RavenGaze: A Dataset for Gaze Estimation Leveraging Psychological Experiment Through Eye Tracker

Tao Xu[1] Bo Wu[1], Yuqiong Bai[1] and Yun Zhou[2]
[1] School of Software, Northwestern Polytechnical University, Xi'an, China
[2] Faculty of Education, Shaanxi Normal University, Xi'an, China

2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG) | 979-8-3503-4544-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/FG57933.2023.10042793

Abstract— **One major challenge in appearance-based gaze estimation is the lack of high-quality labeled data. Establishing databases or datasets is a way to obtain accurate gaze data and test methods or tools. However, the methods of collecting data in existing databases are designed on artificial chasing target tasks or unintentional free-looking tasks, which are not natural and real eye interactions and cannot reflect the inner cognitive processes of humans. To fill this gap, we propose the first gaze estimation dataset collected from an actual psychological experiment by the eye tracker, called the RavenGaze dataset. We design an experiment employing Raven's Matrices as visual stimuli and collecting gaze data, facial videos as well as screen content videos simultaneously. Thirty-four participants were recruited. The results show that the existing algorithms perform well on our RavenGaze dataset in the 3D and 2D gaze estimation task, and demonstrate good generalization ability according to cross-dataset evaluation task. RavenGaze and the establishment of the benchmark lay the foundation for other researchers to do further in-depth research and test their methods or tools. Our dataset is available at https://intelligentinteractivelab.github.io/datasets/RavenGaze/index.html.**

I. INTRODUCTION

"Eye-mind" theory suggests that eye movements show where attention is directed [9][7]. The gaze has been identified as the best indicator of cognitive functioning, thus allowing researchers to arrive at fundamental mechanisms and processes. The goal of gaze estimation is to determine where a person is looking based on the image of their face. In human-computer interaction, robotics, psychology, and educational research and practice, gaze estimation has been applied to many applications [20]. Therefore, gaze estimation algorithms are in greater demand than ever to be accurate and reliable. The lack of high-quality data is a major obstacle for researchers in designing good algorithms and building practical applications.

Benefit from deep learning, appearance-based gaze estimation methods perform better only requiring human facial or eye images [5]. However, it requires a large amount of accurate gaze data to support model training. Establishing the gaze estimation database is a way to obtain accurate gaze data and test methods or tools. Existing gaze estimation databases can be categorized into two types: based on the target and target-free. The former one is based on the experiment that the design is specific tasks and requires subjects to look at different targets so as to collect ground truth data merely by cameras. This method is less expensive; however, the experiment of which is based on artificially constructed scenes and does not conform to people's actual habits of reading, viewing scenes, and watching. Thus, it cannot capture some core features of eye movements in the natural eye interaction process, such as saccades. The latter one is to record eye movements naturally by eye trackers. In the experiment, the subjects are asked to look freely as they want, without any tasks. According to the state-of-the-art, there are only two databases of this type. However, real eye interactions reflect subjects' intentions and inner cognitive processes. The data of real eye interactions and movements differ from unintentional gaze data as mentioned above. In sum, it is urgent to build high-quality gaze estimation databases based on real tasks containing a large amount of accurate-annotated gaze data.

To fill this gap, the present paper proposes the first gaze estimation dataset collected from an actual psychological experiment by the eye tracker, called the RavenGaze dataset. We design an experiment employing the Raven's Advanced Progressive Matrices as visual stimuli and collecting gaze data, facial videos as well as screen content videos simultaneously. The gaze data were collected through a Tobii Pro Nano eye tracker. Thirty-four participants were recruited. The results show that the existing algorithms perform well on our RavenGaze dataset in the 3D and 2D gaze estimation task, and demonstrate good generalization ability according to cross-dataset evaluation task, compared with the MPI-IFaceGaze and EyeDiap datasets.

