Appearance Based Gaze Estimation Using Eye Region Landmarks and Math Approach

Shichao Cheng[1], Bocheng Zhang[1], Jianjun Li[1], Zheng Tang[2], Korhan Cengiz[3]

[1] School of Computer Science and Engineering, Hangzhou Dianzi University, China
[2] Key Laboratory of Data Link Technology of CETC, Xi'An, China
[3] Department of Electrical-Electronics Engineering, Trakya University, Turkey
sccheng@hdu.edu.cn, zbcmars@gmail.com, lijjcan@gmail.com, tangz@gmail.com, korhancengiz@trakya.edu.tr

Abstract

The gaze direction can be defined by the pupil and the center of the eyeball, the latter cannot be observed in the 2D image, which can cause ill-posed problems and cannot achieve highly accurate gaze estimation. Therefore, we try to extract several effective landmarks around the eyeball and iris from the monocular input for gaze estimation. Instead of directly returning the two angles for the pitch and yaw of the eyeball, we return to an intermediate graphical representation, which in turn simplifies the task of 3D gaze estimation. We try to use a novel learning-based method to locate the landmarks of an eyeball with the appearance-based method. Through these high-accuracy feature points, we have proposed a new and effective formula for drawing more accurate gaze directions. As for the individual gaze estimation of independent people, our method is superior to existing model fitting and appearance-based methods.

Keywords: Gaze estimation, Convolutional neural network, Mathematical method

1 Introduction

As an important cue connecting to human behaviors, gaze direction, and gaze changing behaviors both act as necessary tools in many real-world domains, including Human-robot interaction [1-2], Virtual reality [3], human-computer interfaces [4-5] and health care [6]. Therefore, gaze has been an important communication signal which has also been shown to be related to higher-level characteristics such as personality, for which gaze tracking method has been paid much more attention.

To date, most eye trackers like to use off-the-shelf cameras, including those in mobile devices, which can help users with reduced mobility, or can perform crowdsourced visual saliency estimation without having to spend dedicated hardware. While these provide high accuracy, they also place strong constraints on users' head movements. In the meantime, those gaze estimation systems can fail when encountering challenging issues such as low image quality or unusual illumination conditions. In this work, we provide a novel perspective for addressing the problem of gaze estimation from images taken both from Synthetic pictures and from real-world environments.

In general, gaze estimation methods can be divided into two categories: geometry-based methods and appearance-based methods. The basic idea of the geometry-based method is to detect some features of the eye (such as key points such as the corner of the eye and the position of the pupil), and then calculate gaze based on these features. The appearance-based approach is to directly learn a model that maps appearance to gaze. The two types of methods have their own advantages and disadvantages: Geometric methods are relatively more accurate and stable for different domains. However, these methods have high requirements for image quality and resolution. Appearance-based methods have low-resolution and high-noise images. The performance is better, but the training of the model requires a lot of data, and it is easy to domain overfitting. With the rise of deep learning and the publication of a large number of data sets, appearance-based methods have received increasing attention.

Our new method for gaze estimation combines the advantages of those two categories. Since such methods assume the geometry and shape of the eyes, they are sensitive to changes in appearance that are ubiquitous in unconstrained images. Therefore, this method lacks reliable detection of important features in such natural images. We note that the task of eye area landmark detection is similar to the joint detection of human hand and full body posture. Therefore, by displaying a robust eye area landmark detector, we can train on high-quality synthetic eye diagram images, thus providing detailed and accurate location of important landmarks in the eye area (such as the

*Corresponding Author: Jianjun Li; E-mail: lijjcan@gmail.com
DOI: 10.3966/160792642020112106013

eyelid-sclera border) labels, including the limbal area and the corners of the eyes. We use these clear landmarks to train more accurate eye feature point locations. The main advantages of this method are that we integrate multiple robust multi-dimensional networks, retain the characteristics of each dimension, and mix them into a more powerful network to improve accuracy and re-modify our condensed view function. We have shown through experiments that our scheme proposes a different method for predicting the gaze direction than the previous methods, which has obtained significant improvements.

In summary, the main contributions of our work are: (a) improving the accuracy of the tasks of iris positioning and eyelid registration on monocular images, (b) learning robust and accurate landmarks detectors only on synthetic images produced by UnityEyes, and (c) increased gaze estimation accuracies.

