Location and Recognition Fruit Trees Based on Binocular Stereo Vision

Xueguan Zhao[1,2], Yuanyuan Gao[1,2], Songlin Wang[1,2],
Xiu Wang[1,2(✉)], Pengfei Fan[1,2], and Qingcun Feng[1,2]

[1] Beijing Research Center of Intelligent Equipment for Agriculture,
Beijing, China
[2] National Research Center of Intelligent Equipment for Agriculture,
Beijing, China
{zhaoxg,gaoyy,wangsl,wangx,wangxiu,
fanpf,fengqc}@nercita.org.cn

Abstract. In order to improve pesticide utilization rate and reduce the environmental pollution caused by pesticide ground loss, this paper proposes to use binocular vision to recognize the contour and distance information of fruit trees. To improve the recognition accuracy and speed, focusing on the optimization of SIFT stereo matching algorithm. A method for matching the feature points of left and right images base on cosine distance and the vector modulus is proposed. On this basis, two stereo matching algorithms are compared, The accuracy of the Improved SIFT stereo matching algorithm is improved by 1.53%, With this method, the recognition time is almost unchanged, And the stability of depth measurement is analyzed. When the target distance sensor is 180 cm–220 cm, the standard deviation is 1.3592 cm, can meet the requirements of the work.

Keywords: Image recognition · Binocular vision · Stereo match
Visual location

1 Introduction

The misuse of pesticide application is very common in the actual process of agricultural production. Excessive application of will not only cause the waste of pesticide, but also bring pollution to the environment. Automatic target spray technology of fruit trees is an important way to achieve high efficiency and low pollution. The realization of target spray based on machine vision technology has become a research hotspot in the field of precision spray technology at home and abroad due to its high flexibility and low equipment development cost.

Target technology mainly uses sensors for target detection, such as infrared sensors, ultrasonic sensors, laser radar, image sensors (CCD) and so on. Li Li et al. designed a target spraying control system [1], in which the infrared sensor was used to judge whether the target was available or not, and a green sensor was used to judge whether the target was green, so as to achieve the goal of applying pesticide only to green crops, thereby further reducing the waste of pesticides. However, because of the small

detection area of infrared sensor, some switch jitter may be caused by the space between the branches and leaves. As the ultrasonic sensor could detect surfaces, it could solve the similar problem effectively. Gil et al. designed a multi nozzle sprayer consists of 3 ultrasonic sensors and 3 solenoid valves in early 2007 [3]. It could realize variable spraying according to the variation of grape leaves and save 58% pesticide compared with traditional spraying method.

Changyuan Zhai et al. built a target contour detection platform based on the ultrasonic sensor [2, 4], and some target detection experiments were carried out on the regular crown and cherry trees. Good results were obtained, which proved the feasibility of the ultrasonic target detection. Solanelles et al. applied ultrasonic sensors and proportional control valves to the air blower and controlled the spray flow according to the measured tree width, thereby greatly saving the pesticide [5]. Although the cost of CCD image sensor is higher and the processing speed is slower, it can also detect crop diseases and insect pests with a certain image processing technology. Honghui Rao et al. used CCD image sensors to collect target information, and sprayed the target by controlling motor movement after images processing from CCD sensors [6]. In 2010, Tianxiang Hu et al. studied the application of binocular vision technology in intelligent target spray, and further improved the accuracy and efficiency of spray [7].

Zacharie, Doerr used two-dimensional laser radar and GPS as sensors, and successfully developed a tractor autonomous navigation system for orchard operations. The accuracy of the system was higher [8, 9].

We can see that the application and research of machine vision in agriculture and forestry have entered the stage of development, and various theoretical and practical results emerge in an endless stream. But the accuracy of ultrasonic sensors and infrared sensors is poor, and laser ranging is expensive, at the same time target detection of machine vision is mainly used in weed identification in the field, and the research of target detection in orchard based on machine vision is still very few. At the same time, as the nerve center of target spraying system, the vision detection decision system needs to be further improved at the speed and accuracy of the detection algorithm.

2   Real - Time Precision Target System

The hardware of Precision target spray decision system mainly include: Daheng CCD camera, PC (host computer), indoor fruit tree test stand. Image analysis development based on Microsoft Visual Studio 2012, It mainly to achieve image acquisition, segmentation, measurement, ranging, intelligent decision-making and sending results and other functions. The image acquisition function is completed by the CCD camera. The image of the fruit tree collected by the CCD camera is stored in the form of an image sequence to the computer for subsequent image processing.

