Hindawi
International Journal of Aerospace Engineering
Volume 2022, Article ID 5818231, 11 pages
https://doi.org/10.1155/2022/5818231

Research Article

Eye Gaze Estimation Based on Stacked Hourglass Neural Network for Aircraft Helmet Aiming

Quanlong Qiu, Jie Zhu, Chao Gou, and Mengtang Li

School of Intelligent Systems Engineering, Sun Yat-Sen University, Guangzhou, Guangdong 510000, China

Correspondence should be addressed to Mengtang Li; limt29@mail.sysu.edu.cn

Received 26 May 2022; Revised 21 August 2022; Accepted 24 September 2022; Published 6 October 2022

Academic Editor: Xingling Shao

Helmet-mounted display (HMD) systems allow aircraft pilots to aim at targets by using head postures. However, the direct use of helmet orientation to indicate the aiming direction ignores human eye movements, which are more flexible and efficient for interaction. Since the opaqueness of the helmet goggle blocks the sight of external cameras to capture facial or eye images of the pilots, and traditional eye feature extraction methods may fail when encountering conditions such as poor lighting, occlusion, and shaking, which are common on fighter aircrafts. In this work, an eye gaze based aiming solution that adapts to pilots wearing HMDs is proposed, and a deep learning-based method is proposed to extract eye features robustly. The prototype experiments demonstrate the ability to pick and aim at targets in real-time (60FPS) and are capable to accurately locate the target markers on a screen with an average error of fewer than 2 degrees. Conclusively, the proposed method performs the tasks of eye feature extraction on real-person imagery and the estimation of the 3D aiming direction for users with helmets, displaying competitive results with similar research.

1. Introduction

Helmet mounted display (HMD) systems provide versatile functions and information for pilots of advanced fighter aircrafts by assisting flight control and improving control efficiency. A critical function of the HMD systems is to allow pilots to use head posture as the control and guidance direction of weapon systems, thereby simplifying and accelerating the progress of aiming [1–4]. Optical, electromagnetic, ultrasonic, and hybrid sensors are usually installed within the helmet to achieve accurate head aiming function by sensing the orientation of the helmet [3, 5, 6]. The Joint Helmet-Mounted Cueing System (JHMCS) refers to an integrated product of Display and Sight Helmet (DASH) III and Kaiser Agile Eye helmet displays, which uses helmet electromagnetic position sensor to measure the posture of helmet [7–9]. Scorpion, developed by the French company Thales, has been available in the military aviation market since 2008. Its posture is initially measured by using alternating current (AC) and electromagnetic sensors and later replaced with a Hybrid Optical based Inertial Tracker (HObIT) [10–12]. The Eurofighter Typhoon uses the Helmet-Mounted Symbology System (HMSS) developed by the British company BAE and the Japanese company Pilkington Optronics. Similar to the DASH system, the HMSS system uses an integrated helmet position sensor to measure and indicate the aiming direction of the pilot under head movements [13, 14].

Current existing HMDs directly use the posture of helmets to indicate the direction of targeting. However, these methods ignore the eye movements of pilots, which are more flexible and offer a faster reaction rate. Using eye tracking based direction for aiming may reduce pilot workload on head rotations with a heavy helmet, and thus reduces physical fatigue and improves the combat fitness. As the deep learning advances by leaps and bounds, it is promisingly possible to apply the eye-tracking algorithm to the helmet aiming.

Conventional eye-tracking devices are usually placed within a certain distance before the users for gaze estimation [15, 16]. The opaque goggle of a flight helmet inevitably blocks the view from such devices and therefore prevents the direct use of external cameras in a cockpit that senses the eye gaze of a pilot. Therefore, the camera that captures the image of the human eyes must be mounted inside the helmet goggles, which leads to a very close distance between the camera and eyes and the poor lighting condition. Besides, the jitter of a fighter aircraft in a flight process cannot be ignored. All these factors require an accurate and robust eye feature extraction.

Generally, eye-tracking is divided into two steps: eye feature extraction and eye mapping model establishment [16–19]. Usually, eye features with both unchanged (the corner of an eye, Purchin's spot) and changed (the center of the pupil and the edge of the iris) relative positions are extracted [20, 21]. Then feature vectors are constructed based on the relative position of these features and mapped to a 3D gaze direction or a 2D fixation point. Robust eye tracking usually depends on the accurate detection of the eye features such as the iris center or eye corners. In previous eye-tracking works, features are handcrafted by adopting image processing techniques and model fitting [22, 23]. Since such approaches assume the geometry and shape of the eyes, they are sensitive to changes in appearance such as poor lighting conditions, blurriness, and vibrations of the pilot's head. By leveraging advanced neural network architecture, we propose a deep learning-based method for accurate eye features extraction. A preliminary study was done by the authors in a previous preprint [24].

Driven by the aforementioned needs and difficulties, this work attempts to extract eye features of a pilot equipped with an HMD system, and combines eye movements with the head posture of a pilot for aiming. The main contributions and novelties of this paper are as follows:

(1) An eye tracking solution that adapts to pilots wearing HMDs is proposed. The method of space vector transformation is used to combine the head pose and the 3D gaze direction for aiming

(2) To improve the accuracy of eye feature extraction in the eye-tracking based HMD aiming scenario, a deep learning-based eye feature detector is trained, supporting the 3D gaze estimation

(3) HMD prototype is developed and manufactured, and the accuracy of the proposed HMD aiming system is experimentally evaluated. Results show that the system achieves real-time accurate HMD aiming in the laboratory environment

The remainder of this work is arranged as follows: Section 2 presents the eye feature extraction algorithm, and Section 3 presents the design of the eye-tracking based helmet-mounted display system. These are followed by lab environment tests in Section 4, with detailed evaluation and comparison of results with other counterpart researches. Conclusive remarks are drawn in the last Section 5.

2. Eye Feature Extraction

2.1. Neural Network Architecture. Gaze estimation for pilots requires real-time extraction of high-quality eye features. The stacked hourglass network design [25, 26] is used in this work to meet with the above requirement. The hourglass network is initially proposed for human pose estimation, where a key problem is the occlusion. This design of stacked hourglass network attempts to capture long-range context and to ensure a large receptive field. The stacked hourglass network meets the needs of extracting information at different sizes, so it is suitable for extracting eye features. Although eye images contain fewer global structural elements than pose estimation, there are still important spatial contexts that could be developed by large receptive field models. We take advantage of this property to train an eye feature extractor.

Figure 1 shows a single four level hourglass module. This module does not change the input size, but the output fuses features at different sizes. The output is the feature maps extracted from a pilot's eyes images. Feature maps from human eye images are downscaled via pooling operations, and then upscaled using bilinear interpolation. Before downsampling operation, it separates a single route to retain the information in the current size. At every scale level, the task-relevant features are preserved by the residual module during deep network training. The hourglass module only changes the depth of the data without changing the size of data. The hourglass module performs repeated bottom-up, top-down inference. Hence it is able to capture and consolidate information from different scales and resolutions and allows the encoding of spatial relations between landmarks. Leveraging with the hourglass module, high-quality latent features contribute to accurate gaze estimation.

