Claims

1. A driving state monitoring method, comprising:

performing driver state detection on a driver image; and

performing at least one of: outputting a driving state monitoring result of a driver or performing intelligent driving control based on a result of the driver state detection;

wherein the driver state detection comprises at least one of: driver fatigue state detection, driver distraction state detection, or driver scheduled distraction action detection,

wherein in response to the scheduled distraction action being the smoking action, the performing driver state detection on a driver image comprises:

performing face detection on the driver image via a fourth neural network to obtain a face detection frame, and extracting feature information of the face detection frame; and

determining whether the smoking action occurs via the fourth neural network based on the feature information of the face detection frame.

2. The method according to claim 1, wherein the performing driver fatigue state detection on a driver image comprises:

detecting at least part of a face region of the driver in the driver image to obtain state information of the at least part of the face region, the state information of the at least part of the face region comprising at least one of: eye open/closed state information or mouth open/closed state information;

obtaining a parameter value of an index for representing a driver fatigue state based on the state information of the at least part of the face region within a period of time; and

determining a result of the driver fatigue state detection based on the parameter value of the index for representing the driver fatigue state;

wherein the index for representing the driver fatigue state comprises at least one of: an eye closure degree or a yawning degree;

wherein the parameter value of the eye closure degree comprises at least one of: a number of eye closures, an eye closure frequency, eye closure duration, eye closure amplitude, a number of eye semi-closures, or an eye semi-closure frequency; and/or

the parameter value of the yawning degree comprises at least one of: a yawning state, a number of yawns, yawning duration, or a yawning frequency.

3. The method according to any one of claims 1-2, wherein the performing driver distraction state detection on a driver image comprises:

performing face orientation and/or gaze direction detection on the driver in the driver image to obtain face orientation information and/or gaze direction information;

determining a parameter value of an index for representing a driver distraction state based on the face orientation information and/or the gaze direction information within a period of time, the index for representing the driver distraction state comprises at least one of: a face orientation deviation degree or a gaze deviation degree; and

determining a result of the driver distraction state detection based on the parameter value of the index for representing the driver distraction state;

wherein the parameter value of the face orientation deviation degree comprises at least one of: a number of head turns, head turning duration, or a head turning frequency; and/or

the parameter value of the gaze deviation degree comprises at least one of: a gaze direction deviation angle, gaze direction deviation duration, or a gaze direction deviation frequency.

4. The method according to claim 3, wherein the performing at least one of face orientation or gaze direction detection on the driver image comprises:

detecting face key points of the driver image; and

performing at least one of face orientation or gaze direction detection based on the face key points.

5. The method according to claim 4, wherein the performing face orientation detection based on the face key points to obtain the face orientation information comprises:

obtaining feature information of head pose based on the face key points; and

determining the face orientation information based on the feature information of the head pose.

6. The method according to claim 5, wherein the obtaining feature information of head pose based on the face key points, and determining the face orientation information based on the feature information of the head pose comprise:

extracting the feature information of the head pose via a first neural network based on the face key points; and

performing face orientation estimation via a second neural network based on the feature

information of the head pose to obtain the face orientation information.

7. The method according to any one of claims 4-6, wherein the performing gaze direction detection based on the face key points to obtain the gaze direction information comprises:

determining a pupil edge location based on an eye image positioned by an eye key point among the face key points, and computing a pupil center location based on the pupil edge location; and

computing the gaze direction information based on the pupil center location and an eye center location.

8. The method according to claim 7, wherein the determining a pupil edge location based on an eye image positioned by an eye key point among the face key points comprises:

detecting, based on a third neural network, a pupil edge location of an eye region image among images divided based on the face key points, and obtaining the pupil edge location based on information outputted by the third neural network.

9. The method according to any one of claims 1-8,

wherein the performing scheduled distraction action detection on a driver image comprises:

performing target object detection corresponding to the scheduled distraction action on the driver image to obtain a detection frame for a target object; and

determining whether the scheduled distraction action occurs based on the detection frame for the target object.

10. The method according to any one of claim 1-9, further comprising:

if the distraction action occurs, obtaining a determination result indicating whether the scheduled distraction action occurs within a period of time to obtain a parameter value of an index for representing a distraction degree; and

determining the result of the driver scheduled distraction action detection based on the parameter value of the index for representing the distraction degree;

wherein the parameter value of the distraction degree comprises at least one of: a number of occurrences of the scheduled distraction action, duration of the scheduled distraction action, or a frequency of the scheduled distraction action.

