The Acquisition of Kiwifruit Feature Point Coordinates Based on the Spatial Coordinates of Image

Bin Wang, Zixiao Chen, Jianmin Gao, Longsheng Fu, Baofeng Su, Yongjie Cui

The Acquisition of Kiwifruit Feature Point Coordinates Based on the Spatial Coordinates of Image

Bin Wang [a], Zixiao Chen, Jianmin Gao, Longsheng Fu, Baofeng Su, Yongjie Cui [b,*]

(College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China)

[a]wangbinwork@nwsuaf.edu.cn, [b]cuiyongjie@nwsuaf.edu.cn

Abstract: How to obtain the spatial coordinates of kiwi fruit has been one of the key techniques for kiwi fruit harvesting robot. In this paper, the writer proposes a unique way to obtain the spatial coordinates of the features of kiwi fruit from the bottom of the target fruit based on the growth characteristics and scaffolding cultivation pattern characteristicsof kiwi fruit, plus the help of Microsoft camera and Kinect sensor. Also included in this paper is the coordinate conversion between the images come from Microsoft camera and the images of the Kinect sensor, which is followed by an analysis of the precision of the spatial coordinates of Kiwi fruit captured by the Microsoft camera and Kinect sensor. The process is like this: first, capture images of the target fruit from the bottom of the fruit with Microsoft camera, and then extract coordinates of the target fruits' feature points to determine the corresponding target fruit feature point coordinates in the Kinect sensor; second, analyze the correspondence between the Microsoft camera image coordinate system and the Kinect sensor image coordinate system so as to establish amathematical model for the image coordinate conversion; finally,capture target feature points' spatial coordinates with Kinect sensorand conduct tests. The results show that the precision of coordinate conversion mode and Kiwifruit spatial coordinates can meet the requirements of the harvesting robots.

Keywords: Kiwi fruit, harvesting robot, image coordinate, mathematical model, Kinect sensor

1 Introduction[1]

The acreage and production of China's kiwi fruit rank first in the world. However, at present, the kiwi fruit is mainly harvested manually, which is highly labor-intensive. With the progress of urbanization and industrialization, more and more young and middle aged people are attracted to work in cities. As a result, the loss of labor force in agriculture is becoming serious, which in turn raises the cost of agricultural production and lowers market competitiveness of our agricultural products. Therefore, the development of kiwi fruit picking robot is of great significance to the development of China's kiwi fruit industry.

The key techniques of Kiwi picking robot involve three parts: fruit identification, location and nondestructive picking. The widespread adoption of standardized scaffolding pattern in kiwi fruit production makes robot picking fruit feasible. However, there are still several factors that hinders the development of kiwi fruit picking robot. Firstly, kiwi plants grow in clusters, each of which is usually composed of 3-5fruits, and the fruits usually grow too close to one another and even overlap. Moreover, foliage sheltering and similar color between fruits and the background make the harvesting robot difficult to perform precise fruit identification and separation as well as feature extraction of fruits. Secondly, kiwi

fruit positioning and spatial coordinate acquisition are also problems to be solved for the development of the harvesting robot. The existing fruit and vegetable harvesting robot positioning system is low in precision, time-consuming, complex in structure and high in cost. So it is imperative to develop a new, efficient positioning system.

Among the existing fruit and vegetable harvesting robots at home and abroad, some can harvest fruits whose colors are greatly different from the background colors, such as strawberry picking robot [1], tomato picking robot [2-4], citrus harvesting robots [5-6], and some can harvest the target fruits whose colors are similar to the background colors, such as cucumber picking robots [7]. Such robots usually adopt near-infrared spectroscopy or laser technology for detection. In terms of detection and identification of kiwi fruits, Zhan Wentian [8]used Adaboot algorithm. Ding Yalan [9] used RB color component method to separate kiwi fruits. These two methods can only separate the regions with fruits from the ones without fruits, but they failed to identify the fruit individually. Cui Yongjie [10] used 0.9R-G color features in the fruit image segmentation, but in a complex background environment, this method involves a large amount of calculation and time. In terms of kiwi fruit positioning and coordinate acquisition, some other methods are used, such as monocular vision, binocular vision, multi-purpose vision, hyper-spectrum, laser etc. However, these methods have drawbacks, such as complex computation, low accuracy, high cost, poor reliability and so on. Meanwhile, the conversion among pixel coordinates, spatial coordinates and mechanical arms is still a problem to be solved.

