An Eye Tracking based Aircraft Helmet Mounted Display Aiming System

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

09-01-2022 / 12-01-2022

CITATION

Li, Mengtang; Qiu, Quanlong; Zhu, Jie; Gou, Chao (2022): An Eye Tracking based Aircraft Helmet Mounted Display Aiming System. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.18093233.v1>

DOI

[10.36227/techrxiv.18093233.v1](https://doi.org/10.36227/techrxiv.18093233.v1)

An Eye Tracking based Aircraft Helmet Mounted Display Aiming System

Quanlong Qiu, Jie Zhu, Chao Gou, Mengtang Li

Abstract—Helmet display system is a device used in fighter aircrafts to provide situational awareness, scene enhancement images and other information to a pilot. The system is not limited by the field of view and integrates radar and fire control systems into the helmet, allowing the pilot to choose and aim targets by pointing his/her head. However, the direct use of helmet orientation to indicate the direction of interest ignores the more flexible and efficient way of interacting via the pilot's eye movements. The opaque goggle of the helmet prevents the direct usage of a camera in a cockpit to sense a pilot's eye gaze. Therefore, eye gaze direction based helmet display system is innovatively introduced to realize the spatial combination of head posture and eye gaze direction. Wide angle monocular camera installed inside a helmet is used to capture human eye images under helmet coordinate frame in real time. Robust eye gaze directions (namely, pitch angle and yaw angle) are obtained through the combination of appearance-based and model-based eye tracking algorithms. The eye gaze direction is combined with the head posture sensor to acquire the pilot's line of sight relative to the cockpit. Prototype experiments demonstrate that the proposed helmet display system allows user to pick and aim targets faster, and can accurately locate the gazing point in a simulated cockpit environment with less than 2.1 degree averaged error.

Index Terms—helmet mounted display system, gaze direction, head direction.

I. INTRODUCTION

Helmet mounted display (HMD) is a device used in fighter aircrafts to provide visual information to pilots, such as situational awareness, enhanced images of the scene, and strategic information. Furthermore, a HMD allows usage of a pilot's head pointing direction as the control and guidance direction of weapon systems, accelerating the speed of information exchange and flight control [1]. In the military weapon indication system of a jet fighter, HMD uses inertial sensors, optical sensors, electromagnetic sensors, ultrasonic sensors and hybrid sensor methods to accurately measure the position and orientation of the helmet. The hybrid sensor approach works the best via using a combination of inertial and optical sensors to improve tracking accuracy, enhancing update rate and reducing delay.

However, direct use of a HMD's orientation instead of a pilot's eye gaze direction to indicate the target direction ignores the pilot's eye movement, which is a more flexible, more efficient and more freedom of interaction. Eye tracking technology is a hot research topic in the field of computer vision in recent years. Through various sensors and algorithms,

the gaze direction of users' eyes can be identified and estimated readily, improving the efficiency of various human-computer interaction applications [2], [3]. Yet, an opaque goggle of the flight helmet prevents the direct usage of a camera in a cockpit to sense a pilot's eye gaze.

Motivated by the aforementioned needs, this work attempts to upgrade the current HMD system with eye gaze direction through combination of eye tracking with head orientation to acquire the final aiming direction. Conventional eye tracking uses a fixed camera mounted on a reference surface to capture and calculate the direction of the person's eyes, while the helmet goggles worn by fighter pilots will block the view of the external camera towards the driver's face. Therefore, the camera that captures the eye image should be installed inside the goggles. For above reasons, the captured images are possibly dark and blurry due to poor lighting conditions (if no additional light source is added) and the vibration of the helmet. Therefore, in order to obtain an accurate gaze direction of the pilot, it is critically important to detect the pilot's eye features robustly. In this paper, a deep neural network was trained to perform the task of detecting interpretable eye features. The core of this detector is a stacked hourglass convolutional neural network (CNN) architecture that initially proposed for human pose estimation task [4]. The task of eye-region landmark detection bears similarities to the problem of joint detection in human hand and full body pose, where a key problem is the occlusion of landmarks due to other body parts or decoration. The hourglass architecture tries to capture long-range context by performing repeated bottom-up, top-down inference which ensures a large effective receptive field and allows for the encoding of spatial relations between landmarks, even under occlusion. This eye-region landmark detector is trained solely on high-quality synthetic eye images [5] which are accurately labeled for the location of important landmarks in the eye region, such as eyelid-sclera border, limbus regions (iris-sclera border), and eye corners. The key advantage of this approach is that model-based and feature based gaze estimation methods can be applied even to eye images for which iris localization and ellipse fitting can be highly challenging with traditional methods. Three hourglass modules were trained on eye images and annotations provided by UnityEyes [5], and the trained model allows for a real-time implementation (60Hz). Then, those detected features were used to fit a 3D eyeball model which directly estimates the gaze direction in 3D [2]. Figure 1 shows the whole pipeline of the proposed system. As the eye tracking system is head-mounted, the measured gaze direction is essentially relative to the helmet. To deduce the line of sight relative to the fixed world frame system,

Qiu, Q., Zhu, J., Gou C., and Li M., are with School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou, China.

