Machine Learning-Based Gaze-Tracking and Its Application in Quadrotor Control on Mobile Device

HU Jiahui[1], *LU Yonghua*[1*], *LIU Jiangwei*[2], *YAN Changkai*[2], *LIU Tao*[1]

1. College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, P. R. China;

2. China Aeronautical Control System Research Institute, Wuxi 214000, P. R. China

(Received 12 December 2022; revised 30 May 2023; accepted 6 October 2023)

Abstract: A machine learning-based monocular gaze-tracking technology for mobile devices is proposed. This non-invasive, convenient, and low-cost gaze-tracking method can capture the gaze points of users on the screen of mobile devices in real time. Combined with the quadrotor's 3D motion control, the user's gaze information is converted into the quadrotor's control signal, solving the limitations of previous control methods, which allows the user to manipulate the quadrotor through visual interaction. A complex quadrotor track is set up to test the feasibility of this method. Subjects are asked to intervene their gaze into the control flow to complete the flight tasks. Flight performance is evaluated by comparing with the joystick-based control method. Experimental results show that the proposed method can improve the smoothness and rationality of the quadrotor motion trajectory, and can introduce diversity, convenience, and intuitiveness to the quadrotor control.

Key words: gaze-tracking; UAV control; machine learning; HRI; eye-gaze drive

CLC number: TP249 **Document code:** A **Article ID:** 1005-1120(2023)05-0547-08

0 Introduction

In our daily life, eyes are not only an important organ for us to obtain information, but also an important source for us to transmit our thoughts and emotions to the outside world. Recently, the gaze-tracking has been applied to the direct control of graphical interfaces.

Using machine learning techniques, the mapping relationship between eye images and gaze information can be obtained. Among them, the method using convolutional neural network (CNN) is proven to be effective. In this method, information such as human eye image and head pose is input into CNN, and the gaze vector is decoded at the last fully connected layer. Theoretically, the network can be trained as long as there is sufficient data[1-2].

However, even using deep neural network for regression analysis, its accuracy is usually limited to about six to ten degrees with high interindividual variance. This is due to many factors, including sparse calibration data, differences in human eye anatomy, and the introduction of head posture to complicate the model[3]. In addition, unrestricted head motion is crucial for the generalization of gaze-tracking, and gaze trackers that improve prediction accuracy by fixing the head tend to have a very narrow application in reality[4-5].

Advanced machine learning techniques are applied to this field. Recently, Huang et al.[6] used a residual network for feature extraction of eye images and treated the gaze difference as auxiliary information to improve the prediction accuracy. Zhuang et al.[7] proposed to use an attention mechanism to enhance the network effect and obtained excellent performance in a multi-camera multi-screen system.

*Corresponding author, E-mail address: nuaa_lyh@nuaa.edu.cn.

How to cite this article: HU Jiahui, LU Yonghua, LIU Jiangwei, et al. Machine learning-based gaze-tracking and its application in quadrotor control on mobile device[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2023, 40(5): 547-554.

http://dx.doi.org/10.16356/j.1005-1120.2023.05.004

Nagpure et al.[8] proposed a compact model to accurately and efficiently solve the problem of gaze estimation by using a multi-resolution fusion transformer and improve the network performance. However, these large or complex inference process models make these technologies almost impossible to deploy on edge processors and mobile devices. In addition, easy personalization of the model is necessary for the application scenarios corresponding to this paper.

The practical application of gaze-tracking technology has always been a vexing problem. Applications of this technology in fields such as psychology and cognition began more than a decade ago, but there are not many studies or products that use gaze information to drive mobile robots, especially in the field of eye-gaze driven quadrotors.

In an earlier study, Hansen et al.[9] combined eye-gaze drive and a keyboard to control the quadrotor, but the gaze was only able to control two degrees of freedom (DOF) of the quadrotor, and it still could not get rid of the keyboard. Kim et al.[10] combined gaze-tracking and brain-computer interfaces to control quadrotors and obtained good results, but this work can only control a single DOF of the quadrotor at the same moment, and complex wearable devices seriously limit the diffusion of this control method.