The image acquisition function is completed by CCD camera, and the fruit images collected by the CCD camera are stored in the form of image sequences for the computer to be used for subsequent image processing The applicable image

segmentation, recognition and measurement algorithm is sought and develop software used for image segmentation and processing, realize describerion of fruit tree by gravity, perimeter, shape complexity, depth and other parameters. And can recover the real object of fruit tree from the image information to lay the foundation for the precise application of pesticides, functional flow chart shown in Fig. 1.
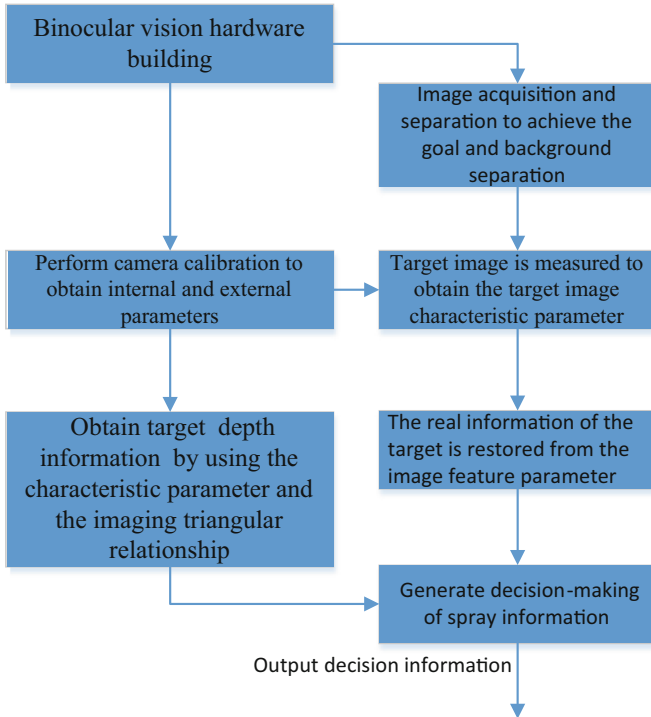

Fig. 1. Functional flow chart

3 System Calibration

MATLAB camera calibration toolbox has higher calibration accuracy, and the calibration method is simple, therefore this paper uses MATLAB to about camera calibration, the mercury series Daheng camera MER-500-7UC, 2592*1944 resolution, spatial position and binocular vision sensor calibration plate as shown in Fig. 2:
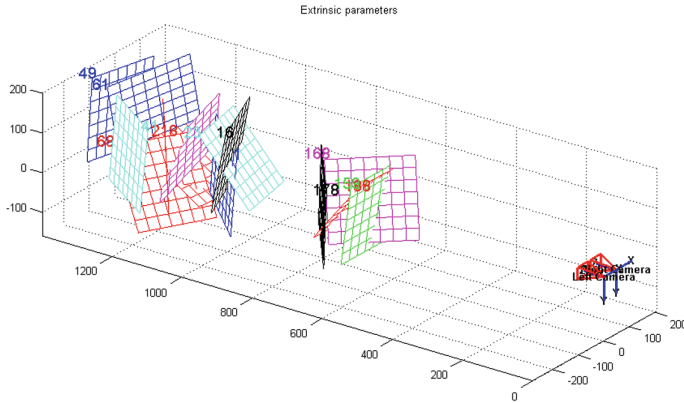
Fig. 2. Spatial position of calibration board and binocular vision sensor

MATLAB's camera calibration toolbox has a high calibration accuracy [10, 11], and calibration method is simple, so use MATLAB to calibrate the left and right cameras in this paper, and the calibration results are as follows, the internal parameters of the left and right camera are:

$$A = \begin{bmatrix} 3503.12 & 0 & 1207.31 \\ 0 & 3504.61 & 967.84 \\ 0 & 0 & 1 \end{bmatrix} B = \begin{bmatrix} 3511.25 & 0 & 1198.93 \\ 0 & 3511.71 & 960.13 \\ 0 & 0 & 1 \end{bmatrix}$$