In this work, we made two salient contributions:

- We designed and built the target-free RavenGaze dataset, which used an eye tracker device to record eye movements while subjects were doing the Raven's tests. The process is an entirely natural eye interaction based on real tasks. Additionally, Raven's tests aim to evaluate people's ability to reason. Thus, the collected eye movement data can reflect human's real interaction intention and cognitive processes.
- We provided a benchmark performance from various perspectives based on the RavenGaze dataset. We used recently popular and state-of-the-art gaze estimate methods to evaluate our dataset. The methods contained 2D gaze estimation across personal tasks and 3D gaze estimation across dataset tasks. This establishment of the benchmark lays the foundation for other researchers to do further in-depth research and test their methods

979-8-3503-4544-5/23$31.00 ©2023 IEEE

or tools.

II. Related of works

A. Gaze estimation datasets

One of the biggest challenges of gaze estimation is to provide a high-quality labeled dataset. Researchers have many attempts to build different kinds of datasets. According to experiment contents, we divide existing eye estimation datasets into two categories: the target paradigm experiment-based dataset and the target-free paradigm experiment-based dataset. Comparison with existing gaze estimation datasets is shown in Table. I.

The target paradigm experiment's main idea is to ask participants to watch the designed targets like dots and circles for a period. The designed target could be moving or fixed, while one or several cameras are used to record subjects' gestures and gazes. Some research work [23], [11], [16] uses a fixed target, for example, in the dataset of Columbia [16], a grid of dots was attached to a wall in front of subjects as the visual target. Some of the work [10], [17], [7] uses a dynamic target. In the dataset of UTMV [17] uses the white circle on the screen as the gaze target, which is moved to the center of each grid shown on the screen in random order during experiments. This target paradigm has relatively low requirements on hardware, only cameras, and the acquisition process is simple and low cost. This paradigm is widely adopted by most datasets. But the gaze data in the target paradigm collected comes from deliberately designed points or trajectories, not the subject's natural and authentic eye movements. It can not contain some essential actual eye activity features such as saccades. It can be regarded as a kind of simulation data, which is quite different from eye movement in the natural state. The gaze estimate method built based on this paradigm has limited generalization ability and is difficult to apply in practice.

The target-free paradigm experiment doesn't need any specific target to watch. Therefore, the ground truth gaze data is collected from the eye tracker. The advantage of this kind of paradigm is that it is closer to people's natural interaction. Gaze360 [10] attempts to use the eye-tracking glass to record subjects' gazes. Subjects are asked to vary their head poses and eye gazes as much as possible and move within the motion capture area. It is no specific task to complete in this experiment. EVE Dataset [14] is the first dataset that used a screen-based eye tracker to label the gaze. It employed image, video, and Wikipedia pages as visual stimuli. This experiment has also no specific task besides asking subjects freely view these visible contents. These two collection processes can be considered passive viewing of the subject, which lacks a clear purpose and intent. We believe that eye movement in authentic and natural interactions has apparent intentions, so there is still a gap between this collection of eye movement data and natural eye movement. To solve these problems, We present a novel gaze estimation dataset called RevenGaze. It adopts a screen-based eye tracker to record subjects' eye movement during a task, completing Raven test. It records subjects' eye movement during a real job, which is associated with subjects' problem-solving thinking process. It provides a new option for researchers to explore eye movements during fundamental interactions.

