Research on the Gaze Direction of Head-Eye Data Fusion

Xin Xu*

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, Shaanxi, China
E-mail: Ariel970504@163.com
*corresponding author

Changyuan Wang

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, Shaanxi, China
E-mail: Cyw901@163.com

***Abstract*—The line of sight refers to the gaze direction of the human eye and reflects the focus of human attention. Head movement is an important accompanying behavior in the process of human gaze, and it is of great significance to human visual attention. This paper intends to combine the gaze focus calculation modeling of head movement and eye movement combined with deep learning data fusion. By combining the gaze direction calculation model of data fusion and neural network algorithm, deep learning technology is used to reveal the relationship between head movement and eye movement, and the data of head movement and eye movement are merged to realize accurate and fast real-time gaze spatial direction calculation. New ideas for improving the efficiency, reliability, usability and functionality of the gaze tracking system. In this paper, the convolutional neural network method is used, and the classification accuracy of the line of sight direction reaches 99% when the head posture is free.**

Keywords-Head Pose Estimation; Pupil Center Detection; Line of Sight Direction; Data Fusion; Neural Network

I. INTRODUCTION

The cockpits of modern advanced fighter jets mostly adopt a one-level three-down display mode. With the development of avionics technology, the down-view display gradually adopts a large-size overall touch screen, which improves ergonomics by integrating visual information. The large number of manual operations that still exist are an important factor that limits the current efficiency of human-computer interaction. The gaze tracking technology obtains the direction of the gaze by measuring the positioning and posture of the human eye. Based on human physiological characteristics, the gaze response is fast and accurate, and is not affected by high overload. It has obvious advantages in the field of aviation human-computer interaction.

A. Eye tracking technology

Eye tracking technology refers to the use of certain features that are relatively unchanged during eye movement to obtain pupil data, including pupil center coordinates, contours and other parameters. Due to the superior performance of deep learning technology in the field of computer vision, currently, eye tracking technology is mainly developed around this technology. Fuhl et al. [1] proposed a coarse-to-fine pupil detection model-pupilnet based on convolutional neural networks. The team input the eye pictures into the rough recognition network in blocks, and the area with the highest score is the pupil. Rough area, and then input it into the precise recognition network to extract pupil parameters. Due to the strong dependence of the neural network on the data set, the unbalanced distribution of the image types in the data set will cause the model to have larger detection errors for the relatively small image types . Based on this feature, Shaharam Eivazi et al. [2] proposed a targeted image enhancement method for pupil detection errors caused by mirror reflection, sunlight reflection, blurring, etc., and used the enhanced image as a data set to train classic convolution The neural network recognition model achieves the best simultaneous detection model. Fuhl et al. [3] proposed to use Cycle GANs [4] to enhance the image, increase the richness of the data set, and perform pupil segmentation on it. The segmentation result only retains the shape and position of the pupil, and filters out all the noise,

DOI: 10.21307/ijanmc-2021-031

thereby improving the pupil Detection accuracy. The above models are all based on the pupil characteristics of a single frame image, which is also a characteristic of convolutional neural networks. It fails to introduce the pupil movement characteristics embodied in consecutive frames. At the same time, the high complexity of the model makes it difficult to apply to real-time tracking tasks. In order to solve the above Two problems, this project intends to introduce Long Short-Term Memory (LSTM) [5], use its ability to process continuous sequences to introduce pupil movement features to improve recognition accuracy, and add a "pruning" operation [6] Reduce model calculation speed to achieve real-time tracking performance.

B. Head pose estimation

Head pose refers to the use of computer vision and pattern recognition technology to estimate the orientation of the head in a digital image. The head movement data mainly includes the three-axis spatial position and the three-axis posture. Among them, the three-axis posture data is composed of three sets of yaw, pitch and roll data, reflecting the head space rotation state. Image-based head motion tracking mainly uses facial features as reference points to match the three-dimensional model of the face through the recognition of the head and face.

Bao et al. [7] proposed using a three-layer convolutional neural network for head pose estimation, and used a coarse-to-fine method in the model training process. For the application of deep learning in the field of head pose estimation, Patacchiola et al.[8] conducted experiments on four convolutional neural networks and showed that the shallow architecture shows better performance on small-scale data sets, while the deep architecture is more Suitable for large-scale data.

C. Line of sight estimation

The line of sight estimation mainly solves the nonlinear mapping relationship between the collected human body information and the line of sight. Based on the difference of the collection equipment, it is mainly divided into wearable and non-wearable line of sight estimation methods. For wearable devices, Thiago Santini et al. [9] used glasses-type acquisition devices and assumed that the gaze tracker was still a rigid body after calibration. On this basis, the tracker coordinates were used to quickly estimate the gaze angle, and then a second-order polynomial regression model was used. Mapping it to the two-dimensional field of view image collected by the tracker to obtain a precise landing point. Although the model has a good detection effect on its experimental data set, its dependence on the pupil contour leads to occlusion and other interference which will cause a larger line of sight Estimated error.

