are not limited in the embodiments of the present application.

It can be understood that before the training the neural network according to the first gaze direction, the first detected gaze direction, the second detected gaze direction, and a second gaze direction, the method further includes:

normalizing the first gaze direction, the first detected gaze direction, the second detected gaze direction, and the second gaze direction respectively.

The training the neural network according to the first gaze direction, the first detected gaze direction, the second detected gaze direction, and the second gaze direction includes:

training the neural network according to the normalized first gaze direction, the normalized second gaze direction, a normalized first detected gaze direction, and a normalized second detected gaze direction.

In the embodiments of the present application, by normalizing the first gaze direction, the first detected gaze direction, the second gaze direction, and the second detected gaze direction of the vector, the loss function can be simplified, the computing accuracy of the loss function can be improved, and the computing complexity of loss function can be avoided. The loss function may be a loss of the first gaze direction and the first detected gaze direction, may also be a loss of a first offset vector and a second offset vector, and may also be a loss of the second gaze direction and the second detected gaze direction.

That is, the network parameters of the neural network may be adjusted according to a third loss of the normalized first gaze direction and the normalized first detected gaze direction, and the fourth loss of the normalized second gaze direction and the normalized second detected gaze direction.

Assuming that the first gaze direction is (x3, y3, z3) and the first detected gaze direction is (x4, y4, z4), the mode of normalization may be as shown in equations (3) and (4):

$$normalize\ ground\ truth = \left(\frac{(x3)}{\sqrt[2]{(x3)^2+(y3)^2+(z3)^2}}, \frac{(y3)}{\sqrt[2]{(x3)^2+(y3)^2+(z3)^2}}, \frac{(z3)}{\sqrt[2]{(x3)^2+(y3)^2+(z3)^2}}\right) \quad (3)$$

where $normalize\ ground\ truth$ is the normalized first gaze direction.

$$normalize\ prediction\ gaze = \left(\frac{(x4)}{\sqrt[2]{(x4)^2+(y4)^2+(z4)^2}}, \frac{(y4)}{\sqrt[2]{(x4)^2+(y4)^2+(z4)^2}}, \frac{(z4)}{\sqrt[2]{(x4)^2+(y4)^2+(z4)^2}}\right) \quad (4)$$

where $\mathrm{normalize\ prediction\ gaze}$ is the normalized first detected gaze direction.

The third loss may be calculated as shown in equation (5):

$$loss = \| normalize\ ground\ truth - normalize\ prediction\ gaze \| \quad (5)$$

where "loss" is the third loss.

It can be understood that the above expressions by the various letters or parameters are merely examples, and should not be construed as limiting the embodiments of the present application.

By normalizing the first gaze direction, the first detected gaze direction, the second gaze direction, and the second detected gaze direction, the influence of the magnitude in each gaze direction can be eliminated, so that only the gaze direction is focused on, and thus, the accuracy of training the neural network can be further improved.

Implementation II

The training the neural network according to the first gaze direction, the first detected gaze direction, the second detected gaze direction, and a second gaze direction includes:

determining a first loss of the first gaze direction and the first detected gaze direction;

determining a second loss of a first offset vector and a second offset vector, where the first offset vector is an offset vector between the first gaze direction and the second gaze direction, and the second offset vector is an offset vector between the first detected gaze direction and the second detected gaze direction; and

adjusting network parameters of the neural network according to the first loss and the

second loss.

In the embodiments of the present application, the neural network is trained not only according to the loss of the first gaze direction and the first detected gaze direction, but also according to the loss of the first offset vector and the second offset vector. By enhancing the input image data, not only the problem of gaze jitter during the gaze tracking process can be effectively prevented, but also the stability and accuracy of training the neural network can be improved.

Assuming that the first gaze direction is (x3, y3, z3), the first detected gaze direction is (x4, y4, z4), the second detected gaze direction is (x5, y5, z5), and the second gaze direction is (x6, y6, z6), the first offset vector is (x3-x6, y3-y6, z3-z6), and the second offset vector is (x4-x5, y4-y5, z4-z5).

