Aim Where You Look: Eye-Tracking-Based UAV Control Framework for Automatic Target Aiming

Minqiang Yang, *Member, IEEE*, Jintao Wang, Yujie Gao, and Bin Hu, *Fellow, IEEE*

***Abstract*—Unmanned aerial vehicles (UAVs) often require real time and accurate control in military, rescue, and transportation applications. There is an urgent demand from operators for usability and interoperability. Human–machine collaboration enhancement and tiny machine learning aimed at edge computing hold the potential to greatly improve the operational efficiency of UAVs. In this article, we propose a novel UAV control framework based on eye-tracking technology, i.e., eye-tracking-based UAV (ETUAV), which combines with lightweight object detection to assist the control of the UAV for attitude adjustment through the gaze information of the UAV operator. The ETUAV control method we propose involves real-time tracking of the operator's gaze using our self-developed head-mounted eye tracker, combined with object detection to achieve automatic targeting. We also propose an incremental proportion integration differentiation (PID) control algorithm for adjusting the UAV's attitude, which provides automatic and real-time UAV control. We develop a system prototype based on the DJI UAV and conduct performance benchmarks and comparisons. The experimental results indicate that the operational latency of our method is significantly less than manual operation and the built-in automatic tracking function in the DJI application. The "Aim Where You Look" UAV operation framework proposed in this article significantly streamlines the tasks for operators in time-critical scenarios.**

***Index Terms*—Eye-tracking, incremental proportion integration differentiation (PID), unmanned aerial vehicle (UAV).**

I. Introduction

UNMANNED aerial vehicles (UAVs) are aircraft operated using radio remote control devices and autonomous programmable control systems, which have a wide range of applications covering many different fields, such as agriculture [1], communication [2], security [3], forestry [4], mapping [5], film and media [6], mobile edge computing (MEC) [7], [8], and logistics [9]. Traditional UAV control involves the operator using control sticks to adjust the UAV's flight and attitude, and the operator's proficiency greatly influences the efficiency and accuracy. Some special or complex scenarios that require the real time and precise operation of UAVs, such as search and rescue [10], defense and military [11], and the management of UAV swarms, which pose significant challenges to operators. The challenges include as follows.

1) Proficient mastery of UAV control is required, with the ability to operate UAVs accurately within a short timeframe.
2) Manual operation demands simultaneous and coordinated use of both hands, leaving the operator with no time for other tasks.

Although some advanced UAVs offer automatic tracking, setting the tracking target still requires manual configuration, and there are still cumbersome operations involved when switching targets. Sensing human intentions through physiological signals or behavioral data to achieve real time and efficient control of UAVs has emerged as a pioneering research area [12], [13]. This innovative approach aims to enhance the capabilities and efficiency of UAVs by leveraging real-time data from human operators to adapt and optimize UAV operations. The fundamental premise of this research lies in the synergy between human operators and UAVs, intending to create a more intuitive and responsive control system. By monitoring and analyzing human physiological signals, such as electroencephalography (EEG) [14] and electrooculography (EOG) [15], as well as behavioral data like gaze direction, hand movements, and vocal commands, UAVs can better understand and anticipate the intentions and needs of their operators.

In the application of UAVs, the collaboration between humans and UAVs has evolved beyond simple control to become more intelligent. To solve the problem of criminals evading detection and capture, Zheng et al. [16] constructed a human-UAV collaborative search system for fugitive criminals and proposed a hybrid evolutionary algorithm, which was shown to have the advantages of practicality and performance. Zhou and Liu [17] demonstrated a hybrid human–machine augmented tracking system for improving the ability of UAVs to track people in complex environments. Through an eye-movement interaction paradigm, humans can guide and refine UAV tracking by providing guidance and making necessary corrections. Object detection using UAVs has garnered substantial interest in various fields, such as disaster rescue [18], traffic management [19], and search operations for missing individuals [20]. Additionally, UAVs are employed for

Manuscript received 22 November 2023; revised 26 January 2024 and 18 March 2024; accepted 3 April 2024. Date of publication 17 April 2024; date of current version 7 June 2024. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFA0706200; in part by the National Natural Science Foundation of China under Grant 62227807; in part by the Natural Science Foundation of Gansu Province, China, under Grant 22JR5RA488; in part by the Fundamental Research Funds for the Central Universities under Grant lzujbky-2023-16; and in part by the Supercomputing Center of Lanzhou University. *(Corresponding authors: Minqiang Yang; Bin Hu.)*

The authors are with the School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China (e-mail: yangmq@lzu.edu.cn; bh@lzu.edu.cn).

Digital Object Identifier 10.1109/JIOT.2024.3390115

location estimation [21], further enhancing their utility and effectiveness.

The integration of edge computing with artificial intelligence models significantly enhances human–machine interaction in the context of Internet of Things (IoT) environments [22], [23], [24]. This article combines visual attention and object detection based on UAVs to explore more efficient operations. By using eye-tracking and object detection to identify visual points of interest and adjust the UAV's attitude, it is expected to achieve real time, precise control, overcoming the delays and errors associated with manual operation. This has significant applications in special scenarios like emergency rescuing or battlefield. The main contributions of this article are as follows.

1) We design a novel eye-tracking-based UAV (ETUAV) control framework and implement a prototype based on a DJI UAV. This framework enables rapid and accurate control of the UAV through eye fixation, thereby facilitating target localization.
2) We propose a field-of-view transformation (FVT) method, which converts the coordinates from the UAV's field of view to those of the eye tracker's field of view. This method solves the problem of object detection failure caused by blurry and reflective images captured by the eye tracker. Additionally, this method mitigates the errors introduced during the control process by combining eye fixation points with object detection.
3) We propose an UAV control algorithm based on incremental proportion integration differentiation (PID). The inclusion of differential and integral control terms based on the error value allows the UAV to be efficiently and accurately adjusted.
4) We conduct comparison experiments with the manual control mode and the intelligent following control mode of the DJI official app. The experimental result shows that our designed eye-movement control UAV framework is outstanding in terms of efficiency and accuracy, which is superior to the traditional manual control of the UAV as well as the official DJI control program.

II. Related Work

Intent perception-based UAV control via brain–computer interface (BCI) technology allows for more reliable and precise operations, which has gained widespread research attention [25]. Coogan and He [26] combined BCI with virtual reality (VR) and IoT technologies to provide real-time control of the user's virtual and physical environment. A growing body of research demonstrates the feasibility of BCI technology to control UAVs. BCI control of UAVs can simplify the control process and achieve better accuracy and real-time performance, especially when combined with IoT technology. Chung et al. [27] introduced the steady-state visual evoked potential (SSVEP) method to control UAVs with less error correction and improved accuracy. Deng et al. [28] integrated BCI technology and VR to create an intelligent group control method for UAV swarms, addressing the limitations of traditional ground station control and offering new paradigms for swarm UAV management. Shi et al. [29] proposed a noninvasive hybrid computer interface system utilizing EOG and EEG for indoor target searching. In [30], the brain patterns of an user while performing different motion-image tasks were analyzed using a motion-image-based BCI and applied to the navigation of a simulated UAV.

Gesture recognition is an integral component of human–computer interaction (HCI) and IoT systems [31]. Gesture control is a technology that enables device or system control through the recognition and interpretation of hand movements by individuals. It offers an intuitive, flexible, and versatile means of interaction, thereby enhancing the quality and effectiveness of the overall user experience. In the work presented in [32], a gesture recognition method is proposed for UAV control. This method exhibits remarkable accuracy in recognizing a wide range of gestures and efficiently translates them into corresponding control commands for the UAV. As a result, real time and stable UAV control is achieved, highlighting the effectiveness of the proposed approach. Yeh et al. [33] constructed an intuitive gesture recognition control system based on deep learning, and outputs labels as control commands for the UAV.