The heatmaps acquired from the hourglass module represent the probabilities of the eye landmarks at every pixel. As for the last hourglass module, its output of heatmaps is further processed via a soft-argmax layer to find the subpixel coordinates of these landmarks. The eye landmarks are utilized for gaze estimation. Finally, the three linear fully-connected layers as well as one final regression layer are used for predicting the eyeball radius.

In this work, three hourglass modules are stacked. Single eye images (150×90) are used as input, and 18 heatmaps are generated (75×45): 8 in the limbus region, 8 on the iris edge, and 1 at the iris center as well as 1 at the eyeball center. The proposed framework predicts 18 landmarks and an eyeball radius. Figure 2 shows the whole architecture of the entire network. Accurate eye localization facilitates real-time gaze targeting.

2.2. Training Data. UnityEyes, a high-quality synthetic eye image dataset with rich annotation, is utilized for training [27]. These synthesized data provide accurate eye feature coordinates, which include the eyelid-sclera border, limbus regions (iris-sclera border), and eye corners, as shown in Figure 3. Besides, it additionally includes other information such as pupil size, head posture, and gaze direction. UnityEyes is efficient and infinite in size to present excellent variations in iris color, eye region shape, and head pose as well as illumination conditions. Yet, the appearance variations do not contain visual artifacts which are common in webcam images or common eye decorations like eyeglasses or make-up. To address this issue, this paper performs the

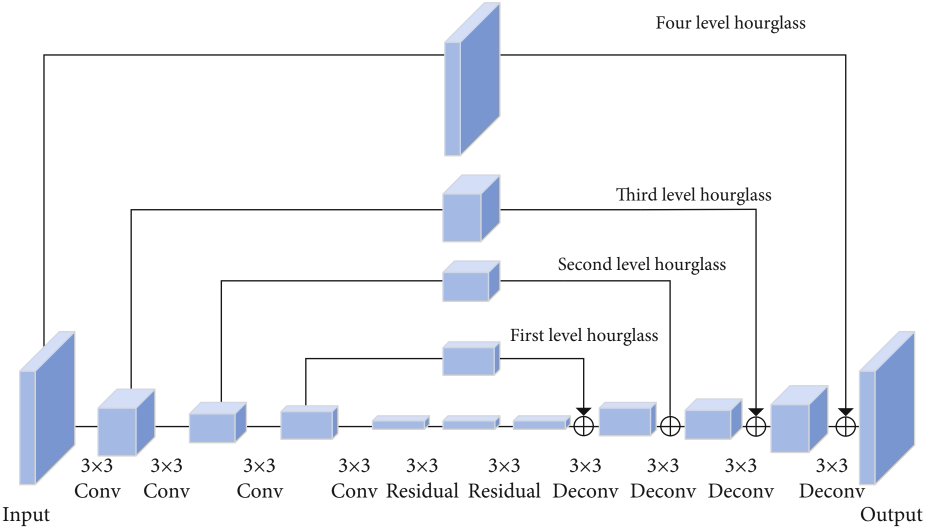

Figure 1: The four-level hourglass module.

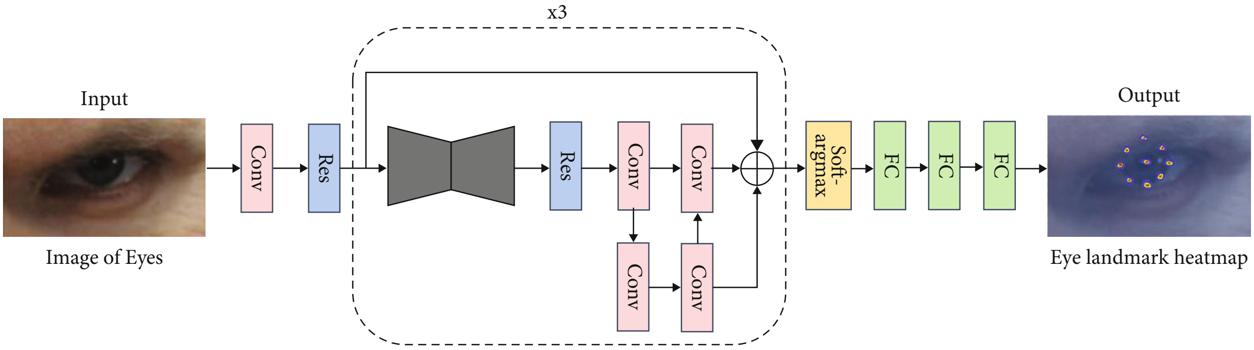

Figure 2: The whole network architecture.

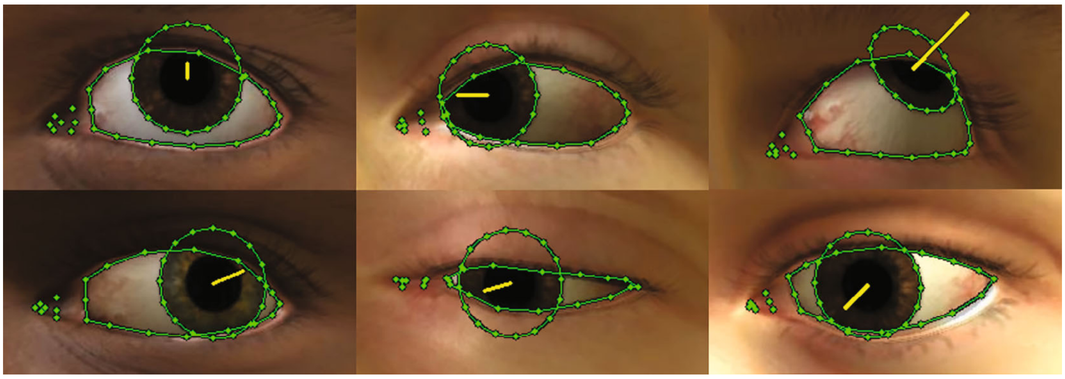

Figure 3: Synthetic eye images from UnityEyes with rich annotations.

training data augmentation so that a robust model can be trained solely on synthetic eye images. The subsequent augmentations are applied (range in brackets represents scaling coefficients of the value sampled from N (0, 1)): translation (2–10 pixels), rotation (0.1– 2.0 rad), intensity (0.5–20.0), blur (0.1–1.0 std. dev. on 7 × 7 Gaussian filter), scale (1.01–1.1), downscale-then-upscale (1–5 times), and the addition of artifact lines (0–2) for artificial occlusions. Image flipping is not performed during training but easily guarantees that the inner eye corner locates on the left side of the input image.

