

MECHANICAL ARM CONTROL METHOD AND DEVICE BASED ON MONOCULAR VISION POSITIONING

CROSS-REFERENCE TO RELATED APPLICATION(S)

[0001] This application claims priority to Chinese Patent Application No. 202010349776.X filed April 28, 2020, entitled “MECHANICAL ARM CONTROL METHOD AND DEVICE BASED ON MONOCULAR VISION POSITIONING”, the entire disclosure of which is hereby incorporated by reference.

TECHNICAL FIELD

[0002] This disclosure relates to the technical field of machine vision, and particularly to a mechanical arm control method and device based on monocular vision positioning.

BACKGROUND

[0003] With the improvement of the level of intelligence and the popularization of the use of robots, mechanical arm movement control based on machine vision has attracted widespread attention. The mechanical arm movement control based on machine vision is to obtain target positioning through machine vision and to control an end of a mechanical arm to move to a target object to perform related operations such as grabbing.

[0004] The inventor realizes that the currently commonly used machine vision positioning method is a binocular vision positioning method. That is, two monocular cameras with a certain distance are used to collect images, and a distance (that is, parallax) between a pair of matching points in images collected by the two monocular cameras is obtained through image processing. Then, according to a correspondence relationship between parallaxes and depth information, distance information of the target object can be obtained. According to the distance information and a focal length, coordinate information in X and Y directions is further obtained. However, more cameras are used in this method, resulting in a complex positioning system, inconvenient operations, and a reduced positioning efficiency.

SUMMARY

[0005] In view of this, this application provides a mechanical arm control method and device based on monocular vision positioning, which aims to solve a problem of a complex positioning system which is inconvenient to operate and has a low positioning efficiency due to usage of more cameras in a binocular vision positioning method of the prior art.

[0006] According to an aspect of this application, a mechanical arm control method based on monocular vision positioning is provided. The method is implemented in a mechanical arm control system based on monocular vision positioning. The mechanical arm control system includes a mechanical arm and a monocular camera installed at an end of the mechanical arm. The method includes the following.

[0007] A pixel coordinate system, a camera coordinate system, and a world coordinate system are established. Pixel coordinates of a target point of a target object in the pixel coordinate system, an intrinsic parameter matrix of the camera, an extrinsic parameter matrix of the camera, and a homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system are obtained. A line-of-sight vector of the target point relative to the world coordinate system are obtained according to the pixel coordinates of the target point, the intrinsic parameter matrix, the extrinsic parameter matrix, and the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system. A line-of-sight linear equation is obtained according to a homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system and the line-of-sight vector of the target point relative to the world coordinate system. A planar equation of a plane where the target point is located is obtained in the world coordinate system. World coordinates of the target point are obtained according to the line-of-sight linear equation and the planar equation. The mechanical arm is controlled, according to the world coordinates of the target point, to operate the target object according to a preset strategy.

[0008] According to another aspect of this application, a mechanical arm control device based on monocular vision positioning is provided. The device includes a coordinate system establishing module configured to establish a pixel coordinate system, a camera coordinate system, and a world coordinate system; a first obtaining module configured to obtain pixel coordinates of a target point of a target object in the pixel coordinate system, an intrinsic parameter matrix of a camera, an extrinsic parameter matrix of the camera, and a homogeneous

transformation matrix of an end of a mechanical arm relative to the world coordinate system; a line-of-sight vector determining module configured to obtain a line-of-sight vector of the target point relative to the world coordinate system according to the pixel coordinates of the target point, the intrinsic parameter matrix, the extrinsic parameter matrix, and the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system; a line-of-sight linear equation determining module configured to obtain a line-of-sight linear equation according to a homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system and the line-of-sight vector of the target point relative to the world coordinate system; a second obtaining module configured to obtain, in the world coordinate system, a planar equation of a plane where the target point is located; a world coordinates determining module configured to obtain world coordinates of the target point according to the line-of-sight linear equation and the planar equation; and a control module configured to control, according to the world coordinates of the target point, the mechanical arm to operate the target object according to a preset strategy.

[0009] According to yet another aspect of this application, a storage medium is provided. The storage medium stores at least one executable instruction, which causes a processor to execute the following operations.

[0010] A pixel coordinate system, a camera coordinate system, and a world coordinate system are established. Pixel coordinates of a target point of a target object in the pixel coordinate system, an intrinsic parameter matrix of the camera, an extrinsic parameter matrix of the camera, and a homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system are obtained. A line-of-sight vector of the target point relative to the world coordinate system are obtained according to the pixel coordinates of the target point, the intrinsic parameter matrix, the extrinsic parameter matrix, and the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system. A line-of-sight linear equation is obtained according to a homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system and the line-of-sight vector of the target point relative to the world coordinate system. A planar equation of a plane where the target point is located is obtained in the world coordinate system. World coordinates of the target point are obtained according to the line-of-sight linear equation and the planar equation. The mechanical arm is controlled, according to the world coordinates of the target point, to operate

the target object according to a preset strategy.

[0011] According to still another aspect of this application, a computer device is provided. The computer device includes a processor, a memory, a communication interface, and a communication bus. The processor, the memory, and the communication interface communicate with each other through the communication bus and the memory is configured to store at least one executable instruction, which causes the processor to execute the following.

[0012] A pixel coordinate system, a camera coordinate system, and a world coordinate system are established. Pixel coordinates of a target point of a target object in the pixel coordinate system, an intrinsic parameter matrix of the camera, an extrinsic parameter matrix of the camera, and a homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system are obtained. A line-of-sight vector of the target point relative to the world coordinate system are obtained according to the pixel coordinates of the target point, the intrinsic parameter matrix, the extrinsic parameter matrix, and the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system. A line-of-sight linear equation is obtained according to a homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system and the line-of-sight vector of the target point relative to the world coordinate system. A planar equation of a plane where the target point is located is obtained in the world coordinate system. World coordinates of the target point are obtained according to the line-of-sight linear equation and the planar equation. The mechanical arm is controlled, according to the world coordinates of the target point, to operate the target object according to a preset strategy.

[0013] According to implementations of this application, there is no need to set two monocular cameras, which can achieve a simplified positioning system and convenient operations, and comparison of image feature points and calculation of an optimal distance are not involved, which simplifies a processing algorithm and reduces requirements for hardware equipment.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] FIG. 1 illustrates a flow chart of a mechanical arm control method based on monocular vision positioning provided in implementations of this application.

[0015] FIG. 2 illustrates a flow chart of a mechanical arm control method based on

monocular vision positioning provided in other implementations of this application.

[0016] FIG. 3 illustrates an application scenario of implementations of this application.

[0017] FIG. 4 illustrates a block diagram of a mechanical arm control device based on monocular vision positioning provided in implementations of this application.

[0018] FIG. 5 illustrates a block diagram of a mechanical arm control device based on monocular vision positioning provided in other implementations of this application.

[0019] FIG. 6 illustrates a schematic structural diagram of a computer device provided in implementations of this application.

DETAILED DESCRIPTION

[0020] Hereinafter, exemplary implementations of the disclosure will be described in detail with reference to accompanying drawings. Although the accompanying drawings illustrate the exemplary implementations of the disclosure, it should be understood that the disclosure can be implemented in various forms and should not be limited by the implementations set forth herein. On the contrary, these implementations are provided for a more thorough understanding of the disclosure and to fully convey the scope of the disclosure to those skilled in the art.