Eye movement is an intuitive reflection of human visual search and attention allocation. The eye movement events, such as blinking, saccades, fixation, and pupil dilation can not only be used to detect affective disorders [34], [35] but also can be coded as control signaling. Eye tracking can be used as input for HCI systems to enhance real-time performance and usability of the system [36] [37]. Jun et al. [38] designed a system of controlling UAVs with eye movement patterns. Karas et al. [39] presented a novel BCI approach to control robots with eye artifacts for assistive applications. In this work, the eye movement events, such as blinks are detected to auxiliary control. Wu et al. [40] provided a command and control technique that combines a BCI with eye tracking. It can assist UAV controllers in command selection, thus improving the accuracy of brain–computer signals relying on visual stimuli. To solve the problem of UAVs performing multidimensional motions in 3-D space and to improve control accuracy, Wu et al. [41] proposed an UAV control system that combines EEG, EOG, and head attitude information with good control accuracy and high information transfer rate (ITR).

III. Proposed Method

A. Framework Overview

In this article, we propose a novel framework for controlling the UAV by eye movement automatically, i.e., ETUAV, and implement a prototype with our own designed eye tracker and DJI UAV. Fig. 1 shows a quick overview of the proposed method, which allows the operator to control the UAV by gaze direction automatically. Our proposed framework leverages wearable eye-tracking technology and object detection to predict the target. When the operator's visual focus shifts from the building to the lady, the wearable eye-tracking device worn by the operator can predict his gaze direction. Combined with object detection, the new target of visual focus is identified. UAV attitude adjustments are made by

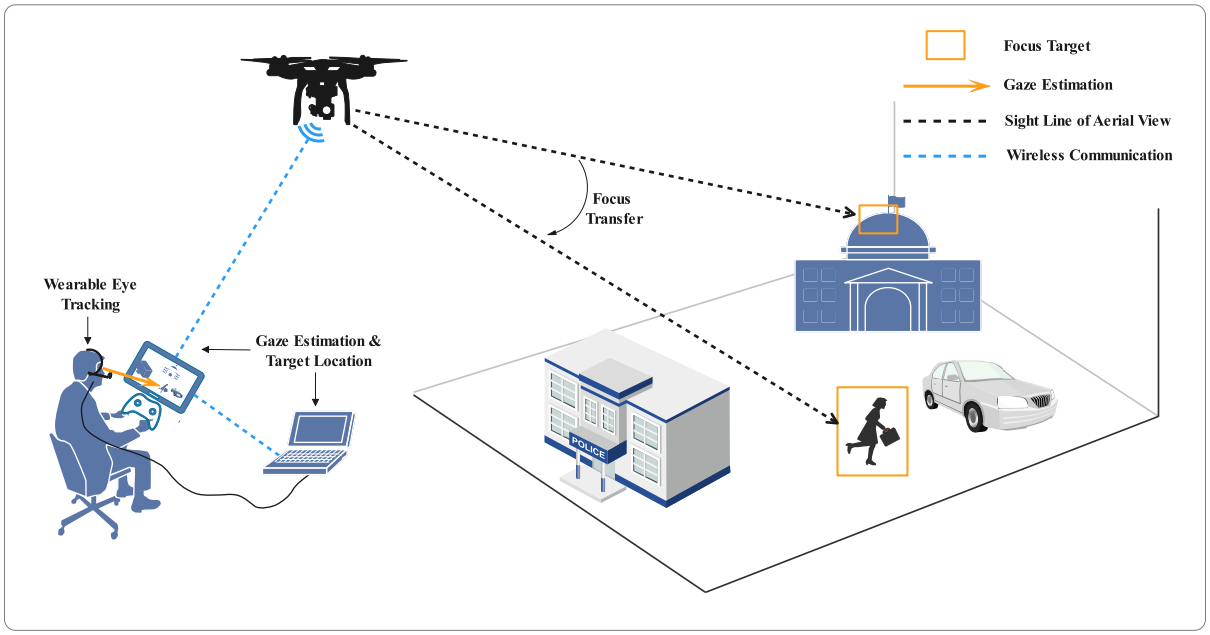

Fig. 1. Overview of the proposed method. The core purpose of the method is to implement Aim Where You Look by eye-tracking and object detection.

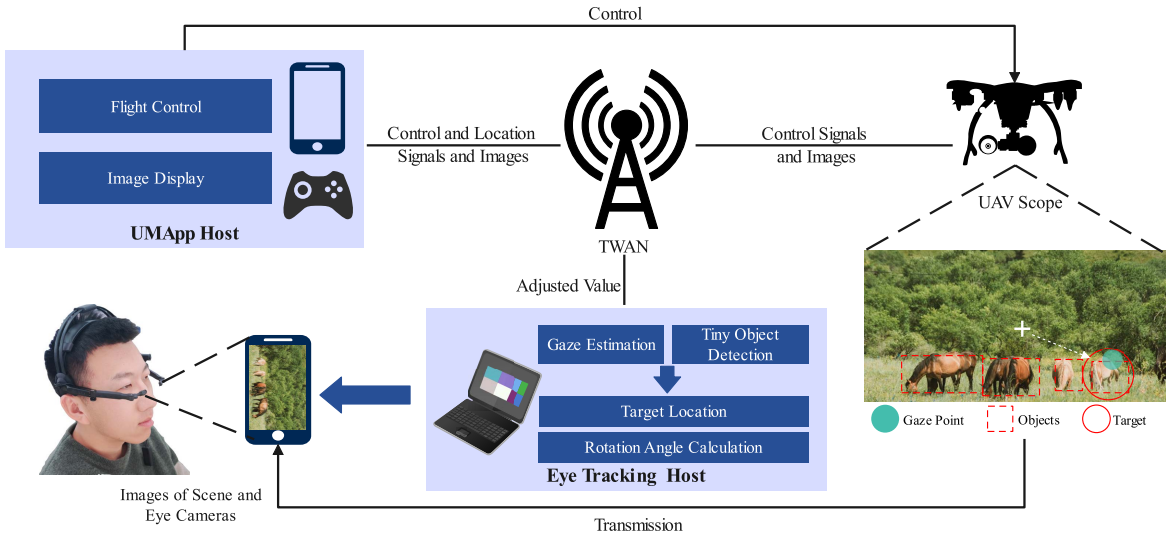

Fig. 2. Framework of the proposed method.

calculating the relative positions of the new and old targets within the scope of the UAV. This approach significantly reduces the time required for UAV attitude adjustment while maintaining a high level of accuracy. It is well-suited for assisting nonprofessional operators in time-critical situations, providing them with valuable support.

The overall ETUAV framework is divided into three parts, which are the eye tracking host, UMApp host, and UAV. The terminals in the framework exchange data and control signal through a two-tier wireless ad-hoc network (TWAN) formed by UAV, remote controller, eye tracking host, and UMApp host. Through our proposed FVT method and a closed-loop control algorithm based on incremental PID, the system achieves the automatic targeting of the UAV on the designated objective. The framework of the proposed method is shown in Fig. 2.

1) UMApp Host: The UMApp host consists of a mobile terminal, an UAV remote controller, and a mobile application. The mobile application mentioned in the framework, i.e., unified fly (UFly), is capable of receiving data information from TWAN and generating control commands. UFly needs to be installed on a smartphone or tablet and connected to the remote controller via a wired or wireless link. The main functionality of UFly is to display real-time images captured by the UAV, receive adjustment values from the network, generate control commands, and control the UAV's corresponding attitude adjustments. During the TWAN communication process, it additionally transmits both the location information and real-time image of the UAV.