It can be understood that before the training the neural network according to the first gaze direction, the first detected gaze direction, the second detected gaze direction, and a second gaze direction, the method further includes:

normalizing the first gaze direction, the first detected gaze direction, the second detected gaze direction, and the second gaze direction respectively.

The training the neural network according to the first gaze direction, the first detected gaze direction, the second detected gaze direction, and the second gaze direction includes:

training the neural network according to the normalized first gaze direction, the normalized second gaze direction, a normalized first detected gaze direction, and a normalized second detected gaze direction.

In the embodiments of the present application, by normalizing the first gaze direction, the first detected gaze direction, the second gaze direction, and the second detected gaze direction of the vector, the loss function can be simplified, the computing accuracy of the loss function can be improved, and the computing complexity of loss function can be avoided. The loss function may be a loss of the first gaze direction and the first detected gaze direction, may also be a loss of the first offset vector and the second offset vector, and may also be a loss of the second gaze direction and the second detected gaze direction.

That is, the network parameters of the neural network may be adjusted according to the first loss of the normalized first gaze direction and the normalized first detected gaze direction and the second loss of a normalized first offset vector and a normalized second offset vector. The normalized first offset vector is an offset vector between the normalized first gaze direction and the normalized second gaze direction, and the normalized second offset vector is an offset vector between the normalized first detected gaze direction and the normalized second detected gaze direction.

For the specific implementation of normalization, reference may be made to the implementation shown in implementation I, and details are not described herein again.

By normalizing the first gaze direction, the first detected gaze direction, the second gaze direction, and the second detected gaze direction, the influence of the magnitude in each gaze direction can be eliminated, so that only the gaze direction is focused on, and thus, the accuracy of training the neural network can be further improved.

In a possible implementation, before the normalizing the first gaze direction, the first detected gaze direction, the second detected gaze direction, and the second gaze direction respectively, the method further includes:

determining eye positions in the first image; and

rotating the first image according to the eye positions so that the two eye positions in the first image are the same on a horizontal axis.

It can be understood that, in the embodiments of the present application, determining the eye positions in the first image may specifically include determining a left eye position and a right eye position in the first image respectively, capturing an image corresponding to the left eye position and an image corresponding to the left eye position, and then respectively rotating the image corresponding to the right eye position and the image corresponding to the left eye position to make the two eye positions the same on the horizontal axis.

It can be understood that, in order to further improve the smoothness of the gaze direction, the detecting, by the neural network, the gaze direction in the first image to obtain the first directed gaze direction includes:

respectively detecting gaze directions in N adjacent image frames through the neural network if the first image is a video image, where N is an integer greater than or equal to 1; and

determining the gaze direction in the N-th image frame as the first detected gaze direction according to the gaze directions in the N adjacent image frames.

The specific value of N is not limited in the embodiments of the present application. The N adjacent image frames may be N image frames before the N-th image frame (including the N-th frame), or may be N image frames after the N-th image frame, or may be N image frames before and after the N-th image frame, and the like, which is not limited in the embodiments of the present application.

In the embodiments of the present application, in gaze tracking in a video, there may still be jitter occurring to the gaze direction output by the neural network. Therefore, by determining the gaze direction in the N-th image frame according to the gaze directions in N image frames, and performing a smoothing process based on the gaze direction detected by the neural network, the stability of the gaze direction detected by the neural network can be improved.

Optionally, the gaze direction in the N-th image frame may be determined according to an average sum of the gaze directions in N adjacent image frames, so as to smooth the gaze direction, making the obtained first detected gaze direction more stable.

It can be understood that the method for determining the second detected gaze direction may also be obtained by the method described above, and details are not described herein again.

In the embodiments of the present application, by obtaining the first detected gaze direction and the second detected gaze direction and training the neural network according to the first gaze direction, the first detected gaze direction, and the second detected gaze direction, on the one hand, the accuracy of training the neural network can be improved, and on the other hand, the neural network can be trained efficiently.