Manuscript received February 6, 2018; revised July 31, 2018.

Fig. 1. Stacked hourglass network trained on synthetic data estimates eye region landmarks. The landmark coordinates are used for model-based gaze estimation. Head pose is combined with gaze direction to indicate the final aiming direction.

instantaneous kinematic information of the helmet must be acquired as well to transform the eye-to-head direction to the final gaze direction. Various methods exist as stated previously to measure the instantaneous orientation of the helmet under versatile acceleration situations. Here, a gyroscope, which cannot measure the correct orientation under accelerations of an aircraft, was used to simply illustrate the working principle of the proposed eye tracking based HMD aiming system under lab environment. Then the posture of the helmet relative to the cockpit and the gaze direction relative to the head are combined to represent the final aiming direction. To evaluate the feasibility and precision of this system, experiments were conducted upon an HMD prototype to estimate the final gaze point on the screen of a desktop computer. To obtain the relative pose of the head and the screen, an additional camera is used on top border of the screen and an Aruco marker [6] is added in front of the prototype helmet. The prototype experiment shows that the average angular error is less than 2.1 degree when the distance from the screen is about 60cm, which demonstrates the feasibility that eye tracking based HMD aiming system.

This paper has the following three contributions: First, using a large number of synthesized eye images to train an eye region landmark extractor. Second, the eye movements are added to the helmet mounted display aiming system for the final aiming direction. Third, experimental prototype was made and the robustness of our HMD aiming was evaluated under lab environment.

II. RELATED WORK

A. Helmet Mounted Display Aiming System

Modern advanced fighters mostly use a helmet mounted display (HMD) aiming system to ameliorate human-system interactions. Its main working principle is to accurately measure the orientation (pitch, yaw and roll) of a pilot's helmet to indicate the aiming direction. In some cases, the position (x, y, and z) of the pilot's helmet relative to the cockpit must also be measured. The current HMD aiming system generally uses hybrid sensor methods such as inertial and optical sensors to accurately measure the position and orientation of the helmet, which improves tracking accuracy, update rate and delay.

Joint Helmet-Mounted Cueing System (JHMCS) is a derivative product of DASH III and Kaiser Agile Eye helmet displays, developed by Vision Systems International (VSI). The JHMCS system retains the same helmet position electromagnetic sensor as the DASH (Display and Sight Helmet) system, and uses a newer and faster digital processing software package [7]–[10]. The JHMCS integrated with night vision goggles, which can provide pilots with visual information about the cockpit and environment at night.

French company Thales put its Helmet Mounted Display System, code-named Scorpion, into the military aviation market in 2008. The attitude of the Scorpion was initially measured by alternating current (AC) electromagnetic sensors, and then replaced with a Hybrid Optical based Inertial Tracker (HOBIT) [11]–[13]. The HOBIT was developed by InterSense and tested by Thales [14]. Besides, compared with the JHMCS, the scorpion can display colorful numbers and symbols. In the aircraft mission system, the helmet is used to indicate and guide the aiming sensor and over-the-horizon missiles.

The Eurofighter Typhoon uses the Helmet-Mounted Symbolology System (HMSS) developed by British company BAE and Japanese company Pilkington Optronics, code-named Striker and an upgraded version of Striker II. It can display raster images and symbol information, and can be integrated into a night vision goggles. Similar to the DASH system, the HMSS system uses an integrated helmet position sensor to measure and indicate the direction of the pilot's line of sight, and to ensure that the displayed image information symbols are consistent with the pilot's head movement [15], [16].

Existing helmet aiming systems all directly use the direction of the helmet to indicate the direction of the targeting, ignoring the pilot's eye movement, which is a more flexible, more efficient interaction method. Therefore, studying the eye tracking based helmet mounted display aiming system has important strategic significance for improving the combat effectiveness of fighters.

B. Eye Tracking Technology

Feature-based eye tracking methods generally detect certain visual features of a human's eye from the captured images, such as pupil, iris, eye corners and corneal reflection points (with light source), and then extract the relevant eye gaze parameters to estimate the gaze point by a mapping model. Sesma et al. [17] introduced a simple and common approach that use the pupil-center-eye-corner vector or PC-EC vector to represent the eye movements. Huang et al. [18] formulate a feature vector form estimated head pose and distance between 6 landmarks detected on a single eye. Aiming at the difficulty of accurately detecting the corner of the eye, researches have added a light source to the system, and replaces the eye corner point with the Purkin spot formed by the light source reflected by the cornea [19]–[22]. In addition, methods have been proposed that use a deep neural network to detect the eye features [2], [23]. Park et al. [2] used deep convolutional neural network trained solely on synthetic eye images to extract 18 eye region landmarks. Rakshit et al. [23] proposed training

a convolutional neural network to directly segment entire elliptical structures and demonstrate that such a framework is robust to occlusions

Model-based methods attempt to fit a known 3D model to the eye image by minimizing a suitable energy [24]–[28]. The symmetry axis of the eyeball is the optical axis, and there is a fixed deflection angle between the optical axis of the eyeball and the visual axis that varies from person to person. This angle is called the Kappa and is usually around 5 degrees. Eye tracking refers to estimating the direction of the visual axis and determining the gaze point based on the information of the observed scene. The eyeball can generally be regarded as two intersecting spheres with deformations and the center and radius of the eyeball as well as the angular offset between visual and optical axes are determined during user calibration procedures. Further, approaches also proposed to use a neural network to fit an eyeball to an eye image [2], [24], [29], [30].

