

DESCRIPTION

GESTURE RECOGNITION METHOD AND APPARATUS, ELECTRONIC DEVICE, AND STORAGE MEDIUM

Technical Field

The present disclosure relates to computer vision technologies, and in particular, to a gesture recognition method, an apparatus, an electronic device, and a storage medium.

Background

In recent years, with continuous development of an artificial intelligence technology, the artificial intelligence technology achieves good effects in aspects such as computer vision and speech recognition. When target recognition is performed in some relatively special scenes (such as a tabletop game scene), it would be easy to cause misjudgment during recognition by using a single frame of picture, while an excessively large deep learning model during recognition by using multiple frames of videos will cause large system time consumption.

Summary

The present disclosure provides solutions for gesture recognition.

According to one aspect of the present disclosure, provided is a gesture recognition method. The method includes: obtaining a hand detection result by detecting a to-be-processed image; in response to the hand detection result including a bounding box of hand, determining that a hand pose in the bounding box of hand conforms to a hand pose in a target gesture; determining that the hand is located in a pre-defined area of the to-be-processed image; and determining that the hand makes the target gesture.

In combination with any implementation provided by the present disclosure, the in response to the hand detection result including a bounding box of hand, determining that a hand pose in the bounding box of hand conforms to a hand pose in a target gesture includes: in response to the hand detection result including 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 the hand pose in the cropped image conforms to the hand pose in the target gesture.

In combination with any implementation provided by the present disclosure, the hand pose in the target gesture is two hands both facing up; and the determining that a hand pose in the bounding box of hand conforms to a hand pose in a target gesture includes: 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 including a hand facing up or the hand with other poses; and if hand

classification results of two bounding boxes of hand are hands facing up, determining that the hand poses in the bounding boxes of hand conform to the hand pose in the target gesture.

In combination with any implementation provided by the present disclosure, the hand classification neural network is obtained by training with hand image samples, a hand image sample including 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.

In combination with any implementation provided by the present disclosure, the determining the hand is located in a pre-defined area of the to-be-processed image includes: determining that a center point of the bounding box of hand is located in the pre-defined area.

In combination with any implementation provided by the present disclosure, the to-be-processed image is an image obtained by capturing a target area, and the pre-defined area includes two set sub-areas corresponding to two sub-areas in the target area respectively in the to-be-processed image; and the determining that a center point of the bounding box of hand is located in the pre-defined area includes: determining that the center points of the two bounding boxes of hand are located in the two set sub-areas, respectively.

In combination with any implementation provided by the present disclosure, the to-be-processed image is an image of the target area captured at an overhead view.

According to one aspect of the present disclosure, provided is a gesture recognition apparatus. The apparatus includes: an obtaining unit, configured to obtain a hand detection result by detecting a to-be-processed image; a first determining unit, configured to, in response to the hand detection result including a bounding box of hand, determine that a hand pose in the bounding box of hand conforms to a hand pose in a target gesture; a second determining unit, configured to determine that the hand is located in a pre-defined area of the to-be-processed image; and a third determining unit, configured to determine that the hand makes the target gesture.

In combination with any implementation provided by the present disclosure, the first determining unit is specifically configured to, in response to the hand detection result including 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 the hand pose in the cropped image conforms to the hand pose in the target gesture.

In combination with any implementation provided by the present disclosure, the hand pose in the target gesture is two hands both facing up; and the first determining unit configured to determine that a hand pose in the bounding box of hand conforms to a hand pose in a target gesture, is specifically 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 including a hand facing up or the hand with other poses; and if hand classification results of two bounding boxes of hand are hands facing up, determine that the hand poses in the bounding boxes of hand conform to the hand pose in the target gesture.

In combination with any implementation provided by the present disclosure, the hand classification neural network is obtained by training with hand image samples, a hand image sample including 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.

In combination with any implementation provided by the present disclosure, the second determining unit is specifically configured to: determine that a center point of the bounding box of hand is located in the pre-defined area.

