2018 IEEE International Conference on Systems, Man, and Cybernetics

# Robust pupil segmentation and center detection from visible light images using convolutional neural network

Kazunari Kitazumi
*Graduate School of Infomatics*
*Kyoto University*
Kyoto, Japan
kitazumi@ii.ist.i.kyoto-u.ac.jp

Atsushi Nakazawa
*Graduate School of Infomatics*
*Kyoto University*
Kyoto, Japan
nakazawa.atsushi@i.kyoto-u.ac.jp

***Abstract*—In this paper, we present a robust pupil detection method from visible light (VL) images using a convolutional neural network (CNN). In contrast to existing pupil detection algorithms, our method does not require infrared (IR) illuminations and cameras, and robustly works even for the images of dark-brown irises (black eyes) and/or images including strong corneal reflections such as outdoor scenes. Thus it has potential application scenarios. Our method first detects an eye region from an input image and applies CNN-based pupil segmentation which has a composition and decomposition structure. To learn the relationship between visible eye images and pupil segments, we construct two datasets and augmentation algorithms. One dataset is based on an existing eye image dataset (UBIRIS.v2), and the other consists of images taken by ourselves by using a corneal imaging camera that can take eye images in mobile environments. Applying color and corneal reflection augmentations to these image datasets and using them for learning the CNN, we built a robust pupil segmentation neural network. The performance is evaluated in three ways. First, we evaluate the segmentation accuracy. Second, we evaluate the pupil center detection accuracy using GI4E facial image sets. Third, we developed an eye gaze tracking (EGT) algorithm that uses the pupil detection and evaluated its accuracy. From the result, the proposed method detects pupil centers more accuracy than the state-of-the arts, and shows similar EGT accuracy to the commercial systems that use IR-active lighting setups.**

***Index Terms*—pupil detection, segmentation, eye gaze tracking, convolutional neural net, visible light**

## I. Introduction

The human pupil is a hole located at the center of the iris of the eye. We observe our surroundings by accepting the light through the pupil to the retina, thus human visual information, such as eye gaze or pupillary responses, is obtained by observing the pupil. For example, eye gaze tracking (EGT) systems are commonly used in many applications. They mainly use the pupil positions for computing gaze directions. Another important application is biometrics since the pupil size (pupillary response) reflects internal states such as interest, concentration, and habituation [1].

These applications use infrared (IR) illuminations and cameras to obtain imaged pupils since IR illuminations clearly distinguish the pupil boundary as darker (dark pupil) or brighter (bright pupil) image segments, as shown in the existing literature [2]–[4]. However, the requirement for IR cameras and illuminations has several drawbacks such as additional cost and installation space, battery use, noisy results under sunlight conditions, and eye safety problems.

To solve these problems, we present a method to robustly detect pupils from visible light (VL) images using a convolutional neural network (CNN). In contrast to existing pupil detection algorithms, our method does not require infrared (IR) illuminations and cameras and robustly works even for the images of dark-brown irises (black eyes) and/or images including strong corneal reflections such as outdoor scenes. It thus has many potential application scenarios.

Several existing studies have been conducted for similar tasks. Morita et al. [5] developed a pupil detection technique using the images under visible light conditions. It first detects iris edges and then finds the pupil boundary by using Hough transform. However, this method assumes close-up views of eyes that are not contaminated by the corneal reflections. In addition, they experimentally validated the method by using only their original image sets.

In the iris segmentation literature, methods using CNNs have been studied and have demonstrated significant performances [6], [7]. Iris segmentation complements pupil detection, thus we also use a CNN for pupil segmentation/center detection. However, we newly introduce image augmentation algorithms that simulate the corneal reflections and then demonstrate their effectiveness in experiments. In addition, we also show that the same network, not only its network structure but also the learned weights, can be used to both segment pupils and detect pupil centers by using different open datasets. Our main contributions with this paper are as follows.

1) We show that our proposed method robustly segments pupils by using a CNN. In contrast to existing approaches, our method can detect an iris that has black (meramin) pigment the boundary of which even humans cannot recognize. As far as we know, this is the first work to use a CNN for pupil segmentation.

2577-1655/18/$31.00 ©2018 IEEE
DOI 10.1109/SMC.2018.00154

2) We show the same network can be used for pupil segmentation and center detection tasks. The proposed algorithm detects the pupil center better than state-of-the-art methods.
3) We build eye image and pupil segmentation datasets. One is based on an existing dataset (UBIRIS.v2), and another consists of images taken by our corneal imaging camera in outdoor environments.
4) Our corneal reflection augmentation makes the pupil detection more robust in the outdoor environment.
5) We built an EGT system that uses the proposed pupil detection method. Experimental results revealed the proposed system tracks eye gaze within 1 degree accuracy, similar to commercial EGT systems that use IR lighting/cameras.