[0021] The technical solution of this application can be applied to the technical field of artificial intelligence and/or smart city, involving computer vision technology. The mechanical arm control scheme based on machine vision positioning of this application helps to simplify the positioning effect, improve the positioning efficiency, and realize smart life.

[0022] A mechanical arm control method based on monocular vision positioning is provided in implementations of this application. The method is implemented in a mechanical arm control system based on monocular vision positioning. The mechanical arm control system includes a mechanical arm and a monocular camera installed at an end of the mechanical arm. As illustrated in FIG. 1, the method includes the following.

[0023] At block 101, a pixel coordinate system, a camera coordinate system at the end of the mechanical arm, and a world coordinate system are established.

[0024] The world coordinate system can be built according to the actual location of the target object. The camera coordinate system is established as follows. In space, the optical center of the monocular camera, that is, the center of the optical axis of the monocular camera is determined as the origin of the camera coordinate system, the Z_C axis coincides with the optical axis of the

camera and is perpendicular to the imaging plane, and the photographing direction is taken as the positive direction, to establish the camera coordinate system $O_C - X_C Y_C Z_C$. The pixel coordinate system is established on the plane of the image collected by the monocular camera. Specifically, the vertex of the upper left corner of the plane of the image can be determined as the origin of the pixel coordinate system, and the horizontal line and vertical line can be determined as the u-axis and v-axis, respectively, to establish the pixel coordinate system. The pixel coordinate system is in unit of the number of pixels.

[0025] At block 102, pixel coordinates of a target point of a target object in the pixel coordinate system, an intrinsic parameter matrix of the camera, an extrinsic parameter matrix of the camera, and a homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system are obtained.

[0026] The target object is an object operated by the mechanical arm, and the target point is a specific point on the target object, generally a center point of the target object. The intrinsic parameter matrix (or called camera intrinsic parameter matrix) and the extrinsic parameter matrix (or called camera extrinsic parameter matrix) can be obtained by camera calibration.

[0027] At block 103, a line-of-sight vector of the target point relative to the world coordinate system are obtained according to the pixel coordinates of the target point, the intrinsic parameter matrix, the extrinsic parameter matrix, and the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system.

[0028] At block 104, a line-of-sight linear equation is obtained according to a homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system and the line-of-sight vector of the target point relative to the world coordinate system.

[0029] At block 105, a planar equation of a plane where the target point is located is obtained in the world coordinate system.

[0030] At block 106, world coordinates of the target point are obtained according to the line-of-sight linear equation and the planar equation.

[0031] At block 107, the mechanical arm is controlled, according to the world coordinates of the target point, to operate the target object according to a preset strategy.

[0032] The implementations of this application provide the mechanical arm control method based on monocular vision positioning. The pixel coordinates of the target point of the target object in the pixel coordinate system are obtained. The homogeneous transformation matrix of

the camera coordinate system relative to the end of the mechanical arm and the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system are used for transformation, to obtain the line-of-sight linear equation. In the world coordinate system, the plane equation of the plane where the target point is located is obtained. According to the line-of-sight linear equation and the plane equation, the world coordinates of the target point are obtained. Thus, compared with the prior art, in the implementations of this application, there is no need to set two monocular cameras, which can achieve a simplified positioning system and convenient operations thereof. Furthermore, in the implementations of this application, comparison of image feature points and calculation of an optimal distance are not involved, which simplifies a processing algorithm and reduces requirements for hardware equipment.

[0033] A mechanical arm control method based on monocular vision positioning is provided in other implementations of this application. The method is implemented in a mechanical arm control system based on monocular vision positioning. The mechanical arm control system includes a mechanical arm and a monocular camera installed at an end of the mechanical arm. As illustrated in FIG. 2, the method includes the following.

[0034] At block 201, a pixel coordinate system, a camera coordinate system at the end of the mechanical arm, and a world coordinate system are established.

[0035] The world coordinate system can be constructed according to the actual location of the target object. Taking FIG. 3 as an example, the target object is a teacup and the teacup is placed on the workbench. The plane formed by an x-axis and a y-axis of the world coordinate system is parallel to the upper surface of the worktable, and a z-axis of the world coordinate system is perpendicular to the upper surface of the worktable. That is, the world coordinate system is the coordinate system $O_b-X_bY_bZ_b$.

[0036] The camera coordinate system is established as follows. In space, the optical center of the monocular camera, that is, the center of the optical axis of the monocular camera is determined as the origin of the camera coordinate system, the Z_C axis coincides with the optical axis of the camera and is perpendicular to the imaging plane, and the photographing direction is taken as the positive direction, to establish the camera coordinate system $O_C-X_CY_CZ_C$. The pixel coordinate system is established on the plane of the image collected by the monocular camera. Specifically, the vertex of the upper left corner of the plane of the image can be

determined as the origin of the pixel coordinate system, and the horizontal line and vertical line can be determined as the u-axis and v-axis, respectively, to establish the pixel coordinate system. The pixel coordinate system has a unit of the number of pixels.

[0037] At block 202, pixel coordinates of a target point of a target object in the pixel coordinate system, an intrinsic parameter matrix of the camera, an extrinsic parameter matrix of the camera, and a homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system are obtained.

[0038] The target object is an object operated by the mechanical arm, and the target point is a specific point on the target object, generally a center point of the target object. Taking FIG. 3 as an example, the target object is a teacup and the center point of the mouth of the teacup is determined as the target point.

[0039] The intrinsic parameter matrix and the extrinsic parameter matrix can be obtained by collecting multiple images of a calibration target with the monocular camera and calibrating the monocular camera with an image set. The specific calibration process is as follows. Focus the monocular camera on a checkerboard calibration target, where the size of the checkerboard is known. Take images of the calibration target from different angles. For each image of the calibration target, extract the pixel position of the corner point of the calibration target in the image. Thereafter, according to the pixel position of the corner point of the calibration target in the image and the parameters of the calibration target, perform camera calibration by using Zhang's calibration method, to obtain the intrinsic parameter matrix.

[0040] The target object is an object operated by the mechanical arm, and the target point is a specific point on the target object, generally a center point of the target object. The intrinsic parameter matrix can be obtained by camera calibration.

[0041] At block 203, a three-dimensional (3D) coordinate expression of the target point in the camera coordinate system is obtained according to the pixel coordinates of the target point and the intrinsic parameter matrix, where the 3D coordinate expression of the target point in the

camera coordinate system is $s \begin{bmatrix} \mu \\ \nu \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & \alpha \\ 0 & f_y & \beta \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ and where $\begin{bmatrix} \mu \\ \nu \\ 1 \end{bmatrix}$ is the pixel coordinates of

the target point, s is a homogeneous transformation matching coefficient, $\begin{bmatrix} f_x & 0 & \alpha \\ 0 & f_y & \beta \\ 0 & 0 & 1 \end{bmatrix}$ is the intrinsic parameter matrix, f_x is a focal length coefficient of an x-axis of a pixel plane, f_y is a focal length coefficient of a y-axis of the pixel plane, α is an offset of the x-axis of the pixel

plane, β is an offset of the y-axis of the pixel plane, and $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ is 3D coordinates of the target point in the camera coordinate system.