2 Related Work

2.1 Appearance-based Gaze Estimation

Mathematical methods rely on eye feature extraction to learn the geometric model of the eye, and then use these features to infer the direction of the line of sight (LOS) [7-9]. However, these methods usually require high-resolution eye diagrams for robust and accurate feature extraction, are prone to noise or lighting interference, and cannot handle well head pose variabilities.

Appearance-based gaze estimation methods directly return the gaze estimation from the image. In principle, given enough training data, they may be able to resolve huge differences in real-world situations. Early works based on the appearance-based approach were limited to a laboratory environment with a fixed head pose [10]. These initial constraints have gradually been relaxed, especially recent datasets collected in daily settings or simulated environments [11].

The increasing scale and complexity of training data to solve gaze estimation leads to K-Nearest Neighbors [12], Support Vector Regression [13], and Random Forests [12]. Recently, the application of deep CNN on this issue has received more attention. The first deep CNNs for gaze estimation was proposed in [14] by Zhang et al, and they provided significant accuracy. To further improve the accuracy, others proposed enhancements, such as employed the information outside the eye region [15], focused on the head-eye relationship [16], and extracted better information from the eye images [17-18]. Person-independent gaze estimation was performed without user calibration and directly applied to areas such as visual attention analysis on unmodified devices, interaction on public displays, and gaze target recognition, although it is based on increasing training requirements and calculation costs.

2.2 Eyeball Landmark Localization

Detecting 3D landmarks from 2D image is one of the most important subjects in computer vision. In particular, there are a lot of studies using deep convolutional neural networks to perform eyeball landmark and skeletal joint detection for human pose estimation tasks [19]. Although face (or eyeball) landmarks and human joint positions are usually occluded, remote dependencies and global context can make up for the lack of information based on local appearance. Therefore, recent works attempted to learn the spatial relationship among joint positions. The stacked multi-scale architecture was simple, while exhibiting low model complexity (a small number of model parameters), its performance has surpassed other latest technologies. The architecture, which is originally developed for pose estimation, has successfully adopted to the eyeball landmark positioning task in the new eyaball landmark positioning challenge.

3 The Proposed Approach

We first predict the coordinates and eyeball radius of the eye feature points through a synthetic network that is supervised by the hourglass network and the multi-dimensional feature retention network. Then, a more accurate gaze direction is estimated by combining the coordinate points and radius into the proposed novel mathematical module.

3.1 Mixed Multi-dimensional Feature Retention Network

Figure 1 shows the framework of the overall model. The first stage is divided into upper and lower parts, and converge into the gaze estimation model in the second part. The first part is to extract feature points of the eye contours.

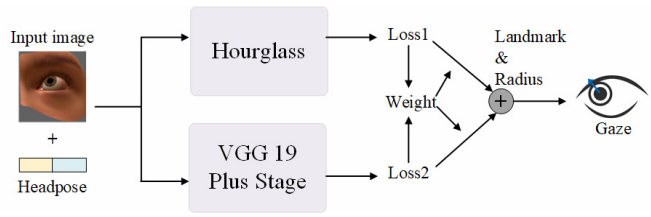


Figure 1. We use monocular images and head pose as input. The eye features are obtained through the feature point extraction network of the upper and lower parts

The e_l and e_u indicate angular errors of upper and lower parts. The g_h and g_r are the ground truth of landmarks and radius, and g'_h and g'_r are the predicted landmarks and radius. We then compute the weighted average of the two-mean-prediction error:

$$e = \lambda_l \cdot e_l + \lambda_u \cdot e_u. \quad (1)$$

To represent the loss of landmarks and radius prediction accuracy for both parts. The weights λ_l and λ_u determine which part should be regarded as the upper part in terms of the mean landmarks estimation error. When $\lambda_l \neq \lambda_u$, the loss becomes asymmetric.

Following the key idea of asymmetric regression, we will increase the weight of that part to optimize the network if one part is more likely to get better performance than the other part. Hence, the weights are calculated as:

$$\frac{\lambda_l}{\lambda_u} = \frac{1/e_l}{1/e_u}, \quad (2)$$

$$\lambda_l + \lambda_u = 1, \quad (3)$$

whose solutions are

$$\lambda_l = \frac{1/e_l}{1/e_l + 1/e_u}, \quad (4)$$

$$\lambda_u = \frac{1/e_r}{1/e_l + 1/e_u}, \quad (5)$$

which encourages the mixed multi-dimensional feature retention network to rely on the superior part while training.