In on-site investigation, it is found that the places below the fruits are spacious with less sheltering and the background is simple, so the writer proposes that the fruits be identified, positioned, and picked from the bottom parts of the plants. The principle is like this: determine the sequence of fruit identification, feature point extraction and fruit picking by using elliptic Hough conversion; acquire the feature point coordinates of the images with Kinect sensors made by Microsoft Company; obtain the image coordinates of feature points by Kinect sensor referring the foreign research results for Microsoft Kinect sensor in robot navigation [11-12] and feature recognition [13-15]; finally, conduct coordinates conversion between the camera and sensors and construct mathematical model of the coordinates conversion to obtain the 3D coordinates of the feature points.

2 Information Perception

2.1 Feature Extraction and Image Acquisition

The kiwi fruit pictures were taken in October 2014 during harvest time at Kiwi Experimental Station of Northwest Agriculture and Forestry University and the breed is "Hayward". Camera used is Microsoft Life Camera studio with COMS sensor and auto-focus. Image pixel acquisition is of 640 × 360, jpg format. Each picture containing 2-5 fruitswas taken from the bottom with a distance of 20cm to the fruits. Image acquisition mainly comes from the side and bottom, as shown in Figure 1.

(a) Picture shotfrom side(b)Picture shoot frombottom

Fig. 1. Two different shooting directions

In Figure 1, due to the greater scene depth, and complicated background, it can be seen that the images of fruits taken from the side contain not only the leaves of the near-byplant branches, but also distant non-target fruits, and serious mutual occlusion between the target fruits. All of these affect the accuracy of target fruit segmentation and recognition. In contrast, the picture shot from the bottom has less mutual occlusion between fruits and no interference of other distant non-target fruits, which is favorable for extracting target fruits. Due to the mutual occlusion between fruits, fruit

identification can only be performed from outside to inside, picking one by one, which results in low harvest efficiency. In contrast, the shadow area between fruits is less in image shot from the bottom, which makes it possible that all the fruits can be identified at a time. As a result, the picking sequence can be determined and efficiency improved.

In order to improve fruit recognition success rate under complex background, Cui Yongjieet al [16], researchers of Northwest Agriculture and Forestry University, presented a comprehensive method to identify fruits and extract fruit features according to kiwi fruit characteristics and color features, and elliptical Hough conversion. This method can minimize the impact of different complex background and illumination on the identification and the extraction of fruit features. The specific steps are shown in Figure 2.
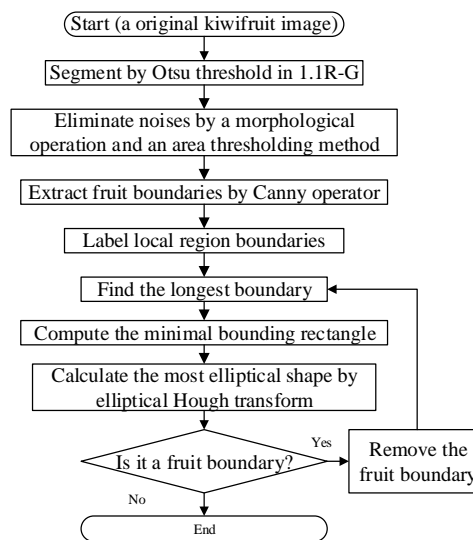

Fig.2. Flow chart of fruit recognition

2.2Picking Order

Figure 3 is the pixel coordinates of the feature points when each fruit is identified. X-axis is within the range of 0-360, Y axis 0-640.
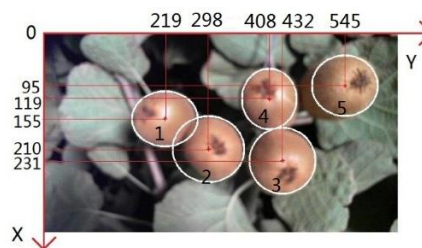

Fig.3. Coordinates of target fruit

The picking sequence is determined according to the values of the feature point coordinates. In figure 3, 1,2,3,4,5 represent picking sequence, which is determined according to the Y coordinate values of the feature points from small to big, thus the picking arm of the robot reaches minimum stroke, and maximum efficiency in the whole picking process.