[0042] At block 204, a line-of-sight vector of the target point relative to the camera coordinate system is obtained according to the 3D coordinate expression of the target point in the camera coordinate system, where the line-of-sight vector of the target point relative to the camera

$$\begin{bmatrix} \gamma_c^x \\ \gamma_c^y \\ \gamma_c^z \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{\mu / f_x}{\sqrt{\left(\frac{\mu}{f_x}\right)^2 + \left(\frac{\nu}{f_y}\right)^2} + 1} \\ \frac{\nu / f_y}{\sqrt{\left(\frac{\mu}{f_x}\right)^2 + \left(\frac{\nu}{f_y}\right)^2} + 1} \\ 1 \\ 1 \end{bmatrix},$$

coordinate system is

[0043] where γ_c^x is a projection value (i.e., a value of a projection) of the target point on an x-axis of the camera coordinate system, γ_c^y is a projection value of the target point on a y-axis of the camera coordinate system, γ_c^z is a projection value of the target point on a z-axis of the camera coordinate system, f_x is the focal length coefficient of the x-axis of the pixel plane, f_y is the focal length coefficient of the y-axis of the pixel plane, and μ and ν are the pixel coordinates of the target point.

[0044] At block 205, a homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system is obtained according to the extrinsic parameter matrix

and the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system and according to the following formula,

$$H_w^c = H_w^e \times H_e^c,$$

[0045] where H_w^c is the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system, H_w^e is the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system, and H_e^c is the extrinsic parameter matrix.

[0046] The homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system H_w^e can be solved based on mechanical arm kinematics. In mechanical arm kinematics, the position and posture of the end of the mechanical arm relative to the world coordinate system is calculated according to various known joint angles of the mechanical arm. Specifically, the mechanical arm kinematics modeling uses the D-H (Denavit-Hartenberg) four-parameter method, which is a matrix method for establishing relative position and posture. Use homogeneous transformation to describe the spatial geometric relationship of each link relative to the fixed reference coordinate system and use a 4×4 homogeneous transformation matrix to describe the spatial relationship between two adjacent links, to derive matrix the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system H_w^e .

[0047] At block 206, a line-of-sight vector of the target point relative to the world coordinate system is obtained according to the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system and the line-of-sight vector of the target point relative to the camera coordinate system and according to the following formula,

$$\begin{bmatrix} r_w^x \\ r_w^y \\ r_w^z \\ 1 \end{bmatrix} = H_w^c \begin{bmatrix} r_c^x \\ r_c^y \\ r_c^z \\ 1 \end{bmatrix},$$

[0048] where $\begin{bmatrix} r_w^x \\ r_w^y \\ r_w^z \\ 1 \end{bmatrix}$ is the line-of-sight vector of the target point relative to the world coordinate system, r_w^x is a projection value of the target point on the x-axis of the world coordinate system, r_w^y is a projection value of the target point on the y-axis of the world coordinate system, r_w^z is a projection value of the target point on the z-axis of the world coordinate system, H_w^c is the homogeneous transformation matrix of the camera coordinate

system relative to the world coordinate system, $\begin{bmatrix} r_c^x \\ r_c^y \\ r_c^z \\ 1 \end{bmatrix}$ is the line-of-sight vector of the target point relative to the world coordinate system, γ_c^x is the projection value of the target point on the x-axis of the camera coordinate system, γ_c^y is the projection value of the target point on the y-axis of the camera coordinate system, and γ_c^z is the projection value of the target point on the z-axis of the camera coordinate system.

[0049] At block 207, coordinates of an origin of the camera in the world coordinate system are obtained according to the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system.

[0050] The coordinates of the origin of the camera in the world coordinate system is

$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} H_w^c(1,4) \\ H_w^c(2,4) \\ H_w^c(3,4) \end{bmatrix}$, where $H_w^c(1,4)$ is a projection value of the origin of the camera on the x-axis of the world coordinate system, $H_w^c(2,4)$ is a projection value of the origin of the camera on the y-axis of the world coordinate system, and $H_w^c(3,4)$ is a projection value of the origin of the

camera on the x-axis of the world coordinate system.

[0051] At block 208, a line-of-sight linear equation is obtained according to the coordinates of the origin of the camera in the world coordinate system and the line-of-sight vector of the target point relative to the world coordinate system.

[0052] The line-of-sight linear equation is
$$\begin{cases} x = x_c + r_w^x t \\ y = y_c + r_w^y t \\ z = z_c + r_w^z t \end{cases}$$
, where x, y, and z are coordinates of points on the line-of-sight linear equation, r_w^x is a projection value of the target point on the x-axis of the world coordinate system, r_w^y is a projection value of the target point on the y-axis of the world coordinate system, r_w^z is a projection value of the target point on the z-axis of the world coordinate system, t is a dependent variable, and x_c , y_c , and z_c are the coordinates of the origin of the camera in the world coordinate system.

[0053] At block 209, a planar equation of a plane where the target point is located is obtained in the world coordinate system.

[0054] Taking FIG. 3 as an example, the plane where the target point is located is the plane where the mouth of the teacup is located, and its plane equation is $Z=H+h$. H is the height of the top surface of the workbench in the world coordinate system, and h is the height of the plane where the mouth of the teacup is located in the world coordinate system.

[0055] At block 210, world coordinates of the target point are obtained according to the line-of-sight linear equation and the planar equation.

[0056] Continuing to take FIG. 3 as an example, the line-of-sight linear equation in block 208 and the plane equation are combined, to finally obtain the world coordinates of the target

point:
$$\begin{cases} x = x_c + \frac{r_w^x}{r_w^z} (H + h - z_c) \\ y = y_c + \frac{r_w^y}{r_w^z} (H + h - z_c) \\ z = H + h \end{cases}.$$

[0057] At block 211, the mechanical arm is controlled, according to the world coordinates of

the target point, to operate the target object according to a preset strategy.

[0058] The preset strategy can be set by the staff according to actual needs, such as grabbing the target point, etc. Taking a teacup as an example, the preset strategy can be grabbing the teacup or filling the teacup with water.

[0059] According to the mechanical arm control method based on monocular vision positioning of the implementations of this application, the pixel coordinates of the target point of the target object in the pixel coordinate system are obtained. The homogeneous transformation matrix of the camera coordinate system relative to the end of the mechanical arm and the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system are used for transformation, to obtain the line-of-sight linear equation. In the world coordinate system, the plane equation of the plane where the target point is located is obtained. According to the line-of-sight linear equation and the plane equation, the world coordinates of the target point are obtained. Thus, compared with the prior art, in the implementations of this application, there is no need to set two monocular cameras, which can achieve a simplified positioning system and convenient operations. Furthermore, comparison of image feature points and calculation of an optimal distance are not involved, which simplifies the processing algorithm and reduces requirements for hardware equipment.

[0060] Furthermore, as an implementation of the method illustrated in FIG. 1, implementations of this application provide a mechanical arm control device based on monocular vision positioning. As illustrated in FIG. 4, the device includes the following.