A novel object detection-based multi-rotor micro aerial vehicle (MAV) localization method in a human sensor framework has been proposed in recent years, which uses gaze to assist the quadrotor for spatial localization, but does not directly control the motion of the quadrotor and still uses a head-mounted gaze-tracking device[11].

Wang et al.[12] proposed GPA-teleoperation, an assisted teleoperation framework for gaze-enhanced perception that enables intent control and improves safety, but the wearing of VR glasses and the many requirements for quadrotor systems limit the application scenarios of this technology.

To enhance the role of eye-gaze drive in real life, we apply the proposed gaze-tracking network to mobile devices. Therefore, this research work aims to develop a simple, easy-to-use, non-wearable, and low-cost gaze-tracking platform that interprets eye movements and enables real-time control of quadrotors in 3D environments.

Therefore, the contribution of this study is to address the limitations of previous systems in a single system and provide the user with an additional, complete, and safe method of quadrotor control. The main contributions of this work are as follows:

(1) A machine learning-based monocular gaze-tracking technique is proposed and deployed on mobile devices to improve the application prospects of eye-gaze drive.

(2) An easy-to-learn and easy-to-use system: Users can convert their gaze information into control information for mobile robots in 3D space.

(3) A non-intrusive, portable, low-cost device: Users can plan the flight trajectory of the quadrotor by gaze.

1 System Overview

In this section, we discuss the hardware components and software pipeline of our system. The system's framework is shown in Fig.1, where the

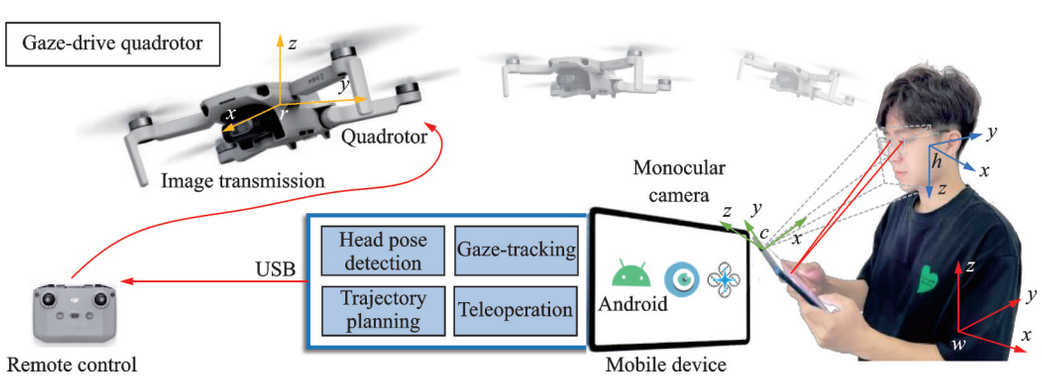


Fig.1 Illustration of controlling a quadrotor using gaze-tracking on mobile platform

green, blue, and red coordinate systems represent the camera coordinate system, the head coordinate system, and the world coordinate system, respectively. This system needs to deal with the relationship between these coordinate systems.

1.1 Hardware setup

Our novel system is based on HONOR V7, an inexpensive Android tablet. This device is chosen because its front-facing camera is located in the middle of the long side of the screen for gaze-tracking. It has a MediaTek 1300T CPU that is capable of achieving the computing power needed for machine learning. The controlled object is DJI Mini2, a small quadcopter drone with a two-axis gimbal, a takeoff mass of less than 249 g, a maximum flight time of 31 min, support for satellite positioning and optical flow positioning, real-time image transmission at the maximum bit rate of 8 Mb/s.

1.2 Algorithm pipeline

As shown in Fig.2, we used the TNN inference framework provided by Tencent to provide a variety of different acceleration options for the mobile terminals on the premise of ensuring uniform models and interfaces. The optimized adaptation of face recognition and head pose detection based on the single shot multibox detector (SSD) machine learning model is finally achieved, and the computing speed of 50 Hz is reached for 1080P images.