2.3. Intermediate Supervision. Intermediate supervision is performed to calculate the loss of output of the heatmaps in each hourglass module. Three hourglass modules are stacked so that the network can continue to repeat the bottom-up and top-down processes through the intermediate supervision. The loss of each hourglass module is computed independently, such that the subsequent hourglass module can be reevaluated, and the higher-level spatial relationships can be reassessed. The network predicts heatmaps, one per eye feature point, a total of eighteen. The heatmaps encode the per-pixel confidence on a specific eye feature. Two-dimensional Gaussian distributions are centered at the subpixel positions of eye features with the peak value of unity. The neural network minimizes the l_2 distance between the predicted and ground-truth heatmaps per eye feature through the loss term as shown in

$$L_h = \alpha \sum_{i=1}^{18} \sum_{p} \left\| \tilde{h}_i(p) - h_i(p) \right\|_2^2, \qquad (1)$$

where $h(p)$ refers to the confidence at pixel p and $\tilde{h}(p)$ represents the heatmap predicted by the network. Besides, the weight coefficient α is empirically set to 1. The loss term for predicting the radius of the eyeball is

$$L_r = \beta \| \tilde{r}_{uv} - r_{uv} \|_2^2, \qquad (2)$$

where $\tilde{r}_{uv}$ is the predicted eyeball radius and r_{uv} is the ground truth and β is set to 10^{-7}.

2.4. Training Process. The training scheme implements curriculum learning, which is imitates the learning process of human beings, advocates that models start from easy samples, and gradually progress to difficult samples. To make it easier to control, a difficulty measure with the range from 0 to 1 is implemented. In the training process, the difficulty of the samples is related to eye rotation angles (pitch and roll), head posture, data enhancement, and so on. For instance, the greater the eye movement, the larger difficulty of the sample. The process starts training with difficulty 0 and linearly enhances difficulty until 10^6 training steps have passed. Afterward, the difficulty is maintained at 1. Training according to the difficulty of samples from simple to difficult can achieve better performance with fewer number of iteration steps.

During the training process, the ADAM optimizer [28] is used, with a learning rate of 5×10^{-4}, batch size of 16, l_2 regularization coefficient of 10^{-4}, and ReLU activation. The model is trained for 10^6 steps on an Nvidia GTX 1660 super GPU, which consists of less than 1 million model parameters and allows for a real-time implementation (60FPS). Figure 4 shows eye feature extraction results under very challenging conditions.

3. HMD Aiming System

3.1. Gaze Direction Estimation. A simple model of the human eyeball can be considered as a large sphere with a small sphere intersecting each other [29], as shown in Figure 5. Suppose the predicted pixel coordinates of the 8 iris landmarks in a given eye image are $(u_{i1}, v_{i1}), \cdots, (u_{i8}, v_{i8})$. In addition, the eyeball center (u_c, v_c) and the iris center (u_{i0}, v_{i0}) are also detected. Furthermore, the network predicts the eyeball radius in pixels, r_{uv}. Having the eyeball and iris center coordinates and eyeball radius in terms of pixels enables it to fit a 3D model without acquiring any camera intrinsic parameters.

In the case that the camera intrinsic parameters are unknown, the coordinates can only be projected into 3D space in pixel units. As a result, the radius remains $r_{xy} = r_{uv}$ in 3D model space and $(x_c, y_c) = (u_c, v_c)$. Assuming the gaze direction is expressed by pitch and yaw angles $g_c = (\theta, \phi)$, the iris center coordinates can be represented as

$$\begin{aligned} u_{i0} &= x_{i0} = x_c - r_{xy} \cos\theta \boldsymbol{sin}\phi, \\ v_{i0} &= y_{i0} = y_c + r_{xy} \boldsymbol{sin}\theta. \end{aligned} \qquad (3)$$

To write similar expressions for the 8 iris edge feature points, the angular iris radius δ and an angular offset γ which equals to eye roll are jointly estimated. For the j-th iris edge landmarks (with $j = 1 \cdots 8$) is as follows:

$$\begin{aligned} u_{ij} &= x_{ij} = x_c - r_{xy} \boldsymbol{cos}\theta'_j \boldsymbol{sin}\phi'_j, \\ v_{ij} &= y_{ij} = y_c + r_{xy} \boldsymbol{sin}\theta'_j, \end{aligned} \qquad (4)$$

where

$$\begin{aligned} \theta'_j &= \theta + \delta \boldsymbol{sin}\left(\frac{\pi}{4} j + \gamma\right), \\ \phi'_j &= \phi + \delta \boldsymbol{cos}\left(\frac{\pi}{4} j + \gamma\right). \end{aligned} \qquad (5)$$

For this model-based gaze estimation, θ, ϕ, γ, and δ are unknown, while other variables are provided by the eye region feature points localization step of the network. The conjugate gradient method, serves as an iterative optimization approach, is used to solve this problem, and the minimized loss function can be written as follows:

$$L_{opt} = \sum_{0 \le j \le 8} \left(u_{ij} - u'_{ij}\right)^2 + \left(v_{ij} - v'_{ij}\right)^2, \qquad (6)$$

where (u'_{ij}, v'_{ij}) refers to the estimated pixel coordinates of the j-th iris landmarks at each iteration. The calculation of person-specific parameters based on calibration samples adopts the proposed model to a certain person. Gaze correction can be implemented with $(\widetilde{\theta}, \widetilde{\phi}) = (\theta + \triangle\widetilde{\theta}, \phi + \triangle\widetilde{\phi})$, in which $(\triangle\widetilde{\theta}, \triangle\widetilde{\phi})$ refers to the person-specific angular offset between optical and visual axes.

3.2. Aiming Direction Estimation. The gaze direction relative to the camera that captures the eye image is so far acquired, which is represented as pitches and yaw angles. The roll angle (φ) can be regarded as 0 since the eyeball does not roll. To prevent the interference of the helmet goggles, a custom-made wide-angle camera that captures the image of the eyes is installed inside the helmet goggles and is fixed with the

