American Journal of Biochemistry and Biotechnology

Original Research Paper

A Vision System Based on Time-of-Flight and RGB Cameras Applied to Robotic Tree Peony Fruit Harvesting

Jiaming Liu, Dong Zhao, Tianxing Li and Jian Zhao*

School of Technology, Beijing Forestry University, Beijing, 100083, P.R. China

Article history
Received: 17-07-2020
Revised: 04-08-2020
Accepted: 18-08-2020

Corresponding Authors:
Jian Zhao
School of Technology, Beijing Forestry University, Beijing, 100083, P.R. China
Email: zhaojian1987@bjfu.edu.cn

Abstract: Tree peony is a deciduous shrub endemic to China and the Peony Seed Oil (PSO) is an important plant oil resource. However, at present, fruits harvesting of peony are mainly completed by manual work with low efficiency. In response to the need for a mechanized operation, a multi-sources vision system based on Time-Of-Flight (TOF) and RGB cameras was set up in this study. To achieve this, an RGB camera and a TOF camera were used to capture tree peony images over the same time period. A method based on Speeded-Up Robust Features (SURF) algorithm, nearest neighbor and Random Sample Consensus (RANSAC) algorithm was carried out to detect and match the feature points of grayscale images and intensity images. Then, the Normalized Direct Linear Transformation (NDLT) algorithm was used to achieve image registration of RGB images and depth images through the matched feature points. Based on Multi-Layer Perceptron (MLP) algorithm, by using the depth image and RGB image, the localization and maturity classification for peony fruits were achieved in this study. In our research, 90 groups of tree peony fruit images captured by this vision system were used to verify the feasibility of the algorithm. The result shows that in these images, 152 of 173 fruits were correctly recognized and the fruit recognition rate was 85.74%. The average of localization errors was 3.53, which is accuracy for harvesting operation. As for maturity classification, this system achieved a high recognition rate, 91.68% in total. The results show that the vision system achieved extracting location and color information of the fruit at the same time and it is not easy to be affected by environmental illumination and other factors. The proposed method can achieve high efficiency and high accuracy in terms of fruit localization and maturity classification.

Keywords: Tree Peony Fruit, Multi-Sources, Feature Points, Vision System, SURF, NDLT, MLP

Introduction

Tree peony, which belongs to the family Paeoniaceae, genus Paeonia and Sect. Moutan DC is a woody deciduous shrub endemic to China, which has grown there since the Eastern Jin Dynasty (Sun *et al*., 2016). Its pod contains dark oval seeds characterized by various Unsaturated Fatty Acids (UFAs) and a high proportion of n-3 fatty acids (Su *et al*., 2016). Peony Seed Oil (PSO) was declared as a new resource food in China in 2011, owing to its high level of α-Linolenic Acid (ALA). In the past few years, tree peony has been considered as an economic plant and some species (such as Paeoniarockii and Paeoniaostii) are widely planted in China with a potential annual seed production of 60000 tons in the next decade (Mao *et al*., 2017). The harvesting of peony pods is a time-consuming task and is currently performed by hand, accounting for more than 60% of the total labor time of this crop in China. Moreover, the harvesting period is very short, around 5 weeks. The key to increase the output of PSO is to prune branches reasonably in the growing process and to harvest fruits in time during the frutescence (Liu *et al*., 2020). However, increasing costs and decreasing supply of skilled labor force are restrictive factors in the development of the peony industry in China. The development of a peony pod harvesting robot is an effective way to address the labor shortage and high labor cost.

Harvesting robots are designed to sense the complex field environment by various sensors and to employ this information effectively to perform harvesting actions.

The bottleneck to promote the application of harvesting robot lies in the performance of the vision system. Three major functions should be achieved in a vision system applied to a harvesting robot: Recognizing the peony fruit from the tree, identifying the maturity of the fruit and locating the fruit. These functions provide important information to the harvesting robot to guide the harvesting actuating mechanism.

The basic technology employed in this study is computer vision, which has become a mature theory with many achievements in several fields, such as features extraction and image segmentation (Bulanon and Kataoka, 2010; Bulanon *et al*., 2010; Park *et al*., 2017; Chen and Hashimoto, 2017; Sharma *et al*., 2020). With the developments in the areas of Digital Image Processing and Intelligent Control technologies, machine vision is extensively used in agriculture (Kumar and Rajpurohit, 2019). For example, (Wu, 2020) reported a method for determining the chlorophyll content of rice based on computer vision, by the three color characteristic parameters of GR, BR and R/(G+B). Nowadays, a vision scheme for a harvesting robot is achieved by various visual sensors (De-An *et al*., 2011; Zhao *et al*., 2016). As an example, (Rakun *et al*., 2011) described a multi-view computer vision-based model for object detection that can serve as a preliminary step in fruit prognosis, which involves the estimation of the number, diameter and yield of apple fruits. This vision system uses three features-color, texture and three-Dimensional (3D) shape of possible areas-to detect and verify all areas fruits. Similar work was done by (Ding, 2009). In her work, a matching method using centroid characters of the fruit was described, so as to match the binocular images of kiwifruit. With precise dimensional measurements on the position data, the experiments showed that the calculation error of the special position was 9.03 mm when the orientation depth was near 800 mm. However, most of the studies presented drawbacks of timeliness and veracity and the results could be disturbed by environmental features such as illumination intensity. This resulted in limitations of the effect of the vision system in some situations.

