CLAIMS

1. A driver attention monitoring method, comprising:

capturing, by multiple cameras arranged on a vehicle, a video of a driving area of the vehicle;

determining, according to each of multiple frames of face images of a driver in the driving area comprised in the video, a type of a gazing area of the driver in the frame of face image, wherein the gazing area of the frame of face image is one of multiple types of defined gazing areas obtained by dividing a space area of the vehicle in advance; and

determining an attention monitoring result of the driver according to a type distribution of gazing areas of the frames of face images comprised within at least one sliding time window in the video,

wherein determining, according to each of the multiple frames of face images of the driver in the driving area comprised in the video, the type of the gazing area of the driver in the frame of face image comprises:

inputting each of the multiple frames of face images into a neural network and outputting the type of the gazing area of the driver in each frame of face image from the neural network, wherein the neural network is pre-trained by using a face image set comprising gazing area type labeling information, or the neural network is pre-trained by using a face image set comprising gazing area type labeling information and an eye image cropped based on each face image in the face image set; and the gazing area type labeling information comprises one of the multiple types of defined gazing areas.

2. The method according to claim 1, wherein the multiple types of defined gazing areas obtained by dividing the space area of the vehicle in advance comprise two or more of: a left front windshield area, a right front windshield area, a dashboard area, an in-vehicle rearview mirror area, a center console area, a left rearview mirror area, a right rearview mirror area, a visor area, a shift lever area, an area below a steering wheel, a front passenger seat area, or a glove compartment area in front of a front passenger seat.

3. The method according to claim 1 or 2, wherein determining the attention monitoring result of

the driver according to the type distribution of gazing areas of the frames of face images comprised within at least one sliding time window in the video comprises:

determining, according to the type distribution of gazing areas of the frames of face images comprised within at least one sliding time window in the video, an accumulated gazing duration of each type of gazing area within the at least one sliding time window; and

determining the attention monitoring result of the driver according to a comparison result between the accumulated gazing duration of each type of gazing area within the at least one sliding time window and a predetermined time threshold, wherein the attention monitoring result comprises whether the driver is in distracted driving and/or a distracted driving level.

4. The method according to claim 3, wherein the time threshold comprises multiple time thresholds corresponding to respective types of defined gazing areas, wherein the time thresholds corresponding to at least two different types of defined gazing areas in the multiple types of defined gazing areas are different; and

determining the attention monitoring result of the driver according to the comparison result between the accumulated gazing duration of each type of gazing area within the at least one sliding time window and the predetermined time threshold comprises: determining the attention monitoring result of the driver according to a comparison result between the accumulated gazing duration of each type of gazing area within the at least one sliding time window and a respective time threshold of the type of defined gazing area.

5. The method according to any one of claims 1 to 4, wherein determining, according to each of the multiple frames of face images of the driver in the driving area comprised in the video, the type of the gazing area of the driver in the frame of face image comprises:

performing line of sight and/or head pose detection on the multiple frames of face images of the driver in the driving area comprised in the video; and

determining the type of the gazing area of the driver in each frame of face image according to the line of sight and/or head pose detection result for the frame of face image.

6. The method according to claim 1, wherein a training method for the neural network comprises:

obtaining a face image comprising the gazing area type labeling information in the face image set;

cropping the eye image of at least one eye in the face image, wherein the at least one eye comprises the left eye and/or the right eye;

respectively extracting a first feature of the face image and a second feature of the eye image of the at least one eye;

fusing the first feature and the second feature to obtain a third feature;

determining a gazing area type detection result of the face image according to the third feature; and

adjusting a network parameter of the neural network according to a difference between the gazing area type detection result and the gazing area type labeling information.

7. The method according to any one of claims 1 to 6, further comprising:

in the case that the attention monitoring result of the driver is the distracted driving, giving the driver a prompt for distracted driving, wherein the prompt for distracted driving comprises at least one of: a text prompt, a voice prompt, a smell prompt, or a low-current stimulation prompt; or

in the case that the attention monitoring result of the driver is the distracted driving, determining a distracted driving level of the driver according to a preset mapping relationship between the distracted driving level and the attention monitoring result and to the attention monitoring result of the driver; and determining, according to a preset mapping relationship between the distracted driving level and a prompt for distracted driving and to the distracted driving level of the driver, a prompt from among prompts for distracted driving to give the driver the prompt for distracted driving.

8. The method according to any one of claims 1 to 7, wherein the preset mapping relationship between the distracted driving level and the attention monitoring result comprises: in the case

that the monitoring results within multiple consecutive sliding time windows are all the distracted driving, the distracted driving level is positively correlated to a number of the sliding time windows.

9. The method according to any one of claims 1 to 8, wherein capturing, by the multiple cameras arranged on the vehicle, the video of the driving area of the vehicle comprises: respectively capturing, by multiple cameras respectively arranged on multiple areas on the vehicle, videos of the driving area from different angles; and

determining, according to each of the multiple frames of face images of the driver in the driving area comprised in the video, the type of the gazing area of the driver in the frame of face image comprises: respectively determining, according to an image quality evaluation index, an image quality score of each frame of face image in the multiple frames of face images of the driver in the driving area comprised in each of multiple captured videos; respectively determining a face image having a highest image quality score among the frames of face images aligned in time in the multiple videos; and respectively determining the type of the gazing area of the driver in each face image having the highest image quality score.