2) Eye Tracking Host: The eye tracking host is a computer that is connected to a head-mounted eye tracker. The eye tracker consists of three cameras, with two eye cameras

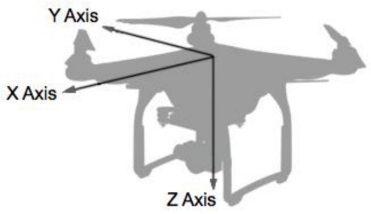

Fig. 3. Body coordinate system structure of the UAV.

located in the front for pupil detection, and another camera for capturing the operator's field of view. In the eye tracking host, the eye tracker captures the operator's fixation data and transmits it to the computer. The eye tracking host serves as the server in the TWAN and performs tasks, such as target detection and adjustment value calculation. We deploy a tiny model on the computer to detect the targets in the images captured by the UAV and calculate the coordinates of the target that the operator is looking at. After obtaining the coordinates of the target, we need to calculate the deviation angle based on the current attitude and feedback from the UMApp host.

3) Overall Process: In the preparation phase, the operator needs to connect the mobile terminal with UFly installed to the UAV's remote controller and make sure that the UFly and the eye tracking host are connected through TWAN. The operator then puts on the eye tracker and performs adjustment and calibration. During the control phase, the operator simply gazes at the target on the UFly interface and clicks on the targeting button to perform automatic targeting. The UFly communicates with the eye tracking host several times until the UAV is adjusted in place. The UFly sends real-time attitude information and imagery of the UAV and receives the adjustments computed by the eye tracking host to generate control commands to maneuver the UAV. The eye tracking host receives the data sent by the UFly and processes it using target detection, eye tracking techniques, FVT methods, and incremental PID control algorithms to generate adjustments value sent to the UFly.

B. Interactive UAV Control by Eye Movement

For eye-controlled UAV operation, it is essential to understand the rotational dynamics of the UAV. The three vertical axes in the body coordinate system of the UAV are illustrated in Fig. 3. The body coordinate system is relative to the UAV, with the center of gravity as the origin. The x-axis points forward, the y-axis passes through the right side of the UAV, and the z-axis passes through the bottom of the UAV using the right-hand rule. When describing rotational motion, the x-axis corresponds to the roll axis, the y-axis to the pitch axis, and the z-axis to the *yaw* axis. The UAV rotates around these respective axes, generating roll, pitch, and yaw angles that are positive in the clockwise direction.

In the framework we proposed, by rotating the UAV in two directions, yaw control and pitch control, it becomes possible to lock onto a target in 3-D space, achieving automatic targeting of the UAV. We retain the yaw control of the UAV and use the pitch attitude of the UAV gimbal camera instead of the pitch attitude of the UAV itself. Since the UAV and the gimbal camera are connected, it is possible to use the camera's pitch instead of the UAV's pitch, making the UAV more stable when operating the UAV. Next, the detailed process of eye movement control will be introduced in three parts: 1) FVT; 2) deviation angle calculation; and 3) incremental PID control algorithm.

1) FVT Method: In practical applications, two significant challenges arise. The images captured by the eye tracker's scene camera can suffer from blurring and distortion, a consequence of screen reflections from the mobile terminal. Additionally, sole reliance on gaze point data for UAV adjustment leads to inaccuracies, affecting its reliability. To overcome these issues, we propose the FVT method, consisting of a three-step approach. The initial phase detects the target in the scene image. This is followed by a transformation of the image's coordinate information. The final phase leverages gaze data to pinpoint the object of interest. In this procedure, the image captured by the UAV is termed the first field-of-view image (FFVI), while the image taken by the scene camera of the eye tracker is named the second field-of-view image (SFVI). Fig. 4(a) represents the SFVI and Fig. 4(b) depicts the FFVI. The internal box of Fig. 4(a) shows the image taken by the UAV, which is displayed on the screen of the UMApp host. Fig. 4(b) shows the image sent to the eye tracking host by the UMApp host.

In the initial phase, we substitute SFVI with FFVI as the model's input. The FFVI, leveraging high-definition camera images from the UAV, presents numerous advantages over SFVI. Primarily, FFVI yields clearer and more detailed imagery, facilitating the model's precision in object detection and localization. Furthermore, the enhanced contrast in FFVI aids in distinguishing target objects from their background, significantly improving the model's object detection capabilities. Additionally, FFVI circumvents prevalent issues like reflection and distortion common in SFVI, ensuring a more stable and reliable image input for the model. This shift notably bolsters the model's performance and accuracy in detecting objects. As illustrated in Fig. 4(b), the model processes FFVI to determine the coordinates of the upper-left and lower-right corners of multiple objects. Considering the rightmost object as the target, the model computes its coordinates (x_4, y_4) and (x_5, y_5), along with the FFVI's length v_u and width u_u.

In the second step, our objective is to map the coordinate information from the FFVI to the SFVI, ensuring their alignment within an unified coordinate system. This process necessitates converting the coordinates of each object in FFVI to their corresponding values in SFVI. Before executing this coordinate mapping, it is essential to identify the UMApp host within the SFVI and ascertain its bounding box. Following this, the coordinates derived from the FFVI are mapped to fit within this identified bounding box. For instance, considering our target object, we must compute its SFVI coordinate values (x_2, y_2) and (x_3, y_3). The formula for this calculation is as follows:

$$(x_2, y_2) = \left(\frac{x_4 \cdot v_e}{v_u} + x_1, \frac{y_4 \cdot u_e}{u_u} + y_1 \right) \tag{1}$$

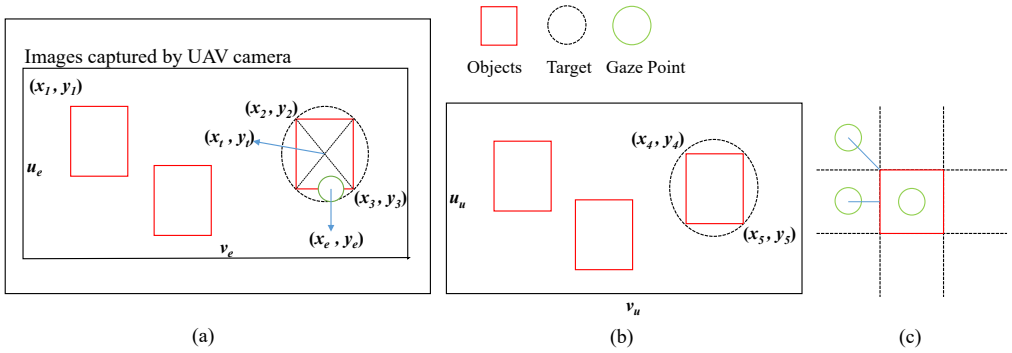

Fig. 4. Schematic of FVT method. (a) Images captured by eye tracker. (b) Images captured by UAV camera. (c) Distance calculation.

$$(x_3, y_3) = \left(\frac{x_5 \cdot v_e}{v_u} + x_1, \frac{y_5 \cdot u_e}{u_u} + y_1 \right) \tag{2}$$

where (x_4, y_4) and (x_5, y_5) represent the coordinate values in the FFVI obtained in the initial phase, respectively. v_u and u_u denote the length and width of the FFVI, while v_e and u_e represent the length and width of the image displayed on the UMApp host in the SFVI. Finally, (x_1, y_1) denotes the coordinates of the starting vertex of this image in the SFVI.

In the third step, we integrate gaze data with object information to define the UAV's adjustment target. Due to the inherent inaccuracies in the eye tracker's gaze point data, relying exclusively on these gaze points for UAV adjustments can lead to significant errors. As depicted in Fig. 4(a), the gaze point of the UAV operator in the SFVI is represented by (x_e, y_e). Our goal is to identify the object in the SFVI nearest to this gaze point as our primary interest. Continuing with our initial assumption, the rightmost object is selected as our focal target. We then compute its geometric centroid (x_t, y_t), which serves as the final adjustment point for the UAV. Illustrated in Fig. 4(c), by extending the edges of its bounding box, we divide the image into nine sections. This spatial arrangement of the expanded image regions relative to the gaze point allows us to categorize the detection process into several scenarios.

