

TARGET TRACKING METHOD AND APPARATUS, AND ELECTRONIC DEVICE AND STORAGE MEDIUM

The present application claims priority to Chinese Patent Application No.
5 CN201810184967.8, entitled "TARGET TRACKING METHOD AND
APPARATUSE, ELECTRONIC DEVICE, PROGRAM, AND STORAGE MEDIUM",
and Chinese Patent Application No. CN201810182797.X, entitled "TARGET
TRACKING METHODS AND APPARATUSES, ELECTRONIC DEVICES,
PROGRAMS, AND STORAGE MEDIA", which are filed on March 6, 2018. The
10 disclosures of these applications are hereby incorporated by reference in their entirety.

TECHNICAL FIELD

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

BACKGROUND

20 Target tracking constantly refers to that, under the condition that position
information of a target object in a first frame of a video sequence is given, predicting
the position of the target object in a subsequent image frame of the video sequence.
25 Target tracking technologies are widely applied to application scenarios such as video
monitoring, unmanned driving, and human-computer interaction. Since the amount of
time consumed by the detection technology and the segmentation technology for a
single frame are relatively high, the target tracking technology is usually used
together with a detection/segmentation technology for increasing the video processing

speed and ensuring a better real-time performance. Therefore, the requirement on the speed of a target tracking algorithm for a particular task in practical engineering is high.

5

SUMMARY

Embodiments of the present disclosure provide a target tracking technology.

According to an aspect of the embodiments of the present disclosure, a target
10 tracking method is provided, including:

obtaining feature data of a reference frame of a first image frame, where the first image frame and at least one second image frame have a same reference frame; and

determining a position of a tracking target in the first image frame based on
15 the feature data of the reference frame.

According to one or more embodiments of the present disclosure, a first image frame is separated from a reference frame by at least one image frame.

According to one or more embodiments of the present disclosure, determining the position of a tracking target in the first image frame based on the feature data of
20 the reference frame includes:

performing feature extraction on the first image frame to obtain feature data of the first image frame; and

obtaining the position of the tracking target in the first image frame based on the feature data of the reference frame and the feature data of the first image frame.

25 According to one or more embodiments of the present disclosure, obtaining the position of the tracking target in the first image frame based on the feature data of the reference frame and the feature data of the first image frame includes:

connecting the feature data of the first image frame to the feature data of the reference frame to obtain a connection feature; and

obtaining the position of the tracking target in the first image frame based on the connection feature.

According to one or more embodiments of the present disclosure, the obtaining the position of the tracking target in the first image frame based on the
5 feature data of the reference frame includes:

obtaining, based on the feature data of the reference frame, relative position information of the tracking target in the first image frame relative to the tracking target in the reference frame; and

obtaining the position of the tracking target in the first image frame based on
10 the relative position information and position information of the tracking target in the reference frame.

According to one or more embodiments of the present disclosure, the obtaining feature data of a reference frame of a first image frame includes:

obtaining the feature data of the reference frame of the first image frame from
15 a memory.

According to one or more embodiments of the present disclosure, the obtaining feature data of a reference frame of a first image frame further includes:

determining, based on the position of the tracking target in the first image frame and the position of the tracking target in the reference frame, whether to use the
20 first image frame as an updated reference frame; and/or

using the first image frame as an updated reference frame if an interval between the first image frame and the reference frame meets a preset condition.

According to one or more embodiments of the present disclosure, determining, based on the position of the tracking target in the first image frame and the position of
25 the tracking target in the reference frame, whether to use the first image frame as an updated reference frame includes:

determining, based on an intersection over union between a first bounding box of the tracking target in the first image frame and a second bounding box of the tracking target in the reference frame, whether to use the first image frame as an

updated reference frame.

According to one or more embodiments of the present disclosure, the determining, based on an intersection over union between a first bounding box of the tracking target in the first image frame and a second bounding box of the tracking
5 target in the reference frame, whether to use the first image frame as an updated reference frame includes:

in response to the intersection over union between the first bounding box and the second bounding box being less than or equal to a preset threshold, determining to use the first image frame as an updated reference frame.

10 According to one or more embodiments of the present disclosure, the determining whether to use the first image frame as an updated reference frame based on an intersection over union between a first bounding box of the tracking target in the first image frame and a second bounding box of the tracking target in the reference frame further includes:

15 storing feature data of the first image frame in response to determining to use the first image frame as an updated reference frame.

According to another aspect of the embodiments of the present disclosure, a target tracking apparatus is provided, including:

a feature obtaining unit, configured to obtain feature data of a reference frame
20 of a first image frame, where the first image frame and at least one second image frame have a same reference frame; and

a position determining unit, configured to determine a position of a tracking target in the first image frame based on the feature data of the reference frame.

According to one or more embodiments of the present disclosure, the position
25 determining unit includes:

a feature extraction module, configured to perform feature extraction on the first image frame to obtain feature data of the first image frame; and

a position module, configured to obtain the position of the tracking target in the first image frame based on the feature data of the reference frame and the feature

data of the first image frame.

According to one or more embodiments of the present disclosure, the position module is configured to connect the feature data of the first image frame to the feature data of the reference frame to obtain a connection feature; and

- 5 obtain the position of the tracking target in the first image frame based on the connection feature.

According to one or more embodiments of the present disclosure, the position determining unit is configured to obtain, based on the feature data of the reference frame, relative position information of the tracking target in the first image frame
10 relative to the tracking target in the reference frame; and

obtain the position of the tracking target in the first image frame based on the relative position information and position information of the tracking target in the reference frame.

According to one or more embodiments of the present disclosure, the feature
15 obtaining unit is configured to obtain the feature data of the reference frame of the first image frame from a memory.

According to one or more embodiments of the present disclosure, the target tracking apparatus further includes:

a reference determining unit, configured to determine, based on the position of
20 the tracking target in the first image frame and the position of the tracking target in the reference frame, whether to use the first image frame as an updated reference frame; and/or

use the first image frame as an updated reference frame if an interval between the first image frame and the reference frame meets a preset condition.

25 According to one or more embodiments of the present disclosure, the reference determining unit is configured to determine, based on an intersection over union between a first bounding box of the tracking target in the first image frame and a second bounding box of the tracking target in the reference frame, whether to use the first image frame as an updated reference frame.

According to one or more embodiments of the present disclosure, the reference determining unit is configured to, in response to the intersection over union between the first bounding box and the second bounding box being less than or equal to a preset threshold, determine to use the first image frame as an updated reference
5 frame.

According to one or more embodiments of the present disclosure, the reference determining unit is further configured to store the feature data of the first image frame in response to determining to use the first image frame as an updated reference frame.

10 According to another aspect of the embodiments of the present disclosure, an electronic device is provided, including a processor, where the processor includes the foregoing target tracking apparatus.

According to another aspect of the embodiments of the present disclosure, an electronic device is provided, including: a memory, configured to store executable
15 instructions;

and a processor, configured to communicate with the memory to execute the executable instruction so as to implement the foregoing target tracking method.

According to another aspect of the embodiments of the present disclosure, a computer storage medium is provided, configured to store computer-readable
20 instructions that, when being executed, cause the processor to implement the foregoing target tracking method.