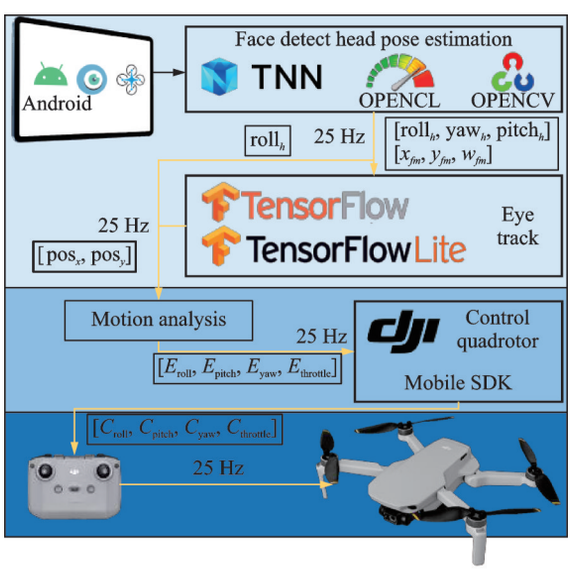


Fig.2 Diagram of our control system architecture

Using the OpenCV and OpenCL libraries, the human eye image is cropped and transmitted together with the head pose and head position information to the gaze-tracking module. The Tensorflow library is used to build the gaze tracking module proposed in this paper, and the TensorflowlLite library is used to convert it into a mobile device-compatible model (.tflite) for inference.

The result of the gaze-tracking model inference is an estimation of the user's gaze point on the tablet screen at a rate of 25 Hz. And then the estimation of the gaze point is input to the motion analysis program module to get the expected value of the quadrotor motion, and the result is input to the quadrotor control module to get the actual amount of flight control.

2 Method

In this section, we describe the proposed method of gaze-tracking and the method for converting gaze information into a quadrotor control signal.

2.1 2D monocular gaze tracking

In this paper, a CNN model for free-head gaze point (2D) estimation is proposed. It has the characteristics of low computational demand and fast computation, as well as good prediction accuracy, and supports free rotation of the head within a certain range. The model architecture is shown in Fig.3.

Before inference, the images captured by the front camera are processed by the face recognition model and the head pose detection model to obtain the left and right eye images, face frame and head pose. We flip one of the eye images horizontally and scale the two images to a size of 64×64. In particular, the coordinates of the upper left corner of the face frame in the image are used to indicate the position of the face relative to the screen, which is denoted by $[x_m, y_m]$. The width of the face frame is used to indicate the distance of the face relative to the screen, which is denoted by w_m. Finally, the eye images, face frame information, and head pose are fed into the three corresponding CNN channels of the network, and four fully connected layers are added at the end for obtaining the prediction results.

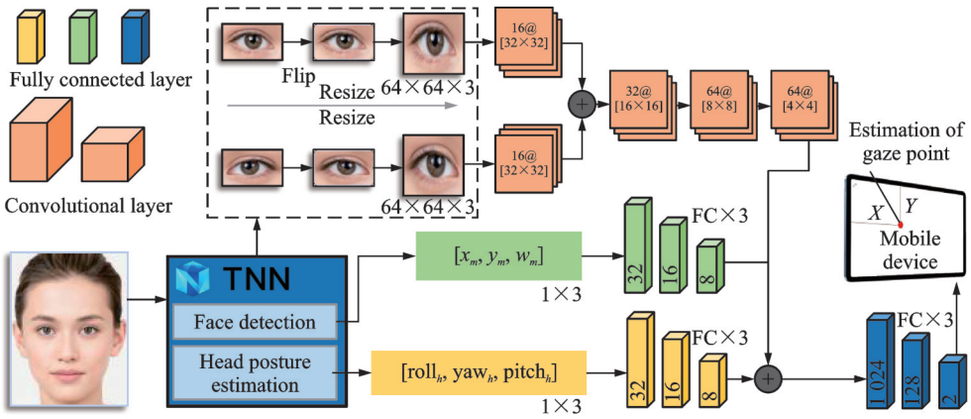


Fig.3 Our gaze point estimation network structure

In addition, we test the model performance on a generic dataset. The accuracy of the model tested on the MPIIFaceGaze dataset is 5.23 cm. It is superior to ITracker[2], Gaze-Net[13] and Mnist[1].

