2019-01-01

Accurate eye center localization via hierarchical adaptive convolution

Cai, H

British Machine Vision Conference 2018, BMVC 2018

Accurate Eye Center Localization via Hierarchical Adaptive Convolution

Haibin Cai[1]
haibin.cai@port.ac.uk
Bangli Liu[1]
bangli.liu@port.ac.uk
Zhaojie Ju[1]
zhaojie.ju@port.ac.uk
Serge Thill[23]
serge.thill@plymouth.ac.uk
Tony Belpaeme[24]
tony.belpaeme@plymouth.ac.uk
Bram Vanderborght[5]
bram.vanderborght@vub.ac.be
Honghai Liu[1]
honghai.liu@port.ac.uk

[1] School of Computing
University of Portsmouth,UK
[2] School of Computing, Electronics, and Mathematics
University of Plymouth,UK
[3] School of Informatics
University of Skövde, Sweden
[4] IDLab-imec
Ghent University, Belgium
[5] Faculty of Applied Sciences
Vrije Universiteit Brussel and Flanders Make, Belgium

Abstract

Eye center localization has been an active research topic for decades due to its important biological properties, which indicates human's visual focus of attention. However, accurate eye center localization still remains challenging due to the high degree appearance variation caused by different kinds of viewing angles, illumination conditions, occlusions and head pose. This paper proposes a hierarchical adaptive convolution method (HAC) to localize the eye center accurately while consuming low computational cost. It mainly utilizes the dramatic illumination changes between the iris and sclera. More specifically, novel hierarchical kernels are designed to convolute the eye images and a differential operation is applied on the adjacent convolution results to generate various response maps. The final eye center is localized by searching the maximum response value among the response maps. Experimental results on several publicly available datasets demonstrate that HAC outperforms the start-of-the-art methods by a large margin. The code is made publicly available at https://github.com/myopengit/HAC

1 Introduction

Eye center localization has attracted much attention due to its importance in gaze estimation, virtual reality, human-robot interaction, human-machine interfaces, psychology, and cognitive linguistics [12]. Although accurate eye center location can be obtained through high-quality wearable eye-gaze tracking systems, the unconformable user experience or expensive devices make these methods unattractive [22]. Thus, this paper focuses on localizing eye centers from images captured in practical non-wearable scenarios such as interaction with a desktop, a laptop, a phone or a robot equipped with cameras.

As a similar research with eye center localization, face alignment aims to locate several outstanding facial landmarks such as eye corners, mouth lips, noses and eyebrows. The last five years have seen a great improvement of face alignments techniques [15, 29]. However, the eye center is often excluded from these facial landmarks due to its high degree of appearance variation caused by the occlusion of eyelids, viewing angles, ethnicity, illumination conditions and head poses. For example, Vicente *et al.* [23] suggested using other shape based methods for eye center localization rather than jointly locating it with other landmarks for more accurate performance. Although many eye center localization methods have been proposed and significant improvements have been achieved in the past three decades, fast and accurate eye center localization is still very challenging [11], especially for images with low resolution.

One of the most popular eye center localization methods is the Integral Differential Operator (IDO) [8], which locates the eye center by searching the maximum differential response along with a predefined circle boundary. However, as pointed out in [27], IDO is too computationally intensive to achieve real-time performance. To reduce the computational complexity of IDO, Cai *et al.* [5] proposed a Convolution based Integral Differential Operator (CIDO) by coding the integral operation into kernels which contain different circular boundaries. Although CIDO improves the localization speed to a large extent, the requirement of frontal 2D circular boundaries limits its performance in dealing with different viewing angles. The circle boundary property of the iris is also utilized by many researchers [1, 7, 22] to locate the eye center. The limitation of these methods lies in that they assume a frontal view condition.

To overcome the circular boundary limitation in non-frontal view condition, this paper proposes to model the 3D viewing pose into the designed kernels. The main contributions of this paper are listed as follows:

1) Novel hierarchical kernels are constructed according to different viewing angles and they are adaptively selected according to the obtained 3D head pose in the localization stage. The designed kernel enables the algorithm to effectively deal with the situations when the boundary of the iris is not circular.

