CLAIMS

What is claimed is:

1. A method for detecting abnormality for unmanned checkout, the method comprising:

obtaining, by a depth camera in a detection area for unmanned checkout, image data that comprises a first depth image and a corresponding RGB image;

detecting, using gesture detection and based on the first depth image and the corresponding RGB image, one or more gestures of customers that enter the detection area for unmanned checkout, wherein determining the one or more gestures of the customers comprises:

identifying a first image region within the first depth image, wherein the first image region i) corresponds to hands of the customers and ii) comprises one or more image portions within the first depth image;

in response to identifying the first image region within the first depth image,

mapping the first image region in the first depth image to a second image region in the RGB image, the second image region corresponding to the hands of the customers; and

determining, based on the second image region in the RGB image, the one or more gestures of the customers;

determining that the one or more gestures comprise an abnormal gesture; and

in response to determining that the one or more gestures comprise an abnormal gesture, notifying a central control system to stop the unmanned checkout and trigger an alarm.

2. The method according to claim 1, further comprising:

determining, based on the first depth image, a quantity of customers entering the detection area for the unmanned checkout, wherein determining the quantity of customers comprises:

comparing, the first depth image with a background model, to obtain a foreground image, wherein the foreground image is associated with one or more moving objects, and the background model is obtained based on a plurality of second depth images including stationary objects;

identifying, in the foreground image, one or more regions representing the one or more moving objects, wherein each of the one or more regions is identified based on identifying a cluster of pixel values within a predetermined range of one another; and

determining the quantity of customers based on the one or more regions,

wherein determining the quantity of customers or the gesture detection comprises:

determining that the one or more moving object are the customers; and

performing connected component analysis on the foreground image based on one or more depth values of the first depth image to obtain connected components representing the one or more regions.

3. The method according to claim 2, wherein determining the quantity of customers further comprises:

obtaining a headcount of the customers in the detection area based on results of the connected component analysis.

4. The method according to claim 2, wherein the first depth image and the corresponding RGB image are obtained by the depth camera at substantially a same time of a same scenario, and the gesture detection comprises:

detecting each of the customers in the foreground image based on results of the connected component analysis to obtain the connected components that represent the customers; and

determining the one or more gestures of the customers based on the connected components in the first depth image and one or more regions corresponding to the connected components in the RGB image.

5. The method according to claim 4, wherein determining the one or more gestures further comprises:

locating the first image region corresponding to the hands of the customers in the connected components in the first depth image; and

determining whether the first image region corresponding to the hands are located.

6. The method according to claim 5, further comprising:

performing skeleton detection of the hands on the second image region in the RGB image; and

determining the one or more gestures of the customers based on results of the skeleton detection.

7. The method according to claim 5, further comprising:

in response to determining that the first image region corresponding to the hands are not located, identifying regions in the RGB image corresponding to the connected components in the first depth image;

performing skeleton detection of the hands on the regions in the RGB image; and

determining the one or more gestures of the customers based on results of the skeleton detection.

8. The method according to claim 1, further comprising:

detecting, based on the first depth image, that the customers enter the detection area that is unoccupied; and

notifying the central control system to start unmanned checkout.

9. The method according to claim 8, further comprising:

detecting, based on the first depth image, that the customers exit the detection area and the detection area becomes unoccupied; and

notifying the central control system to stop unmanned checkout.

10. A non-transitory, computer-readable medium storing one or more instructions executable by a computer system to perform operations comprising:

obtaining, by a depth camera in a detection area for unmanned checkout, image data that comprises a first depth image and a corresponding RGB image;

detecting, using gesture detection and based on the first depth image and the corresponding RGB image, one or more gestures of customers that enter the detection area for unmanned checkout, wherein determining the one or more gestures of the customers comprises:

identifying a first image region within the first depth image, wherein the first image region i) corresponds to hands of the customers and ii) comprises one or more image portions within the first depth image;

in response to identifying the first image region within the first depth image,

mapping the first image region in the first depth image to a second image region in the RGB image, the second image region corresponding to the hands of the customers; and

determining, based on the second image region in the RGB image, the one or more gestures of the customers;

determining that the one or more gestures comprise an abnormal gesture; and

in response to determining that the one or more gestures comprise an abnormal gesture, notifying a central control system to stop the unmanned checkout and trigger an alarm.

11. The non-transitory, computer-readable medium according to claim 10, wherein the operations further comprise:

determining, based on the first depth image, a quantity of customers entering the detection area for the unmanned checkout, wherein determining the quantity of customers comprises:

comparing, the first depth image with a background model, to obtain a foreground image, wherein the foreground image is associated with one or more moving objects, and the background model is obtained based on a plurality of second depth images including stationary objects;

identifying, in the foreground image, one or more regions representing the one or more moving objects, wherein each of the one or more regions is identified based on identifying a cluster of pixel values within a predetermined range of one another; and

determining the quantity of customers based on the one or more regions,

wherein determining the quantity of customers or the gesture detection comprises:

determining that the one or more moving object are the customers; and

performing connected component analysis on the foreground image based on one or more depth values of the first depth image to obtain connected components representing the one or more regions.

12. The non-transitory, computer-readable medium according to claim 11, wherein determining the quantity of customers further comprises:

obtaining a headcount of the customers in the detection area based on results of the connected component analysis.