## II. Approach

Our pupil segmentation and center detection (pupil detection) algorithm is shown in Figure 1. When the input image is a close-up view of the eye like the images in the UBIRIS.v2 dataset, we directly input the image to the segmentation network (Figure 1(a)). When the input image is a facial image such as images in the Gi4E dataset, we first find the eye regions from the entire image by using dlib [8] and input the eye regions to the pupil segmentation network (Figure 1(b)).

For the pupil segmentation network, we use the five-layer U-Net architecture [9], which consists of composition and decomposition architecture as well as short-cut connections between the same layers. The pupil segmentation network is learned by using the pupil segment databases. Here, we developed several augmentation techniques that simulate the variation of camera color sensitivities, iris color, and corneal reflections of the scenes. We describe the image datasets, data augmentations and network structure below.

### A. Eye images and pupil segment datasets

To learn and evaluate the pupil segmentation network, we prepared three eye images datasets captured under visible lighting setups and manually annotated ground-truth pupil segmentation.

**UBIRIS.v2 [10]:** The first dataset is UBIRIS.v2, which includes 11,102 iris images of 261 subjects. The images are taken using a commercial single-lens reflex camera (SLR, Canon EOS 5D), and image sizes are (H,V) = (400,300) [pixels] as shown in Figure 3(a). We compute the ground-truth pupil segments from the manually labeled iris segmentation datasets developed by Hofbauer et al. Specifically, we find a circular hole from a binalized iris mask image and assume it to be the pupil region. When a pupil is not completely surrounded by the iris due to the eyelids partly covering it, we manually obtain the pupil mask. Simultaneously, we prepare the iris boundary mask for the latter corneal reflection augmentations from this dataset. We prepare 2,250 eye images as well as their pupil/iris masks from 104 subjects.

**GI4E database [11]:** GI4E is a facial image dataset consisting of 1339 images acquired with a standard webcam,

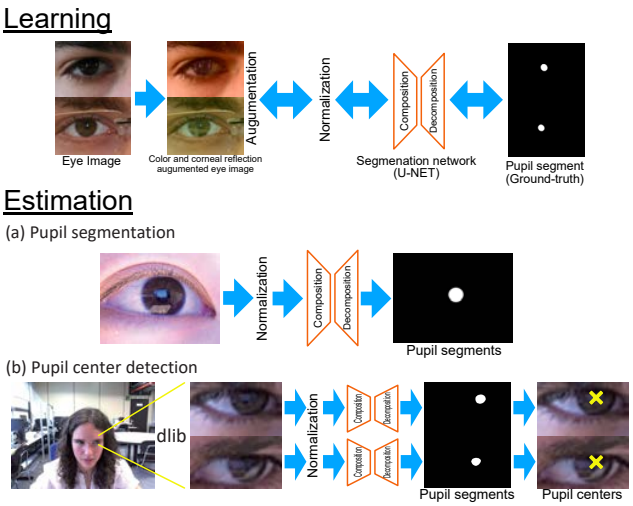


Fig. 1. Overview of our pupil segmentation and pupil center detection (pupil detection) algorithm. (a) When the input image is the close-up view of the eye (e.g. UBIRIS v2 dataset), the image is directly input to the pupil segmentation network. (b) When the input image is an entire facial image (e.g., images in the Gi4E dataset), the eye regions are cropped and put into the segmentation network separately. We can obtain pupil segments or pupil centers using the outputs.

corresponding to 13 images each for 103 different subjects. Every set of 13 images consists of 12 images in which the user gazes at different points on the screen and another image in which the users eyes are closed. GI4E does not provide pupil/iris segment information but includes manually annotated eye corners and pupil centers of both eyes. We used this database for evaluating the accuracy of pupil center detections.

**Outdoor eye image datasets:** We also prepare our original dataset consisting of images taken from about 10 cm away from eyes by using our corneal imaging camera, which is designed to capture the eye images as well as the corneal reflections in visible lighting conditions (Fig. 2) [12]. The camera module is mounted on the head coordinate and can take images of high-definition (HD) resolution (1920×1080 pixels) in 30 Hz. While images were taken, subjects walked around an outdoor environment, thus the eye images include many scene reflections at the corneal surface. We collected 148 eye images as well as manually annotated pupil segments from 11 subjects.

