IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 13, 2020, accepted January 25, 2020, date of publication February 3, 2020, date of current version February 12, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2971349

Design and Test of a Sorting Device Based on Machine Vision

KAI ZHOU[1,2], ZHIWEI MENG[1], MINGYANG HE[1], JIALIN HOU[1,2], AND TIANHUA LI[1,2]
[1]College of Mechanical and Electronic Engineering, Shandong Agricultural University, Tai'an 271018, China
[2]Shandong Provincial Key Laboratory of Horticultural Machinery and Equipment, Tai'an 271018, China

Corresponding author: Tianhua Li (lth5460@163.com)

This work was supported in part by the Project funded by China Postdoctoral Science Foundation under Grant 2019M662410, in part by the Key Research and Development Program of Shandong Province under Grant 2019JZZY020620, in part by the National Key Research and Development Program of China under Grant 2017YFD0701500, and in part by the National Characteristic Vegetable Industry Technology System Project under Grant CARS-24-D-01.

ABSTRACT At present, the sorting of agricultural products in China mainly relies on manual labour, which results in low efficiency, and the development of corresponding automatic equipment lags behind. The grasping method based on machine vision has been widely used in industry, and can provide a reference for the automatic sorting of agricultural products. In this paper, an automatic sorting device for agricultural products was designed. The grasping mechanism adopted a 4-degree-of-freedom (4-DOF) manipulator, and the machine vision control system adopted a monocular camera, which can realize the positioning and classification of the grasp-target. First, the geometric model of the manipulator was established, and the kinematics model of the manipulator was established via the Denavit-Hartenberg (D-H) parameter method. Next, the kinematics analysis and verification were carried out. Then, Zhang Zhengyou calibration method was used to calibrate the camera. An image processing method based on histogram correction was proposed. Based on this, a target positioning algorithm based on the pinhole imaging principle and a target classification algorithm based on the area threshold were established. Finally, an automatic sorting test platform for agricultural products using a visual servo was built. Target classification, positioning and sorting tests were conducted using tomatoes and oranges as the test objects. The test results show that the success rate of the target positioning is close to 98%, that of the target classification is close to 98% and that of the grasping is close to 95%. Furthermore, the sorting time of a single target object can be as fast as 1 second, which can meet the requirements of automatic sorting for common agricultural products. The automatic sorting device for agricultural products has a simple structure, reliable performance and low costs. The structure and algorithms proposed in this paper are simple, reliable, and highly efficient and thus can easily realize technology transplantation. These relevant methods provide a theoretical reference for the development of an automatic sorting device for agricultural products.

INDEX TERMS Machine vision, manipulator, target positioning, target classification, automatic sorting.

I. INTRODUCTION

Robots are typical mechatronics equipment. It benefits from the latest research results of many subjects, such as micro-electronics and computers, automatic control and driving, and artificial intelligence. In recent years, robots have played increasingly important role in scientific research and engineering applications. It has become a common industry practice to use robots to reduce labour costs and improve operating efficiency [1]–[3]. Due to the continuous developments in science and technology, the application range of robots has expanded from the traditional industrial field to many aspects, such as the service industry and agriculture. However, because of their relatively simple perception and recognition, traditional robots can only be used in some specific environments.

In the field of robotics, a machine vision control system can make a robot move more flexibly and faster [4]. With the developments in computer technology, image processing technology and control theory, a robot machine vision system can integrate visual data into the controller and provide feedback control signals for the robot controller. It will

The associate editor coordinating the review of this manuscript and approving it for publication was Yingxiang Liu.

improve the control performance and make the robot's actions more flexible, including robot assembly, driverless cars, robot tracking, detection systems, etc. [5]–[9]. Due to the advantages of large amounts of information, strong perception and high reliability, vision sensors have been widely used in robot control systems [10]–[12]. Therefore, robots based on machine vision have gradually become one of the active research directions.

Robots have been widely used in industrial automation production processes. And they can improve production efficiency and ensure product accuracy [13]–[15]. For example, industrial robots have been commonly used in industrial production for sorting [16]–[18]. Robots based on machine vision have fewer applications in agriculture, and there is no relatively mature sorting device for agricultural products. At present, sorting agricultural products is labour-intensive, but it mainly relies on manual labour that has with low efficiency and high costs. Compared with industrial robots, agricultural robots have lower precision requirements, and so a simple algorithm can meet its requirements. A simple algorithm can reduce the costs and improve the efficiency via its simple operation [19], [20]. Therefore, the sorting agricultural products sorting devices have broad application prospects and research value.

In this paper, a sorting device was designed that can position and classify for target objects. Camera calibration and image processing were carried out. A target positioning algorithm and a target classification algorithm were established. Then, the test platform was built. Finally, the target classification, target positioning and target sorting tests were carried out, respectively.

II. SYSTEM ANALYSIS AND DESIGN

The sorting device based on a visual servo can be divided into two parts: the hardware system and the software system. As shown in Fig. 1, the hardware system is composed of a manipulator, an Arduino control board, a servo control board, five servos, a CCD camera and an upper computer. The upper computer is the processing centre of the whole system. Its function is to control the Arduino control board, process the images captured by the CCD camera, and calculate the motion parameters of the manipulator. The Arduino control board that connected with the servo control board receives the control signal of the upper computer, and then controls of the servo. The angular deviation of the servo motor includes the return error of 0° and the left and right errors of $\leq 3°$. The servo is a closed-loop control system with a fast response and high precision. The servo of the detection circuit can judge the turning angle using a potentiometer and adjust the servo rotation according to the deviation. The sorting of agricultural products does not require high precision, so there is no need to measure the joint angle again.

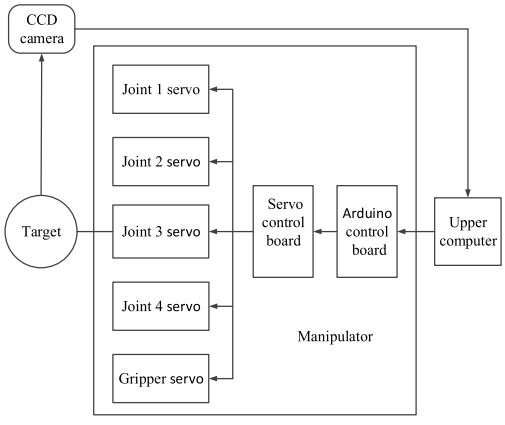


FIGURE 1. Manipulator hardware system.