In combination with any implementation provided by the present disclosure, the to-be-processed image is an image obtained by capturing a target area, and the pre-defined area includes two set sub-areas corresponding to two sub-areas in the target area respectively in the to-be-processed image; and the second determining unit is specifically configured to: determine that the center points of the two bounding boxes of hand are located in the two set sub-areas, respectively.

In combination with any implementation provided by the present disclosure, the to-be-processed image is an image of the target area captured at an overhead view.

According to one aspect of the present disclosure, provided is an electronic device, including: a memory and a processor, where the memory is configured to store computer instructions which can be run on a processor, and the processor is configured to implement the gesture recognition method according to any implementation of the present disclosure when executing the computer instructions.

According to one aspect of the present disclosure, provided is a computer-readable storage medium having a computer program stored thereon, where when the program is executed by a processor, the gesture recognition method according to any implementation of the present disclosure is implemented.

In a gesture recognition method, an apparatus, an electronic device, and a storage medium according to one or more embodiments of the present disclosure, the hand pose in the bounding box of hand in the hand detection result is detected, and on the basis of determining that the hand pose conforms to the hand pose in the target gesture, plus constraint of detection position information, target gesture recognition is performed. Fast and high-precision target gesture recognition can be realized by using a single image frame.

Brief Description of the Drawings

To describe the technical solutions in one or more embodiments of the description or the prior art more clearly, the accompanying drawings required for describing the embodiments or the prior art are briefly introduced below. Apparently, the accompanying drawings in the following description show merely some embodiments described in one or more embodiments of the description, and a person of ordinary skill in the art can still derive other accompanying drawings from these accompanying drawings without creative efforts.

FIG. 1 is a flowchart of a gesture recognition method provided according to at least one embodiment of the present disclosure;

FIG. 2 is a schematic diagram of an application scene of an image processing method provided according to at least one embodiment of the present disclosure;

FIGS. 3A-3F are schematic diagrams of a hand classification result provided according to at least one embodiment of the present disclosure;

FIG. 4 is a schematic diagram of an application scene of an image processing method provided according to at least one embodiment of the present disclosure;

FIG. 5 is a schematic structural diagram of a gesture recognition apparatus provided according to at least one embodiment of the present disclosure; and

FIG. 6 is a schematic structural diagram of an electronic device provided according to at least one embodiment of the present disclosure.

Detailed Description

To make a person skilled in the art better understand the technical solutions in one or more embodiments of the description, the technical solutions in the one more embodiments of the description are clearly and fully described below with reference to the accompanying drawings in the one or more embodiments of the description. Apparently, the described embodiments are merely some of the embodiments of the description, but not all the embodiments. Based on the one or more embodiments of the description, all other embodiments obtained by a person of ordinary skill in the art without involving an inventive effort shall fall within the scope of protection of the present disclosure.

At least one embodiment of the present disclosure provides a gesture recognition method. The gesture recognition method may be executed by an electronic device such as a terminal device or a server. The terminal device may be a User Equipment (UE), a mobile device, a user terminal, a terminal, a cellular phone, a cordless phone, a Personal Digital Assistant (PDA), a handheld device,

a computing device, a vehicle-mounted device, a wearable device, or the like. The method may be implemented by a processor by invoking computer-readable instructions stored in a memory.

FIG. 1 shows a flow of the gesture recognition method. As shown in FIG. 1, the method may include steps 101-104.

At step 101, a hand detection result is obtained by detecting a to-be-processed image.

In the embodiments of the present disclosure, the to-be-processed image may be an image of a tabletop area captured by an image acquisition device (e.g., a camera). The image may include a complete tabletop image or a partial tabletop image.

The to-be-processed image captured by the camera may be one frame in a video stream or an image captured in real time, which is not limited in the present disclosure.