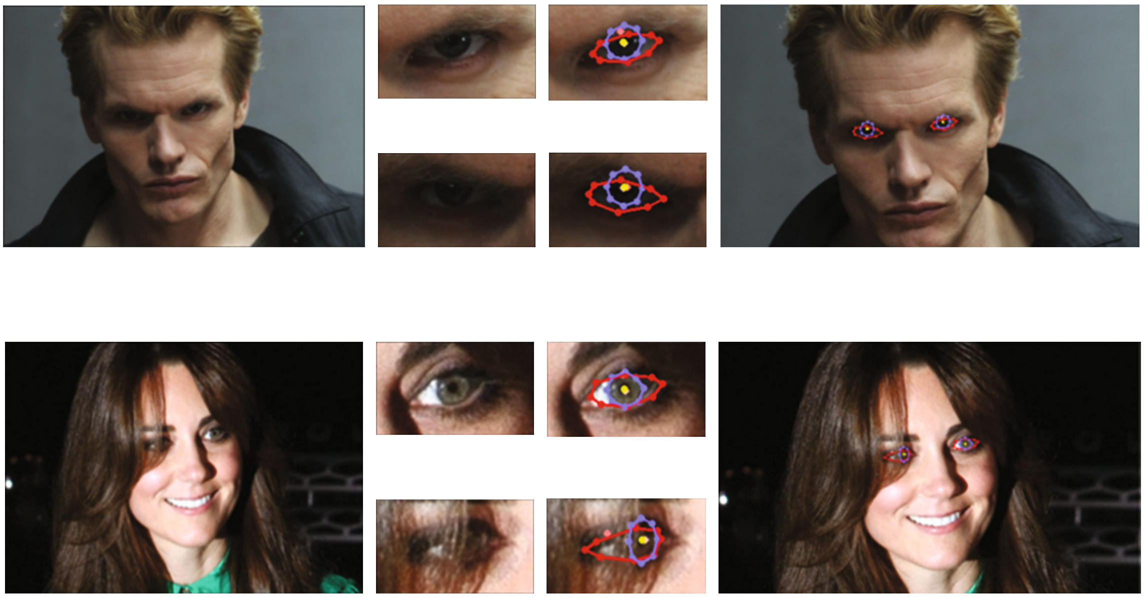

Figure 4: Eye features are extracted under dark condition and occlusion condition.

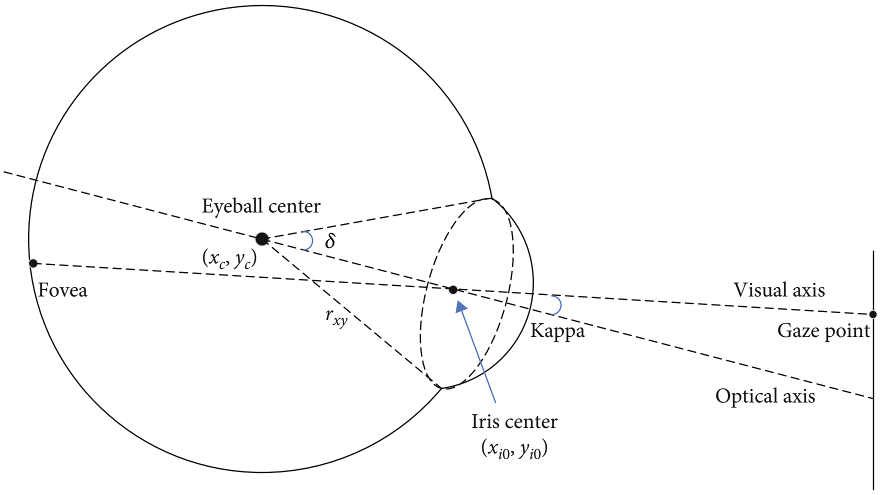

Figure 5: 3D eyeball model.

helmet. Thus, the gaze direction is relative to the helmet coordinate frame.

Assume the helmet attitude (denoted as h in subscript) relative to the aircraft cockpit (denoted as c in superscript) is $p_h^c = (\alpha, \beta, \gamma)$. The helmet attitude uniquely determines a unit vector $e_h^c = (\boldsymbol{cos}\alpha, \boldsymbol{cos}\beta, \boldsymbol{cos}\gamma)$, and both p_h^c and e_h^c indicate the direction of the helmet. Since the gaze direction is relative to the camera mounted on the helmet, it can be considered an additional rotation of the helmet attitude. The gaze direction can be converted into a rotation matrix from eyes (denoted as e in subscript) to helmet$R_e^h \in SO(3)$:

$$R_e^h = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \boldsymbol{cos}\theta & -\boldsymbol{sin}\theta \\ 0 & \boldsymbol{sin}\theta & \boldsymbol{cos}\theta \end{bmatrix} \begin{bmatrix} \boldsymbol{cos}\phi & 0 & \boldsymbol{sin}\phi \\ 0 & 1 & 0 \\ -\boldsymbol{sin}\phi & 0 & \boldsymbol{cos}\phi \end{bmatrix} \begin{bmatrix} \boldsymbol{cos}\varphi & -\boldsymbol{sin}\varphi & 0 \\ \boldsymbol{sin}\varphi & \boldsymbol{cos}\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{7}$$

The unit vector for the final aiming direction can be further expressed as $v_e^c = e_h^c \cdot R_e^h = [x_e \, y_e \, z_e]$ in the world coordinate frame, and the aiming angle can be calculated as

$$\begin{aligned} g_e^c &= \boldsymbol{arccos}\left[\frac{x_e}{\sqrt{x_e^3 + y_e^2 + z_e^2}} \; \frac{y_e}{\sqrt{x_e^3 + y_e^2 + z_e^2}} \; \frac{z_e}{\sqrt{x_e^3 + y_e^2 + z_e^2}}\right] \\ &= [\theta_e^c \, \phi_e^c \, \varphi_e^c], \end{aligned} \tag{8}$$

where g_e^c indicates the final aiming direction of the pilot's eyes relative to the cockpit considering eye movements and helmet posture.

3.3. System Implementation. The schematic diagram of the system is shown in Figure 6. Unlike dataset-based algorithm test, no ground truth of 3D aiming direction is available in

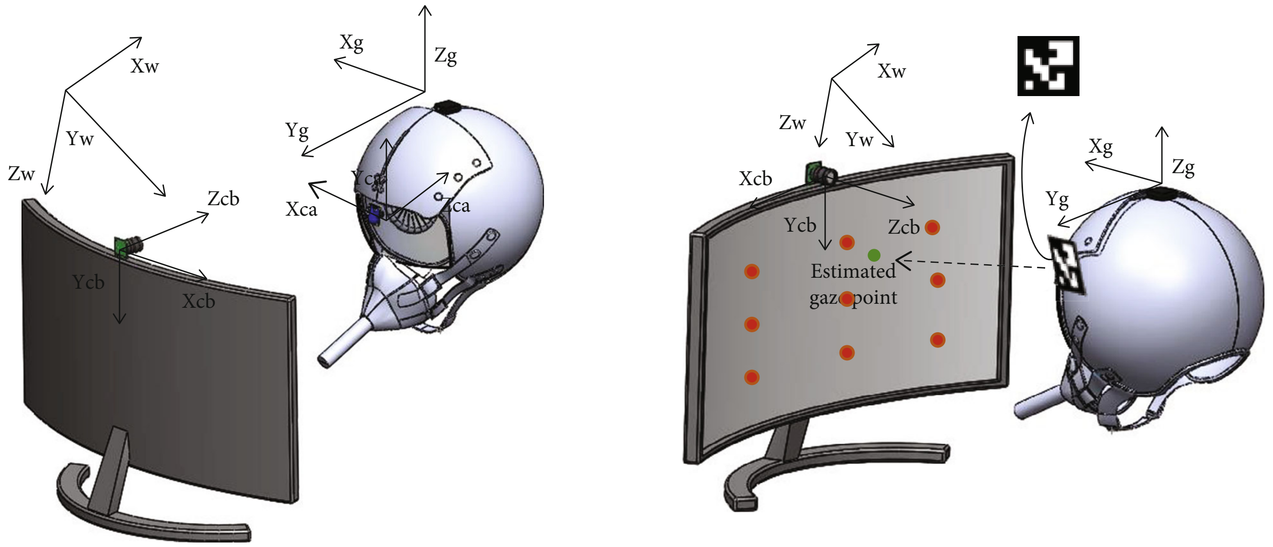

Figure 6: System diagram.