A Time-Of-Flight (TOF) camera is a 3D camera that able to capture the 3D geometry of a scene. Objects can be then segmented and recognized from their 3D geometry (Conde, 2020). It's fundamental principles are using the time-of-flight method, which continually sends a light pulse to the target object and records the time the sensor receives the reflected pulse to calculate the distance between the object and camera (Chiabrando *et al*., 2009; Schwarz *et al*., 2014; Falie and Buzuloiu, 2007). There are several advantages to using a TOF camera to locate an object, including its smaller size, lighter weight and better signal-to-noise ratio (Conde, 2020). TOF cameras have been widely used in several fields such as gesture recognition, face detection and object localization (Kollorz *et al*., 2008; Takahashi *et al*., 2011; Lee *et al*., 2011). For example, some researchers reported a flexible sensor fusion approach to retrieve scale information in monocular Visual Odometry (VO) through integrating range measurements from a wide variety of depth sensors spanning from small-resolution Time-Of-Flight (TOF) cameras to 2-D and potentially 3-D LiDARs. (Chiodini *et al*., 2020). In addition, many studies focus on the accuracy of the TOF camera. A correction method based on Delay Line (DLL) was demonstrated to improve the linearity and accuracy of the TOF camera (Li *et al*., 2020). However, the depth image is a pseudo-color image as the color of each pixel represents the depth data instead of the actual color, which means that the color information of the target object cannot be shown in the obtained image. For robotic harvesting, this drawback prevents the identification of the maturity of the fruit.

This study intends to develop a multi-source vision system based on RGB and time-of-flight cameras that can be applied to robotic harvesting for tree peony fruit. This system is designed to combine the advantages of the two cameras for achieving fruit recognizing, locating and maturity classifying. RGB images captured from an RGB camera contain the full color and texture information of the fruit, which can be used to achieve maturity identification. The depth image captured from a TOF camera contains location information of fruit, which can be used to achieve fruit detection and localization. An algorithm is described in this study to match RGB images captured from an RGB camera with depth images captured from a TOF camera. A classifier based on Multi-Layer Perceptron (MLP) neural networks algorithm for fruit recognition as well as an MLP classifier for maturity classification are set up to locate the fruits and classify their maturity.

Image Registration

Images captured from the RGB camera and TOF camera were processed in several stages, as coarsely illustrated in Figure 1. In this system, the RGB camera captures RGB images and the TOF camera captures three images (depth image, amplitude image and intensity image). The RGB image is transformed to a grayscale image and the intensity image is enhanced by histogram equalization to improve its quality for the next stages. The next stage is to extract and match features of the two images. Considering the nature of the image, we used a Speeded-Up Robust Features (SURF) algorithm to detect the feature points of these images and presented an algorithm based on the nearest-neighbor algorithm, Hessian matrix trace accelerates and the RANSAC algorithm to achieve image features matching. When this work was accomplished, we finished the image registration of the two cameras using the Normalized Direct Linear Transformation (NDLT) algorithm.

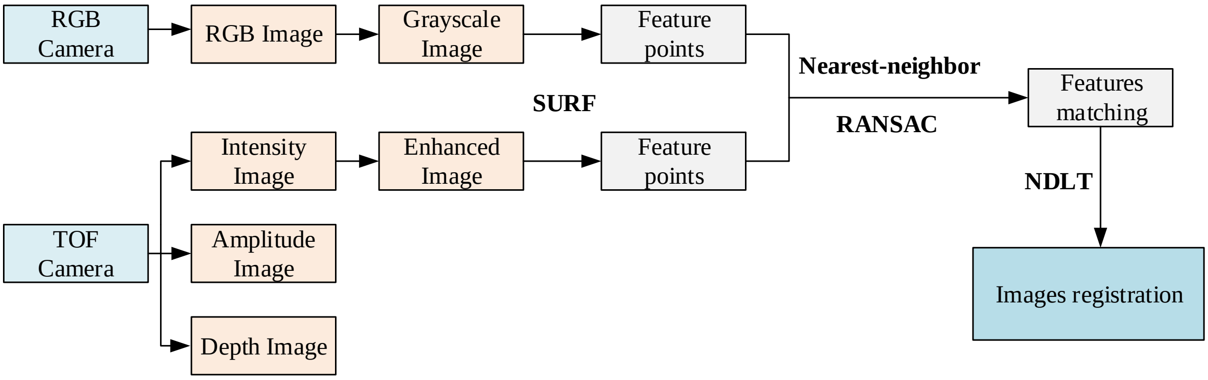


Fig. 1: The conceptual algorithm of image registration of the vision system

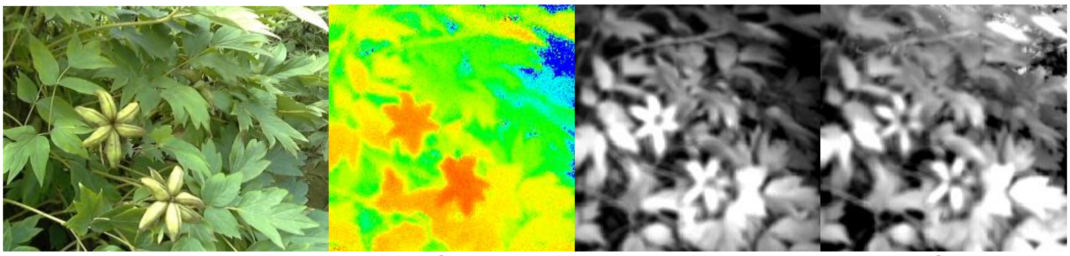


Fig. 2: Images captured from the RGB camera and Time-of-Flight (TOF) camera at the same time (a) RGB image; (b) depth image; (c) amplitude image; (d) intensity image

Image Preprocessing

This study used a vision system to simultaneously capture four images (RGB image, depth image, amplitude image and intensity image), which are illustrated in Fig. 2. The RGB image was captured from the RGB camera, others were captured from the TOF camera.