3.2 Modified Hourglass Extraction Network

Figure 2 shows the architecture of the original hourglass model. The hourglass network architecture [20] has previously used in human pose estimation, and the key issue is the landmark occlusion problem. In this case, the appearance of the landmark is no longer useful for accurate positioning, and only prior knowledge can be used. The hourglass architecture attempts to transfer the remote environment by repeatedly improving the overcome solution at multiple scales using hourglass modules. The proposal is designed to capture information at each scale and finally require fusing the whole information. In order to capture the features of the image at multiple scales, we use multiple pipelines to separately process information at different scales, and then combine these features in the back part of the network. The proposal method aims to use skip layers with a single pipeline to save spatial information at each scale.

Figure 2. The original hourglass model

Feature map of the hourglass module manual automatic encoder is reduced by the pooling operation, and then expanded using bilinear interpolation. At each scale level, the corresponding crossover from the other side of the hourglass is calculated and applied by skipping connections and applying the residuals. Therefore, when 64 feature maps are given, the network will be refined multiple times with 4 different image ratios. This repeated dimension reduction and enlargement inference ensure a large effective perceptual field, and even can encode the spatial relationship between landmarks under the occlusion.

In our work, we adapt the original architecture to the task of landmarks detection in eye images. Although the eye diagram contains fewer global structural elements than the global structural elements in pose estimation, there are still significant spatial contexts that can be utilized by large acceptance field models. We use this attribute to detect the center of the eyeball and the occluded iris edge landmarks to achieve logical accuracy, sometimes even in the case of complete occlusion. Most high-order features only appear at lower resolutions, unless at the end of the upsampling. If you supervise after the network is upsampled, you cannot re-evaluate these features in a larger global context. If the goal of the network is to make the best predictions, these predictions should not be made in a local range.

Figure 3 shows the original multi-dimensional feature extraction network. There are 64 refined feature maps which can be combined through a 1×1 convolutional layer to generate 18 heat maps, each of which represents the estimated position of the landmark of a particular eyeball region. Intermediate supervision is performed by calculating the sum of squares of pixel differences on each predicted heat map. The original paper proved that the use of eight hourglass modules with intermediate supervision can significantly improve the accuracy of landmark positioning compared to both the 2-stack and 4-stack models using the same number and size of parameters.

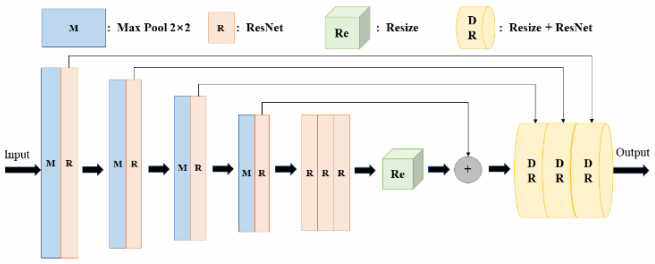


Figure 3. The original multi-dimensional feature extraction network

In our work, we propose a 3 cascaded hourglass network separated by 1 ResNet shown as Figure 4, training on UnityEyes-produced eye images and annotations. We use single eye images (150×90 or 36×60) as input in order to generate 18 heatmaps, including 8 on the marginal zone, 8 on the iris's edge,

2 at the iris center and eyeball center. These coordinates are then passed to the weight calculation module and rounding module.

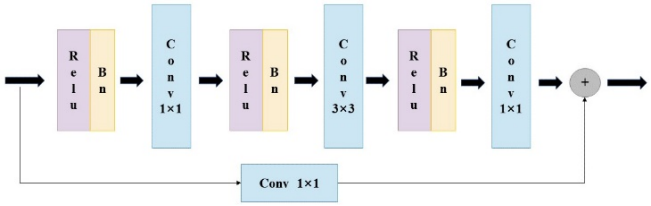


Figure 4. The proposed multi-dimensional feature extraction network

3.3 VGG-19 Plus Feature Extraction Network

In the lower branch of Figure 2, the input picture passes through the classic VGG-19 structure and a repeatedly convolutional network. The output feature-map has 18 layers, each layer represents a heatmap. The feature-map and label are used to calculate the loss at this stage. At the end of the network, the loss of each layer is added up as a total loss for back propagation to achieve intermediate supervision and avoid the disappearance of the gradient. The architecture of the VGG-19 plus feature extraction network is shown as Figure 5.