real applications. In practical applications, it is required to know the intersection of the aiming direction and a certain plane. To do this, we installed an additional camera above the screen and put an ArUco marker [30] on the helmet. The camera at the top of the screen detects the ArUco marker to determine the position of the helmet in relation to the screen. Other available methods such as depth cameras can also be adopted to accomplish this job. Knowing the relative position between the helmet and the screen and the aiming direction, aiming points on the screen can be calculated. Figure 7 shows how ArUco marker is detected.

Although optical or mixed sensor methods are used in real aircrafts during heavy maneuvers, a 6-axis gyroscope is used and installed inside the helmet to detect the helmet posture since there is no acceleration in the lab environment. In addition, to capture the image of the pilot's eyes, a monocular wide-angle camera is mounted inside the helmet. Note that in order to detect ArUco marker and de-distort images captured by the wide-angle camera, both the wide-angle camera and the camera above the screen need to be calibrated.

Nine marker points are stationarily set up on a computer screen, covering most of the screen area. The screen is 54 cm long and 30 cm wide, the left and right spacing of each point is 18 cm, and the upper and lower spacing is 5 cm, as illustrated in Figure 6.

Since the relative position of the helmet to the cockpit (environment) $p_h^c = [x_h \, y_h \, z_h]$ is acquired via detecting the ArUco marker and the aiming direction $[\theta_e^c \, \phi_e^c \, 0]$ are known, the aiming points on the screen can be calculated as

$$\begin{aligned} X_c &= -z_h \times \sin\left(\phi_e^c\right) + x_h, \\ Y_c &= z_h \times \sin\left(\theta_e^c\right) + y_h, \end{aligned} \tag{9}$$

where z_h, $x_{h,}$ and y_h represent the coordinates of the helmet in the screen coordinate above the screen; θ_e^c represents yaw and ϕ_e^c represents pitch of the eyeball. The accuracy of the eye gaze tracking system is quantitatively evaluated by calculating the angular value E_{dg} as

$$E_{dg} = \boldsymbol{arctan}\left(\frac{E_d}{E_g}\right), \tag{10}$$

where E_d is the distance between the estimated gaze position and the real aimed position, and E_g stands for the distance between the subjects and the screen plane.

4. Experimental Evaluation

4.1. Gaze Estimation. The accuracy of the proposed gaze estimation is experimentally evaluated via assessing the precision of gaze points landing on a screen as mentioned in the previous section. To better analyze the performance of the proposed method, experiments without and with helmets are conducted successively. Resulting differences presented below give the community a straightforward understanding of the sources (eye gaze estimation, helmet pose estimation, etc.) and magnitudes of estimation errors. The diagram of the experiment setup is shown in Figure 8.

Note that an ArUco marker is attached to a subject's forehead or the front of the helmet's goggle to determine the relative position between the screen and the subject for the sake of simplicity. Other methods such as depth camera or optical sensors can be used in real applications. The webcam above the screen captures the ArUco images and the orientation of the subject's head is simultaneously calculated. The estimation of gaze points landing on the screen is then calculated via equation (9), replacing the aim direction with gaze direction for no-helmet scenarios. Also, subjects are allowed to move their heads while aiming at the points on the screen since head posture is taken into account when calculating aiming direction.

Twenty subjects, eight wearing glasses, voluntarily and anonymously participate in the experimental evaluation

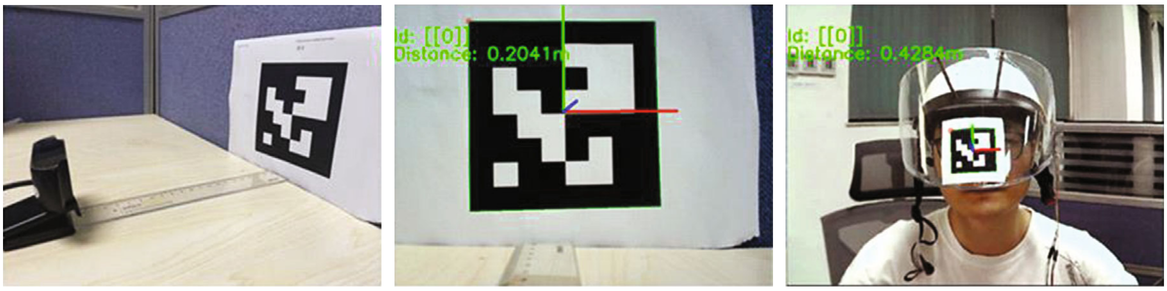

Figure 7: ArUco detection in the lab environment. Left: the camera is about 20 cm away from the ArUco marker. Middle: after calibration, the camera can accurately detect the posture of the ArUco marker. Right: an ArUco marker is detected to determine the head posture relative to the screen.

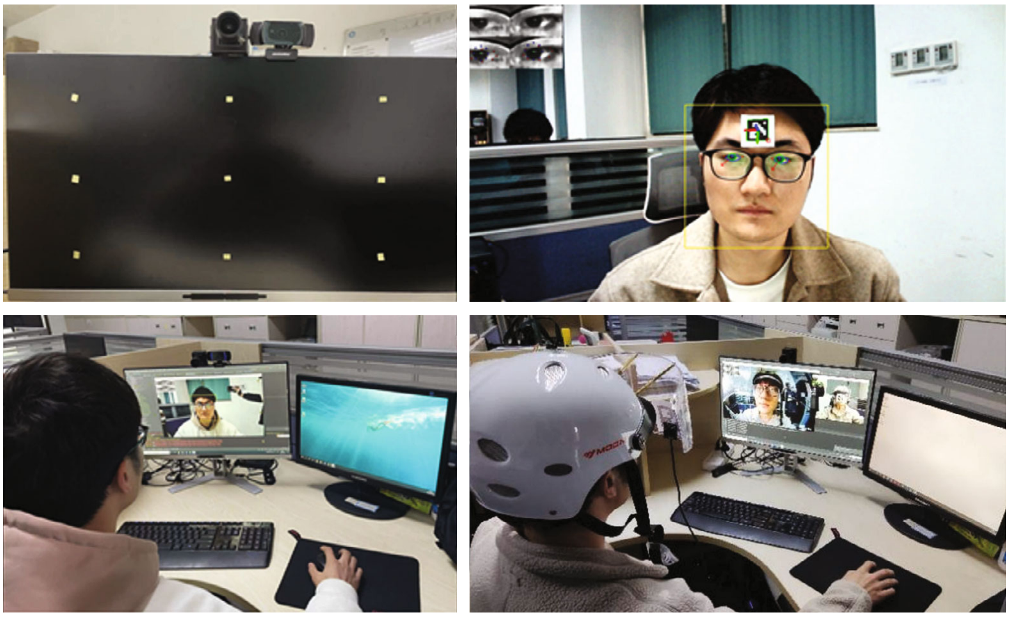

Figure 8: Gaze estimation under lab environment. Upper left: nine markers on the screen. Upper right: a subject sat in front of a screen and looked at a marker on the screen. Lower left: the eye feature points are extracted from the webcam and 3D gaze direction is estimated. Lower right: the eye feature points are extracted from the proposed HMD helmet and 3D gaze direction combined with helmet orientation is estimated.