According to another aspect of the embodiments of the present disclosure, a computer program is provided, including computer readable codes that, when being executed on a device, cause a processor in the device to execute the target tracking
25 method.

According to yet another aspect of the embodiments of the present disclosure, a computer program product is provided, configured to store computer-readable instructions that, when being executed, cause a computer to execute the target tracking method according to any one of the foregoing possible implementations.

In one optional embodiment, the computer program product is specifically a computer storage medium. In another optional embodiment, the computer program product is specifically a software product, such as SDK.

5 The embodiments of the present disclosure further provide another target tracking method and a corresponding apparatus thereof, an electronic device, a computer storage medium, a computer program, and a computer program product. The method includes: obtaining feature data of a reference frame of a first image frame, where the first image frame and at least one second image frame correspond to the same reference frame; and determining a position of a tracking target in the first
10 image frame based on the feature data of the reference frame.

According to the target tracking method and apparatus, the electronic device, and the storage medium provided in the embodiments of the present disclosure, feature data of a reference frame of a first image frame is obtained, and the position of a tracking target in the first image frame is determined based on the feature data of the
15 reference frame. Since the reference frame corresponds to at least two image frames, reuse of the feature data is implemented, thereby reducing the number of times calculation is performed on the feature data, and increasing the tracking speed.

The following further describes in detail the technical solutions of the present disclosure with reference to the accompanying drawings and embodiments.
20

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings constituting a part of the specification describe
25 the embodiments of the present disclosure and are intended to explain the principles of the present disclosure together with the descriptions.

According to the following detailed descriptions, the present disclosure may be understood more clearly with reference to the accompanying drawings.

FIG. 1 is a flowchart for a target tracking method provided in embodiments of

the present disclosure.

FIG. 2 is a schematic structural diagram of a target tracking apparatus provided in embodiments of the present disclosure.

FIG. 3 is a schematic structural diagram of an electronic device suitable for
5 implementing a terminal device or a server in the embodiments of the present disclosure.

DETAILED DESCRIPTION

10

Various exemplary embodiments of the present disclosure are now described in detail with reference to the accompanying drawings. It should be noted that, unless otherwise stated specifically, relative arrangement of the components and steps, the numerical expressions, and the values set forth in the embodiments are not intended to
15 limit the scope of the present disclosure.

In addition, it should be understood that, for ease of description, the size of each part shown in the accompanying drawings is not drawn in actual proportion.

The following descriptions of at least one exemplary embodiment are merely illustrative actually, and are not intended to limit the present disclosure and the
20 applications or uses thereof.

Technologies, methods and devices known to a person of ordinary skill in the related art may not be discussed in detail, but such technologies, methods and devices should be considered as a part of the specification in appropriate situations.

It should be noted that similar reference numerals and letters in the following
25 accompanying drawings represent similar items. Therefore, once an item is defined in an accompanying drawing, the item does not need to be further discussed in the subsequent accompanying drawings.

The embodiments of the present disclosure may be applied to a computer system/server, which may operate with numerous other general-purpose or

special-purpose computing system environments or configurations. Examples of well-known computing systems, environments, and/or configurations suitable for use together with the computer system/server include, but are not limited to, Personal Computer (PC) systems, server computer systems, thin clients, thick clients, handheld or laptop devices, microprocessor-based systems, set top boxes, programmable consumer electronics, network PCs, small computer systems, large computer systems, distributed cloud computing environments that include any one of the foregoing systems, and the like.

The computer system/server may be described in the general context of computer system executable instructions (for example, program modules) executed by the computer system. Generally, the program modules may include routines, programs, target programs, components, logics, data structures, and the like for performing specific tasks or implementing specific abstract data types. Generally, the program modules may include routines, programs, target programs, components, logics, data structures, and the like for performing specific tasks or implementing specific abstract data types. In the distributed computing environments, the program modules may be located in local or remote computing system storage media including storage devices.

FIG. 1 is a flowchart for a target tracking method provided in embodiments of the present disclosure. The method may be performed by any electronic device, such as a terminal device, a server, or a mobile device.

At 110, feature data of a reference frame of a first image frame is obtained.

In an optional example, the operation 110 may be executed by a processor by invoking a corresponding instruction stored in a memory, or may be executed by a feature obtaining unit 21.

In order to track a tracking target, in some embodiments, multiple image frames of the tracking target are collected by means of video and the like, and the position of the tracking target in these image frames is determined, so that the trajectory of the tracking target in these target images by time may be determined by combining the time for obtaining the image frames, thereby realizing tracking of the

tracking target.

In some embodiments, in order to track a tracking target, it is required to know the position of the tracking target in one image frame, and the tracking target currently required to be tracked is determined based on the position; or it is known that the tracking target is a certain particular target (for example, a face or human body), multiple image frames are collected for the determined tracking target.

According to one or more embodiments of the present disclosure, according to the definition of a target tracking issue, tracking of a tracking target may be realized based on the determined position of the tracking target in one or more images. The tracking target in the embodiments of the present disclosure may be a known or an unknown particular object, human or an object of other types, which is not defined in the embodiments of the present disclosure.

In the embodiments of the present disclosure, at least two image frames use a same reference frame, and in some embodiments, the at least two image frames are adjacent to each other. For example, in the embodiments of the present disclosure, the reference frame is updated once at every M frames, and adjacent (M-1) image frames correspond to a same reference frame, where M is a preset value greater than or equal to 2, and the preset value may be determined according to actual application requirements. However, the value of M and the implementations of the multiple image frames using a same reference frame are not defined in the embodiments of the present disclosure.

In the embodiments of the present disclosure, a certain image frame may be adjacent to or not adjacent to a reference frame thereof. For example, a first image frame is separated from a reference frame thereof by one or more image frames. For example, a first image frame is separated from the reference frame thereof by N image frames, where N is a value greater than or equal to 1. The value of N is not defined in the embodiments of the present disclosure.

In this way, a first image frame and other image frames (for example, one or more second image frames) use the same reference frame, where a second image

frame may be located in front of or behind the first image frame, and compared with other target tracking approaches in which different reference frames are used for each frame of target image, the processing efficiency of the target tracking may be improved.

5 According to one or more embodiments of the present disclosure, the reference frame of the first image frame may include one or more image frames, where the multiple image frames may be adjacent to or not adjacent to each other in the video stream. For example, the multiple reference frames may be K adjacent reference frames in the video stream; or, a part of adjacent reference frames
10 in the multiple reference frames are adjacent to each other in the video stream, and the other part of adjacent reference frames in the multiple reference frames are not adjacent to each other in a video; or, any two adjacent reference frames in the multiple reference frames are not adjacent to each other in the video stream. In some embodiments, multiple reference frames may be separated by one or more images. In
15 some embodiments, any two adjacent image frames in the multiple image frames are separated by a same number of image frames, for example, M image frames, where M is a preset value greater than or equal to 2. The value of M may be determined according to application scenarios and requirements, and is not defined in the embodiments of the present disclosure. Or, different adjacent image frames in the
20 multiple image frames may also be separated by different number of image frames, which is not defined in the embodiments of the present disclosure.