B. Gaze estimation methods

Gaze estimation is divided into model-based methods and appearance-based methods. Model-based methods focus on modeling eye geometry, and gaze estimation is performed by extracting geometric features such as corneal reflection [8], pupil center[18], etc. Compared with model-based methods, appearance-based methods learn a mapping function from face images to gaze directions or gaze points, which do not need to extract features manually, Since it can directly estimate gaze from facial images or eye images captured by cameras in the wild, the appearance-based method becomes the mainstream in this research field. One of the earliest related research works can be traced back to 1994. Baluja et al. developed an artificial neural network-based gaze tracking system that can only estimate gaze by using an image of the pupil and cornea as the input. With the advent of AlexNet [12], Convolutional Neural Networks (CNN) demonstrate powerful capabilities in computer vision, and it also achieved superior performance in the field of gaze estimation [5]. Zhang et al.[21] first proposed a single-branch input CNN method to estimate the gaze direction from a single-eye image. Since head pose can affect gaze estimation, some methods for obtaining gaze from eye images [23], [6] additionally input the head pose vector, which is spliced into a fully connected layer for gaze regression. The facial image contains head pose information but also contains other redundant information. To preserve the gaze-related information, iTracker[11] first uses face images, left and right eye images, and face mesh information as multi-channel input. Zhang et al. [22] proposed a spatial weighting mechanism to encode the importance of the entire facial image. Liu et al.[13] used multi-scale channels and spatial attention to select important features from face images for gaze regression. Cheng et al.[3] proposed a self-adversarial network to remove facial image features and retain gaze-related features. Some widely used appearance-based methods are adopted to evaluate our RavenGaze Dataset.

III. The RavenGaze Dataset

A. Experiment design

We design experiments based on two guiding ideas:

- Record a completely actual interaction process replacing labeling gaze by specific target.
- Set specific task goals to ensure that the eye movement data is a true reflection of cognitive activities.

We choose an psychological cognitive experiment, Raven Progressive Matrices (RPM) Test [15] as experiment materials. It is one of common used experiment to measure human intelligence quotient, widely adopted in cognitive testing and psychological experiments. RPM is a non-verbal test, all of the questions being comprised of visual geometric design with a missing piece, an example of RPM test is shown in

TABLE I

OVERVIEW OF POPULAR GAZE ESTIMATION DATASETS

Datasets	Year	#Subjects	Image	#Data	Label	Gaze Device	Screen-based	Paradigm	Visual Stimuli
Columbia [16]	2013	56	Face	5,880	3D	Canon EOS Rebel T3i camera	N	Target	Gird of dots on the wall
UTMV[17]	2014	50	Eyes	64,000	2D+3D	Color camera	Y	Target	White circle of grid center on the monitor
EyeDiap[7]	2014	16	Face	237min	2D+3D	RGB-D camera,HD camera	Y	Target	Random circle and curve in the screen,3D floating ball
MPIIGaze[23]	2015	15	Eyes	213,659	2D+3D	Laptop camera	Y	Target	Random on-screen positions
GazeCapture[11]	2016	1,474	Face	2,445,504	2D	IPhone camera	Y	Target	Random dots on the screen
MPIIFaceGaze[22]	2017	15	Face	213,659	2D+3D	Laptop camera	Y	Target	Random on-screen positions
RT-GENE[6]	2018	15	Face	122,531	3D	RGB-D camera,pupil Labs mobile eyetracking glasses	N	Target-free	Free-viewing within the motion capture area
Gaze360[10]	2019	238	Face	172,000	3D	Ladybug5 360° camera	N	Target	Moving rigid target board
EVE [14]	2020	54	Face	12,308,334	2D+3D	Logitech C922 webcam,Tobii Pro Spectrum Eye Tracker	Y	Target-free	Image,video,wikipedia page
RavenGaze(Ours)	**2022**	**31**	**Face**	**309min**	**2D+3D**	**Tobii Pro Nano Eye Tracker**	**Y**	**Target-free**	**The Raven's Advanced Progressive Matrices**