1) If the gaze point is inside or on the edges of the rectangular area, the object within this region is immediately identified as the target.
2) Should the gaze point be within the top, bottom, left, or right sections surrounding the rectangle, we calculate the Euclidean distance from the gaze point to the nearest edge of the bounding box.
3) If the gaze point is located in one of the four corner areas, the distance between the gaze point and the closest corner of the target on the bounding box is determined. If scenario (1) does not occur, it becomes necessary to compute the minimum distance from the set of potential target distances. The object corresponding to this minimum distance is then identified as the final target. The calculation is executed as follows:

$$d_{\min} = \min(d_1, d_2, \ldots, d_n) \tag{3}$$

where d_i $(i = 1, 2, \ldots, n)$ is the distance between each object and the gaze point, and $d_{\min}$ is the shortest distance which is also the final distance of the target object. After obtaining the fixation object, we calculate its geometric centroid coordinates (x_t, y_t), which represent the objective coordinates to be adjusted next.

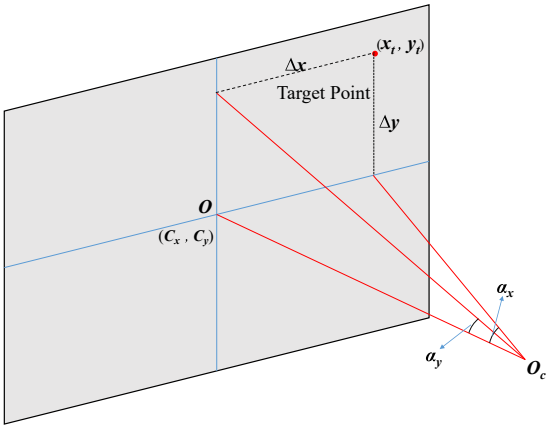

Fig. 5. Calculated deviation angle.

2) Calculated Deviation Angle: After obtaining the coordinates (x_t, y_t) of the target that the UAV operator is gazing at, we need to calculate the value of the deviation angle from the coordinates of the center point of the image (c_x, c_y) to (x_t, y_t). As shown in the imaging plan in Fig. 5, point O is the center point of the image with coordinates (c_x, c_y), the target point is the coordinates of the target of gaze, and Δx and Δy are the deviation distances between the center point of the image and the target point, respectively. O_c is the camera center. The UAV yaw angle adjustment value α_x and pitch angle adjustment value α_y are calculated as follows:

$$\alpha_x = \arctan \frac{n_x \cdot d_x}{f} + \text{err}_x \tag{4}$$

$$\alpha_y = \arctan \frac{n_y \cdot d_y}{f} + \text{err}_y \tag{5}$$

where f is the focal length of the UAV's camera. d_x and d_y correspond to the physical lengths of the pixel points along the horizontal and vertical directions, respectively. Similarly, n_x and n_y denote the number of pixel points in the horizontal and vertical directions, respectively. err_x and err_y are the error terms, which are due to camera distortion and positional distance errors.

3) Control Algorithm Based on Incremental PID: We control the rotation of the UAV based on an incremental PID control algorithm. Incremental PID control is a commonly used control algorithm to regulate the output of a system to bring it close to the desired value. It is an improved form of the traditional PID control algorithm and offers several advantages [42]. Incremental PID control controls the system by calculating the amount of output change, rather than calculating the output value directly.

A traditional PID control algorithm consists of three parts: 1) a proportional term (P); 2) an integral term (I); and 3) a differential term (D) [43]. Incremental PID control algorithms still include these three components, but they are calculated in a slightly different way. The classical PID control rule is written as

$$u_t = K_p\left(e_t + \frac{1}{T_I}\int_0^t e_t dt\right) + T_D\frac{de_t}{dt} \tag{6}$$

$$e_t = r_t - c_t \tag{7}$$

where e_t is the difference between the set target value r_t and the actual measured value c_t, K_p is the proportional gain, T_I is the time constant of the integral, and T_D is the time constant of the differential. When using a computer program to control, the differential and integral in (6) can not be used directly, we need to use the form of discretization. We replace the integral with a sum, the differential with a first-order difference, and the continuous time with a series of sampling points, as follows:

$$\int_0^t e_t dt \approx \sum_{i=0}^{k} Te_i \tag{8}$$

$$\frac{de_t}{dt} \approx \frac{e_k - e_{k-1}}{T} \tag{9}$$

where T is the sampling period and k is the sampling time. Then, replacing the integral and differential in (6) gives the sampling output for u_k and u_{k-1} as

$$u_k = K_p\left(e_k + \frac{T}{T_I}\sum_{i=0}^{k} e_i + T_D\frac{e_k - e_{k-1}}{T}\right) \tag{10}$$

$$u_{k-1} = K_p\left(e_{k-1} + \frac{T}{T_I}\sum_{i=0}^{k-1} e_i + T_D\frac{e_{k-1} - e_{k-2}}{T}\right). \tag{11}$$

Then, the two equations are differenced to obtain the PID increment Δu_k

$$\begin{aligned}\Delta u_k &= u_k - u_{k-1}\\ &= K_p(e_k - e_{k-1}) + \frac{K_p T}{T_I}e_k\\ &\quad + \frac{K_p T_D}{T}(e_k - 2e_{k-1} + e_{k-2})\\ &= K_p(e_k - e_{k-1}) + K_I e_k + K_D(e_k - 2e_{k-1} + e_{k-2})\end{aligned} \tag{12}$$

where K_I represents the integral coefficient, K_p represents the proportional coefficient, and K_D represents the differential coefficient.

When controlling the UAV for an attitude adjustment, we need to calculate the error between the target value V_t and the feedback value V_f each time, and after setting the three parameters K_p, K_I, and K_D, we use (12) to find the increment of the next angle adjustment. Algorithm 1 is the pseudocode for this UAV control. We mark the targets in the image during the initial image processing. In the subsequent phase of verifying whether the UAV has adjusted to the correct attitude, additional recognitions are required. The marking helps reduce the time spent on recognition. We use the perceptual hash algorithm (PHA) to calculate the similarity between two images to determine the target.

Algorithm 1 Control Algorithm Based on Incremental PID

Require: UAV adjusted target value Vt_x, Vt_y, UAV adjusted feedback value Vf_x, Vf_y, PID parameters K_{p1}, K_{p2}, K_{I1}, K_{I2}, K_{D1}, K_{D2}, error threshold D_t, and three times error value in x-direction and y-direction Ex_0, Ex_1, Ex_2, Ey_0, Ey_1, Ey_2

Initialize $Ex_0 \leftarrow 0$, $Ex_1 \leftarrow 0$, $Ex_2 \leftarrow 0$, $Ey_0 \leftarrow 0$, $Ey_1 \leftarrow 0$, $Ey_2 \leftarrow 0$
while true **do**
  Receive UAV location data from the mobile terminal
  if First data received **then**
    Detecting mobile terminal borders using models
    Using the FVT method to obtain the target coordinates (x_t, y_t)
    Calculate the yaw angle α_x and the pitch angle α_y
    Send UAV adjustment value α_x and α_y
  else
    Use the model to detect the objects in the image
    Secondary detection of the target using PHA
    Calculate the distance Δd from the center of the target (x_o, y_o) to the center of the image
    if $\Delta d < D_t$ **then**
      break
    end if
    $Ex_0 \leftarrow Vt_x - Vf_x$; $Ey_0 \leftarrow Vt_y - Vf_y$
    $\Delta\alpha_x \leftarrow K_{p1}(Ex_0 - Ex_1) + K_{I1}Ex_0 + K_{D1}(Ex_0 - 2Ex_1 + Ex_2)$
    $\Delta\alpha_y \leftarrow K_{p2}(Ey_0 - Ey_1) + K_{I2}Ey_0 + K_{D2}(Ey_0 - 2Ey_1 + Ey_2)$
    $Ex_2 \leftarrow Ex_1$; $Ex_1 \leftarrow Ex_0$
    $Ey_2 \leftarrow Ey_1$; $Ey_1 \leftarrow Ey_0$
    Send $\Delta\alpha_x$ and $\Delta\alpha_y$ as new adjustments
  end if
end while

C. Gaze Sensing and Target Location

In this study, we used our self-developed head-mounted eye tracker for gaze tracking, i.e., UEYE headset [44], which is shown in Fig. 6. The head-mounted eye tracker provides fast, accurate, and robust gaze acquisition. The design of the UEYE headset ensures stable wear, making it suitable for operating UAVs in a variety of environments.