For non-wearable devices, if the freedom of the user's head movement is restricted, the line of sight direction can be estimated based on facial features alone. Su Haiming et al. [10] proposed to determine the position of the human eye through a face positioning algorithm and use the pupil template to detect it. The pupil area is then used to establish the gaze point mapping relationship using the neural network model. When the head posture is fixed, the recognition accuracy rate of 96.74% is achieved. This type of method limits the range of head movement. Although it has achieved high recognition performance, it can only be applied to specific fields. Regarding the line of sight estimation in free pose, S Park et al. [11] pointed out that in this case, high-precision line of sight estimation based on the eye picture is not suitable for the task. The spatial position of the pupil center of the human eye that determines the direction of the line of sight is not observable in the picture. . Therefore, this type of method adds head posture parameters to the eye movement image technology to establish a head-eye-line-of-sight model. Xucong Zhang et al. [12] proposed GazeNet, which takes normalized eye pictures and head pose parameters as parameters, uses VGG16 [13] as the basic network model to extract the feature spectrum of the eye pictures, and cascades the head pose angle to The output vector of the first fully connected layer of the line-of-sight estimation network is compensated, and the line-of-sight angle is estimated through the line-of-sight estimation network.

Based on the existing research foundation, it is

found that in the human visual gaze system, head movement expands the person's field of vision. At the same time, if the head space movement and posture cannot be clarified, the eye movement coordinate system cannot be clarified. Therefore, this article starts from the non-wearable head movement and eye movement detection and tracking, through image recognition measurement, head space movement and eye movement measurement, combined with data fusion and deep learning neural network algorithm technology to establish a line of sight calculation model, Clarify the mathematical relationship between head movement, eye movement and the direction of the line of sight, and realize the acquisition of the direction of the binocular line of sight.

II. RELATED WORK

A. Based on non-wearable multi-lens camera data acquisition and simulation flight experimental platform construction

In this paper, a non-wearable device is used to collect data from the head and eyeballs, that is, to set up multiple cameras and infrared light sources in the environment to collect images of the head, face and eyeballs. The control system is divided into upper computer and lower computer. The upper computer controls the multi-eye camera and infrared light source through the control chip. The lower computer is responsible for processing the image data and measuring the direction of sight. At the same time, the simulated flight platform is improved. Based on the original six-axis full-motion platform, according to the design of the J10 cockpit, a head-up display, a down-view display and a visual simulation display are added. The simulated flight experiment platform is shown in Figure 1.

Figure 1. Flight simulation platform

B. Head movement measurement

Head movement measurement refers to measuring the movement trajectory of the subject's head relative to an absolute reference point (usually the spatial coordinates of the back of the head, forehead center or neck in the initial sitting position) in real or virtual space, and calculating the movement paradigm to obtain The mathematical model of the environment where the research object is located [14]. In the actual environment, it is necessary to pay attention to the vector components of the three-dimensional orthogonal coordinate system with the calibration point as the origin, and use the orthogonal vector to describe the movement characteristics of the head; and for the head movement measurement in the virtual space, the two-dimensional image is mainly used. , One of the vectors needs to be converted into a head depth function to obtain the motion behavior in the actual environment through the two-dimensional image, as shown in 2.

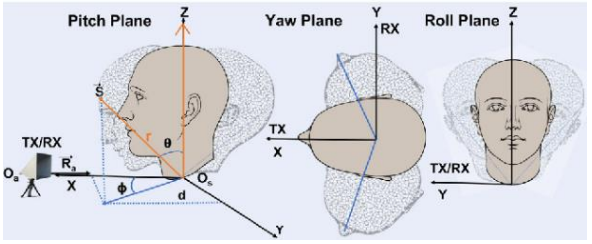

Figure 2. Spatial coordinates of head movement measurement (neck origin)

With the rapid development of computer vision and the decline in the price of motion sensor hardware, many image and sensor-based head motion measurement solutions have also been proposed. As a part of the human torso, the development of motion measurement and the field of motion capture of the head are advancing almost at the same time [15-17], and head motion measurement has driven the development of human-computer interaction, assisted driving, and machine-human collaboration; similarly, Since the head image contains information such as eyes, facial expressions, and facial features, head movement measurement is also used in live body detection, focus tracking, and target recognition. At present, the more accurate solution in the above-mentioned applications is motion measurement based on motion sensors, the more common is motion measurement based on image displacement algorithms, and the more cutting-edge is image-based deep learning motion measurement.

The sensor-based motion detection scheme is based on the motion sensor placed on the head. The head motion is obtained by calculating the relative position of the current head's spatial position and the initial position (or calibration point) through the acceleration and angular velocity vectors built into the sensor. Way. Existing hardware can easily reach the sampling rate of 30Hz and above. The movement track can be obtained by recording the displacement of continuous time. Taking the sensor hardware used in this article as an example, its appearance is shown in 3.