In the embodiments of the present disclosure, a hand detection result is obtained by detecting a to-be-processed image by using a deep learning network, for example, a faster RCNN network. If a hand is detected, the hand detection result may include a bounding box of hand, the position of the bounding box of hand and the like. A person skilled in the art should understand that the faster RCNN network is merely an example, and other networks may be used for hand detection, which is not limited in the present disclosure.

At step 102, in response to the hand detection result including a bounding box of hand, it is determined that a hand pose in the bounding box of hand conforms to a hand pose in a target gesture.

In some embodiments, the to-be-processed image may be captured at an overhead view, i.e., the to-be-processed image is captured by a birdview camera. For an application scene, i.e., the tabletop, the overhead view is a good photography view angle, and the view angle can generally avoid other objects from blocking a detection target. However, a person skilled in the art should understand that the image captured by the camera at other view angles may also be applied to the gesture recognition method provided according to the embodiments of the present disclosure.

In the embodiments of the present disclosure, the image the bounding box of hand is detected, and a detected feature is compared with a feature of the hand pose in the target gesture to determine whether the hand pose in the bounding box of hand conforms to the hand pose in the target gesture; or hand poses in the bounding box of hand are classified, and it is determined whether the hand poses conform to the hand pose in the target gesture through a classification result. The specific determining method would be described in detail below.

At step 103, it is determined that the hand is located in a pre-defined area of the to-be-

processed image.

In the embodiments of the present disclosure, if the hand detection result includes a bounding box of hand, the hand detection result also includes position information of the bounding box of hand, i.e., position information of the hand in the to-be-processed image is determined.

For the target gesture which appears in a specific area, position constraint is provided for the hand in the bounding box of hand by using the pre-defined area of the to-be-processed image, and it may be further determined whether the detected hand meets the requirements for the target gesture. The pre-defined area of the to-be-processed image may be determined according to the position and internal parameters of the camera for photographing the to-be-processed image, and the specific area corresponding to the target gesture in an actual space.

At step 104, it is determined that the hand makes the target gesture.

If the hand pose in the bounding box of hand conforms to the hand pose in the target gesture, and the hand is located in the pre-defined area of the to-be-processed image, it is determined that the hand makes the target gesture.

In the embodiments of the present disclosure, the hand pose in the bounding box of hand in a hand detection result is detected, and on the basis of determining that the hand pose conforms to a hand pose in a target gesture, plus constraint of detection position information, target gesture recognition is performed. Fast and high-precision target gesture recognition can be realized by using a single image frame.

In some embodiments, if a hand is detected, i.e., the hand detection result includes a bounding box of hand, an image located in the bounding box of hand in the to-be-processed image may be cropped; and according to the cropped image, it is determined that the hand pose in the cropped image conforms to the hand pose in the target gesture.

By cropping the image in the bounding box of hand, and detecting the hand pose according to the cropped image, the efficiency and accuracy of hand gesture detection may be improved.

In some embodiments, a center point of the bounding box of hand is determined as the position of the hand. A person skilled in the art should understand that other position points of the bounding box of hand may also be determined as the position of the hand, for example, when a target gesture to be recognized has a closest relationship with one of fingers, the position of the finger in the bounding box of hand may be determined as the position of the hand. The specific determining mode of the hand position may be determined according to the relationship between the target gesture to be recognized and the bounding box of hand.

In some embodiments, it may be determined that the hand pose in the bounding box of hand conforms to the hand pose in the target gesture in the following ways: inputting the image in the bounding box of hand to a pre-trained hand classification neural network to obtain a hand classification result, and determining whether the hand pose in the bounding box of hand conforms to the hand pose in the target gesture through the classification result.

The used hand classification neural network may be a convolutional classification model, for example, a convolutional classification model using Resnet 18 as a backbone network. A person skilled in the art should understand that the used hand classification neural network may be determined according to a classification type to be obtained; the classification type to be obtained may be determined according to a target gesture to be recognized. No limitation is made thereto in the present disclosure.