[0061] The device includes a coordinate system establishing module 401 configured to establish a pixel coordinate system, a camera coordinate system, and a world coordinate system; a first obtaining module 402 configured to obtain pixel coordinates of a target point of a target object in the pixel coordinate system, an intrinsic parameter matrix of a camera, an extrinsic parameter matrix of the camera, and a homogeneous transformation matrix of an end of a mechanical arm relative to the world coordinate system; a line-of-sight vector determining module 403 configured to obtain a line-of-sight vector of the target point relative to the world coordinate system according to the pixel coordinates of the target point, the intrinsic parameter matrix, the extrinsic parameter matrix, and the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system; a line-of-sight linear equation determining module 404 configured to obtain a line-of-sight linear equation according to a

homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system and the line-of-sight vector of the target point relative to the world coordinate system; a second obtaining module 405 configured to obtain, in the world coordinate system, a planar equation of a plane where the target point is located; a world coordinates determining module 406 configured to obtain world coordinates of the target point according to the line-of-sight linear equation and the planar equation; and a control module 407 configured to control, according to the world coordinates of the target point, the mechanical arm to operate the target object according to a preset strategy.

[0062] Furthermore, as an implementation of the method illustrated in FIG. 2, other implementations of this application provide a mechanical arm control device based on monocular vision positioning. As illustrated in FIG. 5, the device includes the following.

[0063] The device includes a coordinate system establishing module 501 configured to establish a pixel coordinate system, a camera coordinate system, and a world coordinate system; a first obtaining module 502 configured to obtain pixel coordinates of a target point of a target object in the pixel coordinate system, an intrinsic parameter matrix of a camera, an extrinsic parameter matrix of the camera, and a homogeneous transformation matrix of an end of a mechanical arm relative to the world coordinate system; a line-of-sight vector determining module 503 configured to obtain a line-of-sight vector of the target point relative to the world coordinate system according to the pixel coordinates of the target point, the intrinsic parameter matrix, the extrinsic parameter matrix, and the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system; a line-of-sight linear equation determining module 504 configured to obtain a line-of-sight linear equation according to a homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system and the line-of-sight vector of the target point relative to the world coordinate system; a second obtaining module 505 configured to obtain, in the world coordinate system, a planar equation of a plane where the target point is located; a world coordinates determining module 506 configured to obtain world coordinates of the target point according to the line-of-sight linear equation and the planar equation; and a control module 507 configured to control, according to the world coordinates of the target point, the mechanical arm to operate the target object according to a preset strategy.

[0064] Furthermore, the line-of-sight vector determining module 503 includes a line-of-sight

vector determining unit 5031 configured to obtain a line-of-sight vector of the target point relative to the camera coordinate system according to the pixel coordinates of the target point and the intrinsic parameter matrix, where a direction of the line-of-sight vector is from an origin of the camera coordinate system to a center of the target object; and a line-of-sight vector transforming unit 5032 configured to obtain the line-of-sight vector of the target point relative to the world coordinate system according to the extrinsic parameter matrix, the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system, and the line-of-sight vector of the target point relative to the camera coordinate system.

[0065] Furthermore, the line-of-sight vector determining unit 5031 includes a 3D coordinate expression determining sub-unit 50311 configured to obtain a 3D coordinate expression of the target point in the camera coordinate system according to the pixel coordinates of the target point and the intrinsic parameter matrix, where the 3D coordinate expression of the target point in the

camera coordinate system is $s \begin{bmatrix} \mu \\ \nu \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & \alpha \\ 0 & f_y & \beta \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ and where $\begin{bmatrix} \mu \\ \nu \\ 1 \end{bmatrix}$ is the pixel coordinates of

the target point, s is a homogeneous transformation matching coefficient, $\begin{bmatrix} f_x & 0 & \alpha \\ 0 & f_y & \beta \\ 0 & 0 & 1 \end{bmatrix}$ is the intrinsic parameter matrix, f_x is a focal length coefficient of an x-axis of a pixel plane, f_y is a focal length coefficient of a y-axis of the pixel plane, α is an offset of the x-axis of the pixel

plane, β is an offset of the y-axis of the pixel plane, and $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ is 3D coordinates of the target point in the camera coordinate system; and a line-of-sight vector calculating sub-unit 50312 configured to obtain the line-of-sight vector of the target point relative to the camera coordinate system according to the 3D coordinate expression of the target point in the camera coordinate system, where the line-of-sight vector of the target point relative to the camera coordinate system is

$$\begin{bmatrix} \gamma_c^x \\ \gamma_c^y \\ \gamma_c^z \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{\mu / f_x}{\sqrt{\left(\frac{\mu}{f_x}\right)^2 + \left(\frac{\nu}{f_y}\right)^2 + 1}} \\ \frac{\nu / f_y}{\sqrt{\left(\frac{\mu}{f_x}\right)^2 + \left(\frac{\nu}{f_y}\right)^2 + 1}} \\ 1 \\ 1 \end{bmatrix}, \text{ where } \gamma_c^x \text{ is a projection value of the target point on an x-axis of}$$

the camera coordinate system, γ_c^y is a projection value of the target point on a y-axis of the camera coordinate system, γ_c^z is a projection value of the target point on a z-axis of the camera coordinate system, f_x is the focal length coefficient of the x-axis of the pixel plane, f_y is the focal length coefficient of the y-axis of the pixel plane, and μ and ν are the pixel coordinates of the target point.

[0066] Furthermore, the line-of-sight vector transforming unit 5032 includes a homogeneous transformation matrix determining sub-unit 50321 configured to obtain the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system according to the extrinsic parameter matrix and the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system and according to the following formula: $H_w^c = H_w^e \times H_e^c$, where H_w^c is the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system, H_w^e is the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system, and H_e^c is the extrinsic parameter matrix; and a line-of-sight vector transforming sub-unit 50322 configured to obtain the line-of-sight vector of the target point relative to the world coordinate system according to the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system and the line-of-sight vector of the target point relative to

$$\begin{bmatrix} r_w^x \\ r_w^y \\ r_w^z \\ 1 \end{bmatrix} = H_w^c \begin{bmatrix} r_c^x \\ r_c^y \\ r_c^z \\ 1 \end{bmatrix},$$

the camera coordinate system and according to the following formula:

$\begin{bmatrix} r_w^x \\ r_w^y \\ r_w^z \\ 1 \end{bmatrix}$ is the line-of-sight vector of the target point relative to the world coordinate system, r_w^x is a projection value of the target point on an x-axis of the world coordinate system, r_w^y is a projection value of the target point on a y-axis of the world coordinate system, r_w^z is a projection value of the target point on a z-axis of the world coordinate system, H_w^c is the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system, $\begin{bmatrix} r_c^x \\ r_c^y \\ r_c^z \\ 1 \end{bmatrix}$ is the line-of-sight vector of the target point relative to the camera coordinate system, r_c^x is a projection value of the target point on an x-axis of the camera coordinate system, r_c^y is a projection value of the target point on a y-axis of the camera coordinate system, and r_c^z is a projection value of the target point on a z-axis of the camera coordinate system.

[0067] Furthermore, the line-of-sight linear equation determining module 504 includes a coordinates of origin of camera determining unit 5041 configured to obtain coordinates of an origin of the camera in the world coordinate system according to the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system.

[0068] The coordinates of the origin of the camera in the world coordinate system is

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} H_w^c(1,4) \\ H_w^c(2,4) \\ H_w^c(3,4) \end{bmatrix},$$
 where $H_w^c(1,4)$ is a projection value of the origin of the camera on an x-axis of the world coordinate system, $H_w^c(2,4)$ is a projection value of the origin of the camera on the y-axis of the world coordinate system, and $H_w^c(3,4)$ is a projection value of the origin of the camera on the z-axis of the world coordinate system.

