses, scales, shadows and other outside of controlled laboratory conditions. In this work, we incorporate shape information and local appearance features for gaze estimation at the first step of each iteration. By leveraging the cascade regression framework, we can iteratively update gaze direction based on the updated eye related key point locations, where we can extract more representable features for gaze information.

TABLE I
EYE LOCALIZATION COMPARISON ON GI4E DATABASE

Method	$d_{eye} \leq 0.05$	$d_{eye} \leq 0.1$
Timm *et al.* [22]	92.4%	96.0%
Villanueva *et al.* [19]	93.9%	97.3%
Gou *et al.* [1]	94.2%	98.3%
Proposed	**97.3%**	**99.8%**

Fig. 4. Eye detection and gaze estimation results on GI4E database, red dot represents the detected eye center and green points are the ground truth.

D. Further discussions

In our experiment, we conduct an experiment with training and testing on pure synthetic data for gaze estimation as

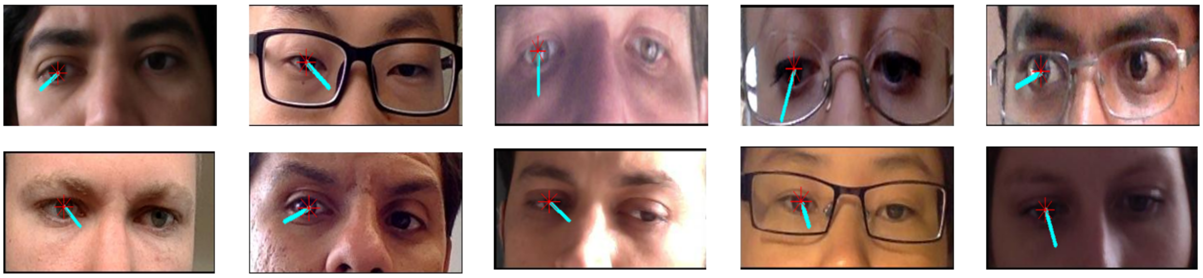

Fig. 5. Eye deteciton and gaze estimation examples results on MPIIGaze database.

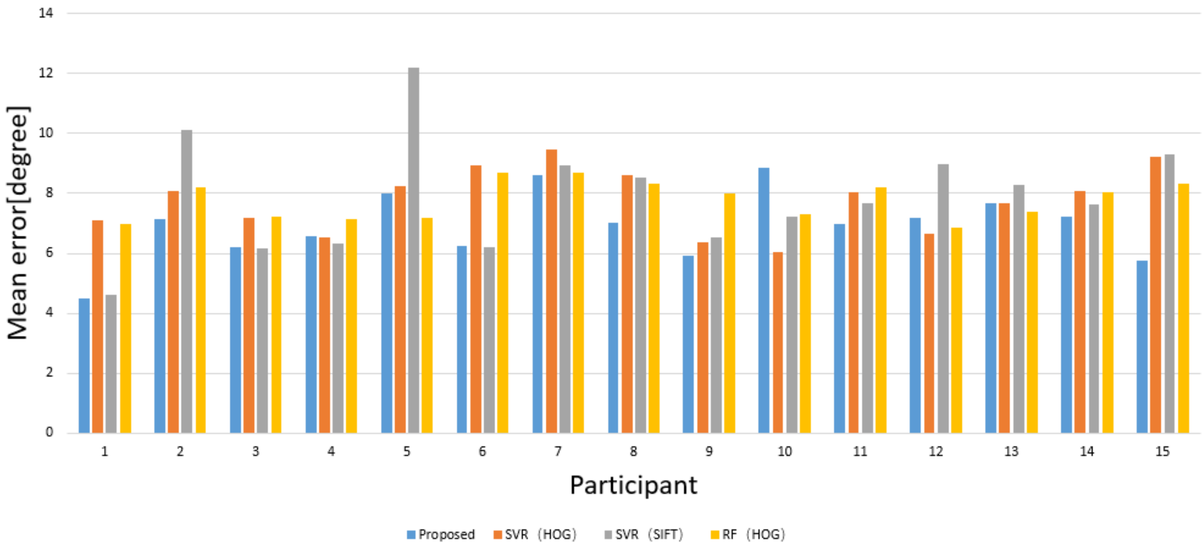

Fig. 6. Gaze estimation compared with baselines.

TABLE II
MEAN ERRORS ON MPIIGAZE IN DEGREES

Method	Proposed	SVR(SIFT)	SVR(HOG)	RF(HOG)
Error	**6.9**	7.9	7.8	7.8

well. And the experimental result shows that our proposed method achieves a MAE of 5.6 degrees and performs better than baselines where we perform regression based on the appearance features. It further demonstrates the effectiveness of the proposed method.

As presented in [1], different training samples will result in different eye detection results on GI4E, where they achieve a eye detection rate of 94.2% and 98.2% with 2218 and 10730 synthetic eye images, respectively. We also learn the model from 10730 synthetic images and achieve better eye detection rate than compared methods. Since GI4E does not provide the ground truth gaze information, we conduct qualitative experiment for gaze estimation as show in Fig.4. We also test the models learned from MPIIGaze on GI4E, the qualitative gaze estimation on GI4E is significantly better and the quantitative eye detection on GI4E achieves detection rate of 96.9% and 99.3% with $d_{eye} \leq 0.05$ and $d_{eye} \leq 0.1$, which show the superiority of proposed approach than compared methods. By further investigation, the synthetic eyes are limited to cover the variations of real images with different illuminations, head poses, and subjects with glasses. More efforts will be undertaken to add realism to the synthetic training samples in the future.