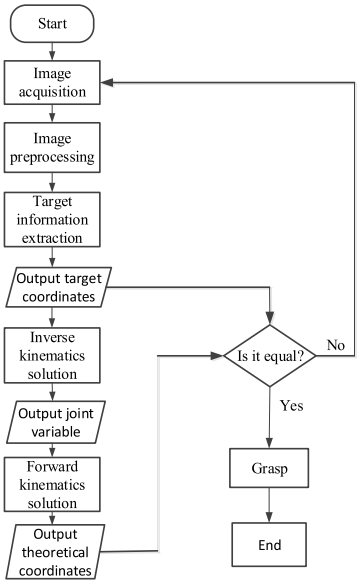


FIGURE 2. Machine vision system grasping process.

The software system can be divided into the three parts: image processing module, motion processing module and pose transformation module. The image processing module is used to obtain the target image and then transmit the target image to the upper computer. The upper computer analyzes the target image in real time, extracts the target position information, and transmits the target position information to the motion processing module. After obtaining the target position information, the motion processing module obtains the joint variable information through the inverse kinematics, and transmits the joint variable information to the pose transformation module. After obtaining the joint variable information, the pose transformation module gives the corresponding servo control information to realize the fixed-point grasping task of the manipulator. The grasping process implemented by the machine vision control system is shown in Fig. 2. After adequate tests, it is shown that the solution of kinematics is reliable, and the consistency between target coordinates and theoretical coordinates is high. So, the grasping operation can be successful almost once.

A. STRUCTURAL DESIGN AND MODELING

The manipulator is mainly composed of the base, the waist, the upper arm, the lower arm, the wrist, the palm and the finger, as shown in Fig. 3.

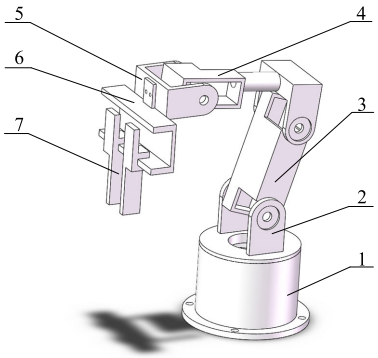


FIGURE 3. The structure of the manipulator.

The D-H parameter method is used to establish the manipulator model. The link coordinate system is shown in Fig. 4, and the D-H parameters are shown in Table 1.

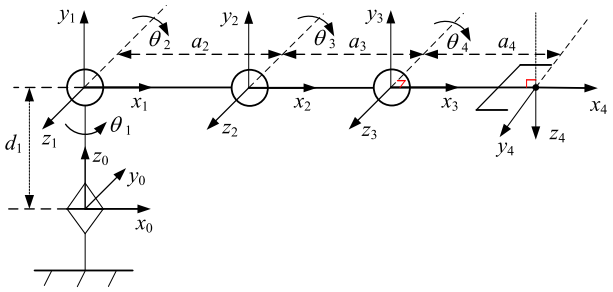


FIGURE 4. Link-pole coordinate system of the manipulator.

TABLE 1. The D-H parameters of the manipulator.

J_i	a_i/mm	d_i/mm	α_i/°	θ_i/°
1	0	95	+90	θ_1
2	80	0	0	θ_2
3	91.92	0	0	θ_3
4	112.36	0	+90	θ_4

In Table 1, a_i represents the length of the vertical line between the z_{i-1} axis and the z_i axis in the x_i direction, d_i represents the distance between the x_{i-1} axis and the x_i axis along the z_{i-1} direction, α_i represents the rotation angle from the z_{i-1} axis to the z_i axis around the x_i axis according to the right hand rotation rule, and θ_i represents the rotation angle from the x_{i-1} axis to the x_i axis around the z_{i-1} axis according to the right hand rotation rule.

B. KINEMATICS ANALYSIS OF THE MANIPULATOR

1) FORWARD KINEMATICS PROBLEM

Kinematics establishes the mapping relationship between the joint variables and the end pose of the manipulator. The forward kinematics problem can obtain the end pose form the joint variables.

After specifying the manipulator link-pole coordinate system and D-H parameters, according to homogeneous transformation theory, the coordinate system $\{O_{i-1}x_{i-1}y_{i-1}z_{i-1}\}$ and the coordinate system $\{O_ix_iy_iz_i\}$ can achieve the transformation via the four times homogeneous transformation matrix.

${}^{i}\mathbf{T}_j$ is the transformation matrix from the coordinate system $\{O_ix_iy_iz_i\}$ to the coordinate system $\{O_jx_jy_jz_j\}$.The forward kinematics equation can be obtained by combining and simplifying the problem as follows:

$$ {}^{0}\mathbf{T}_4 = {}^{0}\mathbf{T}_1 \cdot {}^{1}\mathbf{T}_2 \cdot {}^{2}\mathbf{T}_3 \cdot {}^{3}\mathbf{T}_4 = \begin{bmatrix} n_x & o_x & a_x & p_x \\ n_y & o_y & a_y & p_y \\ n_z & o_z & a_z & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix} = \text{fkine}(\Theta) \tag{1} $$

Here, we can get the following equations from equation (1), (2) and (3), shown at the bottom of this page.

2) INVERSE KINEMATICS PROBLEM

According to the orientation of the end-effector, the joint variables can be inversely solved, which is expressed as equation (4).

$$ \Theta = \text{ikine}\ ({}^{0}\mathbf{T}_4) \tag{4} $$

The solution of the forward kinematics problem is unique, and the solution of the inverse kinematics problem is not unique, or it may even not exist. The inverse problem kinematics is usually solved via the algebraic method.