Figure 3. Xsens® MTi-G motion sensor used in the article

The known acceleration is shown in Eq. (1). The natural coordinate system decomposes to obtain the tangential acceleration and the normal acceleration, and the acceleration decomposes in the circular motion to obtain the tangential acceleration and the centripetal acceleration.

$$\alpha = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{dv}{dt} \tag{1}$$

When studying the problem, the sensor fixed on the head can be regarded as a mass point in space, then the sensor acceleration is decomposed into a_t tangential acceleration and a_n normal acceleration, as shown in Eq. (2).

$$a = a_t + a_n \tag{2}$$

Furthermore, the acceleration modulus, angular velocity vector, and angular acceleration vector shown in Eq. (3) to Eq. (5) are obtained.

$$|a| = \sqrt{a_t^2 + a_n^2} \tag{3}$$

$$\omega = \lim_{\Delta t \to 0} \frac{\Delta n}{\Delta t} = \frac{dn}{dt} \tag{4}$$

$$\alpha = \lim_{\Delta t \to 0} \frac{\Delta \omega}{\Delta t} = \frac{d\omega}{dt} \tag{5}$$

According to the vector obtained above, the trajectory of the mass point (the center of gravity of the sensor) in space can be obtained as shown in Figure 4. Figure 5 shows the research process of head movement measurement based on motion sensors.

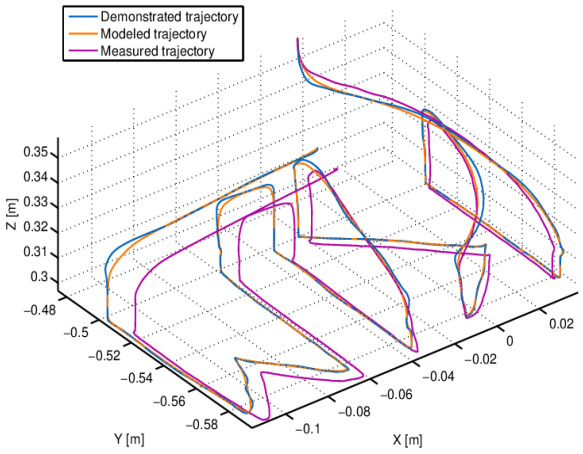

Figure 4. Schematic diagram of particle motion in space

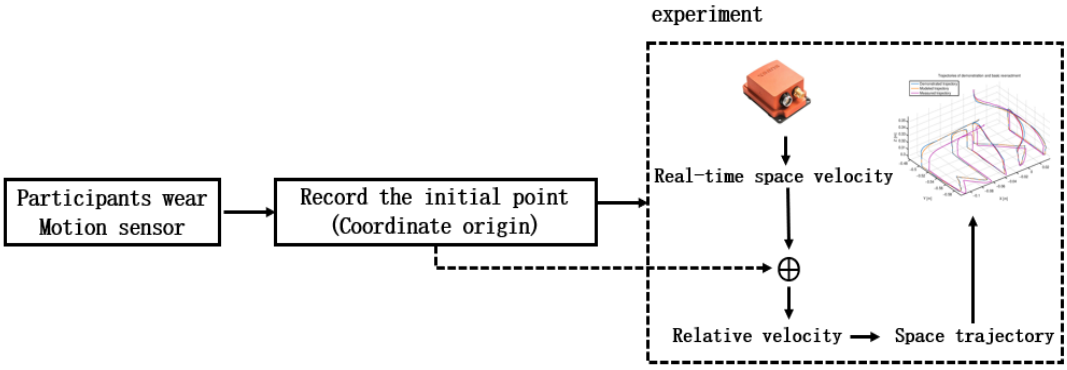

Figure 5. Experiment process of head movement measurement based on motion sensor

C. Eye movement detection

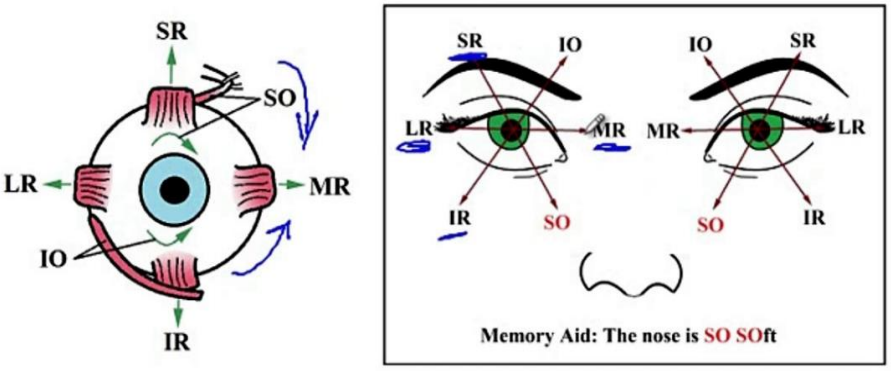

Figure 6. Eye movement measurement plane coordinates (origin of nose tip)

Eye movement measurement refers to measuring the relative movement of the eyeball and this point with a relative reference point as the origin of the coordinates (usually the center of the brow and the tip of the nose). Because the eyeball is restricted to close to two degrees of freedom by the human skull structure, four straight eyeball muscles and two oblique muscles [18], the displacement of the eyeball perpendicular to the longitudinal section of the skull is less than 10 % [19-21]. At the same time, most of the tasks of eye movement are to quantify the center of the pupil and the point of sight. The measurement of the task can be simplified into a two-dimensional domain, as shown in 6.