 It should be understood that, the term "adjacent images or adjacent reference images in multiple reference images" in this document refer to reference images that are adjacent to each other after the multiple reference images are sorted by a certain
25 sorting rule, where the sorting rule may be sorting by the collection time, the detection time, etc.; and the term "adjacent images in a video stream" in this document may refer to images that are adjacent to each other after multiple images in a video stream or a video stream clip are sorted, where the sorting may refer to the original order in the video stream or a certain particular sorting rule. The specific implementation of

the sorting is not defined in the embodiments of the present disclosure.

In the embodiments of the present disclosure, feature data of a reference frame may be obtained by multiple manners. In some embodiments, feature data of a reference frame may be obtained from other apparatuses. For example, a server
5 receives feature data of a reference frame sent by a terminal device. For another example, an electronic device obtains feature data of a reference frame from a database, where the database may be provided in the local electronic device or other devices. In other embodiments, feature extraction may be performed on a reference frame to obtain feature data thereof.

10 In one or more embodiments, an electronic device may obtain the stored feature data of a reference frame. For example, feature data of a reference frame is obtained from a memory. In some embodiments, feature data of a reference frame may be stored in a buffer in the process of processing the reference frame (for example, determining position information of a tracking target in the reference frame),
15 and is then obtained from the buffer, but the embodiments of the present disclosure are not limited hereto.

In the embodiments of the present disclosure, a target image may be adjacent to multiple reference images thereof. Or a target image may be separated from the reference images thereof by at least one image frame (which is called as a second
20 image). In some embodiments, a target image may be separated from any reference image in multiple reference images thereof by one or more images; or a target image may be adjacent to a part of reference images in multiple reference images, and is separated from the other part of reference images in multiple reference images by one or more images, which is not defined in the embodiments of the present disclosure.

25 Compared with the approach which use a previous image frame of a target image as a reference image, the technical solution provided in the embodiments of the present disclosure may improve the prediction precision due to the fact that the prediction on an earlier image is more accurate, and the error accumulation problem may be efficiently solved by using an earlier image as a reference image for tracking.

In the embodiments of the present disclosure, different images may have different reference images, or multiple images may use the same reference image. For example, a target image and at least one third image may have the same reference images. For example, multiple reference images of a target image are identical to those of a third image, or, a part of reference images in multiple reference images of a target image are also reference images of the third image, which is not defined in the embodiments of the present disclosure. In some embodiments, the at least one third image may be adjacent to the target image, but the embodiments of the present disclosure are not limited hereto. In this case, in some embodiments, features of a reference image may be stored in a buffer, and it is merely needed to invoke the stored feature of the reference image when predicting the position of a tracking target of an image, thereby increasing the prediction speed of the position.

In the embodiments of the present disclosure, the feature of a reference frame may be obtained by multiple manners. In some embodiments, the feature of a reference image may be obtained from other apparatuses. For example, a server receives a feature of a reference image sent by a terminal device. For another example, an electronic device obtains a feature of a reference image from a database, where the database may be provided in the local electronic device or other devices. In other embodiments, feature extraction may be performed on a reference image to obtain a feature thereof after the reference image is obtained.

In one or more optional embodiments, features of multiple reference images stored by a memory of an electronic device may be obtained. For example, features of multiple reference images of a target image are obtained from a second buffer. In some embodiments, a feature of a reference image may be stored in a second buffer, and obtained from same, where the second buffer may also be called as a location exemplar buffer. However, the name of the second buffer is not defined in the embodiments of the present disclosure.

In the embodiments of the present disclosure, the second buffer may be used for storing information of a reference image. In some embodiments, the second buffer

may store information of multiple images currently used as reference images, i.e., information of multiple reference images currently included in a reference image list. For example, the second buffer may store a feature of a reference image, or may further store position information of a tracking object in the reference image, which is
5 not defined in the embodiments of the present disclosure.

In a target tracking method implemented based on a depth recurrent network, the depth recurrent network consists of a reference image branch and a target image branch. The reference image branch may obtain valuable information of a location exemplar by using a reference image feature extractor, so as to guide the target image
10 branch to predict the position of a target object. In some embodiments, the policy of using a $(t-1)^{\text{th}}$ frame to predict a $(t)^{\text{th}}$ frame is generally used. However, in fact, when predicting the position of the target object in the $(t)^{\text{th}}$ frame, all the images in front of a $(t-2)^{\text{th}}$ frame has passed through the feature extractor of the reference image branch, and an effective location exemplar has been obtained. Therefore, the reuse of the
15 location exemplar may be implemented, thereby increasing the speed of a tracking system.

In order to increase the tracking speed, feature data obtained based on a reference frame is stored in a memory. The feature data may be obtained by performing feature extraction on the reference frame. An image frame for obtaining a
20 final position of a tracking target may also be stored in the memory as a reference frame so as to determine the position of the tracking target in subsequent image frames.

At 120, the position of a tracking target in the first image frame is determined based on the feature data of the reference frame.

25 In an optional example, the operation 120 may be executed by a processor by invoking a corresponding instruction stored in a memory, or may be executed by a position determining unit 22.

According to one or more embodiments of the present disclosure, the position of a tracking target may include coordinate information of the tracking target in a first

image frame, for example, coordinate information of a bounding box of the tracking target in the first image frame, etc., which is not defined in the embodiments of the present disclosure.

According to one or more embodiments of the present disclosure, a first image
5 frame is separated from a reference frame by at least one image frame.

In some optional embodiments, feature extraction may be performed on a first image frame to obtain feature data of the first image frame, and the position of a tracking target in the first image frame is determined based on the feature data of the first image frame and feature data of the reference frame.

10 According to one or more embodiments of the present disclosure, feature extraction may be performed on an image frame by means of a neural network. For example, a convolution operation is performed on a first image frame by means of a convolutional layer in a convolutional neural network, to obtain feature data of the first image frame, where the feature data may optionally be a feature map or in other
15 formats. The manner of extracting the feature data and the specific implementation formats of the obtained feature data are not limited in the embodiments of the present disclosure.

According to one or more embodiments of the present disclosure, feature data of a first image frame and that of a reference frame may be connected (or spliced) to
20 obtain a connection feature, and the position of a tracking target in the first image frame is obtained based on the connection feature or the splicing feature.

Connection (or splicing) here refers to overlaying of channels. For example, a first image frame outputs three feature maps via three channels, a reference frame outputs three feature maps via three channels, and six feature maps may be obtained
25 by means of the feature connection.

In the embodiments of the present disclosure, there may be multiple manners for obtaining the position of a tracking target based on the connection feature. In some optional embodiments, the position of a tracking target may be obtained by using the processing of a full-connection layer and/or other network layers, which is not defined

in the embodiments of the present disclosure.

According to one or more embodiments of the present disclosure, a connection feature may be processed to obtain coordinates corresponding to a tracking target in a first image frame.

5 The range corresponding to the obtained coordinates is the position of the tracking target in the first image frame.

 In some optional embodiments, relative position information of a tracking target in a reference frame relative to a tracking target in a first image frame may be obtained based on feature data of the reference frame, and the
10 position of the tracking target in the first image frame is obtained based on the relative position information and position information of the tracking target in the reference frame.