2) The design of the hierarchical kernels and convolutional framework greatly improves the eye center localization accuracy. Experimental results on two of the most frequently used publicly available datasets demonstrate that HAC outperforms the state-of-the-art eye center localization methods by a large margin.

The remainder of this paper is organized as follows. Section 2 briefly reviews the related work of eye center localization. The proposed HAC is explained in detail in Section 3. The experimental results are presented in Section 4 . Finally, Section 5 concludes this paper.

2 Related Work

Eye center localization has been an area of active research for many decades. Recent years have seen great improvements in both the localization accuracy and computational cost. Generally speaking, the existing eye center localization methods can be classified into shape-based and appearance-based according to whether a learning progress is involved.

2.1 Shape-Based Methods

Shape-based methods rely on prior eye shapes information such as rotation invariant points, edges, or filtered responses for the eye center localization. Daugman [8] proposed IDO which makes use of the large intensity change between the iris and sclera. IDO achieves good performance when the captured eye image has a high resolution and the iris's boundary is near circular [4]. However, IDO suffers from computation cost too high to be applied into real-time eye tracking applications as mentioned in [27] and requires high-resolution with a minimum of 50 pixels in iris radius [13]. Cai *et al.* [5] proposed a convolutional variation of IDO method to reduce the computational complexity of IDO for real-time eye center localization. Timm *et al.* [21] proposed to model the radial symmetry information as a form of means of gradient and developed a simple yet efficient eye center localization method by calculating the dot products of the gradient for each point in the image. George *et al.* [10] proposed a fast eye center localization method by convoluting a series of Hough transform kernels with the eye image, which can be seen as a convolution version of the classic Hough transform based eye center localization method [26]. Asadifard *et al.* [2] calculated the histogram cumulative density function of the eye region, followed by a minimum intensity pixels' filter to locate the eye center. Valenti *et al.* [22] proposed to locate the eye center by using a voting strategy where each pixel of the eye image has the ability to vote a potential center according to the gradient information. They further constructed a pyramid by resizing the images and linearly summing the response maps of the pyramid images to determine the final eye center position. Skodras *et al.* [20] applied the fast radial symmetry transform method [18] on a constructed eye region color map for eye center localization. Due to the model restriction, most of the shape-based eye center localization methods are not well adapted to extreme low-resolution images.

2.2 Appearance-Based Methods

Unlike the shape-based methods, appearance-based methods take the entire eye image as an input and try to learn a mapping function for the target position. Jesorsky *et al.* [14] presented a method of detecting faces in an image using Hausdorff distance as a similarity measure and then localized the pupil by a multilayer perceptron trained with pupil centered images. Kroon *et al.* [17] located the eye position by searching the maximum response of a trained fisher linear discriminant classifier. Kim *et al.* [16] proposed an eye localization method based on multi-scale Gabor feature vectors. Two support vector machines trained on properly selected Haar wavelet coefficients were used to localize the eye position in [6]. Valenti *et al.* [22] added scale invariance to their former isophote based method using a scale space pyramid and finally determined the eye location by matching the SIFT vector of each candidate with a database. Markuvs *et al.* [19] proposed to train an ensemble of randomized regression trees for the eye center localization. Wu *et al.* [28] utilized a Deep Boltzmann Machine to learn the eye features and trained a Neutral Network to detect the eye centers. Encouraged by the great success of supervised descent method [29] in face alignment, Gou *et al.* [11] proposed to employ a cascaded regression framework to jointly detect the eye center and eye state. The state of the eye can also provide useful information for the potential eye center localization. Compared to model-based methods, the learning-based methods might achieve better performance in low-resolution images and in the situations where the eyes are near fully closed due to the modality in the training data. On the other hand, appearance-based methods require a large amount of data for training and their performance is largely related to the training data.

3 Method

This paper proposes a hierarchical adaptive convolution method (HAC) to localize the eye center accurately while consuming low computational cost. The following subsections firstly introduce the theory of IDO and then present HAC which improves the accuracy of IDO by overcoming the circular boundary limitation.