The RGB camera captured images with a resolution of 320×240 pixels and the TOF camera captured images with a resolution of 200×200 pixels. The TOF camera contains two kinds of sensors: CMOS based sensor and IR light emitter. The amplitude image reflects the infrared reflection characteristics of each object in the image and the intensity image is generated by CMOS components capturing ambient light and infrared light emitted by the camera. In this study, we achieved image registration between the RGB camera and TOF camera by matching the RGB image with the intensity image.

In the natural growth environment situation, some parts of the intensity image were blurring as the image had infrared features. Therefore, it was necessary to conduct image enhancement to improve the image quality before image registration.

In this study, considering the similar features of the intensity images and grayscale images, we enhanced the intensity images by histogram equalization, which can improve the contrast ratio and brightness of the image. This algorithm can enhance most details of a zone with a large area and combine the pixel points with the similar grayscale of a zone with a small area:

$$s = T\left(r_k\right) = \sum_{j=0}^{k} n_j / N = \sum_{j=0}^{k} p_r\left(r_k\right) \quad (1)$$

The effect of the histogram equalization algorithm is illustrated in Fig. 3. This figure shows that the discrepancy between the target fruits and background environment was obviously improved without losing information and feature points after the treatment of the original image.

Images captured from the RGB camera were transformed to grayscale images to match to the intensity images, as illustrated in Fig. 4.

Features Detection of Images

In order to achieve image registration, it was crucial to detect the features of each image. Point features, as a simple feature of an image, are widely used for features detection. Many researches have been conducted to detect the point features of an image and several algorithms have been described. Speeded-Up Robust Features (SURF) is a feature extraction algorithm described by (Bay *et al*., 2008). SURF algorithms are suitable for real-time information perception of non-structural situations and has invariance of rotation, size, brightness and high execution efficiency (Li *et al*., 2011; 2012; 2017).

(a) (b)

Fig. 3: The effect of the enhancement of the image intensity; (a) Source image; (b) enhanced image

(a) (b)

Fig. 4: Gray preprocessing of the color image; (a) RGB image; (b) grayscale image transformed from (a)

SURF algorithms detect interest points by the Hessian matrix, using the Gaussian filter to conduct a convolution operation. For a point $X = (x, y)$ in the image, the Hessian matrix of $X = (x, y)$ in σ dimension is as follows:

$$H'(X,\sigma) = \begin{bmatrix} D_{xx}(X,\sigma) D_{xy}(X,\sigma) \\ D_{xx}(X,\sigma) D_{yy}(X,\sigma) \end{bmatrix}. \quad (2)$$

In Equation (2), D_{xx} is the convolution of the *x*-direction filter, D_{xy} is the convolution of the *xy*-direction filter and D_{yy} is the convolution of the y-direction filter. The determinant form is:

$$\det(H'(X,\sigma)) = D_{xx}(X,\sigma) D_{yy}(X,\sigma) - (wD_{xy}(X,\sigma))^2 \quad (3)$$

In Equation (3), *w* denotes the weight coefficient, which was set as 0.9 in this study.

Image Features Matching

Nearest-Neighbor Algorithm

In this study, we performed rough matching by using the nearest-neighbor algorithm, described by (Muja and Lowe, 2014). Defining ND as the nearest distance and NND as the next nearest distance, we calculated Rod as a specific value of ND and NND using Equation (4). We then set a threshold value and used the algorithm expressed in Equation (3.4) to match the point features of each image:

$$Rod = ND / NND \quad (4)$$

$$\begin{cases} if\ Rod \le threshold\ success \\ \quad else\ failure \end{cases} \quad (5)$$

The smaller threshold value leads to a higher reliability of the result of feature registration. In this study, considering of the image size and empirical value, we set the threshold value as 0.6.

Accelerates Matching Algorithm

In our research, to improve the matching efficiency and reduce the matching time, the trace of Hessian matrix was used to determine the neighborhood brightness to speed up the matching

The stages of this algorithm are described as follows. First, the brightness of the feature point and its neighborhood was compared with the background region. If the feature point and its neighborhood had a higher brightness, the Hessian matrix trace was a positive value; otherwise, the Hessian matrix trace was a negative value. When matching the feature points, it was considered whether the brightness of two feature points were the same by the Hessian matrix trace value.

We then calculated the Euclidean distance of each point with the same brightness to improve the matching efficiency (Bishop, 1992).

The results of 10 feature point matching experiments are listed in Table 2. It is shown that the algorithm decreases the processing time by 31.9%, which is proved to be efficient.

RANSAC Algorithm

After the processing described before, it is necessary to check for mismatching and remove false matching points to output the final result. In this study, we used the Random Sample Consensus (RANSAC) algorithm (Matas and Chum, 2004; Schnabel *et al*., 2007), the steps for which are described as follows:

Step 1. Set $I_{\max}$ as the maximum number of iterations, T_{01} as the model fault tolerance and $N_{\lim}$ as the minimum number of consistent points.
Step 2. Extract the numbers of matched feature points by random(nonlinearity) and set up the initial model by estimating the parameter of coordinate transformation.
Step 3. Verify the rest of the points with the initial model and record the number of the matched points as $N_{con}(I)$.
Step 4. If $N_{con}(I) \le N_{\lim}$, reset the initial model, using the least squares method to improve it.
Step 5. Repeat steps 1 to 4 until the cycle-index reaches $I_{\max}$. Use the model of the largest number of $N_{con}(I)$ as the optimal model and remove any false matching points.