 According to one or more embodiments of the present disclosure, the relative difference between a first image frame and a reference frame may be determined
15 based on feature data of the reference frame and that of the first image frame, and relative position information of a tracking target in the first image frame and the reference frame may be determined based on the relative difference between the feature data. The relative position information herein may indicate the relative distance of the tracking target moving from the reference frame to the first image
20 frame. Since position information of the tracking target in the reference frame is known, the position of the tracking target in the first image frame may be determined based on the known position information of the reference frame as well as the relative position information.

 In other target tracking methods, the policy of using a $(t-1)^{\text{th}}$ frame to predict a
25 $(t)^{\text{th}}$ frame is generally used. In fact, when determining the position of the of target object in the $(t)^{\text{th}}$ frame, all the images in front of a $(t-2)^{\text{th}}$ frame has passed through the feature extractor of the reference image branch, and an effective location exemplar has been obtained. The speed of a tracking system may be increased by using the forward property of the depth recurrent network.

Since the object displacement between two adjacent frames is not large, the position of a $(t-2)^{\text{th}}$ frame, a $(t-3)^{\text{th}}$ frame, a $(t-K)^{\text{th}}$ frame, or a $(t)^{\text{th}}$ frame may be utilized. Therefore, if let the $(t-1)^{\text{th}}$ frame not to pass through the reference image branch, a large amount of time may be saved. Considering the actual industrial
5 production scenario, if updating a location exemplar in the reference image branch at every K frames is utilized for the collected video frames, nearly $2K/(K+1)$ times of increase on the speed may be realized. Such significant improvement may further improve the real-time performance in actual applications, and reduce the power consumption of a device.

10 In the embodiments of the present disclosure, multiple target image frames may use the same reference frame. In this way, when predicting the position of a tracking target in a certain target image, the stored feature data of a reference frame may be invoked, thereby increasing the prediction speed of the position. Moreover, the policy of not using a previous image frame as a reference frame may improve the
15 precision due to the fact that the precision of the position of a tracking target in a reference frame obtained by detection in the embodiments of the present disclosure is higher, and the prediction on the position of a tracking target in an earlier image frame is more accurate. The error accumulation problem may be efficiently solved by using an earlier image frame as a reference frame for tracking.

20 According to the target tracking method provided in the embodiments of the present disclosure, feature data of a reference frame of a first image frame is obtained, and the position of a tracking target in the first image frame is determined based on the feature data of the reference frame. Since the first image frame and at least one second image frame have a same reference frame, reuse of the feature data is
25 implemented, thereby reducing the number of times calculation is performed on the feature data, and increasing the tracking speed.

In some optional examples, the operation 120 may include: determining multiple initial predicted positions of a tracking target in a first image frame based on feature data of a reference frame; and

determining a final position of the tracking target in the first image frame based on the multiple initial predicted positions.

According to one or more embodiments of the present disclosure, the initial predicted position and/or the final position may refer to the position of a bounding
5 box corresponding to the tracking target, for example, the position of a box of a face, of a human body or of a particular object or part. The specific form of the position of the tracking target is not defined in the embodiments of the present disclosure.

In some embodiments, average processing is performed on multiple initial predicted positions, such as, for example, mathematical average or weighted average
10 and the like, to obtain a final position of a tracking target in a first image frame. In other embodiments, at least one initial predicted position may be selected from multiple initial predicted positions, and a final position of a tracking target in a first image frame is obtained based on the at least one initial predicted position.

In one or more optional embodiments, a final position of a tracking target in a
15 first image frame is obtained based on the confidence of multiple initial predicted positions.

According to one or more embodiments of the present disclosure, at least one initial predicted position is selected from multiple initial predicted positions based on the confidence of the multiple initial predicted positions, and a final position is
20 determined based on the at least one initial predicted position. For example, at least one initial predicted position having a higher confidence is selected from multiple initial predicted positions, where the at least one initial predicted position may be a preset number of positions having the highest confidence in the multiple initial predicted positions, and may also be the position having the confidence greater than a
25 second preset threshold in the multiple initial predicted positions, which are not defined in the embodiments of the present disclosure.

The confidence refers to the probability that the predicted position corresponds to the actual position of the tracking target in the first image frame. In some embodiments, the confidence is obtained by a scoring network. The scoring network

is a trained neural network. For example, a network for binary classification is used, when the predicted position overlaps with the actual position, a true value of 1 and a predicted value of 0 are output, and when the predicted position does not overlap with the actual position, a real value of 0 and a predicted value of 1 are output, etc. The specific implementation of the scoring network is not defined in the embodiments of the present disclosure.

In some embodiments, the average value of the selected at least one initial predicted position is used as a final position, where the average value may be an arithmetic average value, a geometric average value, or a weighted average value. For example, weighted average is performed on the selected at least one initial predicted position to obtain a final position of a tracking target in a first image frame. The weighted value of the initial predicted position is relevant to the confidence thereof. The higher the confidence, the larger the weighted value, and the lower the confidence, the smaller the weighted value. However, the embodiments of the present disclosure are not limited hereto.

The initial predicted position having a larger deviation in the obtained initial predicted positions may be deleted according to the confidence. Accordingly, the obtained final position is more approximate to the actual position of the tracking target in the target image, thereby realizing better tracking.

According to one or more embodiments of the present disclosure, determining the confidence of at least one initial predicted position in multiple initial predicted positions includes:

separately processing the multiple initial predicted positions by a confidence scoring network to obtain the confidence corresponding to the multiple initial predicted positions,

where the confidence scoring network is obtained through training. The trained confidence scoring network corresponding to the input initial predicted position may output a corresponding confidence. In some embodiments, an error may be obtained based on the predicted confidence and the actual confidence. The scoring

network is trained by a reverse gradient propagation method according to the error. The trained scoring network may better predict the probability that the predicted position is the actual position (the known target position). The scoring network trains a particular object required to be tracked. For example, if a face needs to be tracked,
5 the face is used to train the scoring network.

According to one or more embodiments of the present disclosure, before separately processing the multiple initial predicted positions by a confidence scoring network to obtain the confidence corresponding to the multiple initial predicted positions, further includes:

10 processing at least one initial predicted position of a sample image by the confidence scoring network to obtain a predicted confidence corresponding to the at least one initial predicted position;

determining the actual confidence of the at least one initial predicted position based on an annotated position of the sample image; and

15 training the confidence scoring network based on the predicted confidence and the actual confidence of the at least one initial predicted position.

In order to improve the accuracy of the confidence scoring network, the confidence scoring network needs to be trained. The training process is the same as that of a neural network, in which the confidence scoring network is trained based on
20 the known actual confidence of a sample image. In some embodiments, targeted trainings may be used for different tracking objects. For example, when a face needs to be tracked, the confidence scoring network is trained by a face image having the known confidence in the process of training the confidence scoring network, so that the obtained confidence score of the face image from the confidence scoring network
25 is more accurate.

In some embodiments, whether to use a first image frame as an updated reference frame is further determined.

According to one or more embodiments of the present disclosure, if it is determined not to update the reference frame, i.e., it is determined that a first image

frame is not used as an updated reference frame, the same reference frame is continually used to perform target tracking processing on the image frame after the first image frame. In some embodiments, if it is determined to update the reference frame, i.e., it is determined that a first image frame is used as an updated reference frame, the updated reference frame, i.e., a first image frame, is used to perform target tracking processing on the image frame behind the first image frame.