The equivalent focal length of the camera in the X direction is fx, the inner parameter matrix of the left camera obtained from the test shows that the fx of the left camera is 3503.12,This article adopts Computer brand 8 mm focal length lens, the physical dimensions of a camera pixel in the X direction have been given by the camera: dx = 2.2 um. According to fx = f/dx, can get f l = 7.706 mm, the same can get f r = 7.724 mm, A computer series of 8 mm focal length lenses are used. Which can be obtained left and right camera calibration error were 0.37% and 0.34% respectively. Then the rotation matrix and the translation matrix are calculated according to the common feature points, and R and T are

$$R = \begin{bmatrix} 0.0220 & 0.6674 & 0.7446 \\ 0.9964 & -0.0759 & 0.0386 \\ 0.0822 & 0.7408 & -0.0666 \end{bmatrix}$$

$$T = \begin{bmatrix} -166.1646 & -84.2903 & 955.0699 \end{bmatrix}$$

The obtained R and T results represent the matrix required for the left camera and the right camera to achieve coplanar, write R and T as XML files for cvstereoRectify function calls in Opencv for stereo correction, the results of the stereo correction are shown in Fig. 3:
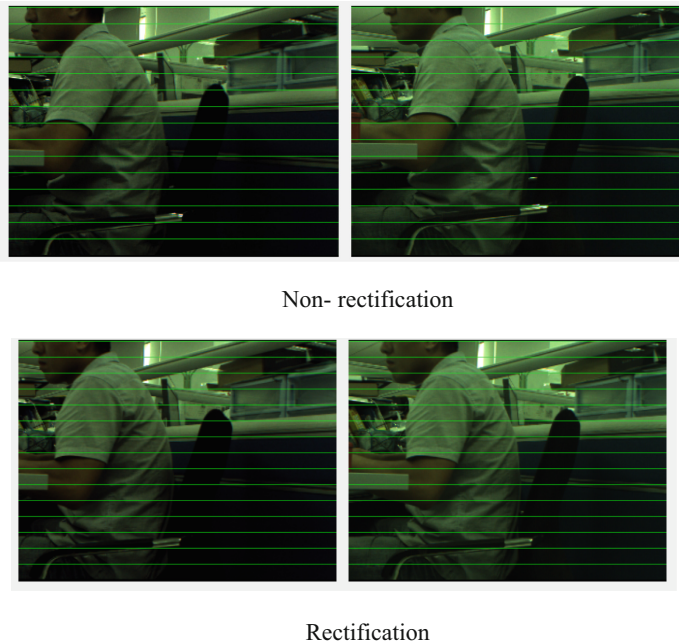
Non- rectification

Rectification

Fig. 3. Image before and after the rectification

4 Location

According to the principle of stereo vision, the simplest binocular stereo vision model is a stereo camera made up of two parallel lenses to capture the same scene image, as shown in Fig. 4. Since there is a distance between the two lenses (Fig. 4b), the target perceived through these two lenses will produce bias in the captured image (Fig. 4, dl, dr). According to the triangulation principle [12, 13], these deviations are proportional to the distance between the camera and the target (Z in Fig. 4), so that these deviations can be used to calculate the depth information of the target.
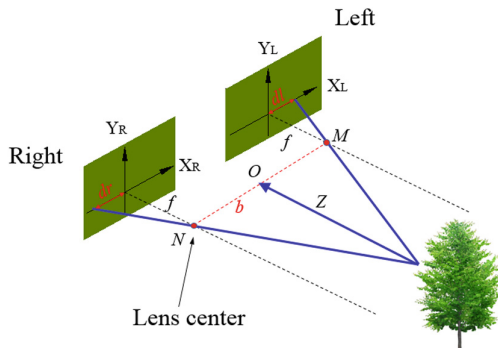

Fig. 4. Parallel optical axis geometric model of fruit trees Stereoscopic vision

b represents the horizontal distance between the two cameras, the baseline of the stereo vision system, the F is focal length, and the Z is depth. The parallax d can be calculated according to the formula (2-1) once the deviation (dl and dr) of the horizontal direction of the two images is determined.

$$d = d\,l - d\,r \tag{1}$$

See from the coordinates in Fig. 4, the dr is negative, so in fact the parallax d is the sum of dl and dr. Where the relevant parameters from the triangular similarity relationship can be obtained from the baseline mathematical expression:

$$b = (dl + dr) \cdot \frac{Z}{f} = d \cdot \frac{Z}{f} \tag{2}$$

Considering the relationship between length and pixels, the unit unity of parallax and depth calculation results is:

$$Z_{[mm]} = \frac{b_{[mm]} \cdot f_{[mm]}}{k_{[mm/pixel]} \cdot d_{[pixels]}} \tag{3}$$

In Eq. (3), Z, b, f units are mm, d is represented by pixels, and k is the size of each pixel.