NDLT Algorithm

The Normalized Direct Linear Transformation (NDLT) algorithm is used to match two images after feature points matching is completed (Guo, 2009). Compared to the DLT algorithm (Guo, 2009; Abdel-Aziz *et al*., 2015), the NDLT algorithm has an invariance property of similarity transformation while having high calculation accuracy. The algorithm needs to perform the orthogonal transformation of the matched feature points before DLT to calculate the projection matrix and adjust the images. The steps of the NDLT algorithm are listed as follows:

Step 1. Conduct the orthogonalization of feature point x: set R as the rotation matrix and t as the translation vector; then, the similarity transformation of the feature point in the other image can be described as $T = \begin{bmatrix} sR & t \\ 0^T & 1 \end{bmatrix}$; next, transform x_i to X_i by $X_i = Tx_i$.
Step 2. Conduct the orthogonalization of feature point x' by the same method as step 1.
Step 3. Calculate the orthogonalization projection matrix by $\{x_i \leftrightarrow x_i'\}$.
Step 4. Calculate the real homography projection matrix by H: $H = T'^{-1}HT$.

Experiment of Image Registration

Setup of the Vision System

We firstly put the RGB camera on the tripod and then setup the TOF camera on the top of the RGB camera. So, these two cameras were pointed at the same space. The vision system we set up in this study contained an RGB camera, a TOF camera and computer, as shown in Fig. 5. The parameters of the two cameras are listed in Table 1.

We set the frame rate of both camera to 30F/S and captured images of tree peony fruits. We then used the vision system to simultaneously capture four images, as illustrated in Fig. 1. We selected 15 groups of images to perform image rectification.

Fig. 5: Multi-source vision system for peony fruit oil

Table 1: The effects of speeded-up matching

Method	SURF matching time
Matching without speeding-up	1.580928s
Speeded-up matching	1.076284s

Table 2: The effects of speeded-up matching

Camera	Name	Visual angle	Resolution ratio	Frame rate
RGB camera	Logitech C270	60°	320×240 pixels	40 *F/S*
TOF camera	PMD Camcube3.0	40°	200×200 pixels	30 *F/S*

Result of Image Features Matching

In order to verify the feasibility and accuracy of the image features matching algorithm, an experiment using 15 groups of RGB images and intensity images was conducted. We evaluated the results by such parameters as those listed as follows:

- M_{SE} (Mean squared error):

$$M_{SE} = \frac{1}{N}\sum_{(r,c)}\left(R(r,c)-S(r,c)\right)^2 \tag{6}$$

- C_{EF} (Correlation coefficient):

$$C_{EF} = \frac{\sum_{(r,c)}\left(R(r,c)-\bar{R}\right)\left(S(r,c)-\bar{S}\right)}{\sqrt{\sum_{(r,c)}\left(R(r,c)-\bar{R}\right)^2\sum_{(r,c)}\left(S(r,c)-\bar{S}\right)^2}} \tag{7}$$

- N_{MI} (Normalized mutual information):

$$N_{MI} = \frac{H\left(R(r,c)\right)+H\left(S(r,c)\right)-H\left(R(r,c)S(r,c)\right)}{\max\left(H\left(R(r,c)\right),H\left(S(r,c)\right)\right)} \tag{8}$$

In Equations (6)–(8), N is the number of image pixels; $S(r,c)$ is the grayscale of the standard image; $R(r,c)$ is the grayscale of the matched image; $H(S(r,c))$ is the information entropy of the standard image; and $H(R(r,c))$, $H(S(r,c))$ is the information entropy of the matched image.

The result of image features matching is shown in Fig. 6 and Table 3. The results are better when M_{SE} is as small as possible, C_{EF} is closest to 1 and N_{MI} is as large as possible.

Figure 6 shows the numbers of feature points were matched between the grayscale images and intensity images. Table 3 also shows that the method is efficient and accurate when using the above-described algorithm to achieve feature points detecting and matching.

Result of Image Registration

In this study, the TOF camera captured depth image, amplitude image and intensity image at the same time and the rectification of each image was performed automatically by the camera. This means that we could match the depth image with the RGB image by matching the intensity image with the RGB image.

After the feature matching process, we matched the feature points of the grayscale image with the intensity image. In this experiment, the NDLT algorithm as previously described was used to match the grayscale image (transformed by RGB image) with the intensity image. Since the grayscale image was transformed by the RGB image captured from the RGB camera and the intensity image could be matched with the depth image automatically, we achieved image registration between the RGB camera and TOF camera.

Table 3: Result of feature points matching

Algorithm	M_{SE}	C_{EF}	N_{M1}	Time
SURF	0.0073	0.9528	0.7904	1.86927s

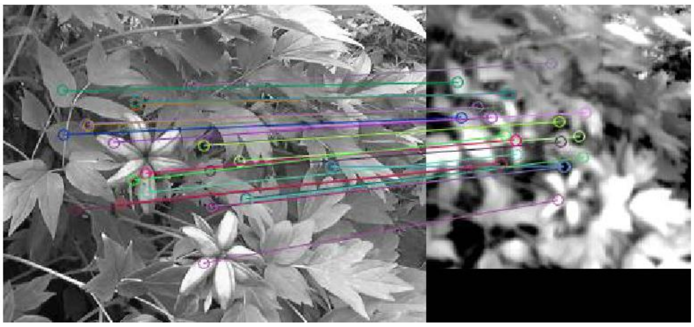
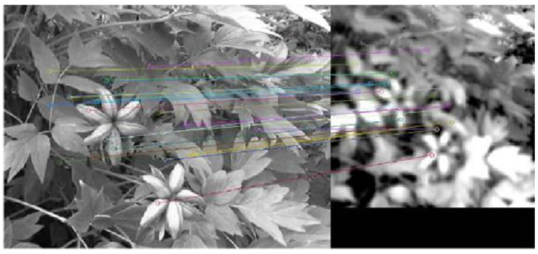
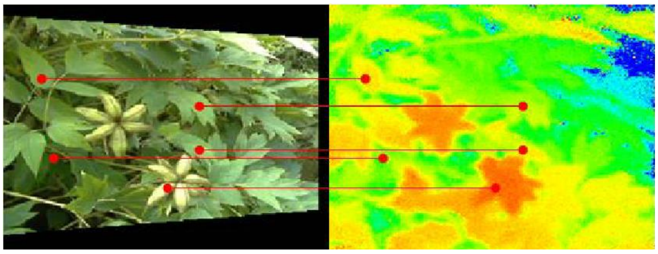