For the MPIIGaze, there are only 3 key points are labeled where we can not capture representable shape and appearance information for final eye detection and gaze estimation. It can be optimized by annotating more key points locations. As presented in [3], more eye related key points could result in higher accuracy where it is more representative for local appearance and shape information.

We empirically set the maximal iteration in cascade regression to 4. All experiments are conducted with non-optimized Matlab codes on a standard PC with Intel Core (TM) i7 CPU 3.60 GHZ. The proposed method can achieve real time application at a FPS of 15 where we perform facial landmark detection, eye detection and gaze estimation.

V. Conclusions

In this paper, we propose a new effective learning based framework, which enables effective and efficient eye detection and gaze estimation simultaneously. Based on incorporating local appearance features and shape information of eye related landmarks, gaze estimation can be achieved through regression models. The eye center can be estimated on the basis of shape, appearance and estimated gaze at previous step. By leveraging the cascade regression framework to incorporate the shape, structural, and local appearance for simultaneous eye detection and gaze estimation, our proposed method performs better over conventional regression based methods. In the future, we will focus on applying the proposed method to more applications, such as driver fatigue detection and attention estimation, cognitive process and designing powerful deep models as the regression model.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 61773381, 61304200, 61533019 and 91520301, Chinese Guangdong's S&T project (2015B010103001, 2016B090910001, 2017B090912001), Dongguan's Innovation Talents Project (Gang Xiong).

References

[1] C. Gou, Y. Wu, K. Wang, F.-Y. Wang, and Q. Ji, "Learning-by-synthesis for accurate eye detection," in *Pattern Recognition (ICPR), 2016 23rd International Conference on*. IEEE, 2016, pp. 3362–3367.

[2] E. Wood, T. Baltrusaitis, X. Zhang, Y. Sugano, P. Robinson, and A. Bulling, "Rendering of eyes for eye-shape registration and gaze estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3756–3764.

[3] C. Gou, Y. Wu, K. Wang, K. Wang, F.-Y. Wang, and Q. Ji, "A joint cascaded framework for simultaneous eye detection and eye state estimation," *Pattern Recognition*, vol. 67, pp. 23–31, 2017.

[4] C. Gou, Y. Wu, F.-Y. Wang, and Q. Ji, "Coupled cascade regression for simultaneous facial landmark detection and head pose estimation," in *Proc. IEEE International Conference on Image Processing*, 2017.

[5] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 3, pp. 478–500, 2010.

[6] S. Chen and C. Liu, "Eye detection using discriminatory haar features and a new efficient svm," *Image and Vision Computing*, vol. 33, pp. 68–77, 2015.

[7] W. Zhang, M. L. Smith, L. N. Smith, and A. Farooq, "Eye center localization and gaze gesture recognition for human–computer interaction," *JOSA A*, vol. 33, no. 3, pp. 314–325, 2016.

[8] N. S. Karakoc, S. Karahan, and Y. S. Akgul, "Regressor based estimation of the eye pupil center," in *German Conference on Pattern Recognition*. Springer, 2015, pp. 481–491.

[9] M. Zhou, X. Wang, H. Wang, J. Heo, and D. Nam, "Precise eye localization with improved sdm," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 4466–4470.

[10] X. Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 532–539.

[11] L. Florea, C. Florea, and C. Vertan, "Robust eye centers localization with zero-crossing encoded image projections," *Pattern Analysis and Applications*, vol. 20, no. 1, pp. 127–143, 2017.

[12] K. Wang and Q. Ji, "Real time eye gaze tracking with 3d deformable eye-face model," in *IEEE International Conference on Computer Vision*, 2017, pp. 1003–1011.

[13] A. Kar and P. Corcoran, "A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms," *IEEE Access*, vol. 5, pp. 16 495–16 519, 2017.

[14] T. Schneider, B. Schauerte, and R. Stiefelhagen, "Manifold alignment for person independent appearance-based gaze estimation," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*. IEEE, 2014, pp. 1167–1172.

[15] F. Lu, Y. Sugano, T. Okabe, and Y. Sato, "Inferring human gaze from appearance via adaptive linear regression," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 153–160.

[16] Z. Guo, Q. Zhou, and Z. Liu, "Appearance-based gaze estimation under slight head motion," *Multimedia Tools and Applications*, vol. 76, no. 2, pp. 2203–2222, 2017.

[17] X. Zhang, Y. Sugano, M. Fritz, and 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.

[18] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, and A. Torralba, "Eye tracking for everyone," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2176–2184.

[19] A. Villanueva, V. Ponz, L. Sesma-Sanchez, M. Ariz, S. Porta, and R. Cabeza, "Hybrid method based on topography for robust detection of iris center and eye corners," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 9, no. 4, p. 25, 2013.

[20] E. Wood, T. Baltrušaitis, L.-P. Morency, P. Robinson, and 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.

[21] O. Jesorsky, K. J. Kirchberg, and R. W. Frischholz, "Robust face detection using the hausdorff distance," in *International Conference on Audio-and Video-Based Biometric Person Authentication*. Springer, 2001, pp. 90–95.

[22] F. Timm and E. Barth, "Accurate eye centre localisation by means of gradients." *Visapp*, vol. 11, pp. 125–130, 2011.