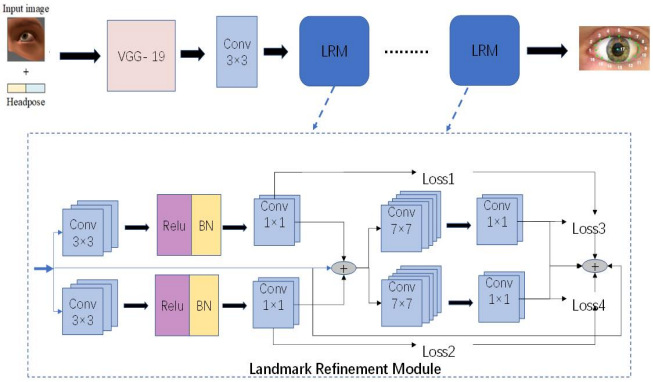


Figure 5. The architecture of VGG-19 plus feature extraction network

The blue LRM block in Figure 5 is a repeated part called landmark refinement module, which can obtain more accuracy 18 landmarks by iteratively adding the output results of the current LRM block into the input of the next LRM block. We obtained a better result by two iterations in our experiments.

3.4 Mathematical Method of Gaze Calculation

In view of the fact that the mathematical calculation methods of the line-of-sight estimation given in the previous papers never considered the offset of the pupil center so that they could not achieve a good result in gaze estimation.

We propose a two-stage method in this paper. In the first stage, the images with an eye in each picture are input and the feature maps can be obtained. Then the feature maps along with eighteen landmarks are fed into the second stage so that the gaze direction can be achieved by mathmatic calculations. Therefore, we reconstruct the gaze estimation task into two specific tasks: (1) process the input monucular image to minimum normalized form (with gaze and headpose), (2) estimate the gaze direction from 18 detected landmarks.

Figure 6 shows the eyeball calculation model. The gaze view of a given input eye image should be visually similar to the input. We assume a simple model of the human eye and iris, where the eyeball is a perfect sphere and the iris is a perfect circle. As for the size of output image is $m \times n$, we assume the projected eyeball diameter $2r = 1.2n$, where r is a long axis diameter. Therefore, ϕ and θ can be calculated as the followings:

$$u_i = m/2 - r' \sin\phi \cos\theta, \tag{6}$$

$$v_i = n/2 - r' \sin\theta. \tag{7}$$

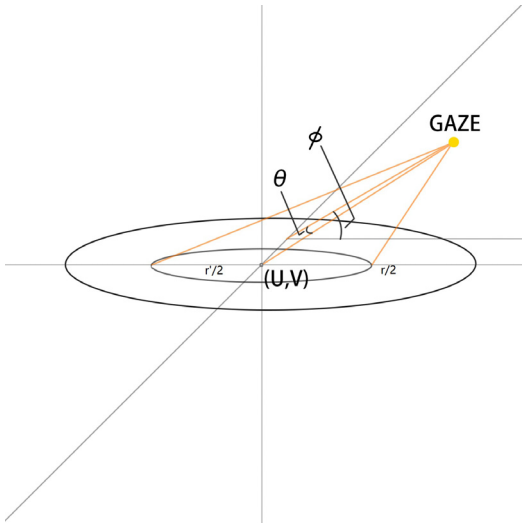


Figure 6. Eyeball calculation model

With ϕ and θ, we can obtain ϕ' and θ' using the following equations:

$$\theta' = \arctan \frac{n/2 - r' \sin\theta - \chi}{m/2 - \sin\phi r' \sin\theta}, \tag{8}$$

$$\theta' = k \cdot \frac{\sin\theta'}{\sin\phi}, \tag{9}$$

where gaze direction $g = (\theta', \phi')$. The iris is drawn as an ellipse, with a long axis diameter r and a short axis diameter $r|\cos\theta\cos\varphi|$. Due to the obstruction of the eyelids and the dim light of the upper part of the eyeball, gaze predictions have always been relatively upward. So we introduce a constant k to correct this situation. Here, k is $5r/12$, based on the positive result from $r/2$ to $r/3$, $5r/12$ is the best choice empirically.

4 Experiments

4.1 Dataset

Since our method is designed to process only eye images and requires head pose information, but does not require the extra eyeball features. We considered the following three public eye gaze data sets for gaze verification and each manually generates eye gaze data set for landmarks verification.