11. The method according to any one of claims 1-10, wherein the scheduled distraction action comprises at least one of : an eating action, a drinking action, a phone call action, or an

entertainment action, and the method further comprises:

performing preset target object detection corresponding to the scheduled distraction action on the driver image via a fifth neural network to obtain a detection frame for a preset target object; the preset target object comprising: hands, mouth, eyes, or a target item; and the target item comprising at least one of following types: containers, foods, or electronic devices; and

determining a detection result of the distraction action based on the detection frame for the preset target object; the detection result of the distraction action comprising at least one of: no eating action occurs, no drinking action occurs, no phone call action occurs, no entertainment action occurs, the eating action occurs, the drinking action occurs, the phone call action occurs, or the entertainment action occurs.

12. The method according to claim 11, wherein the determining a detection result of the distraction action based on the detection frame for the preset target object comprises:

determining the detection result of the scheduled distraction action based on whether a detection frame for the hands, a detection frame for the mouth, a detection frame for the eyes, or a detection frame for the target item are detected, whether the detection frame for the hands overlaps the detection frame for the target item, a type of the target item, and whether a distance between the detection frame for the target item and the detection frame for the mouth or the detection frame for the eyes satisfies preset conditions.

13. The method according to claim 12, wherein the determining the detection result of the distraction action based on whether the detection frame for the hands overlaps the detection frame for the target item, and whether a location relationship between the detection frame for the target item and the detection frame for the mouth or the detection frame for the eyes satisfies preset conditions comprises:

if the detection frame for the hands overlaps the detection frame for the target item, the type of the target item is a container or food, and the detection frame for the target item overlaps the detection frame for the mouth, determining that the eating action or the drinking action occurs; and/or

if the detection frame for the hands overlaps the detection frame for the target item, the type of the target item is an electronic device, and the minimum distance between the detection frame for the target item and the detection frame for the mouth is less than a first preset distance, or the minimum distance between the detection frame for the target item and the detection frame for the eyes is less than a second preset distance, determining that the entertainment action or the phone

call action occurs.

14. The method according to claim 12 or 13, further comprising:

if the detection frame for the hands, the detection frame for the mouth, and the detection frame for any one target item are not detected simultaneously, and the detection frame for the hands, the detection frame for the eyes, and the detection frame for any one target item are not detected simultaneously, determining that the detection result of the distraction action is that no eating action, drinking action, phone call action and entertainment action is detected; and/or

if the detection frame for the hands does not overlap the detection frame for the target item, determining that the detection result of the distraction action is that no eating action, drinking action, phone call action, and entertainment action is detected; and/or

if the type of the target item is a container or food and the detection frame for the target item does not overlaps the detection frame for the mouth, and/or the type of the target item is an electronic device and the minimum distance between the detection frame for the target item and the detection frame for the mouth is not less than the first preset distance, or the minimum distance between the detection frame for the target item and the detection frame for the eyes is not less than the second preset distance, determining that the detection result of the distraction action is that no eating action, drinking action, phone call action, and entertainment action is detected.

15. The method according to any one of claims 1-14, further comprising:

if the result of the driver scheduled distraction action detection is that a scheduled distraction action is detected, prompting the detected distraction action;

wherein the method further comprises:

outputting distraction prompt information based on at least one of the result of the driver distraction state detection or the result of the driver scheduled distraction action detection.

16. The method according to any one of claims 1-15, wherein the outputting a driving state monitoring result of a driver based on the result of the driver state detection comprises:

determining a driving state level according to a preset condition that the result of the driver fatigue state detection, the result of the driver distraction state detection, and the result of the driver scheduled distraction action detection satisfy; and

using the determined driving state level as the driving state monitoring result.

17. The method according to any one of claims 1-16, further comprising:

performing a control operation corresponding to the driving state monitoring result;

wherein the performing a control operation corresponding to the driving state monitoring result comprises at least one of:

if the determined driving state monitoring result satisfies a predetermined prompting/warning condition, outputting prompting/warning information corresponding to the predetermined prompting/warning condition; and/or

if the determined driving state monitoring result satisfies a predetermined driving mode switching condition, switching a driving mode to an automatic driving mode.