5   Binocular Matching

Identification of fruit trees in the environment in real time and accurately, it is the key to the target spraying vision system, and the essence of object recognition is image segmentation. In this study, indoor simulation experiments were conducted to collect images of fruit trees [14]. Because there is a big difference between the surface color and the background color of fruit trees, there are different distribution characteristics in color space, so that image segmentation can be used to extract the target fruits and vegetables from the background. In this paper, 2G-R-B is used as a partition factor of the super green method. The picture of the fruit tree captured by the binocular vision system is shown in Fig. 5.

Fig. 5. A Pair of original stereo images

5.1 SIFT Feature Point Matching Algorithm

The appearance of fruit trees is complicated and prone to mismatching. Due to a hypothetical of region matching is the space plane is the plane, which is parallel to the camera plane, in the actual orchard scene there are a lot of non-positive planes, so people began to consider the use of some of the more significant feature points (points of interest) to match, this method can also be called feature matching method [15, 16].

SIFT feature matching algorithm is an image local feature description algorithm based on scale space, invariance to image scaling, rotation, and even affine transformations, therefore, it is widely used in matching technology. Because the number of feature points detected by SIFT feature point detection algorithm is larger, and each feature point descriptor is a vector of 128 dimensions. The feature utilization rate of the algorithm is not very high, it takes longer time, and there are some problems such as matching error or repetition.

5.2 Improved Stereo Matching Strategy

As the most common distance measurement standard, "euclidean distance "is also used as a matching criterion for SIFT based matching, and good results are obtained. However, there are still problems such as matching errors or repetitions. From the analysis of the eigenvector itself: if the direction of the two vectors is the same, the smaller the angle is, on the basis of that, the modulus of the vector is taken into account. If the modulus is equal or close, the two vectors are considered equal. To determine whether the two vectors are in the same direction, it is necessary to use the cosine theorem to compute the angle of the vector. The cosine similarity of vector a and B is calculated as follows

$$\cos\theta = \frac{a \cdot b}{|a||b|} \tag{4}$$

From the above SIFT feature descriptor, we can see that the dimension of vectors a and b is a = (a1, a2, …an)T, b = (b1, b2, …bn)T, then the cosine of vectors a and b is:

$$\cos\theta = \frac{\sum\limits_{i=1}^{n} a_i y_i}{\sqrt{\sum\limits_{i=1}^{n} a_i^2 \sum\limits_{i=1}^{n} b_i^2}} \tag{5}$$

Compared to Euclidean distance, cosine distance pays more attention to the difference of the two vectors in the direction. If we compare the length of the two vectors on the basis of the direction, we can accurately extract the feature matching points, the length of the vector, that is, the norm of the vector:

$$||x|| = |\sqrt{x_1^2 + x_2^2 + \cdots x_n^2} \tag{6}$$

The specific implementation steps of the proposed stereo matching strategy are as follows: Firstly, the key feature points of the left/right image are extracted, and the image with the key feature points is used as the reference image, and the image with

less key feature points is used as the image to be matched. The vector matrix of the K1 critical feature points of the baseline image is X, the vector matrix of the K2 critical feature points of the matched image is Y, X and Y are shown below:

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,1} \\ x_{2,1} & x_{2,2} & x_{2,3} & x_{2,1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{k1,1} & x_{k1,2} & x_{k1,3} & x_{k1,4} \end{bmatrix} Y = \begin{bmatrix} y_{1,1} & y_{1,2} & y_{1,3} & y_{1,1} \\ y_{2,1} & y_{2,2} & y_{2,3} & y_{2,1} \\ \vdots & \vdots & \vdots & \vdots \\ y_{k2,1} & y_{k2,2} & y_{k2,3} & y_{k2,4} \end{bmatrix} \quad (7)$$

Take the first row vector of A $\begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,1} \end{bmatrix}$, that is, the first critical feature point vector of the reference image, calculates cosine distance from all row vectors in B and takes inverse cosine, obtain an angle sequence of two vectors {θ1, 1, θ1, 2, θ1, 3, ⋯ θ1, k}. To increase robustness,