According to equation (1), θ_1 can be obtained by $[o_x o_y]^T$, and substituting $[o_x o_y]^T$ into $[a_x a_z]^T$ can obtain $\theta_{2+}\theta_{3+}\theta_4$. θ_3 can be obtained by substituting $[\theta_1\theta_{2+}\theta_{3+}\theta_4]^T$ into $[p_x p_z]^T$. The inverse matrix $({}^{0}\mathbf{T}_1)^{-1}$ is left multiplied by ${}^{0}\mathbf{T}_4 = {}^{0}\mathbf{T}_1 \cdot {}^{1}\mathbf{T}_2 \cdot {}^{2}\mathbf{T}_3 \cdot {}^{3}\mathbf{T}_4$, and substituting into $[\theta_1\theta_3]^T$ can get

$$ \begin{bmatrix} n_x & o_x & a_x \\ n_y & o_y & a_y \\ n_z & o_z & a_z \end{bmatrix} = \begin{bmatrix} \cos\theta_1\cos(\theta_2+\theta_3+\theta_4) & \sin\theta_1 & \cos\theta_1\sin(\theta_2+\theta_3+\theta_4) \\ \sin\theta_1\cos(\theta_2+\theta_3+\theta_4) & -\cos\theta_1 & \sin\theta_1\sin(\theta_2+\theta_3+\theta_4) \\ \sin(\theta_2+\theta_3+\theta_4) & 0 & -\cos(\theta_2+\theta_3+\theta_4) \end{bmatrix} \tag{2} $$

$$ \begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix} = \begin{bmatrix} \cos\theta_1[112.36\cos(\theta_2+\theta_3+\theta_4)+91.92\cos(\theta_2+\theta_3)+80\cos\theta_2] \\ \sin\theta_1[112.36\cos(\theta_2+\theta_3+\theta_4)+91.92\cos(\theta_2+\theta_3)+80\cos\theta_2] \\ 112.36\sin(\theta_2+\theta_3+\theta_4)+91.92\sin(\theta_2+\theta_3)+80\sin\theta_2+95 \end{bmatrix} \tag{3} $$

θ_2 because the corresponding elements are equal. Substituting $[\theta_2\theta_3]^{\mathrm{T}}$ into $[a_{\mathrm{x}}n_{\mathrm{x}}]^{\mathrm{T}}$ can get θ_4. Finally, the equation of the inverse solution can be obtained as follows:

$$\theta_1 = \arctan2(o_{\mathrm{x}}, o_{\mathrm{y}}) \tag{5}$$

$$\theta_2 = -\arctan2(A, B) + \arctan2(A^2 + B^2 + a_2^2 - a_3^2, 2a_2 \times \sqrt{A^2 + B^2 - (\frac{A^2 + B^2 + a_2^2 - a_3^2}{2a_2})^2}) \tag{6}$$

$$\theta_3 = -\arctan2(\sqrt{4a_2^2a_3^2 - (A^2 + B^2 - a_2^2 - a_3^2)^2}, A^2 + B^2 - a_2^2 - a_3^2) \tag{7}$$

$$\theta_4 = \arctan2(a_{\mathrm{x}}, n_{\mathrm{x}}) - \theta_2 - \theta_3 \tag{8}$$

Here, we get the following:

$$\begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} p_{\mathrm{x}} \cos\theta_1 + p_{\mathrm{y}} \sin\theta_1 - a_4 \\ p_{\mathrm{z}} - d_1 \end{bmatrix} \tag{9}$$

3) KINEMATICS VERIFICATION

Using the manipulator model established above, the joint values are changed, and the relationship between the coordinates and the joint variables is verified using the random locations. The results are as follows,

Taking the joint variable as $\boldsymbol{\theta} = [-33\ 42\ \ -47\ 29]^{\mathrm{T}}$, and the transformation matrix $\mathbf{T}$ is

$$\mathbf{T} = \begin{bmatrix} 0.7662 & -0.5446 & 0.3411 & 212.1991 \\ -0.4976 & -0.8387 & -0.0015 & -138.9960 \\ 0.4067 & 0 & -0.9135 & 186.2200 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

The orientation under the random joint variable is shown in Fig. 5.

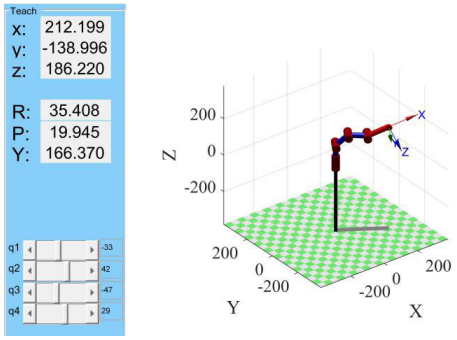


FIGURE 5. The orientation under random joint variables.

After substituting the end pose matrix into the analytical solution formula of the inverse kinematics, the results of the inverse solution are $\theta_1 = -33.9972°$, $\theta_2 = 42.0988°$, $\theta_3 = -47.1202°$, and $\theta_4 = 29.0193°$.

After randomly setting the joint variable as $\theta = [13\ 42\ 26\ 47]^{\mathrm{T}}$, the correctness shall also be verified. The results of the inverse solution are $\theta_1 = 13.0234°$, $\theta_2 = 42.1136°$, $\theta_3 = 25.9654°$, and $\theta_4 = 47.1365°$.

C. CAMERA CALIBRATION AND IMAGE PROCESSING

1) CAMERA CALIBRATION

a: IMAGING PRINCIPLE

An image captured by the camera is the reflection projection of a three-dimensional space object on a two-dimensional plane, namely, the optical centre projection, or pinhole imaging. Assuming that all reflected light strictly follows the theory of the rectilinear propagation of light, the whole model shall consist of the optical centre o, the imaging plane π_1 and the optical axis oO. The image point on the imaging plane and the point on the object are collinear with the optical centre. To facilitate the representation and calculation, four coordinate systems are defined, which are the world coordinate system $\{O_{\mathrm{W}}X_{\mathrm{W}}Y_{\mathrm{W}}Z_{\mathrm{W}}\}$, the camera coordinate system $\{oxyz\}$, the pixel coordinate system $\{O_{\mathrm{f}}X_{\mathrm{f}}Y_{\mathrm{f}}\}$ and the image coordinate system $\{OXY\}$. For convenience, move the final projection plane to the front of the optical centre. The relationship between the four coordinate systems is shown in Fig. 6.

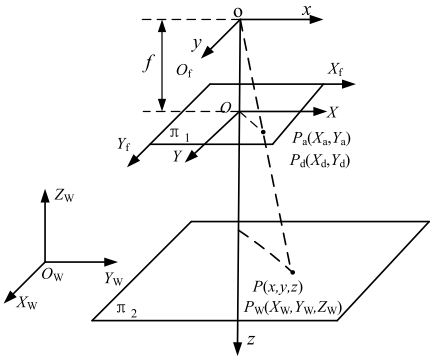


FIGURE 6. Relationship between the four coordinate systems.