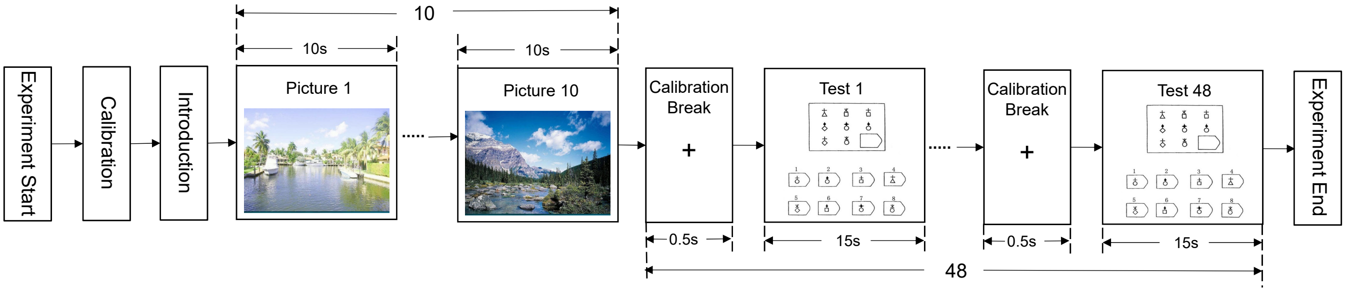


Fig. 1. The procedure of experiment

Fig. 2. Because this test is independence of language and reading, it is easy to be used by people of different ages, cultures and education level.

We used the 48-question from RPM as the subject's visual stimuli. In addition, to ensure subjects are not distracted as much as possible during the task time, we set answer time for questions(15s) shorter than the average time to work out the questions. The procedure of the experiment is shown in Fig. 1.

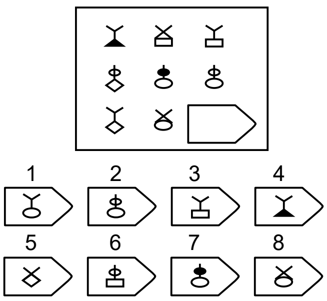


Fig. 2. The procedure of experiment

Before each subject starts the investigation, first, the eye tracker is calibrated the gaze of the current subject to reduce the gaze error of the collection process. Then ten landscape pictures help the subject calm down and enter the test state. After that, it begins the test, which includes 48 questions. We set each question to have a maximum answering time of 15 seconds. The subject must press the button corresponding to the answer option preset by the program within the answering time to complete the question and automatically jump to the next question. If the subject does not meet the answer within the set time, it will automatically jump to the next question. Before jumping to the next question, a calibration break is set to ensure high-quality gaze data. It is a fixation point that appeared in the center of the screen for 0.5 seconds, and subjects were asked to focus on this fixation point as much as possible. During the answering process, the program simultaneously records the facial video, screen video, and eye movement data of each subject during completing the 48 reasoning questions.

B. Data collection

The experiment setting is shown in Fig. 3, and the visual stimuli materials are displayed on a 27-Inch monitor. The subjects' videos are captured from a commercial webcam fixed on the top of the screen. Logitech C270 HD Webcam is chosen, since it is one of low-cost and most widely used webcam, supporting 1280×720 video recording with light correction technology.

The ground truth eye movements are recorded by the screen-based eye tracker, the Tobii Pro Nano eye tracker, setting on the bottom of the screen. The eye tracker is high precision and non-contact, supported by Pupil Center Corneal Reflection (PCCR) technology. It gradually becomes the primary collection tool of the eye movement data in gaze estimation datasets. The Tobii Pro Nano eye tracker is one of the world's smallest research-grade eye trackers

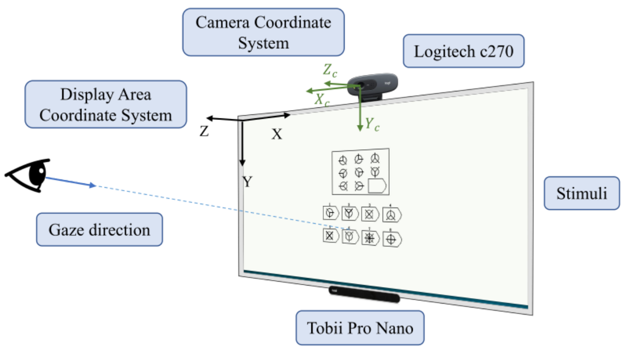


Fig. 3. Experiment setting

with a proven capability of tracking. It is easy-to-use and robust to eye tracking research, which collects gaze data at 60 hertz (Hz) after a quick calibration. It can recognize the gaze direction, screen gaze point, pupil size, saccade, and other information.