### B. Data Augmentations

Data augmentations for the eye images are needed to increase the accuracy of pupil detection. This is because of 1) the color variations of the human iris, 2) color sensitivity of the cameras, and 3) the reflections of environmental light at the corneal surface. Thus, the data augmentation algorithms need to support these variations/noises. We introduce color augmentations (CA) to simulate 1 and 2 and corneal reflection augmentation (CRA) to simulate 3.

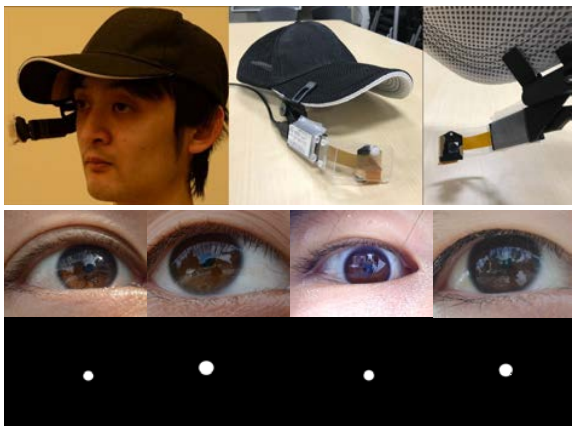

Fig. 2. Corneal imaging camera (upper) and outdoor eye image dataset (lower) and its groud-truth pupil segments. Outdoor eye images include a lot of corneal scene reflections.

*1) Color augmentation:* To simulate the variations of color sensitivity and white balance parameters of a camera, we apply CA in the Lab color space. Assuming the original RGB-color of a pixel as $[r, g, b]$, we first convert the color into CIE-Lab color space as in $[L, a, b] = \text{rgb2Lab}([r, g, b])$ and apply random CA for each image:

$$[L', a', b'] = [L \pm \text{rand}(L_d), a \pm \text{rand}(a_d), b \pm \text{rand}(b_d)]\,,$$

where $rand(x)$ is a uniform random variable ranging $-x$ to $x$. We set $[L_d, a_d, b_d] = [20, 30, 30]$ in our experiment in 8-bit Lab color space. Finally, we back-project from Lab to RGB color space and obtain final augmented images (Figure 3(d)(e)).

*2) Corneal Reflection augmentation:* Corneal reflections (CR) are the reflection of the environment light at the corneal surface and are the major image noise in iris/pupil images taken under visible lighting conditions. To more robustly detect pupils, we augmented eye images that include CR. Assuming an eye image as $\mathbf{I}_{\text{e}}$, binary iris mask as $\mathbf{I}_{\text{iris}} \in \{0, 1\}$ (Figure 3(c)), and a randomly chosen scene image as $\mathbf{I}_{\text{sc}}$, the corneal reflection augmented image $\mathbf{I}_{\text{CRA}}$ is given as,

$$\mathbf{I}_{\text{CRA}} = \alpha\mathbf{I}_{\text{e}} + (1 - \alpha)\mathbf{I}_{\text{iris}} \cdot \mathbf{I}_{\text{sc}},$$

where $\alpha$ is the randomly set blending parameter ranging from 0 to 0.7. The resulting images after applying CA and CRA and scene images used for augmentations are shown in Figure 3(f)(g).

### C. Learning pupil segmentation CNN

By using the augmented eye image datasets and ground-truth pupil images, a pupil segmentation network is learned. The structure of the network is illustrated in Figure 4. The input image is first converted into $224 \times 224$ pixels and input to a two-stage $3 \times 3$ convolution network and then composed into an output of 1/2 size. These stages are overlaid for five layers, and last layer is a $7 \times 7$ full connected network. The decomposition network is the inverse structure of the composition network, namely, two convolution networks as well as $\times 2$ up-sampling layers. The output and input of composition and decomposition at the same layer are directly connected (short-cut connection) so that the filter properties at the same level can be shared together. In our experiment, we used an Adam optimizer [13] with default parameters $([lr, \beta_1, \beta_2] = [0.001, 0.9, 0.999])$ and 100 epochs for learning. We used 2,250 eye images and pupil masks for UBIRIS.v2 dataset. From one image, we generated five images by CA and CA+CRA, thus, 13,500 images are prepared to learn the networks using UBIRIS.v2+CA and UBIRIS.v2+CA+CRA.