2.2 User interface

The user interface consists of eight parts, as shown in Fig.4, in which the view is returned by the on-board camera. The gimbal camera on the quadcopter streams the video back through the image transmission module and displays it full screen on the monitor. The transmission delay is around 200 ms, which is within the acceptable range.

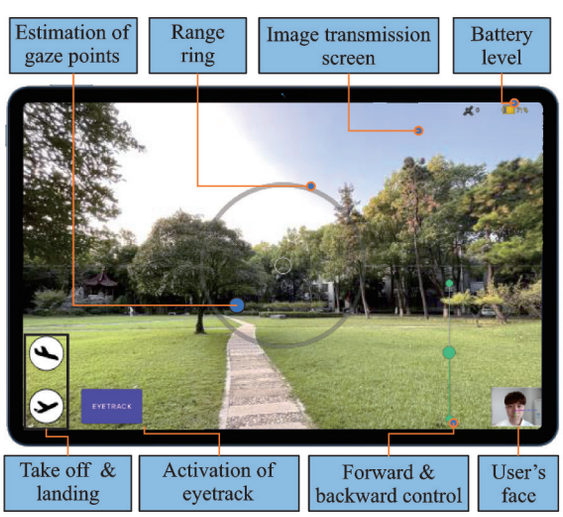


Fig.4 Components of the user interface

There is a small box showing a face in the bottom right corner of the interface, allowing the user to determine whether they have the tablet in a reasonable position. We display the results of gaze-tracking (the user's gaze point on the tablet screen) as a blue dot in the interface. The role of the distance ring is to limit the effect of the eye-gaze drive. The user can realize eye drive when the estimated result of the gaze point is outside the distance ring, otherwise the control of the quadrotor will not be triggered.

Another prerequisite for starting eye-gaze control is that the activation button in the bottom left corner of the interface is pressed. To ensure the security of the control, the user needs to keep the button pressed. Note that the quadcopter's DOF in the forward and backward directions are controlled manually. The forward speed of the quadcopter is adjusted by sliding up the green slider in the lower right corner, while sliding down the slider has the opposite effect.

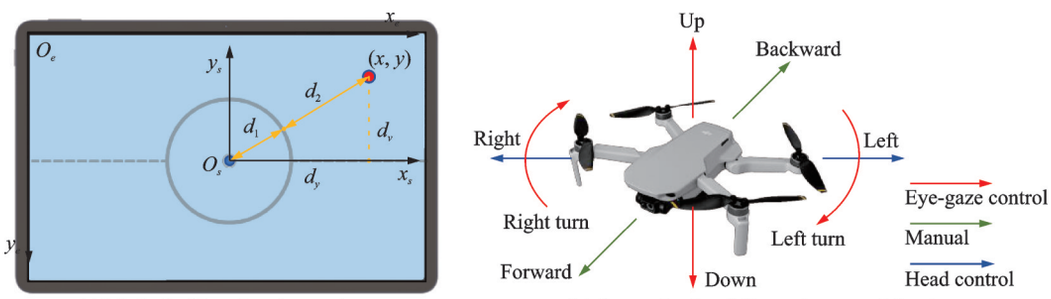
2.3 Quadrotor flight control

In this work, the predicted result of the gaze-tracking model is the user's gaze point (x, y) on the tablet display. Since 2D gaze-tracking is used, the quadrotor can only be controlled simultaneously by the human eye in two DOF of motion.

By summarizing previous research works, we find a better mapping logic: (1) The motion of gaze in the vertical direction maps to the motion of the quadrotor in the altitude direction. (2) The motion of gaze in the horizontal direction maps to the motion of the quadrotor in the yaw direction. We believe that such a mapping method is the most intuitive and more in line with the user's operation habits.

Because the motion of the quadrotor in the vertical direction and its yaw have been determined by the gaze direction, other control methods are needed to determine the motion of the quadrotor in other directions.

We use the roll angle of the head to determine the roll angle of the quadrotor, and use the slider on the interface to control the movement of the quadrotor in the forward and backward directions. The overall control method is shown in Fig.5.



(a) Method of interface interaction (b) Constrains for different degrees of freedom

Fig.5 The overall control method

We first introduce the implementation of gaze control of the quadrotor motion in the vertical and yaw directions. In Fig.5(a), the blue gaze point is located outside the distance ring with coordinates (x, y), so it can trigger eye-gaze drive.