In order to obtain accurate data and simplify the data labeling process, we developed a time-synchronized experimental protocol. When subjects performed the Raven advanced reasoning experiment, the program started and ended synchronously to record screen video, facial video, and eye movement data. It can label the time stamps on frame sequences at the beginning and end of each question, which can be used to segment the video and eye-tracking data to form the subject's independent data for each question.

C. Dataset characteristic

We collected gaze data from a total of 34 subjects(18 female and 16 male), with 22 wearing glasses and 12 having normal vision. Three subjects' gaze data were screened out due to improper experimental operation. The data were collected under well-lit indoor conditions, where the sampling rate of face video was 30Hz, and the sampling rate of Tobii Pro Nano eye tracker was 60Hz. The time for each subject to complete RMP test is about 8-11 minutes. The entire dataset contains 31 subjects' facial videos with a total length of 309 minutes. There are 556,476 images in total, some face image examples are shown in Fig.4.

Fig. 4. Face image from the dataset.

For head posture, we did not ask them to turn their heads deliberately to create a wide range of head angles, all actual data was collected when they were browsing the screen content usually. During the answering process, within 15 seconds of each question, the subject's head posture did not change significantly, and they almost maintained a head posture.

The gaze angle in the horizontal direction is 10°, and the rise in the vertical direction is 12° shown in Fig.5(a). The gaze range is demonstrated in Fig.5(b), which is smaller than another gaze dataset. It also means that RavenGaze places higher demands on the recognition algorithms.

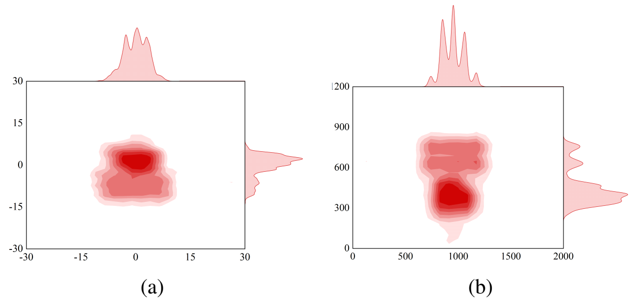


Fig. 5. Gaze direction and gaze point distribution of RavenGaze,(a) Gaze direction range of RavenGaze,(b) Gaze point range of RavenGaze

IV. Evaluation protocol

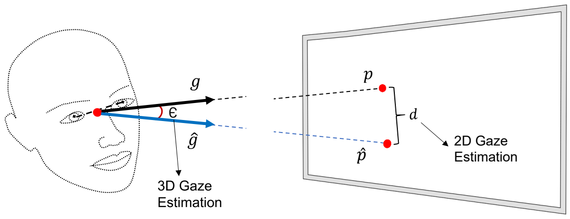
To easy for further explore and study RevenGaze, we conduct a benchmark performance based on recent state-of-the-art methods from different aspects to evaluate our RavenGaze dataset. The current appearance-based gaze estimation methods are mainstream methods and usually have a good performance on the exited datasets. We have selected eight recently state-of-the-art appearance-based gaze estimation methods: GazeNet[23], Spatial Weights CNN[22], Dilated-Net[2], RT-Gene[6], Gaze360[10], GazeTR-Hybird[5], Itracker[11], Spatial Weights CNN[22], AFF-Net[1], and conduct experiments from the evaluation within the dataset and the evaluation across the dataset from 3D gaze estimation task and 2D gaze estimation task. To fully insight the performance, we select two famous screen-based gaze estimate datasets: MPIIFaceGaze[22] and EyeDiap[7] for comparison. Finally, the benchmark of cross-dataset experiments are provided.