TABLE I
PUPIL SEGMENTATION ACCURACY (UBIRIS.V2, 10-FOLD CROSS VALIDATION)

| Learned Dataset | Precision (av.) | Recall (av.) | $F_1$ score (av.) |
|---|---|---|---|
| UBIRIS.v2 | 0.640 | 0.774 | 0.786 |
| UBIRIS.v2+CA | 0.773 | **0.902** | **0.815** |
| UBIRIS.v2+CA+CRA | **0.801** | 0.855 | 0.808 |

TABLE II
PUPIL SEGMENTATION ACCURACY (LEARNED BY UBIRIS.V2, TESTED BY USING OUTDOOR EYE IMAGE DATASET.)

| Learned Dataset | Precision (av.) | Recall (av.) | $F_1$ score (av.) |
|---|---|---|---|
| UBIRIS.v2 | 0.554 | 0.602 | 0.531 |
| UBIRIS.v2+CA | 0.641 | 0.658 | 0.604 |
| UBIRIS.v2+CA+CRA | **0.670** | **0.707** | **0.630** |

## III. EXPERIMENTS

We conducted three experiments. First, we evaluated the segmentation performance by using the UBIRIS.v2 dataset. The network is learned by using the close-up eye images and their ground-truth images. The second experiment is conducted to evaluate the performance of pupil center detection by using GI4E dataset. The third experiment shows the applicability of the proposed approach for real EGT. Using the segmentation results, we built an EGT system that uses the proposed pupil detection method.

### A. Evaluation of Segmentation Accuracy

We first evaluate the pupil segmentation accuracy using the UBIRIS.v2 dataset and our original dataset. We train the network by using the different sets of the datasets, namely, UBIRIS.v2 eye images, UBIRIS.v2 and CA images, and UBIRIS.v2, CA, and CRA images. The training and evaluation is performed by 10-fold cross validation.

We evaluate the accuracy as follows. We compute the number of pixels of True Positives (TP), False Positives (FP), and False Negatives (FN) and obtain the precision $(TP/(TP + FP))$, recall $(TP/(TP + FN))$, and $F_1$ score $2 \cdot \frac{precision \cdot recall}{precision + recall}$.

The experimental results are shown in Table I and Figure 5. From the results using the UBIRIS.v2 dataset, the network learned by UBIRIS.v2+CA performed best ($F_1$ = 0.815) followed by that learned by U2+CA+CRA ($F_1 = 0.808$).

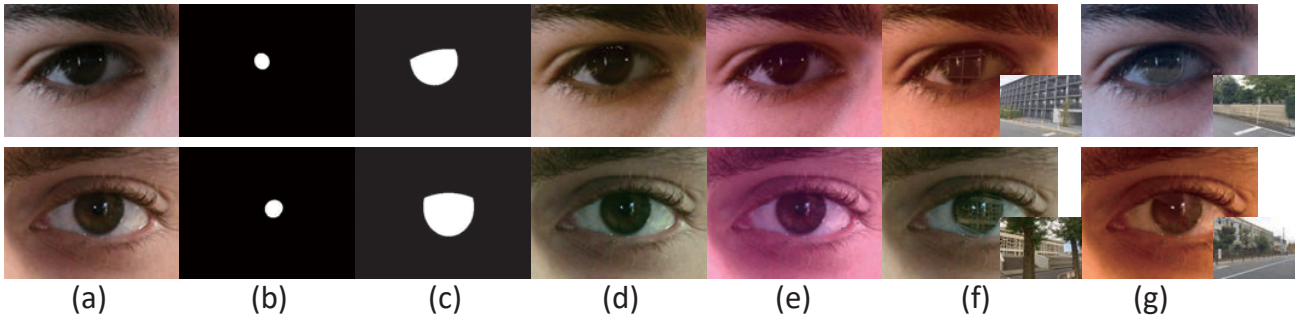


Fig. 3. Examples of eye image (UBIRIS.v2), pupil, iris segments and augmentations. (a) Eye images, (b) pupil mask, (c) iris mask, (d)(e) color augmentations (CA), (f)(g) color and corneal reflection augmentations (CA+CRA). Thumbnails overlaid in right-bottom are the scene pictures used for corneal reflection augmentations. This picture is best viewed in color.

