

CLAIMS

1. A method for recognizing a dynamic gesture, comprising:
 - positioning a dynamic gesture in a video stream to be detected to obtain a dynamic gesture box;
 - capturing an image block corresponding to the dynamic gesture box from each of multiple image frames of the video stream, wherein the captured image block is smaller than the image frame of the video stream, and is larger than or equal to the dynamic gesture box;
 - generating a detection sequence based on the captured image blocks; and
 - performing dynamic gesture recognition according to the detection sequence.
2. The method according to claim 1, wherein the positioning a dynamic gesture in a video stream to be detected to obtain a dynamic gesture box comprises:
 - positioning a static gesture in at least one image frame of the multiple image frames of the video stream to obtain a static gesture box of the at least one image frame; and
 - determining the dynamic gesture box according to the static gesture box of the at least one image frame.
3. The method according to claim 2, wherein the determining the dynamic gesture box according to the static gesture box of the at least one image frame comprises:
 - enlarging the static gesture box of the at least one image frame to obtain the dynamic gesture box.
4. The method according to claim 2 or 3, wherein the static gesture box of the at least one image frame of the multiple image frames of the video stream meets the following condition:

the static gesture box is located within the dynamic gesture box, or the static gesture box is as same as the dynamic gesture box.

5. The method according to any one of claims 1 to 4, wherein the performing dynamic gesture recognition according to the detection sequence comprises:

determining multiple inter-frame image differences in the detection sequence;
generating an image difference sequence based on the multiple inter-frame image differences; and

performing the dynamic gesture recognition according to the detection sequence and the image difference sequence.

6. The method according to claim 5, wherein one of the multiple inter-frame image differences is an image difference between two adjacent reference frames in the detection sequence.

7. The method according to claim 5 or 6, wherein the performing the dynamic gesture recognition according to the detection sequence and the image difference sequence comprises:

inputting the detection sequence into a first dynamic gesture recognition model to obtain a first dynamic gesture category prediction probability output by the first dynamic gesture recognition model;

inputting the image difference sequence into a second dynamic gesture recognition model to obtain a second dynamic gesture category prediction probability output by the second dynamic gesture recognition model; and

determining a dynamic gesture recognition result according to the first dynamic gesture category prediction probability and the second dynamic gesture category prediction probability.

8. The method according to claim 7, wherein the first dynamic gesture recognition model is a first neural network, the second dynamic gesture recognition model is a second neural network, and the first neural network and the second neural network have a same structure or different structures.

9. The method according to any one of claims 5 to 8, further comprising:

performing the capturing a plurality of times to obtain the detection sequence, generating the image difference sequence a plurality of times, and performing the dynamic gesture recognition a plurality of times according to the detection sequence and the image difference sequence; and

determining the dynamic gesture recognition result according to a probability of a dynamic gesture category, the probability being obtained by dynamic gesture recognition each time.

10. The method according to claim 7 or 8, wherein before the performing the dynamic gesture recognition according to the detection sequence and the image difference sequence, the method further comprises: establishing the first dynamic gesture recognition model by:

collecting one or more sample video streams involving different categories of dynamic gestures;

annotating dynamic gesture boxes of the different categories of dynamic gestures;

capturing image blocks corresponding to annotation information of the dynamic gesture boxes from multiple image frames of the sample video stream to form an image sequence; and

training the first dynamic gesture recognition model by using categories of the dynamic gestures as supervision data and using the image sequence as training data.

11. The method according to claim 10, wherein the training the first dynamic gesture recognition model by using categories of the dynamic gestures as supervision data and using the image sequence as training data comprises:

dividing the image sequence into at least one segment;

extracting a preset number of image frames from the at least one segment, and stacking the image frames to form image training data; and

training the first dynamic gesture recognition model by using the categories of the dynamic gestures as the supervision data and using the image training data.

12. The method according to any one of claims 7, 8, 10, and 11, wherein before the performing dynamic gesture recognition according to the detection sequence and the image difference sequence, the method further comprises: establishing the second dynamic gesture recognition model by the following means:

collecting one or more sample video streams involving different categories of dynamic gestures;

annotating dynamic gesture boxes of the different categories of dynamic gestures;

capturing image blocks corresponding to annotation information of the dynamic gesture boxes from multiple image frames of the one or more sample video streams to form the image sequence;

determining multiple inter-frame image differences in the image sequence;

generating an image difference sequence based on the multiple inter-frame image differences; and

training the second dynamic gesture recognition model by using categories of the dynamic gestures as supervision data and using the image difference sequence as training data.

13. The method according to claim 12, wherein the training the second dynamic gesture recognition model by using categories of the dynamic gestures as supervision data and using the image difference sequence as training data comprises:

dividing the image difference sequence into at least one segment;

extracting a preset number of image frames from the at least one segment, and stacking the image frames to form image difference training data; and

training the second dynamic gesture recognition model by using the categories of the dynamic gestures as the supervision data and using the image difference training data.

14. A control method using gesture interaction, comprising:

obtaining a video stream;

determining a dynamic gesture recognition result of the video stream by the method according to any one of claims 1 to 13; and

controlling a device to execute an operation corresponding to the dynamic gesture recognition result.

15. The method according to claim 14, wherein the controlling a device to execute an operation corresponding to the dynamic gesture recognition result comprises:

obtaining an operation instruction corresponding to the dynamic gesture recognition result according to a predetermined correspondence between the dynamic gesture recognition result and the operation instruction, and

controlling the device to execute a corresponding operation according to the operation instruction; or

in response to the dynamic gesture recognition result being a predefined dynamic action, controlling a vehicle to execute an operation corresponding to the predefined dynamic action.

16. The method according to claim 15, wherein the device comprises a vehicle, the controlling the device to execute a corresponding operation according to the operation instruction comprises:

controlling a window, a door, or a vehicle-mounted system of the vehicle according to the operation instruction.

17. The method according to claim 15, wherein the predefined dynamic action comprises a dynamic gesture comprising at least one of: single-finger clockwise/counterclockwise rotation, palm left/right swing, two-finger poke, extending the thumb and pinky finger, press-down with the palm downward, lift with the palm upward, fanning to the left/right with the palm, left/right movement with the thumb extended, long slide to the left/right with the palm, changing a fist into a palm with the palm upward, changing a palm into a fist with the palm upward, changing a palm into a fist with the palm downward, changing a fist into a palm with the palm downward, single-finger slide, pinch-in with multiple fingers, single-finger double click, single-finger single click, multi-finger double click, or multi-finger single click; and

the operation corresponding to the predefined dynamic action comprises at least one of: volume up/down, song switching, song pause/resume, call answering or initiation, hang-up or call rejection, air conditioning temperature increase or decrease, multi-screen interaction, sunroof opening, sunroof closing, door lock locking, door lock unlocking, drag for navigation, map zoom-out, or map zoom-in.

18. An electronic device, comprising: a memory, configured to store executable instructions; and

a processor, configured to communicate with the memory to execute the executable instructions to complete operations of the method for recognizing a

dynamic gesture according to any one of claims 1 to 13, or operations of the control method using gesture interaction according to any one of claims 14 to 17.

19. A computer-readable storage medium, configured to store computer-readable instructions, wherein when the instructions are executed, cause a computer to execute operations of the method for recognizing a dynamic gesture according to any one of claims 1 to 13, or operations of the control method using gesture interaction according to any one of claims 14 to 17.

20. A program, configured to cause a computer to perform the method for recognizing a dynamic gesture according to any one of claims 1 to 13, or the control method using gesture interaction according to any one of claims 14-17.