It can be understood that after the neural network is obtained by neural network training by the above method, the neural network training apparatus may directly apply the neural network to predict the gaze direction, or the neural network training apparatus may also transmit the trained neural network to other apparatuses so that other apparatuses utilize the trained neural network to predict the gaze direction. Which apparatuses the neural network training apparatus specifically transmits the neural network to are not limited in the embodiments of the present application.

Referring to FIG. 4a, FIG. 4a shows a schematic flowchart of a method for determining a first gaze direction provided in embodiments of the present application. As shown in FIG. 4a, the method for determining a first gaze direction includes the following steps.

At step 401, the first camera is determined from a camera array, and coordinates of the pupil in a first coordinate system are determined, where the first coordinate system is a coordinate system corresponding to the first camera.

In the embodiments of the present application, the coordinates of the pupil in the first coordinate system may be determined according to the focal length and principal point position of the first camera.

Optionally, the determining the coordinates of the pupil in the first coordinate system includes:

determining coordinates of the pupil in the first image; and

determining the coordinates of the pupil in the first coordinate system according to the coordinates of the pupil in the first image and the focal length and principal point position of the first camera.

In the embodiments of the present application, for a captured 2D picture of the eyes, i.e., the first image, points around the edge of the pupil of an eye may be extracted directly by a network model for detecting edge points of the pupil, and the coordinates of the pupil position, such as (m, n), are then calculated according to the points around the edge of the pupil. The calculated

Optionally, the detection unit 602 is specifically configured to detect gaze directions in the first image and a second image through the neural network to obtain the first detected gaze direction and a second detected gaze direction, respectively, where the second image is obtained by adding noise to the first image; and

the training unit 603 is specifically configured to train the neural network according to the first gaze direction, the first detected gaze direction, the second detected gaze direction, and a second gaze direction, where the second gaze direction is a gaze direction obtained by adding noise to the first gaze direction.

Optionally, the training unit 603 is specifically configured to adjust network parameters of the neural network according to a third loss of the first gaze direction and the first detected gaze direction and a fourth loss of the second gaze direction and the second detected gaze direction.

Optionally, as shown in FIG. 7, the training unit 603 includes:

a first determination sub-unit 6031, configured to determine a first loss of the first gaze direction and the first detected gaze direction;

a second determination sub-unit 6032, configured to determine a second loss of a first offset vector and a second offset vector, where the first offset vector is an offset vector between the first gaze direction and the second gaze direction, and the second offset vector is an offset vector between the first detected gaze direction and the second detected gaze direction; and

an adjustment sub-unit 6033, configured to adjust the network parameters of the neural network according to the first loss and the second loss.

Optionally, as shown in FIG. 8, the apparatus further includes:

a normalization unit 604, configured to normalize the first gaze direction, the first detected gaze direction, the second detected gaze direction, and the second gaze direction respectively; and

the training unit 603, specifically configured to train the neural network according to the normalized first gaze direction, the normalized second gaze direction, a normalized first detected gaze direction, and a normalized second detected gaze direction.

Optionally, as shown in FIG. 8, the apparatus further includes:

a second determination unit 605, configured to determine eye positions in the first image; and

a rotation unit 606, configured to rotate the first image according to the eye positions so that the two eye positions in the first image are the same on a horizontal axis.

Optionally, as shown in FIG. 9, the detection unit 602 includes:

a detection sub-unit 6021, configured to respectively detect gaze directions in N adjacent image frames through the neural network if the first image is a video image, where N is an integer greater than or equal to 1; and

a third determination sub-unit 6022, configured to determine the gaze direction in the N-th image frame as the first detected gaze direction according to the gaze directions in the N adjacent image frames.

Optionally, the third determination sub-unit 6022 is specifically configured to determine the gaze direction in the N-th image frame as the first detected gaze direction according to the average sum of the gaze directions in the N adjacent image frames.