Eye movement measurement is different from head movement measurement, which requires the algorithm to have higher calculation accuracy. There are two main types of eye movement measurement methods: measurement based on head-mounted infrared camera and measurement based on non-contact infrared camera. The reason why the infrared camera is selected as the eye movement capture device is that the reflection effect of the human eye pupils on infrared wavelengths is better than that of visible light [22-24]. Figure 7 shows the study of pupil reflectivity at different infrared wavelengths by scholars; Figure 8 compares the imaging effects of RGB cameras and infrared cameras on eye features. In the measurement range of the eyeball diameter, the state of pupil zoom cannot be easily ignored, especially in terms of gaze tracking, the pupil zoom represents the change of the focus of the line of sight, but also the change of the range of attention [25].

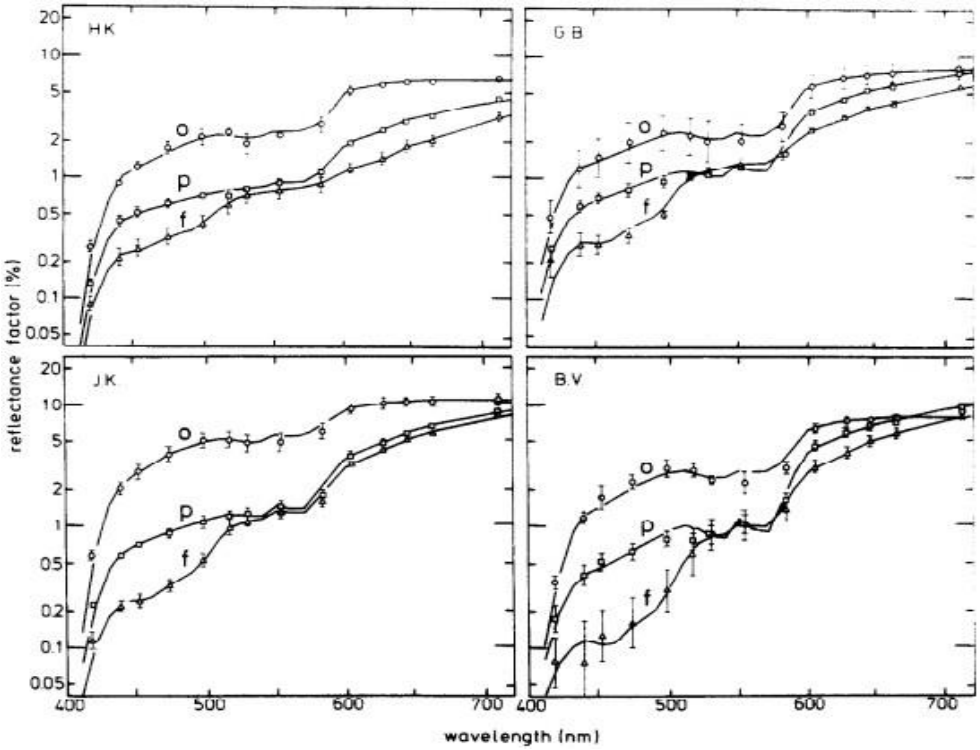

Figure 7. Pupil reflectance curve at different wavelengths

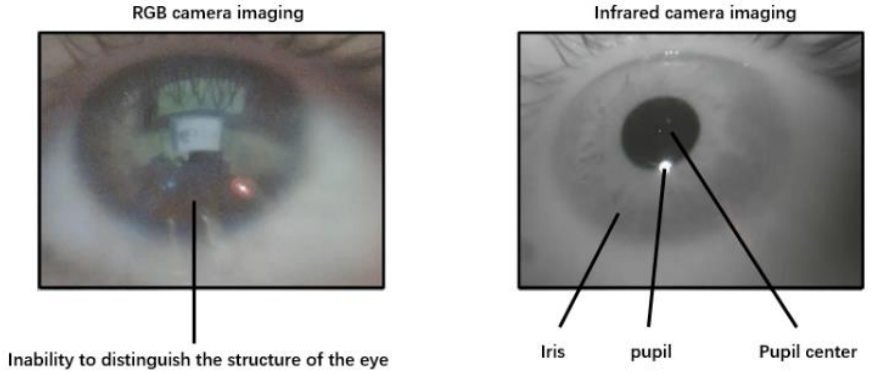

Figure 8. Comparison of pupil imaging between RGB camera and infrared camera

Because the head space position and posture data can clarify the eye movement coordinate system, they are indispensable data in the calculation of the line of sight direction. This paper intends to fuse head space movement and eye movement data, establish a gaze direction calculation model combining data fusion and deep learning neural network algorithm, and conduct machine learning training neural network through calibration experiments to clarify the relationship between head movement and gaze direction. Realize the measurement of the line of sight.