Appearance-based method directly takes raw eye images (or other auxiliary information, such as face) as input, and trains a mapping model between the eye appearance and the gaze direction. Appearance-based methods often require the introduction of large, diverse training datasets [31]–[34] due to the changes of head movement, changing illumination, (partial) occlusions, and eye occlusions and typically leverage some form of convolutional neural network (CNN) architecture as well as different input data modalities. Krafka et al. [32] introduced a large scale dataset GazeCapture. By using this dataset, a convolutional neural network that performs the task of 2D gaze estimation is trained and achieves accuracy of less than 1.34cm error. Zhang et al. [33] presented the MPIIGaze which has become a commonly used public dataset. Zhang et al. [35], [36] adapted the LeNet-5 and VGG-16 architectures and introduced spatial weights to encode the face image. Besides, weakly supervised gaze estimation from video [37] and using eye landmark heatmaps to obtain gaze estimation [38] were also proposed.

III. METHOD

A. Neural Network Architecture

Since a custom-made camera that captures the polit's eye image is installed inside the goggle, relatively dark environment and the vibration of the helmet require the robustness of the eye feature detection. Here, a deep convolutional neural network is trained to perform the task of eye region landmarks localization, which has been demonstrated as a robust method in previous works [2], [4], [39]. The primary module used in this network is Residual Module, named after the skip connections, as shown in Fig. 2. The first path is the residual mapping, which is composed of three convolutional layers (white) with different kernel scales in series, with batch normalization (BN, blue) and ReLU (orange) interleaved. The second path is the identity mapping, which only contains a convolutional layer with a core scale of 1. The step size of all convolutional layers is 1, and that of the padding is 1 as well. Hence the data size will not change, only the data depth (channel) will be changed. The residual module is controlled by two parameters: input depth M and output depth N . The

Fig. 2. Residual module, the blue, orange and white boxes represent batch normalization, active layer and convolution layer, respectively.

residual module extracts higher-level features, while retaining the original level of information so as to be able to train deeper networks.

The hourglass module consists of the residual module, which is the core component of the network. Hourglass model has different levels of complexity, depending on the order. Both the upper and lower paths contain several residual modules (gray), which gradually extract deeper features. The upper path is performed at the original scale, and the lower one has undergone a process of down-sampling (divided by 2) and then up-sampling (multiplied by 2). Down-sampling (green) uses max pooling with a kernel size of 2×2 and stride of 2, and up-sampling (light orange) uses bilinear interpolation. The input and output depth of each residual module is 32. The feature maps are downscaled via pooling operations, then upsampled using bilinear interpolation, thereby forming an hourglass-shaped network structure. The second-order hourglass only needs to replace the middle residual module of the first-order hourglass with a first-order hourglass module. Then replace the middle residual module of the second-order hourglass module with a first-order hourglass module, which is a third-order hourglass. Therefore, the increase in the order of the hourglass module can be regarded as a recursive process. This work uses a fourth-order hourglass module, as shown in Fig. 3. Three fourth-order hourglass modules were stacked to perform feature extraction tasks. The architecture of the entire network is shown in Fig. 4.

The stacked hourglass network architecture has previously been applied to human pose estimation and facial landmarks detection where complex spatial relations need to be modeled at various scales [4], [39]. The hourglass architecture has been proven to be effective in tasks of estimating the locations of occluded joints or key points. The architecture performs repeated multi-scale refinement of feature maps, which ensures a large effective receptive field and is capable of encoding of spatial relations between landmarks, even under occlusion. The network calculates the coordinates of each eye feature point through the soft-argmax layer, and then further appending 3 linear fully-connected layers with 100 neurons each (with batch normalization and ReLU activation) and one final regression layer with 1 neuron to predict eyeball radius.

B. Training Data

High-quality synthetic eye images, UnityEyes [5], were utilized for training. These synthesized data provide rich and accurate eye feature point coordinates, including eyelid-sclera

Fig. 3. Fourth-order hourglass module, the green ,gray and light orange boxes represent maximum pooling layer, residual module and bilinear interpolation upsampling.

Fig. 4. The whole network architecture. The white, orange and gray boxes represent convolution layer, BN plus ReLU layer and residual module. The hourglass-shaped module is the fourth-order hourglass module. The white and red framed convolution module performs intermediate supervision, the light blue is the soft-argmax module, and purple is the fully connected layer.