EyeDiap: It contains 94 videos from 16 subjects [21]. Video is divided into three categories: continuous screen (CS) targets, discrete screen (DS) targets, and floating targets (FT). CS videos were used in our experiments, which include static posture recording (subjects maintain approximately the same posture when observing the target) and dynamic posture (MP, subjects perform other important head movements while observing). From these data, We used data from screen targets, involving 14 objects. Sampling 6 sets of VGA video at a ratio of ten frames, a total of about 3600 images. In the test, we took a sample of ten sets of three sets of VGA video for verification (about 2000 images). The truth of the labeled world gaze on the ground is transformed accordingly in the "Head Posture Coordinate System" (HCS).

MPIIgaze: It contains full-face images of 15 subjects (six women, five with glasses). It provides an "evaluation subset" that contains 3,000 randomly selected images for each topic. In this subset, half of the image is flipped horizontally. We use manual annotation of the eye image of the corner of the eye to ensure that the landmark of the eye area is detected. The size of the input eye image of the hourglass network is 150×90. We trained the data of 14 people (the left eye and the right eye each input 1500) and tested one person at random.

UnityEyes: It is effectively infinite in size and is designed to exhibit good variations in iris color, eye region shape, head pose, and illumination conditions. The model contains less than one million model parameters, enough to demonstrate our method and allow real-time implementation (approximately 20 Hz).

4.2 Further Experiments

For the extraction of our feature points, our network model has a significant improvement in accuracy compared to the ELG method. Figure 7 and Figure 8 show the landmarks and radius error in UnityEyes.

The x axis in the Figure 7 represents the prediction error of the eye radius, which calculated by:

$$L_{radius} = \beta \| \tilde{r} - r \|_2^2, \tag{10}$$

where $\beta = 10^{-7}$. While the y axis refers to the number of iterations. The x axis in the Figure 8 represents the prediction error of the eyeball's landmarks, which calculated by:

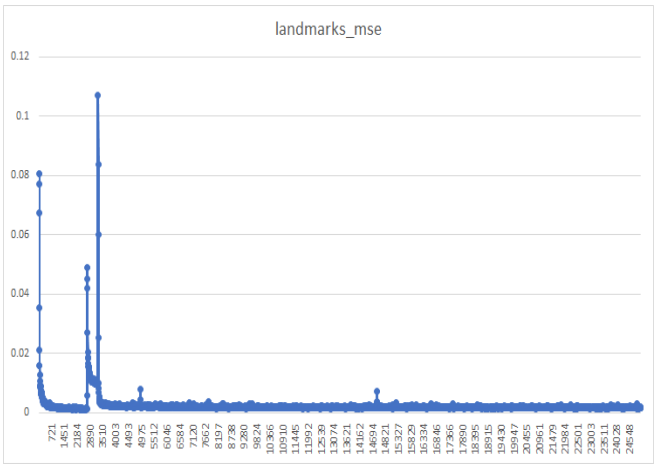


Figure 7. Errors of landmarks

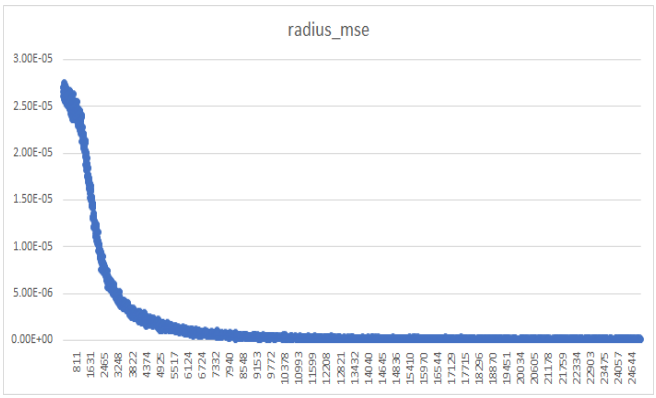


Figure 8. Errors of radius

$$\alpha \sum_{i=1}^{18} \Sigma_p \| \tilde{h}_i(p) - h_i(p) \|_2^2, \tag{11}$$

where h is the confidence at pixel p and $\tilde{h}_i$ is a landmarks predicted by the network. We set the weighting coefficient α = 1 based on experience.