III. TECHNICAL ROUTE

This article puts forward requirements for multiple indicators such as head movement eye movement measurement accuracy, field of view, device portability, non-contact conditions, and rapid output of results. There is currently no such highly integrated research reference and hardware conditions in related fields , Especially the field of view in the head movement measurement and the calculation accuracy in the eye movement measurement. These two indicators are the technical difficulties and innovative entry points of

this article. The portability of the equipment requires that the entire system can be easily and quickly deployed on various platforms, that is, to reduce the proportion of customized equipment; non-contact conditions require the use of the testees to reduce the burden and learning costs; the rapid output of the results requires that the text can be implemented, the traditional The image processing method takes a lot of time to calculate the head movement, and involves the solution of a large number of nonlinear equations.

The motion sensor can accurately capture the motion trajectory, speed, acceleration and torsion angle of the participant's limbs in space. These data will provide accurate labels during the text construction stage to modify the performance of the model. The trinocular infrared camera layout is selected as the only hardware in the application stage of this article. In the application process, the motion sensor will no longer be used, but the infrared images of different angles collected by the trinocular camera will be directly calculated.

Deep neural networks have developed rapidly in recent years. One of the characteristics is that training models are time-consuming and the amount of calculation in application is generally less than traditional methods. This is because the nature of deep neural networks is to fit piecewise nonlinear functions through multiple linear equations. The concept of gradient, divergence, etc. is converted to linear operations. This attribute determines that deep neural networks will not generate a large number of partial differential equations at the application stage, but instead are alternative multiplication operations. This article chooses deep neural network as a tool, responsible for deriving the technical difficulties mentioned above.

In summary, the technical route of the construction phase of this article is shown in Figure 9, and the process of embedding the motion paradigm into the deep neural network model is shown in Figure 10.