Optionally, the first determination unit 601 is specifically configured to: determine the first camera from a camera array, and determine coordinates of the pupil in a first coordinate system, where the first coordinate system is a coordinate system corresponding to the first camera; determine coordinates of the pupil in a second coordinate system according to a second camera in the camera array, where the second coordinate system is a coordinate system corresponding to the second camera; and determine the first gaze direction according to the coordinates of the pupil in the first coordinate system and the coordinates of the pupil in the second coordinate system.

Optionally, the first determination unit 601 is specifically configured to: determine coordinates of the pupil in the first image; and determine the coordinates of the pupil in the first coordinate system according to the coordinates of the pupil in the first image and the focal length and principal point position of the first camera.

CLAIMS

1. A neural network training method, comprising:

determining a first gaze direction according to a first camera and a pupil in a first image, wherein the first camera is a camera that captures the first image, and at least an eye image is comprised in the first image;

detecting a gaze direction in the first image through a neural network to obtain a first detected gaze direction; and

training the neural network according to the first gaze direction and the first detected gaze direction.

2. The method according to claim 1, wherein the detecting a gaze direction in the first image through a neural network to obtain a first detected gaze direction comprises:

detecting gaze directions in the first image and a second image respectively through the neural network to obtain the first detected gaze direction and a second detected gaze direction respectively, wherein the second image is obtained by adding noise to the first image; and

the training the neural network according to the first gaze direction and the first detected gaze direction comprises:

training the neural network according to the first gaze direction, the first detected gaze direction, the second detected gaze direction, and a second gaze direction, wherein the second gaze direction is a gaze direction obtained by adding noise to the first gaze direction.

3. The method according to claim 2, wherein the training the neural network according to the first gaze direction, the first detected gaze direction, the second detected gaze direction, and a second gaze direction comprises:

determining a first loss of the first gaze direction and the first detected gaze direction;

determining a second loss of a first offset vector and a second offset vector, wherein the first offset vector is an offset vector between the first gaze direction and the second gaze direction, and the second offset vector is an offset vector between the first detected gaze direction and the second detected gaze direction; and

adjusting network parameters of the neural network according to the first loss and the second loss.

4. The method according to claim 2, wherein the training the neural network according to the first gaze direction, the first detected gaze direction, the second detected gaze direction, and a second gaze direction comprises:

adjusting the network parameters of the neural network according to a third loss of the first gaze direction and the first detected gaze direction and a fourth loss of the second gaze direction and the second detected gaze direction.

5. The method according to claim 3 or 4, wherein before the training the neural network according to the first gaze direction, the first detected gaze direction, the second detected gaze direction, and a second gaze direction, the method further comprises:

normalizing the first gaze direction, the first detected gaze direction, the second detected gaze direction, and the second gaze direction respectively; and

the training the neural network according to the first gaze direction, the first detected gaze direction, the second detected gaze direction, and the second gaze direction comprises:

training the neural network according to the normalized first gaze direction, the normalized second gaze direction, a normalized first detected gaze direction, and a normalized second detected gaze direction.

6. The method according to claim 5, wherein before the normalizing the first gaze direction, the first detected gaze direction, the second detected gaze direction, and the second gaze direction respectively, the method further comprises:

determining eye positions in the first image; and

rotating the first image according to the eye positions so that the two eye positions in the first image are the same on a horizontal axis.

7. The method according to any one of claims 1 to 6, wherein the detecting a gaze direction in the first image through a neural network to obtain a first detected gaze direction comprises:

respectively detecting gaze directions in N adjacent image frames through the neural network if the first image is a video image, wherein N is an integer greater than or equal to 1; and

determining a gaze direction in a N-th image frame as the first detected gaze direction according to the gaze directions in the N adjacent image frames.

8. The method according to claim 7, wherein the determining a gaze direction in a N-th image frame as the first detected gaze direction according to the gaze directions in the N adjacent image

frames comprises:

determining the gaze direction in the N-th image frame as the first detected gaze direction according to an average sum of the gaze directions in the N adjacent image frames.