Fig. 6: Result of feature points matching



(a)



(b)

Fig. 7: SURF feature point matching and NDLT; (**a**) Result of matching of SURF feature points; (**b**) Result of matching of the RGB image and the depth image by NDLT

As shown in Fig. 7, this algorithm solved the problem of failed matches caused by the different resolutions of images captured from two cameras. It was able to verify the target object position in two cameras.

Localization and Maturity Classification of Tree Peony Fruit

Localization of Fruit

Image Segmentation

Image segmentation is the basis of the recognition and localization of peony fruit. In this study, the depth information of objects (the distance to the camera) was represented by different colors in the depth image, shown as a pseudo-color image, as shown in Fig. 8. Thus, the differences in location of peony fruits, stems and leaves could be shown in different colors in the depth image. The depth image was transformed into a grayscale image in three RGB color channels, as illustrated in Fig. 8. The difference between the area of oil peony fruit and the background was obvious in the G channel image, which was used to achieve image segmentation in this research.

To segment the fruit and background area, a grayscale threshold value needed to be defined. In this study, the range of the grayscale value of the fruit area of the *G* channel image was 100-180 and the other values represented the background area, as observed by numerous experiments. To achieve image segmentation, the threshold value was defined using Equation (9) and the result of the segmentation is shown in Fig. 9:

$$g(x,y)=\begin{cases}255 & G\in[100,180]\\ 0 & other\end{cases} \tag{9}$$

There were some noise and residue in the image after fixed threshold segmentation (Fig. 9), which was caused by the leaves and other barriers next to the fruits. Therefore, a morphological image processing based on a two-value graph shape operation was used to remove these barriers and the noise of the image to recognize the oil tree fruits accurately. The main stages of the process are described as follows.

Dilation:

$$A\oplus B=\left\{z\,|\,(B)_z\cap A\neq\varnothing\right\} \tag{10}$$

Erosion:

$$A\Theta B=\left\{z\,|\,(B)_z\cap A^c\neq\varnothing\right\} \tag{11}$$

In Equations (10) and (11), A is the image and B is the structure element. The result of the fixed threshold segmentation is illustrated in Fig. 10.

MLP Algorithm

To classify the maturity of tree peony fruit, a classifier of peony fruits based on the Multi-Layer Perceptron (MLP) neural networks algorithm (Mühlenbein, 1990; Kim and Adali, 2002; Rossi and Conan-Guze, 2005) was used in this study, as shown in Fig. 11. The activation function of neural networks is described by Equation (12):

$$h_0\left(x^{(i)}\right)=\begin{bmatrix}P\left(y^{(i)}1\,|\,x^{(i)};\theta\right)\\ P\left(y^{(i)}2\,|\,x^{(i)};\theta\right)\\ \cdots\\ P\left(y^{(i)}k\,|\,x^{(i)};\theta\right)\end{bmatrix}=\frac{1}{\sum_{j=1}^{k}e^{\theta_j^{Tx^{(i)}}}}=\begin{bmatrix}e^{\theta_j^{Tx^{(i)}}}\\ e^{\theta_j^{Tx^{(i)}}}\\ \cdots\\ e^{\theta_j^{Tx^{(i)}}}\end{bmatrix}=\frac{e^{\theta_j^{Tx^{(i)}}}}{\sum_{i=1}^{k}e^{\theta_j^{Tx^{(i)}}}} \tag{12}$$

where the dimension of [θ] represents the group of maturity classification, for which the value is 3, representing the number of possible classifications. Setting *X* as the input sample and *Y* as the output sample, Y_j can be described by Equation (13) while W_{ij} is the weight function from element *i* to element *j*:

$$Y_j=f\left(\sum_{i=1}^{n}W_{ij}X_i\right). \tag{13}$$

(a) (b) (c)

Fig. 8: Grayscale images in three RGB channels of the depth image; (a) R channel image; (b) G channel image; (c) B channel image