In one or more optional embodiments, whether to use a first image frame as an updated reference frame is determined based on the position of a tracking target in a first image frame and the position of a tracking target in a reference frame.

10 In other embodiments, whether to use a first image frame as an updated reference frame is determined based on an interval between the first image frame and the reference frame.

For example, the first image frame is used as an updated reference frame in response to the interval between the first image frame and the reference frame meeting a preset condition. In some embodiments, the preset condition may be reaching a preset value. For example, a reference frame is updated once at every M frames, but the embodiments of the present disclosure are not limited hereto.

In the actual target tracking, since the motion of a tracking target is complex, the speed thereof is generally not constant. A reference frame may be dynamically adjusted for a $(t+1)^{\text{th}}$ frame of image frame by comparing the position of a tracking target in a $(t)^{\text{th}}$ frame with the position thereof in a reference frame, thereby obtaining a more appropriate location exemplar feature.

According to one or more embodiments of the present disclosure, if a $(t)^{\text{th}}$ frame is used to update a reference frame, i.e., a $(t)^{\text{th}}$ frame is used as an updated reference frame, a $(t+1)^{\text{th}}$ frame of image and its reference frame are adjacent to each other. While if the $(t)^{\text{th}}$ frame is not used as a reference frame, the reference frame is not updated, and in this case, the $(t+1)^{\text{th}}$ frame of image and its reference frame are separated by one or more image frames. For example, an image frame and the reference frame thereof are separated by N image frames, where N is a preset value

greater than or equal to 2. The value of N may be determined according to application scenarios, and is not defined in the embodiments of the present disclosure.

According to one or more embodiments of the present disclosure, whether to use a first image frame as an updated reference frame is determined based on the distance between the position of a tracking target in a first image frame and the position of a tracking target in a reference frame. For example, it is determined that the first image frame is used as an updated reference frame in response to the distance reaching a particular value. For another example, it is determined that the first image frame is not used as an updated reference frame in response to the distance being less than a particular value, i.e., the reference frame is not updated. The distance between two positions may be determined by multiple approaches. For example, the distance between particular points such as a center point or a certain vertex of a bounding box of a tracking target in the first image frame, and a corresponding point of a bounding box of a tracking target in the reference frame as the distance between the position of the tracking target in the first image frame and the position of the tracking target in the reference frame, etc., which is not defined in the embodiments of the present disclosure.

According to one or more embodiments of the present disclosure, whether to use a first image frame as an updated reference frame is determined based on an intersection over union between a first bounding box of a tracking target in a first image frame and a second bounding box of a tracking target in a reference frame.

If the position of an image frame is predicted by always using feature data of the same reference frame, the position is inaccurate if the movement position of the tracking target is relatively large. Therefore, the feature data of the reference frame needs to be updated. When to update is determined optionally according to whether the intersection over union between the first bounding box of the tracking target and the second bounding box thereof in the reference frame is greater than a preset value, where the first bounding box is obtained based on at least one coordinate (for example, all the vertex coordinates) corresponding to the predicted position, while the

second bounding box is obtained based on at least one coordinate (for example, all the vertex coordinates) corresponding to the tracking target in the reference frame.

According to one or more embodiments of the present disclosure, it is determined that a first image frame is used as an updated reference frame in response to an intersection over union between a first bounding box and a second bounding box being less than or equal to a preset threshold.

According to one or more embodiments of the present disclosure, it is determined not to update the reference frame in response to an intersection over union between a first bounding box and a second bounding box reaching a preset threshold.

The intersection over union refers to the ratio of the intersection to the union of two areas, and in the embodiments of the present disclosure, refers to the ratio of the intersection between the first bounding box and the second bounding box to the union of the first bounding box and the second bounding box. The lower the intersection over union, the larger the distance where a tracking target moves from a reference frame to a first image frame. In this case, the predicted position is inaccurate if the current reference frame is still used as the reference frame. Therefore, when the intersection over union is less than a set value, the first image frame is used as a reference frame.

According to one or more embodiments of the present disclosure, feature data corresponding to a reference frame is obtained by performing feature extraction on a first image frame to obtain the feature data of the updated reference frame; or a position coordination of a tracking target in a first image frame is determined, and is performed feature extraction to obtain the feature data of the updated reference frame.

If the absolute distance between the predicted position and the position in the reference frame is relatively close (the intersection over union between a first bounding box and a second bounding box is greater than or equal to a set threshold), the reference frame does not need to be updated, and the currently buffered feature data of the reference frame may be continually used for guiding the next image frame. If the absolute distance between the predicted position and the position in the

reference frame is relatively far (the intersection over union between a first bounding box and a second bounding box is less than a set threshold), the position extraction feature of the first image frame may be used to update the feature data of the reference frame. By means of the dynamic adjustment mechanism, the previously calculated
5 feature data (e.g. an intermediate feature map) may be fully utilized so as to reduce a large variety of unnecessary feature extraction processes.

For example, if a target tracking method is used to track an object having the absolute position stay unchanged basically in a video clip, a first image may always be used as a reference frame, which significantly increases the position speed of a
10 tracking target in each video frame, thereby improving the tracking efficiency.

For a particular problem (tracking a particular target, such as, for example, face tracking), a targeted reference frame update policy may be used in an application to adjust the reference frame in the target tracking method, which has a strong promotion value.

15 In one or more optional embodiments, the following procedure is further included: storing feature data of a first image frame in response to determining to use the first image frame as an updated reference frame.

The generated feature data may be generated by means of buffer in the embodiments of the present disclosure. In order to save space, the feature data of the
20 updated reference frame may be stored each time by replacing the feature data of the previous reference frame. When the position is required to be predicted, only the currently stored feature data needs to be invoked from a buffer. In some embodiments, the feature data of at least one reference frame may also be saved by the update time, and at least one feature data (for example, all includes the firstly generated feature
25 data and the updated feature data) is stored. In this way, when the position is required to be predicted, the latest stored feature data is invoked as the feature data of the reference frame.

When a user uses a cell phone or a camera to detect a face and a common object, the common approach is to combine with a target tracking algorithm to

increase the speed and reduce the power consumption of a device. According to the embodiments of the present disclosure, a feature buffer and feature reuse mechanism is used for the target tracking technology of a depth recurrent network, which may significantly increase the speed of a traditional recurrent tracker and implement the
5 high-rate tracking at 1,500 fps on a particular task (face tracking). This is critical in the actual product application, and significantly increases the forward speed of the target tracking algorithm.

The core of the embodiments of the present disclosure is to propose a feature buffer-based architecture and implement a practical high-precision and high-speed
10 target tracking method. According to the present disclosure, a mechanism of maintaining and reusing an intermediate feature map of a recurrent network by means of feature buffer is proposed by analyzing the structure of the recurrent network. The embodiments of the present disclosure propose that the intermediate feature obtained through calculation is stored by a location exemplar buffer (a location exemplar
15 obtained by a reference frame branch through calculation), and the previously buffered location exemplar is reused in the subsequent tracking process, thereby quickly tracking an object by using multiple reference frame exemplars. The method proposed in the embodiments of the present disclosure may be applied to multiple types of target tracking scenarios, increases the running speed of a depth recurrent
20 network-based target tracking algorithm as nearly twice as the original speed, significantly improves the real-time tracking performance, and reduces the power consumption of a device.