3Coordinate Conversion

The principle of Kiwi fruit extraction is shown in Fig 4. On the left is a front view and on the right is atop view, where 1 stands for kiwi fruits, 2 for Microsoft camera, and 3 for Kinect sensors. The camera is used to identify fruits, extract features and determine the picking sequence. Kinect sensorsare used to obtain the spatial coordinates of the feature points of kiwi fruits. The intersection between the optical center and the outer surface of the infrared camera is

used as the origin of the coordinates; the intersection between Microsoft camera lens surface and the optical center is used as the coordinate origin of its coordinate system as is shown in Figure 4, point 'O' and point 'E' .
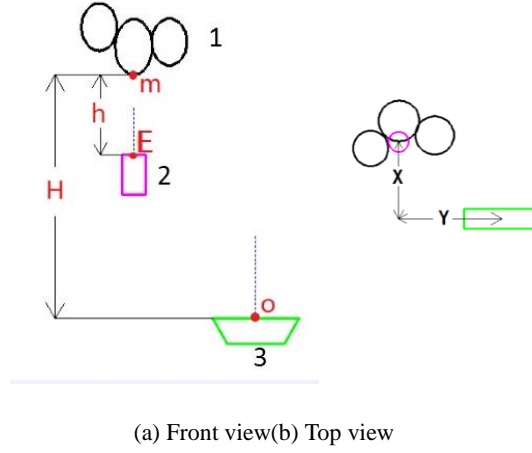

(a) Front view(b) Top view

Fig.4. The Schematic of fruit space coordinates acquisition

As is shown in Fig 4, the horizontal distance between the feature point and the Microsoft camera is 'h'; the distance between the feature point and Kinect sensor is 'H', then the distance between the Microsoft camera and Kinect sensor is 'H-h'. When the distance between the Microsoft camera and Kinect sensor remains unchanged, let the spatial coordinates between the camera and Kinect sensor be (X, Y, Z), then we get Figure 5.
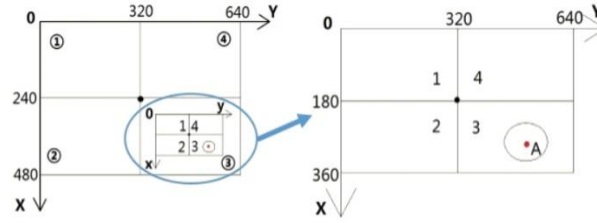

(a) Microsoft camera image coordinate system (b) Kinect sensor image coordinate system

Fig.5. Relationship between two coordinate systems

The coordinate system diagram of Microsoft camera image is shown in Figure 5(a) with a pixel area of 640×360. The value of the center coordinates is (320,180), which is also the projection position of the camera's optical center in the image. Letthe pixel coordinates of the feature point A recognized by the camera at the image be (x, y), so the value of the relative pixel coordinates in the image is ($\Delta x, \Delta y$). Then:

$$\Delta x = x - 320 \quad （1）$$

$$\Delta y = y - 180 \quad （2）$$

The relationship between Δx and Δy is positive and negative, and the image can be divided into our regions, that is, '1', '2', '3', '4'. When$\Delta x < 0$ and $\Delta y < 0$, it corresponds to region '1', representing that the feature point is on the upper left of the projection point. When $\Delta x < 0$and $\Delta y > 0$, it corresponds to region '2', showing that the feature point is on the lower left of the projection point. When $\Delta x > 0$and $\Delta y > 0$, it corresponds to region '3', meaning that the feature point is at the right bottom of the projection point; When $\Delta x > 0$ and $\Delta y < 0$, it corresponds to region '4', indicating that the feature point is on the upper right of the projection point.

When the distance between the camera and the feature point is h, letone of the image pixels be a (mm), then the three 3D coordinates X_w, Y_w, Z_w of the feature pointsrecognized relative to the origin are respectively as follows:

$$X_w = a\Delta x \quad （3）$$

$$Y_w = a\Delta y \quad (4)$$

$$Z_w = h \quad (5)$$

In addition, the positive or negative values of X_w and Y_w determines the regions where the feature points distribute.