3.1 Integro-Differential Operator

Observing the tremendous illumination changes along the iris and scalar, IDO [8] located the eye center via searching the biggest differential radius of the mean value along the circles. The mathematical definition of IDO is as follows:

$$max_{(r,x_0,y_0)}\left|G_\sigma(r)*\frac{\partial}{\partial r}\oint_{r,x_0,y_0}\frac{I(x,y)}{2\pi r}\mathrm{d}s\right| \tag{1}$$

where the $G_\sigma(r)$ represents for a smoothing function. $I(x,y)$ means the extracted eye images. The integral operation $\oint_{r,x_0,y_0}$ is calculated by averaging the pixels around the contour which is constructed by a circle with radius r and center of (x_0,y_0).

3.2 Hierarchical Adaptive Convolution

Originally designed for iris recognition, IDO has an assumption that the iris always has a circular boundary since it requires the users to frontally place their eye towards a camera at a short distance. However, in human-machine interaction scenarios, this assumption is not always true due to different viewing angles and head poses. Besides the frontal view assumption, IDO also suffers from high computational cost as mentioned in [27]. Recently, some researchers [5, 7] proposed to adapt IDO to human-machine interaction scenarios by reducing the computational load or adjusting the energy function. However, the assumption of circular boundary still exists in the constructed models, which greatly affects the localization accuracy. This paper proposes to remove this assumption by modeling the viewing angle property into the designed hierarchical kernels.

3.2.1 Framework Description

Fig. 1 shows the framework of HAC. The input of the algorithm is an image and the output is the localized eye centers. The top left quarter of the figure shows an illustration of the pre-constructed hierarchical kernels, which will be introduced in detail in Section 3.2.2. The input of the algorithm is an image located in the bottom left corner of the image. Once an image is captured, the classic cascade face detector [25] is employed for face detection. Then the supervised descent method [29] is used to detect the face landmarks. Based on the localized eye region landmarks, a rough eye region can be extracted. The detected eye region is then normalized to a 50×50 eye patch image. It should be noted that other size of eye patch should possibly result in similar performance and HAC is robust to different initial eye regions as long as the eye is within the patch. By accurately detecting and normalizing

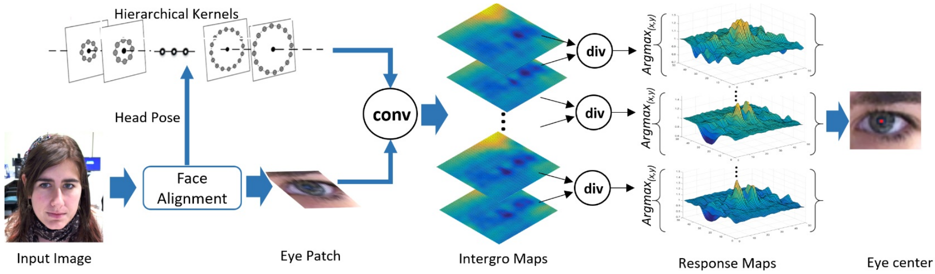

Figure 1: Framework of hierarchical adaptive convolution method.

the eye region, we can set a fixed range for the eye radius. The obtained facial landmark positions can also be used to calculate the 3D head pose information by solving the classic PnP problem. In our case, we utilize the POSIT algorithm [9] with a pre-defined 3D face model to estimate the person's head pose. The estimated head pose is used to adaptively select the nearest viewing hierarchical kernels. The final hierarchical kernels consist of both the selected hierarchical kernels and the frontal viewing hierarchical kernels. Then, a hierarchical of integral maps shown in the middle part of the figure can be obtained by convoluting the normalized eye region images with the final hierarchical kernels. The convolution operation simulates the integral operation as proposed in IDO, which improves the executing speed by a large margin. After using an element-wise division operation, we can obtain different response maps as shown in the right part of the figure. The final eye center can be localized by searching the maximum response pixel in the response maps.