The target tracking algorithm proposed in the embodiments of the present application increases the speed of the target tracking system of the traditional depth
25 recurrent network as nearly once while guaranteeing the precision. The feature reuse-based recurrent architecture proposed in the present disclosure significantly improves the precision-speed equilibrium curve of a tracker for a particular task (face tracking). On one hand, the speed may be increased while maintaining the precision. On the other hand, the network expression capability may also be improved without

changing the speed, and the precision of a target tracker is improved.

The experiment result of face tracking on a long-sequence video (for example, 7,289 frames) is as follows: Table 1 is a comparison table of indicator values obtained by applying different tracking methods in the experiment to track. As shown in Table 1, the policy of dynamically updating the location exemplar feature of a reference frame is used in the experiment. It may be seen that the speed is increased as nearly once while maintaining the precision unchanged. In fact, the policy of not determining to use the previous frame as a reference frame may slightly improve the precision due to the fact that the precision of the detected frame is high, and the earlier frame is more accurate. The error accumulation problem may be efficiently solved by using an earlier frame as a reference frame for tracking.

Table 1 Comparison table of indicator values for tracking by different tracking methods

Indicator Tracking method	Speed (ms)	Robustness ([0.2:0.5])	Precision ([0.5:0.95])
16 Channel tracker	0.4015	0.771	0.568
16 Channel + feature reuse	0.2944	0.772	0.579
8 Channel tracker	0.2716	0.768	0.540
8 Channel + feature reuse	0.1808	0.768	0.545

A person of ordinary skill in the art may understand that: all or some operations for implementing the foregoing method embodiments are achieved by a program by instructing related hardware; the foregoing program may be stored in a computer readable storage medium; and when the program is executed, the operations including the foregoing method embodiments are executed. Moreover, the foregoing storage medium includes various media capable of storing program codes, such as a ROM, a RAM, a magnetic disk, or an optical disk.

FIG. 2 is a schematic structural diagram of a target tracking apparatus provided in embodiments of the present disclosure. The apparatus is configured to

implement the foregoing method embodiments of the present disclosure. As shown in FIG. 2, the apparatus includes:

a feature obtaining unit 21, configured to obtain feature data of a reference frame.

5 The reference frame corresponds to at least two image frames. In order to track a tracking target, in some embodiments, multiple image frames of the tracking target are collected by means of video and the like, and the position of the tracking target in these image frames is determined, so that the trajectory of the tracking target in these target images by time may be determined by combining the time for obtaining
10 the image frames, thereby realizing tracking of the tracking target.

 In some embodiments, in order to track a tracking target, it is required to know the position of the tracking target in one image frame, and the tracking target currently required to be tracked is determined based on the position; or it is known that the tracking target is a certain particular target (for example, a face or human body),
15 multiple image frames are collected for the determined tracking target.

 According to one or more embodiments of the present disclosure, according to the definition of a target tracking issue, tracking of a tracking target may be realized based on the determined position of the tracking target in one or more images. The tracking target in the embodiments of the present disclosure may be a known or an
20 unknown particular object, human or an object of other types, which is not defined in the embodiments of the present disclosure.

 In the embodiments of the present disclosure, at least two image frames use a same reference frame, and in some embodiments, the at least two image frames are adjacent to each other. For example, in the embodiments of the present disclosure, the
25 reference frame is updated once at every M frames, and adjacent (M-1) image frames correspond to a same reference frame, where M is a preset value greater than or equal to 2, and the preset value may be determined according to actual application requirements. However, the value of M and the implementations of the multiple image frames using a same reference frame are not defined in the embodiments of the

present disclosure.

In one or more optional embodiments, a feature obtaining unit 21 may obtain the stored feature data of a reference frame. For example, feature data of a reference frame is obtained from a memory. In some embodiments, feature data of a reference frame may be stored in a buffer, and is then obtained from the buffer, but the embodiments of the present disclosure are not limited hereto.

A position determining unit 22 is configured to determine the position of a tracking target in a first image frame based on feature data of a reference frame.

According to one or more embodiments of the present disclosure, the first image frame is separated from the reference frame by at least one image frame.

In some optional embodiments, the position determining unit 22 may include:
a feature extraction module, configured to perform feature extraction on a first image frame to obtain feature data of the first image frame; and

a position module, configured to obtain the position of a tracking target in the first image frame based on the feature data of a reference frame and the feature data of the first image frame.

According to one or more embodiments of the present disclosure, feature extraction may be performed on an image frame by means of a neural network. For example, a convolution operation is performed on a first image frame by means of a convolutional layer in a convolutional neural network, to obtain feature data of the first image frame, where the feature data may optionally be a feature map or in other formats. The manner of extracting the feature data and the specific implementation formats of the obtained feature data are not limited in the embodiments of the present disclosure.

According to one or more embodiments of the present disclosure, the position module may be configured to connect (or splice) feature data of a first image frame and that of a reference frame to obtain a connection feature or a splicing feature, and the position of a tracking target in the first image frame is obtained based on the connection feature or the splicing feature.

In some optional embodiments, the position determining unit 22 is configured to obtain, based on feature data of the reference frame, relative position information of a tracking target in a first image frame relative to the tracking target in a reference frame; and

- 5 obtain the position of the tracking target in the first image frame based on the relative position information and position information of the tracking target in the reference frame.

According to one or more embodiments of the present disclosure, the relative difference between a first image frame and a reference frame may be determined
10 based on feature data of the reference frame and that of the first image frame, and relative position information of a tracking target in the first image frame and the reference frame may be determined based on the relative difference between the feature data. The relative position information herein may indicate the relative distance of the tracking target moving from the reference frame to the first image
15 frame. Since position information of the tracking target in the reference frame is known, the position of the tracking target in the first image frame may be determined based on the known position information of the reference frame as well as the relative position information.

According to the target tracking apparatus provided in the embodiments of the
20 present disclosure, feature data of a reference frame of a first image frame is obtained, and the position of a tracking target in the first image frame is determined based on the feature data of the reference frame. Since the first image frame and at least one second image frame have a same reference frame, reuse of the feature data is implemented, thereby reducing the number of times calculation is performed on the
25 feature data, and increasing the tracking speed.

In one or more optional embodiments, the apparatus further includes:

a reference determining unit, configured to determine whether to use a first image frame as an updated reference frame based on the position of a tracking target in the first image frame and the position of the tracking target in the reference frame.

According to one or more embodiments of the present disclosure, a first image frame is used as an updated reference frame if an interval between the first image frame and the reference frame meets a preset condition.

In the actual target tracking, since the motion of a tracking target is complex,
5 the speed thereof is generally not constant. A reference frame may be dynamically adjusted for a $(t)^{\text{th}}$ frame of image frame by comparing the absolute position obtained by means of a reference frame with the absolute position thereof, thereby obtaining a more appropriate location exemplar feature.