The Kinect sensor is installedbelow the Microsoft camera with a fairly great distance between them, so that images acquired by the Kinect sensor include the shooting area of the Microsoft camera. As is shown in Figure 5 (b), 'XOY' is the Kinect image screen coordinates, and $'xoy'$ is the Microsoft camera image screen coordinates. When the pixel range of the Kinect sensor in capturing image is 640×480, the projection of the optical center of the infrared video camera in the image is the pixel coordinates (320×240) of the center point of image projected.

When the spatial position of the feature points recognized by the Microsoft camera remains unchanged, and let the coordinates of feature point A in the screen coordinate system of the Kinect sensor be x', y', then its relative pixel coordinates to the center point in the image are:

$$\Delta x' = x' - 320 \quad (6)$$

$$\Delta y' = y' - 240 \quad (7)$$

Supposing one pixel of the image on plane H of the Kinect sensor represents the actual length b (mm), the 3Dcoordinates X_k, Y_k, Z_k of feature point A relative to the origin of the Kinect sensor are respectively:

$$X_k = b\Delta x' \quad (8)$$

$$Y_k = b\Delta y' \quad (9)$$

$$Z_k = H \quad (10)$$

Similarly, the positive and negative of X_k, Y_k correspond to the locations of four images①②③④ and the actual locations of information. Since the spatial position of the Microsoft camera and Kinect sensors remain unchanged, pixel coordinates of feature point A in the imaged captured by Kinect sensor can be derived from the pixel coordinates of the feature point 'A' in the camera image shot by the Microsoft camera. That is

$$a(\text{x} - 320) + \text{X} = b(\text{x}' - 320) \quad (11)$$

$$a(\text{y} - 180) + \text{Y} = b(\text{y}' - 240) \quad (12)$$

From equations (11) and (12), we can get the following formulas:

$$\text{x}' = \frac{a}{b}(\text{x} - 320) + \frac{X}{b} + 320 \quad (13)$$

$$\text{y}' = \frac{a}{b}(\text{y} - 180) + \frac{Y}{b} + 240 \quad (14)$$

When the values of distance H and h remain unchanged, the values of X, Y, a and b would be constant, so would be the values of $\frac{a}{b}, \frac{X}{b}, \frac{Y}{b}$. When Microsoft camera recognizes and extracts pixel coordinates(x, y) of the feature point of kiwi fruits, the corresponding pixel on the Kinect sensor screen is（x', y'）in theory. However, in practice, image acquired by the obtained by RGB video camera on the Kinect sensor is just opposite along the left-right direction. That is to say, it is reversed along the direction of axis X. In this case, the value of the coordinates (x, y) of the point corresponding to the coordinates (x", y") is $(640 - X', y')$.

4Obtaining Spatial Coordinates

As far as the existing fruit and vegetable harvesting robots at home and abroad are concerned, a variety of methods are adopted for target positioning and coordinates extraction, such as monocular vision, binocular vision, multi-purpose visualization, close infrared spectroscopy, laser scanners etc. but each of them has some problems to be solved.
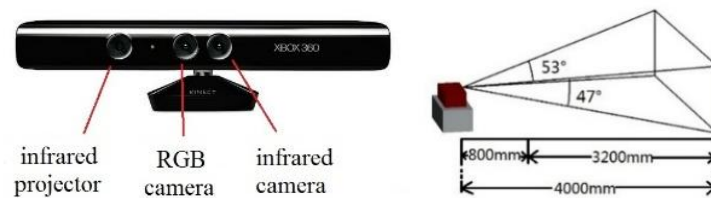
(a) Hardware composition(b) View field

Fig.6. Hardware components and field of view of the Kinect sensor

In this study, Microsoft's Kinect sensor is used to obtain the spatial coordinates of the feature points of the fruits, and the development platform is Kinect for Windows SDK. The sensing device is shown in Figure 6 (a). It mainly consists of three parts. They are, from left to right, an infrared projector, a RGB camera, and an infrared camera. The function of the infrared projector is to project near-infrared spectrum actively. As is known, when the infrared spectrum is projected ontothe objects with rough surfaces or ground glass,there would be distorted spectrum, which in turn would will generate random points of reflected light (also called speckles). The speckles are then read by the infrared camera.The infrared camera is used to analyze the close infrared spectrum and to create depth images of the objects within our vision. RGB camera is used to shoot colored images within our vision. The measurement range is shown in Figure 6 (b).The range centers on the infrared camera with upper angle43° and lower angle 43°, 400-4000mm away in front of the video camera. The precision of the depth images captured within this area can reach millimeter.

