

METHOD AND APPARATUS FOR POSITIONING TARGET, AND ELECTRONIC DEVICE, STORAGE MEDIUM AND PROGRAM

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims priority to Chinese Patent Application No. 202110208854.9 filed on Feb. 24, 2021, applied by Shanghai Sensetime Lingang Intelligent Technology Co., Ltd., and entitled" METHOD AND APPARATUS FOR POSITIONING TARGET, AND ELECTRONIC DEVICE AND STORAGE MEDIUM", the content of which is hereby incorporated by reference in its entirety.

TECHNICAL FIELD

[0002] The present disclosure relates to the field of computer technologies, and in particular, to a method for positioning a target, an apparatus, an electronic device, a storage medium, and a program.

BACKGROUND

[0003] In the related art, a global positioning system based on computer vision is generally used to position a movable object. Based on the positioning of the object based on computer vision, the plane positioning of the object (i.e., assuming that the object is positioned at the same height on the ground) can be realized. In order to realize the spatial positioning of the object, it is necessary to use an additional device such as a high-performance infrared camera to assist the positioning.

SUMMARY

[0004] The embodiments of the present disclosure provide a method for positioning a target, an apparatus, an electronic device, a storage medium, and a program.

[0005] A method for positioning a target is provided in an embodiment of the present disclosure. The method is performed by an electronic device and includes: obtaining a first image of a target scene, where the first image is obtained by at least one image acquisition device; performing target detection on the first image, to determine a first image position of a target, which is in the target scene, in the first image; determining first spatial positions of a plurality of first reference points for the target in the target scene; determining second image positions of the plurality of first reference points in the first image according to the first spatial positions of the plurality of first reference points and a camera parameter of the at least one image acquisition device; and determining a second spatial position of the target in the target scene according to the first image position of the target, the second image positions of the plurality of first reference points, and the first spatial positions of the plurality of first reference points. In this way, the spatial positioning of the target can be realized without arranging any additional hardware device on the target, which saves costs and has high universality.

[0006] In some embodiments, the determining the second spatial position of the target in the target scene according to the first image position of the target, the second image positions of the plurality of first reference points, and the first spatial positions of the plurality of first reference points includes: for any one of the first reference points, determining a weight of the first reference point according to a similarity between the second image position of the first reference point and the first image position of the target; and determining the second spatial position of the target according to weights of the plurality of first reference points and the first spatial positions of the plurality of first reference points. In this way, the weight of the reference point can be determined according to the similarity between the image position of the reference point and the image position of the target, and then the spatial position of the target can be precisely determined based on the weights of the reference points and the spatial positions of the reference points, thereby realizing the spatial positioning of the target.

[0007] In some embodiments, the determining the second spatial position of the target according to the weights of the plurality of first reference points and the first spatial positions of the plurality of first reference points includes: determining a weighted average of the first spatial positions of the plurality of first reference points based on the weights of the plurality of first reference points; and determining the weighted average as the second spatial position of the target. In this way, in the process of real-time spatial positioning of the target, the determined second spatial position can be prevented from fluctuating greatly, and the stability and reliability of the spatial positioning are higher.

[0008] In some embodiments, the method further includes: determining an initial spatial position of the target on a ground of the target scene; adding random quantities to the initial spatial position, to generate third spatial positions of a plurality of initial reference points; and obtaining the first spatial positions of the plurality of first reference points according to the third spatial positions of the plurality of initial reference points. In this way, the initial spatial position of the target and the initial reference points for the target can be conveniently determined, so as to realize real-time tracking and positioning of the target.

[0009] In some embodiments, the determining the initial spatial position of the target in the target scene includes: obtaining a second image of the target scene, where the second image includes a reference object preset on the ground and the target on the ground, and an acquisition time of the second image is before an acquisition time of the first image; determining a homography matrix between an imaging plane of the at least one image acquisition device and the ground according to a fourth spatial position of the reference object on the ground and a third image position of the reference object in the second image; and determining the initial spatial position of the target according to the homography matrix and a fourth image position of the target on the second image. In this way, the homography matrix between the imaging plane and the ground can be determined by using the spatial position and the image position of the reference object, and the initial spatial position of the target on the ground can be effectively determined based on the homography matrix and the image position of the target on the second

image.

[0010] In some embodiments, the method further includes: selecting a second reference point whose weight exceeds a preset threshold from the plurality of first reference points according to weights of the plurality of first reference points; adding random quantities to a fifth spatial position of the second reference point, to generate sixth spatial positions of a plurality of third reference points; and performing spatial positioning on the target in a third image of the target scene according to the sixth spatial positions of the plurality of third reference points, to obtain the second spatial position of the target in the target scene, where an acquisition time of the third image is after an acquisition time of the first image. In this way, based on the currently generated reference points, the reference point for spatially positioning the target in the image acquired at the next moment can be generated, so that the target in the target scene can be tracked and positioned in real time.

[0011] In some embodiments, for any one of the first reference points, the determining the weight of the first reference point according to the similarity between the second image position of any one of the first reference points and the first image position of the target includes: for any one of the first reference points, determining the weight of the first reference point according to a similarity between a first position vector formed by a plurality of second image positions of the first reference point and a second position vector formed by a plurality of first image positions of the target, where the weight is positively correlated with the similarity. In this way, the similarity between the plurality of second image positions and the plurality of first image positions can be easily calculated.

[0012] According to an embodiment of the present disclosure, an apparatus for positioning a target is provided, including: an obtaining module, configured to obtain a first image of a target scene, where the first image is obtained by at least one image acquisition device; a detection module, configured to perform target detection on the first image, to determine a first image position of a target, which is in the target scene, in

the first image; a first spatial position determination module, configured to determine first spatial positions of a plurality of first reference points for the target in the target scene; an image position determination module, configured to determine second image positions of the plurality of first reference points in the first image according to the first spatial positions of the plurality of first reference points and a camera parameter of the at least one image acquisition device; and a second spatial position determination module, configured to determine a second spatial position of the target in the target scene according to the first image position of the target, the second image positions of the plurality of first reference points, and the first spatial positions of the plurality of first reference points.

[0013] In some embodiments, the second spatial position determination module includes: a weight determination sub-module, configured to, for any one of the first reference points, determine a weight of the first reference point according to a similarity between the second image position of the first reference point and the first image position of the target; and a spatial position determination sub-module, configured to determine the second spatial position of the target according to weights of the plurality of first reference points and the first spatial positions of the plurality of first reference points.

[0014] In some embodiments, the spatial position determination sub-module is configured to: determine a weighted average of the first spatial positions of the plurality of first reference points based on the weights of the plurality of first reference points; and determine the weighted average as the second spatial position of the target.

[0015] In some embodiments, the apparatus further includes: an initial spatial position determination module, configured to determine an initial spatial position of the target on a ground of the target scene; an initial reference point determination module, configured to add random quantities to the initial spatial position, to generate third spatial positions of a plurality of initial reference points; and a first spatial position determination module, configured to obtain the first spatial positions of the plurality of first reference points

according to the third spatial positions of the plurality of initial reference points.

[0016] In some embodiments, the initial spatial position determination module includes: an acquisition sub-module, configured to obtain a second image of the target scene, where the second image includes a reference object preset on the ground and the target on the ground, and an acquisition time of the second image is before an acquisition time of the first image; a homography matrix determination sub-module, configured to determine a homography matrix between an imaging plane of the at least one image acquisition device and the ground according to a fourth spatial position of the reference object on the ground and a third image position of the reference object in the second image; and an initial spatial position determination sub-module, configured to determine the initial spatial position of the target according to the homography matrix and a fourth image position of the target on the second image.

[0017] In some embodiments, the apparatus further includes: a reference point determination module, configured to: select a second reference point whose weight exceeds a preset threshold from the plurality of first reference points according to weights of the plurality of first reference points; add random quantities to a fifth spatial position of the second reference point, to generate sixth spatial positions of a plurality of third reference points; and perform spatial positioning on the target in a third image of the target scene according to the sixth spatial positions of the plurality of third reference points, to obtain the second spatial position of the target in the target scene, where an acquisition time of the third image is after an acquisition time of the first image.

