CLAIMS

1. A target tracking method, comprising:

detecting a to-be-processed image to obtain a hand detection result;

in response to the hand detection result comprising a bounding box of hand, determining a hand in the bounding box of hand, with a hand pose conforming to a hand pose in a target gesture and in a pre-defined area in the to-be-processed image, as a target hand; and

tracking the target hand in a video stream according to the target hand in the to-be-processed image, wherein images in the video stream and the to-be-processed image are obtained by capturing a same target area, the images in the video stream are obtained by capturing subsequent to capture of the to-be-processed image, the target area comprises two sub-areas, the pre-defined area in the to-be-processed image comprises two set sub-areas corresponding to the two sub-areas respectively, and the hand in the bounding box in the pre-defined area refers to that the hand in the bounding box is located in any of the two set sub-areas.

2. The method according to claim 1, wherein determining that the hand pose in the bounding box of hand conforms to the hand pose in the target gesture comprises:

in response to the hand detection result comprising the bounding box of hand, cropping an image located in the bounding box of hand in the to-be-processed image; and

determining, according to the cropped image, that a hand pose in the cropped image conforms to the hand pose in the target gesture.

3. The method according to any one of claims 1 or 2, wherein the hand pose in the target gesture is that two hands both face up;

determining that the hand pose in the bounding box of hand conforms to the hand pose in the target gesture comprises:

inputting the image in the bounding box of hand to a hand classification neural network to obtain a hand classification result, the hand classification result comprising a hand facing up or the hand with other poses; and

if classification results in two bounding boxes of hand are that two hands both face up and two hands in the two bounding boxes are respectively located in the two set sub-areas, determining that the hand poses in the two bounding boxes of hand conform to the hand pose in

the target gesture.

4. The method according to claim 3, wherein the hand classification neural network is obtained by training with hand image samples, a hand image sample comprising an image sample with the hand facing up, and an annotation result of the image sample with the hand facing up being the palm of the hand.

5. The method according to claim 1, wherein determining that the hand in the bounding box of hand is located in the pre-defined area in the to-be-processed image comprises:

determining that a center point of the bounding box of hand is located in the pre-defined area of the to-be-processed image.

6. The method according to any one of claims 1 to 5, wherein the tracking the target hand in a video stream according to the target hand in the to-be-processed image comprises:

for an image frame in the video stream, determining a first matrix and a second matrix of the image frame according to a first matrix and a second matrix of a previous image frame of the image frame,

wherein the first matrix of the image frame comprises position information of the center of the bounding box of hand of the target hand in the image frame, and the second matrix of the image frame represents an error of each element in the first matrix of the image frame.

7. The method according to claim 6, wherein the determining a first matrix and a second matrix of the image frame according to a first matrix and a second matrix of a previous image frame of the image frame comprises:

in response to a hand detection result of the image frame not comprising the bounding box of hand of the target hand, determining the first matrix of the previous image frame as the first matrix of the image frame, and correcting the second matrix of the previous image frame with a correction matrix to obtain the second matrix of the image frame, wherein the correction matrix represents an error of each element in the second matrix of the previous image frame.

8. The method according to claim 6, wherein the determining a first matrix and a second matrix of the image frame according to a first matrix and a second matrix of a previous image

frame of the image frame comprises:

in response to the hand detection result of the image frame comprising the bounding box of hand of the target hand, predicting, according to a time interval between the image frame and the previous image frame, the first matrix and the second matrix of the image frame according to the first matrix of the previous image frame;

obtaining a third matrix and a fourth matrix of the image frame according to the position information of the bounding box of hand of the target hand, wherein the third matrix comprises the position information of the center of the bounding box of hand of the target hand in the image frame, and the fourth matrix represents an error of each element in the third matrix; and

multiplying Gaussian distributions corresponding to the second matrix by Gaussian distributions corresponding to the fourth matrix to obtain new Gaussian distributions, updating the first matrix according to an average value of the obtained new Gaussian distributions, and updating the second matrix according to covariance of the obtained new Gaussian distributions.

9. A target tracking apparatus, comprising:

a detecting unit, configured to detect a to-be-processed image to obtain a hand detection result;

a determining unit, configured to: in response to the hand detection result comprising a bounding box of hand, determine in the bounding box of hand, with a hand pose conforming to a hand pose in a target gesture and in a pre-defined area in the to-be-processed image, as a target hand; and

a tracking unit, configured to track the target hand in a video stream according to the target hand in the to-be-processed image, wherein images in the video stream and the to-be-processed image are obtained by capturing a same target area, the images in the video stream are obtained by capturing subsequent to capture of the to-be-processed image, the target area comprises two sub-areas, the pre-defined area in the to-be-processed image comprises two set sub-areas corresponding to the two sub-areas respectively, and the hand in the bounding box in the pre-defined area refers to that the hand in the bounding box is located in any of the two set sub-areas.