The flow chart of Kiwi fruit spatial coordinates acquisition is shown in Fig.7. First, register color image data and depth image data flow to get respective color image and depth image, and to map RGB image onto the depth image.Then load the mapped image into the Kinect built-in spatial coordinates system, flowed by loading Map Depth Point to Skeleton Point function.Afterwards, judge whether the coordinate values are within the range with higher accuracy. If the values are within the range, then output the spatial coordinates with the infrared camera as its origin; if not, the output distance is beyond range.
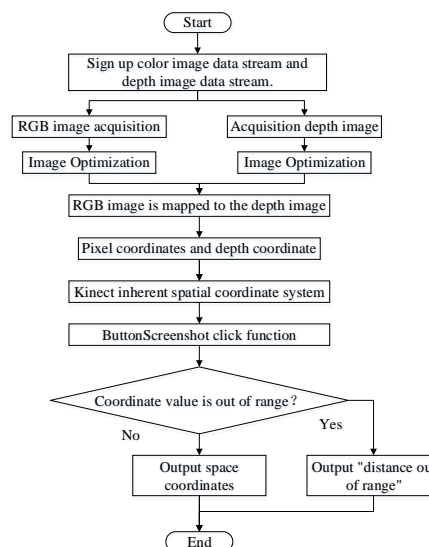
5Test and Analysis

5.1Test Method

In order to verify the conversion relationship between feature point coordinates acquired by Microsoft camera and the coordinates of the Kinect sensor, as well as the accuracy of the mathematical model, verification test was carried out in the laboratory. Firstly, graph paper was used to calibrate the length represented by one pixel (e.g.numerical values of 'a' and 'b') in the images captured by Microsoft camera and Kinect sensor in fixed positions. Then the feature point images and 3D coordinates were acquired with the Microsoft camera and the Kinect sensor.The specific steps are as follows:
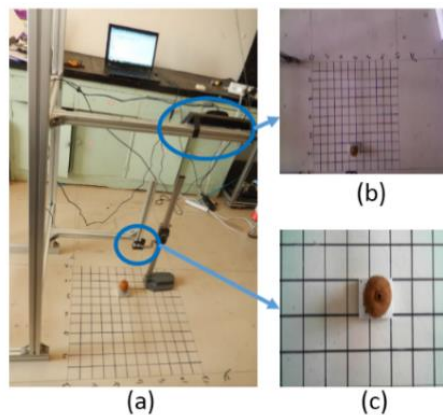
(a) Testing apparatus(b) The image of the Kinect sensor(c) Pictures taken by Microsoft camera.

Fig.8. Test equipment and images acquired

1）In order to facilitate verification testing, the whole coordinate conversion system was inverted: Kiwi fruit was placed at the bottom, and Microsoft camera was placed at the upper part with the Kinect sensor on top. The testing platform constructed is shown in Figure 8 (a).Kinect sensor and Microsoft camera are fixed onto the bracket on the same level with desktop supporting Kiwi fruit. In this test, in order to facilitate verification, the kiwi fruit surface was randomly marked with a point as the recognition feature point. Figure 8 (b) is an image acquired by the sensor and Figure 8 (c) is the image captured by Microsoft camera.

2）In this test, the vertical distance from the Microsoft camera to the feature point is 200mm, and the vertical distance from the Kinect sensor to the feature point is 928mm. Graph paper was used to calibrate the actual length between the image plane and the place 200mm away from the Microsoft camera as well as the actual length of one pixel 928mm away from the Kinect sensor (i.e. the values of a and b). The way of calibration is shown in Figure 9, where Microsoft camera is fastened to the height gauge parallel to the coordinate plane. Kinect sensor calibration method is same with Microsoft camera calibration method.

3）Microsoft camera was used to obtain the images of feature points and the pixel coordinates of the feature point. Kinect sensor was used to obtain the images of feature points, pixel coordinates and spatial coordinates.