In practical applications, due to the limitation of the lens design, light is deflected in the lens, and it causes lens distortion. If the theoretical imaging point is $(X_{\mathrm{f}}, Y_{\mathrm{f}})$, the final imaging point caused by the camera distortion is $(X_{\mathrm{d}}, Y_{\mathrm{d}})$. For ordinary CCD cameras, the calibration accuracy can be satisfied by considering only the first-order radial distortion. The mathematical model between the theoretical imaging point and the final imaging point is as follows:

$$\begin{aligned} X_{\mathrm{f}} &= X_{\mathrm{d}}(1 + k_1 r_{\mathrm{d}}^2) \\ Y_{\mathrm{f}} &= Y_{\mathrm{d}}(1 + k_1 r_{\mathrm{d}}^2) \end{aligned} \tag{10}$$

Here, k_1 is the first-order radial distortion coefficient of the camera, r_{d} is the distortion radius and $r_{\mathrm{d}} = (X_{\mathrm{d}}^2 + Y_{\mathrm{d}}^2)^{1/2}$.

b: CAMERA PARAMETERS

The imaging projection process of a point P on a three-dimensional space object to the image point p involves three transformations of the above four coordinate systems. First, the world coordinate system is transformed into the camera coordinate system via a homogeneous transformation. Then, the camera coordinate system is transformed into the

image coordinate system according to the geometry. Finally, using the conversion relationship between pixels and other units, the coordinates in the final pixel coordinate system are obtained. The transformation from the world coordinate system to the pixel coordinate system is as follows:

$$\begin{bmatrix} X_f \\ Y_f \\ 1 \end{bmatrix} = \frac{1}{z}\begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{O}_{3\times1} & 1 \end{bmatrix}\begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} = \frac{1}{z}\mathbf{M}_{in}\mathbf{M}_{out}\begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \quad (11)$$

Equation (11) is a mathematical model of the pinhole imaging principle, including 10 camera parameters. Here, the effective focal length f represents the vertical distance between the projected imaging plane and the optical centre. f_x and f_y represent the number of pixels of the actual focal length f in the x-axis and y-axis directions, respectively, that is, the equivalent focal length in pixels. The principal point, which is also called the image centre, is the pixel coordinates of the intersection of the camera optical axis and the projected imaging plane, and it is (u_0, v_0). The variables f_x, f_y, u_0 and v_0 are the internal parameters of the camera. The transformation matrix $\mathbf{R}$ is the orthogonal rotation matrix, and $\mathbf{T}$ represents the position coordinate of the origin of world coordinate system in the camera coordinate system. Rotation matrix $\mathbf{R}$ contains three orientation independent variables, and translation matrix $\mathbf{T}$ contains three position independent variables. These six variables are the external parameters of the camera, and they determine the position of the camera in the world coordinate system.

When the internal parameters are known, the external parameters can be solved using the several given spatial coordinates and pixel coordinates. The established corresponding relationship is uniquely determined by the internal parameters of the camera, and it is only related to the camera itself.

c: CAMERA CALIBRATION TEST

The camera calibration adopts the Zhang Zhengyou calibration method, which uses the different pose relationships of the calibration board in a three-dimensional space to construct the camera parameter equation. Then, the camera parameters can be obtained by the equation [21], [22].

The camera position is fixed and the pose of the camera calibration board in the three-dimensional space is changed for the sampling. The camera calibration boards are sampled, and the sampling results are shown in Fig. 7.

The pre-sampled calibration board samples are imported, and the sample to be calibrated will be obtained. The corners are positioned for each image. Then, the camera parameters are solved, and the final results of the camera calibration can

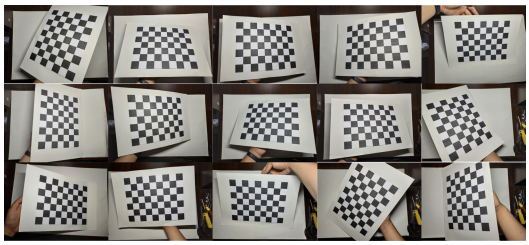


FIGURE 7. The sampling of the camera calibration board.

be obtained as follows:

$$\begin{aligned} fc &= [3095.269;\ 3097.026] \\ cc &= [1980.104;\ 1485.479] \\ alpha &= 0.000; \\ kc &= [-0.002;\ -0.030;\ 0.003;\ 0.000;\ 0.000] \end{aligned}$$

Equation (12) is the internal parameter matrix of the camera.

$$\mathbf{M}_{in} = \begin{bmatrix} 3095.269 & 0 & 1980.104 \\ 0 & 3097.026 & 1485.479 \\ 0 & 0 & 1 \end{bmatrix} \quad (12)$$

The z-axis of the world coordinate system is collinear with the z-axis of the camera coordinate system, and the x-axis and the y-axis of the world coordinate system are parallel to the x-axis and the y-axis of the camera coordinate system, respectively. The orientation $\mathbf{R}$ in the external parameter matrix is obtained, which is the third-order identity matrix $\mathbf{E}$. The position vector is measured by the vertical height of the camera's optical centre.

2) IMAGE PROCESSING

An image acquired by the visual system contains interference, which cause noise and distortion in the captured image. Noise and distortion greatly affect the extraction of the final image information and result in random interference. Therefore, noise and distortion must be removed before image processing.

In the case of insufficient exposure or overexposure, the pixel greyscale values in the image will be highly concentrated within a certain range, resulting in blurred details and low contrast. At this time, the greyscale transformation enhancement theory is used to solve this question. The greyscale value of each point of the input image is mapped to another grey function through a mapping relationship, and the original pixel position is kept unchanged. A histogram is a statistical graph that reflects the relationship between the greyscale level in an image and its appearance probability, and it represents the distribution of the pixel brightness in the image. The sharpness, contrast and brightness of an image can be judged by analysing the histogram. For images lacking detailed information, the histogram is corrected to a uniform histogram to increase the image contrast [23], [24].

In the image generation or transmission process, there are some interferences, which make the image contain much noise. Median filtering theory is used to enhance the low frequency component and suppress the high frequency noise. After the target image information is obtained, useful information is extracted and useless information is removed. Image segmentation is used to divide the image into several disjoint regions to extract the useful regions. After the useful area is extracted, the target object is detected in real time. According to the prominent characteristics of the greyscale level of the target object and the surrounding environment, the target will be captured, and the edges will be detected.