[0018] In some embodiments, the weight determination sub-module is configured to: for any one of the first reference points, determine the weight of the first reference point according to a similarity between a first position vector formed by a plurality of second image positions of the first reference point and a second position vector formed by a plurality of first image positions of the target, where the weight is positively correlated with the similarity.

[0019] An embodiment of the present disclosure provides an electronic device, including: a processor; and a memory for storing instructions executable by the processor. The processor is configured to invoke instructions stored in the memory to perform the method for positioning the target described above.

[0020] An embodiment of the present disclosure provides a computer readable storage medium having stored thereon computer program instructions that, when executed by a processor, implement the method for positioning the target described above.

[0021] An embodiment of the present disclosure provides a computer program including computer readable codes that, when run in an electronic device, is executed by a processor of the electronic device to implement the method for positioning the target described above.

[0022] In the embodiments of the present disclosure, the second image positions of the plurality of first reference points can be obtained based on the estimated first spatial positions of the plurality of first reference points, and the second spatial position of the target in the target scene can be determined according to the second image positions of the plurality of first reference points in combination with the first image position where the target is actually positioned in the first image and the estimated first spatial positions, so that the spatial positioning of the target can be realized without arranging any additional hardware device on the target, which saves costs and has high universality.

[0023] It should be understood that the above general description and the following detailed description are illustrative and explanatory only and are not intended to limit the present disclosure. Other features and aspects of the present disclosure will become apparent from the following detailed description of exemplary embodiments with reference to the drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0024] In order to describe the technical solutions of the embodiments of the present disclosure more clearly, the drawings that need to be used in the embodiments will be briefly introduced below, the drawings here are incorporated into the description and constitute a part of the description. The drawings illustrate embodiments in accordance with the present disclosure and together with the description serve to illustrate the technical solutions of the present disclosure. It should be understood that the following drawings only illustrate certain embodiments of the present disclosure and should not be construed as a limitation of scope. Other related drawings may be obtained from these drawings without creative efforts by those skilled in the art.

[0025] FIG. 1 illustrates a flowchart of a method for positioning a target according to an embodiment of the present disclosure.

[0026] FIG. 2 is a schematic diagram of a system architecture for positioning a target according to an embodiment of the present disclosure.

[0027] FIG. 3 illustrates a block diagram of an apparatus for positioning a target according to an embodiment of the present disclosure.

[0028] FIG. 4 illustrates a block diagram of an electronic device according to an embodiment of the present disclosure.

[0029] FIG. 5 illustrates a block diagram of an electronic device according to an embodiment of the present disclosure.

DETAILED DESCRIPTION

[0030] Various exemplary embodiments, features, and aspects of the present disclosure will be described in detail below with reference to the drawings. The same reference numerals in the drawings denote elements having the same or similar

functions. Although various aspects of the embodiments are illustrated in the drawings, the drawings need not be drawn to scale unless specifically noted.

[0031] The word "exemplary" is used exclusively herein to mean "serving as an example, embodiment, or illustration". Any embodiment described herein as "exemplary" is not necessarily to be interpreted as preferred or advantageous over other embodiments.

[0032] As used herein, the term "and/or" represents only an association relationship describing associated objects and represents that three relationships may exist. For example, A and/or B may represent three conditions: i.e., independent existence of A, existence of both A and B and independent existence of B. In addition, the term "at least one" herein refers to any one or any combination of at least two of the plurality, e.g., including at least one of A, B, C, may refer to any one or more elements selected from the set of A, B, and C.

[0033] It should be understood that the terms "first", "second" and "third" in the claims, description and drawings of the present disclosure are used to distinguish different objects, rather than to describe a specific order. The terms "comprise" and "include" used in the description and claims of the present disclosure indicate the presence of the described features, integers, steps, operations, elements and/or components, but do not exclude one or more other features, integers, steps, operations, elements, components and/or the presence or addition of a collection thereof.

[0034] In addition, in order to better illustrate the present disclosure, numerous specific details are given in the specific implementations below. It should be understood by those skilled in the art that the present disclosure may likewise be implemented without certain specific details. In some examples, methods, means, elements, and circuits well known to those skilled in the art are not described in detail in order to highlight the subject matter of the present disclosure.

[0035] FIG. 1 illustrates a flowchart of a method for positioning a target according to

an embodiment of the present disclosure. As shown in FIG. 1, the method for positioning the target includes the following operation.

[0036] At S11, a first image of a target scene is obtained, where the first image is obtained by at least one image acquisition device.

[0037] At S12, target detection is performed on the first image, to determine a first image position of a target in the target scene, the target is in the first image.

[0038] At S13, first spatial positions of a plurality of first reference points for the target in the target scene are determined.

[0039] At S14, second image positions of the plurality of first reference points in the first image are determined according to the first spatial positions of the plurality of first reference points and a camera parameter of the at least one image acquisition device.

[0040] At S15, a second spatial position of the target in the target scene is determined according to the first image position of the target, the second image positions of the plurality of first reference points, and the first spatial positions of the plurality of first reference points.

[0041] In one possible implementation, the method for positioning the target 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, an in-vehicle device, a wearable device, or the like. The method may be implemented by a processor to invoke computer readable instructions stored in a memory. Alternatively, the method may be performed by a server.

[0042] FIG. 2 is a schematic diagram of a system architecture to which the method for positioning a target according to an embodiment of the present disclosure can be applied. As shown in FIG. 2, the system architecture includes an image acquisition device 201, a network 202, and a control terminal 203. To support an exemplary

application, the image acquisition device 201 and the control terminal 203 establish a communication connection through the network 202. The image acquisition device 201 reports the first image in the target scene to the control terminal 203 through the network 202. In response to the first image, the control terminal 203 performs the target detection on the first image to determine a first image position of a target, which is in the target scene, in the first image. Secondly, first spatial positions of a plurality of first reference points for the target in the target scene are determined. Second image positions of the plurality of first reference points in the first image are determined according to the first spatial positions of the plurality of first reference points and a camera parameter of the image acquisition device. Thirdly, a second spatial position of the target in the target scene is determined according to the first image position of the target, the second image positions of the plurality of first reference points, and the first spatial positions of the plurality of first reference points. Finally, the control terminal 203 uploads the second spatial position of the target to the network 202 and sends it to the image acquisition device 201 through the network 202. As such, the spatial positioning of the target can be realized without arranging any additional hardware device on the target, which saves costs and has high universality.

[0043] As an example, the control terminal 203 may include a visual processing device or a remote server with visual information processing capabilities. The network 202 may adopt a wired or wireless connection. When the control terminal 203 is a visual processing device, the image acquisition device 201 may be communicatively connected to the visual processing device through a wired connection. For example, data is communicated through a bus. When the control terminal 203 is a remote server, the image acquisition device 201 may exchange data with the remote server over a wireless network.

[0044] Alternatively, in some scenarios, the image capture device 201 may be a visual processing device with a video capture module, and may be a host with a camera. At this time, the display method in an augmented reality scene according to the embodiments of the present disclosure may be executed by the image acquisition device

201, and the above system architecture may not include the network 202 and the control terminal 203.

[0045] In one possible implementation, in operation S11, the target scene may be a geospatial area where the target is movable, for example, may be an area such as an open space, a square, a classroom, or the like. The target may include at least one of an unmanned aerial vehicle, an intelligent vehicle, or an intelligent robot. The embodiments of the present disclosure are not limited to specific types of objects.

[0046] The user may control the movement of the target in the target scene, and the target may also be autonomously moved based on visual positioning, which is not limited in the embodiments of the present disclosure. It should be understood that the spatial position of the target can be positioned, both by controlling the movement of the target and by autonomous movement of the target.

[0047] In one possible implementation, in operation S11, the image acquisition device (e.g., a camera) may be arranged in or around the target scene. One or more image acquisition devices may be used, and the plurality of image acquisition devices may acquire the images of different regions of the target scene. The arrangement mode, quantity and the like of the image acquisition device may be set according to actual requirements, which is not limited in the embodiments of the present disclosure. In this way, the spatial positioning of the target in the target scene can be realized.