In one example, if the target gesture to be recognized relates only to a hand facing up, the classification type to be obtained may include the hand facing up or with other poses, where the other poses include a hand pose other than the hand facing up and the hand being blocked. The hand classification neural network may be trained with image samples including the hand facing up, and an annotation result of the image sample with the hand facing up is the palm of the hand.

In one example, if the target gesture to be recognized relate to a hand facing up or back, the classification type to be obtained may include the hand facing up, the hand facing back, and the hand with other poses, where the other poses include hand poses other than the hand facing up and the hand facing back, and the hand being blocked. The hand classification neural network may be trained with image samples including a hand facing up or an image sample including the hand facing back, where an annotation result of the image sample with the hand facing up is the palm of the hand, and the annotation result of the image sample with the hand facing back may be an identifier indicating the back.

In one example, if the target gesture to be recognized relates to a specific pose, the classification type needs to include the specific pose, and the hand classification neural network may be trained with sample hand images including the pose, so that the hand classification neural network recognizes the specific pose from the image in the bounding box of hand.

A person skilled in the art should understand that it may also be determined that the hand pose in the bounding box of hand conforms to the hand pose in the target gesture in other ways, gesture detection is performed by using the hand pose in the bounding box of hand, and the like. No limitation is made thereto in the present disclosure.

By taking the fact that both hands of a target gesture to be recognized face front, and the hands

are located above two specific sub-areas of the tabletop respectively as an example, the gesture recognition method provided according to the embodiments of the present disclosure is described in detail below. A person skilled in the art should understand that the gesture recognition method is not limited to recognizing the target gesture and may also be used for recognizing other gestures.

FIG. 2 is a schematic diagram of an application scene of a gesture recognition method provided according to at least one embodiment of the present disclosure. As shown in FIG. 2, the target gesture to be recognized is the gesture with two palms facing up made above two areas 221 and 222 of a tabletop 20.

First, a hand detection result is obtained by detecting a to-be-processed image including a complete area or a partial area of the tabletop 20 photographed at an overhead view.

If a hand is detected, i.e., the hand detection result includes a bounding box of hand, the image in the bounding box of hand is input to a hand classification neural network to obtain a hand classification result. The hand classification result is used for classifying the input image based on whether the input image is a hand facing up or the hand with other poses, so as to obtain a hand classification result including the hand facing up or the hand with other poses. If the classification results of the two bounding boxes of hand are hands facing up, it is determined that the hand poses in the bounding boxes of hand are hands facing up, and conform to the hand pose in the target gesture.

If both hands are determined to face front, it is determined whether the positions of the two bounding boxes of hand are located in two set sub-areas of the to-be-processed image. The two set sub-areas of the to-be-processed image are determined according to the position and internal parameters of the camera for photographing the to-be-processed image, and the positions of the two areas 221 and 222 of the tabletop 20. In response to the center points of the two bounding boxes of hand being located in the two set sub-areas respectively, it is determined that the hand makes a target gesture, where the two set sub-areas respectively correspond to two sub-areas in a target area in the to-be-processed image. When the target area is a game table, the two sub-areas in the target area may be two sub-areas of the game table, and the two set sub-areas respectively correspond to the two sub-areas of the game table in the to-be-processed image.

In one example, the image cropped according to the bounding box of hand may be input into the hand classification neural network to obtain a hand classification result. If the to-be-processed image includes multiple bounding boxes of hand, the multiple cropped images are input into the hand classification neural network, respectively.

The hand classification neural network used for classifying the input image based on whether

the input image is a hand facing up or the hand with other poses may be trained in the following ways.

First, an image sample including a hand facing up is obtained, and the palm of the hand in the image sample is annotated.