18. The method according to any one of claims 1-17, further comprising:

performing facial recognition on the driver image; and

performing authentication control based on a result of the facial recognition.

19. The method according to claim 18, wherein the performing facial recognition on the driver image comprises:

performing face detection on the driver image via a sixth neural network, and performing feature extraction on the detected face to obtain a face feature;

performing face matching between the face feature and face feature templates in a database; and

if a face feature template matching the face feature exists in the database, outputting identity information corresponding to the face feature template matching the face feature.

20. The method according to claim 19, further comprising:

if no face feature template matching the face feature exists in the database, prompting the driver to register;

in response to receiving a registration request from the driver, performing face detection on the collected driver image via the sixth neural network, and performing feature extraction on the detected face to obtain a face feature; and

establishing user information of the driver in the database by using the face feature as the face feature template of the driver, the user information comprising the face feature template of the driver and the identity information inputted by the driver;

21. The method according to claim 19 or 20, further comprising:

storing the driving state monitoring result in the user information of the driver in the

database.

22. The method according to any one of claims 1-21, further comprising:

performing image collection using an infrared camera to obtain the driver image;

wherein the performing image collection using an infrared camera comprises:

performing image collection using the infrared camera deployed in at least one location within a vehicle;

wherein the at least one location comprises at least one of the following locations: a location above or near a dashboard, a location above or near a center console, an A-pillar or nearby location, or a rear-view mirror or nearby location.

23. The method according to any one of claims 1-22, further comprising:

performing driver gesture detection based on the driver image; and

generating a control instruction based on a result of the driver gesture detection;

24. The method according to claim 23, wherein the performing driver gesture detection based on the driver image comprises:

detecting a hand key point in a driver image of a current frame; and

using a static gesture determined based on the detected hand key point as the result of the driver gesture detection; or

wherein the performing driver gesture detection based on the driver image comprises:

detecting hand key points of a plurality of driver image frames in a driver video; and

using a dynamic gesture determined based on the detected hand key points of the plurality of driver image frames as the result of the driver gesture detection.

25. A driver monitoring system, comprising:

a display module, configured to display a driver image and a driving state monitoring result of a driver; and

a driver state detection module, configured to perform driver state detection on the driver image, and output the driving state monitoring result of the driver based on a result of the driver state detection;

the driver state detection comprising at least one of: driver fatigue state detection, driver distraction state detection, or driver scheduled distraction action detection,

wherein in response to the scheduled distraction action being the smoking action, the driver state detection module is further configured to perform face detection on the driver image via a fourth neural network to obtain a face detection frame, and extract feature information of the face detection frame, and determine whether the smoking action occurs via the fourth neural network based on the feature information of the face detection frame.

26. The system according to claim 25, wherein the display module comprises:

a first display region, configured to display the driver image and prompting/warning information corresponding to the driving state monitoring result; and

a second display region, configured to display a scheduled distraction action;

wherein the driver state detection module is further configured to perform facial recognition on the driver image;

the first display region is further configured to display a result of the facial recognition;

wherein the driver state detection module is further configured to perform driver gesture detection based on the driver image;

the display module further comprises: a third display region, configured to display a result of the gesture detection, the result of the gesture detection comprising a static gesture or a dynamic gesture.

27. An electronic device, comprising:

a memory, configured to store a computer program; and

a processor, configured to execute the compute program stored in the memory, and when the computer program is executed, the processor is configured to:

perform driver state detection on a driver image; and

perform at least one of: outputting a driving state monitoring result of a driver or performing intelligent driving control based on a result of the driver state detection, wherein the driver state detection comprises at least one of: driver fatigue state detection, driver distraction state detection, or driver scheduled distraction action detection,

wherein in response to the scheduled distraction action being the smoking action, the processor is further configured to perform face detection on the driver image via a fourth neural network to obtain a face detection frame, and extract feature information of the face detection frame, and determine whether the smoking action occurs via the fourth neural network based on

the feature information of the face detection frame.

28. A computer readable storage medium having a computer program stored thereon, wherein when the computer program is executed by a processor, the method according to any one of claims 1-24 is implemented.

29. A computer program, comprising computer instructions, wherein when the computer instructions run in a processor of a device, the method according to any one of claims 1-24 is implemented.