D. TARGET POSITIONING AND TARGET CLASSIFICATION

1) TARGET POSITIONING

Two points P_1 and P_2 in a three-dimensional space form a vector $\mathbf{P_1P_2} = (x_2\text{-}x_1, y_2\text{-}y_1, z_2\text{-}z_1)$. Assume that the general equation of plane π is $Ax + By + Cz + D = 0$, and the normal vector of the plane is $\mathbf{n} = (A, B, C)$. If there is a coefficient λ, the coordinates of point H can be determined via λ. The following formula can be obtained from the geometric relation,

$$\mathbf{OH} = \mathbf{OP1} + \mathbf{P1H} = \mathbf{OP1} + \lambda\mathbf{P1P2} \tag{13}$$

According to the point-to-plane distance formula,

$$|\mathbf{P1Q}| = \frac{Ax_1 + By_1 + Cz_1 + D}{\sqrt{A^2 + B^2 + C^2}} \tag{14}$$

In $\Delta P_1P_2Q_2$, because the vector $\mathbf{P_1P_2}$ is parallel to the plane normal vector $\mathbf{n}$, according to the point-to-plane distance formula and the cosine theorem,

$$|\mathbf{P1Q2}| = \frac{|\mathbf{P1P2} \cdot \mathbf{n}|}{|\mathbf{n}|} = \frac{A(x_2 - x_1) + B(y_2 - y_1) + C(z_2 - z_1)}{\sqrt{A^2 + B^2 + C^2}} \tag{15}$$

In $\Delta P_1P_2Q_2$ and ΔP_1HQ, according to equation (14) and equation (15), the λ value is as follows:

$$\lambda = \frac{|\mathbf{P_1Q_2}|}{|\mathbf{P_1Q}|} = \frac{A(x_2 - x_1) + B(y_2 - y_1) + C(z_2 - z_1)}{Ax_1 + By_1 + Cz_1 + D} \tag{16}$$

According to the plane equation and the coordinates of the spatial point, the coordinates of the intersection for any plane and the line outside the plane can be obtained. The intersection of line P_1P_2 and plane π is shown in Fig. 8.

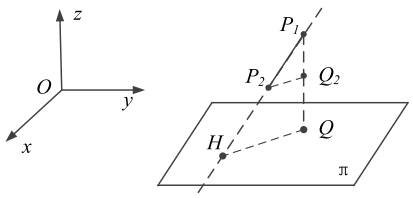


FIGURE 8. Line $\mathbf{P_1P_2}$ intersects plane π.

The light reflected from any point on the object passes through the optical centre, and the projection point can be obtained on the projection plane. The point P on the object, the optical centre o and the projection point p are collinear. Given a space coordinate system, the coordinates of two points, and the target plane, the intersection of the ray and the target plane can be found. The intersection is the position in the final projected image on the actual three-dimensional space [25]. Fig. 9 is a schematic diagram of target positioning.

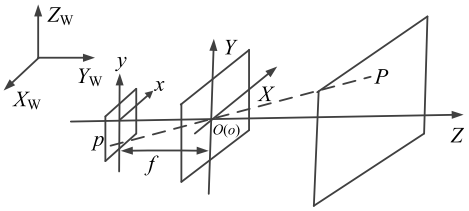


FIGURE 9. Schematic diagram of target positioning.

In the target positioning algorithm, in order to simplify the model, the target is replaced by the centroid P, the coordinates of projection point p and optical centre o are known, and the coordinates of object point P are the final task solution quantities. In actual image processing, the most easily obtained is the pixel coordinates $({}^pX_f, {}^pY_f)$. Therefore, it is necessary to perform coordinate transformation first, so that the coordinates of the point $({}^pX, {}^pY)$ in the image coordinate system can be obtained, which occurs as follows:

$$\begin{bmatrix} {}^pX \\ {}^pY \end{bmatrix} = \begin{bmatrix} d_x({}^pX_f - u_0) \\ d_y({}^pY_f - v_0) \end{bmatrix} \tag{17}$$

If the imaging coordinates are affected by the camera distortion, they can be corrected using the camera distortion model. The theoretical pixel coordinate (X_f, Y_f) is obtained by the correction, and then the coordinate transformation can be carried out via equation 10.

To simplify the coordinate transformation process and make it easy to establish a connection between the world coordinate system and the manipulator coordinate system, the intersection of the camera's optical axis and the object plane in the imaging model is selected as the origin of the world coordinate system, and the object plane is X_WOY_W. At this time, the object's vertical projection plane is also the plane where the point P lies, and the equation of the plane is $z = 0$. As shown in Fig. 10, the distance of the image coordinate system and the camera coordinate system in the z-axis direction is the effective focal length f, and the point $({}^pX, {}^pY, f)$ in the camera coordinate system is obtained. For a vertically fixed camera, the distance between the optical centre and the target plane is fixed. Therefore, the coordinate on the projection plane in the world coordinate system is $({}^pX, {}^pY, f + {}^PZ_W)$, and the coordinate of the camera's optical centre in the world coordinate system is $(0,0, {}^PZ_W)$.

The coefficient λ is as follows:

$$\lambda = \frac{f}{f + {}^PZ_W} \tag{18}$$

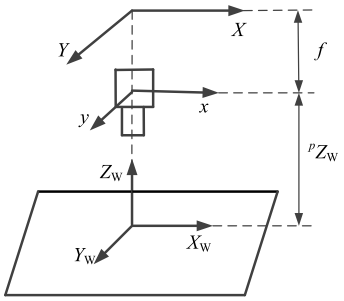


FIGURE 10. Coordinate transformation relationship.

2) TARGET CLASSIFICATION

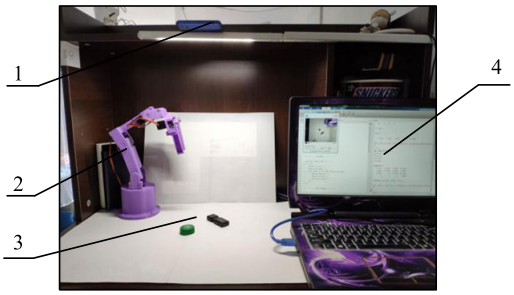
The pattern matching algorithm has a fast recognition speed and a high success rate for a single object. However, the recognition effect for multiple objects is not good, which may cause misjudgements due to the coincidence or similarity of targets. The recognition speed will be too slow due to the repeated matching. For simple object classification and detection in an industrial production process or target detection tasks with low precision requirements, using a pattern matching algorithm will increase the operating time and costs. Therefore, the area threshold method can be used to classify the objects in the common recognition task without considering the target orientation. In actual industrial production, most parts are produced in batches with uniform specifications, and so the area threshold method can be used to classify the target objects. For example, objects with a photosensitive area less than 0.06 can be classified as the first type of object, and those with a photosensitive area greater than 0.06 can be classified as the second type of object.