under diverse illumination situations. The subjects are asked to focus on nine markers on the screen about 60 cm away from them in turn. Same tests are requested for another round except that the subjects wear the develop HMD prototype helmet. The results from twenty subjects, without and with helmets, are plotted in Figure 9.

The angular error of gaze landing estimation without helmet from twenty subjects ranges from 0.94 to 1.94 degrees, with an average error of 1.36 degrees. The variation originates from different subjects' physical characteristics, such as gender, glasses, eyelashes, and sitting position. With helmet, the angular error of gaze landing estimation is between 1.65 to 2.23 degrees, with an average error of 1.99 degrees. The slightly increased error is largely caused by the helmet pose estimation and difficulty in detecting the eye features at a very close distance. Since the HMD prototype is head-mounted and has determined its relative position to the world coordinate frame, subject's head is theoretically free to move spatially. In fact, the head movement is limited in a range simply because the ArUco marker is used in this work to calculate the position of the helmet and to show the proof-of-concept. Once the detection of ArUco marker fails, it leads to wrong aiming point estimation. Optical sensors can be used in real applications to overcome this limitation. From this point of view, the HMD prototype achieves free movement of the head.

Next, the same group of volunteers, without and with helmets, is requested to sit at 40 cm, 60 cm, and 80 cm away from the screen to study the impact of distance onto the gaze estimation accuracy. Corresponding results are plotted in Figures 10–12, respectively. The red crosses represent the ground-truth positions of the nine markers on the screen. The blue and red dots are estimated gaze marker landing positions without and with helmets obtained within one second. It is assumed that the coordinates of these estimated gaze landing positions obey a 2D Gaussian distribution as

$$X \sim \left(\mu_x, \sigma_x^2\right), Y \sim \left(\mu_y, \sigma_y^2\right), \quad (11)$$

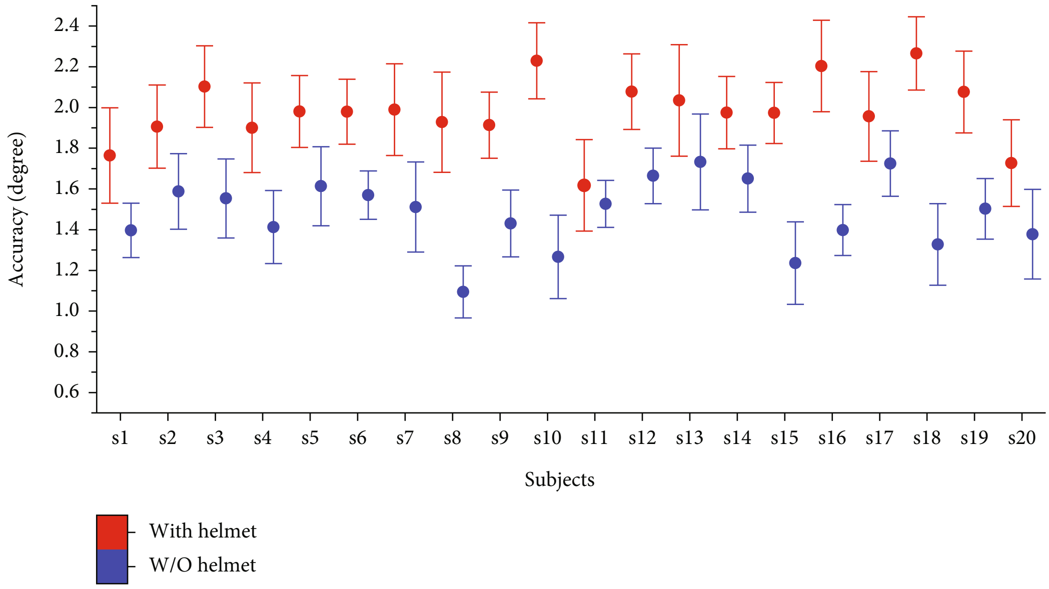

Figure 9: The average error from 20 subjects, without (blue) and with (red) helmets, looking at 9 markers at 60 cm away from the screen.

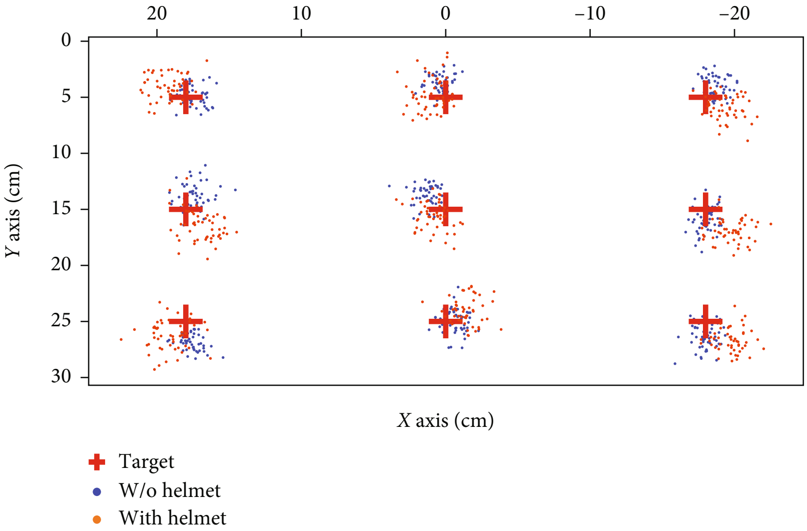

Figure 10: Visualized results of gaze marker landing estimation with subjects at 40 cm away from the screen.

where μ_x, σ_x^2, μ_y, and σ_y^2 are sampling means and sampling standard deviations for horizontal and vertical directions.

Straightforward results can be seen from the above figures. The Euclidean distance error increases as the distance between head and screen increases, but the angular error does not change significantly. Although the results of each experiment fluctuate, the angle error is below 2 degrees. Note that the error in the y-direction is slightly greater than in the x-direction, and the accuracy toward the screen edge is decreased slightly. With the eyeball moving to the edge of the eye socket when subjects focus on the screen edge markers, the iris could be overlapped by the eyelids. Thereby, the detection accuracy of the gaze vector is slightly weakened.