[0048] In a possible implementation, the image acquisition device may be connected to the electronic device that performs the method for positioning a target through a wired connection or a wireless connection, and may be specifically set according to actual requirements, which is not limited in the embodiments of the present disclosure. As such, the image(s) acquired by one or more image acquisition devices may be sent to the electronic device to realize the spatial positioning of the target.

[0049] In one possible implementation, the first image acquired by the image acquisition device may be an image acquired by the image acquisition device in real

time, or an image frame in a video stream acquired by the image acquisition device in real time. It should be understood that there may be more than one image acquisition device, and the corresponding first image may be more than one, and the target may be more accurately positioned in space based on the plurality of first images acquired in real time.

[0050] In a possible implementation, in operation S12, the first image position of the target, which is in the target scene, in the first image may be determined by performing the target detection on the first image through the target detection network. The target detection network may be obtained by using any known neural network training, such as a convolutional neural network, a residual network, a feature pyramid network, or the like. The embodiments of the present disclosure are not limited to the network structure and training mode of the target detection network.

[0051] In a possible implementation, an area where the target is located in the first image (e.g., an area indicated by a rectangular frame) can be detected through the target detection network, and then a center point of the area where the target is located can be determined based on the position of the rectangular frame (e.g., four vertex coordinates), and the pixel coordinate of the center point in the image is used as the first image position of the target. Further, a plurality of key points on the target may be detected through the target detection network, and an average value of pixel coordinates of the plurality of key points may be used as the first image position of the target, and the plurality of key points may all indicate the center of gravity of the target.

[0052] That is, the first image position of the target may be the pixel coordinate of the center point or the center of gravity of the target in the image, that is, the two-dimensional coordinate position. In this way, the pixel coordinate of the center point or the center of gravity of the detected target in the image can be used as the first image position of the target, so as to determine the spatial position of the target.

[0053] It should be noted that determining the image position of the target by using

the center point or the center of gravity of the target is an implementation provided by the embodiments of the present disclosure. It should be understood by those skilled in the art that the present disclosure is not limited thereto. In practice, as long as it is a point on a target, it may indicate an image position of the target, for example, a point on the boundary of the target may be used, and the embodiments of the present disclosure is not limited thereto.

[0054] In a possible implementation, in operation S13, the plurality of first reference points for the target which may also be referred to as a plurality of particles may be used to indicate a position where the target may be located in the target scene. The first spatial positions of the plurality of first reference points may be estimated spatial positions where the target may be located.

[0055] The number of first reference points may be set according to actual requirements, for example, at least 10000 first reference points may be used, which is not limited in the embodiments of the present disclosure.

[0056] In one possible implementation, the second spatial position of the target in the target scene may be considered as the actual position of the target in the target scene. The actual position of the target in the target scene may be similar to the estimated spatial positions indicated by some of the plurality of first reference points. Based on this, the second spatial position of the target in the target scene may be obtained at least according to the estimated first spatial positions of the plurality of reference points.

[0057] It should be noted that the spatial position in the embodiments of the present disclosure may refer to a three-dimensional coordinate position in the target scene. It should be understood that a world coordinate system may be established for the target scene, and the spatial position of the target in the target scene may be a three-dimensional coordinate position of the target in the world coordinate system.

[0058] In one possible implementation, in operation S13, the first spatial positions of the plurality of first reference points may be the estimated spatial positions where the

target may be located. The first spatial positions of the plurality of first reference points may be obtained based on the initial spatial position of the target (e.g., the position of the target on the ground of the target scene) in combination with the particle filtering mode. A particle filtering method such as a Bayesian filtering algorithm, a Monte Carlo sampling algorithm, and an importance filtering algorithm may be used, which is not limited in the embodiments of the present disclosure.

[0059] In a possible implementation, in operation S14, the camera parameter of the image acquisition device may be obtained by any known camera calibration method, for example, by the Zhang Zhengyou camera calibration method. The camera parameter may include an external parameter of the camera and an internal parameter of the camera. To calibrate the camera parameter, the external parameter of the camera may be calibrated only, or the external parameter of the camera and the internal parameter of the camera may be calibrated. The camera calibration method of the image acquisition device is not limited in the embodiments of the present disclosure.

[0060] In a possible implementation, in operation S14, the determining of the second image positions of the plurality of first reference points according to the first spatial positions of the plurality of first reference points and the camera parameter may be implemented based on the imaging principle of the image acquisition device, such as the aperture imaging principle.

[0061] In a possible implementation, the imaging principle of the formula (1) may be used to implement the determination of the second image positions of the plurality of the first reference points according to the first spatial positions of the plurality of the first reference points and the camera parameter.

[0062]

$$z \begin{bmatrix} u & v & 1 \end{bmatrix}^T = K \begin{bmatrix} R & t \\ \vec{0} & 1 \end{bmatrix} \begin{bmatrix} X & Y & Z & 1 \end{bmatrix}^T \quad (1);$$

[0063] where z is a scale factor; (u, v) may denote the second image position, i.e.,

the pixel coordinate of the first reference point; (X,Y,Z) may denote the first spatial position, i.e., the world coordinate of the first reference point; K denotes an internal parameter matrix of the camera, $\begin{bmatrix} R & t \\ \tilde{0} & 1 \end{bmatrix}$ denotes an external parameter matrix of the camera, R denotes a rotation matrix, t denotes a translation matrix, and $[\cdot]^T$ denotes a matrix transpose.

[0064] Where z can be obtained by the formula $\begin{bmatrix} x & y & z & 1 \end{bmatrix}^T = \begin{bmatrix} R & t \\ \tilde{0} & 1 \end{bmatrix} \begin{bmatrix} X & Y & Z & 1 \end{bmatrix}^T$;
 (x,y,z) denotes the coordinate of the first reference point in the camera coordinate system of the image acquisition device.

[0065] In a possible implementation, in operation S15, the second spatial position of the target may be determined according to the similarities or the distances between the first image position of the target and the second image positions of the plurality of first reference points, and the first spatial positions of the plurality of reference points.

[0066] In one possible implementation, the higher the similarity or the distance between the second image position of a first reference point and the first image position of the target, the closer the first reference point to the target. Based on this, the weighted average value of the first spatial positions of the first reference point may be determined according to the similarities or the distances, and the weighted average value is further used as the second spatial position of the target. Also, reference points each having a similarity or a distance exceeding a certain threshold may also be selected from the plurality of first reference points, and an average value of spatial positions of the reference points each having the similarity or the distance exceeding the certain threshold may be used as the second spatial position of the target.

[0067] It should be understood that the above methods of similarity, distance, calculation of the weighted average value, and calculation of the average value are

specific implementations disclosed in the embodiments of the present disclosure. It should be understood by those skilled in the art that the present disclosure is not limited thereto. In practice, as long as the second spatial position of the target can be determined according to the first image position of the target, the second image positions of the plurality of first reference points, and the first spatial positions of the plurality of first reference points, it falls within the scope of the present disclosure.

[0068] According to the embodiments of the present disclosure, the second image positions of the plurality of first reference points can be obtained based on the estimated first spatial positions of the plurality of first reference points, and then the second spatial position of the target in the target scene can be determined according to the second image positions of the plurality of first reference points in combination with the first image position where the target is actually located in the first image and the estimated first spatial positions, so that the spatial positioning of the target can be realized without arranging any additional hardware device on the target, which saves costs and has high universality.

[0069] In one possible implementation, in operation S15, the determination of the second spatial position of the target in the target scene according to the first image position of the target, the second image positions of the plurality of first reference points, and the first spatial positions of the plurality of first reference points may include the following operations.

[0070] For any one of the first reference points, a weight of the first reference point is determined according to a similarity between the second image position of the first reference point and the first image position of the target.

[0071] The second spatial position of the target is determined according to weights of the plurality of first reference points and the first spatial positions of the plurality of first reference points.