A. Data preprocessing

In gaze estimation tasks, the context of an image is gaze-independent information. To improve the generalization of the dataset, we crop out face images and local eye images from selected video frames to avoid redundant image backgrounds from affecting gaze estimation. The widly used preprocessing method is normalizing the gaze vector and projecting eye image to the virtual camera coordinate system[17]. In our data collection experiment, the camera recording the video is located at the center of the top of the screen, perpendicular to the screen and facing the participant's face, which allows the camera coordinate system and the display area coordinate system to only have a translation relationship without rotation transformation. At the same time, the participants had almost fixed head poses, which allowed us to use a simpler preprocessing method. We detect facial keypoints through MTCNN [19], and crop the eye image and face image by scaling the image according to the center landmark of the left and right eyes and the center landmark of the connection between the left and right eyes, respectively.

B. Within-dataset evaluation

For gaze estimation, it can be considered as a task to find where a person is looking at based on image of his face. There are two main directions for evaluation: 3D gaze estimation and 2D gaze estimation.



(a) 3D gaze estimation (b) 2D gaze estimation

Fig. 6. Two task of gaze estimation

1) 3D gaze estimation: This task refers to predict the gaze direction given human's face image, as shown Fig. . It can be considered as a regression task from input image I to 3D gaze vector g.

$$g = f(I) \tag{1}$$

In 3D gaze estimation, the angular error is commonly used to evaluate the performance of 3D gaze estimation algorithm [5].The actual gaze direction is a vector denoted as g, and estimated gaze direction is a vector denoted as $\hat{g}$. The angular error is the angle between the actual gaze vector and the estimated gaze vector g, computed as follows:

$$\epsilon = \frac{g \cdot \hat{g}}{\|g\|\|\hat{g}\|} \tag{2}$$

To fully evaluate our RavenDataset, we apply five popular methods mentioned above on RavenDataset based on a cross-validation strategy, and compared with two renowned screen based gaze estimation dataset: MPII Facegaze and Eyediap. Table II summarizes the evaluation results of 3D gaze estimation within the dataset.

TABLE II
COMPARISON WITH 3D APPEARANCE-BASED METHODS

Methods	RavenGaze	MPIIFaceGaze	EyeDiap
GazeNet[23]	4.86°	-	7.29°
Spatial Weights CNN[22]	4.80°	4.87°	6.40°
Dilated-Net[2]	4.88°	4.83°	6.65°
RT-Gene[6]	4.16°	4.68°	6.31°
Gaze360[10]	3.31°	4.05°	5.36°
GazeTR-Hybird[4]	3.44°	4.00°	5.17°

From the performance of the 3D gaze estimation method [23], [22], [2], [6], [4], RavenGaze has basically same trend of angular error with the public datasets MPIIFaceGaze and EyeDiap. Besides, the Gaze360 [10] performans better than our RavenGaze dataset. In conclusion, these existing 3D gaze estimation methods show good results on our RavenGaze dataset.

2) 2D gaze estimation: The other task is 2D gaze estimation, which is to predict coordinates of Point of Gaze (PoG) on a 2-D screen. It also can be considered as a regression task from input image I to 2D screen gaze position.

$$p = f(I) \tag{3}$$

The Euclidean distance d between the screen on screen gaze point predicted by gaze estimation algorithm and the actual gaze point labelled by the eye tracker is used as the evaluation standard. The actual gaze point on screen is denoted as p, and the estimated gaze point is denoted as $\hat{p}$. The Euclidean distant can be computed as follow:

$$d = \|p - \hat{p}\| \tag{4}$$

For 2D gaze estimation , Results in Table III indicates [11], [22], [1] achieves average Euclidean distances of 4.61cm, 5.53cm and 4.96cm on the RavenGaze Dataset respectively. The overall performance is better than the EyeDiap, with MPIIFaceGaze basically equivalent. Since RavenGaze has a smaller gaze range, this performance is expected.