Inspired by [11] which separates the open status and close status of the eye to improve the localization performance, this paper proposes a simple yet effective way to check if the eye is fully closed. The height of an eye is measured by using the distance of the eyelids and the length of the eye can be determined using the two eye corners. If the ratio of the height and the length of the eye is smaller than 0.08, the status of the eye is regarded as fully closed and the eye center is calculated by using the average of four landmarks alongside the eyelids.

3.2.2 Hierarchical Kernels

This section introduces the design of different hierarchical kernels, which model the 3D viewing pose to improve the localization accuracy. As the first step, a 2D circular boundary is constructed using the following equation:

$$\begin{cases} T_r(x,y) = \begin{cases} 1, & \text{if } (x,y) = (r\cos(\theta)+r, r\sin(\theta)+r) \\ 0, & \text{otherwise} \end{cases} \\ \theta \varepsilon [-t,t] \cup [180^0 - t, 180^0 + t] \\ K_r = \frac{T_r}{\sum_{x=0}^{2r+1} \sum_{y=0}^{2r+1} T_r} \end{cases} \tag{2}$$

where r represents the radius. The size of the kernel is $2r+1$. (x,y) is the location of a pixel inside the kernel. The value of a specific pixel $T_r(x,y)$ depends on its location. For example, the pixel's value will be assigned to 1 if its location is alongside the circle boundary. θ is the sampling angle which is limited to the right and left part of the circle to cope with the obscure of eyelids as with IDO. t is the range of the sampling angle. The sampling interval $\Delta\theta$ is set to be small enough to locate every pixel around the selected part of the circular boundary. Finally, the selected pixels along the circular are normalized to a unit value.

By assigning a zero depth value to the constructed 2D circular boundary, we can obtain a 3D circular boundary. The pixels along the 3D circular boundary are then rotated from three different axis to handle the different viewing angles. For each axis, we set the rotate range from -45^o to 45^o with an interval of 15^o. In total, there are 343 viewing angles including the frontal viewing angle. By multiplying the 3D circular boundary with the constructed rotation matrix and projecting it back to the frontal view, the designed kernels can cover a wide range of viewing angles.

Since each eye patch is normalized to a fixed size of 50×50, the range of the radius in the designed kernels can also be fixed. In practical, we set the range of the radius from 9 to 12 and found that it is sufficient to cover the correct radius for different people with different distance. Thus, each set of hierarchical kernels consists of 4 convolution kernels with different radii. It should be noted that the construction of these kernels are calculated ahead and thus the computation load is not increased at the localization stage.

3.2.3 Convolution and Differential Operation

Once the kernels are constructed, they can be used to convolute the eye patches. As mentioned ahead, we selected two sets of hierarchical kernels to convolute the eye patches. One is the nearest viewing hierarchical kernels and the other is the frontal viewing hierarchical kernels. The employment of frontal viewing kernels is for the covering of situations where the person is still viewing frontal direction with his head turned away. For each set of hierarchical kernels, the convolution operation and the differential operation are conducted according to the following equation:

$$\begin{cases} M_r = K_r \otimes I_e \\ D_r = div(M_{r+1}, M_r) \\ argmax_{(r,x_0,y_0)}(D_r) \\ r\varepsilon[r_{min}, r_{max}] \end{cases} \tag{3}$$

where I_e represents for the input eye images and K_r is the designed kernels when the radius is r. The radius has a searching range from r_{min} to r_{max}. The convolution operation is denoted as $\otimes$ and the convolution result M_r is named as the integral map. The element-wise division of the integral map is the differential map D_r. The localized eye center can be found by searching the maximum response in the two sets of response maps.

4 Experimental results

This section describes the details of the evaluation procedure and the analysis of experimental results. Experimental evaluation of HAC has been conducted on two of the most popular datasets for eye center localization including the BioID database [14] and the GI4E database [24]. During the testing, we use the same set of model parameters and accuracy measure equation for all the databases. We uploaded the localization results together with the code of HAC in the aforementioned link.