Take the minimum value of the sequence of five values corresponding to the vector for norm comparison, take X first row feature vector and Y line k feature vector as an example:

$$\delta = abs(||a|| - ||b||) = |\sqrt{x_{1,1}^2 + x_{1,2}^2 + \cdots x_{1,n}^2} - \sqrt{y_{k,1}^2 + y_{k,2}^2 + \cdots y_{k,n}^2}| \quad (8)$$

Take the vector corresponding to the minimum of the δ as the feature point. And so on, until all of the row vectors in A have been completed in turn to calculate the cosine distance of all row vectors in B, and the comparison of the norm. That is, all the key K1 feature points of the reference image are matched with all the K2 key feature points of the image to be matched.

5.3 Key Feature Point Matching Contrast Experiment

According to the research scheme described in this paper, the application of improved SIFT algorithm in stereo matching is studied, and simulation experiments are carried out to verify and analyze the feasibility of the improved stereo matching strategy. Computer configuration in the test are: CPU core i5-2410 M, Memory is 2 GB, the operating system is Win7, and simulation platform is Matlab 2013. According to this method, the key feature points of left and right images are generated based on the improved SIFT algorithm, and the feature vectors extracted from the key feature points of left and right pictures are shown in Fig. 6(a) and (b).
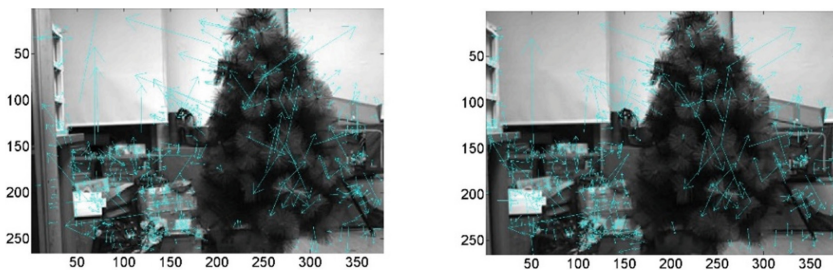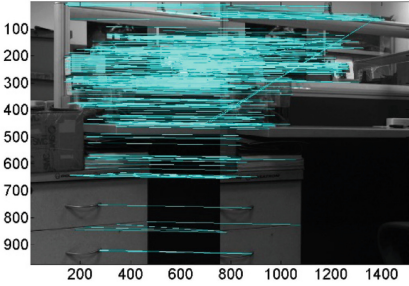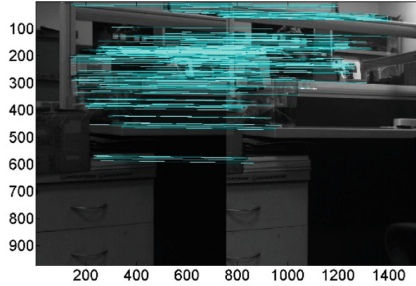

Fig. 6. Feature extraction result of the key feature points of the left/right image
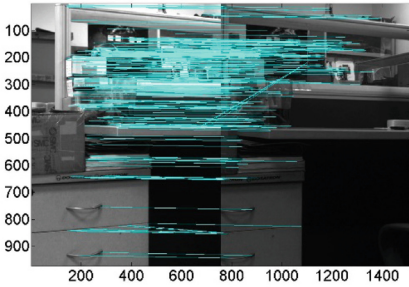
In this paper, the indoor environment image is selected, and the improved SIFT algorithm is tested under the illumination changes, scale changes and rotation changes. The matching results are shown in Fig. 7.
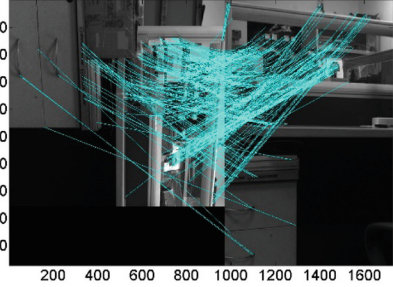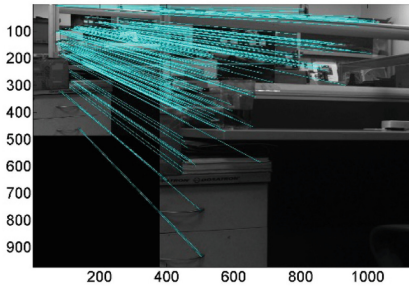

Control group

light is reduced by 30%.

light is increased by 30%.