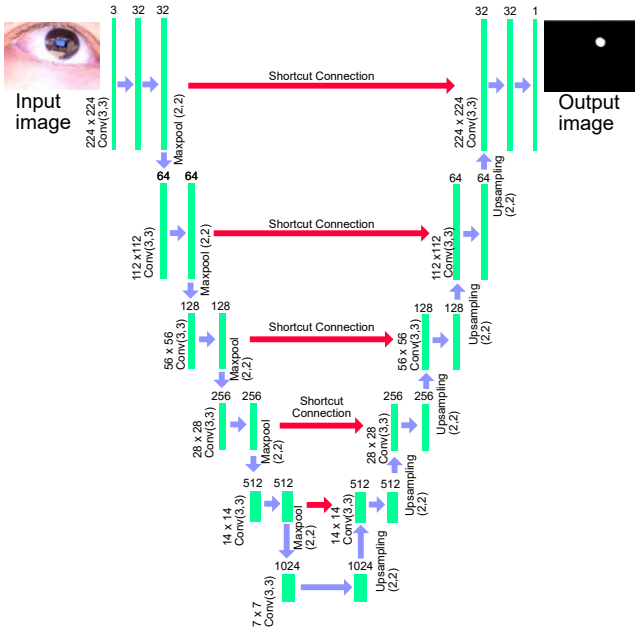


Fig. 4. Architecture of segmentation network (U-Net) used in our pupil detection method.

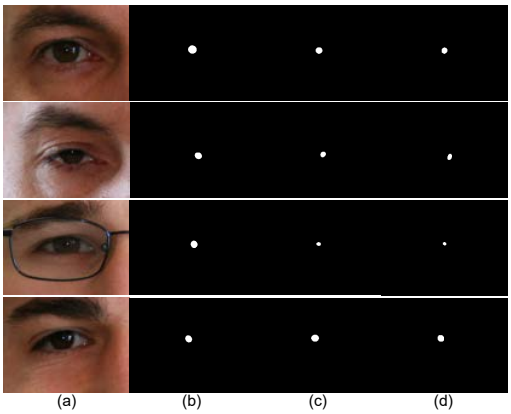


Fig. 5. Example results of the pupil segmentation using UBIRIS.v2 datasets (10-fold cross validation). (a) Original eye images, (b) ground-truth pupil segment, (c) results using UBIRIS.v2, (c) results using UBIRIS.v2 + color augmented eye images, and (d) results using UBIRIS.v2 + color augmented + corneal reflection augmented images.

### B. Evaluation of Pupil Center Estimation Accuracy

The second evaluation is the accuracies of the pupil center detection. For this evaluation, we used the network that is learned by UBIRIS.v2 datasets and tested by using GI4E datasets [11]. Specifically, we first learned the network by using the images in UBIRIS.v2, then applied the segmentation to the cropped eye regions in the facial images of the GI4E dataset, and lastly obtained the pupil centers.

Table III and Figure 6 show the experimental results. The GI4E database evaluates the results by the number of images whose errors $\epsilon$ are within particular thresholds: $\epsilon < 0.025$, $\epsilon < 0.05$, and $\epsilon < 0.1$. The $\epsilon$ is computed by $\frac{|PD_{estimated} - PD_{gt}|}{PD_{gt}}$ where $PD_{estimated}$ is the estimated pupil distance of both eyes (PD) and $PD_{gt}$ is the ground-truth PD. According to the results, the proposed algorithm performs best (98.62%) for $\epsilon < 0.05$ and performed 96.26% in the case of $\epsilon < 0.025$, which is a case that has not been clearly investigated in existing literature.

TABLE III
PUPIL CENTER ESTIMATION USING GI4E DATASETS

| Learned Dataset | $\epsilon < 0.025$ | $\epsilon < 0.05$ | $\epsilon < 0.1$ |
|---|---|---|---|
| Proposed (UBIRIS.v2) | 92.15% | 98.38% | 99.59% |
| Proposed (UBIRIS.v2+CA) | 94.50% | 98.06% | 98.46% |
| Proposed (UBIRIS.v2+CA+CRA) | **96.28%** | **98.62%** | 98.95% |
| Baek2013 [14] | – | 79.50% | 88.00% |
| Villanueva2013 [15] | – | 93.90% | 97.30% |
| Zhang2016 [16] | – | 97.90% | 99.60% |
| Gou2016 [17] | – | 98.20% | **99.80%** |
| Gou2017 [18] | – | 94.20% | 99.10% |

### C. Evaluation of Point of Gaze estimation (Eye gaze tracking)

To apply the proposed algorithm to an EGT system, we construct an eye image dataset assuming the eye tracking setup. Figure 7 shows the experimental setup. A subject uses the chin-rest to fix the facial position to 140 cm away from a 70-inch display. Six subjects looked at the 12 markers shown