13. The non-transitory, computer-readable medium according to claim 11, wherein the first depth image and the corresponding RGB image are obtained by the depth camera at substantially a same time of a same scenario, and the gesture detection comprises:

detecting each of the customers in the foreground image based on results of the connected component analysis to obtain the connected components that represent the customers; and

determining the one or more gestures of the customers based on the connected components in the first depth image and one or more regions corresponding to the connected components in the RGB image.

14. The non-transitory, computer-readable medium according to claim 13, wherein determining the one or more gestures further comprises:

locating the first image region corresponding to the hands of the customers in the connected components in the first depth image; and

determining whether the first image region corresponding to the hands are located.

15. The non-transitory, computer-readable medium according to claim 14, the operations further comprising:

performing skeleton detection of the hands on the second image region in the RGB image; and

determining the one or more gestures of the customers based on results of the skeleton detection.

16. The non-transitory, computer-readable medium according to claim 14, the operations further comprising:

in response to determining that the first image region corresponding to the hands are not located, identifying regions in the RGB image corresponding to the connected components in the first depth image;

performing skeleton detection of the hands on the regions in the RGB image; and

determining the one or more gestures of the customers based on results of the skeleton detection.

17. The non-transitory, computer-readable medium according to claim 10, the operations further comprising:

detecting, based on the first depth image, that the customers enter the detection area that is unoccupied; and

notifying the central control system to start unmanned checkout.

18. The non-transitory, computer-readable medium according to claim 17, the operations further comprising:

detecting, based on the first depth image, that the customers exit the detection area and the detection area becomes unoccupied; and

notifying the central control system to stop unmanned checkout.

19. A computer-implemented system, comprising:

one or more computers; and

one or more computer memory devices interoperably coupled with the one or more computers and having tangible, non-transitory, machine-readable media storing one or more instructions that, when executed by the one or more computers, perform operations comprising:

obtaining, by a depth camera in a detection area for unmanned checkout, image data that comprises a first depth image and a corresponding RGB image;

detecting, using gesture detection and based on the first depth image and the corresponding RGB image, one or more gestures of customers that enter the detection area for unmanned checkout using gesture detection, wherein determining the one or more gestures of the customers comprises:

identifying a first image region within the first depth image, wherein the first image region i) corresponds to hands of the customers and ii) comprises one or more image portions within the first depth image;

in response to identifying the first image region within the first depth image,

mapping the first image region in the first depth image to a second image region in the RGB image, the second image region corresponding to the hands of the customers; and

determining, based on the second image region in the RGB image, the one or more gestures of the customers;

determining that the one or more gestures comprise an abnormal gesture; and

in response to determining that the one or more gestures comprise an abnormal gesture, notifying a central control system to stop the unmanned checkout and trigger an alarm.

20. The computer-implemented system according to claim 19, wherein the operations further comprise:

determining, based on the first depth image, a quantity of customers entering the detection area for the unmanned checkout, wherein determining the quantity of customers comprises:

comparing, the first depth image with a background model, to obtain a foreground image, wherein the foreground image is associated with one or more moving objects, and the background model is obtained based on a plurality of second depth images including stationary objects;

identifying, in the foreground image, one or more regions representing the one or more moving objects, wherein each of the one or more regions is identified based on identifying a cluster of pixel values within a predetermined range of one another; and

determining the quantity of customers based on the one or more regions,

wherein determining the quantity of customers or the gesture detection comprises:

determining that the one or more moving object are the customers; and

performing connected component analysis on the foreground image based on one or more depth values of the first depth image to obtain connected components representing the one or more regions.

21. The computer-implemented system according to claim 20, wherein determining the quantity of customers further comprises:

obtaining a headcount of the customers in the detection area based on results of the connected component analysis.

22. The computer-implemented system according to claim 20, wherein the first depth image and the corresponding RGB image are obtained by the depth camera at substantially a same time of a same scenario, and the gesture detection comprises:

detecting each of the customers in the foreground image based on results of the connected component analysis to obtain the connected components that represent the customers; and

determining the one or more gestures of the customers based on the connected components in the first depth image and one or more regions corresponding to the connected components in the RGB image.

23. The computer-implemented system according to claim 22, wherein determining the one or more gestures further comprises:

locating the first image region corresponding to the hands of the customers in the connected components in the first depth image; and

determining whether the first image region corresponding to the hands are located.

24. The computer-implemented system according to claim 23, the operations further comprising:

performing skeleton detection of the hands on the second image region in the RGB image; and

determining the one or more gestures of the customers based on results of the skeleton detection.

25. The computer-implemented system according to claim 23, the operations further comprising:

in response to determining that the first image region corresponding to the hands are not located, identifying regions in the RGB image corresponding to the connected components in the first depth image;

performing skeleton detection of the hands on the regions in the RGB image; and

determining the one or more gestures of the customers based on results of the skeleton detection.

26. The computer-implemented system according to claim 19, the operations further comprising:

detecting, based on the first depth image, that the customers enter the detection area that is unoccupied; and

notifying the central control system to start unmanned checkout.

27. The computer-implemented system according to claim 26, the operations further comprising:

detecting, based on the first depth image, that the customers exit the detection area and the detection area becomes unoccupied; and

notifying the central control system to stop unmanned checkout.