4.1 Databases

The BioID database has 1521 grayscale images of 23 different people with a resolution of 384 * 288 pixels. Eye center localization in this dataset is considered as challenging due to the low resolution of eye images, different viewing conditions, strong glints and various eye statuses. The dataset also contains some fully closed eye images where it is even impossible for human beings to point out the actual eye center. In some images, the eyes are completely hidden by reflections on the glasses. The face detection rate on this dataset is around 96.5%.

The GI4E database contains 1236 images of 103 different subjects with a resolution of 800*600 pixels. Each subject has 12 images corresponding to different gaze points in the screen. Due to the relative high resolution, this dataset is considered as a normal desktop or laptop setup nowadays. The face detection rate of this dataset is around 97.4%.

4.2 Accuracy measurement

The accuracy measurement of eye location is calculated in normalized error which records the maximum error of both eye points. It is introduced by Jesorsky *et al.* [14] and is defined as follows:

$$e = \frac{\max(d_l, d_r)}{d} \tag{4}$$

where d_l is the Euclidean distance between the detected left eye center and the one in the ground truth. d_r is the corresponding Euclidean distance for the right eye. d is the Euclidean distance between the left and right eyes in the ground truth. The normalized error of $e \leq 0.05$ means the localization result should be within the length of pupil to the ground truth.

4.3 Qualitative results

Table 1 shows a comparison of maximum normalized error in the BioID database. Considering the low-resolution property of the images in this dataset, the normalized error of $e \leq 0.05$ means that the Euclidean distance between the detected eye center and labeled eye center distance should be within 2 to 3 pixels. The proposed method achieved 92.8% accuracy which outperforms the state-of-the-art methods by 1.6%. The boost of performance is owing to the utilization of the easily detected illumination changes between the iris and sclera and the design of the hierarchical adaptive convolution framework. The importance of iris circulary boundary in eye center localization has been extensively studied in [5, 8, 21, 22]. The new results in this paper show that better performance can still be achieved via carefully designing the convolutional kernels to remove the circular boundary assumption.

Table 1: Comparison of maximum normalized error in the BioID database.

Method	$e \leq 0.05$
Timm2011 [21]	82.5%
Valenti2012 [22]	86.1%
Markus2014 [19]	89.9%
George2016 [10]	85.1%
Daugman1993 [8]	80.3%
Cai2017 [5]	86.8%
Gou2017 [11]	91.2%
Proposed HAC	**92.8%**

Table 2 shows a comparison of maximum normalized error in the GI4E database. Compared to the BioID database, its resolution is higher and most of the methods achieved above 90% accuracy with the normalized error of $e \leq 0.05$. The proposed method has achieved 99.5% accuracy which outperforms the highest reported accuracy by 1%. In this database, the normalized error of $e \leq 0.025$ corresponds to the 2 to 3 pixels distances error. HAC achieved 86.4% accuracy which outperforms existing methods by a large margin. Considering that the dataset is constructed for the task of gaze estimation, we argue that the normalized error of $e \leq 0.025$ is a better evaluation criteria since even one pixel's deviation might result in several degrees gaze angle deviation.

Table 2: Comparison of maximum normalized error in the GI4E database.

Method	$e \leq 0.025$	$e \leq 0.05$
Timm2011 [21]	–	92.4%
George2016 [10]	69.1%*	89.3%
Baek2013 [3]	57.4%*	81.4%
Villanueva2013 [24]	–	93.9%
Gou2017 [11]	–	94.2%
Skodras2015 [20]	–	98.5%
Proposed HAC	**85.7%**	**99.5%**

* means data is estimated from the curve in [10, 11].

By combining Table 1 and Table 2, we can see that the cascade based regressor proposed by Geo *et al.* [11] achieved the relatively high accuracy on the low-resolution dataset, however its performance on the high-resolution dataset drops dramatically. The decrease of the performance might due to its over-fitting in the low-resolution images. On the other hand, HAC achieves stable and leading performance on both low-resolution images and high-resolution images.