FIGS. 3A and 3B are exemplary image samples with hands facing up. The hand image with the palm facing up may be taken as an image sample with the hand facing up, and an annotation result of the image sample of the hand facing up is the palm of the hand. White points in the palms in FIGS. 3A and 3B are annotations of the image samples with the hands facing up. FIGS. 3C-3F are exemplary hand images with other poses which are a hand image with the hand facing back or the palm being blocked, and the hand images may be annotated as other poses.

The hand classification neural network is trained by using the image sample, and the trained neural network can classify the image in the bounding box of hand so as to obtain a classification result with the hand facing up or the hand with other poses.

In some embodiments, the image in the bounding box of hand may be classified by using the hand classification neural network, so as to obtain a classification result with the hand facing up, or the hand facing back, or the hand with other poses. The hand classification neural network may be trained by obtaining an image sample with the hand facing up, the hand facing back, or the hand with other poses. FIGS. 3A-3F are still taken as examples, where FIGS. 3A and 3B are exemplary image samples with hands facing up, and annotation results of the image samples with hands facing up are palms of the hands; FIGS. 3C and 3D are exemplary image samples with hands facing back. The hand image with a complete hand back and five fingers seen may be taken as a back-side sample hand image, and the hand image is annotated as a back side. FIGS. 3E and 3F are exemplary image samples with hands with other poses. The hand image with a palm, a complete hand back, or five fingers not seen may be taken as an image sample with the hand with other poses, and the hand image is annotated as other poses. In FIGS. 3E and 3F, although the overall poses face up, palms cannot be seen, and therefore, the poses are other poses. The hand classification neural network is trained by using the image sample, and the trained neural network can classify the image in the bounding box of hand so as to obtain a classification result with the hand facing up, or the hand facing back, or the hand with other poses.

In the embodiments of the present disclosure, the image in the bounding box of hand is classified by using the hand classification neural network, so as to recognize whether the hand is presents a pose facing up, i.e., the pose with the palm facing up. In this case, for the used hand classification neural network, compared with a video classification mode, it is easy to obtain and

annotate a training sample, the recognition speed of the model is high, and thus, real-time recognition for the hand pose can be implemented.

On the basis that the hand classification result conforms to the hand pose in the target gesture, position constraint is performed on a pre-defined area of the to-be-processed image to further determine that the detected hand makes the target gesture.

The gesture recognition method provided according to the embodiments of the present disclosure is described below by taking a target gesture as a palm-up gesture of a croupier as an example.

FIG. 4 is a schematic diagram of an application scene of a gesture recognition method provided according to at least one embodiment of the present disclosure. As shown in FIG. 4, in a game scene, games can be played through a game table 40, and players 421, 422, 423 and a croupier 43 are located on opposite sides of the game table. The game table mainly includes: an exchanged object area 47 of the croupier (an area in front of the croupier), exchanged object areas 441, 442 of the players (areas in front of the players), a game playing area 46, and a pre-defined area 47 (which can also be referred to as a hot area).

The palm-up gesture of the croupier indicates that after completing a specific operation or before executing the specific operation, the palms of both hands of the croupier face up and are horizontally placed on the two sides of the exchanged object area of the croupier (as the exchanged object area 47 in FIG. 4) to show others that there is nothing in the hands, and to indicate being capable of proceeding to the next game stage.

In the embodiments of the present disclosure, two sub-areas (e.g., sub-areas 411 and 412 in FIG. 4) on the two sides of the exchanged object area in front of the croupier on the game table may be taken as the specific area of the actual space (i.e., the two sub-areas of the target area). The two set sub-areas corresponding to the to-be-processed image are determined through the position and the internal parameters of the camera for photographing the to-be-processed image, and the position of the specific area so as to perform position constraint on the hand classification result.

The to-be-processed image including a complete area or a partial area of the game table 40 is detected. If the obtained hand detection result includes a bounding box of hand, the image in the bounding box of hand is input into a pre-trained hand classification neural network. The hand classification neural network is capable of classifying the input image based on whether the input image is a hand facing up or the hand with other poses. In response to the classification results of the images of the two hand detections bounding boxes are hands both facing up, it is further determined whether the positions of the two bounding boxes of hand are located in the two set

sub-areas of the to-be-processed image. If the two bounding boxes of hand are respectively located in the two set sub-areas, it is indicated that the two hands respectively enter the hot area on the game table 40, and thus, it may be determined that the detected hands make a palm-up gesture.