For the EyeDiap and MPIIGaze datasets, we applied the leave-one-out method to exit the protocol. Please note that for these dataset, we train real data and synthetic data, but only test real data. We compared our method with the baseline using a single model [14], KNN [11], RF [11], VGG-16 [18], GazeNet [22], SVR-Ad [15], RT-GENE [21], iTracker [15] and spatial weights CNN [23]. The baseline was defined as the average of the average gaze angular error computed for each fold. These methods use the single eye (which labeled as e) or head pose (which labeled as h) as input. On the MPIIGaze dataset, our proposed network achieves an average angular error of 4.4°, and its performance is 8.3% higher than 4.8° [21]. Figure 9 shows our result in MPIIGaze and Table 1 shows the input of our experiment.

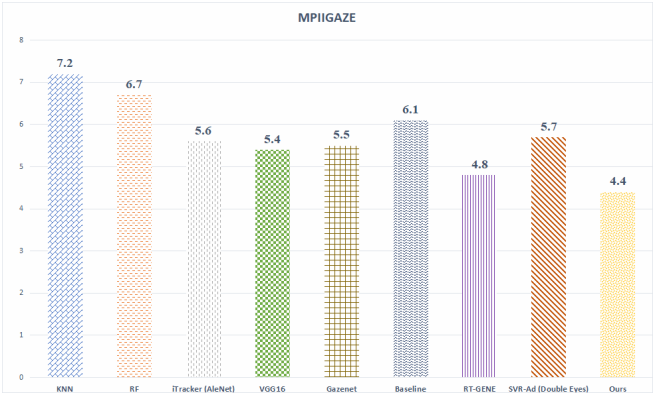


Figure 9. Result in MPIIGaze

Table 1. Inputs of MPIIGaze

Model	Input
KNN	e + h
RF	e + h
iTrack (AlexNet)	E
VGG16	E
GazeNet	e + h
Baseline	N/A
SVR-Ad (Double eyes)	N/A
Ours	e + h

We also compared our method with state-of-the-art approaches on the EyeDiap dataset. Figure 10 shows the gaze estimation results of different methods, and it can be seen that our gaze estimation accuracy reaches 5.9°, which is 1.6% lower than the latest one-eye gaze estimation method [23]. The accuracy improvement is partially due to the use of improved data normalization methods and our proposed gaze decomposition.

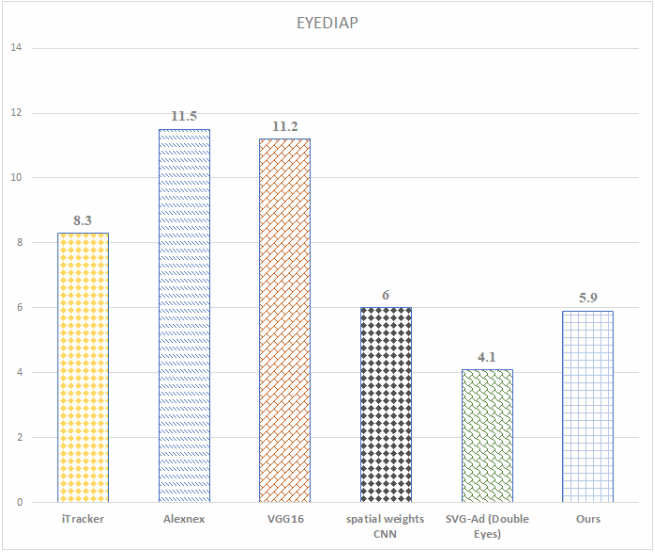


Figure 10. Result in EyeDiap

For the extraction of landmarks points, we introduced different layers in the first part of the network to explore the impact of different dimensions or different numbers on the extraction of landmarks points shows in Figure 11. The first one refers to our original basic network, and then by extracting features of different dimensions, we found that extracting one more layer of network features will increase the accuracy of landmarks by 1.8% shown in Table 2.