Fig. 5. Synthetic eye images from UnityEyes with rich annotations.

border, limbus regions (iris-sclera border), and eye corners, as shown in Fig. 5. Specifically, eyelid-sclera border has 16 annotation points, iris-sclera border has 32 annotation points, and eye corner has 7 annotation points. It also additionally includes pupil size, head posture, gaze direction, etc. UnityEyes is effectively infinite in size and was designed to exhibit good variations in iris color, eye region shape, head pose and illumination conditions.

C. Intermediate Supervision

Multiple hourglass modules are stacked in the entire network structure, so that the network can continue to repeat the bottom-up and top-down processes. The key to adopte this structure is to use intermediate supervision to calculate the loss of the heatmaps output by each hourglass module. The location for performing intermediate supervision is in the red box convolutional layer in Fig. 4. Hence, the loss of each hourglass module is calculated separately, so that the subsequent hourglass module can be better re-evaluated and

reassess higher order spatial relationships. In this work, we selected a total of 18 eye feature points: 8 on the eyelid-sclera border, 8 on the iris-sclera border, 1 at the iris center, and 1 at the eyeball center. The network performs the task of predicting heatmaps, one per eye feature point. The heatmaps encode the per-pixel confidence on a specific feature point. Two-dimensional Gaussian distributions were centered at the sub-pixel feature point positions and the peak value is 1. The neural network minimizes the l_2 distance between the predicted and ground-truth heatmaps per feature point via the following loss term:

$$L_h = \sum_{i=1}^{18} \sum_p \|\tilde{h}_i(p) - h_i(p)\|_2^2 \quad (1)$$

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

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

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

D. Training Process

To increase the robustness of the model, data augmentation was employed. The following augmentations were applied. Range in brackets are scaling coefficients of value sampled from $N(0, 1)$: translation (2 – 9 px), rotation (0.2 – 2.0 rad), intensity (0.5 – 20.0), blur (0.1 – 1.0 std. dev. on 7×7 Gaussian filter), scale (1.01 – 1.2), downscale-then-upscale (1x – 5x), and addition of lines (0 – 2) for artificial occlusions.

During the training process, ADAM optimizer was used, with a learning rate of 5×10^{-4} , batch size of 16, l_2 regularization coefficient of 10^{-4} and ReLU activation. Model was trained for 6M 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).

IV. HMD AIMING SYSTEM

A. Model-based Gaze Estimation

A simple model of the human eyeball can be generally regarded as a large sphere with two small spheres intersecting with each other to represent the corneal bulge, as shown in Fig. 6. Suppose the predicted coordinates of the 8 iris landmarks in a given eye image are $(u_{i1}, v_{i1}), \dots, (u_{i8}, v_{i8})$. In addition, the eyeball center (u_c, v_c) and the iris center (u_{i0}, v_{i0}) are also detected. Eventually the network predicts the eyeball radius in pixels, r_{uv} . Knowing the eyeball and iris center coordinates and eyeball radius in pixels makes it possible to fit a 3D model without access to any camera intrinsic parameters.

Since the intrinsic parameters of the camera are not known, the coordinates can only be unprojected into 3D space in pixel units. Thus, 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 $g_c = (\theta, \phi)$, the iris center coordinates can be represented as:

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

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

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

where,

$$\begin{aligned} \theta'_j &= \theta + \sigma \sin\left(\frac{\pi}{4}j + \gamma\right) \\ \phi'_j &= \phi + \sigma \cos\left(\frac{\pi}{4}j + \gamma\right) \end{aligned} \quad (5)$$

For this model-based gaze estimation, θ , ϕ , γ and δ are unknown whereas other variables are provided by the eye region feature points localization step of the network. An iterative optimization method was used to solve this problem, such as conjugate gradient method. And the minimized loss function is represented as:

$$\sum_{0 \leq j \leq 8} (u_{ij} - u'_{ij})^2 + (v_{ij} - v'_{ij})^2 \quad (6)$$

where $u_{ij} - u'_{ij}$ is the estimated pixel coordinates of the j -th iris feature point at each iteration. Calculating person-specific parameters based on calibration samples will adapt this model to a specific person. Gaze correction can be applied with $(\tilde{\theta}, \tilde{\phi}) = (\theta + \Delta\theta, \phi + \Delta\phi)$, where $(\Delta\theta, \Delta\phi)$ is the person-specific angular offset between optical and visual axes.

Fig. 6. 3D eyeball model.