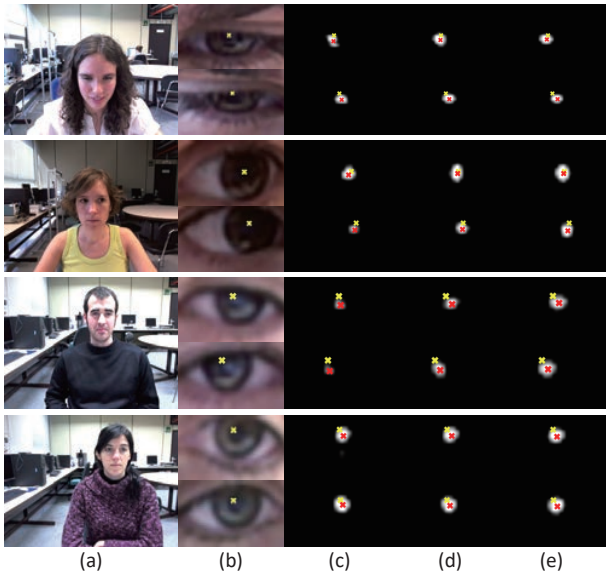


Fig. 6. Example results of the pupil center detection using GI4E datasets. (a) Original facial images, (b) cropped eye images (upper:left, lower:right), pupil detection results for the proposed method by using the DNN learned by (c) UBIRIS.v2 dataset, (d) UBIRIS.v2 dataset and color augmentations (CA), and (e) UBIRIS.v2 dataset, color augmentations and corneal reflection augmentations (CA+CRA). In all images, yellow and red crosses indicate the ground-truth pupil centers and the estimation results, respectively.

on the display, where the vertical and horizontal intervals of markers are 20 cm. A subject did two rounds of a gazing task. The eye images of the first and second round are used for calibration and evaluating gaze estimation, respectively. In the calibration step, we obtain the mapping functions $[m_x, m_y] = [f_x(p_x, p_y), f_y(p_x, p_y)]$ that transfer a pupil center $[p_x, p_y]$ to a marker point $[m_x, m_y]$. We used 2nd-order approximation for $f_x$ and $f_y$ [19], namely,

$$f_x(x,y) = \sum_{j=0}^{2}\sum_{k=0}^{j} a_{jk} {p_x}^k {p_y}^{j-k},$$

$$f_y(x,y) = \sum_{j=0}^{2}\sum_{k=0}^{j} b_{jk} {p_x}^k {p_y}^{j-k}.$$

The calibration parameters $a_{jk}, b_{jk}$ are computed by solving the least square minimization of following residual functions.

$$E_{f_x} = \sum_{i=0}^{12} \| f_x(p_x^i, p_y^i) - m_x^i \|^2,$$

$$E_{f_y} = \sum_{i=0}^{12} \| f_y(p_x^i, p_y^i) - m_y^i \|^2,$$

where $[p_x^i, p_y^i]$ and $[m_x^i, m_y^i]$ are the $i$-th numbers of the pupil center and marker points of the calibration round. In the experiment, we used the right eye for estimation and the network used is learned by using the UBIRIS.v2+CA+CRA dataset.

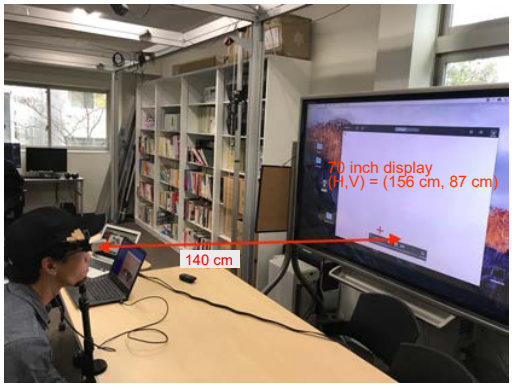


Fig. 7. Data capturing environment assuming eye gaze tracking task. A subject uses the chin-rest to fix the facial position to 140 cm away from a 70 inch display. Subjects gaze at 12 marker points shown on the display.

TABLE IV
GAZE ESTIMATION ERROR [DEG.]

| Subject | 1 | 2 | 3 | 4 | 5 | 6 | Avg. |
|---|---|---|---|---|---|---|---|
| Accuracy | 0.91 | 0.64 | 0.70 | 1.15 | 1.49 | 0.63 | 0.92 |
| Accuracy [20] | – | – | – | – | – | – | 1.81 |

The experimental results of ten subjects are shown in Table IV and Figures 8 and 9.

The gaze estimation is accurate to 0.92 [deg] on average, ranging 0.63 - 1.49 [deg] between individuals. According to Figure 8, there is no significant relationship between gaze marker points and errors, but lower-right markers show relatively larger errors (1.18 and 1.99 [deg]). We think this is because the right bottom markers are hard to look at with the right eye, and thus subjects tend to move their faces for these markers. Despite this, the system using our proposed method performs nearly as accurately as the commercial systems using IR lighting that typically are accurate to within 1 degree.