According to the imaging principle of the camera, there is a one-to-one correspondence between the points on the target object and the target pixel points in the final projected image. The pixel coordinates of the target object image can be converted into actual length coordinates in the world coordinate system. Furthermore, the digitized image is based on the discretization of successive greyscale images, which is advantageous for generating a single pixel block. Therefore, the target area can correspond to the one-sided surface area of the actual object in the world coordinate system. The scale factor d_x and d_y represent the actual size of a single pixel.

In the digital image of the target object, the number of pixels included in the outline range is N, and the actual area of the surface is as follows:

$$S_W = N \cdot {}^{0}S_W = N \cdot d_x \cdot d_y \tag{19}$$

The target area is estimated using the pixel statistical method, polygon fitting method and circumscribed rectangle method. It can be seen from the comparison tests that all three methods can meet the requirements, but the pixel statistics method has the highest efficiency. In addition, the precision of sorting for agricultural products is not high, and so the pixel statistics method is used to estimate the area in this paper. The projected areas of different sized target objects are also different, and the classification of the target objects can be realized based on this. Because the numbers of pixels in the target areas of captured target images are different, the areas of the target objects are different. Regarding the statistics of the number of pixels contained in the target region of a captured target image, it is often necessary to scan the digital matrix of the image row by row, judge individual pixel blocks one by one, classify the judgement results, and then count the number of pixels.



1. Monocular camera 2. 4-DOF manipulator 3. Target identification area
4. Upper computer

FIGURE 11. Machine vision manipulator test platform.

FIGURE 12. Target positioning and grasping test.

III. TEST VERIFICATION

The structure of the manipulator is printed via 3D printing technology, and the test platform is built, as shown in Fig. 11. On the top of the test platform is a monocular camera for capturing the target image. The right side is the upper computer for the calculation, the left side is the 4-DOF manipulator, and the middle desktop area is the target recognition area where the target object is located.

The configuration of the upper computer is as follows: a 64-bit operating system, an i5-5200U central processing unit with a 2.20 GHz CPU clock speed, 4.0 GB of memory, an NVIDIA GeForce 930M video card.

A. TARGET POSITIONING AND GRASPING TEST

A visual platform is built, and the camera is calibrated in advance. Then, the image processing is carried out to obtain the coordinates of the target centre point. Then, the calculated coordinates of the target in the final three-dimensional space are obtained by using the target positioning algorithm. The verification test of the target positioning algorithm is conducted by following these steps.

Image pre-processing is performed on the captured target image, and the target area is scanned row-by-row and column-by-column. Then the pixel position coordinates of the centre of the target object are obtained. The calibrated camera and the coordinates of the captured centre point are verified using the correspondence between the pixel coordinates and the actual three-dimensional space that is proposed in this paper.

It is stipulated that the origin of the world coordinate system is the intersection of the camera's optical axis and the three-dimensional space object plane. The direction of the coordinate axis is the same as the manipulator base coordinate system $\{x_0y_0z_0\}$, that is, the world coordinate system coincides with the manipulator base coordinate system $\{x_0y_0z_0\}$. The camera calibration board with the known actual length information is used to verify the positioning algorithm, and the corner point is used to replace the target. After the origin is determined, the position of the corner point on the calibration board, which is the actual target coordinate, is determined accordingly. The measured target coordinates can be obtained using positioning test of this algorithm. The unit of the coordinate length in the world coordinate system is mm. The calculated results are shown in Table 2.

TABLE 2. Calculated results.

Target actual coordinates	Target calculated coordinates	Error
(28.5,28.5)	(29.1123,29.1451)	(-0.6123,-0.6451)
(-28.5,28.5)	(-27.6652,28.1684)	(-0.8348,0.3316)
(57,-57)	(57.9465,-56.2484)	(-0.9465,-0.7516)
(85.5,-85.5)	(85.6524,-84.5488)	(-0.1524,-0.9512)
(-199.5,199.5)	(-200.2934,200.0645)	(0.7934,-0.5645)
(199.5,199.5)	(198.6486,200.3466)	(0.8514,-0.8466)

The analysis results show that the target positioning algorithm proposed in this paper can map the points of the pixel coordinate system to the actual three-dimensional space, and the error can meet the accuracy requirements. This algorithm can realize the accurate positioning of the target object.

The target positioning and grasping test, which is shown in Fig. 12, mainly assesses the whole machine vision system. Several tomatoes are placed by randomly throwing them into the target recognition area, and the target positioning algorithm is used for the target positioning. The joint variables are obtained by the kinematic equations, and then the manipulator grasps the target objects.

The pixel coordinates of target are obtained by visual device, and the pixel coordinates are converted into the actual coordinates by the transformation relationship. Then, the motion processing module of the upper computer is used to calculate the pose information and joint variables which required by the manipulator.

A total of 66 positioning grasping tests of objects at random locations were carried out, and the test results are shown in Table 3.

For larger tomatoes, there is some unsuccessful grasping because the driving force of the servo is too small, and the mechanical claw does not insufficiently open. There is also some lack of success for small tomatoes. The reason is that there is still more noise after image preprocessing, resulting in image information extraction errors, and the manipulator is not designed to possess high precision. The superposition of multiple errors led to the failure of positioning and grabbing tasks. In the future, the grasping mechanism of the manipulator can be adjusted according to the sizes of agricultural products.

TABLE 3. Results of the target positioning and grasping test.

Target type	Large	Medium	Small
Number of tests	22	22	22
Number of successful positioning	22	22	21
Number of successful grasping	21	22	20
Positioning success rate	100%	100%	95.45%
Grasping success rate	95.45%	100%	90.90%

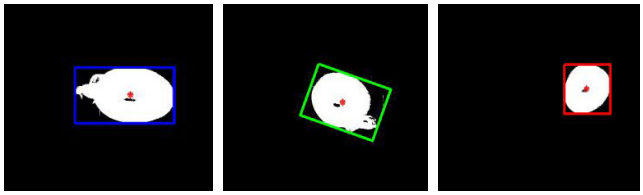


FIGURE 13. Tomato size classification results.

The test results show that for the different sized tomatoes placed at random locations, the positioning success rate is close to 98%, and the grasping success rate after positioning is close to 95%. The success rate can meet the basic requirements of positioning and grasping for the target objects.