9. The method according to any one of claims 1 to 8, wherein the determining a first gaze direction according to a first camera and a pupil in the first image comprises:

determining the first camera from a camera array, and determining coordinates of the pupil in a first coordinate system, wherein the first coordinate system is a coordinate system corresponding to the first camera;

determining coordinates of the pupil in a second coordinate system according to a second camera in the camera array, wherein the second coordinate system is a coordinate system corresponding to the second camera; and

determining the first gaze direction according to the coordinates of the pupil in the first coordinate system and the coordinates of the pupil in the second coordinate system.

10. The method according to claim 9, wherein the determining coordinates of the pupil in a first coordinate system comprises:

determining coordinates of the pupil in the first image; and

determining the coordinates of the pupil in the first coordinate system according to the coordinates of the pupil in the first image and a focal length and principal point position of the first camera.

11. The method according to claim 9 or 10, wherein the determining coordinates of the pupil in a second coordinate system according to a second camera in the camera array comprises:

determining a relationship between the first coordinate system and the second coordinate system according to the first coordinate system and a focal length and principal point position of each camera in the camera array; and

determining the coordinates of the pupil in the second coordinate system according to the relationship between the second coordinate system and the first coordinate system.

12. A gaze tracking method, comprising:

performing face detection on a third image comprised in video stream data;

performing key point positioning on a detected face region in the third image to determine an eye region in the detected face region;

capturing an image of the eye region in the third image; and

inputting the image of the eye region to a pre-trained neural network and outputting a gaze direction in the image of the eye region.

13. The method according to claim 12, wherein after the inputting the image of the eye region to a pre-trained neural network and outputting a gaze direction in the image of the eye region, the method further comprises:

determining a gaze direction in the third image according to the gaze direction in the image of the eye region and a gaze direction in at least one adjacent image frame of the third image.

14. The method according to claim 12 or 13, wherein the performing face detection on a third image comprised in video stream data comprises:

performing face detection on the third image comprised in the video stream data when a trigger instruction is received; or

performing face detection on the third image comprised in the video stream data during vehicle running; or

performing face detection on the third image comprised in the video stream data if a running speed of the vehicle reaches a reference speed.

15. The method according to claim 14, wherein

the video stream data is a video stream of a driving region of the vehicle captured by a vehicle-mounted camera, and the gaze direction in the image of the eye region is a gaze direction of a driver in the driving region of the vehicle; or, the video stream data is a video stream of a non-driving region of the vehicle captured by a vehicle-mounted camera, and the gaze direction in the image of the eye region is a gaze direction of a person in the non-driving region of the vehicle.

16. The method according to claim 15, wherein after the outputting a gaze direction in the image of the eye region, the method further comprises:

determining a region of interest of the driver according to the gaze direction in the image of the eye region; determining a driving behavior of the driver according to the region of interest of the driver, wherein the driving behavior comprises whether the driver is distracted from driving; or outputting, according to the gaze direction, control information for the vehicle or a vehicle-mounted device provided on the vehicle.

17. The method according to claim 16, further comprising:

storing one or more of the image of the eye region and a predetermined number of image frames before and after the image of the eye region if the driver is distracted from driving; or

transmitting one or more of the image of the eye region and the predetermined number of image frames before and after the image of the eye region to the terminal connected to the vehicle if the driver is distracted from driving.

18. The method according to any one of claims 12 to 17, wherein before the inputting the image of the eye region to a pre-trained neural network, the method further comprises: training the neural network by the method according to any one of claims 1 to 11.

19. An electronic device, comprising a processor and a memory which are connected to each other by a line, wherein the memory is used for storing program instructions that, when executed by the processor, cause the processor to execute the method according to any one of claims 1 to 18.

20. A computer-readable storage medium, which stores a computer program therein, wherein the computer program comprises program instructions that, when executed by a processor, cause the processor to execute the method according to any one of claims 1 to 11, and/or cause the processor to execute the method according to any one of claims 12 to 18.