4）Equations (13), (14) were verified with the pixel coordinates acquired by Microsoft camera and Kinect sensor and the values of a, b.Errors in coordinate conversion were derived according to the actual pixel coordinates of the feature points while Kinect sensor acquiring images, D-values between the images coordinates derived from equations, D-values between pixel coordinates from Kinect sensor image and calculation, and the actual length represented by each D-value.

5.2 Result and Analysis

Through calibration, it is found that the actual length represented by one pixel 200mm away from at Microsoft camera is 0.445mm, that is to say, a=0.445m and that the actual length represented by one pixel 928mm away from the Kinect sensor is 1.32mm, that is to say, b=1.778. In this experiment,we got the pixel coordinates of 24 groups of feature points in different positions on the images captured by Microsoft camera and the images captured by Kinect sensor. In addition, with the help of Kinect sensor, we got the spatial coordinates of the feature points at such positions.

By plugging into formulas (13), (14) the pixel coordinates of point 1 and point 24 on the images acquired by Microsoft camera and the symmetricalimage of the image acquired by Kinect sensor, we have the following results:

$$\frac{a}{b} = 0.25$$

$$\frac{X}{b} = 32.00$$

$$\frac{Y}{b} = 186.374$$

The value of a/b coincides with the previous calibration. By plugging into formula (13), (14)the values of a/b, X/b, Y/b and the coordinate values of the rest 22 points on the images acquired by the Microsoft camera, we can work out the coordinates of these 22 points on the symmetrical images acquired of the Kinect sensor image.

Table 1. The coordinates of points in different coordinate systems

	(x,y)	(x',y')	(x'',y'')	The coordinates after the formula extrapolated	The spatial coordinates of the feature point coordinates obtained by Kinect (x,y,z)
1	591,42	218,393	422,393	420，391	-214,-247,928
2	594,154	218,423	422,423	421，420	-216,-301,928
3	597,265	218,451	422,451	421，448	-213,-349,928
4	597,324	218,467	422,467	421，462	-214,-375,926
5	481,42	246,393	394,393	392，391	-166,-249,928
6	482,151	246,421	394,421	393，419	-161,-294,928
7	486,265	248,451	392,451	394，448	-158,-349,928
8	484,322	248,465	392,465	393，462	-162,-375,928
9	369,41	276,391	364,391	364，392	-112,-244,928
10	372,153	276,421	364,421	365，420	-112,-299,928
11	373,264	276,449	364,449	365，447	-112,-348,928
12	373,319	276,464	364,464	365，461	-112,-371,926
13	260,38	304,385	336,385	337，391	-62,-242,928
14	260,153	304,419	336,419	337，420	-62,-294,928
15	261,266	304,449	336,449	337，448	-63,-344,928

16	259,321	306,463	334,463	337，462	-62,-369,928
17	144,38	332,389	308,389	308，391	-10,-244,928
18	145,152	334,419	306,419	308，419	-11,-294,928
19	144,266	334,447	306,447	308，448	-8,-343,928
20	144,326	334,463	306,463	208，463	-10,-370,926
21	26,39	362,388	278,388	278，391	41,-240,928
22	25,153	364,417	276,417	278，420	44,-289,928
23	26,268	364,447	276,447	278，448	42,-345,928
24	26,323	364,461	276,461	278，462	40,-364,927

Figure 13 is a diagram drawn with MATLAB to express the calculated coordinate values and the actual coordinate values.
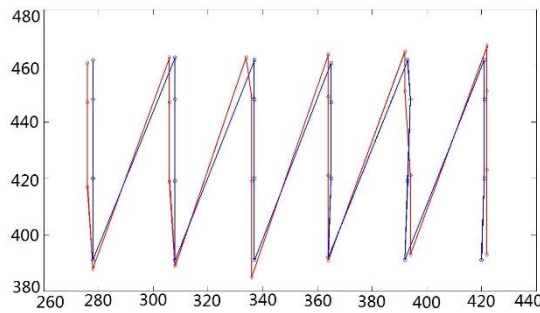

Fig.10. The calculated and actual coordinate distribution curves

In the diagram, the red curves represent the actual distribution of points, and the blue ones represent the distribution of calculated points. It can be seen that the two curves coincide with each other. Table 1 indicates that the error between the calculated point coordinates and the actual coordinates is less than 3 pixels or 5mm, thus formulas (13), (14) can accurately reflect the correspondence of a same point on the image acquired by Microsoft camera and on the image acquired by the Kinect sensor.