B. HMD Aiming

In early works, the pilot's head posture was used to choose and aim targets. This work tries to combine the gaze direction with head posture for more accurate and flexible human-machine interaction. Through the above derivation, we have acquired the gaze angle (θ, ϕ) of the eye relative to the camera that captures human facial features, which is presented as the pitch and the yaw angles. Since the eyeball does not roll (φ), the roll angle can be regarded as 0. Conventional eye tracking uses a fixed camera mounted on a reference surface to capture and calculate the direction of the human eyes, while the helmet goggles worn by fighter pilots blocks the view of the external camera towards the pilot's face. In order to prevent the interference of the helmet goggles, the camera that captures the image of eyes is installed inside the helmet goggles and fixed with the helmet. Through optical or mixed sensor methods are used in a real aircraft since a gyroscope cannot measure the correct orientations under heavy maneuvers, a 6-axis gyroscope is installed above the helmet to detect head posture for straightforward and easy proof of concept illustration. For the follow-up experiment, a desktop screen and a web-cam directly installed above the screen were used to obtain the aiming point on the screen for the aiming accuracy evaluation. Four coordinate systems, namely the HMD camera frame, the HMD frame, the web-cam frame and space (the cockpit) frame, are established as shown in Fig. 7.

Through the 3D gaze estimation process, the eye gaze direction relative to the camera mounted on the helmet ($g_e^h = \theta, \phi, \varphi$) is acquired. Head attitude relative to the aircraft cockpit is measured by gyroscope $p_h^c = (\alpha, \beta, \gamma)$. The head attitude uniquely determines a unit vector $e_h^c = (\cos\alpha, \cos\beta, \cos\gamma)$, both the p_h^c and e_h^c indicate the direction of the head. Since the gaze direction is relative to the camera mounted on the helmet, it can be considered that the gaze direction is an additional rotation of the head attitude. The gaze angle can be converted into a rotation matrix $R_e^h \in SO(3)$:

$$R_e^h = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{bmatrix} \begin{bmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (7)$$

The unit vector for the final gaze direction can be calculated as, or further expressed as $v_e^c = e_h^c \cdot R_e^h = [x_e \ y_e \ z_e]$:

$$g_e^c = \arccos \left[\frac{x_e}{\sqrt{x_e^2 + y_e^2 + z_e^2}} \frac{y_e}{\sqrt{x_e^2 + y_e^2 + z_e^2}} \frac{z_e}{\sqrt{x_e^2 + y_e^2 + z_e^2}} \right] \quad (8)$$

$$= \begin{bmatrix} \theta_e^c & \phi_e^c & \varphi_e^c \end{bmatrix}$$

And g_e^c indicates the final aiming direction relative to the cockpit considering eye movement and head posture.

V. EXPERIMENTAL EVALUATION

A. HMD Aiming System Setup

To evaluate the effectiveness and robustness of the proposed eye tracking based helmet aiming system, a series of experiments were conducted to quantitatively demonstrate how well the system performs. A total of nine marker points were stationarily set up on the screen, covering most of the screen area, of which the screen is 54cm long and 30cm wide, as shown in Fig. 7. A camera was additionally installed above the screen for estimating the coordinates of the head and marker points coordinates related to the screen. Note that additional camera calibration was required here. In order to distinguish between the camera on the screen and the camera on the helmet, we denote the camera on the helmet C_a and the camera on the screen as C_b . Assume that the x-y plane of the camera C_b coordinate system coincides with the screen, so the coordinates of each landmarks under camera C_b are known. ArUco is a synthetic marker [6], usually used for target detection and positioning, here it is used to detect the position of subject's head. Other available methods such as depth camera can also be adopted to accomplish this job. Since the position of the head $p_h^c = [x_h \ y_h \ z_h]$ is acquired via detecting the ArUco marker and aiming direction, the position of the line of sight on the screen can be calculated:

$$\begin{aligned} X_c &= -Z_h \times \sin(\phi_e^c) + x_h \\ Y_c &= Z_h \times \sin(\theta_e^c) + y_h \end{aligned} \quad (9)$$

where Z_h represents the vertical distance of the head in the C_b coordinate system, that is, the vertical distance between the head and the screen; x_h and y_h are the horizontal and vertical coordinates of the head in the C_b coordinate system; 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} :

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

where E_d is the distance between the estimated gaze position and the real observed position, and E_g represents the distance between the subject and the screen plane.

B. Aiming at Different Distances

To calculate the point of gaze of the user on the screen, the positional relationship between locations on the screen and the head position must be determined. Subjects were first asked to stay still and look at several points for which the positions on the screen are known. While the user is fixating each point on the screen, the eye movements and the head rotation are

TABLE I
AVERAGE ERROR AT DIFFERENT DISTANCES

Distance/cm	Error/cm	Error/degree
40	1.65	2.00
60	2.48	2.07
80	3.02	2.10

measured. The head posture is an absolute value, we need to know the rotation of the head relative to the screen. Then a mapping between the two sets of points is generated through the calibration of the gyroscope. The subjects were about 60cm away from the screen and were requested to look at different marker points on the screen while the estimated gaze points were recorded. Angular value was computed with respect to the target point positions. The experiment tested the impact of the gaze point error at head distances of 40, 60, and 80 cm, where results are shown in Fig. 8. and summarized in Table I.