[0069] The line-of-sight linear equation determining module 504 further includes a line-of-sight linear equation determining unit 5042 configured to obtain the line-of-sight linear equation according to the coordinates of the origin of the camera in the world coordinate system and the line-of-sight vector of the target point relative to the world coordinate system.

$$\begin{cases} x = x_c + r_w^x t \\ y = y_c + r_w^y t \\ z = z_c + r_w^z t \end{cases},$$
 where x, y, and z are coordinates of points on the line-of-sight linear equation, r_w^x is a projection value of the target point on the x-axis of the world coordinate system, r_w^y is a projection value of the target point on a y-axis of the world coordinate system, r_w^z is a projection value of the target point on a z-axis of the world coordinate system, t is a dependent variable, and x_c , y_c , and z_c are the coordinates of the origin of the camera in the world coordinate system.

[0071] A storage medium is provided according to implementations of this application. The storage medium stores at least one executable instruction. The at least one executable instruction is configured to execute the mechanical arm control method based on monocular vision positioning of any of the above method implementations. Optionally, the storage medium involved in this application may be a computer-readable storage medium. The storage medium, such as the computer-readable storage medium, may be non-volatile or volatile.

[0072] FIG. 6 illustrates a schematic structural diagram of a computer device provided in implementations of this application. The specific implementation of the computer device is not

limited in the implementations of this application.

[0073] As illustrated in FIG. 6, the computer device may include: a processor 602, a communication interface 604, a memory 606, and a communication bus 608. The processor 602, the communication interface 604, and the memory 606 communicate with each other through the communication bus 608.

[0074] The communication interface 604 is used to communicate with other devices, for example, network elements such as clients or other servers.

[0075] The processor 602 is configured to execute programs 610, and specifically can execute the relevant operations in the foregoing implementations of the mechanical arm control method based on monocular vision positioning.

[0076] Specifically, the programs 610 may include program codes, and the program codes include computer operation instructions.

[0077] The processor 602 may be a central processing unit (CPU), or an application specific integrated circuit (ASIC), or one or more integrated circuits configured to implement the implementations of this application. The computer device includes one or more processors. The one or more processors may be processors of the same type, such as one or more CPUs; or processors of different types, such as one or more CPUs and one or more ASICs.

[0078] The memory 606 is used to store the programs 610. The memory 606 may include a high-speed random access memory (RAM) memory, and may also include a non-volatile memory, for example, at least one magnetic disk memory.

[0079] The programs 610 may be specifically used to cause the processor 602 to perform the following operations. A pixel coordinate system, a camera coordinate system at the end of the mechanical arm, and a world coordinate system are established. Pixel coordinates of a target point of a target object in the pixel coordinate system, an intrinsic parameter matrix of the camera, an extrinsic parameter matrix of the camera, and a homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system are obtained. The target object is an object operated by the mechanical arm, and the target point is a specific point on the target object, generally a center point of the target object. The intrinsic parameter matrix and the extrinsic parameter matrix can be obtained by camera calibration. A line-of-sight vector of the target point relative to the camera coordinate system is obtained according to the pixel coordinates of the target point and the intrinsic parameter matrix, where a direction of the line-

of-sight vector is from an origin of the camera coordinate system to a center of the target object. A line-of-sight vector of the target point relative to the world coordinate system is obtained according to the extrinsic parameter matrix, the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system, and the line-of-sight vector of the target point relative to the camera coordinate system. A line-of-sight linear equation is obtained according to a homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system and the line-of-sight vector of the target point relative to the world coordinate system. A planar equation of a plane where the target point is located is obtained in the world coordinate system. World coordinates of the target point are obtained according to the line-of-sight linear equation and the planar equation. The mechanical arm is controlled, according to the world coordinates of the target point, to operate the target object according to a preset strategy.

[0080] Obviously, those skilled in the art should understand that the aforementioned modules or operations of this application can be implemented by a general computing device. They can be integrated in a single computing device or distributed on a network composed of multiple computing devices. Optionally, they can be implemented with program codes executable by the computing device. Thus, they can be stored in the storage device and executed by the computing device. In some cases, the operations illustrated or described can be performed in a different order than here. Alternatively, they are separately made into respective integrated circuit modules. Alternatively, multiple modules or operations of them can be made into a single integrated circuit module. In this way, this application is not limited to any specific combination of hardware and software.

[0081] The above are merely the preferred implementations of this application and are not used to limit this application. For those skilled in the art, various modifications and changes can be made to this application. Any modification, equalization, improvement, etc. made within the spirit and principle of this application shall be included in the scope of protection of this application.

What is claimed is:

1. A mechanical arm control method based on monocular vision positioning, implemented in a mechanical arm control system based on monocular vision positioning, the mechanical arm control system comprising a mechanical arm and a monocular camera installed at an end of the mechanical arm, the method comprising:

establishing a pixel coordinate system, a camera coordinate system, and a world coordinate system;

obtaining pixel coordinates of a target point of a target object in the pixel coordinate system, an intrinsic parameter matrix of the camera, an extrinsic parameter matrix of the camera, and a homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system;

obtaining a line-of-sight vector of the target point relative to the world coordinate system according to the pixel coordinates of the target point, the intrinsic parameter matrix, the extrinsic parameter matrix, and the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system;

obtaining a line-of-sight linear equation according to a homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system and the line-of-sight vector of the target point relative to the world coordinate system;

obtaining, in the world coordinate system, a planar equation of a plane where the target point is located;

obtaining world coordinates of the target point according to the line-of-sight linear equation and the planar equation; and

controlling, according to the world coordinates of the target point, the mechanical arm to operate the target object according to a preset strategy.

2. The method of claim 1, wherein obtaining the line-of-sight vector of the target point relative to the world coordinate system according to the pixel coordinates of the target point, the intrinsic parameter matrix, the extrinsic parameter matrix, and the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system comprises:

obtaining a line-of-sight vector of the target point relative to the camera coordinate system according to the pixel coordinates of the target point and the intrinsic parameter matrix,

wherein a direction of the line-of-sight vector is from an origin of the camera coordinate system to a center of the target object; and

obtaining the line-of-sight vector of the target point relative to the world coordinate system according to the extrinsic parameter matrix, the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system, and the line-of-sight vector of the target point relative to the camera coordinate system.

3. The method of claim 2, wherein obtaining the line-of-sight vector of the target point relative to the camera coordinate system according to the pixel coordinates of the target point and the intrinsic parameter matrix comprises:

obtaining a three-dimensional (3D) coordinate expression of the target point in the camera coordinate system according to the pixel coordinates of the target point and the intrinsic parameter matrix, wherein the 3D coordinate expression of the target point in the camera

coordinate system is $s \begin{bmatrix} \mu \\ \nu \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & \alpha \\ 0 & f_y & \beta \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ and wherein $\begin{bmatrix} \mu \\ \nu \\ 1 \end{bmatrix}$ is the pixel coordinates of the

target point, s is a homogeneous transformation matching coefficient, $\begin{bmatrix} f_x & 0 & \alpha \\ 0 & f_y & \beta \\ 0 & 0 & 1 \end{bmatrix}$ is the

intrinsic parameter matrix, f_x is a focal length coefficient of an x-axis of a pixel plane, f_y is a focal length coefficient of a y-axis of the pixel plane, α is an offset of the x-axis of the pixel