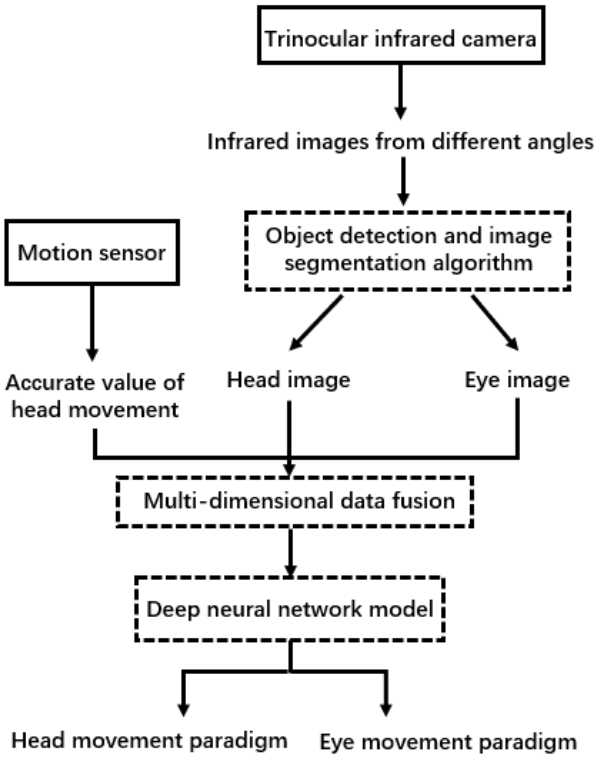

Figure 9. Technical route

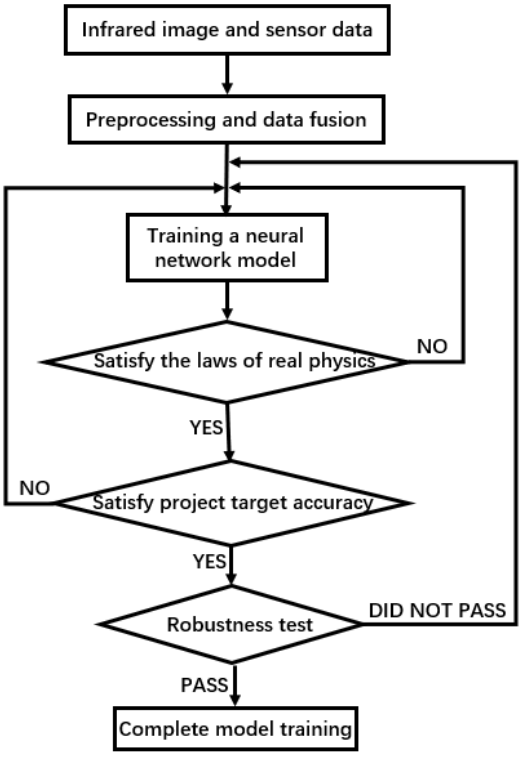

Figure 10. Model training process

IV. EXPRIMENT

A. Data collection

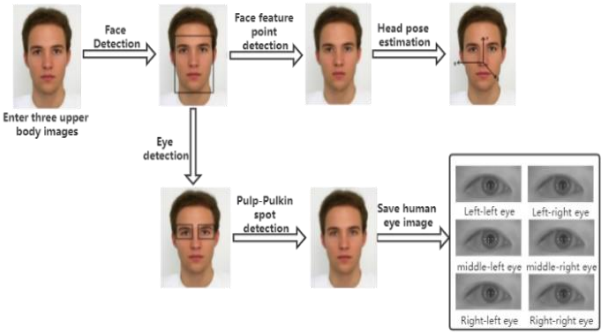

Figure 11. Training set collection process

The data set in this paper is collected in the laboratory, and the process of collecting eye images and head posture is shown in Figure 11.

The data collection process in this article is as follows: A real-time data collection program is implemented, three pure black pictures of 2560×1440 size are displayed on the full screen, and a red circle appears at a random position on the screen, and the image collector is required to look at the red circle target and tap Space bar, this is to ensure that the collector is focused. There are no other requirements for the image collector, and the head can move freely. While gazing at the point, press the space bar, the program will save the three upper body photos of the image collector at that moment, and save the corresponding head posture and gaze point to a text file (one-to-one correspondence with the number), and then click Another red circled target appears at a random location on the screen. The fixation point is the coordinate value of the red circle target randomly appearing on the computer. The 3D coordinates of the fixation point can be determined by the camera posture and the position of the computer screen in the world coordinate system. The direction of the line of sight is the distance from the 3D coordinates of the eye to the 3D position of the fixation point Connect. A schematic diagram of the data collection process is shown in Figure 12, and part of the collected face pictures are shown in Figure 13.

Figure 12. Data collection diagram

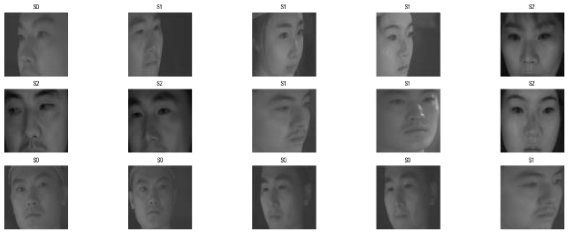

Figure 13. Part of the data set display

B. Facial feature recognition

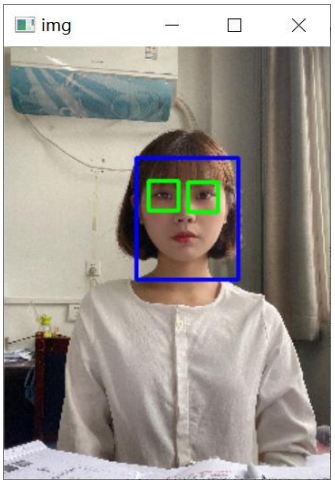

Figure 14. Facial feature recognition

After comparative research, the Adaboost algorithm is selected for face detection. The Adaboost algorithm selects the best features from a large number of Haar features and converts them into weak classifiers for classification and use, so as to achieve the purpose of classifying the target. The face recognition result based on the Adaboost algorithm is shown in Figure 14.

C. Human eye feature recognition

The facial features have been successfully recognized through the above process, and the next step is to recognize the eye area. Recognize

the eye area based on the Adaboost algorithm. As shown in the figure, the white point indicated by 15 is the light spot (Pulchin spot) formed on the human cornea by the near-infrared light source on the screen, and the darkest area in the center of the human eye area pointed to by 2 is the human eye pupil area. The slightly shallower area pointed to by 3 is the iris area of the human eye.

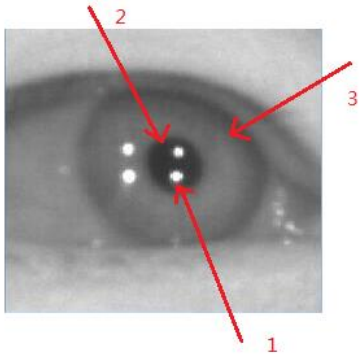

Figure 15. Eye area feature map

Take a screenshot of the identified eye area. The three photos at the same time are divided into one group, each group gets a total of 6 eye photos, and the pupil-Pulchin spot picture can be clearly seen, as shown in Figure 16.