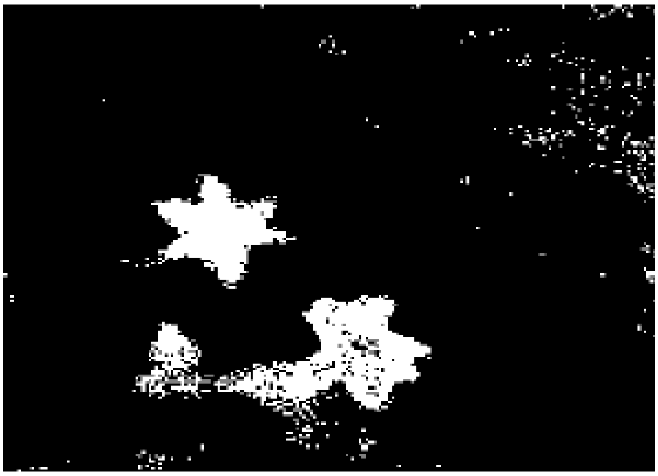


Fig. 9: Result of the fixed threshold segmentation

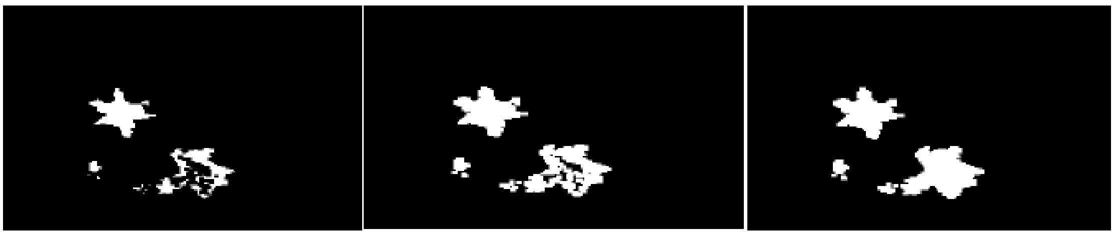


Fig. 10: Fixed threshold segmentation

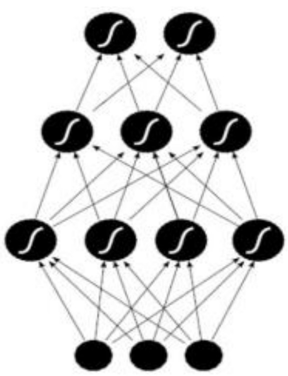


Fig. 11: Multi-Layer Perceptron (MLP) neural networks

The study process from the input layer to hidden layer is shown as follows, while η is the learning rate:

$$W_{ij}(n+1) = W_{ij}(n) + \eta \cdot \delta_j \cdot X_i . \tag{14}$$

The function of the output node is shown in Equation (15), while T represents the expectation output sample:

$$\delta_j = (T_i - Y_i) f'\left(\sum_{i=1}^{n} W_{ij} \cdot X_i\right). \tag{15}$$

The output layer includes two 0-1 neurons to recognize the target as an oil tree peony fruit or not.

Fruit Recognition

To recognize the area of an oil tree peony fruit, the geometry features of the target area were input into the MLP classifier. There were five features extracted from the images after segmentation in this study, including roundness and four normalization center moments.

The mathematical expression of roundness is expressed by Equation (16), in which σ is the mean deviation and $\bar{d}$ is the mean distance:

$$e = \frac{1-\sigma}{d}, \tag{16}$$

$$\sigma^2 = \sum\left(\| P - P_i \| - \bar{d}\right)^2 / F, \tag{17}$$

$$\bar{d} = \sum\left(\| P - P_i \|\right) / F. \tag{18}$$

In Equations (17) and (18), P is the center of the area P_i is the pixel point and F is the perimeter of the area outline.

The four normalization center moments used in this study are described as:

$$\begin{cases} \eta_1 = \dfrac{\mu_{20}\mu_{02} - \mu_{11}^2}{\mu^4} \\ \eta_2 = \dfrac{\left(\mu_{30}\mu_{03} - \mu_{21}\mu_{12}\right)^2 - 4\left(\mu_{30}\mu_{12} - \mu_{21}^2\right)\left(\mu_{21}\mu_{03} - \mu_{12}^2\right)}{\mu^{10}} \\ \eta_3 = \dfrac{\left(\mu_{30}\mu_{03} - \mu_{21}\mu_{12}\right)^2 - 4\left(\mu_{30}\mu_{12} - \mu_{21}^2\right)\left(\mu_{21}\mu_{03} - \mu_{12}^2\right)}{\mu^{10}} \\ \eta_4 = \dfrac{\mu_{30}^2\mu_{02}^2 - 6\mu_{30}\mu_{21}\mu_{11}\mu_{02}^2 + 6\mu_{30}\mu_{12}\mu_{02}\left(2\mu_{11}^2 - \mu_{20}\mu_{02}\right) + \cdots + \mu_{03}^2\mu_{20}^3}{\mu^{11}} \end{cases} \tag{19}$$

$$\eta_3 = \frac{\mu_{nm}}{\mu_{00}^{\frac{(m+n)}{2}}} \tag{20}$$

The results of the features extraction of some training samples are shown in Table 4.

These five geometry features of each area segmented from the images of the training samples were input into the MLP to recognize the area as background (0) or fruit (1).

Localization of Fruits

In order to guide the mechanical arm to harvest the target fruit, it is necessary to locate the fruits after recognition in the images. The centroid coordinates of the contour line of fruits are the target coordinates in this study. As the RGB images had been matched with depth images, the fruit recognition in RGB images could be achieved by image registration. The process of localization is shown in Fig. 12 and the results are shown in Fig. 13.

Fruit Maturity Classification

The RGB images contain complete color and texture information of the oil tree peony fruits, which can be used to classify the maturity of the fruits. The maturity of peony fruits can be divided into three levels: Green ripeness, yellow ripeness and dead ripeness, as illustrated in Fig. 14.

As shown in Fig. 14, the main discrepancy between green ripeness and yellow ripeness is the color feature of the fruits and the difference of texture feature between yellow ripeness and dead ripeness is obvious. Therefore, in the natural environment, color and texture features can be used to classify the maturity of the oil tree fruits. In this study, to achieve the classification, a classifier was set up based on an MLP, described in Section 3.1.2 and the input layer contained color feature and texture feature. This process is shown in Fig. 15. To improve the accuracy and efficiency of the classification model, 60 images of three kinds of peony fruit were captured as the training samples to be used for model training, as shown in Fig. 16.