TABLE III
COMPARISON WITH 2D APPEARANCE-BASED METHODS

Methods	RavenGaze	MPIIFaceGaze	EyeDiap
Itracker[11]	4.61cm	7.67cm	10.13cm
Spatial Weights CNN[22]	5.53cm	4.20cm	8.56cm
AFF-Net[1]	4.96cm	4.21cm	9.25cm

C. Cross-dataset evaluation

The cross-dataset evaluation reflects the generalization ability of the dataset. We train gaze estimation methods on one dataset, and test on the remaining dataset. For convenient, RavenGaze, MPIIFaceGaze[22], EyeDiap[7] are represented as R, M, E for short respectively. Table IV summarizes the cross-dataset evaluation results.

TABLE IV
COMPARISON WITH CROSS-DATASET

Methods	R→M	R→E	M→R	M→E	E→R	E→M
GazeNet[23]	13.01°	17.13°	8.7°	20.02°	13.38°	21.25°
Spatial Weights CNN[22]	14.09°	16.07°	16.18°	17.60°	19.11°	25.43°
Dilated-Net[2]	15.20°	13.72°	13.95°	17.38°	12.08°	20.47°
Gaze360[10]	13.06°	18.60°	13.23°	14.16°	14.93°	23.88°
GazeTR-Hybird[4]	14.36°	17.40°	17.42°	16.40°	13.59°	22.17°

When RavenGaze is used as the training set, the performances of selected methods have obviously better than EyeDiap as the training set. Compared with MPIIfaceGace, the task of testing on EyeDiap (R→E, M→E), RavenGaze perform slightly better. Given RavenGaze has a smaller gaze range and small changes in head posture, it demonstrates good generalization ability.

V. CONCLUSIONS AND FUTURE WORKS

A. Conclusions

In this work, we build a RevanGaze data, using Eyetracker to record eye movement of the whole process of subjects' Psychological Cognitive Experiment. Because using the eye tracker record eye movement, the data collection can be labeled without requiring the user to gaze at the specific target, and the real interaction process of the user can be collected. We chose the cognitive test: Raven test as the visual stimuli material, replacing free viewing with explicit intent compared to previous experiments with databases. Its gaze data can reflect the cognitive changes in the problem-solving process, which is the same as the natural interaction process. Therefore, when the algorithm performs well on our dataset, it can be more easily generalized to other application scenarios.

Similarly, our database still has some shortcomings. Because the gaze collection procedure is carried out indoors, the light changes are not obvious. In addition, during the whole experiment, the head posture did not change much.There is still a gap with some databases in these aspects.

B. Future Works

As mentioned above, our future work will focus on how to design experiments and collect gaze data in different natural environments and human body gestures. How to collecting gaze data from large-screen interactions involving multiple people, such as teleconferencing, will also be one of our future research directions.

VI. ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China: 62077036; National Key Research and Development Program of China: 2022YFC3303600, 2018AAA0100501; We gratefully acknowledge the support of Ascend CANN (Compute Architecture for Neural Networks) and Ascend 910 AI Processor used for this research. Artificial intelligence computing power and technical support are provided by Xi'an Future Artificial Intelligence Computing Center.

REFERENCES

[1] Y. Bao, Y. Cheng, Y. Liu, and F. Lu. Adaptive Feature Fusion Network for Gaze Tracking in Mobile Tablets. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 9936–9943, 2021. ISSN: 1051-4651.

[2] Z. Chen and B. E. Shi. Appearance-Based Gaze Estimation Using Dilated-Convolutions. In C. Jawahar, H. Li, G. Mori, and K. Schindler, editors, *Computer Vision – ACCV 2018*, pages 309–324, Cham, 2019. Springer International Publishing.

[3] Y. Cheng, Y. Bao, and F. Lu. PureGaze: Purifying Gaze Feature for Generalizable Gaze Estimation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(1):436–443, June 2022.