Table 1 summarizes the mean error for the gaze marker landing estimation for subjects without and with helmets. It can be seen that the algorithm itself (without helmets) is insensitive to the distance between the camera and head, and reaches an average accuracy of 1.35 degrees in the person-independent evaluation, which demonstrates the robustness of the eye-tracking algorithm. Integrating the eye-tracking algorithm with the helmet, the accuracy is slightly decreased due to the error of measuring the helmet pose estimation and detecting the eye features at a very close distance, leading to a slightly inferior accuracy of 2.00 degrees.

4.2. Comparison. To verify the performance of the eye tracking based helmet, the system is compared with other eye tracking algorithms/devices. Table 2 presents the performance of the proposed method with the head movement in comparison with Skodras et al. [31], Cheung et al. [32],

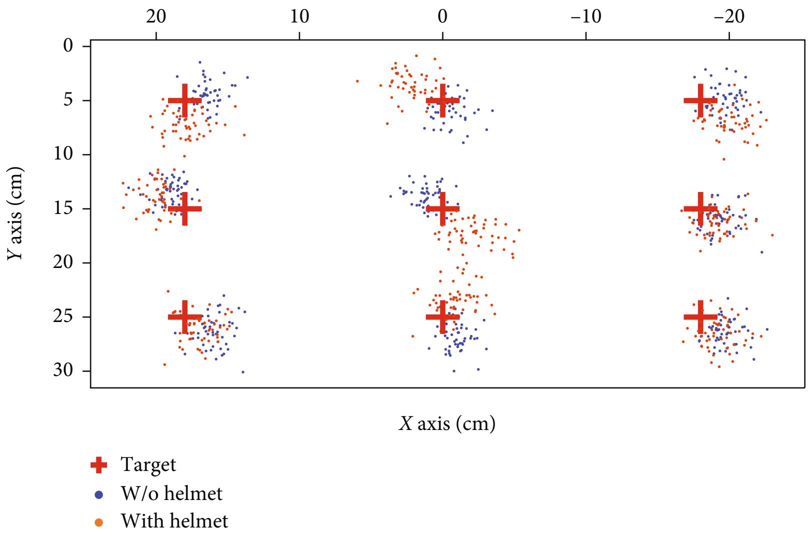

Figure 11: Visualized results of gaze marker landing estimation with subjects at 60 cm away from the screen.

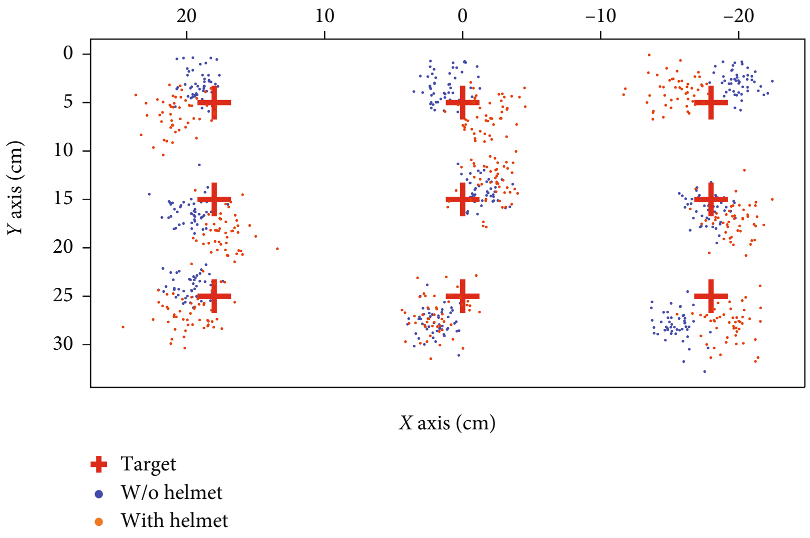

Figure 12: Visualized results of gaze marker landing estimation with subjects at 80 cm away from the screen.

Table 1: Mean error for the gaze marker landing estimation at different distances.

Distance/cm	Without helmet		With helmet	
	Error/cm	Error/degree	Error/cm	Error/degree
40	1.14	1.34	1.72	1.91
60	1.59	1.36	2.28	1.99
80	2.00	1.36	3.00	2.00

Arar et al. [33], Li et al. [34], and Wang et al. [29]. The ranges of head motion in each method have been explicitly shown and are compared under similar experimental conditions for a fair comparison.

Advantageously speaking, the proposed work achieves an average error of less than 2.0 degrees under free head movement without any additional light source and measures the 3D aiming direction. Skodras et al. and Cheung et al. directly maps eye features to screen landing markers, rather than estimating the 3D gaze direction. Therefore, the head's motion range of movement is limited, and the error will become larger when the head deviates from the calibrated position. Arar et al. also directly maps eye feature to gaze point on the screen and achieve an accuracy of 1.15 degrees. But 5 light sources are required which increase the complexity of the system. Li et al. and Wang et al. estimate the gaze

Table 2: Performance comparison with relevant research.

Methods	Error/degree	Number of light sources	Head movement range
Skodras et al. [31]	1.42 ± 1.35	0	Fixed
Cheung et al. [32]	2.27	0	15.5 × 15.5 × 15.5
Arar et al. [33]	1.15	5	Free
Li et al. [34]	3.0 ~ 4.5	0	Free
Wang et al. [29]	1.3	4	Free
Proposed	2.00	0	Free

direction in 3D and both achieve free head movement, where Li et al. uses a Kinect depth camera instead of conventional cheap cameras. While Wang et al. reaches 1.3 degrees of error and free from calibration, 4 additional light sources are needed. In contrast, the proposed gaze system utilizes a single monocular camera capturing the face video without using additional light sources, which reduces the system complexity. By determining the position relationship between the head and the screen, the system estimates the aiming point on the screen with an accuracy of less than 2 degrees.

5. Conclusion

In this paper, an eye gaze estimation-based aircraft helmet aiming methodology is proposed to allow pilots use the more flexible and efficient eye movement for human-machine interaction, other than the current method of rotating the heavy head-mounted helmet, which increases pilot's physical fatigue and decreases combat fitness. The proposed eye gaze estimation method is achieved via a monocular wide-angle camera installed inside a flight helmet to capture real time eye images from a pilot. A stacked hourglass deep learning network is designed to extract real-time high-quality eye features under versatile illumination conditions and to estimate the gaze direction. A proof-of-concept prototype is developed and tested along with the eye gaze estimation algorithm itself under lab environment, providing the community with straightforward understanding of the 3D aiming error's sources and magnitudes. The experiment results have demonstrated the ability to pick and aim at targets in real-time (60FPS) and are capable to accurately locate the target markers on a screen with an average error of fewer than 2 degrees under versatile operating conditions. Compared with relevant eye gaze estimation studies, the proposed method stands out with advantages such as free head movement and no additional light source requirement.