It is assumed that the coordinates of the estimated gaze point obey a Gaussian distribution as:

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

where μ_x , σ_x^2 , μ_y , and σ_y^2 are sampling means and sampling standard deviations for horizontal and vertical directions. The radius of the light blue circle in the figure is:

$$3\sigma = \sqrt{(3\sigma_x)^2 + (3\sigma_y)^2} \quad (12)$$

which represents the estimated gaze point distribution with a 99.7% confidence.

Note that as the head-to-screen distance increases, the estimated gaze point error for distance increases. When the head is 40cm away from the screen, the average error is 2.00 cm, achieving high accuracy. When the head is 80cm away from the screen, the average drop point error is 3.02cm. From the distance point of view, the point error will gradually increase, but from the angle error point of view, the error does not change significantly. From the 2.00 degree error for 40cm case to the 2.10 degree error for 80cm case, the angular error only increases by 0.1 degree, demonstrating a steadily good and satisfactorily good performance of the proposed HMD aiming system prototype.

C. Aiming by Different Persons

Five people of different ages and different gender voluntarily tested the proposed HMD prototype. Box plots from Fig. 9 shows the accuracy of five voluntary subjects. The gaze accuracy varied due to the different physical characteristics such as eyes, head movement, height and sitting position. The subjects were requested to look at 9 positions on the screen. The estimated gaze points were recorded. We computed the angular value with respect to the target point positions.

The proposed HMD system aiming method was also compared related works such as Skodras et al. [40], Cheung et al. [41], Arar et al. [42], Li et al. [43] and Wang et al. [44], with results presented in Table II. Skodras et al. [40] and

Fig. 7. Prototype experiments under lab environment (a) Experiment setup. The system runs in a desktop environment. (b) Frame establishments. (c) Schematic diagram of aiming point calculation.

TABLE II
PERFORMANCE OF DIFFERENT METHODS

Methods	Accuracy/degree	Number of light sources	Head movement range
Skodras et al. [40]	1.42+-1.35	0	fixed
Cheung et al. [41]	2.27	0	15.5*15.5*15.5
Arar et al. [42]	1.15	5	Free
Li et al. [43]	3.0~4.5	0	Free
Wang et al. [44]	1.3	4	Free
Proposed	2.07	0	Free

Cheung et al. [41] directly maps eye features to screen drop points, rather than estimating the three-dimensional line of sight. Therefore, their head's range of movement is limited, and the error will become larger when the head deviates from the calibrated position. Since the proposed method measures the head posture and the camera is fixed relative to the head, it can achieve free head movement. Arar et al. [42] also directly maps eye feature to gaze point on the screen and achieved accuracy of 1.15 degree. But 5 light sources are used which increase the complexity of the system. Li et al. [43] and Wang et al. [44] estimated the line of sight in 3D and both of them achieved free head movement. Li et al. [43] used a Kinect depth camera instead of conventional cheap cameras. Wang et al. [44] reached 1.3 degree of accuracy and free from calibration but also increased 4 additional light sources. Conclusively, the proposed work achieves an accuracy of 2.07 degree under free head movement without any light source. Note that our equipment is head-mounted and used for aiming on fighter aircrafts, as long as the positional relationship between the helmet and the gaze plane is calibrated, the gaze point prediction can be achieved.

VI. CONCLUSION

This work proposed a comprehensive and effective HMD aiming system to provide a freer, faster and more flexible human machine interaction method for aircraft pilots. Detailed design and algorithm behind were presented and discussed. The eye gaze direction is obtained by combining the eye to helmet orientation and helmet to cockpit orientation. A custom made wide angle monocular camera installed inside a flight helmet is used to capture eye images to circumvent

the blocking issues by the outside helmet goggle. Proof of concept prototype experiments illustrates a steadily good and satisfactorily good performance of the proposed HMD aiming system prototype, and illustrates advantages such as free head movement and no additional light source compared to some related works. Conclusively, the proposed HMD aiming system offers a feasible alternative to the current HMD technology and provides practical human machine interaction remarks.