plane, β is an offset of the y-axis of the pixel plane, and $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ is 3D coordinates of the target point in the camera coordinate system; and

obtaining the line-of-sight vector of the target point relative to the camera coordinate system according to the 3D coordinate expression of the target point in the camera coordinate system, wherein the line-of-sight vector of the target point relative to the camera coordinate

$$\begin{bmatrix} \gamma_c^x \\ \gamma_c^y \\ \gamma_c^z \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{\mu / f_x}{\sqrt{\left(\frac{\mu}{f_x}\right)^2 + \left(\frac{\nu}{f_y}\right)^2 + 1}} \\ \frac{\nu / f_y}{\sqrt{\left(\frac{\mu}{f_x}\right)^2 + \left(\frac{\nu}{f_y}\right)^2 + 1}} \\ 1 \\ 1 \end{bmatrix},$$

system is

wherein γ_c^x is a projection value of the target point on an x-axis of the camera coordinate system, γ_c^y is a projection value of the target point on a y-axis of the camera coordinate system, γ_c^z is a projection value of the target point on a z-axis of the camera coordinate system, f_x is the focal length coefficient of the x-axis of the pixel plane, f_y is the focal length coefficient of the y-axis of the pixel plane, and μ and ν are the pixel coordinates of the target point.

4. The method of claim 2, wherein obtaining the line-of-sight vector of the target point relative to the world coordinate system according to the extrinsic parameter matrix, the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system, and the line-of-sight vector of the target point relative to the camera coordinate system comprises:

obtaining the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system according to a homogeneous transformation matrix of the camera coordinate system relative to the end of the mechanical arm and the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system and according to the following formula,

$$H_w^c = H_w^e \times H_e^c,$$

wherein H_w^c is the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system, H_w^e is the homogeneous transformation matrix of the

end of the mechanical arm relative to the world coordinate system, and H_e^c is the extrinsic parameter matrix; and

obtaining the line-of-sight vector of the target point relative to the world coordinate system according to the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system and the line-of-sight vector of the target point relative to the camera coordinate system and according to the following formula,

$$\begin{bmatrix} r_w^x \\ r_w^y \\ r_w^z \\ 1 \end{bmatrix} = H_w^c \begin{bmatrix} r_c^x \\ r_c^y \\ r_c^z \\ 1 \end{bmatrix},$$

wherein $\begin{bmatrix} r_w^x \\ r_w^y \\ r_w^z \\ 1 \end{bmatrix}$ is the line-of-sight vector of the target point relative to the world coordinate system, r_w^x is a projection value of the target point on an x-axis of the world coordinate system, r_w^y is a projection value of the target point on a y-axis of the world coordinate system, r_w^z is a projection value of the target point on a z-axis of the world coordinate system, H_w^c is the homogeneous transformation matrix of the camera coordinate

system relative to the world coordinate system, $\begin{bmatrix} r_c^x \\ r_c^y \\ r_c^z \\ 1 \end{bmatrix}$ is the line-of-sight vector of the target point relative to the world coordinate system, r_c^x is a projection value of the target point on an x-axis of the camera coordinate system, r_c^y is a projection value of the target point on a y-axis of the camera coordinate system, and r_c^z is a projection value of the target point on a z-axis of the camera coordinate system.

5. The method of claim 1, wherein obtaining the line-of-sight linear equation according to the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system and the line-of-sight vector of the target point relative to the world coordinate system comprises:

obtaining coordinates of an origin of the camera in the world coordinate system according to the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system; and

obtaining the line-of-sight linear equation according to the coordinates of the origin of the camera in the world coordinate system and the line-of-sight vector of the target point relative to the world coordinate system.

6. The method of claim 5, wherein the coordinates of the origin of the camera in the world coordinate system is

$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} H_w^c(1,4) \\ H_w^c(2,4) \\ H_w^c(3,4) \end{bmatrix}$, wherein $H_w^c(1,4)$ is a projection value of the origin of the camera on an x-axis of the world coordinate system, $H_w^c(2,4)$ is a projection value of the origin of the camera on the y-axis of the world coordinate system, and $H_w^c(3,4)$ is a projection value of the origin of the camera on the z-axis of the world coordinate system; and

the line-of-sight linear equation is $\begin{cases} x = x_c + r_w^x t \\ y = y_c + r_w^y t \\ z = z_c + r_w^z t \end{cases}$, wherein x, y, and z are coordinates of points on the line-of-sight linear equation, r_w^x is a projection value of the target point on the x-axis of the world coordinate system, r_w^y is a projection value of the target point on a y-axis of the world coordinate system, r_w^z is a projection value of the target point on a z-axis of the world

coordinate system, t is a dependent variable, and x_c , y_c , and z_c are the coordinates of the origin of the camera in the world coordinate system.

7. A mechanical arm control device based on monocular vision positioning:
- a coordinate system establishing module configured to establish a pixel coordinate system, a camera coordinate system, and a world coordinate system;
 - a first obtaining module configured to obtain pixel coordinates of a target point of a target object in the pixel coordinate system, an intrinsic parameter matrix of a camera, an extrinsic parameter matrix of the camera, and a homogeneous transformation matrix of an end of a mechanical arm relative to the world coordinate system;
 - a line-of-sight vector determining module configured to obtain a line-of-sight vector of the target point relative to the world coordinate system according to the pixel coordinates of the target point, the intrinsic parameter matrix, the extrinsic parameter matrix, and the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system;
 - a line-of-sight linear equation determining module configured to obtain a line-of-sight linear equation according to a homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system and the line-of-sight vector of the target point relative to the world coordinate system;
 - a second obtaining module configured to obtain, in the world coordinate system, a planar equation of a plane where the target point is located;
 - a world coordinates determining module configured to obtain world coordinates of the target point according to the line-of-sight linear equation and the planar equation; and
 - a control module configured to control, according to the world coordinates of the target point, the mechanical arm to operate the target object according to a preset strategy.

8. The device of claim 7, wherein the line-of-sight vector determining module comprises:
- a line-of-sight vector determining unit configured to obtain a line-of-sight vector of the target point relative to the camera coordinate system according to the pixel coordinates of the target point and the intrinsic parameter matrix, wherein a direction of the line-of-sight vector is from an origin of the camera coordinate system to a center of the target object; and
 - a line-of-sight vector transforming unit configured to obtain the line-of-sight vector of the

target point relative to the world coordinate system according to the extrinsic parameter matrix, the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system, and the line-of-sight vector of the target point relative to the camera coordinate system.

9. A storage medium storing at least one executable instruction, which causes a processor to execute:

- establishing a pixel coordinate system, a camera coordinate system, and a world coordinate system;

- obtaining pixel coordinates of a target point of a target object in the pixel coordinate system, an intrinsic parameter matrix of a camera, an extrinsic parameter matrix of the camera, and a homogeneous transformation matrix of an end of a mechanical arm relative to the world coordinate system;

- obtaining a line-of-sight vector of the target point relative to the world coordinate system according to the pixel coordinates of the target point, the intrinsic parameter matrix, the extrinsic parameter matrix, and the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system;

- obtaining a line-of-sight linear equation according to a homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system and the line-of-sight vector of the target point relative to the world coordinate system;

- obtaining, in the world coordinate system, a planar equation of a plane where the target point is located;

- obtaining world coordinates of the target point according to the line-of-sight linear equation and the planar equation; and

- controlling, according to the world coordinates of the target point, the mechanical arm to operate the target object according to a preset strategy.