Experiment

An experiment was conducted by using 90 groups of images captured by the vision system in a natural scene of oil tree fruits as the testing samples to test the accuracy of the algorithm. Each group contained four images which have been matched by the algorithm described above.

Table 4: Results of features extraction of some training samples

	Roundness (e)	η_1	η_2	η_3	η_4
Fruit 1	0.785628	0.00873776	-6.52e-10	-2.41e-06	1.69e-07
Fruit 2	0.785628	0.00874776	-6.52e-10	-2.41e-06	1.69e-07
Fruit 3	0.762452	0.0096116	-4.98e-09	-7.51e-06	6.04e-07
Background 1	0.783294	0.00896887	-1.31e-08	-1.18e-05	8.81e-07
Background 2	0.673865	0.00633478	-2.10e-09	-3.67e-06	1.68e-07
Background 3	0.683868	0.00927532	-1.20e-07	-3.42e-05	2.57e-06
Background 4	0.828427	0.00548697	0	0	0

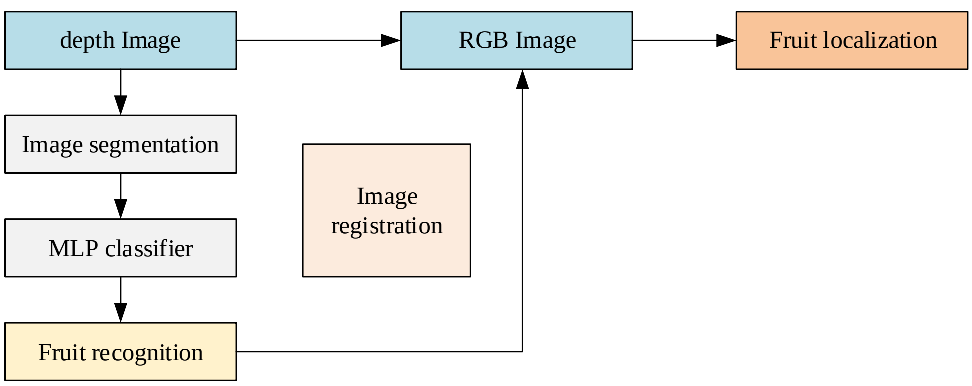


Fig. 12: Process of fruit localization

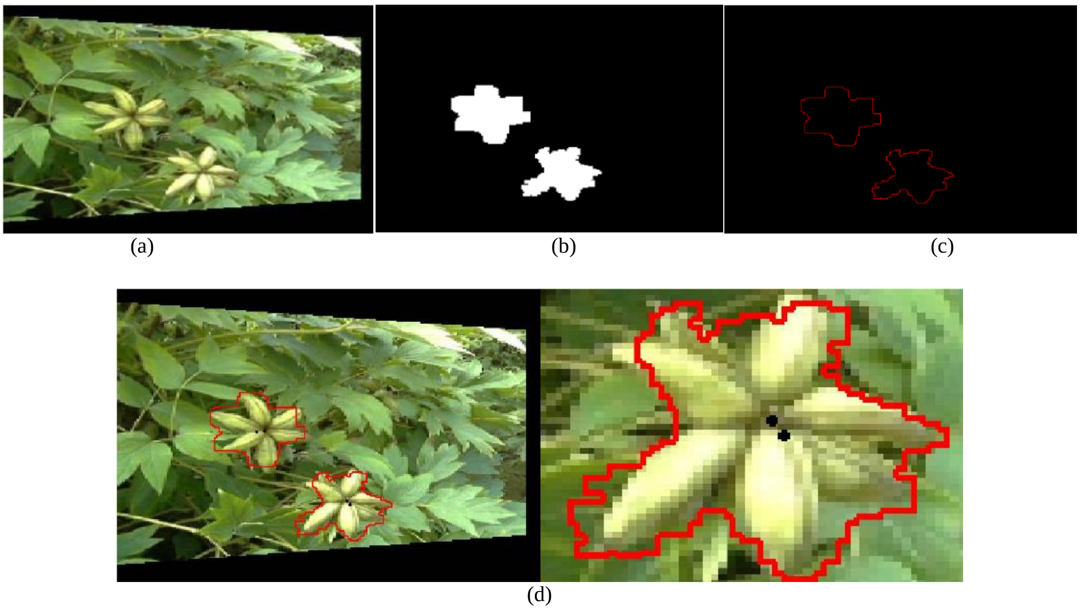


Fig. 13: Results of the localization of oil tree peony fruits. (a) RGB image; (b) result of image segmentation; (c) contour line of fruits; (d)centroid point of fruit

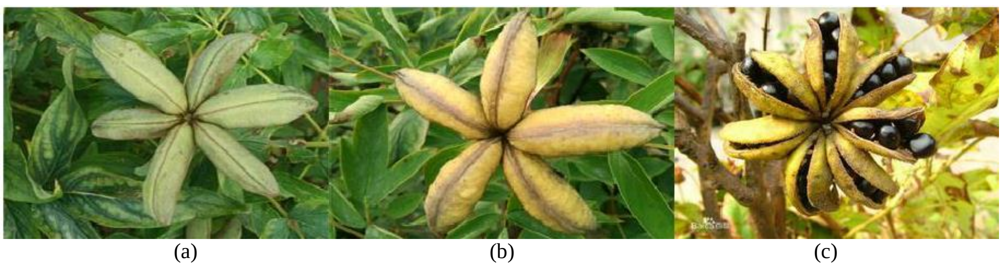


Fig. 14: Different levels of oil tree peony fruit; (a) Green ripeness; (b) yellow ripeness; (c) dead ripeness