Rotate 90 °

Scaling 50%

Fig. 7. Matching results of improved SIFT algorithm

Table 1. Statistical results of feature point matching

Matching method	Number of key feature points		Number of feature points matching	Number of error matching points	Correct matching rate/%	Matching time/ms
	Left	Right				
SIFT matching algorithm	490	498	206	7	96.60	697.21
Improved SIFT matching algorithm	490	498	214	4	98.13	702.63

In order to compare the difference in matching time between the improved SIFT algorithm and the SIFT algorithm, about 20 of the image acquisition in contrast, rotation, zoom, change of illumination conditions, using SIFT algorithm and improved SIFT matching algorithm for feature matching results statistics, statistical results from 20 on average of the image, as shown in the Table 1. The improved rule matching, stereo matching method increases the matching number, improve the matching efficiency, matching repetition or error reduce 1.53%, it is more advantageous to 3D reconstruction and localization of robot vision system.

5.4 Binocular Stereo Vision Experiment

The distance between the fruit tree and the camera is between 1.5 and 2. 5 m, the length of the baseline of the two camera is 50 mm, and the algorithm is used to match. The matching effect is as Fig. 8:
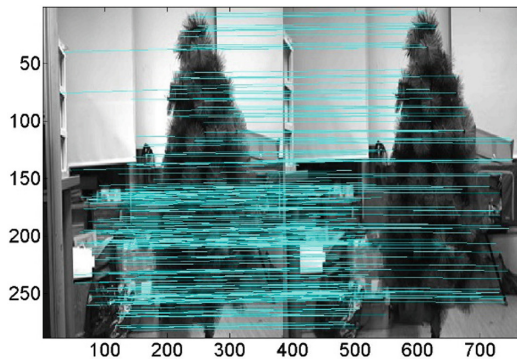

Fig. 8. Matching result of fruit tree binocular image

5.5   Test Results and Discussions

In order to further observe the influence of matching region selection on the region matching effect, the SIFT matching algorithm and the improved SIFT matching algorithm are tested respectively. The binocular vision sensor is used to reconstruct the fruit tree, and then the width, height and depth of the fruit tree are measured in space, and the comparison is made. In order to test the accuracy of the binocular vision sensor in space reconstruction at different depths from the target, the distance between binocular camera and fruit tree from 1.5 m to 2.5 m is measured every 0. 05 m. The measurement results are shown in Fig. 9.

According to the three-dimensional reconstruction method, combined with the camera's internal and external parameters and parallax images, the 3D point cloud is measured at different angles, to get the height, width of the fruit trees and the depth in the vertical direction of the image plane. Where AB is the connection between vertex and diameter center of stem, The CD is the corners of the vertical line at 1/8 of the AB line. and the outer contour,1 is the center of the AB line, 2 is the 1/8 at one end of the AB line, 3,4, respectively, at both ends 1/6 of the CD, the results shown in Table 1.
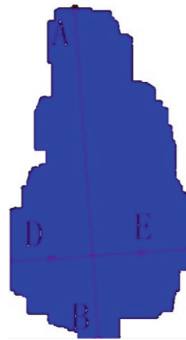

Fig. 9. Test point distribution

The actual distance between the lens and the target 1 points(cm)	Measured value					
	Height (cm)	Width	Target point A	Target point B	Target point D	Target point E
150	120.0	45.5	136.4	147.6	168.6	173.3
160	119.9	46.7	147.3	148.7	176.1	179.7
170	120.4	45.2	159.4	164.5	190.7	194.5
180	121.8	47.1	177.8	182.0	208.2	207.5
190	122.1	48.4	186.6	192.4	219.6	219.2
200	121.4	48.1	210.4	212.8	227.1	241.6
210	120.9	47.8	221.2	224.3	244.0	252.6
220	122.0	48.8	228.9	230.8	260.3	261.9
230	122.1	49.1	236.6	246.8	274.4	271.9

<div align="right">(continued)</div>

(*continued*)

The actual distance between the lens and the target 1 points(cm)	Measured value					
	Height (cm)	Width	Target point A	Target point B	Target point D	Target point E
240	124.5	49.6	258.9	261.0	287.5	287.1
250	123.2	50.3	273.8	274.9	302.6	302.4
True value	121.5	48.0	–	−5.4	+20.3	+22.7

The true value of the target 2–4 points is the relative depth of the target 1, and the "+" represents the direction of great depth.