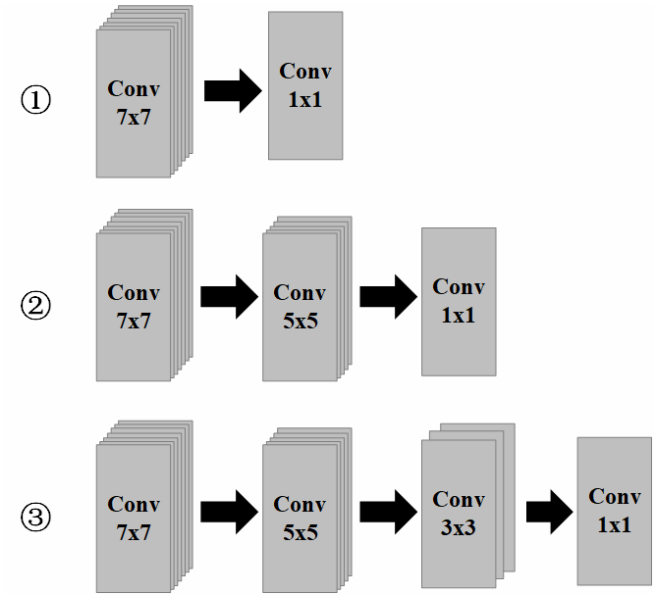


Figure 11. The network model of extracting landmarks

Table 2. Results in MPIIGaze

Landmarks_mse	Radius_mse	Mean angular gaze error (MPIIGaze)
0.0090	4.20e-6	4.48
0.0025	3.00e-6	4.40
0.0027	4.00e-6	4.42

5 Conclusion

This paper aims to improve appearance-based gaze estimation using the subject specific model and novel math method. A two-stage method of gaze estimation has benn proposed efficiently. We considered the offset of the pupil center when using a mathmatic method to calculate the gaze direction. The main contribution of this paper is to propose a Multi-dimensional feature extraction network for predicting gaze landmarks, insteading of gaze directions, to alleviate the impact of factors such as illumination. Experimental results on three public and commonly used data sets prove the effectiveness of this method.

Acknowledgments

This work was supported in part by the National Science Fund of China under Grant 61871170, in part by the National Defense Basic Research Program under Grant JCKY2017210A001, the Natural Science Foundation of Zhejiang Provincial under Grant Q20F020062, and Key Laboratory of Brain Machine Collaborative Intelligence of Zhejiang Province.

References

[1] S. Andrist, X. Z. Tan, M. Gleicher, B. Mutlu, Conversational Gaze Aversion for Humanlike Robots, *ACM/IEEE International Conference on Human-robot Interaction*, Bielefeld, Germany, 2014, pp. 25-32.

[2] C.-M. Huang, B. Mutlu, Anticipatory Robot Control for Efficient Human-Robot Collaboration, *ACM/IEEE International Conference on Human-Robot Interaction*, Christchurch, New Zealand, 2016, pp. 83-90.

[3] T. Pfeiffer, Towards Gaze Interaction in Immersive Virtual Reality: Evaluation of a Monocular Eye Tracking Set-up, *Virtuelle und Erweiterte Realität-Fünfter Workshop der GI-Fachgruppe VR/AR*, Magdeburg, Germany, 2008, pp. 81-92.

[4] Z. Chen, B. E. Shi, Using Variable Dwell Time to Accelerate Gaze-Based Web Browsing with Two-Step Selection, International *Journal of Human-Computer Interaction*, Vol. 35, No. 3, pp. 240-255, March, 2019.

[5] J. Pi, B. E. Shi, Probabilistic Adjustment of Dwell Time for Eye Typing, *International Conference on Human System Interactions*, Ulsan, South Korea, 2017, pp. 251-257.

[6] A. Grillini, D. Ombelet, R. S. Soans, F. W. Cornelissen, Towards Using the Spatio-Temporal Properties of Eye Movements to Classify Visual Field Defects, *ACM Symposium on Eye Tracking Research & Applications*, Warsaw, Poland, 2018, pp. 1-5.

[7] E. Wood, A. Bulling, Eyetab: Model-based Gaze Estimation on Unmodified Tablet Computers, *ACM Symposium on Eye Tracking Research & Applications*, Safety Harbor, Florida, 2014, pp. 3-6.

[8] L. Sun, M. Song, Z. Liu, M.-T. Sun, Real-time Gaze Estimation with Online Calibration, *IEEE MultiMedia*, Vol. 21, No. 4, pp. 28-37, October-December, 2014.

[9] K. Wang, Q. Ji, Real Time Eye Gaze Tracking with Kinect, *IEEE Conference on Pattern Recognition*, Cancún, México, 2016, pp. 2752-2757.

[10] K. H. Tan, D. J. Kriegman, N. Ahuja, Appearance-based Eye Gaze Estimation, *Sixth IEEE Workshop on Applications of Computer Vision*, Orlando, FL, USA, 2002, pp. 191-195.

[11] Y. Sugano, Y. Matsushita, Y. Sato, Learning-by-synthesis for Appearance-Based 3d Gaze Estimation, *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014, pp. 1821-1828.