Fig. 2 shows some snapshots of accurately localized eye center in the two databases. The green dots and red dots represent the ground truth and the localized eye center respectively. In the situation where the localized eye center is exactly same with the ground truth, we can only observe the red dots. As we can see from the figure that although the eyes are partly occluded by the eyelids and glasses, HAC can still locate the eye centers accurately. However, for those images where the faces are wrongly detected or there are strong glints that occlude most of the eyes, HAC will not be able to locate accurately. Fig. 3 shows some wrongly detected eye centers. The firstly line of the image shows some wrong judged cases due to the labeling errors of the dataset. The second line shows some localization errors due to the strong glints of the glasses, the closure of the eyes and wrongly detected faces.

The designed novel hierarchical kernels not only improve the localization accuracy to a large extend, they also greatly reduce the computational cost compared to IDO. To evaluate the computational performance, HAC is implemented in C++ in the aforementioned code link. The average processing time to locate the two eye centers is around 1ms on a laptop equipped with Intel Core i7-8550U CPU.

5 Conclusion

This paper proposed a hierarchical adaptive convolution method to localize the eye center accurately and quickly. Novel hierarchical kernels which model different viewing angles

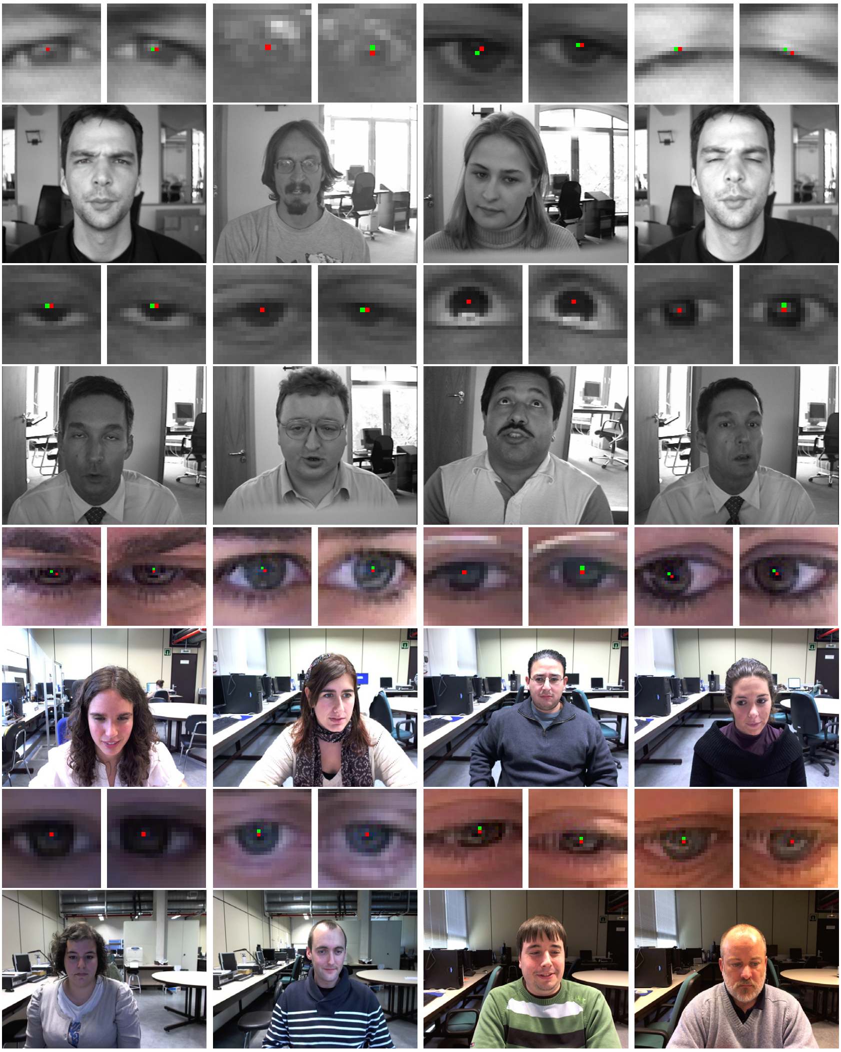

Figure 2: Snapshots of the successful localization results.