Conclusively, the proposed HMD aiming algorithm and developed system achieves the eye-tracking based HMD aiming function. System design and algorithm architecture offer promising novelties and contribute to the current HMD technology.

Data Availability

The data used to support this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Authors' Contributions

Quanlong Qiu and Jie Zhu contributed equally to this work.

References

[1] W. S. Kim, A. Liu, K. Matsunaga, and L. Stark, “A helmet mounted display for telerobotics,” *IEEE Computer Society*, vol. 88, pp. 543–547, 1988.

[2] M. M. Bayer, C. E. Rash, and J. H. Brindle, “Introduction to helmet-mounted displays,” *Helmet-mounted displays: sensation, perception and cognition Issues*, pp. 47–108, 2009.

[3] M. Adamski, M. Adamski, and A. Szelmanowski, “The magnetic field curvature correction algorithm dedicated for helmet mounted cueing systems,” *Journal of KONES*, vol. 25, no. 2, pp. 15–30, 2018.

[4] V. Kumar and S. K. Raghuwanshi, “Implementation of helmet mounted display system to control missile 3D movement and object detection,” in *2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC)*, pp. 175–179, Mathura, India, 2020.

[5] R. Atac and E. Foxlin, “Scorpion hybrid optical-based inertial tracker (HObIT),” *Head-and Helmet-Mounted Displays XVIII: Design and Applications*, vol. 8735, no. 873502, 2013.

[6] G. W. Orf, “Joint helmet-mounted cueing system (JHMCS) helmet qualification testing requirements,” *Helmet-and Head-Mounted Displays III*, vol. 3362, pp. 118–123, 1998.

[7] S. A. Mekonnen, M. B. Asrat, and S. Ramasamy, “A helmet cueing system based firing control for anti-aircraft gun prototype,” *Advances in Military Technology*, vol. 16, no. 1, pp. 19–33, 2021.

[8] S. Michalak, J. Borowski, A. Szelmanowski, and A. Pazur, “The polish helmet mounted display systems for military helicopters,” in *2016 IEEE Metrology for Aerospace (MetroAeroSpace)*, pp. 353–358, Florence, Italy, 2016.

[9] J. M. Barnaba and H. Anthony Orr, “A summary of efforts toward the definition of potential upgrades to the joint helmet mounted cueing system,” in *14th Annual AESS/IEEE Dayton Section Symposium. Synthetic Visualization: Systems and Applications*, Fairborn, OH, USA, 1997.

[10] R. Atac and T. Bugno, “Qualification of the scorpion helmet cueing system,” *Head-and Helmet-Mounted Displays XVI: Design and Applications*, vol. 8041, pp. 182–188, 2011.

[11] R. Atac, "Applications of the Scorpion color helmet-mounted cueing system," *Head-and Helmet-Mounted Displays XV: Design and Applications*, vol. 7688, no. 768803, 2010.

[12] R. Atac, S. Spink, T. Calloway, and E. Foxlin, *Display Technologies and Applications for Defense, Security, and Avionics VIII; and Head-and Helmet-Mounted Displays XIX*, vol. 9086, no. 90860U, 2014.

[13] C. J. Smith, "Design of the Eurofighter human machine interface," *Air & Space Europe*, vol. 1, no. 3, pp. 54–59, 1999.

[14] S. J. Carter and A. A. Cameron, "Eurofighter helmet-mounted display: status update," *Helmet-and Head-Mounted Displays V*, vol. 4021, pp. 234–244, 2000.

[15] N. M. Arar and J.-P. Thiran, "Robust real-time multi-view eye tracking," 2017, https://arxiv.org/abs/1711.05444.

[16] R. Banaeeyan, A. A. Halin, and M. Bahari, "Nonintrusive eye gaze tracking using a single eye image," in *2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pp. 139–144, Kuala Lumpur, Malaysia, 2015.

[17] H. Yamazoe, A. Utsumi, T. Yonezawa, and S. Abe, "Remote gaze estimation with a single camera based on facial-feature tracking without special calibration actions," in *Proceedings of the 2008 symposium on Eye tracking research & applications*, pp. 245–250, New York, NY, USA, 2008.

[18] X. Xiong, Z. Liu, Q. Cai, and Z. Zhang, "Eye Gaze Tracking Using an RGBD Camera: A Comparison with a RGB Solution," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pp. 1113–1121, New York, NY, USA, 2014.

[19] J. Zhu and J. Yang, "Subpixel eye gaze tracking," in *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, pp. 131–136, Washington, DC, USA, 2002.

[20] D. Li, D. Winfield, and D. J. Parkhurst, "Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops*, p. 79, San Diego, CA, USA, 2005.

[21] W. Fuhl, T. Kübler, K. Sippel, W. Rosenstiel, and E. Kasneci, "Excuse: robust pupil detection in real-world scenarios," in *International conference on computer analysis of images and patterns*, pp. 39–51, Cham, 2015.

[22] R. S. Kothari, A. K. Chaudhary, R. J. Bailey, J. B. Pelz, and G. J. Diaz, "Ellseg: an ellipse segmentation framework for robust gaze tracking," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 5, pp. 2757–2767, 2021.

[23] Z. He, T. Tan, Z. Sun, and X. Qiu, "Toward accurate and fast iris segmentation for iris biometrics," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 9, pp. 1670–1684, 2009.

[24] M. Li, Q. Qiu, J. Zhu, and C. Gou, "An Eye Tracking based aircraft helmet mounted display aiming system," 2022.

[25] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *European conference on computer vision*, pp. 483–499, Amsterdam, The Netherlands, 2016.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, Nevada, USA, 2016.

[27] E. Wood, T. Baltrušaitis, L. P. Morency, P. Robinson, and A. Bulling, "Learning an appearance-based gaze estimator from one million synthesised images," in *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, pp. 131–138, New York, NY, USA, 2016.

[28] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," pp. 273–297, 2014, https://arxiv.org/abs/1412.6980.

[29] K. Wang and Q. Ji, "3D gaze estimation without explicit personal calibration," *Pattern Recognition*, vol. 79, pp. 216–227, 2018.

[30] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.

[31] E. Skodras, V. G. Kanas, and N. Fakotakis, "On visual gaze tracking based on a single low cost camera," *Signal Processing: Image Communication*, vol. 36, pp. 29–42, 2015.

[32] Y.-m. Cheung and Q. Peng, "Eye gaze tracking with a web camera in a desktop environment," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 4, pp. 419–430, 2015.

[33] N. M. Arar, H. Gao, and J.-P. Thiran, "Towards convenient calibration for cross-ratio based gaze estimation," in *2015 IEEE Winter Conference on Applications of Computer Vision*, pp. 642–648, Waikoloa, HI, USA, 2015.

[34] J. Li and S. Li, "Gaze estimation from color image based on the eye model with known head pose," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 3, pp. 414–423, 2016.