[12] X. Wang, K. Liu, X. Qian, A Survey on Gaze Estimation, *International Conference on Intelligent Systems and Knowledge Engineering*, Taipei, Taiwan, 2015, pp. 260-267.

[13] T. Schneider, B. Schauerte, R. Stiefelhagen, Manifold Alignment for Person Independent Appearance-Based Gaze Estimation, *International Conference on Pattern Recognition*, Stockholm, Sweden, 2014, pp. 1167-1172.

[14] X. Zhang, Y. Sugano, M. Fritz, A. Bulling, Appearance-based Gaze Estimation in the Wild, *IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 2015, pp. 4511-4520.

[15] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, A. Torralba, Eye Tracking for Everyone, *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 2176-2184.

[16] H. Deng, W. Zhu, Monocular Free-head 3d Gaze Tracking with Deep Learning and Geometry Constraints, *IEEE International Conference on Computer Vision*, Venice, Italy, 2017, pp. 3162-3171.

[17] Z. Chen, B. E. Shi, Appearance-based Gaze Estimation Using Dilated-convolutions, *Asian Conference on Computer Vision*, Perth, Australia, 2018, pp. 309-324.

[18] S. Park, A. Spurr, O. Hilliges, Deep Pictorial Gaze Estimation, *European Conference on Computer Vision*, Munich, Germany, 2018, pp. 721-738.

[19] A. Toshev, C. Szegedy, DeepPose: Human Pose Estimation via Deep Neural Networks, *IEEE International Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014, pp. 1653-1660.

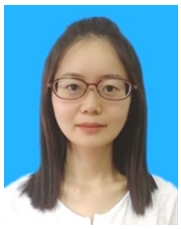
[20] A. Newell, K. Yang, J. Deng, Stacked Hourglass Networks for Human Pose Estimation, *European Conference on Computer Vision*, Amsterdam, The Netherlands, 2016, pp. 483-499.

[21] T. Fischer, H. J. Chang, Y. Demiris, Rt-gene: Real-time Eye Gaze Estimation in Natural Environments, *European Conference on Computer Vision*, Munich, Germany, 2018, pp. 339-357.

[22] X. Zhang, Y. Sugano, M. Fritz, A. Bulling, Mpiigaze: Real-World Dataset and Deep Appearance-Based Gaze Estimation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 41, No. 1, pp. 162-175, January, 2019.

[23] X. Zhang, Y. Sugano, M. Fritz, A. Bulling, It's Written All Over Your Face: Full-Face Appearance-based Gaze Estimation, *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Honolulu, HI, USA, 2017, pp. 2299-2308.

Biographies



Shichao Cheng received the B.E. degree in mathematics and applied mathematics from Henan Normal University. Xinxiang, China, in 2013 and the Ph.D. degree in mathematics from the Dalian University of Technology, Dalian, China, in 2019. She is currently a lecturer in Hangzhou Dianzi University. Her research interests include computer vision, machine learning and optimization.

Bocheng Zhang is an undergraduate in School of Computer Science and Engineering, Hangzhou Dianzi University. His research interests include information security, gaze estimation, image processing, and deep learning.

Jianjun Li received the B.S. degree in information engineering from the Xi'an University of Electronic Science and Technology, Xi'an, China, the M.S. degree in electrical and computer from The University of Western Ontario, Canada, and the Ph.D. degree in electrical and computer from the University of Windsor, Canada. He is currently with Hangzhou Dianzi University, as a Chair Professor. His research interests include micro-electronics, audio, video, and image processing algorithms and implementation.

Zheng Tang received his B.S. degree in electronic engineering and M.S. and Ph.D. degrees in electronic and information in 2008, and all are from North Western Polytechnical University in Xi'An, China. He is currently working in the 20th research institute of China Electronics Technology Group Corporation as a senior engineer. His research interests include information fusion and data link information processing algorithms and implementations.

Korhan Cengiz received the B.S. degr-ee in electronics and communicat-ion en-gineering from Kocaeli University, Ko-caeli, Turkey, in 2008, the M.S. degree in electronics and communica-tion eng-ineering from Namik Kemal University, Tekirdağ, Turkey, in 2011, and the Ph.D. degree in electronics engineering from Kadir Has University, Istanbul, Turkey, in 2016. Since 2013, he has been a Lecturer Doctor with the Electrical-Electronics Engineering Department, Trakya University. His research interests include wireless sensor networks, routing protocols, wireless communications, statistical signal processing, and spatial modulation.