[0072] Where each first reference point may correspond to one or more second image

positions. It should be understood that the number of second image positions corresponding to the first reference point is consistent with the number of image acquisition devices.

[0073] In one possible implementation, the similarity between the second image position of the first reference point and the first image position of the target may be calculated by using any known similarity calculation method, such as cosine similarity, Euclidean distance, Pearson correlation coefficient, and the like, which is not limited in the embodiments of the present disclosure.

[0074] In one possible implementation, the weight of the first reference point may be set to be positively correlated with the similarity. That is, the higher the similarity, the higher the weight assigned to the first reference point. The specific value of the weight assigned may be set according to an actual requirement. For example, the weight sum of a plurality of first reference points may be defined as 1, or the weight sum may be defined as 10, 100, or the like, so that the weight of the first reference point is assigned according to the definition, which is not limited in the embodiments of the present disclosure.

[0075] The image acquisition devices may be multiple. Accordingly, for any one of the first reference points, the second image positions of the first reference point may be multiple and the first image positions of the target are multiple. In one possible implementation, for any one of the first reference points, the determination of the weight of the first reference point according to the similarity between the second image position of the first reference point and the first image position of the target may include the following operation.

[0076] For any one of the first reference points, the weight of the first reference point is determined according to a similarity between a first position vector formed by a plurality of second image positions of the first reference point and a second position vector formed by a plurality of first image positions of the target. In this way, the

calculation of the similarity between the plurality of second image positions and the plurality of first image positions can be facilitated.

[0077] In one example, based on n first images acquired by n image acquisition devices, for any one of the first reference points, n second image positions of the first reference point in n first images may be obtained by operation S14. The first position vector formed by the n second image positions may be denoted as $P_L = [u_{L1} \ v_{L1} \ u_{L2} \ v_{L2} \ \cdots \ u_{Ln} \ v_{Ln}]$, where L denotes the L th first reference point in the plurality of first reference points, and (u_{Ln}, v_{Ln}) denotes the n th second image position of the L th first reference point.

[0078] The second position vector formed by n first image positions of the target in n first images may be denoted as $P_0 = [u_{01} \ v_{01} \ u_{02} \ v_{02} \ \cdots \ u_{0n} \ v_{0n}]$, where (u_{0n}, v_{0n}) denotes the n th first image position of the target in the n first images; the cosine similarity between P_L and P_0 may be denoted as $\frac{P_L \cdot P_0}{\|P_L\| \|P_0\|}$.

[0079] In one possible implementation, the determination of the second spatial position of the target according to weights of the plurality of first reference points and the first spatial positions of the plurality of first reference points may include the following operation.

[0080] A weighted average of the first spatial positions of the plurality of first reference points is determined based on the weights of the plurality of first reference points. The sum of the weights of the plurality of first reference points is 1. The weighted average is determined as the second spatial position of the target. In this way, in the process of real-time spatial positioning of the target, the determined second spatial position can be prevented from fluctuating greatly, and the stability and reliability of the spatial positioning are higher.

[0081] In one example, the first spatial positions of m first reference points may be

denoted as $O_1, O_2, O_3, \dots, O_m$; the weights of m first reference points may be denoted as $a_1, a_2, a_3, \dots, a_m$; the weighted average value of the first spatial positions of m first reference points may be denoted as $\frac{\sum_{i=1}^m O_i a_i}{m}$. Since the sum of the weights of the m first reference points is defined as 1, the weighted average value can be directly determined as the second spatial position of the target. It should be understood that the m th first spatial position is $O_m = (X_m, Y_m, Z_m)$.

[0082] The imaging principle of the image acquisition device is a process of mapping and converting three-dimensional points in the three-dimensional space to two-dimensional points, and therefore the spatial position of the target cannot be obtained only by the image position of the target. According to the embodiments of the present disclosure, the image position of the reference point can be obtained from the spatial position of the reference point based on the imaging principle of the image acquisition device, and then the weights of the reference points can be determined according to the similarity between the image positions of the reference points and the image position of the target, and then the spatial position of the target can be precisely determined based on the weights of the reference points and the spatial positions of the reference points, thereby realizing the spatial positioning of the target.

[0083] The plurality of first reference points for the target may indicate a position where the target may be located in the target scene. The first spatial positions of the plurality of first reference points may be the estimated spatial position where the target may be located. A method for generating a plurality of first reference points will be described below.

[0084] It should be understood that, since the embodiments of the present disclosure may implement the spatial positioning of the target based on the image frames acquired by the image acquisition device in real time, the interval (i.e., the frame rate) between the image frames may be short. During the interval, the spatial position of the target

may change, but to a small extent. In this way, a plurality of initial reference points may be firstly generated according to the initial spatial position of the target. Further, when the target is tracked and positioned in real time, new reference points can be continuously generated based on the initial reference points by using the particle filter mode to realize the tracking and positioning of the target.

[0085] In one possible implementation, the method further includes the following operations.

[0086] An initial spatial position of the target in the target scene is determined. The initial spatial position includes a spatial position of the target on the ground of the target scene. Random quantities are added to the initial spatial position, to generate third spatial positions of a plurality of initial reference points. The first spatial positions of the plurality of first reference points are obtained according to the third spatial positions of the plurality of initial reference points. In this way, the initial spatial position of the target and the initial reference points for the target can be conveniently determined, so as to realize real-time tracking and positioning of the target.

[0087] In one possible implementation, the initial spatial position of the target may be understood as the first spatial position of the target in the target scene. The initial spatial position of the target can be easily determined by taking the spatial position of the target on the ground as the initial spatial position.

[0088] When determining the spatial position of the target, the world coordinate system can be established for the target scene. In order to obtain the initial spatial position of the target, the vertical coordinate (Z-axis) of the object on the ground in the target scene in the world coordinate system can be set to 0, that is, the initial spatial position of the target on the ground in the target scene can be denoted as $(X_c, Y_c, 0)$. In this way, when the horizontal and vertical coordinates (X_c, Y_c) of the target on the ground are obtained, the initial spatial position of the target on the ground can be obtained.

[0089] In one possible implementation, adding the random quantities to the initial spatial position may be generating a plurality of random quantities by using a known particle filter algorithm or the like, and adding the generated plurality of random quantities to the initial spatial position, so as to obtain third spatial positions of the plurality of initial reference points. It should be understood that the third spatial positions of the initial reference points are different, that is, the random quantities corresponding to the initial reference points may be different. The method of generating third spatial positions of a plurality of initial reference points is not limited in the embodiments of the present disclosure.

[0090] It should be understood that the first reference point may be an initial reference point, and may also be reference points generated based on a plurality of initial reference points. The number of first reference points may be the same as the number of initial reference points, to ensure the stability of continuously generating the reference points.

[0091] In one possible implementation, in a case that the first reference point is the initial reference point, obtaining the first spatial positions of the plurality of first reference points according to the third spatial positions of the plurality of initial reference points may include: determining the third spatial positions of the plurality of initial reference points as the first spatial positions of the plurality of first reference points.

[0092] In a possible implementation, in a case that the first reference point is not the initial reference point, obtaining the first spatial positions of the plurality of first reference points according to the third spatial positions of the plurality of initial reference points may include: selecting a fourth reference point whose weight exceeds a preset threshold from the plurality of initial reference points according to weights of the plurality of initial reference points; and adding random quantities to the sixth spatial position of the fourth reference point, to generate the first spatial positions of the plurality of first reference points.

[0093] The weights of the plurality of initial reference points may be determined in the same manner as the manner in which the weights of the first reference points are determined. The specific value of the preset threshold may be set according to actual requirements, which is not limited in the embodiments of the present disclosure.

[0094] By setting the preset threshold, the probability that the initial reference point is selected can be positively correlated with the weight of the initial reference point, so that the fourth reference point close to the initial spatial position can be selected, and the generated first spatial position of the first reference point can be as close to the current second spatial position as possible.

[0095] In one possible implementation, the determination of the initial spatial position of the target on the ground of the target scene may include the following operations.