REFERENCES

- [1] Clarence E Rash, Michael B Russo, Tomasz R Letowski, and Elmar T Schmeisser. Helmet-mounted displays: Sensation, perception and cognition issues. Technical report, ARMY AEROMEDICAL RESEARCH LAB FORT RUCKER AL, 2009.
- [2] Seonwook Park, Xucong Zhang, Andreas Bulling, and Otmar Hilliges. Learning to find eye region landmarks for remote gaze estimation in unconstrained settings. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, pages 1–10, 2018.
- [3] Sina Honari, Pavlo Molchanov, Stephen Tyree, Pascal Vincent, Christopher Pal, and Jan Kautz. Improving landmark localization with semi-supervised learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1546–1555, 2018.
- [4] Alejandro Newell, Kaiyu Yang, and Jia Deng. Stacked hourglass networks for human pose estimation. In *European conference on computer vision*, pages 483–499. Springer, 2016.
- [5] Erroll Wood, Tadas Baltrušaitis, Louis-Philippe Morency, Peter Robinson, and Andreas 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*, pages 131–138, 2016.
- [6] Sergio Garrido-Jurado, Rafael Muñoz-Salinas, Francisco José Madrid-Cuevas, and Manuel Jesús Marín-Jiménez. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition*, 47(6):2280–2292, 2014.
- [7] Solomon Ayalew Mekonnen, MB Asrat, and S Ramasamy. A helmet cueing system based firing control for anti-aircraft gun prototype. *Advances in Military Technology*, 16(1):19–33, 2021.
- [8] Sławomir Michalak, Jerzy Borowski, Andrzej Szelmanowski, and Andrzej Pazur. The polish helmet mounted display systems for military helicopters. In *2016 IEEE Metrology for Aerospace (MetroAeroSpace)*, pages 353–358. IEEE, 2016.
- [9] Garry W Orf. Joint helmet-mounted cueing system (jhmc) helmet qualification testing requirements. In *Helmet and Head-Mounted Displays III*, volume 3362, pages 118–123. International Society for Optics and Photonics, 1998.
- [10] James 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*, pages 20–26. IEEE, 1997.
- [11] Robert Atac and Eric Foxlin. Scorpion hybrid optical-based inertial tracker (hobit). In *Head and Helmet-Mounted Displays XVIII: Design and Applications*, volume 8735, page 873502. International Society for Optics and Photonics, 2013.

Fig. 8. (a) Results of aiming at 40cm. (b) Results of aiming at 60cm. (c) Results of aiming at 80cm. Points of gaze are marked as the blue dots, while the target point is marked as a red cross. Red dots are the average of the blue dots. The size of the coordinate system is consistent with the screen size, and the position of the camera is at the origin of the coordinates. The light blue circle indicates 99.7% confidence

- [12] Robert Atac and Tony Bugno. Qualification of the scorpion helmet cueing system. In *Head-and Helmet-Mounted Displays XVI: Design and Applications*, volume 8041, page 80410J. International Society for Optics and Photonics, 2011.
- [13] Robert Atac. Applications of the scorpion color helmet-mounted cueing system. In *Head-and Helmet-Mounted Displays XV: Design and Applications*, volume 7688, page 768803. International Society for Optics and Photonics, 2010.
- [14] Robert Atac, Scott Spink, Tom Calloway, and Eric Foxlin. Scorpion hybrid optical-based inertial tracker (hobit) test results. In *Display Technologies and Applications for Defense, Security, and Avionics VIII; and Head-and Helmet-Mounted Displays XIX*, volume 9086, page 90860U. International Society for Optics and Photonics, 2014.
- [15] Chris J Smith. Design of the eurofighter human machine interface. *Air & Space Europe*, 1(3):54–59, 1999.
- [16] Stephen J Carter and Alexander A Cameron. Eurofighter helmet-mounted display: status update. In *Helmet-and Head-Mounted Displays*

Fig. 9. box-plot of gaze accuracy.

- V, volume 4021, pages 234–244. International Society for Optics and Photonics, 2000.
- [17] Laura Sesma, Arantxa Villanueva, and Rafael Cabeza. Evaluation of pupil center-eye corner vector for gaze estimation using a web cam. In *Proceedings of the symposium on eye tracking research and applications*, pages 217–220, 2012.
- [18] Michael Xuelin Huang, Tiffany CK Kwok, Grace Ngai, Hong Va Leong, and Stephen CF Chan. Building a self-learning eye gaze model from user interaction data. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 1017–1020, 2014.
- [19] Jose Sigut and Sid-Ahmed Sidha. Iris center corneal reflection method for gaze tracking using visible light. *IEEE Transactions on Biomedical Engineering*, 58(2):411–419, 2010.
- [20] Pieter Blignaut. A new mapping function to improve the accuracy of a video-based eye tracker. In *Proceedings of the south african institute for computer scientists and information technologists conference*, pages 56–59, 2013.
- [21] Pieter Blignaut and Daniël Wium. The effect of mapping function on the accuracy of a video-based eye tracker. In *Proceedings of the 2013 conference on eye tracking south africa*, pages 39–46, 2013.
- [22] Pieter Blignaut. Mapping the pupil-glint vector to gaze coordinates in a simple video-based eye tracker. *Journal of Eye Movement Research*, 7(1), 2014.
- [23] Rakshit S Kothari, Aayush K Chaudhary, Reynold J Bailey, Jeff B Pelz, and Gabriel J Diaz. Ellseg: An ellipse segmentation framework for robust gaze tracking. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2757–2767, 2021.
- [24] Wolfgang Fuhl, Hong Gao, and Enkelejda Kasneci. Neural networks for optical vector and eye ball parameter estimation. In *ACM Symposium on Eye Tracking Research and Applications*, pages 1–5, 2020.
- [25] Kang Wang and Qiang Ji. Real time eye gaze tracking with 3d deformable eye-face model. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1003–1011, 2017.
- [26] Erroll Wood, Tadas Baltrušaitis, Louis-Philippe Morency, Peter Robinson, and Andreas Bulling. A 3d morphable eye region model for gaze estimation. In *European Conference on Computer Vision*, pages 297–313. Springer, 2016.
- [27] Li Sun, Zicheng Liu, and Ming-Ting Sun. Real time gaze estimation with a consumer depth camera. *Information Sciences*, 320:346–360, 2015.
- [28] Erroll Wood and Andreas Bulling. Eyetab: Model-based gaze estimation on unmodified tablet computers. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 207–210, 2014.
- [29] Quan Wen, Derek Bradley, Thabo Beeler, Seonwook Park, Otmar Hilliges, Junhai Yong, and Feng Xu. Accurate real-time 3d gaze tracking using a lightweight eyeball calibration. In *Computer Graphics Forum*, volume 39, pages 475–485. Wiley Online Library, 2020.
- [30] Meng Liu, Youfu Li, and Hai Liu. 3d gaze estimation for head-mounted eye tracking system with auto-calibration method. *IEEE Access*, 8:104207–104215, 2020.
- [31] Petr Kellnhofer, Adria Recasens, Simon Stent, Wojciech Matusik, and Antonio Torralba. Gaze360: Physically unconstrained gaze estimation in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6912–6921, 2019.
- [32] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. Eye tracking for