According to one or more embodiments of the present disclosure, if a
10 $(t)^{\text{th}}$ frame is used to update a reference frame, i.e., a $(t)^{\text{th}}$ frame is used as an updated reference frame, a $(t+1)^{\text{th}}$ frame of image and its reference frame are adjacent to each other. While if the $(t)^{\text{th}}$ frame is not used as a reference frame, the reference frame is not updated, and in this case, the $(t+1)^{\text{th}}$ frame of image and its reference frame are separated by one or more image frames. For
15 example, an image frame and the reference frame thereof are separated by N image frames, where N is a preset value greater than or equal to 2. The value of N may be determined according to application scenarios, and is not defined in the embodiments of the present disclosure.

According to one or more embodiments of the present disclosure, a reference
20 determining unit is configured to determine whether to use a first image frame as an updated reference frame based on an intersection over union between a first bounding box of a tracking target in a first image frame and a second bounding box of a tracking target in a reference frame.

According to one or more embodiments of the present disclosure, a reference
25 determining unit is configured to, in response to the intersection over union between the first bounding box and the second bounding box being less than or equal to a preset threshold, determine to use a first image frame as an updated reference frame.

According to one or more embodiments of the present disclosure, the reference determining unit is further configured to store feature data of a reference

frame in response to using the first image as the reference frame.

According to another aspect of the embodiments of the present disclosure, an electronic device is provided, including a processor, where the processor includes the target tracking apparatus according to any one of the foregoing embodiments of the present disclosure.

According to another aspect of the embodiments of the present disclosure, an electronic device is provided, including a memory, configured to store executable instructions;

and a processor, configured to communicate with the memory to execute an executable instruction to achieve any one of the foregoing embodiments of the target tracking method provided in the present disclosure.

According to another aspect of the embodiments of the present disclosure, a computer storage medium is provided, configured to store computer-readable instructions, where when the instructions are executed by a processor, the processor implements any one of the foregoing embodiments of the target tracking method provided in the present disclosure.

According to another aspect of the embodiments of the present disclosure, a computer program is provided, including a computer readable code, where when the computer readable code runs in a device, a processor in the device executes the target tracking method provided in the present disclosure.

According to yet another aspect of the embodiments of the present disclosure, a computer program product is provided, configured to store computer-readable instructions, where when the instructions are executed, the computer implements the target tracking method according to any one of the foregoing possible implementations.

In one or more optional embodiments, the embodiments of the present disclosure further provide a computer program product, configured to store computer-readable instructions, where when the instructions are executed, a computer execute the target tracking method in any one of the foregoing embodiments.

The computer program product is specifically implemented by means of hardware, software, or a combination thereof. In an optional example, the computer program product is specifically represented by a computer storage medium. In another example, the computer program product is represented by
5 a software product, such as Software Development Kit (SDK).

The embodiments of the present disclosure further provide another target tracking method and a corresponding apparatus thereof, an electronic device, a computer storage medium, a computer program, and a computer program product, where the method includes: obtaining feature data of a
10 reference frame of a first image frame, where the reference frame and the first image frame are separated by at least one second image frame; and determining a position of a tracking target in the first image frame based on the feature data of the reference frame.

In some embodiments, the target tracking indication is specifically an
15 invoking instruction. The first apparatus instructs, by means of invoking, the second apparatus to execute target tracking. Accordingly, in response to receiving the invoking instruction, the second apparatus executes the operations and/process in any one of the embodiments of the foregoing target tracking method.

20 It should be understood that the terms such as "first" and "second" in the embodiments of the present disclosure are only used for distinguishing, and shall not be understood as limitations on the embodiments of the present disclosure.

It should also be understood that, in the present disclosure, "multiple" may
25 refer to two or more, and "at least one" may refer to one, two or more.

It should also be understood that, for any component, data or structure mentioned in the present disclosure, if there is no explicit limitation or no opposite motivation is provided in context, it is generally understood that the number of the component, data or structure is one or more.

It should be further understood that the description of each embodiment of the present disclosure emphasizes differences between the embodiments, and the same or similar points therebetween may be used as reference and omitted for clarity.

The embodiments of the present disclosure further provide an electronic
5 device which, for example, is a mobile terminal, a Personal Computer (PC), a tablet computer, a server, and the like. Referring to FIG. 3 below, a schematic structural diagram of an electronic device 300 suitable for implementing a terminal device or a server in the embodiments of the present application is shown. As shown in FIG. 3, the computer system 300 includes one or more processors, a communication part, and
10 the like. The one or more processors are, for example, one or more Central Processing Units (CPUs) 301 and/or one or more Graphic Processing Units (GPUs) 313, and the processor may execute appropriate actions and processing according to executable instructions stored in a ROM 302 or executable instructions loaded from a storage section 308 to a RAM 303. The communication part 312 may include, but is be
15 limited to, a network card. The network card may include, but is not limited to, an Infiniband (IB) network card.

The processor may communicate with the ROM 302 and/or the RAM 330 to execute executable instructions, is connected to the communication part 312 by means of a bus 304, and communicates with other target devices by means of the
20 communication part 312, so as to complete corresponding operations of any of the methods provided by the embodiments of the present application, for example, obtaining feature data of a reference frame of a first image frame, where the first image frame and at least one second image frame have a same reference frame; and determining the position of a tracking target in the first image frame.

25 In addition, the RAM 303 may further store various programs and data required for operations of the apparatus. The CPU 301, the ROM 302, and the RAM 303 are connected to each other by means of the bus 304. In the presence of the RAM 303, the ROM 302 is an optional module. The RAM 303 stores executable instructions, or writes the executable instructions into the ROM 302 during

running, where the executable instructions cause the processor 301 to execute corresponding operations of the foregoing communication method. An Input/Output (I/O) interface 305 is also connected to the bus 304. The communication part 312 is integrated, or is configured to have multiple sub-modules (for example, multiple IB network cards) connected to the bus.

The following components are connected to the I/O interface 305: an input section 306 including a keyboard, a mouse and the like; an output section 307 including a Cathode-Ray Tube (CRT), a Liquid Crystal Display (LCD), a speaker and the like; the storage section 308 including a hard disk drive and the like; and a communication section 309 of a network interface card including an LAN card, a modem and the like. The communication section 309 performs communication processing via a network such as the Internet. A drive 310 is also connected to the I/O interface 305 according to requirements. A removable medium 311 such as a magnetic disk, an optical disk, a magneto-optical disk, a semiconductor memory or the like is mounted on the drive 310 according to requirements, so that a computer program read from the removable medium is installed on the storage section 308 according to needs.

It should be noted that the architecture shown in FIG. 3 is merely an optional implementation. During specific practice, the number and types of the components in FIG. 3 is selected, decreased, increased, or replaced according to actual requirements. Different functional components are separated or integrated or the like. For example, the GPU and the CPU are separated, or the GPU is integrated on the CPU, and the communication part is separated from or integrated on the CPU or the GPU or the like. These alternative implementations all fall within the scope of protection of this disclosure.