Considering the possible impact of the ambient light when operating the UAVs, we add infrared lights and filters which can decrease the impact and ensure the accuracy and robustness of gaze tracking. In the pupil recognition algorithm, we use a neural network to crop out the iris region and then use a random sample consensus (RANSAC) [45]-based algorithm to

Fig. 6. Eye tracker used in this study: UEYE headset [44]. It captures eye movements through the camera positioned in front of the eyes, enabling accurate gaze tracking.

draw the pupil ellipse. Compared to the RANSAC algorithm and DeepVOG [46] based on deep learning, our algorithm takes only a quarter of the time while ensuring the same or higher precision and accuracy. We design the pupil complementation algorithm for the possible occlusion of the pupil region when operating UAVs, which makes our algorithm more robust.

Based on the acquired pupil ellipse and eye image video, we construct a 3-D two-sphere eyeball model proposed to obtain the optical axis vectors of the two eyes. The nine-point calibration correction model is then used to obtain the visual axis vectors of the two eyes, the intersection of which is the gaze point.

We use PREEMPT_RT [47] as the real-time operating system, and raise the priority of real-time subprocesses, which can ensure synchronization between camera processes. It will ensure that there will not be a large delay between cameras due to system resource utilization when operating UAVs, thus ensuring high accuracy of the gaze point.

Considering the difference in vision, the obtained gaze point can not directly localize the point in the UAV's field of view. We use the YOLOX-tiny model [48] to detect the objects in images of the UAV and match them with the gaze point. Tiny machine learning is possible to use on mobile terminals like UAV controllers. The performance of YOLOX is somewhat improved compared to YOLOV5, and the tiny model has fewer parameters for faster detection and an AP of 32.8%, which can meet the experimental requirements.

IV. Experiments

In this section, we commence with an overview of the equipment required for our experiment. Subjects were assigned to test four distinct targets using three varied control methods with the UAV in a real-world environment. Subsequently, we delve into an analysis of the data gathered during the experiment, emphasizing the time consumption and precision of the outcomes. The real experimental setup is depicted in Fig. 7. Here, the green box symbolizes the operator equipped with an eye tracker, monitoring the UAV's real-time imagery. The red box indicates the four experimental test targets, while the blue box represents the UAV itself.

A. Experimental Equipment

UAV: We chose the DJI Inspire 2 as the experiment UAV, which has a single flight endurance of 27 min, a maximum

Fig. 7. Real experimental environment scene picture.

Fig. 8. Snapshot of UFly app user interface. The red boxes represent the gaze targets at different time intervals for the subjects, and the blue arrows depict the process of the subjects' gaze targets switching during the experiment.

mapping signal distance of 7 km, and can capture 1080p images.

UMApp Host: We employed a Redmi K30 Android phone equipped with the UFly control application for our experiment. The user interface of UFly is illustrated in Fig. 8. The interface layout includes control buttons at the top row, while the bottom section displays the UAV's real-time status information. The button in the bottom right corner is specifically designated for initiating eye movement control. These top buttons are used for standard flight control, while the button in the bottom right corner transmits commands to the UAV, and directs the UAV to rotate toward the target currently under the operator's gaze.

Eye Tracking Host: In our setup, a laptop that is connected to an eye tracker serves as the eye tracking host to perform target detection and adjustment computation. The specifications of the laptop are as follows. It is equipped with an Intel i5-10210U CPU operating at 1.6 GHz, features 8 GB of RAM, and includes a Radeon 550x series GPU. The eye tracker used for the experiment is the UEYE headset, which has a system accuracy of 0.527° to 0.875° [44]. The eye cameras have a resolution of 320 × 240, the frame rate is 480 FPS, and the vertical and horizontal viewing angles are ±55° and ±70°, respectively. The vertical and horizontal viewing angles of the scene camera are ±52° and ±68°, respectively.

TWAN: TWAN in our setup comprises a local area network (LAN) and the UAV's dedicated communication network. The LAN adheres to the 802.11ax protocol, boasting a maximum transmission speed of 3000 Mb/s. It also supports dual-band concurrent operation at both 2.4 and 5 GHz frequencies. For transport layer communication, the TCP protocol is

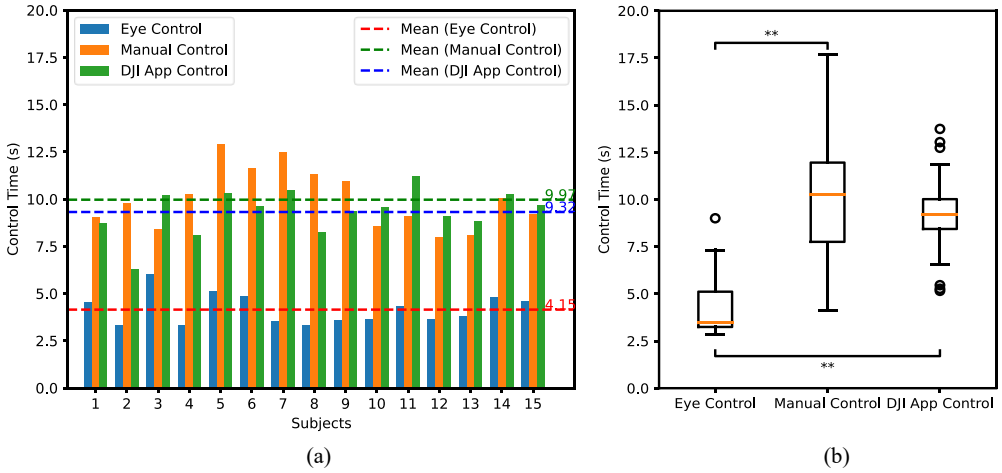

Fig. 9. Time-consuming bar chart and box plot for the ETUAV framework. (a) Bar chart and means of time-consuming in the three control modes for 15 subjects. (b) Box plot of time-consuming for the three control models and comparative results of significance tests.

employed. In this experiment, the UAV's communication is facilitated by OcuSync, a proprietary data transmission technology developed by DJI, ensuring efficient and reliable data exchange.

B. Experimental Steps

In our experiment, 15 subjects were asked to utilize three distinct control methods to operate an UAV in a real-world setting, conducting target aiming tests. The time consumption and accuracy are recorded for further analysis. We selected four objects in the outdoor environment for identification purposes: two people, a bicycle, and a car. These objects were positioned at four different locations within the scene. Benchmarks are conducted in the following three ways.

1) Eye Movement Control: Subjects performed aiming tests on four targets in the environment using the ETUAV framework. The locations of the four targets were selected to ensure that the next target appeared in the UAV's field of view at the end of this target aiming to facilitate successive experiments. Once subjects had aimed at one target, aiming experiments were conducted on subsequent targets in sequence until the entire experiment was completed.

2) Manual Control: Subjects operated the UAV using a remote controller, with an auxiliary center designed at the display center of the UFly application to serve as the UAV imaging's central point. The control procedure mirrored that of the first experiment. A single target sequence concluded when the operator perceived the camera center had aligned with the target center, after which the next target was selected for continuation.