everyone. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2176–2184, 2016.

- [33] Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. Appearance-based gaze estimation in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4511–4520, 2015.
- [34] Kenneth Alberto Funes Mora, Florent Monay, and Jean-Marc Odobez. Eyediap: A database for the development and evaluation of gaze estimation algorithms from rgb and rgb-d cameras. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 255–258, 2014.
- [35] Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. Mpiigaze: Real-world dataset and deep appearance-based gaze estimation. *IEEE transactions on pattern analysis and machine intelligence*, 41(1):162–175, 2017.
- [36] Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. It's written all over your face: Full-face appearance-based gaze estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 51–60, 2017.
- [37] Rakshit Kothari, Shalini De Mello, Umar Iqbal, Wonmin Byeon, Seonwook Park, and Jan Kautz. Weakly-supervised physically unconstrained gaze estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9980–9989, 2021.
- [38] Neelabh Sinha, Michal Balazsia, and François Bremond. Flame: Facial landmark heatmap activated multimodal gaze estimation. *arXiv preprint arXiv:2110.04828*, 2021.
- [39] Jing Yang, Qingshan Liu, and Kaihua Zhang. Stacked hourglass network for robust facial landmark localisation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 79–87, 2017.
- [40] Evangelos Skodras, Vasileios G Kanas, and Nikolaos Fakotakis. On visual gaze tracking based on a single low cost camera. *Signal Processing: Image Communication*, 36:29–42, 2015.
- [41] Yiu-ming Cheung and Qimu Peng. Eye gaze tracking with a web camera in a desktop environment. *IEEE Transactions on Human-Machine Systems*, 45(4):419–430, 2015.
- [42] Nuri Murat Arar, Hua Gao, and Jean-Philippe Thiran. Towards convenient calibration for cross-ratio based gaze estimation. In *2015 IEEE Winter Conference on Applications of Computer Vision*, pages 642–648. IEEE, 2015.
- [43] Jianfeng Li and Shigang Li. Gaze estimation from color image based on the eye model with known head pose. *IEEE Transactions on Human-Machine Systems*, 46(3):414–423, 2015.
- [44] Kang Wang and Qiang Ji. 3d gaze estimation without explicit personal calibration. *Pattern Recognition*, 79:216–227, 2018.

Quanlong Qiu received his B.S. degree from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2016. He is currently a graduate student working towards the M.S. degree at the School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou, China. His research interest covers computer vision and pattern recognition.

Jie Zhu received his B.S. degree from Xiamen University, Xiamen, China, in 2018. He is currently a graduate student pursuing the M.S. degree at School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou, China. His research interests include computer vision and machine learning.

computer vision and machine learning.

Gou Chao received his B.S. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2012 and the Ph.D. degree from the University of Chinese Academy of Sciences (UCAS), Beijing, China, in 2017. From September 2015 to January 2017, he was supported by UCAS as a joint-supervision Ph.D student in Rensselaer Polytechnic Institute, Troy, NY, USA. Dr. Gou is currently an Assistant Professor with School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou, China. His research interests include

Mengtang Li received the B.S. degree with honor in Astronautics Engineering from Northwestern Polytechnical University, Xi'an, China, in 2014, the M.S. degree in Electrical Engineering and the Ph.D. degree in Mechanical Engineering from Vanderbilt University, Nashville, TN, USA, in 2016 and 2020, respectively. He was a visiting scholar at Sarver Heart Center, School of Medicine, University of Arizona, Tucson, AZ, USA from 2018 - 2020. Dr. Li is currently an assistant professor in School of Intelligent Systems Engineering in Sun Yat-Sen University, Guangzhou, China. His research interests include mechanical circulatory support device design, optimization and control; fluid powered robotic system design and control; mobile unmanned robots/vehicles/aircrafts control.