10. The storage medium of claim 9, wherein the at least one executable instruction causing the processor to execute obtaining the line-of-sight vector of the target point relative to the world coordinate system according to the pixel coordinates of the target point, the intrinsic parameter matrix, the extrinsic parameter matrix, and the homogeneous transformation matrix of the end of

the mechanical arm relative to the world coordinate system causes the processor to execute:

obtaining a line-of-sight vector of the target point relative to the camera coordinate system according to the pixel coordinates of the target point and the intrinsic parameter matrix, wherein a direction of the line-of-sight vector is from an origin of the camera coordinate system to a center of the target object; and

obtaining the line-of-sight vector of the target point relative to the world coordinate system according to the extrinsic parameter matrix, the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system, and the line-of-sight vector of the target point relative to the camera coordinate system.

11. The storage medium of claim 10, wherein the at least one executable instruction causing the processor to execute obtaining the line-of-sight vector of the target point relative to the camera coordinate system according to the pixel coordinates of the target point and the intrinsic parameter matrix causes the processor to execute:

obtaining a three-dimensional (3D) coordinate expression of the target point in the camera coordinate system according to the pixel coordinates of the target point and the intrinsic parameter matrix, wherein the 3D coordinate expression of the target point in the camera

coordinate system is $s \begin{bmatrix} \mu \\ \nu \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & \alpha \\ 0 & f_y & \beta \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ and wherein $\begin{bmatrix} \mu \\ \nu \\ 1 \end{bmatrix}$ is the pixel coordinates of the

target point, s is a homogeneous transformation matching coefficient, $\begin{bmatrix} f_x & 0 & \alpha \\ 0 & f_y & \beta \\ 0 & 0 & 1 \end{bmatrix}$ is the

intrinsic parameter matrix, f_x is a focal length coefficient of an x-axis of a pixel plane, f_y is a focal length coefficient of a y-axis of the pixel plane, α is an offset of the x-axis of the pixel

plane, β is an offset of the y-axis of the pixel plane, and $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ is 3D coordinates of the target point in the camera coordinate system; and

obtaining the line-of-sight vector of the target point relative to the camera coordinate

system according to the 3D coordinate expression of the target point in the camera coordinate system, wherein the line-of-sight vector of the target point relative to the camera coordinate

system is

$$\begin{bmatrix} \gamma_c^x \\ \gamma_c^y \\ \gamma_c^z \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{\mu / f_x}{\sqrt{\left(\frac{\mu}{f_x}\right)^2 + \left(\frac{\nu}{f_y}\right)^2 + 1}} \\ \frac{\nu / f_y}{\sqrt{\left(\frac{\mu}{f_x}\right)^2 + \left(\frac{\nu}{f_y}\right)^2 + 1}} \\ 1 \\ 1 \end{bmatrix},$$

wherein γ_c^x is a projection value of the target point on an x-axis of the camera coordinate system, γ_c^y is a projection value of the target point on a y-axis of the camera coordinate system, γ_c^z is a projection value of the target point on a z-axis of the camera coordinate system, f_x is the focal length coefficient of the x-axis of the pixel plane, f_y is the focal length coefficient of the y-axis of the pixel plane, and μ and ν are the pixel coordinates of the target point.

12. The storage medium of claim 10, wherein the at least one executable instruction causing the processor to execute obtaining the line-of-sight vector of the target point relative to the world coordinate system according to the extrinsic parameter matrix, the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system, and the line-of-sight vector of the target point relative to the camera coordinate system causes the processor to execute:

obtaining the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system according to a homogeneous transformation matrix of the camera coordinate system relative to the end of the mechanical arm and the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system and according to the following formula,

$$H_w^c = H_w^e \times H_e^c,$$

wherein H_w^c is the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system, H_w^e is the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system, and H_e^c is the extrinsic parameter matrix; and

obtaining the line-of-sight vector of the target point relative to the world coordinate system according to the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system and the line-of-sight vector of the target point relative to the camera coordinate system and according to the following formula,

$$\begin{bmatrix} r_w^x \\ r_w^y \\ r_w^z \\ 1 \end{bmatrix} = H_w^c \begin{bmatrix} r_c^x \\ r_c^y \\ r_c^z \\ 1 \end{bmatrix},$$

wherein $\begin{bmatrix} r_w^x \\ r_w^y \\ r_w^z \\ 1 \end{bmatrix}$ is the line-of-sight vector of the target point relative to the world coordinate system, r_w^x is a projection value of the target point on an x-axis of the world coordinate system, r_w^y is a projection value of the target point on a y-axis of the world coordinate system, r_w^z is a projection value of the target point on a z-axis of the world coordinate system, H_w^c is the homogeneous transformation matrix of the camera coordinate

system relative to the world coordinate system, $\begin{bmatrix} r_c^x \\ r_c^y \\ r_c^z \\ 1 \end{bmatrix}$ is the line-of-sight vector of the target point relative to the camera coordinate system, r_c^x is a projection value of the target point on an x-

axis of the camera coordinate system, γ_c^y is a projection value of the target point on a y-axis of the camera coordinate system, and γ_c^z is a projection value of the target point on a z-axis of the camera coordinate system.

13. The storage medium of claim 9, wherein the at least one executable instruction causing the processor to execute obtaining the line-of-sight linear equation according to the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system and the line-of-sight vector of the target point relative to the world coordinate system causes the processor to execute:

obtaining coordinates of an origin of the camera in the world coordinate system according to the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system; and

obtaining the line-of-sight linear equation according to the coordinates of the origin of the camera in the world coordinate system and the line-of-sight vector of the target point relative to the world coordinate system.

14. The storage medium of claim 13, wherein the coordinates of the origin of the camera in the world coordinate system is

$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} H_w^c(1,4) \\ H_w^c(2,4) \\ H_w^c(3,4) \end{bmatrix}$, wherein $H_w^c(1,4)$ is a projection value of the origin of the camera on an x-axis of the world coordinate system, $H_w^c(2,4)$ is a projection value of the origin of the camera on the y-axis of the world coordinate system, and $H_w^c(3,4)$ is a projection value of the origin of the camera on the z-axis of the world coordinate system; and

the line-of-sight linear equation is
$$\begin{cases} x = x_c + r_w^x t \\ y = y_c + r_w^y t \\ z = z_c + r_w^z t \end{cases}$$
, wherein x, y, and z are coordinates of

points on the line-of-sight linear equation, r_w^x is a projection value of the target point on the x-axis of the world coordinate system, r_w^y is a projection value of the target point on a y-axis of the world coordinate system, r_w^z is a projection value of the target point on a z-axis of the world coordinate system, t is a dependent variable, and x_c , y_c , and z_c are the coordinates of the origin of the camera in the world coordinate system.

15. A computer device comprising: a processor, a memory, a communication interface, and a communication bus, wherein the processor, the memory, and the communication interface communicate with each other through the communication bus and the memory is configured to store at least one executable instruction, which causes the processor to execute:

establishing a pixel coordinate system, a camera coordinate system, and a world coordinate system;

obtaining pixel coordinates of a target point of a target object in the pixel coordinate system, an intrinsic parameter matrix of a camera, an extrinsic parameter matrix of the camera, and a homogeneous transformation matrix of an end of a mechanical arm relative to the world coordinate system;

obtaining a line-of-sight vector of the target point relative to the world coordinate system according to the pixel coordinates of the target point, the intrinsic parameter matrix, the extrinsic parameter matrix, and the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system;

obtaining a line-of-sight linear equation according to a homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system and the line-of-sight vector of the target point relative to the world coordinate system;

obtaining, in the world coordinate system, a planar equation of a plane where the target point is located;

obtaining world coordinates of the target point according to the line-of-sight linear equation and the planar equation; and

controlling, according to the world coordinates of the target point, the mechanical arm to operate the target object according to a preset strategy.