B. TARGET CLASSIFICATION TEST

The visual platform is built. The camera calibration is carried out in advance, and then the image preprocessing is performed. The target classification algorithm is used to classify the objects, and the verification tests are carried out. The pixel size of the camera is $d_x = d_y = 1.12\ \mu$m, that is, the actual area of a single pixel in the image coordinate system is $1.12\ \mu\text{m} \times 1.12\ \mu$m. The target classification test assesses the visual image processing system. Different sized tomatoes are placed by randomly throwing them into the target recognition area, and target classification using the area threshold method is performed. The information of the target area is statistically analysed. Because the target objects and the image pixels are small, the pixel statistical method is used here, and the size of the areas are $s_1 = 0.0836$, $s_2 = 0.0638$ and $s_3 = 0.0450$, respectively.

After obtaining the area of the target objects, these objects can be classified. Two area thresholds of $M = 0.07$ and $G = 0.05$ are selected as the basis for the classification. If $s < G$, the target object is a small fruit, and it is marked in red.

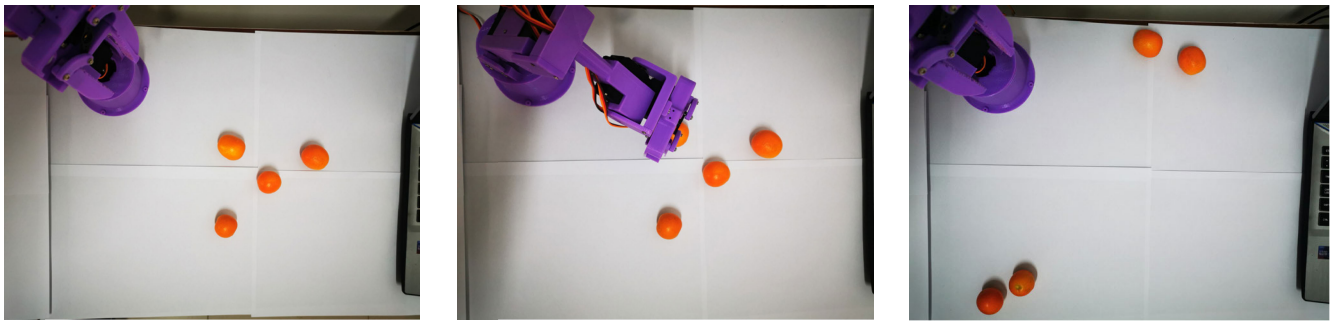


FIGURE 14. Target sorting test.

TABLE 4. Results of the target classification test.

Target type	Large	Medium	Small
Number of tests	22	22	22
Number of successful classifications	22	22	21
Classification success rate	100%	100%	95.45%

If $G < s < M$, the target object is a medium fruit, and it is marked in green. If $s > M$, the target object is a large fruit, and it is marked in blue. They are shown in Fig. 13.

It can be seen from the above results that the visual image processing system can accurately calculate the size of the area. The area threshold method is used to perform 66 classification tests of different target objects. The test results are shown in Table 4.

The test results show that the area-based target classification algorithm can accurately calculate the areas of different target objects, and the classification success rate is close to 98%.

C. TARGET SORTING TEST

The sorting operation mainly includes three processes: positioning, classification and grasping. Several oranges are placed by randomly throwing them into the target recognition area, and the target positioning algorithm and the target classification algorithm are applied. Then, the target objects are positioned and classified, and the manipulator uses the positioning information to grasp the classified target objects, thereby completing the sorting operation. A total of 180 sorting tests were carried out for oranges at random locations. The test site is shown in Fig. 14, and the test results are shown in Table 5.

The test results show that for different sized oranges placed at random locations, the positioning success rate is close to 98%, the grasping success rate is close to 94%, and the sorting success rate is close to 93%. The sorting device has a high sorting success rate and reliable performance. Furthermore, it can meet the requirements of automatic sorting for agricultural products.

TABLE 5. Results of the target sorting test.

Target type	Large	Medium	Small
Number of tests	60	60	60
Number of successful positioning	60	59	57
Number of successful grasping	56	58	55
Number of successful sorting	56	58	54
Positioning success rate	100%	98.33%	95%
Grasping success rate	93.33%	96.67%	91.67%
Sorting success rate	93.33%	96.67%	90%

IV. CONCLUSION

1) An automatic sorting device for agricultural products was designed. The grasping mechanism adopted a 4-DOF manipulator, and the machine vision control system adopted a monocular camera. The device can realize target positioning and classification based on image processing. The automatic sorting device has a simple structure, reliable performance and low costs, and it can realize the positioning, classification and grasping of target objects.

2) The geometric model of the manipulator was established, and then the kinematics model of the manipulator was established via the D-H parameter method. A target positioning algorithm based on the pinhole imaging principle and a target classification algorithm based on the area threshold were established, and these algorithms were verified by tests. The results show that the proposed algorithms can realize target positioning and classification.

3) An automatic sorting test platform for agricultural products based on a visual servo was built. The target classification, positioning and sorting tests were carried out using tomatoes and oranges as the test objects. The test results show that the success rate of target positioning is close to 98%, that of target classification is close to 98% and that of grasping is close to 95%. Furthermore, the sorting time of a single target object can be as fast as 1 second, which can meet the requirements of automatic sorting for common agricultural products.

REFERENCES

[1] M. Zhang, X. Liu, D. Xu, Z. Cao, and J. Yu, "Vision-based target-following guider for mobile robot," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9360–9371, Dec. 2019.

[2] J. Lv, Y. Wang, L. Xu, Y. Gu, L. Zou, B. Yang, and Z. Ma, "A method to obtain the near-large fruit from apple image in orchard for single-arm apple harvesting robot," *Scientia Horticulturae*, vol. 257, Nov. 2019, Art. no. 108758.

[3] F. Gomez-Donoso, S. Orts-Escolano, and M. Cazorla, "Accurate and efficient 3D hand pose regression for robot hand teleoperation using a monocular RGB camera," *Expert Syst. Appl.*, vol. 136, pp. 327–337, Dec. 2019.

[4] H. A. Williams, M. H. Jones, M. Nejati, M. J. Seabright, J. Bell, N. D. Penhall, J. J. Barnett, M. D. Duke, A. J. Scarfe, H. S. Ahn, J. Lim, and B. A. Macdonald, "Robotic kiwifruit harvesting using machine vision, convolutional neural networks, and robotic arms," *Biosyst. Eng.*, vol. 181, pp. 140–156, May 2019.

[5] L.-H. Juang and J.-S. Zhang, "Visual tracking control of humanoid Robot," *IEEE Access*, vol. 7, pp. 29213–29222, 2019.