[0096] A second image of the target scene is obtained, the second image includes a reference object preset on the ground and the target on the ground, and an acquisition time of the second image is before an acquisition time of the first image.

[0097] A homography matrix between an imaging plane of the at least one image acquisition device and the ground is determined according to a fourth spatial position of the reference object on the ground and a third image position of the reference object in the second image.

[0098] The initial spatial position of the target is determined according to the homography matrix and a fourth image position of the target on the second image.

[0099] In one possible implementation, the second image may be the image acquired for determining the initial spatial position of the target. As described above, the fourth image position of the target on the second image may be the pixel coordinate of the center point or center of gravity of the target in the second image.

[00100] In one possible implementation, the reference object may be an object pre-positioned on the ground, and the fourth spatial position of the reference object on

the ground may be known. For example, the user may preset the spatial position of the center of the reference object on the ground. The reference object may be an object of any form, such as a two-dimensional code map, a checkerboard map used to calibrate the camera of the image acquisition device, which is not limited in the embodiments of the present disclosure. The third image position of the reference object in the second image may be the pixel coordinate of the center of the reference object in the second image.

[00101] The vertical coordinate of the object, which is on the ground in the target scene, in the world coordinate system can be set to 0, the fourth spatial position of the reference object on the ground can be denoted as $(X_v, Y_v, 0)$, and then the homography matrix between the imaging plane of the image acquisition device and the ground can be determined according to the horizontal and vertical coordinates (X_v, Y_v) of the fourth spatial position of the reference object and the third image position (x_v, y_v) of the reference object in the second image.

[00102] The method for calculating the homography matrix between the imaging plane and the ground of the image acquisition device is not limited in the embodiments of the present disclosure.

[00103] In one possible implementation, the homography matrix H between the imaging plane of the image acquisition device and the ground can be determined

according to the equation
$$H = K \left(R - \frac{tB^T}{d} \right) K^{-1}$$
. The plane equation of the ground may be denoted as $B^T P + d = 0$. K denotes an internal parameter matrix of the camera, R denotes a rotation matrix, t denotes a translation matrix, B and d are parameters of the plane equation of the ground, B^T denotes the transposition of B , and P denotes the two-dimensional position of the object on the ground, that is, it can denote the horizontal and vertical coordinates of the fourth spatial position of the reference object.

[00104] In one possible implementation, the determined homography matrix may be

$$H = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix},$$

denoted as H , and according to the formula (2), the initial spatial position of the target may be determined according to the homography matrix and the fourth image position of the target on the second image. (X_c, Y_c) denotes the horizontal and vertical coordinates of the initial spatial position of the target, (x_c, y_c) denotes the fourth image position of the target, and when (x_c, y_c) of the target is determined, the initial spatial position $(X_c, Y_c, 0)$ of the target can be obtained.

[00105]

$$(X_c, Y_c) = \left(\frac{H_{11}x_c + H_{12}y_c + H_{13}}{H_{31}x_c + H_{32}y_c + H_{33}}, \frac{H_{21}x_c + H_{22}y_c + H_{23}}{H_{31}x_c + H_{32}y_c + H_{33}} \right) \quad (2);$$

[00106] In the embodiments of the present disclosure, the homography matrix between the imaging plane and the ground can be determined by using the spatial position and the image position of the reference object, and the initial spatial position of the target on the ground can be effectively determined based on the homography matrix and the image position of the target on the second image.

[00107] Considering that, in order to realize real-time tracking and positioning of the target, a plurality of third reference points may be generated based on the generated plurality of first reference points, so as to perform spatial positioning on the target in the image acquired later. In one possible implementation, the method further includes the following operations.

[00108] A second reference point whose weight exceeds a preset threshold is selected from the plurality of first reference points according to weights of the plurality of first reference points. Random quantities are added to a fifth spatial position of the second reference point, to generate sixth spatial positions of a plurality of third reference points. Spatial positioning is performed on the target in a third image of the target scene according to the sixth spatial positions of the plurality of third reference points, to obtain

the second spatial position of the target in the target scene, and an acquisition time of the third image is after an acquisition time of the first image.

[00109] The specific value of the preset threshold may be set according to actual requirements, which is not limited in the embodiments of the present disclosure. The weight assigned to each first reference point may be positively correlated with the similarity between the first reference point and the target. By setting the preset threshold, the probability that the first reference point is selected may be positively correlated with the weight of the first reference point, so that the second reference point close to the current second spatial position of the target may be selected and the generated sixth spatial positions of the plurality of third reference points are closer to the spatial position of the target at the next moment.

[00110] In one possible implementation, the number of third reference points may be the same as the number of first reference points to ensure the stability of continuously generating reference points.

[00111] In one possible implementation, the addition of the random quantities to the fifth spatial position of the second reference point may be performed in the same manner as the manner of the addition of the random quantities to the initial spatial position. For example, a particle filter algorithm may be used to generate the sixth spatial positions of the plurality of third reference points, which is not limited in the embodiments of the present disclosure.

[00112] It may be understood that the first spatial positions of the plurality of first reference points may be generated in the same manner as the manner in which the sixth spatial position of the third reference point is generated in the embodiments of the present disclosure, that is, the first spatial positions of the current plurality of first reference points may be generated based on the spatial positions of the previously generated reference points.

[00113] In one possible implementation, the target in the third image of the target

scene is spatially positioned through the sixth spatial positions of the plurality of third reference points to obtain the second spatial position of the target in the target scene, which may be realized by the method for positioning the target in the embodiments of the present disclosure.

[00114] According to the embodiments of the present disclosure, based on the currently generated reference points, the reference point for spatially positioning the target in the image acquired at the next moment can be generated, so that the target in the target scene can be tracked and positioned in real time.

[00115] According to the embodiments of the present disclosure, stereoscopic positioning (spatial positioning) of a target (such as a robot) in a target scene can be realized by using a plurality of image acquisition devices. In addition, it is not necessary to keep the target captured by each image acquisition device at all times, that is, it is not necessary to ensure that the target appears in the images captured by all cameras, and the stereoscopic positioning of the target can be realized as long as the target is captured by an image acquisition device.

[00116] According to the embodiments of the present disclosure, the robot in the image captured by the camera can be quickly detected by using the related target detection network, and then the key points of the robot can be detected in the plurality of images respectively. The stereoscopic position (spatial position) of the robot is then estimated by using the particle filtering method.

[00117] According to the embodiments of the present disclosure, stereoscopic positioning of a robot in a target scene is implemented by using computer vision without manual calibration of the device used in the method for positioning the target.

[00118] The method for positioning a target according to the embodiments of the present disclosure may be applied in the fields of AI education platforms, AI education robots, and the like.

[00119] It will be understood that the various method embodiments mentioned above in the present disclosure may be combined with each other to form a combined embodiment without violating the principle and logic. It will be understood by those skilled in the art that in the above-described method of the specific implementation, the specific execution order of each step should be determined in terms of its function and possible internal logic.

[00120] In addition, the present disclosure also provides an apparatus for positioning a target, an electronic device, a computer readable storage medium, and a program, all of which may be used to implement any method for positioning a target provided in the present disclosure, and the corresponding technical solutions and descriptions can be found in the corresponding records in the method section.

[00121] FIG. 3 illustrates a block diagram of an apparatus for positioning a target according to an embodiment of the present disclosure. As shown in FIG. 3, the apparatus includes an obtaining module 301, a detection module 302, a first spatial position determination module 303, an image position determination module 304 and a second spatial position determination module 305.

[00122] The obtaining module 301 is configured to obtain a first image of a target scene, where the first image is obtained by at least one image acquisition device.

[00123] The detection module 302 is configured to perform target detection on the first image, to determine a first image position of a target, which is in the target scene, in the first image.

[00124] The first spatial position determination module 303 is configured to determine first spatial positions of a plurality of first reference points for the target in the target scene.

[00125] The image position determination module 304 is configured to determine second image positions of the plurality of first reference points in the first image

according to the first spatial positions of the plurality of first reference points and a camera parameter of the at least one image acquisition device.