In order to analyze the stability of the depth measurement, the difference between the target A and the mean value of target D, E as the length, the difference between the target D and the target E is half of the width, from the three-dimensional reconstruction of the measurement results can be found, compared to its height and width, length and manual measurement results, the standard deviation was 1.3805, 1.6224, 3.6081. And length value corresponds to the depth value in space, and its beating is greater compared with the width and height, the,, and the width and height than the beating larger, When the sensor is away from the target at 180 cm–220 cm, the standard deviation is 1.174, relatively stable, so in the actual positioning of the apple space, the Deviation range of distance from visual sensor to the target 1 point range should be measured within the 40 cm (Fig. 10).

Fig. 10. Real-time measurement results

Through a lot of test analysis, the error is mainly due to the accuracy of the depth information acquisition and several factors related to: the distance between the trees and the camera; the uniformity of light on the regional matching is also greater, due to the uneven light resulting in the match is not accurate, easy to cause the depth of information error, coupled with the impact of random noise, image segmentation results will be biased, affecting the results of distance measurement.

References

1. Li, L., Li, H., He, X., et al.: Development and experiment of automatic detection device for infrared. Trans. Chin. Soc. Agric. Eng. **28**(12), 159–163 (2012)
2. Zhai, C., Zhao, C., Wang, X., et al.: Probing method of tree spray target profile. Trans. CSAE **26**(12), 173–177 (2010)
3. Gile, E., Rosell, J.R., et al.: Variable rate application of plant protection products in vineyard using ultrasonic sensors. Crop. Prot. **26**(8), 1287–1297 (2007)
4. Stajnko, D., Berk, P., Lešnik, M., et al.: Programmable ultrasonic sensing system for targeted spraying in orchards. Sensors **12**(11), 15500–15519 (2012)
5. Solanelles, F., Escola, A., Planas, S., et al.: An electronic control system for pesticide application proportional to the canopy width of tree crops. Biosyst. Eng. **95**(4), 473–481 (2006)
6. Rao, H.: Study on crop spraying control based on machine vision. J. Nanjing Agric. Univ. **30**(1), 120–123 (2007)
7. Hu, T., Zheng, J., Zhou, H.: Distance measurement of tree images based on binocular vision. Proc. Agric. Mech. Soc. **41**(11), 158–162 (2010)
8. Cai, J., Sun, H., Li, Y.: Three dimensional information acquisition and reconstruction based on binocular stereo vision. Chin. J. Agric. Mach. **43**(3), 152–156 (2012)
9. Doerr, Z., Lague, C.: Evaluating the ability to detect foreign objects in crops using range scanners mounted on agricultural vehicles. In: ASABE Annual International Meeting, Nevada, 21–24 June 2009
10. Zou, L., Li, Y.: A method of stereo vision matching based on Open CV. In: Proceedings of 2010 International Conference on Audio Language and Image Processing (ICALIP), pp. 185–190. IEEE, Piscataway (2010)
11. Spampinato, G., Lidholm, J., Ahlberg, C., et al.: An embedded stereo vision module for industrial vehicles automation. In: 2013 IEEE International Conference on Industrial Technology (ICIT), pp. 52–55. IEEE, Piscataway (2013)
12. Ganganath, N., Leung, H.: Mobile robot localization using odometry and Kinect sensor. In: 2010 IEEE International conference on Emerging Signal Processing Applications (ES-PA), NJ, USA, pp. 91–94 (2010)
13. Bradley, D., Heidrich, W.: Binocular camera calibration using rectification error. In: 2010 Canadian Conference on Computer and Robot Vision (CRV), pp. 183–190. IEEE (2010)
14. Shuai, Z.: Research on the Error Theory of Binocular Vision measurement system on Non-cooperative Targe. Changchun: Changchun Institute of Optics, fine Mechanics and Physics, Chinese Academy of Sciences (2014)
15. Ravari, A.R.N., Taghirad, H.D., Tamjidi, A.H.: Vision-based fuzzy navigation of mobile robots in grassland environments. In: Proceedings of International Conference on Advanced Intelligent Mechatronics.[S. I.]. IEEE/ASME Press (2009)
16. Song, T., Tang, B., Zhao, M., et al.: An accurate 3-D fire location method based on sub-pixel edge detection and non-parametric stereo matching. Measurement **50**, 160–171 (2014)