Let the radius of the distance ring be d_1, the distance from the gaze point to O_s is $d_1 + d_2$, the distance from the gaze point to axis x_s is set to d_v, and the distance from the gaze point to axis y_s is set to d_y. Because O_s is the midpoint of the screen and the resolution of the screen is $2\,560 \times 1\,600$, $d_v = 800 - y_g$ and $d_x = 1\,280 - x_g$.

The values of d_v and d_x reflect the user's expectation on the direction of the quadrotor motion. The larger the d_v, the larger the quadrotor motion in the vertical direction should be, and the larger the d_x, the larger the quadrotor motion in the yaw direction should be.

We use C_v and C_y to represent the value of user control over the quadrotor in the vertical and yaw directions, so when d_2 is larger than 0, $C_v = \theta_1 d_v$ and $C_y = \theta_2 d_x$. The coefficients θ_1 and θ_2 indicate the control rate.

The movement of the quadrotor over the roll angle is controlled by the roll of the user's head, which is denoted by roll_h. The user's head angle is detected by the SSD machine learning model. With the head tilted to the left, the quadrotor flies to the left, and the opposite to the right.

We use C_r to represent the value of user control over the quadrotor in the roll angle direction, so $C_r = \theta_3 \text{roll}_h$. The coefficient θ_3 indicates the control rate.

As mentioned above, we manually control the forward and backward of the quadrotor, and the slider on the user interface helps us to achieve this purpose. In this research, the quadrotor is controlled simultaneously by gaze, head pose, and manual. Fig.5(b) shows the functions achieved by each control method.

3 Experiments

In order to conduct flight control experiments, an adequately large physical space is required. We set up the experimental environment in an open area of the school. Fig.6 illustrates the layout of the physical environment.

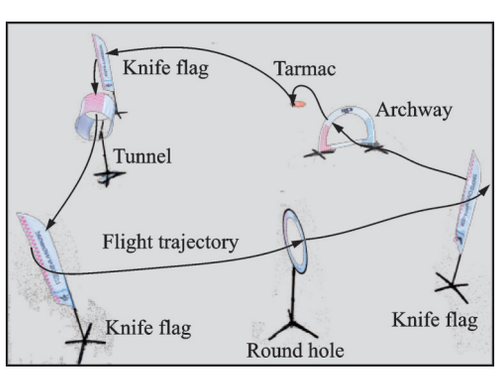


Fig.6 Test grounds with multiple obstacles

3.1 Experimental setup

We place four types of obstacles in the field, six in total: three knife flags, a tunnel, a round hole, and an archway. Subjects are asked to turn their backs to the field and steer the quadcopter from the tarmac and back through each obstacle. They are not allowed to directly observe the field, and could only adjust the quadcopter's flight conditions via video streams from the quadcopter's onboard camera.

In this experiment, each subject is required to control the quadrotor using a joystick and the proposed control method (eye-gaze drive).

3.2 Performance evaluation

To evaluate the effect of eye-gaze drive quadrotors, we set up the following evaluation methods with Ref.[10]: Flight distance, total time, and smooth curve deviation. Our goal is to test whether the proposed system can adequately convert gaze information into control information for the quadrotor, improve the control of the quadrotor, and thus replace the traditional joystick with eye-gaze drive.

To compare the manipulation efficiency of the two control methods, we calculate the total time (TT) and flight distance (FD) of subjects for each completed task.

The smooth curve deviation (SCD) can reflect the smoothness of the quadrotor flight path, as shown in Fig.7. By processing the real flight path, we can get the smoothed path. p_i is the point on the real path at time i, p_i^s is the point on the smoothed path at time i. Therefore, the SCD is calculated as

$$\mathrm{SCD}=\frac{1}{n}\sum_{i=1}^{n}|p_i-p_i^s|$$

where n is the number of quadrotor trajectory points. The quadrotor records its position once every 0.1 s.