10. The method according to claim 9, wherein the image quality evaluation index comprises at least one of: whether an image comprises an eye image, a definition of an eye area in an image, a shielding status of an eye area in an image, or an eye opening/closing status of an eye area in an image.

11. The method according to any one of claims 1 to 8, wherein capturing, by the multiple cameras arranged on the vehicle, the video of the driving area of the vehicle comprises: respectively capturing, by multiple cameras respectively arranged on multiple areas on the vehicle, videos of the driving area from different angles; and

determining, according to each of the multiple frames of face images of the driver in the driving area comprised in the video, the type of the gazing area of the driver in the frame of face image comprises: respectively detecting, for the multiple frames of face images of the driver in the driving area comprised in each of multiple captured videos, gazing area types of the driver in

the frames of face images aligned in time; and determining a majority of obtained gazing area types as the gazing area type of the face images at that time.

12. The method according to any one of claims 1 to 11, further comprising:

sending the attention monitoring result of the driver to a server or a terminal communicationally connected to the vehicle; and/or

performing statistical analysis on the attention monitoring result of the driver.

13. The method according to claim 12, after sending the attention monitoring result of the driver to the server or the terminal communicationally connected to the vehicle, further comprising:

in the case of receiving a control instruction sent by the server or the terminal, controlling the vehicle according to the control instruction.

14. A driver attention monitoring apparatus, comprising:

a first control unit, configured to capture, by multiple cameras arranged on a vehicle, a video of a driving area of the vehicle;

a first determination unit, configured to determine, according to each of multiple frames of face images of a driver in the driving area comprised in the video, a type of a gazing area of the driver in the frame of face image, wherein the gazing area of the frame of face image is one of multiple types of defined gazing areas obtained by dividing a space area of the vehicle in advance; and

a second determination unit, configured to determine an attention monitoring result of the driver according to a type distribution of gazing areas of the frames of face images comprised within at least one sliding time window in the video,

wherein the first determination unit further comprises:

a processing sub-unit, configured to input each of the multiple frames of face images into a neural network and output the type of the gazing area of the driver in each frame of face image from the neural network, wherein the neural network is pre-trained by using a face image set comprising gazing area type labeling information, or the neural network is pre-trained by using a face image set comprising gazing area type labeling information and an eye image cropped based

on each face image in the face image set; and the gazing area type labeling information comprises one of the multiple types of defined gazing areas.

15. The apparatus according to claim 14, wherein the multiple types of defined gazing areas obtained by dividing the space area of the vehicle in advance comprise two or more of: a left front windshield area, a right front windshield area, a dashboard area, an in-vehicle rearview mirror area, a center console area, a left rearview mirror area, a right rearview mirror area, a visor area, a shift lever area, an area below a steering wheel, a front passenger seat area, or a glove compartment area in front of a front passenger seat.

16. The apparatus according to claim 14 or 15, wherein the second determination unit comprises:

a first determination sub-unit, configured to determine, according to the type distribution of gazing areas of the frames of face images comprised within at least one sliding time window in the video, an accumulated gazing duration of each type of gazing area within the at least one sliding time window; and

a second determination sub-unit, configured to determine the attention monitoring result of the driver according to a comparison result between the accumulated gazing duration of each type of gazing area within the at least one sliding time window and a predetermined time threshold, the attention monitoring result comprising whether the driver is in distracted driving and/or a distracted driving level.

17. An electronic device, comprising a memory and a processor, wherein the memory has a computer executable instruction stored thereon, and when the processor runs the computer executable instruction on the memory, the following operations are implemented:

capturing, by multiple cameras arranged on a vehicle, a video of a driving area of the vehicle;

determining, according to each of multiple frames of face images of a driver in the driving area comprised in the video, a type of a gazing area of the driver in the frame of face image, wherein the gazing area of the frame of face image is one of multiple types of defined gazing areas obtained by dividing a space area of the vehicle in advance; and

determining an attention monitoring result of the driver according to a type distribution of gazing areas of the frames of face images comprised within at least one sliding time window in the video,

wherein determining, according to each of the multiple frames of face images of the driver in the driving area comprised in the video, the type of the gazing area of the driver in the frame of face image comprises:

inputting each of the multiple frames of face images into a neural network and outputting the type of the gazing area of the driver in each frame of face image from the neural network, wherein the neural network is pre-trained by using a face image set comprising gazing area type labeling information, or the neural network is pre-trained by using a face image set comprising gazing area type labeling information and an eye image cropped based on each face image in the face image set; and the gazing area type labeling information comprises one of the multiple types of defined gazing areas.

18. A computer-readable storage medium, wherein the computer-readable storage medium has a computer program stored thereon, and when the computer program is executed by a processor, the method according to any one of claims 1 to 13 is implemented.

19. A computer program, comprising a computer-readable code, wherein when the computer-readable code runs in a device, a processor in the device executes instructions for implementing operations of the method according to any one of claims 1 to 14.