After it is determined that the croupier makes a palm-up gesture, the next processing process may be directly entered by generating a trigger signal, for example, monitoring of the game process is enabled, a user does not need to enable monitoring through a button (trigger), and the user experience is improved. Moreover, by recognizing the palm-up gesture of the croupier, whether any forbidden objects are carried in the hands of the croupier may be monitored.

FIG. 5 is a schematic structural diagram of a gesture recognition apparatus provided according to at least one embodiment of the present disclosure. As shown in FIG. 5, the apparatus includes: an obtaining unit 501, configured to obtain a hand detection result by detecting a to-be-processed image; a first determining unit 502, configured to, in response to the hand detection result including a bounding box of hand, determine that a hand pose in the bounding box of hand conforms to a hand pose in a target gesture; a second determining unit 503, configured to determine that the hand is located in a pre-defined area of the to-be-processed image; and a third determining unit 504, configured to determine that the hand makes the target gesture.

In some embodiments, the first determining unit 502 is specifically configured to, in response to the hand detection result including 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 the hand pose in the cropped image conforms to the hand pose in the target gesture.

In some embodiments, the hand pose in the target gesture is two hands both facing up; and the first determining unit configured to determine that a hand pose in the bounding box of hand conforms to a hand pose in a target gesture, is specifically 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 including a hand facing up or the hand with other poses; and if hand classification results of two bounding boxes of hand are hands facing up, determine that the hand poses in the bounding boxes of hand conform to the hand pose in the target gesture.

In some embodiments, the hand classification neural network is obtained by training with hand image samples, a hand image sample including 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.

In some embodiments, the second determining unit is specifically configured to: determine that a center point of the bounding box of hand is located in the pre-defined area.

In some embodiments, the to-be-processed image is an image obtained by capturing a target

area, and the pre-defined area includes two set sub-areas corresponding to two sub-areas in the target area respectively in the to-be-processed image; and the second determining unit 504 is specifically configured to: determine that the center points of the two bounding boxes of hand are located in the two set sub-areas, respectively.

In some embodiments, the to-be-processed image is an image of the game table captured at an overhead view.

FIG. 6 is a schematic structural diagram of an electronic device provided according to at least one embodiment of the present disclosure. The device includes: a processor; and a memory configured to store processor-executable instructions, where the processor is configured to invoke the instructions stored in the memory to execute the gesture recognition method according to at least one embodiment.

The embodiments of the present disclosure further provide a computer readable storage medium having computer program instructions stored thereon, where when the computer program instructions are executed by a processor, the gesture recognition method according to at least one embodiment is implemented.

A person skilled in the art should understand that one or more embodiments of the description may provide a method, a system or a computer program product. Therefore, one or more embodiments of the description may use a form of hardware only embodiments, software only embodiments, or embodiments with a combination of software and hardware. Moreover, one or more embodiments of the description may use a form of a computer program product that is implemented on one or more computer-usable storage media (including but not limited to a disk memory, a CD-ROM, an optical memory, etc.) that include a computer-usable program code.

"And/or" in the description means at least one of the two, e.g., "A and/or B" includes three schemes: A, B, and "A and B".

The embodiments of the description are all described in a progressive way, for same or similar parts in the embodiments, reference can be made to these embodiments, and each embodiment focuses on a difference from other embodiments. In particular, for an embodiment of a data processing device, the embodiment is substantially similar to the method embodiment and therefore is only described briefly, and for related parts, reference can be made to partial descriptions in the method embodiment.