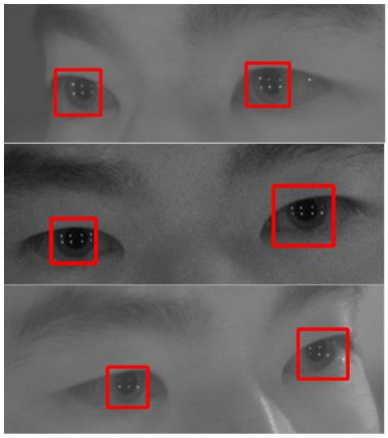

Figure 16. Accuracy comparison chart

D. Training and testing

The two human eye images in the data set collected by this article correspond to a head posture and line of sight direction (assuming that the left eye and the right eye have the same line of sight direction).

The convolutional neural network model used in this article is trained under the deep learning open source framework caffe. The training model requires two configuration files: solver. prototxt and train test. Prototxt

The solver. prototxt mainly includes the setting of training parameters, including the number of iterations, weight attenuation coefficient learning rate, impulse, display test error iteration interval, GPU and CPU settings, etc. The parameter selection in this paper selects the optimal parameters through 10-fold cross-validation, and then retrains all the training sets with the selected parameters to obtain the line-of-sight estimation model.

Train_test. prototxt mainly includes the settings of the convolutional neural network structure, specifically the directory of training data and test data, convolutional layer and pooling layer settings, fully connected layer settings, loss function, pooling type, etc.

On the data set collected by myself, the accuracy graphs of the training set and the test set are shown in Figure 17, where the horizontal axis represents the number of iterations, the blue line represents the training error, and the orange line represents the test error.

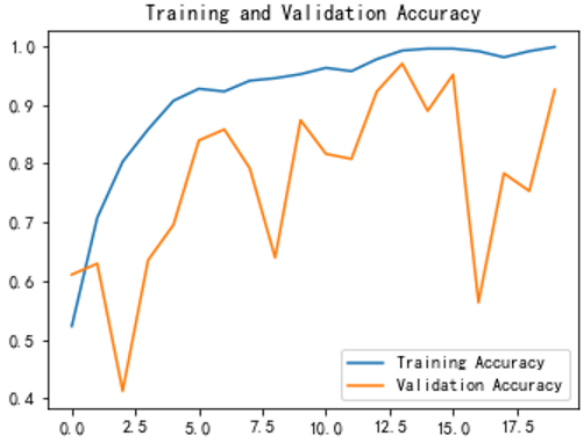

Figure 17. Pupil-Pulchin spot image recognition

The iterative loss curve of training on the self-collected data set is shown in Figure 18, where the horizontal axis represents the number of iterations, the blue line represents the training error, and the orange line represents the test error.

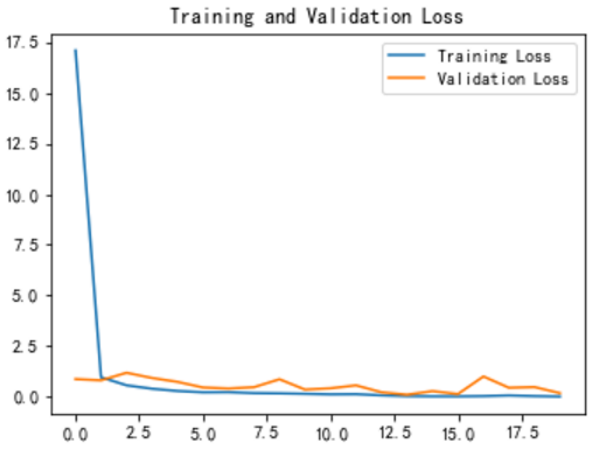

Figure 18. Loss curve

Confusion matrix, also known as error matrix, is a standard format for precision evaluation. Through the confusion matrix, we can clearly see the number of correct identifications and the number of incorrect identifications in each category in the three directions, and then quickly help us analyze the misclassification of each category.In this paper, the convolutional neural network method is used. When the head posture is free, the classification accuracy of the line of sight direction reaches 99%. The classification result is shown in Figure 19.

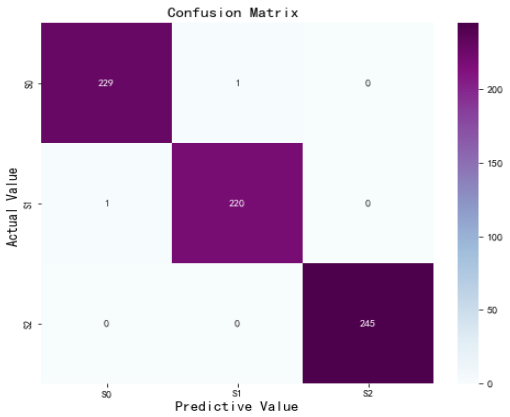

Figure 19. Confusion matrix

V. CONCLUSION

Sight tracking is widely used in psychology, graphics, software engineering, pattern recognition, human-computer interaction, medicine, advertising psychology, military and many other fields. It has strong practical value. Therefore, gaze tracking has become a computer vision and pattern in recent years. Hot topics in the field of identification.In this paper, the pupil-Pulchin spot image method is used to measure the eye movement by the non-wearable image measurement method, combined with the head posture data, through the data fusion combined with the neural network algorithm of the line of sight calculation model, and the deep learning technology is used to reveal the head movement. The relationship with the gaze direction, fusion of head movement and eye movement data, to achieve accurate and fast real-time gaze spatial direction calculation. In this paper, the convolutional neural network method is used, and the classification accuracy of the line of sight direction reaches 99% when the head posture is free.