10. The apparatus according to claim 9, wherein the determining unit is configured to:

in response to the hand detection result comprising the bounding box of hand, crop an image located in the bounding box of hand in the to-be-processed image; and

determine, according to the cropped image, that a hand pose in the cropped image conforms to the hand pose in the target gesture.

11. The apparatus according to any one of claims 9 or 10, wherein the hand pose in the target gesture is that two hands both face up;

the determining unit is configured to:

input the image in the bounding box of hand to a hand classification neural network to obtain a hand classification result, the hand classification result comprising a hand facing up or the hand with other poses; and

if classification results in two bounding boxes of hand are that two hands both face up and two hands in the two bounding boxes are respectively located in the two set sub-areas, determine that the hand poses in the two bounding boxes of hand conform to the hand pose in the target gesture.

12. The apparatus according to claim 11, wherein the hand classification neural network is obtained by training with hand image samples, a hand image sample comprising an image sample with the hand facing up, and an annotation result of the image sample with the hand facing up being the palm of the hand.

13. The apparatus according to claim 9, wherein the determining unit configured to determine that the hand in the bounding box of hand is located in the pre-defined area in the to-be-processed image, is configured to:

determine that a center point of the bounding box of hand is located in the pre-defined area of the to-be-processed image.

14. The apparatus according to any one of claims 9 to 13, wherein the tracking unit is configured to:

for an image frame in the video stream, determine a first matrix and a second matrix of the image frame according to a first matrix and a second matrix of a previous image frame of the image frame,

wherein the first matrix of the image frame comprises position information of the center of

the bounding box of hand of the target hand in the image frame, and the second matrix of the image frame represents an error of each element in the first matrix of the image frame.

15. The apparatus according to claim 14, wherein the tracking unit configured to determine the first matrix and the second matrix of the image frame according to the first matrix and the second matrix of the previous image frame of the image frame, is configured to:

in response to a hand detection result of the image frame not comprising the bounding box of hand of the target hand, determine the first matrix of the previous image frame as the first matrix of the image frame, and correct the second matrix of the previous image frame with a correction matrix to obtain the second matrix of the image frame, wherein the correction matrix represents an error of each element in the second matrix of the previous image frame.

16. The apparatus according to claim 14, wherein the tracking unit configured to determine the first matrix and the second matrix of the image frame according to the first matrix and the second matrix of the previous image frame of the image frame, is specifically configured to:

in response to the hand detection result of the image frame comprising the bounding box of hand of the target hand, predict, according to a time interval between the image frame and the previous image frame, the first matrix and the second matrix of the image frame according to the first matrix of the previous image frame;

obtain a third matrix and a fourth matrix of the image frame according to the position information of the bounding box of hand of the target hand, wherein the third matrix comprises the position information of the center of the bounding box of hand of the target hand in the image frame, and the fourth matrix represents an error of each element in the third matrix; and

multiply Gaussian distributions corresponding to the second matrix by Gaussian distributions corresponding to the fourth matrix to obtain new Gaussian distributions, update the first matrix according to an average value of the obtained new Gaussian distributions, and update the second matrix according to covariance of the obtained new Gaussian distributions.

17. An electronic device, comprising a memory and a processor, wherein the memory is configured to store computer instructions runnable on the processor, and the processor is configured to implement the method according to any one of claims 1 to 8 when executing the computer instructions.

18. A computer-readable storage medium, having a computer program stored thereon, wherein when the program is executed by a processor, the method according to any one of claims 1 to 8 is implemented.

19. A method of tracking a hand of croupier, comprising:

detecting a to-be-processed image comprising some or all of areas of a game table, to obtain a hand detection result;

determining whether a hand in each of two bounding boxes of hand comprised in the hand detection result faces up, and whether two hands in the two bounding boxes are respectively located in two set sub-areas of the to-be-processed image, wherein the game table comprises two sub-areas, the two set sub-areas corresponds to the two sub-areas respectively;

when determining that the hand in each of the two bounding boxes faces up and that the two hands in the two bounding boxes are respectively located in the two set sub-areas of the to-be-processed image, determining that the two hands of the croupier make a palm-up gesture; and

tracking the two hands of the croupier in a video stream, wherein images in the video stream and the to-be-processed image are obtained by capturing the game table, the images in the video stream are obtained by capturing subsequent to capture of the to-be-processed image.