Specific embodiments of the description are described above. Other embodiments are within the scope of the appended claims. In some instances, the actions or steps recited in the claims may be performed in a sequence different from that in the embodiments and still achieve a desirable

result. In addition, the process described in the accompanying drawings does not necessarily need a specific sequence or a consecutive sequence to achieve the desired result. In some implementations, multitask processing and parallel processing can be feasible or can be advantageous.

The embodiments of the subject matter and functional operations described in the description may be implemented in digital electronic circuitry, tangible computer software or firmware, computer hardware including the structures disclosed in the description and structural equivalents thereof, or a combination of one or more thereof. The embodiments of the subject matter described in the description may be implemented as one or more computer programs, i.e., one or more modules of computer program instructions encoded on a tangible non-transitory program carrier to be executed by a data processing apparatus or to control operations of the data processing apparatus. Alternatively or additionally, the program instructions may be encoded on artificially generated propagated signals, such as machine-generated electrical, optical or electromagnetic signals, generated to encode and transmit information to a suitable receiver apparatus for execution by the data processing apparatus. The computer storage medium may be a machine readable storage device, a machine readable storage substrate, a random or serial access memory device, or a combination of one or more thereof.

The processes and logic flows described in the description may be performed by one or more programmable computers for executing one or more computer programs to perform corresponding functions by performing operations according to input data and generating output. The processes and logic flows may also be performed by a dedicated logic circuit, such as an FPGA (field programmable gate array) or an ASIC (application specific integrated circuit), and the apparatus may also be implemented as a dedicated logic circuit.

Computers applicable to executing computer programs include, for example, a universal and/or dedicated microprocessor, or any other type of central processing unit. Usually, the central processing unit will receive instructions and data from a read-only memory and/or a random access memory. Basic components of the computer include a central processing unit for implementing or executing instructions and one or more memory devices storing instructions and data. Usually, the computer may also include one or more mass storage devices for storing data, such as a magnetic disk, a magneto-optical disk or an optical disk, or the computer will be operably coupled to the mass storage device to receive data from the mass storage device or transmit data to the mass storage device, or to do both. However, the computer may not necessarily have such device. Further, the computer may be embedded in another device, such as a mobile phone, a Personal Digital Assistant (PDA), a mobile audio or video player, a game console, a Global

Positioning System (GPS) receiver, or a portable storage device such as a Universal Serial Bus (USB) flash drive, which are only illustrated as examples.

Computer readable media applicable to storing computer program instructions and data include all forms of non-volatile memories, media and memory devices, such as a semiconductor memory device (e.g., an Erasable Programmable Read-Only Memory (EPROM), an Electrically Erasable Programmable Read-Only Memory (EEPROM), and a flash memory device), a magnetic disk (e.g., an internal hard disk or a mobile disk), a magneto-optical disk and CD-ROM and DVD-ROM disks. The processor and the memory may be supplemented by or incorporated in a dedicated logical circuit.

While the description contains many specific implementation details, these should not be construed as limitations on the scope of any invention or of what may be claimed, but rather as descriptions of features that may be specific to particular implementations of particular inventions. Certain features that are described in multiple embodiments in the description can also be implemented in combination in a single implementation. Conversely, various features that are described in the context of a single implementation can also be implemented in multiple implementations separately or in any suitable subcombination. Moreover, although features may be described above as acting in certain combinations and even initially claimed as such, one or more features from a claimed combination can in some cases be excised from the combination, and the claimed combination may be directed to a subcombination or variation of a subcombination.

Similarly, although operations are described in the accompanying drawings in a particular order, this should not be understood as requiring that such operations be performed in the particular order shown or in sequential order, or that all illustrated operations be performed, to achieve desirable results. In certain circumstances, multitasking and parallel processing may be advantageous. Moreover, the separation of various system modules and components in the implementations described above should not be understood as requiring such separation in all implementations, and it should be understood that the described program components and systems can generally be integrated together in a single software product or packaged into multiple software products.