6 Conclusions

（1）In light of the kiwi fruit growth characteristics, an automatic identification method is studied, including how to acquire fruit image from the bottom and an integrated application of fruit shape and color features in recognition.
（2）Considering the drawbacks of the existing fruit and vegetable harvest robots, a fresh method based on Kinect sensor is proposed to acquire the coordinates of the target kiwi fruits.
（3）This paper discusses the coordinate conversion between Microsoft camera and Kinect sensor and the mathematical model constructed can perform accurate coordinates conversion.

Acknowledgment

References

1. Zhang Kailiang, Yang Li, Wang Liangju et al. Design and Experiment of Elevated Substrate Culture Strawberry Picking Robot [J]. Transactions of the Chinese Society for Agricultural Machinery, 2012, 43(9):165-172.

2. Monta M, Kondo N, Shibano Y. Agricultural robot in grape production system [A]. IEEE International Conference on Robotics and Automation, 1995, 2504-2509.

3. Arima S, Kondo N. Cucumber harvesting robot and plant training system [J]. Journal of Robotics and Mechatronics, 1999, 11(3): 208-212.

4. Kondo N, Monta M, Fujiura T. Fruit harvesting robot in Japan [J]. Adv Space Res, 1996.18(1-2):181-184.

5. CaiJianrong, Zhou Xiaojun, Wang Feng et al. Obstacle Identification of Citrus Harvesting Robot [J]. Transactions of the Chinese Society for Agricultural Machinery, 2009, 40(11):171-175.

6. Lu Wei, Song Aiguo, CaiJianrong et al. Structural design and kinematics algorithm research for orange harvesting robot [J]. Journal of Southeast University(Natural Science Edition), 2011, 41(1):95-100.

7. E.J.van Henten.J. Hemming.B.A.J. van Tuijlet a1. An Autonomous Robot for Harvesting Cucumbers in Greenhouses [J]. AutonomousRobots,2002, Vol.13, No.3:241-258

8. Zhan Wentian, He Dongjian, Shi Shilian et al. Recognition of kiwifruit in field based on Adaboost algorithm [J]. Transactions of the Chinese Society of Agricultural Engineering, 2013, (23):140-146.

9. Ding Yalan, Geng Nan, ZhouQuancheng. Reacher on the Object Extraction of Kiwifruit Based on Images [J]. Microcomputer Information, 2009, 25(18) :294-295.

10. Cui Yongjie, Su Shuai ,LvZhihai et al. A Method for Separation of Kiwifruit Adjacent Fruits Based on Hough Transformation [J]. Journal of Agricultural Mechanization Research, 2012, 34(12):166-169.

11 N AfiqahZainuddin, YM Mustafah, YAM Shawgi, and NKAM Rashid, Autonomous Navigation of Mobile Robot Using Kinect Sensor[C]. Computer and Communication Engineering (ICCCE), 2014 International Conference, IEEE, 2014: 28-31.

12 Esmeralda Ruiz, Raúl Acuña, Novel Certad, Angel Terrones, and Maria E Cabrera, Development of a Control Platform for the Mobile Robot Roomba Using Ros and a Kinect Sensor[C]. Robotics Symposium and Competition (LARS/LARC), 2013 Latin American , IEEE, 2013: 55-60.

13 Matthew Clark, David Feldpausch, and Girma S Tewolde, Microsoft Kinect Sensor for Real-Time Color Tracking Robot[C]. Electro/Information Technology (EIT), 2014 IEEE International Conference on,IEEE, 2014:416-421.

14 Tilak Dutta. Evaluation of the Kinect™ Sensor for 3-D Kinematic Measurement in the Workplace[J]. Applied ergonomics, 43 (2012), 645-649.

15 Antonio Sgorbissa, and DamianoVerda, Structure-Based Object Representation and Classification in Mobile Robotics through a Microsoft Kinect[J]. Robotics and Autonomous Systems, 61 (2013), 1665-1679.

16 Cui Yongjie, Su Shuai, Wang Xiaxia et al. Recognition and Feature Extraction of Kiwifruit in Natural Environment Based on Machine Vision [J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(5):247-252.