[00126] The second spatial position determination module 305 is configured to determine a second spatial position of the target in the target scene according to the first image position of the target, the second image positions of the plurality of first reference points, and the first spatial positions of the plurality of first reference points.

[00127] In one possible implementation, the second spatial position determination module 305 includes a weight determination sub-module and a spatial position determination sub-module. The weight determination sub-module is configured to, for any one of the first reference points, determine a weight of the first reference point according to a similarity between the second image position of the first reference point and the first image position of the target. The spatial position determination sub-module is configured to determine the second spatial position of the target according to weights of the plurality of first reference points and the first spatial positions of the plurality of first reference points.

[00128] In one possible implementation, the spatial position determination sub-module is configured to: determine a weighted average of the first spatial positions of the plurality of first reference points based on the weights of the plurality of first reference points; and determine the weighted average as the second spatial position of the target.

[00129] In a possible implementation, the apparatus further includes: an initial spatial position determination module, an initial reference point determination module, and a first spatial position determination module. The initial spatial position determination module is configured to determine an initial spatial position of the target on a ground of the target scene. The initial reference point determination module is configured to add random quantities to the initial spatial position, to generate third spatial positions of a plurality of initial reference points. The first spatial position determination module, configured to obtain the first spatial positions of the plurality of first reference points

according to the third spatial positions of the plurality of initial reference points.

[00130] In a possible implementation, the initial spatial position determination module includes: an acquisition sub-module, a homography matrix determination sub-module and an initial spatial position determination sub-module. The acquisition sub-module is configured to obtain a second image of the target scene, where the second image includes a reference object preset on the ground and the target on the ground, and an acquisition time of the second image is before an acquisition time of the first image. The homography matrix determination sub-module is configured to determine a homography matrix between an imaging plane of the at least one image acquisition device and the ground according to a fourth spatial position of the reference object on the ground and a third image position of the reference object in the second image. The initial spatial position determination sub-module, configured to determine the initial spatial position of the target according to the homography matrix and a fourth image position of the target on the second image.

[00131] In a possible implementation, the apparatus further includes a reference point determination module. The reference point determination module is configured to: select a second reference point whose weight exceeds a preset threshold from the plurality of first reference points according to weights of the plurality of first reference points; add random quantities to a fifth spatial position of the second reference point, to generate sixth spatial positions of a plurality of third reference points; and perform spatial positioning on the target in a third image of the target scene according to the sixth spatial positions of the plurality of third reference points, to obtain the second spatial position of the target in the target scene, where an acquisition time of the third image is after an acquisition time of the first image.

[00132] In a possible implementation, the weight determination sub-module is configured to: for any one of the first reference points, determine the weight of the first reference point according to a similarity between a first position vector formed by a plurality of second image positions of the first reference point and a second position

vector formed by a plurality of first image positions of the target, where the weight is positively correlated with the similarity.

[00133] In the embodiments of the present disclosure, the second image positions of the plurality of first reference points can be obtained based on the estimated first spatial positions of the plurality of first reference points, and the second spatial position of the target in the target scene can be determined according to the first image position where the target is actually located in the first image in combination with the estimated first spatial positions, so that the spatial positioning of the target can be realized without arranging any additional hardware device on the target, which saves costs and has high universality.

[00134] In some embodiments, the apparatus provided by the embodiments of the present disclosure has functions or includes modules that may be used to perform the methods described in the above method embodiments, and the specific implementations thereof may be described with reference to the above method embodiments.

[00135] The embodiments of the present disclosure further provide a computer readable storage medium having stored thereon computer program instructions that, when executed by a processor, implement the method described above. The computer readable storage medium may be a volatile or non-volatile computer readable storage medium.

[00136] The embodiments of the present disclosure further provide an electronic device, including: a processor; and a memory for storing instructions executable by the processor. The processor is configured to invoke instructions stored in the memory to perform the method described above.

[00137] The embodiments of the present disclosure also provide a computer program product including computer readable codes, and when the computer readable codes run in a device, a processor in the device executes instructions for implementing the method for positioning the target provided in any one of the embodiments described above.

[00138] The embodiments of the present disclosure also provide another computer program product storing computer readable instructions that, when executed, cause a computer to perform the operations of the method for positioning the target provided in any one of the embodiments described above.

[00139] The electronic device may be provided as a terminal, a server, or other form of device.

[00140] FIG. 4 illustrates a block diagram of an electronic device 800 according to an embodiment of the present disclosure. For example, the electronic device 800 may be a terminal device such as a mobile phone, a computer, a digital broadcast terminal, a messaging device, a game console, a tablet, a medical device, exercise equipment, a personal digital assistant, and the like.

[00141] Referring to FIG. 4, the electronic device 800 may include one or more of the following components: a processing component 802, a memory 804, a power component 806, a multimedia component 808, an audio component 810, an input/output (I/O) interface 812, a sensor component 814, and a communication component 816.

[00142] The processing component 802 generally controls the overall operation of the electronic device 800, such as operations associated with display, telephone calls, data communications, camera operations, and recording operations. The processing component 802 may include one or more processors 820 to execute instructions to perform all or part of the operations of the abovementioned method. In addition, the processing component 802 may include one or more modules which facilitate interaction between the processing component 802 and other components. For example, the processing component 802 may include a multimedia module to facilitate the interaction between the multimedia component 808 and the processing component 802.

[00143] The memory 804 is configured to store various types of data to support the operation of the electronic device 800. Examples of such data include instructions for any application or method operating on the electronic device 800, contact data,

phonebook data, messages, pictures, video, and the like. The memory 804 may be implemented by using any type of volatile or non-volatile storage device or a combination thereof, such as a static random access memory (SRAM), an electrically erasable programmable read-only memory (EEPROM), an erasable and programmable read-only memory (EPROM), a programmable read-only memory (PROM), a read-only memory (ROM), a magnetic memory, a flash memory, a magnetic or optical disk.

[00144] The power component 806 provides power to various components of the electronic device 800. The power component 806 may include a power management system, one or more power sources, and other components associated with the generation, management, and distribution of power in the electronic device 800.

[00145] The multimedia component 808 includes a screen providing an output interface between the electronic device 800 and the user. In some embodiments, the screen may include a liquid crystal display (LCD) and a touch panel (TP). If the screen includes a touch panel, the screen may be implemented as a touch screen to receive input signals from the user. The touch panel includes one or more touch sensors to sense touches, swipes, and gestures on the touch panel. The touch sensors may not only sense a boundary of a touch or swipe action, but also sense a period of time and a pressure associated with the touch or swipe action. In some embodiments, the multimedia component 808 includes a front camera and/or a rear camera. The front camera and/or the rear camera may receive an external multimedia datum while the electronic device 800 is in an operation mode, such as a photographing mode or a video mode. Each of the front camera and the rear camera may be a fixed optical lens system or have focus and optical zoom capability.

[00146] The audio component 810 is configured to output and/or input audio signals. For example, the audio component 810 includes a microphone (MIC) configured to receive an external audio signal when the electronic device 800 is in an operation mode, such as a call mode, a recording mode, and a voice recognition mode. The received audio signal may be further stored in the memory 804 or transmitted via the

communication component 816. In some embodiments, the audio component 810 further includes a speaker for outputting audio signals.

[00147] The I/O interface 812 provides an interface between the processing component 802 and peripheral interface modules, such as a keyboard, a click wheel, buttons, and the like. The buttons may include, but are not limited to, a home button, a volume button, a starting button, and a locking button.

[00148] The sensor component 814 includes one or more sensors to provide status assessments of various aspects of the electronic device 800. For example, the sensor component 814 may detect an open/closed status of the electronic device 800, relative positioning of components, e.g., the display and the keypad, of the electronic device 800, a change in position of the electronic device 800 or a component of the electronic device 800, a presence or absence of user contact with the electronic device 800, an orientation or an acceleration/deceleration of the electronic device 800, and a change in temperature of the electronic device 800. The sensor component 814 may include a proximity sensor configured to detect the presence of nearby objects without any physical contact. The sensor component 814 may also include a light sensor, such as a complementary metal oxide semiconductor (CMOS) or a charge coupled device (CCD) image sensor, for use in imaging applications. In some embodiments, the sensor component 814 may also include an accelerometer sensor, a gyroscope sensor, a magnetic sensor, a pressure sensor, or a temperature sensor.