Therefore, particular embodiments of the subject matter have been described. Other embodiments are within the scope of the appended claims. In some cases, the actions recited in the claims can be performed in a different order and still achieve desirable results. In addition, the processes depicted in the accompanying drawings do not necessarily require the particular order shown, or sequential order, to achieve desirable results. In certain implementations, multitasking

and parallel processing may be advantageous.

The above descriptions are only preferred embodiments of one or more embodiments of the description and are not intended to limit one or more embodiments of the description. Any modifications, equivalent substitutions and improvements made without departing from the spirit and principle of one or more embodiments of the description are intended to be included within the scope of one or more embodiments of the description.

CLAIMS

1. A gesture recognition method, comprising:

obtaining a hand detection result by detecting a to-be-processed image;

in response to the hand detection result comprising a bounding box of hand, determining that a hand pose in the bounding box of hand conforms to a hand pose in a target gesture;

determining that the hand is located in a pre-defined area of the to-be-processed image; and

determining that the hand makes the target gesture.

2. The method according to claim 1, wherein the in response to the hand detection result comprising a bounding box of hand, determining that a hand pose in the bounding box of hand conforms to a hand pose in a 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 the hand pose in the cropped image conforms to the hand pose in the target gesture.

3. The method according to claim 1 or 2, wherein the hand pose in the target gesture is two hands both facing up; and

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

inputting the image in the bounding box of hand into 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 hand classification results of two bounding boxes of hand are hands facing up, determining that the hand poses in the 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 any one of claims 1 to 4, wherein the determining that the hand is located in a pre-defined area of the to-be-processed image comprises:

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

6. The method according to claim 5, wherein the to-be-processed image is an image obtained by capturing a target area, and the pre-defined area comprises two set sub-areas corresponding to two sub-areas in the target area respectively in the to-be-processed image; and

the determining that a center point of the bounding box of hand is located in the pre-defined area comprises: determining that the center points of the two bounding boxes of hand are located in the two set sub-areas, respectively.

7. The method according to claim 6, wherein the to-be-processed image is an image of the target area captured at an overhead view.

8. A gesture recognition apparatus, comprising:

an obtaining unit, configured to obtain a hand detection result by detecting a to-be-processed image;

a first determining unit, configured to, in response to the hand detection result comprising a bounding box of hand, determine that a hand pose in the bounding box of hand conforms to a hand pose in a target gesture;

a second determining unit, configured to determine that the hand is located in a pre-defined area of the to-be-processed image; and

a third determining unit, configured to determine that the hand makes the target gesture.

9. The apparatus according to claim 8, wherein the first determining unit is specifically 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 the hand pose in the cropped image conforms to the hand pose in the target gesture.

10. The apparatus according to claim 8 or 9, wherein the hand pose in the target gesture is two hands both facing up; and

the first determining unit configured to determine that the hand pose in the bounding box of hand conforms to the hand pose in a target gesture, is specifically configured to:

input the image in the bounding box of hand into 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 hand classification results of two bounding boxes of hand are hands facing up, determine that the hand poses in the bounding boxes of hand conform to the hand pose in the target gesture.

11. The apparatus according to claim 10, 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.

12. The apparatus according to any one of claims 8 to 11, wherein the second determining unit is specifically configured to:

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

13. The apparatus according to claim 12, wherein the to-be-processed image is an image obtained by capturing a target area, and the pre-defined area comprises two set sub-areas corresponding to two sub-areas in the target area respectively in the to-be-processed image; and

the second determining unit is specifically configured to:

determine that the center points of the two bounding boxes of hand are located in the two set

sub-areas, respectively.

14. The apparatus according to claim 13, wherein the to-be-processed image is an image of the target area captured at an overhead view.

15. An electronic device, comprising: a memory and a processor, wherein the memory is configured to store computer instructions which can be run on the processor, and the processor is configured to implement the method according to any one of claims 1 to 7 when executing the computer instructions.

16. 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 7 is implemented.