[6] H. Herrero, A. A. Moughlbay, J. L. Outón, D. Sallé, and K. L. De Ipiña, "Skill based robot programming: Assembly, vision and workspace monitoring skill interaction," *Neurocomputing*, vol. 255, pp. 61–70, Sep. 2017.

[7] S.-C. Chen, "Multimedia for autonomous driving," *IEEE Multimedia-Mag.*, vol. 26, no. 3, pp. 5–8, Jul. 2019.

[8] N. Ando, S. Emoto, and R. Kanzaki, "Insect-controlled robot: A mobile robot platform to evaluate the odor-tracking capability of an insect," *J. Vis. Exp.*, vol. 118, Dec. 2016.

[9] I. Ciric, Z. Cojbasic, D. Ristic-Durrant, V. Nikolic, M. Ciric, M. Simonovic, and I. Pavlovic, "Thermal vision based intelligent system for human detection and tracking in mobile robot control system," *Therm. Sci.*, vol. 20, no. 5, pp. 1553–1559, 2016.

[10] S.-Q. Ji, M.-B. Huang, and H.-P. Huang, "Robot intelligent grasp of unknown objects based on multi-sensor information," *Sensors*, vol. 19, no. 7, p. 1595, Apr. 2019.

[11] M. Heydarzadeh, N. Karbasizadeh, M. T. Masouleh, and A. Kalhor, "Experimental kinematic identification and position control of a 3-DOF decoupled parallel robot," *Proc. Inst. Mech. Eng., C, J. Mech. Eng. Sci.*, vol. 233, no. 5, pp. 1841–1855, Mar. 2019.

[12] Z. Shangguan, L. Wang, J. Zhang, and W. Dong, "Vision-based object recognition and precise localization for space body control," *Int. J. Aerosp. Eng.*, vol. 2019, pp. 1–10, Mar. 2019.

[13] Q. Wei, C. Yang, W. Fan, and Y. Zhao, "Design of demonstration-driven assembling manipulator," *Appl. Sci.*, vol. 8, no. 5, p. 797, May 2018.

[14] L. H. Juang and S. X. Zhang, "Intelligent service robot vision control using embedded system," *Intell. Autom. Soft Comput.*, vol. 25, no. 3, pp. 451–458, Sep. 2019.

[15] W. Chen, T. Xu, J. Liu, M. Wang, and D. Zhao, "Picking robot visual servo control based on modified fuzzy neural network sliding mode algorithms," *Electronics*, vol. 8, no. 6, p. 605, May 2019.

[16] X. Wu, X. Ling, and J. Liu, "Location recognition algorithm for vision-based industrial sorting robot via deep learning," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 33, no. 07, Jun. 2019, Art. no. 1955009.

[17] Q. Wu, M. Li, X. Qi, Y. Hu, B. Li, and J. Zhang, "Coordinated control of a dual-arm robot for surgical instrument sorting tasks," *Robot. Auton. Syst.*, vol. 112, pp. 1–12, Feb. 2019.

[18] Y. Lin, H. Zhou, M. Chen, and H. Min, "Automatic sorting system for industrial robot with 3D visual perception and natural language interaction," *Meas. Control*, vol. 52, nos. 1–2, pp. 100–115, Jan. 2019.

[19] E. Van Henten, D. Van Slot, C. Hol, and L. Van Willigenburg, "Optimal manipulator design for a cucumber harvesting robot," *Comput. Electron. Agricult.*, vol. 65, no. 2, pp. 247–257, Mar. 2009.

[20] M. P. Mann, D. Rubinsterin, I. Shmulevich, R. Linker, and B. Zion, "Motion planning of a mobile Cartesian manipulator for optimal of 2-D crops," *Trans. ASABE*, vol. 57, no. 1, pp. 283–295, 2014.

[21] Y. Ji and J. Wu, "Calibration method of light-field camera for photogrammetry application," *Measurement*, vol. 148, Dec. 2019, Art. no. 106943.

[22] J. Jiang, L. Zeng, B. Chen, Y. Lu, and W. Xiong, "An accurate and flexible technique for camera calibration," *Computing*, vol. 101, no. 12, pp. 1971–1988, Dec. 2019.

[23] A. A. Hayat, R. A. Boby, and S. K. Saha, "A geometric approach for kinematic identification of an industrial robot using a monocular camera," *Robot. Comput.-Integr. Manuf.*, vol. 57, pp. 329–346, Jun. 2019.

[24] S. Kansal and S. Mukherjee, "Automatic single-view monocular camera calibration-based object manipulation using novel dexterous multi-fingered delta robot," *Neural Comput. Appl.*, vol. 31, no. 7, pp. 2661–2678, Jul. 2019.

[25] L. E. Carvalho and A. Von Wangenheim, "3D object recognition and classification: A systematic literature review," *Pattern Anal. Appl.*, vol. 22, no. 4, pp. 1243–1292, Nov. 2019.

KAI ZHOU was born in Tai'an, China, in 1987. He received the B.S. degree in mechanical engineering from the Shandong University of Science and Technology, in 2010, and the M.S. and Ph.D. degrees from the School of Mechatronics Engineering, Harbin Institute of Technology, in 2012 and 2017, respectively. His research interests include intelligent agricultural equipment, agricultural robots, and intelligent manufacturing.

ZHIWEI MENG was born in Jinan, China, in 1996. He received the B.S. degree in vehicle engineering from Shandong Jianzhu University, in 2014. He is currently pursuing the M.S. degree in mechanical engineering with Shandong Agricultural University. His research interests include intelligent agricultural equipment and agricultural robots.

MINGYANG HE was born in Zaozhuang, China, in 1996. He is currently pursuing the B.S. degree in mechanical engineering with Shandong Agricultural University. His research interests include intelligent agricultural equipment and agricultural robot.

JIALIN HOU was born in Weifang, China, in 1963. He received the B.S. degree in mechatronics engineering from Shandong Agricultural University, in 1987, and the M.S. and Ph.D. degrees from the School of Agricultural Electrification and Automation, China Agricultural University, in 2002 and 2005, respectively. His research interests include intelligent agricultural equipment and agricultural machinery design, and theoretical research.

TIANHUA LI was born in Tai'an, China, in 1976. He received the B.S. degree in applied electronic technology from the Shandong University of Technology, in 2000, the M.S. degree in control theory and control engineering from Henan Polytechnic University, in 2003, and the Ph.D. degree in agricultural mechanization engineering from Shandong Agricultural University, in 2015. His research interests include intelligent agricultural equipment, agricultural robot, and agriculture automation.