[00149] The communication component 816 is configured to facilitate communication, wired or wirelessly, between the electronic device 800 and other devices. The electronic device 800 can access a wireless network based on a communication standard, such as a wireless network (WiFi), a second generation mobile communication technology (2G), or a third generation mobile communication technology (3G), or a combination thereof. In one exemplary embodiment, the communication component 816 receives a broadcast signal or broadcast associated information from an external broadcast management system via a broadcast channel. In one exemplary embodiment, the communication

component 816 further includes a near field communication (NFC) module to facilitate short-range communications. For example, the NFC module may be implemented based on a radio frequency identification (RFID) technology, an infrared data association (IrDA) technology, an ultra-wideband (UWB) technology, a bluetooth (BT) technology, and other technologies.

[00150] In exemplary embodiments, the electronic device 800 may be implemented with one or more application specific integrated circuits (ASICs), digital signal processors (DSPs), digital signal processing devices (DSPDs), programmable logic devices (PLDs), field programmable gate arrays (FPGAs), controllers, micro-controllers, microprocessors, or other electronic components, for performing the above described methods.

[00151] In an exemplary embodiment, a non-volatile computer readable storage medium is also provided, such as a memory 804 including computer program instructions that may be executed by a processor 820 in an electronic device 800 to complete the methods described above.

[00152] FIG. 5 illustrates a block diagram of an electronic device 1900 according to an embodiment of the present disclosure. For example, the electronic device 1900 may be provided as a server. Referring to FIG. 5, the electronic device 1900 includes a processing component 1922 which further includes one or more processors; and a memory resource represented by a memory 1932, for storing instructions executable by the processing component 1922, for example an application program. The application program stored in the memory 1932 may include one or more modules, each corresponding to a set of instructions. Additionally, the processing component 1922 is configured to execute instructions, so as to perform the methods described above.

[00153] The electronic device 1900 may also include a power component 1926 configured to perform power management for the electronic device 1900, a wired or wireless network interface 1950 configured to connect the electronic device 1900 to a

network, and an input/output (I/O) interface 1958. The electronic device 1900 may operate based on an operating system stored in the memory 1932, such as a Microsoft server operating system (Windows Server™), a graphical user interface operating system (Mac OS X™) introduced by Apple, a multi-user multi-process computer operating system (Unix™), a free and open-source Unix-like operating system (Linux™), an open-source Unix-like operating system (FreeBSD™), or the like.

[00154] In an exemplary embodiment, a non-volatile computer readable storage medium is also provided, such as a memory 1932 including computer program instructions that may be executed by a processing component 1922 in an electronic device 1900 to complete the methods described above.

[00155] The present disclosure may be a system, a method, and/or a computer program product. The computer program product may include a computer readable storage medium carrying computer readable program instructions that, cause a processor to implement various aspects of the present disclosure.

[00156] The computer readable storage medium may be a tangible device that holds and stores an instruction used by an instruction execution device. The computer readable storage medium may be, for example, but not limited to, an electrical storage device, a magnetic storage device, an optical storage device, an electromagnetic storage device, a semiconductor storage device or any suitable combination of the above. The more specific list (non-exhaustive list) of the computer readable storage media includes: a portable computer disk, a hard disk, a Random Access Memory (RAM), a Read-Only Memory (ROM), an Erasable Programmable Read-Only Memory (EPROM or a flash memory), a static Random Access Memory (SRAM), a portable Compact Disk Read-Only Memory (CD-ROM), a Digital Versatile Disk (DVD), a memory stick, a floppy disk, a mechanical encoding device, such as a punch card or a raised structure in a groove on which an instruction is stored, and any suitable combination of the above. As used herein, the computer readable storage medium is not interpreted as an instantaneous signal itself, such as radio waves or other freely propagating battery

waves, battery waves propagating through waveguides or other transmission media (e.g., optical pulses through fiber optic cables), or electrical signals transmitted through wires.

[00157] The computer readable program instructions described herein may be downloaded from a computer readable storage medium to various computing/processing devices, or to an external computer or external storage device over a network, such as an Internet, a local area network, a wide area network, and/or a wireless network. The network may include copper transport cables, fiber optic transmission, wireless transmission, routers, firewalls, switches, gateway computers, and/or edge servers. A network adapter card or network interface in each computing/processing device receives computer readable program instructions from the network and forwards the computer readable program instructions for storage in a computer readable storage medium in each computing/processing device.

[00158] The computer program instructions used to perform the operations of the present disclosure may be assembly instructions, instruction set architecture (ISA) instructions, machine instructions, machine-related instructions, microcode, firmware instructions, state setting data, or source code or object code written in any combination of one or more programming languages, which include object-oriented programming languages such as Smalltalk, C++, and the like, and conventional process programming languages such as "C" language or similar. The computer readable program instructions may be executed entirely on the user computer, partly on the user computer, as a separate software package, partly on the user computer and partly on the remote computer, or entirely on the remote computer or server. In the case involving a remote computer, the remote computer may be connected to a user computer over any kind of network (which includes a local area network (LAN) or a wide area network (WAN)), or may be connected to an external computer (e.g., via the Internet by using an Internet service provider). In some embodiments, an electronic circuit, such as a programmable logic circuit, a field programmable gate array (FPGA), or a programmable logic array (PLA), may execute computer readable program instructions to implement various aspects of the present disclosure by using status information of the computer readable

program instructions to personalize the electronic circuit.

[00159] Various aspects of the present disclosure are described herein with reference to a flowchart and/or block diagram of a method, apparatus (system), and computer program product according to the embodiments of the disclosure. It should be understood that each block of the flowchart and/or block diagram and a combination of the blocks in the flowchart and/or block diagram may be implemented by computer readable program instructions.

[00160] These computer readable program instructions may be provided to a processor of a general purpose computer, a special purpose computer, or other programmable data processing device, thereby producing a machine, such that when the instructions are executed by the processor of the computer or other programmable data processing device, the apparatus for implementing the functions/actions specified in one or more of the blocks in the flowchart and/or block diagram is produced. These computer readable program instructions may also be stored in a computer readable storage medium to cause a computer, programmable data processing device, and/or other device to operate in a particular manner. So that the computer readable medium storing the instructions includes an article of manufacture that includes instructions to implement various aspects of the functions/actions specified in one or more of the blocks in the flowchart and/or block diagram.

[00161] Computer readable program instructions may also be loaded onto a computer, other programmable data processing device, or other device, such that a series of operational steps are performed on the computer, other programmable data processing device, or other device to produce a computer-implemented process. So that the instructions executed on the computer, other programmable data processing device, or other device implement the functions/actions specified in one or more of the blocks in the flowchart and/or block diagram.

[00162] The flowcharts and block diagrams in the drawings illustrate architectures,

functions, and operations of possible implementations of systems, methods, and computer program products according to various embodiments of the present disclosure. In this regard, each block in a flowchart or block diagram may represent a module, segment, or portion of instructions that contains one or more executable instructions for implementing specified logical functions. In some alternative implementations, the functions indicated in the blocks may also occur in a different order than those indicated in the drawings. For example, two consecutive blocks may actually be executed substantially in parallel, and they may sometimes be executed in reverse order, which depends on the function involved. It is also to be noted that each block in the block diagram and/or flowchart, and a combination of blocks in the block diagram and/or flowchart, may be implemented with a dedicated hardware-based system that performs specified functions or actions, or may be implemented with a combination of dedicated hardware and computer instructions.

[00163] The computer program product may be implemented in particular by means of hardware, software or a combination thereof. In one alternative embodiment, the computer program product is embodied as a computer storage medium. In another alternative embodiment, the computer program product is embodied as a software product, such as a Software Development Kit (SDK) or the like.