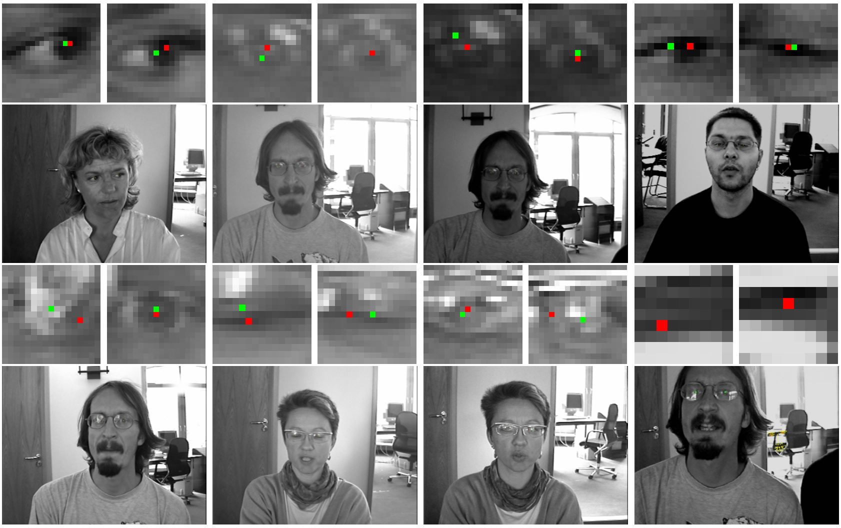

Figure 3: Snapshots of the failed localization results. For the BioID dataset, the localization result considered to be failed if the normalized error is bigger than 0.05.

of the iris were proposed to improve the localization accuracy. The convolution operation greatly reduces the computational cost and enables the algorithm to be integrated into real-time applications. The high accuracy and low combinational cost property make HAC an ideal solution for eye center localization in the common human-machine interaction scenario. HAC was shown to achieve a large performance improvement on two most commonly used eye center localization datasets, covering both low-resolution conditions and high-resolution lab environments. Future direction will focus on exploring its performance in more practical human-robot interaction scenarios.

6 Acknowledgements

This work was supported by EU seventh framework programme DREAM (No. 611391) (corresponding author: H. Liu).

References

[1] Gabriel M. Araujo, Felipe M.L. Ribeiro, Waldir S.S. Júnior, E. A.B. Da Silva, and Siome K. Goldenstein. Weak classifier for density estimation in eye localization and tracking. *IEEE Transactions on Image Processing*, 26(7):3410–3424, 2017.

[2] Mansour Asadifard and Jamshid Shanbezadeh. Automatic Adaptive Center of Pupil Detection Using Face Detection and CDF Analysis. In *Proc. Int. MultiConf. of Engineers and Computer Scientists*, 2010.

[3] Seung Jin Baek, Kang A. Choi, Chunfei Ma, Young Hyun Kim, and Sung Jea Ko. Eyeball model-based iris center localization for visible image-based eye-gaze tracking systems. *IEEE Trans. on Consumer Electronics*, 59(2):415–421, 2013.

[4] Kevin W. Bowyer, Karen Hollingsworth, and Patrick J. Flynn. Image understanding for iris biometrics: A survey. *Computer Vision and Image Understanding*, 110(2): 281–307, 2008.

[5] Haibin Cai, Bangli Liu, Jianhua Zhang, Shengyong Chen, and Honghai Liu. Visual focus of attention estimation using eye center localization. *IEEE Systems Journal*, 11 (3):1320–1325, 2017.

[6] Paola Campadelli, Raffaella Lanzarotti, and Giuseppe Lipori. Precise Eye Localization through a General to specific Model Definition. In *Proc. British Machine Vision Conf.*, 2006.

[7] Yiu-ming Cheung and Qinmu Peng. Eye Gaze Tracking With a Web Camera in a Desktop Environment. *IEEE Trans. Human Machine Systems*, 45(4):419–430, 2015.

[8] John Daugman. High Confidence Visual Recognition of Persons by a Test of Statistical Independence. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(11):1148 – 1161, 1993.