16. The computer device of claim 15, wherein the at least one executable instruction causing the processor to execute obtaining the line-of-sight vector of the target point relative to the world coordinate system according to the pixel coordinates of the target point, the intrinsic parameter matrix, the extrinsic parameter matrix, and the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system causes the processor to execute:

obtaining a line-of-sight vector of the target point relative to the camera coordinate system according to the pixel coordinates of the target point and the intrinsic parameter matrix, wherein a direction of the line-of-sight vector is from an origin of the camera coordinate system to a center of the target object; and

obtaining the line-of-sight vector of the target point relative to the world coordinate system according to the extrinsic parameter matrix, the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system, and the line-of-sight vector of the target point relative to the camera coordinate system.

17. The computer device of claim 16, wherein the at least one executable instruction causing the processor to execute obtaining the line-of-sight vector of the target point relative to the camera coordinate system according to the pixel coordinates of the target point and the intrinsic parameter matrix causes the processor to execute:

obtaining a three-dimensional (3D) coordinate expression of the target point in the camera coordinate system according to the pixel coordinates of the target point and the intrinsic parameter matrix, wherein the 3D coordinate expression of the target point in the camera

coordinate system is $s \begin{bmatrix} \mu \\ \nu \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & \alpha \\ 0 & f_y & \beta \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ and wherein $\begin{bmatrix} \mu \\ \nu \\ 1 \end{bmatrix}$ is the pixel coordinates of the

target point, s is a homogeneous transformation matching coefficient, $\begin{bmatrix} f_x & 0 & \alpha \\ 0 & f_y & \beta \\ 0 & 0 & 1 \end{bmatrix}$ is the intrinsic parameter matrix, f_x is a focal length coefficient of an x-axis of a pixel plane, f_y is a focal length coefficient of a y-axis of the pixel plane, α is an offset of the x-axis of the pixel

plane, β is an offset of the y-axis of the pixel plane, and $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ is 3D coordinates of the target point in the camera coordinate system; and

obtaining the line-of-sight vector of the target point relative to the camera coordinate system according to the 3D coordinate expression of the target point in the camera coordinate system, wherein the line-of-sight vector of the target point relative to the camera coordinate

system is

$$\begin{bmatrix} \gamma_c^x \\ \gamma_c^y \\ \gamma_c^z \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{\mu / f_x}{\sqrt{\left(\frac{\mu}{f_x}\right)^2 + \left(\frac{\nu}{f_y}\right)^2 + 1}} \\ \frac{\nu / f_y}{\sqrt{\left(\frac{\mu}{f_x}\right)^2 + \left(\frac{\nu}{f_y}\right)^2 + 1}} \\ 1 \\ 1 \end{bmatrix},$$

wherein γ_c^x is a projection value of the target point on an x-axis of the camera coordinate system, γ_c^y is a projection value of the target point on a y-axis of the camera coordinate system, γ_c^z is a projection value of the target point on a z-axis of the camera coordinate system, f_x is the focal length coefficient of the x-axis of the pixel plane, f_y is the focal length coefficient of the y-axis of the pixel plane, and μ and ν are the pixel coordinates of the target point.

18. The computer device of claim 16, wherein the at least one executable instruction causing the processor to execute obtaining the line-of-sight vector of the target point relative to the world coordinate system according to the extrinsic parameter matrix, the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system, and the line-of-sight vector of the target point relative to the camera coordinate system causes the processor to execute:

obtaining the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system according to a homogeneous transformation matrix of the

camera coordinate system relative to the end of the mechanical arm and the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system and according to the following formula,

$$H_w^c = H_w^e \times H_e^c,$$

wherein H_w^c is the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system, H_w^e is the homogeneous transformation matrix of the end of the mechanical arm relative to the world coordinate system, and H_e^c is the extrinsic parameter matrix; and

obtaining the line-of-sight vector of the target point relative to the world coordinate system according to the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system and the line-of-sight vector of the target point relative to the camera coordinate system and according to the following formula,

$$\begin{bmatrix} r_w^x \\ r_w^y \\ r_w^z \\ 1 \end{bmatrix} = H_w^c \begin{bmatrix} r_c^x \\ r_c^y \\ r_c^z \\ 1 \end{bmatrix},$$

wherein $\begin{bmatrix} r_w^x \\ r_w^y \\ r_w^z \\ 1 \end{bmatrix}$ is the line-of-sight vector of the target point relative to the world coordinate system, r_w^x is a projection value of the target point on an x-axis of the world coordinate system, r_w^y is a projection value of the target point on a y-axis of the world coordinate system, r_w^z is a projection value of the target point on a z-axis of the world coordinate system, H_w^c is the homogeneous transformation matrix of the camera coordinate

system relative to the world coordinate system, $\begin{bmatrix} r_c^x \\ r_c^y \\ r_c^z \\ 1 \end{bmatrix}$ is the line-of-sight vector of the target point relative to the world coordinate system, γ_c^x is a projection value of the target point on an x-axis of the camera coordinate system, γ_c^y is a projection value of the target point on a y-axis of the camera coordinate system, and γ_c^z is a projection value of the target point on a z-axis of the camera coordinate system.

19. The computer device of claim 15, wherein the at least one executable instruction causing the processor to execute obtaining the line-of-sight linear equation according to the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system and the line-of-sight vector of the target point relative to the world coordinate system causes the processor to execute:

obtaining coordinates of an origin of the camera in the world coordinate system according to the homogeneous transformation matrix of the camera coordinate system relative to the world coordinate system; and

obtaining the line-of-sight linear equation according to the coordinates of the origin of the camera in the world coordinate system and the line-of-sight vector of the target point relative to the world coordinate system.

20. The computer device of claim 19, wherein

the coordinates of the origin of the camera in the world coordinate system is

$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} H_w^c(1,4) \\ H_w^c(2,4) \\ H_w^c(3,4) \end{bmatrix}$, wherein $H_w^c(1,4)$ is a projection value of the origin of the camera on an x-axis of the world coordinate system, $H_w^c(2,4)$ is a projection value of the origin of the camera on the y-axis of the world coordinate system, and $H_w^c(3,4)$ is a projection value of the origin of the

camera on the x-axis of the world coordinate system; and

$$\text{the line-of-sight linear equation is } \begin{cases} x = x_c + r_w^x t \\ y = y_c + r_w^y t \\ z = z_c + r_w^z t \end{cases}, \text{ wherein } x, y, \text{ and } z \text{ are coordinates of}$$

points on the line-of-sight linear equation, r_w^x is a projection value of the target point on the x-axis of the world coordinate system, r_w^y is a projection value of the target point on a y-axis of the world coordinate system, r_w^z is a projection value of the target point on a z-axis of the world coordinate system, t is a dependent variable, and x_c , y_c , and z_c are the coordinates of the origin of the camera in the world coordinate system.