3) DJI Intelligent Follow Control: DJI's intelligent follow mode can recognize objects, such as people, animals, bikes, and cars. We use the DJI GO 4 app on our mobile terminal to select a target, and the UAV will adjust its attitude and track the target.

C. Experimental Results

We compared the three aforementioned control methods across 15 subjects, yielding a total of 60 samples. Our analysis included a time bar chart for targeting four distinct objects and a box plot that demonstrated the deviation from the center point after each control session. This comparative evaluation offers insights into the variations in time efficiency and distance accuracy among the three control methods.

The bar chart in Fig. 9(a), presents the average time consumption using various control methods, with values calculated from aiming at four different targets. The dashed lines indicate the mean time for each method. The ETUAV control framework shows the lowest average time, markedly outperforming the other methods. Following this, the DJI app control method ranks second in efficiency, while manual control is the most time intensive. The bar chart reveals that the ETUAV method has the least variability, indicating remarkable stability. In comparison, the DJI app's time variability is better than manual control but does not match ETUAV's consistency. Conversely, manual control exhibits significant time fluctuations. Fig. 9(b) presents the box plot of time for different control methods. Mann–Whitney U tests were conducted to compare eye-tracking control data with manual control data and DJI app control data. The $*$ denotes $p < 0.05$, while $**$ indicate $p < 0.001$. The time distribution of the ETUAV framework is centered around 4 s, with only one outlier observed. This contrasts with the more varied time distributions of the other two methods. The results from the significance tests clearly demonstrate that the ETUAV framework significantly surpasses the other methods in terms of time efficiency.

Fig. 10 shows the results from the subjects testing four different targets using three control methods. The orange dot indicates the ideal alignment of the camera's imaging center with the target's center. The other dots represent the actual coordinate values measured using different control methods during the experiment. Fig. 11 displays a box plot of all test data, recording the distance discrepancies between the experiment's end coordinates and the target centers on both the x and y axes. Mann–Whitney U tests reveal that the ETUAV framework significantly surpasses the other methods in accuracy. The scatter plot clearly shows that adjustments via the ETUAV framework are more closely clustered around

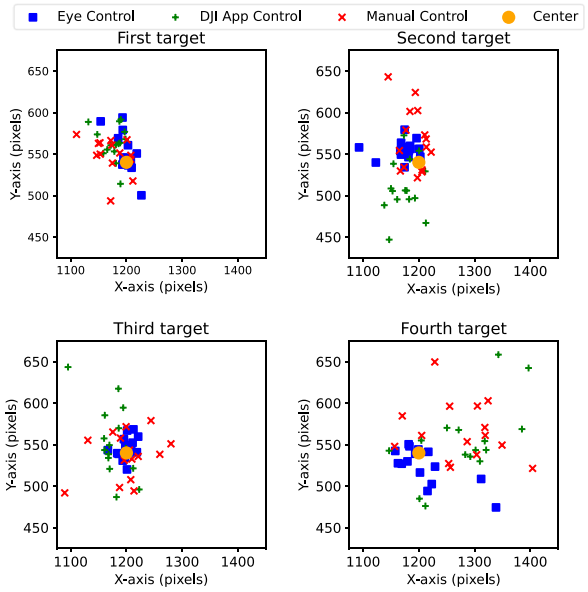

Fig. 10. Scatter plot of experimental results for the four targets.

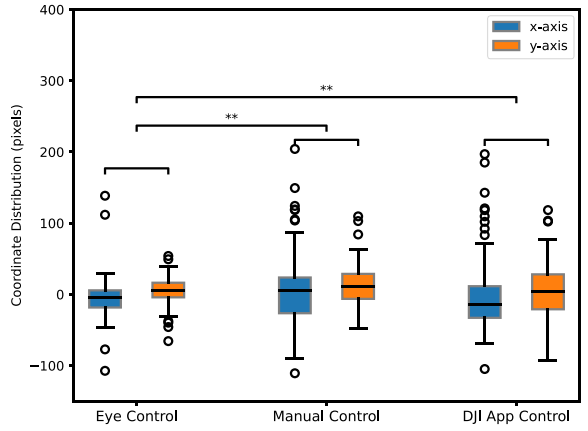

Fig. 11. Box plot of distance difference for all test data results and the results of significance tests, where $*$ denotes $p < 0.05$ and $**$ denotes $p < 0.001$.

the central point compared to other methods. Additionally, the box plot indicates that ETUAV framework data points are more tightly clustered near zero in both dimensions, with fewer outliers, underscoring its precision. The scatter plot's increased scatter for the fourth target, a car, is due to its larger size compared to other targets, leading to greater distance variations.

Table I presents the time allocation for each phase of the eye movement control experiment. Complementing this, Fig. 12 provides a stacked area plot that is consistent with the data in Table I, delivering a more detailed visual representation of time distribution. The total control time is categorized into UAV rotation time, network communication time, and execution algorithm time. An examination of both Table I and Fig. 12 reveals that within the ETUAV framework, the bulk of the time is consumed by UAV rotation and algorithm execution, while network communication occupies a smaller fraction. In terms of variability, rotation time shows the greatest fluctuation, indicating its sensitivity to environmental factors. On the other hand, network communication and algorithm execution times are more consistent, with minor

TABLE I
TIME CONSUMPTION IN EACH STAGE OF ETUAV CONTROL FRAMEWORK

Subjects	Rotation (s)	Network (s)	Algorithm (s)	Total (s)
1	1.508	1.155	1.854	4.517
2	1.005	0.702	1.582	3.288
3	2.512	1.146	2.327	5.984
4	1.005	0.760	1.562	3.327
5	1.758	1.145	2.217	5.121
6	1.760	1.118	1.965	4.843
7	1.006	0.953	1.558	3.516
8	1.006	0.675	1.630	3.310
9	1.008	0.989	1.595	3.592
10	1.255	0.669	1.714	3.638
11	1.507	0.966	1.822	4.295
12	1.256	0.704	1.693	3.653
13	1.256	0.899	1.627	3.781
14	1.762	1.009	2.007	4.777
15	1.762	0.903	1.938	4.603
Mean	1.424	0.920	1.806	4.150
Variance	0.186	0.033	0.059	0.645

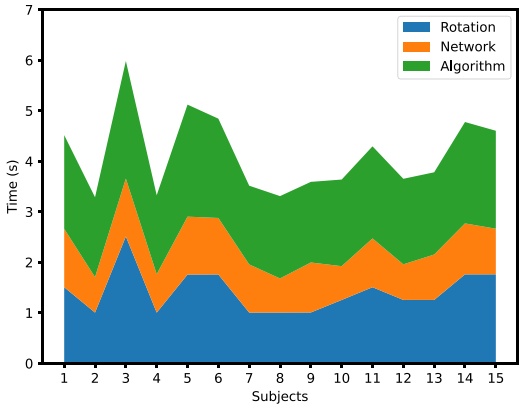

Fig. 12. Stacked area plot of average time spent in each stage of eye movement control for 15 subjects.

variance. The deviations in algorithm execution time are primarily due to varying qualities of input images, leading to differences in processing time.

Relative to the other two control methods, the ETUAV framework demonstrates outstanding efficiency and accuracy in UAV control, clearly outperforming in terms of overall effectiveness and distinct advantages. Nevertheless, the ETUAV framework is not without its limitations. Its reliance on object detection technology means it becomes ineffective in the absence of targets within the field of view. Moreover, the framework's stability is heavily contingent on the performance of the object detection model. Additionally, the framework's reliance on multiple physical devices renders the setup phase somewhat cumbersome. This complexity presents a significant opportunity for further research, particularly in the integration of mobile and computer interfaces, to streamline the overall process.