REFERENCES

[1] Fuhl W, Santini T, Kasneci G. Pupilnet: Convolutional neural networks for robust pupil detection[J]. arXiv preprint arXiv:1601.04902, 2016.

[2] Eivazi S, Santini T, Keshavarzi A. Improving real-time CNN-based pupil detection through domain-specific data augmentation[C]. Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications, 2019: 1-6.

[3] Fuhl W, Geisler D, Rosenstiel W. The applicability of Cycle GANs for pupil and eyelid segmentation, data generation and image refinement[C]. Proceedings of the IEEE International Conference on Computer Vision Workshops, 2019: 0-0.

[4] Zhu J-Y, Park T, Isola P. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]. Proceedings of the IEEE international conference on computer vision, 2017: 2223-2232.

[5] Tsironi E, Barros P, Weber C. An analysis of convolutional long short-term memory recurrent neural networks for gesture recognition[J]. Neurocomputing, 2017, 268: 76-86.

[6] Liu Z, Li J, Shen Z. Learning efficient convolutional networks through network slimming[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2736-2744.

[7] Bao J, Ye M. Head pose estimation based on robust convolutional neural network[J]. Cybernetics and Information Technologies, 2016, 16(6): 133-145.

[8] Patacchiola M, Cangelosi A. Head pose estimation in the wild using convolutional neural networks and adaptive gradient methods[J]. Pattern Recognition, 2017, 71: 132-143.

[9] Santini T, Niehorster D C, Kasneci E. Get a grip: slippage-robust and glint-free gaze estimation for real-time pervasive head-mounted eye tracking[C]. Proceedings of the 11th ACM symposium on eye tracking research & applications, 2019: 1-10.

[10] Su Haiming, Hou Zhenjie, Liang Jiuzhen. A gaze tracking method using geometric features of human eyes [J]. Journal of Image and Graphics, 2019(201906): 914-923.

[11] Park S, Spurr A, Hilliges O. Deep pictorial gaze estimation[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 721-738.

[12] Zhang X, Sugano Y, Fritz M. Mpiigaze: Real-world dataset and deep appearance-based gaze estimation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 41(1): 162-175.

[13] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv:1409.1556, 2014.

[14] Mase K, Watanabe Y, Suenaga Y. Real-time head motion detection system. Sensing and Reconstruction of Three-Dimensional Objects and Scenes: International Society for Optics and Photonics; 1990. p. 262-8.

[15] Rolland JP, Davis LD, Baillot Y. A survey of tracking technologies for virtual environments. Fundamentals of wearable computers and augmented reality: CRC Press; 2001. p. 83-128.

[16] Zhou H, Hu HJBsp, control. Human motion tracking for rehabilitation—A survey. 2008;3:1-18.

[17] Al-Rahayfeh A, Faezipour MJIjoteih, medicine. Eye tracking and head movement detection: A state-of-art survey. 2013; 1:2100212.

[18] Von Lüdinghausen MJCATOJotAAoCA, Anatomists tBAoC. Bilateral supernumerary rectus muscles of the orbit. 1998; 11:271-7.

[19] Raudonis V, Simutis R, Narvydas G. Discrete eye tracking for medical applications. 2009 2nd International Symposium on Applied Sciences in Biomedical and Communication Technologies: IEEE; 2009. p. 1-6.

[20] Botha CP, de Graaf T, Schutte S, Root R, Wielopolski P, van der Helm FC, et al. MRI-based visualisation of orbital fat deformation during eye motion. Visualization in medicine and life sciences: Springer; 2008. p. 221-33.

[21] Glarin RK, Nguyen BN, Cleary JO, Kolbe SC, Ordidge RJ, Bui BV, et al. Mr-eye: high-resolution mri of the human eye and orbit at ultrahigh field (7t). 2021; 29:103-16.

[22] Uhl A, Wild P. Multi-stage visible wavelength and near infrared iris segmentation framework. International Conference Image Analysis and Recognition: Springer; 2012. p. 1-10.

[23] Loskutova E, Butler JS, Hernandez Martinez G, Flitcroft I, Loughman JJCER. Macular Pigment Optical Density Fluctuation as a Function of Pupillary Mydriasis: Methodological Considerations for Dual-Wavelength Autofluorescence. 2021; 46:532-8.

[24] Jan F, Usman IJO. Iris segmentation for visible wavelength and near infrared eye images. 2014; 125:4274-82.

[25] Chen Y, Davoine F. Simultaneous Tracking of Rigid Head Motion and Non-rigid Facial Animation by Analyzing Local Features Statistically. BMVC2006.