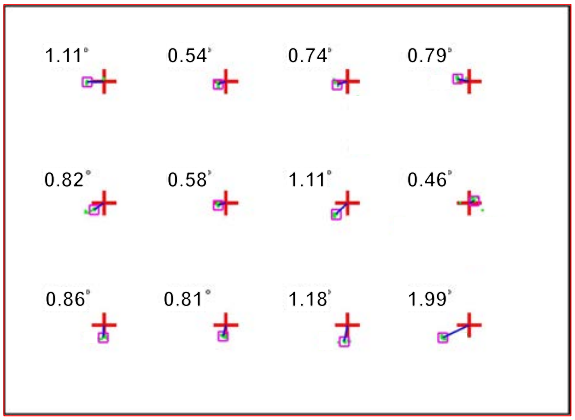


Fig. 8. Eye gaze tracking results for all subjects. The red crosses are the ground truth markers, and green points are the results of the estimations. Magenta squares are the means of the estimations. The horizontal and vertical intervals between markers are 20 cm.

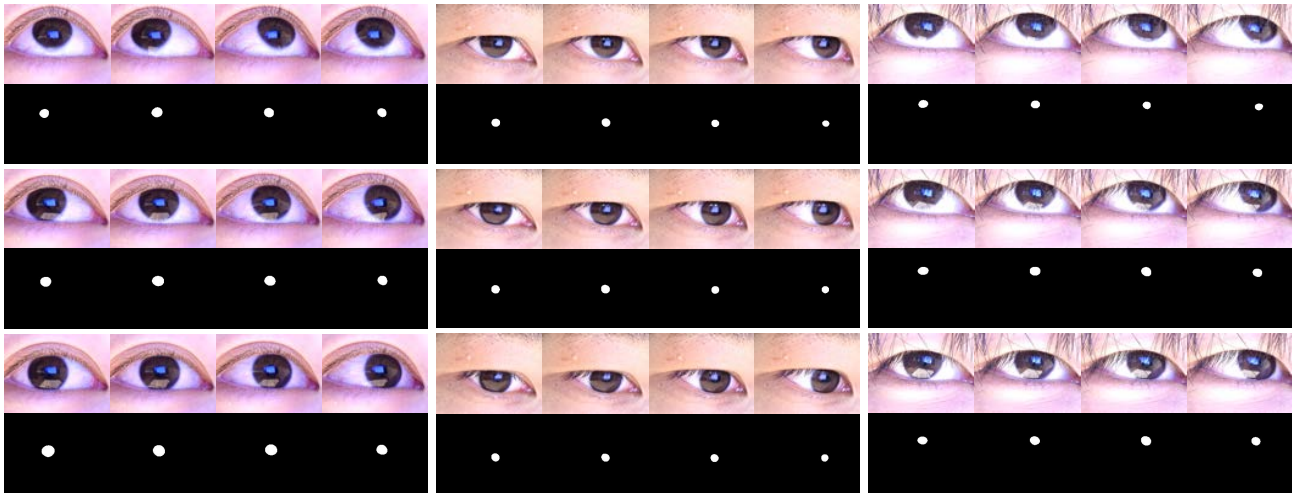

Fig. 9. Pupil detection results of the eye gaze tracking experiments of three subjects.

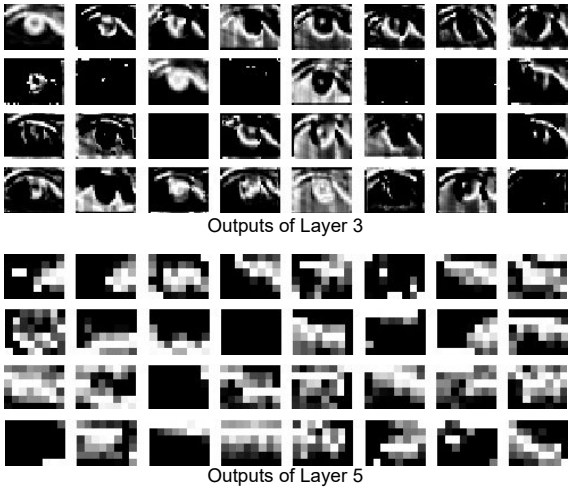


Fig. 10. Output of CNN of layer 3 and 5 in the composition side.