Particularly, a process described above with reference to a flowchart according to the embodiments of the present disclosure may be implemented as a computer software program. For example, the embodiments of the present disclosure include a

computer program product, including a computer program tangibly included on a machine readable medium; the computer program includes program codes for executing the method shown in the flowchart; the program codes may include corresponding instructions for executing operations of the method provided in the
5 embodiments of the present application, for example, obtaining feature data of a reference frame of a first image frame, where the first image frame and at least one second image frame have a same reference frame; and determining the position of a tracking target in the first image frame. In such embodiments, the computer program is downloaded and installed from the network through the communication section 309,
10 and/or is installed from the removable medium 311. The computer program, when being executed by the CPU 301, executes the foregoing functions defined in the methods of the present application.

The embodiments in the specification are all described in a progressive manner, for same or similar parts in the embodiments, refer to these embodiments,
15 and each embodiment focuses on a difference from other embodiments. The system embodiments correspond to the method embodiments substantially and therefore are only described briefly, and for the associated part, refer to the descriptions of the method embodiments.

The methods and apparatuses of the present disclosure are implemented in
20 many manners. For example, the methods and apparatuses of the present disclosure are implemented with software, hardware, firmware, or any combination of software, hardware, and firmware. Unless otherwise specially stated, the foregoing sequences of operations of the methods are merely for description, and are not intended to limit the operations of the methods of the present disclosure. In addition, in some embodiments,
25 the present disclosure may also be implemented as programs recorded in a recording medium. The programs include machine-readable instructions for implementing the methods according to the present disclosure. Therefore, the present disclosure further covers the recording medium storing the programs for executing the methods according to the present disclosure.

The descriptions of the present disclosure are provided for the purpose of examples and description, and are not intended to be exhaustive or limit the present disclosure to the disclosed form. Many modifications and changes are obvious to a person of ordinary skill in the art. The embodiments are selected and described to

5 better describe a principle and an actual application of the present application, and to make a person of ordinary skill in the art understand the present disclosure, so as to design various embodiments with various modifications applicable to particular use.

CLAIMS

1. A target tracking method, comprising:

obtaining feature data of a reference frame of a first image frame, wherein the
5 first image frame and at least one second image frame have a same reference frame;
and

determining a position of a tracking target in the first image frame based on
the feature data of the reference frame.

10 2. The method according to claim 1, wherein determining the position of the tracking
target in the first image frame based on the feature data of the reference frame
comprises:

performing feature extraction on the first image frame to obtain feature data of
the first image frame; and

15 obtaining the position of the tracking target in the first image frame based on
the feature data of the reference frame and the feature data of the first image frame.

3. The method according to claim 2, wherein obtaining the position of the tracking
target in the first image frame based on the feature data of the reference frame and the
20 feature data of the first image frame comprises:

connecting the feature data of the first image frame to the feature data of the
reference frame to obtain a connection feature; and

obtaining the position of the tracking target in the first image frame based on
the connection feature.

25

4. The method according to any one of claims 1 to 3, wherein obtaining the position
of the tracking target in the first image frame based on the feature data of the
reference frame comprises:

obtaining, based on the feature data of the reference frame, relative position

information of the tracking target in the first image frame relative to the tracking target in the reference frame; and

obtaining the position of the tracking target in the first image frame based on the relative position information and position information of the tracking target in the reference frame.

5. The method according to any one of claims 1 to 4, wherein obtaining the feature data of the reference frame comprises:

obtaining the feature data of the reference frame from a memory.

6. The method according to any one of claims 1 to 5, further comprising:

determining whether the first image frame is used as an updated reference frame.

7. The method according to claim 6, wherein determining whether the first image frame is used as the updated reference frame comprises:

determining, based on the position of the tracking target in the first image frame and the position of the tracking target in the reference frame, whether the first image frame is used as the updated reference frame; and/or

determining, based on an interval between the first image frame and the reference frame, whether the first image frame is used as the updated reference frame.

8. The method according to claim 7, wherein determining, based on the position of the tracking target in the first image frame and the position of the tracking target in the reference frame, whether the first image frame is used as the updated reference frame comprises:

determining, based on an intersection over union between a first bounding box of the tracking target in the first image frame and a second bounding box of the tracking target in the reference frame, whether the first image frame is used as the

updated reference frame.

9. The method according to claim 8, wherein determining, based on the intersection over union between the first bounding box of the tracking target in the first image frame and the second bounding box of the tracking target in the reference frame, whether to use the first image frame as the updated reference frame comprises:

in response to the intersection over union between the first bounding box and the second bounding box being less than or equal to a preset threshold, determining that the first image frame is used as the updated reference frame.

10. The method according to any one of claims 6 to 9, further comprising:

storing feature data of the first image frame in response to determining that the first image frame is as the updated reference frame.

11. A target tracking apparatus, comprising:

a feature obtaining unit, configured to obtain feature data of a reference frame of a first image frame, wherein the first image frame and at least one second image frame; and

a position determining unit, configured to determine a position of a tracking target in the first image frame based on the feature data of the reference frame.

12. The apparatus according to claim 11, wherein the position determining unit comprises:

a feature extraction module, configured to perform feature extraction on the first image frame to obtain feature data of the first image frame; and

a position module, configured to obtain the position of the tracking target in the first image frame based on the feature data of the reference frame and the feature data of the first image frame.

13. The apparatus according to claim 12, wherein the position module is configured to connect the feature data of the first image frame to the feature data of the reference frame to obtain a connection feature; and

5 obtain the position of the tracking target in the first image frame based on the connection feature.

14. The apparatus according to any one of claims 11 to 13, wherein the position determining unit is configured to obtain, based on the feature data of the reference frame, relative position information of the tracking target in the first image frame
10 relative to the tracking target in the reference frame, and

obtain the position of the tracking target in the first image frame based on the relative position information and position information of the tracking target in the reference frame.

15 15. The apparatus according to any one of claims 11 to 14, wherein the feature obtaining unit is configured to obtain the feature data of the reference frame of the first image frame from a memory.

16. The apparatus according to any one of claims 11 to 15, further comprising:

20 a reference determining unit, configured to determine, based on the position of the tracking target in the first image frame and the position of the tracking target in the reference frame, whether the first image frame is used as the updated reference frame; and/or

use the first image frame as an updated reference frame if an interval between
25 the first image frame and the reference frame meets a preset condition.

17. The apparatus according to claim 16, wherein the reference determining unit is configured to determine, based on an intersection over union between a first bounding box of the tracking target in the first image frame and a second bounding box of the

tracking target in the reference frame, whether the first image frame is used as the updated reference frame.

18. The apparatus according to claim 17, wherein the reference determining unit is
5 configured to, in response to the intersection over union between the first bounding box and the second bounding box being less than or equal to a preset threshold, determine to use the first image frame as the updated reference frame.

19. The apparatus according to claim 18, wherein the reference determining unit is
10 further configured to store the feature data of the first image frame in response to determining that the first image frame is used as the updated reference frame.

20. An electronic device, comprising a processor, wherein the processor comprises the target tracking apparatus according to any one of claims 11 to 19.

15 21. 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 so as to implement the target tracking method according to
20 any one of claims 1 to 10.

22. A computer storage medium, configured to store computer-readable instructions that, when being executed by a processor, cause the processor to execute the target tracking method according to any one of claims 1 to 10.

25 23. A computer program product, configured to store computer-readable instructions that, when being executed, cause a computer to execute the target tracking method according to any one of claims 1 to 10.