[00164] The embodiments of the present disclosure have been described above, the above description being exemplary and not exhaustive, and are also not limited to the disclosed embodiments. Many modifications and alterations will be apparent to those skilled in the art without departing from the scope and spirit of the illustrated embodiments. The selection of terms used herein is intended to best explain the principles of the embodiments, practical applications or improvements to the technology in the market, or to enable other persons of ordinary skill in the art to understand the embodiments disclosed herein.

INDUSTRIAL PRACTICALITY

[00165] The present disclosure provides a method for positioning a target, an apparatus, an electronic device, a storage medium, and a program. Where obtaining a first image of a target scene, where the first image is obtained by at least one image acquisition device; performing target detection on the first image, to determine a first image position of a target, which is in the target scene, in the first image; determining first spatial positions of a plurality of first reference points for the target in the target scene; determining second image positions of the plurality of first reference points in the first image according to the first spatial positions of the plurality of first reference points and a camera parameter of the at least one image acquisition device; and determining a second spatial position of the target in the target scene according to the first image position of the target, the second image positions of the plurality of first reference points, and the first spatial positions of the plurality of first reference points.

CLAIMS

1. A method for positioning a target, performed by an electronic device, the method comprising:

obtaining a first image of a target scene, wherein the first image is obtained by at least one image acquisition device;

performing target detection on the first image, to determine a first image position of a target, which is in the target scene, in the first image;

determining first spatial positions of a plurality of first reference points for the target in the target scene;

determining second image positions of the plurality of first reference points in the first image according to the first spatial positions of the plurality of first reference points and a camera parameter of the at least one image acquisition device; and

determining a second spatial position of the target in the target scene according to the first image position of the target, the second image positions of the plurality of first reference points, and the first spatial positions of the plurality of first reference points.

2. The method of claim 1, wherein the determining the second spatial position of the target in the target scene according to the first image position of the target, the second image positions of the plurality of first reference points, and the first spatial positions of the plurality of first reference points comprises:

for any one of the first reference points, determining a weight of the first reference point according to a similarity between the second image position of the first reference point and the first image position of the target; and

determining the second spatial position of the target according to weights of the plurality of first reference points and the first spatial positions of the plurality of first reference points.

3. The method of claim 2, wherein the determining the second spatial position of the target according to the weights of the plurality of first reference points and the first spatial positions of the plurality of first reference points comprises:

determining a weighted average of the first spatial positions of the plurality of first reference points based on the weights of the plurality of first reference points; and

determining the weighted average as the second spatial position of the target.

4. The method of any one of claims 1 to 3, wherein the determining the first spatial positions of a plurality of first reference points for the target in the target scene comprises:

determining an initial spatial position of the target on a ground of the target scene;

adding random quantities to the initial spatial position, to generate third spatial positions of a plurality of initial reference points; and

obtaining the first spatial positions of the plurality of first reference points according to the third spatial positions of the plurality of initial reference points.

5. The method of claim 4, wherein the determining the initial spatial position of the target on the ground of the target scene comprises:

obtaining a second image of the target scene, wherein the second image comprises a reference object preset on the ground and the target on the ground, and an acquisition time of the second image is before an acquisition time of the first image;

determining a homography matrix between an imaging plane of the at least one image acquisition device and the ground according to a fourth spatial position of the reference object on the ground and a third image position of the reference object in the second image; and

determining the initial spatial position of the target according to the homography matrix and a fourth image position of the target on the second image.

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

selecting a second reference point whose weight exceeds a preset threshold from the plurality of first reference points according to weights of the plurality of first reference points;

adding random quantities to a fifth spatial position of the second reference point, to generate sixth spatial positions of a plurality of third reference points; and

performing spatial positioning on the target in a third image of the target scene according to the sixth spatial positions of the plurality of third reference points, to obtain the second spatial position of the target in the target scene, wherein an acquisition time of the third image is after an acquisition time of the first image.

7. The method of claim 2 or 3, wherein for any one of the first reference points, determining the weight of the first reference point according to the similarity between the second image position of the first reference point and the first image position of the target comprises:

determining the weight of any one of the first reference point according to a similarity between a first position vector formed by a plurality of second image positions of the first reference point and a second position vector formed by a plurality of first image positions of the target, wherein the weight is positively correlated with the similarity.

8. An apparatus for positioning a target, comprising:

an obtaining module, configured to obtain a first image of a target scene, wherein the first image is obtained by at least one image acquisition device;

a detection module, configured to perform target detection on the first image, to determine a first image position of a target, which is in the target scene, in the first image;

a first spatial position determination module, configured to determine first spatial positions of a plurality of first reference points for the target in the target scene;

an image position determination module, configured to determine second image positions of the plurality of first reference points in the first image according to the first spatial positions of the plurality of first reference points and a camera parameter of the at least one image acquisition device; and

a second spatial position determination module, configured to determine a second spatial position of the target in the target scene according to the first image position of the target, the second image positions of the plurality of first reference points, and the first spatial positions of the plurality of first reference points.

9. The apparatus of claim 8, wherein the second spatial position determination module comprises:

a weight determination sub-module, configured to, for any one of the first reference points, determine a weight of the first reference point according to a similarity between the second image position of the first reference point and the first image position of the target; and

a spatial position determination sub-module, configured to determine the second spatial position of the target according to weights of the plurality of first reference points and the first spatial positions of the plurality of first reference points.

10. The apparatus of claim 9, wherein the spatial position determination sub-module is configured to: determine a weighted average of the first spatial positions of the plurality of first reference points based on the weights of the plurality of first reference points; and determine the weighted average as the second spatial position of the target.

11. The apparatus of any one of claims 8 to 10, further comprising:

an initial spatial position determination module, configured to determine an initial spatial position of the target on a ground of the target scene;

an initial reference point determination module, configured to add random quantities

to the initial spatial position, to generate third spatial positions of a plurality of initial reference points; and

a first spatial position determination module, configured to obtain the first spatial positions of the plurality of first reference points according to the third spatial positions of the plurality of initial reference points.

12. The apparatus of claim 11, wherein the initial spatial position determination module comprises:

an acquisition sub-module, configured to obtain a second image of the target scene, wherein the second image comprises a reference object preset on the ground and the target on the ground, and an acquisition time of the second image is before an acquisition time of the first image;

a homography matrix determination sub-module, configured to determine a homography matrix between an imaging plane of the at least one image acquisition device and the ground according to a fourth spatial position of the reference object on the ground and a third image position of the reference object in the second image; and

an initial spatial position determination sub-module, configured to determine the initial spatial position of the target according to the homography matrix and a fourth image position of the target on the second image.

13. The apparatus of any one of claims 8 to 12, further comprising:

a reference point determination module, configured to: select a second reference point whose weight exceeds a preset threshold from the plurality of first reference points according to weights of the plurality of first reference points; add random quantities to a fifth spatial position of the second reference point, to generate sixth spatial positions of a plurality of third reference points; and perform spatial positioning on the target in a third image of the target scene according to the sixth spatial positions of the plurality of third reference points, to obtain the second spatial position of the target in the target scene, wherein an acquisition time

of the third image is after an acquisition time of the first image.

14. The apparatus of claim 9, wherein the weight determination sub-module is configured to: for any one of the first reference points, determine the weight of the first reference point according to a similarity between a first position vector formed by a plurality of second image positions of the first reference point and a second position vector formed by a plurality of first image positions of the target, wherein the weight is positively correlated with the similarity.

15. An electronic device, comprising:

a processor; and

a memory for storing instructions executable by the processor;

wherein the processor is configured to invoke the instructions stored in the memory to perform the method for positioning the target of any one of claims 1 to 7.

16. A computer readable storage medium having stored thereon computer program instructions that, when executed by a processor, implement the method for positioning the target of any one of claims 1 to 7.

17. A computer program comprising computer readable codes that, when run in an electronic device, cause a processor of the electronic device to implement the method for positioning the target of any one of claims 1 to 7.