## IV. Discussion and Conclusion

The results show the proposed pupil detection method robustly works for hard conditions including dark-brown (black) eyes and contamination of scene lights. The proposed algorithm outperformed the existing approaches at detecting pupil centers using the GI4E dataset. The important feature is that the same network architecture/learned weights are used to both segment the pupil and detect its center. This shows the scalability of the proposed algorithm. We also found the data augmentation techniques are quite important to increase the performances.

The eye gaze tracking (EGT) system using the proposed method is accurate to within 1 degree, which is better than that of Mori et al. that also uses VL images (1.805 degrees [20]). Moreover, it has nearly the same accuracy as the commercial video-based EGT systems that use infrared IR lighting. Since the proposed EGT system does not require special hardware and is robust under sunny conditions where IR lights cannot be used, we believe it has high potential to be used in practical applications.

It will be interesting to see what the convolutional neural network (CNN) learns from the dataset. Figure 10 visualizes the output of the third ($56 \times 56$) and fifth ($14 \times 14$) layers from the input in the composition side. As we can see, the pupil feature is roughly detected at the third layer, thus we expect the pupil can be detected only in the shallow layers. However, the deep layer may contribute to the global robustness (reducing false detection) of the pupils. We need to explore the optimal network depth relative to the eye/pupil image size in the image in the future.

**Acknowledgements** This work was supported by JSPS KAKENHI and JST CREST Grant Number JPMJCR17A5, Japan.

## References

[1] A. Matsumoto, Y. Tange, A. Nakazawa, and T. Nishida, "Estimation of task difficulty and habituation effect while visual manipulation using pupillary response," in *Video Analytics. Face and Facial Expression Recognition and Audience Measurement - Third International Workshop, VAAM 2016, and Second International Workshop, FFER 2016, Cancun, Mexico, December 4, 2016, Revised Selected Papers*, 2016, pp. 24–35.

[2] C. H. Morimoto, D. Koons, A. Amir, and M. Flickner, "Pupil detection and tracking using multiple light sources," *Image and vision computing*, vol. 18, no. 4, pp. 331–335, 2000.

[3] H. Proença and J. C. Neves, "Irina: Iris recognition (even) in inaccurately segmented data," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 538–547.

[4] Z. Zhao and K. Ajay, "An accurate iris segmentation framework under relaxed imaging constraints using total variation model," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3828–3836.

[5] Y. Morita, H. Takano, and K. Nakamura, "Pupil diameter measurement in visible-light environment using separability filter," in *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on*. IEEE, 2016, pp. 000 934–000 939.

[6] Muhammad Arsalan, Hyung Hong, Rizwan Naqvi, Min Lee, Min Kim, Dong Kim, Chan Kim, and Kang Park, "Deep Learning-Based Iris Segmentation for Iris Recognition in Visible Light Environment," *Symmetry*, vol. 9, no. 11, p. 263, Nov. 2017. [Online]. Available: http://www.mdpi.com/2073-8994/9/11/263

[7] S. Bazrafkan, S. Thavalengal, and P. Corcoran, "An End to End Deep Neural Network for Iris Segmentation in Unconstraint Scenarios," *arXiv preprint arXiv:1712.02877*, 2017.

[8] D. E. King, "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[10] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre, "The ubiris. v2: A database of visible wavelength iris images captured on-the-move and at-a-distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1529–1535, 2010.

[11] 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 Trans. Multimedia Comput. Commun. Appl.*, vol. 9, no. 4, pp. 25:1–25:20, Aug. 2013. [Online]. Available: http://doi.acm.org/10.1145/2501643.2501647

[12] A. Nakazawa, C. Nitschke, and T. Nishida, "Non-calibrated and real-time human view estimation using a mobile corneal imaging camera," in *Multimedia & Expo Workshops (ICMEW), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1–6.

[13] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014. [Online]. Available: http://arxiv.org/abs/1412.6980

[14] S.-J. Baek, K.-A. Choi, C. Ma, Y.-H. Kim, and S.-J. Ko, "Eyeball model-based iris center localization for visible image-based eye-gaze tracking systems," *IEEE Transactions on Consumer Electronics*, vol. 59, no. 2, pp. 415–421, 2013.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] A. Goshtasby, "Image registration by local approximation methods," *Image and Vision Computing*, vol. 6, no. 4, pp. 255–261, 1988.

[20] H. Mori, E. Sumiya, T. Mashita, K. Kiyokawa, and H. Takemura, "A wide-view parallax-free eye-mark recorder with a hyperboloidal half-silvered mirror and appearance-based gaze estimation," *IEEE transactions on visualization and computer graphics*, vol. 17, no. 7, pp. 900–912, 2011.