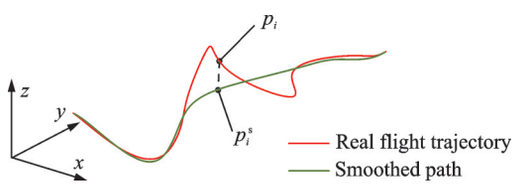


Fig.7 The smooth curve deviation

4 Results and Discussion

In this section, we analyze and compare the effectiveness of the two control methods. We collect data from five subjects, and for each control method, each subject has 20 opportunities. And the average test results are shown in Table 1.

Table 1 The summarized performance of two control methods

Subject	Joystick			Eye-gaze control		
	TT/min	FD/m	SCD/m	TT/min	FD/m	SCD/m
1	0.97	24.67	13.65	1.12	23.02	12.97
2	1.46	29.77	16.21	1.69	27.31	15.04
3	1.21	32.44	21.13	1.46	30.22	19.87
4	1.37	27.63	15.63	1.73	30.19	14.98
5	1.17	28.53	19.68	1.35	26.39	17.77

For the TT, all ten sets of data are within 2 min. The comparison reveals that all five subjects are faster in completing the flight task using the joystick than using the eye-gaze drive with average of about 15.9%. Four of the subjects show little divergence in the two control modes, but the fourth subject shows a significant difference in TT because this subject could not adapt to eye-gaze drive in a short time.

In our control system, the forward speed of the quadrotor is determined by the position of the slider on the screen. For safety reasons, we set the speed corresponding to the slider at the maximum position to be relatively small, which, we believe, is one of the reasons for the larger TT obtained by the eye-gaze control relative to that obtained by the joystick.

Generally speaking, the shorter the flight time, the shorter the flight distance, but the experimental results of FD are counter-intuitive. The FD obtained using the eye-gaze control is nearly 4.13% lower than the FD obtained using the joystick. Using eye-gaze control mode, the subject can control the UAV to complete the flight mission through a shorter flight distance. This phenomenon is difficult to understand, but combined with the experimental results of SCD, the reason can be found out.

Using the eye-gaze control, we can get lower FD and SCD, where SCD is reduced by almost 6.57%, and SCD can reflect the degree of trajectory fluctuation. This shows that although the TT obtained by this control method is larger, the flight trajectory of the controlled quadrotor is shorter and the trajectory is smoother. Therefore, we can conclude to a certain extent that the eye-gaze control method is smoother and more controllable, and the quadrotor travels a more efficient trajectory.

In the experiment, we also find that by using the eye-gaze drive, subjects are able to plan their routes more proactively based on the obstacles. Because of the reduced reliance on hand movements, subjects could focus more on the route.

The results from this study show that using gaze movements and simple body motions is still sufficient to perform a challenging task: Controlling a quadcopter in 3D physical space. The self-developed software and hardware find that an inexpensive interface is possible.

We assign two DOF of the quadrotor to the eye to achieve intuitive gaze intervention. However, the other DOF of the quadrotor still requires limb intervention, which is believed as an area in dire need of improvement.

In addition to using brain-computer interfaces or other bio-signals, we believe that with the interface setup, the eye is capable of controlling the quadrotor flight alone.

5 Conclusions

We present a mobile platform-based gaze interaction system that tracks eye movements while converting gaze information into control information for a quadrotor. The proposed interaction enables the user to manipulate the quadrotor through the eyes to accomplish complex flight tasks in 3D space. With this low-cost and mobile device, people can control their flying machines naturally and easily in their daily lives. From the results of our study, we have succeeded in demonstrating the potential of this interaction method. We believe that our solution can expand new ways of human-computer interaction and create a new dimension of quadrotor control.

References

[1] ZHANG X, SUGANO Y, FRITZ M , et al. Appearance-based gaze estimation in the wild[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2015: 4511-4520.

[2] KRAFKA K, KHOSLA A, KELLNHOFER P, et al. Eye tracking for everyone[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 2176-2184.

[3] AKINYELU A A, BLIGNAUT P. Convolutional neural network-based methods for eye gaze estimation: A survey[J]. IEEE Access, 2020, 8: 142581-142605.

[4] MORA K A F, MONAY F, ODOBEZ J M. EYEDIAP: A database for the development and evaluation of gaze estimation algorithms from RGB and RGB-D cameras[C]//Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA' 14). New York: [s.n.], 2014: 255-258.