V. CONCLUSION

In this article, we propose a novel UAV control framework ETUAV based on eye-tracking technology. This framework combines lightweight object detection with the gaze information of the UAV operator to assist in adjusting the UAV's attitude. We also propose an incremental PID algorithm for UAV attitude adjustment and an FVT method for solving image reflection and reducing error. We develop a system prototype using DJI UAV and compare it with traditional

manual control and DJI's official intelligent follow mode through comparative experiments. The experimental results demonstrate that the ETUAV framework outperforms the manual way and DJI intelligent follow in terms of operational efficiency and accuracy. It is anticipated that this framework can achieve the Aim Where You Look capability in the military, rescue, and other time-critical scenarios, yielding the expected outcomes.

During the development of our framework, we encountered several challenges. One key issue was achieving a balance between precision and speed in object detection. We tested various models to find an optimal solution that provides both accurate target detection and rapid inference. Designing the PID-based control algorithm also presented difficulties, particularly in determining the right parameters. To overcome this, we engaged in code simulations and conducted extensive real-world tests. This iterative process allowed us to fine tune the parameters, ultimately leading to the ideal configuration for the control algorithm.

In the future, we will focus on multimodal control modes for the UAV. In addition to eye movement control, we will add more control modes, such as gesture, EEG, etc., and continue to improve the existing control framework to intelligently control the UAV. Second, we will investigate deploying the model on mobile terminals, thus reducing the control time and making the framework more efficient. Finally, we will investigate and apply other target detection models that can further reduce the algorithm time as well as further improve the accuracy.

REFERENCES

[1] A.-F. Perrin, L. Zhang, and O. Le Meur, "Inferring visual biases in UAV videos from eye movements," *Drones*, vol. 4, no. 3, p. 31, 2020.

[2] J. Tan, X. Dai, F. Tang, M. Zhao, and N. Kato, "Intelligent configuration method based on UAV-driven frequency selective surface for communication band shielding," *IEEE Internet Things J.*, vol. 10, no. 13, pp. 11507–11517, Jul. 2023, doi: 10.1109/JIOT.2023.3243387.

[3] Y. Tan, J. Wang, J. Liu, and N. Kato, "Blockchain-assisted distributed and lightweight authentication service for industrial unmanned aerial vehicles," *IEEE Internet Things J.*, vol. 9, no. 18, pp. 16928–16940, Sep. 2022.

[4] A. J. Chadwick, N. C. Coops, C. W. Bater, L. A. Martens, and B. White, "Species classification of automatically delineated regenerating conifer crowns using RGB and near-infrared UAV imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[5] Y. Zhong et al., "Mini-UAV-borne hyperspectral remote sensing: From observation and processing to applications," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 4, pp. 46–62, Dec. 2018.

[6] I. Mademlis, N. Nikolaidis, A. Tefas, I. Pitas, T. Wagner, and A. Messina, "Autonomous unmanned aerial vehicles filming in dynamic unstructured outdoor environments [applications corner]," *IEEE Signal Process. Mag.*, vol. 36, no. 1, pp. 147–153, Jan. 2019.

[7] X. Wang, J. Li, Z. Ning, Q. Song, L. Guo, and A. Jamalipour, "Wireless powered metaverse: Joint task scheduling and trajectory design for multi-devices and multi-UAVs," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 3, pp. 552–569, Mar. 2024, doi: 10.1109/JSAC.2023.3345433.

[8] Z. Ning et al., "Dynamic computation offloading and server deployment for UAV-enabled multi-access edge computing," *IEEE Trans. Mobile Comput.*, vol. 22, no. 5, pp. 2628–2644, May 2023, doi: 10.1109/TMC.2021.3129785.

[9] J. Yi, H. Zhang, S. Li, O. Feng, G. Zhong, and H. Liu, "Logistics UAV air route network capacity evaluation method based on traffic flow allocation," *IEEE Access*, vol. 11, pp. 63701–63713, 2023.

[10] M. Zhang, W. Li, M. Wang, S. Li, and B. Li, "Helicopter–UAVs search and rescue task allocation considering UAVs operating environment and performance," *Comput. Ind. Eng.*, vol. 167, May 2022, Art. no. 107994.

[11] R. Akter, M. Golam, V.-S. Doan, J.-M. Lee, and D.-S. Kim, "IoMT-Net: Blockchain-integrated Unauthorized UAV Localization using lightweight convolution neural network for Internet of Military Things," *IEEE Internet Things J.*, vol. 10, no. 8, pp. 6634–6651, Apr. 2022.

[12] Y. Zhang, H. Zhang, X. Gao, S. Zhang, and C. Yang, "UAV target detection for IoT via enhancing ERP component by brain computer interface system," *IEEE Internet Things J.*, vol. 10, no. 19, pp. 17243–17253, Oct. 2023, doi: 10.1109/JIOT.2023.3273163.

[13] K. Baizid et al., "Behavioral control of unmanned aerial vehicle manipulator systems," *Auton. Robots*, vol. 41, pp. 1203–1220, Jun. 2017.

[14] J. Li, Z. Li, Y. Feng, Y. Liu, and G. Shi, "Development of a human–robot hybrid intelligent system based on brain teleoperation and deep learning SLAM," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 4, pp. 1664–1674, Oct. 2019.

[15] S.-L. Wu, L.-D. Liao, S.-W. Lu, W.-L. Jiang, S.-A. Chen, and C.-T. Lin, "Controlling a human–computer interface system with a novel classification method that uses electrooculography signals," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 8, pp. 2133–2141, Aug. 2013.

[16] Y.-J. Zheng, Y.-C. Du, H.-F. Ling, W.-G. Sheng, and S.-Y. Chen, "Evolutionary collaborative human-UAV search for escaped criminals," *IEEE Trans. Evol. Comput.*, vol. 24, no. 2, pp. 217–231, Apr. 2020.

[17] T. Zhou and Y. Liu, "Long-term person tracking for unmanned aerial vehicle based on human-machine collaboration," *IEEE Access*, vol. 9, pp. 161181–161193, 2021.

[18] X. Liu and N. Ansari, "Resource allocation in UAV-assisted M2M communications for disaster rescue," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 580–583, Apr. 2019.

[19] F. Ho et al., "Decentralized multi-agent path finding for UAV traffic management," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 997–1008, Feb. 2022.

[20] T. Niedzielski et al., "A real-time field experiment on search and rescue operations assisted by unmanned aerial vehicles," *J. Field Robot.*, vol. 35, no. 6, pp. 906–920, 2018.

[21] N. Dilshad, A. Ullah, J. Kim, and J. Seo, "LocateUAV: Unmanned aerial vehicle location estimation via contextual analysis in an IoT environment," *IEEE Internet Things J.*, vol. 10, no. 5, pp. 4021–4033, Mar. 2023.

[22] Z. Ning et al., "5G-enabled UAV-to-community offloading: Joint trajectory design and task scheduling," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 11, pp. 3306–3320, Nov. 2021.

[23] M. Yang, Y. Ma, Z. Liu, H. Cai, X. Hu, and B. Hu, "Undisturbed mental state assessment in the 5G era: A case study of depression detection based on facial expressions," *IEEE Wireless Commun.*, vol. 28, no. 3, pp. 46–53, Jun. 2021.

[24] Z. Ning et al., "Mobile edge computing enabled 5G health monitoring for Internet of Medical Things: A decentralized game theoretic approach," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 2, pp. 463–478, Feb. 2021.

[25] A. Nourmohammadi, M. Jafari, and T. O. Zander, "A survey on unmanned aerial vehicle remote control using brain–computer interface," *IEEE Trans. Human-Mach. Syst.*, vol. 48, no. 4, pp. 337–348, Aug. 2018.

[26] C. G. Coogan and B. He, "Brain-computer interface control in a virtual reality environment and applications for the Internet of Things," *IEEE Access*, vol. 6, pp. 10840–10849, 2018.