[4] Y. Cheng and F. Lu. Gaze Estimation using Transformer, May 2021. arXiv:2105.14424 [cs].

[5] Y. Cheng, H. Wang, Y. Bao, and F. Lu. Appearance-based Gaze Estimation With Deep Learning: A Review and Benchmark, Apr. 2021. arXiv:2104.12668 [cs].

[6] T. Fischer, H. J. Chang, and Y. Demiris. RT-GENE: Real-Time Eye Gaze Estimation in Natural Environments. In V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, editors, *Computer Vision – ECCV 2018*, volume 11214, pages 339–357. Springer International Publishing, Cham, 2018. Series Title: Lecture Notes in Computer Science.

[7] K. A. Funes Mora, F. Monay, and J.-M. Odobez. EYEDIAP: a database for the development and evaluation of gaze estimation algorithms from RGB and RGB-D cameras. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 255–258, Safety Harbor Florida, Mar. 2014. ACM.

[8] E. Guestrin and M. Eizenman. General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Transactions on Biomedical Engineering*, 53(6):1124–1133, June 2006. Conference Name: IEEE Transactions on Biomedical Engineering.

[9] M. A. Just and P. A. Carpenter. A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4):329–354, 1980. Place: US Publisher: American Psychological Association.

[10] P. Kellnhofer, A. Recasens, S. Stent, W. Matusik, and A. Torralba. Gaze360: Physically Unconstrained Gaze Estimation in the Wild. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6911–6920, Seoul, Korea (South), Oct. 2019. IEEE.

[11] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, and A. Torralba. Eye Tracking for Everyone. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2176–2184, Las Vegas, NV, USA, June 2016. IEEE.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[13] S. Liu, D. Liu, and H. Wu. Gaze Estimation with Multi-Scale Channel and Spatial Attention. In *Proceedings of the 2020 9th International Conference on Computing and Pattern Recognition*, pages 303–309, Xiamen China, Oct. 2020. ACM.

[14] S. Park, E. Aksan, X. Zhang, and O. Hilliges. Towards End-to-End Video-Based Eye-Tracking. In A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, editors, *Computer Vision – ECCV 2020*, pages 747–763, Cham, 2020. Springer International Publishing.

[15] J. Raven. The raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, 41(1):1 – 48, 2000.

[16] B. A. Smith, Q. Yin, S. K. Feiner, and S. K. Nayar. Gaze locking: passive eye contact detection for human-object interaction. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*, pages 271–280, St. Andrews Scotland, United Kingdom, Oct. 2013. ACM.

[17] Y. Sugano, Y. Matsushita, and Y. Sato. Learning-by-Synthesis for Appearance-Based 3D Gaze Estimation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1821–1828, Columbus, OH, USA, June 2014. IEEE.

[18] R. Valenti, N. Sebe, and T. Gevers. Combining Head Pose and Eye Location Information for Gaze Estimation. *IEEE Transactions on Image Processing*, 21(2):802–815, 2012. Conference Name: IEEE Transactions on Image Processing.

[19] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016. Conference Name: IEEE Signal Processing Letters.

[20] X. Zhang, S. Park, T. Beeler, D. Bradley, S. Tang, and O. Hilliges. ETH-XGaze: A Large Scale Dataset for Gaze Estimation Under Extreme Head Pose and Gaze Variation. In A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, editors, *Computer Vision – ECCV 2020*, Lecture Notes in Computer Science, pages 365–381, Cham, 2020. Springer International Publishing.

[21] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling. Appearance-based gaze estimation in the wild. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4511–4520, Boston, MA, USA, June 2015. IEEE.

[22] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling. It's Written All Over Your Face: Full-Face Appearance-Based Gaze Estimation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2299–2308, Honolulu, HI, USA, July 2017. IEEE.

[23] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling. MPIIGaze: Real-World Dataset and Deep Appearance-Based Gaze Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(1):162–175, 2019. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.