[5] SUGANO Y, MATSUSHITA Y, SATO Y. Learning by synthesis for appearance-based 3D gaze estimation[C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 1821-1828.

[6] HUANG L, LI Y, WANG X, et al. Gaze estimation approach using deep differential residual network[J]. Sensors, 2022, 22: 5462.

[7] ZHUANG J, WANG C. Attention mechanism based full-face gaze estimation for human-computer interaction[C]//Proceedings of 2022 International Conference on Computer Network, Electronic and Automation (ICCNEA). Xi'an, China: [s.n.], 2022: 6-10.

[8] NAGPURE V, OKUMA K. Searching efficient neural architecture with multi-resolution fusion transformer for appearance-based gaze estimation[C]//Proceedings of 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA: IEEE, 2023: 890-899.

[9] HANSEN J P, ALAPETITE A, MACKENZIE I S, et al. The use of gaze to control drones[C]//Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA' 14). New York: [s.n.], 2014: 27-34.

[10] KIM B H, KIM M, JO S. Quadcopter flight control using a low-cost hybrid interface with EEG-based classification and eye tracking[J]. Computers in Biology and Medicine, 2014, 51: 82-92.

[11] YUAN L, REARDON C, WARNELL G, et al. Human gaze-driven spatial tasking of an autonomous MAV[J]. IEEE Robotics and Automation Letters, 2019, 4(2): 1343-1350.

[12] WANG Q, HE B, XUN Z, et al. GPA-teleoperation: Gaze enhanced perception-aware safe assistive aerial teleoperation[J]. IEEE Robotics and Automation Letters, 2022, 7(2): 5631-5638.

[13] ZHANG X, SUGANO Y, FRITZ M, et al. MPIIGaze: Real-world dataset and deep appearance-based gaze estimation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(1): 162-175.

Acknowledgements This work was supported by the National Natural Science Foundation of China (No.51975293), the Aeronautical Science Foundation of China (No. 2019ZD052010), and the Zhangjiagang Pre-research Fund of China (No.ZKCXY2101).

Authors Mr. **HU Jiahui** is currently a Ph.D. candidate in Nanjing University of Aeronautics and Astronautics (NUAA). His main research interests include eye-tracking and eye-driving.

Prof. **LU Yonghua** received his Ph.D. degree in 2005 from NUAA, now he is a professor in NUAA. His main research interests include intelligent measurement and control, measurement system and robotics.

Author contributions Mr. **HU Jiahui** designed the study, complied the models, conducted the analysis, interpreted the results and wrote the manuscript. Prof. **LU Yonghua** contributed to the data and model components for the gaze-tracking model. Dr. **LIU Jiangwei** contributed to the hardware processing. Dr. **YAN Changkai** contributed to the discussion and background of the study. Dr. **LIU Tao** contributed to the software and hardware design. All authors commented on the manuscript draft and approved the submission.

Competing interests The authors declare no competing interests.

(Production Editor: XU Chengting)

基于机器学习的视线追踪方法及其在移动端四旋翼控制的应用

胡佳辉[1], 陆永华[1], 刘江伟[2], 严长凯[2], 刘 韬[1]

(1.南京航空航天大学机电学院, 南京 210016, 中国; 2.中国航发控制系统研究所, 无锡 214000, 中国)

摘要:提出了一种基于机器学习的移动设备单目视线追踪技术。这种非侵入、便捷、低成本的注视追踪方法可以实时捕捉用户在移动设备屏幕上的注视点。结合四旋翼飞行器的三维运动控制,将用户的视线信号转化为四旋翼飞行器的控制信号,解决了以往控制方法的局限性,使用户可以通过视觉交互操控飞行器。为了明确该方法的可行性,本文设置了一条复杂的四旋翼飞行赛道。受试者被要求将视线介入到控制流中,以完成飞行任务。通过与基于操纵杆的控制方法进行比较,对飞行性能进行了评估。实验结果表明,本文方法能提高四旋翼飞行器运动轨迹的平滑性和合理性,并能为四旋翼飞行器控制引入多样性、便利性和直观性。

关键词:视线追踪;无人机控制;机器学习;人机交互;眼球驱动