[27] M.-A. Chung, C.-W. Lin, and C.-T. Chang, "The human—Unmanned aerial vehicle system based on SSVEP—Brain computer interface," *Electronics*, vol. 10, no. 23, p. 3025, 2021.

[28] T. Deng et al., "A VR-based BCI interactive system for UAV swarm control," *Biomed. Signal Process. Control*, vol. 85, Aug. 2023, Art. no. 104944.

[29] T. Shi, H. Wang, W. Cui, and L. Ren, "Indoor space target searching based on EEG and EOG for UAV," *Soft Comput.*, vol. 23, Nov. 2019, pp. 11199–11215.

[30] J. W. Choi, B. H. Kim, and S. Jo, "Asynchronous motor imagery brain-computer interface for simulated drone control," in *Proc. 9th Int. Winter Conf. Brain-Comput. Interface (BCI)*, 2021, pp. 1–5, doi: 10.1109/BCI51272.2021.9385309.

[31] W. Meng, X. Chen, W. Cui, and J. Guo, "WiHGR: A robust WiFi-based human gesture recognition system via sparse recovery and modified attention-based BGRU," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 10272–10282, Jun. 2022.

[32] C. Yu, S. Fan, Y. Liu, and Y. Shu, "End-side gesture recognition method for UAV control," *IEEE Sensors J.*, vol. 22, no. 24, pp. 24526–24540, Dec. 2022.

[33] Y.-P. Yeh, S.-J. Cheng, and C.-H. Shen, "Research on intuitive gesture recognition control and navigation system of UAV," in *Proc. IEEE 5th Int. Conf. Knowl. Innov. Invent. (ICKII)*, 2022, pp. 5–8, doi: 10.1109/ICKII55100.2022.9983607.
[34] M. Yang, Y. Wu, Y. Tao, X. Hu, and B. Hu, "Trial selection tensor canonical correlation analysis (TSTCCA) for depression recognition with facial expression and pupil diameter," *IEEE J. Biomed. Health Inform.*, early access, Oct. 5, 2023, doi: 10.1109/JBHI.2023.3322271.
[35] M. Yang, Z. Weng, Y. Zhang, Y. Tao, and B. Hu, "Three-stream convolutional neural network for depression detection with ocular imaging," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 4921–4930, 2023.
[36] A. Murata and W. Karwowski, "Automatic lock of cursor movement: Implications for an efficient eye-gaze input method for drag and menu selection," *IEEE Trans. Human-Mach. Syst.*, vol. 49, no. 3, pp. 259–267, Jun. 2019.
[37] X. Zhang, X. Liu, S.-M. Yuan, and S.-F. Lin, "Eye tracking based control system for natural human-computer interaction," *Comput. Intell. Neurosci.*, vol. 2017, Dec. 2017, Art. no. 5739301.
[38] J. Jun, B. Zhao, P. Zhang, Z. Chen, and F. Peng, "Research on UAV control method based on eye tracking," in *Proc. 33rd Chin. Control Decis. Conf. (CCDC)*, 2021, pp. 3281–3286, doi: 10.1109/CCDC52312.2021.9602071.
[39] K. Karas, L. Pozzi, A. Pedrocchi, F. Braghin, and L. Roveda, "Brain-computer interface for robot control with eye artifacts for assistive applications," *Sci. Rep.*, vol. 13, no. 1, 2023, Art. no. 17512.
[40] X. Wu, H. Li, and J. Chen, "Intelligent command and control of UAV based on brain computer and eye tracking combined technology," in *Proc. IEEE 4th Adv. Inf. Manage., Commun., Electron. Autom. Control Conf. (IMCEC)*, vol. 4, 2021, pp. 212–221.
[41] Q. Wu, S. Xie, Z. Zeng, Q. Huang, and J. Pan, "A multiple command UAV control system based on a hybrid brain-computer interface," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2023, pp. 1–8, doi: 10.1109/IJCNN54540.2023.10191621.
[42] K. Jagatheesan, B. Anand, S. Samanta, N. Dey, A. S. Ashour, and V. E. Balas, "Design of a proportional-integral-derivative controller for an automatic generation control of multi-area power thermal systems using firefly algorithm," *IEEE/CAA J. Automatica Sinica*, vol. 6, no. 2, pp. 503–515, Mar. 2019.
[43] Z. Wang and J. Zhang, "Incremental PID controller-based learning rate scheduler for stochastic gradient descent," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Oct. 26, 2022, doi: 10.1109/TNNLS.2022.3213677.
[44] M. Yang, Y. Gao, L. Tang, J. Hou, and B. Hu, "Wearable eye-tracking system for Synchronized multimodal data acquisition," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Nov. 14, 2023, doi: 10.1109/TCSVT.2023.3332814.
[45] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[46] Y.-H. Yiu et al., "DeepVOG: Open-source pupil segmentation and gaze estimation in neuroscience using deep learning," *J. Neurosci. Methods*, vol. 324, Aug. 2019, Art. no. 108307.
[47] F. Reghenzani, G. Massari, and W. Fornaciari, "The real-time Linux kernel: A survey on preempt_RT," *ACM Comput. Surveys*, vol. 52, no. 1, pp. 1–36, 2019.
[48] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.

Minqiang Yang (Member, IEEE) received the Ph.D. degree in computer science from Lanzhou University, Lanzhou, China, in 2022.

He is an Associate Professor with Gansu Provincial Key Laboratory of Wearable Computing, School of Information Science and Engineering, Lanzhou University. He has published more than 20 papers on IEEE magazines, IEEE journals, and leading conferences. His current research interests include affective computing, image processing, machine learning, and automatic depression detection.

Jintao Wang received the B.S. degree from the School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China, in 2021. He is currently pursuing the M.S. degree with Gansu Provincial Key Laboratory of Wearable Computing, School of Information Science and Engineering, Lanzhou University, Lanzhou, China.

His main research interests are intelligent control of UAVs and machine learning.

Yujie Gao received the B.S. degree from Beijing Institute of Technology, Beijing, China, in 2021. He is currently pursuing the M.S. degree with Gansu Provincial Key Laboratory of Wearable Computing, School of Information Science and Engineering, Lanzhou University, Lanzhou, China.

His main research interests are affective computing, image processing, and machine learning.

Bin Hu (Fellow, IEEE) received the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 1998.

He is currently a Full Professor and the Dean of the School of Medical Technology, Beijing Institute of Technology, Beijing. He is also an Adjunct Professor and the former Dean of the School of Information Science and Engineering, Lanzhou University, Lanzhou, China. He (co)authored more than 400 publications (10 000+ citations and H-index 51). He is a Chinese National Distinguished Expert, the Chief Scientist of 973 projects, and the National Advanced Worker in 2020.

Prof. Hu awards include the 2014 China Overseas Innovation Talent Award, the 2016 Chinese Ministry of Education Technology Invention Award, the 2018 Chinese National Technology Invention Award, and the 2019 WIPOCNIPA Award for Chinese Outstanding Patented Invention. He serves as the Editor-in-Chief for the IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS. He is also the TC Co-Chair of Computational Psychophysiology in the IEEE SYSTEMS, MAN, AND CYBERNETICS SOCIETY (SMC) and the TC Co-Chair of Cognitive Computing in IEEE SMC. He is a Principal Investigator for large grants, such as the National Transformative Technology "Early Recognition and Intervention Technology of Mental Disorders Based on Psychophysiological Multimodal Information," which have extensively promoted the development of objective, quantitative diagnosis, and nondrug interventions for mental disorders. He is a Member of the Steering Council of the ACM China Council and the Vice-Chair of the China Committee of the International Society for Social Neuroscience. He is a Member of the Computer Science Teaching and Steering Committee as well as the Science and Technology Committee. He was elected as a Fellow of Institution of Engineering and Technology.