[9] Daniel F Dementhon and Larry S Davis. Model-Based Object Pose in 25 Lines of Code. *International Journal of Computer Vision*, 15(1-2):123–141, 1995.

[10] A George and A Routray. Fast and accurate algorithm for eye localisation for gaze tracking in low-resolution images. *IET Computer Vision*, 10(7):660–669, 2016.

[11] Chao Gou, Yue Wu, Kang Wang, Kunfeng Wang, Fei Yue Wang, and Qiang Ji. A joint cascaded framework for simultaneous eye detection and eye state estimation. *Pattern Recognition*, 67:23–31, 2017.

[12] Dan Witzner Hansen and Qiang Ji. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 32(3):478–500, 2010.

[13] Farmanullah Jan, Imran Usman, and Shahrukh Agha. A non-circular iris localization algorithm using image projection function and gray level statistics. *Optik*, 124(18): 3187–3193, 2013.

[14] Oliver Jesorsky, Klaus J Kirchberg, and Robert W Frischholz. Robust Face Detection Using the Hausdorff Distance. In *Proc. Int. Conf. Audio- and Video-based Biometric Person Authentication*, 2001.

[15] Amin Jourabloo and Xiaoming Liu. Large-Pose Face Alignment via CNN-Based Dense 3D Model Fitting. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2016.

[16] Sanghoon Kim, Sun-Tae Chung, Souhwan Jung, Dusik Oh, Jaemin Kim, and Seongwon Cho. Multi-Scale Gabor Feature Based Eye Localization. *World Academy of Science, Engineering and Technology*, 19:483–487, 2007.

[17] Bart Kroon, Alan Hanjalic, and Sander M.P. Maas. Eye localization for face matching: Is It Always Useful and Under What Conditions? *Proc. Int. Conf. Content-based Image and Video Retrieval*, 2008.

[18] Gareth Loy and Alexander Zelinsky. Fast radial symmetry for detecting points of interest. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(8):959–973, 2003.

[19] Nenad Markuš, Miroslav Frljak, Igor S. Pandžić, Jörgen Ahlberg, and Robert Forchheimer. Eye pupil localization with an ensemble of randomized trees. *Pattern Recognition*, 47(2):578–587, 2014.

[20] Evangelos Skodras and Nikos Fakotakis. Precise localization of eye centers in low resolution color images. *Image and Vision Computing*, 36:51–60, 2015.

[21] Fabian Timm and Erhardt Barth. Accurate eye centre localisation by means of gradients. In *Proc. Computer Vision Theory and Applications*, 2011.

[22] Roberto Valenti and Theo Gevers. Accurate eye center location through invariant isocentric patterns. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 34(9): 1785–1798, 2012.

[23] Francisco Vicente, Zehua Huang, Xuehan Xiong, Fernando la Torre, Wende Zhang, and Dan Levi. Driver Gaze Tracking and Eyes Off the Road Detection System. *IEEE Trans. Intelligent Transportation Systems*, 16(4):2014–2027, 2015.

[24] Arantxa Villanueva, Victoria Ponz, Laura Sesma-Sanchez, Mikel Ariz, Sonia Porta, and Rafael Cabeza. Hybrid method based on topography for robust detection of iris center and eye corners. *ACM Trans. Multimedia Computing, Communications and Applications*, 9(4):25, 2013.

[25] Paul Viola and Michael J Jones. Robust real-time face detection. *Int. Journal of Computer Vision*, 57(2):137–154, 2004.

[26] Richard P Wildes. Iris recognition: an emerging biometric technology. In *Proc. of the IEEE*, pages 1348—-1363, 1997.

[27] Erroll Wood and A Bulling. Eyetab: Model-based gaze estimation on unmodified tablet computers. In *Proc. Symp. Eye Tracking Research and Applications*, pages 3–6, 2014.

[28] Yue Wu and Qiang Ji. Learning the deep features for eye detection in uncontrolled conditions. In *Proc. Int. Conf. Pattern Recognition*, 2014.

[29] Xuehan Xiong and Fernando De La Torre. Supervised descent method and its applications to face alignment. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2013.