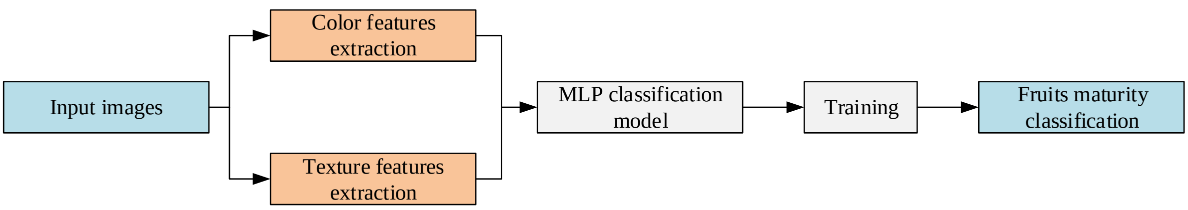


Fig. 15: Algorithm for tree peony fruits maturity classification

The first step was recognition of the fruits. Depth images of each group were input into the MLP classifier to recognize the peony fruits and the total number of peony fruits in the 90 groups of images was 173. The result of fruits recognition is illustrated in Table 5. As can be seen, in the 90 groups of images, there are 173 peony fruits in total. The system recognized 162 fruits, in which contains 5 mistaken recognition. Therefore, 157 fruits were correctly recognized and 11 fruits were missing in this experiment. The total fruit recognition rate was 85.74%. This result shows the vision system achieved high accuracy recognition and can meet the needs of the actual operation.

The centroid coordinates were calculated after fruits recognition and some of the results are shown in Table 6. In this table, (x, y, z) is the calculated centroid coordinates of each fruit recognized by the system and the$e$ value represents the error between calculated coordinates and actual coordinates. The z-coordinate shows the depth information of the peony fruits. In general, the average localization error is 3.53, which is accuracy for harvesting operation.

(a) (b) (c)

Fig. 16: Samples of different maturity levels of peony fruits; (a) Green ripeness; (b) yellow ripeness; (c) dead ripeness

Table 5: Recognition result of oil tree fruits

Image number	Fruits number	Recognition number	Mistake number	Missing number	Fruit recognition rate
90	173	162	5	11	85.74%

Table 6: Example of centroid coordinates of recognized peony fruits

Number	(*x*, *y*, *z*)	*e*	Number	(*x*, *y*, *z*)	*e*
1	(52, 141, 52.91)	3.04	11	(134, 124, 40.54)	3.84
2	(94, 75, 58.72)	4.21	12	(169, 85, 45.87)	2.89
3	(105, 147, 52.98)	3.19	13	(131, 82, 43.94)	2.54
4	(150, 81, 5919)	3.64	14	(51, 127, 38.53)	3.10
5	(65, 94, 51.23)	3.39	15	(145, 47, 39.43)	3.87
6	(110, 143, 47.93)	3.24	16	(122, 141, 48.13)	2.65
7	(68, 120, 37.70)	3.22	17	(38, 133, 35.84)	4.78
8	(82, 140, 44.23)	4.89	18	(69, 56, 39.12)	4.29
9	(110, 60, 49.92)	3.22	19	(184, 122, 4068)	3.75
10	(30, 79, 41.13)	3.64	20	(78, 119, 50.44)	3.14

Table 7: Result of the experiment of maturity classification of tree peony fruits.

Maturity	Sample number	Green ripeness	Result Yellow ripeness	Dead ripeness	Recognition rate	Total recognition rate
Green ripeness	62	57	5	0	91.94%	
Yellow ripeness	55	8	47	0	85.45%	91.68%
Dead ripeness	40	0	0	40	100%	

In the experiment, 157 peony fruits were recognized by using depth images, which were matched with RGB images. Therefore, the fruit areas in the RGB images were recognized and could be used to classify the maturity of peony fruits by the MLP classifier.

The result of maturity classification of these fruits is described in Table 7. As can be seen, for fruits in green ripeness, 57 of 62 fruits were classified correctly with the recognition rate was 91.94%; for fruits in yellow ripeness, 47 of 55 fruits were classified correctly with the recognition rate was 85.45%; for fruits in dead ripeness, all of the 40 fruits were classified correctly. In general, this system achieved a high recognition rate, 91.68% in total, which verified the feasibility and accuracy of classification algorithm of this study.

Discussion

Concerning Images Registration

In section 2, we described an image registration algorithm and set up a multi-source vision system based on RGB and TOF cameras. The algorithm included several stages. SURF was used to detect and extract feature points of intensity and grayscale images. Then, a matching method was described using the nearest-neighbor algorithm, Hessian matrix trace accelerates and RANSAC algorithm, which was used to match the feature points. When feature points matching were achieved, it was feasible to match the grayscale images with the intensity images using the NDLT algorithm, which achieved the matching of RGB and TOF cameras.

The experiment showed that the algorithm can realize high-efficiency and precise image registration applied using the vision system we set up. It achieved the matching of RGB and TOF cameras, which can extract the color and position information of the tree peony fruits, which is the precondition for performing fruits recognition and maturity classification.

Concerning Fruit Localization and Maturity Classification

In section 3, we set up an MLP classifier to recognize and locate peony fruits as well as an MLP maturity classifier to classify the maturity of peony fruits. The G channel images of depth images were used to segment the peony fruits and background input into the MLP classifier, which could recognize the fruits. Then, by the image registration, we located the fruits in the RGB images. These fruits were classified into three maturity levels using a maturity classifier based on MLP to decide whether the fruits were ready to be harvested.

The experiment showed that the algorithm achieved a high fruit recognition rate and maturity recognition rate and is feasible to apply to robotic tree peony fruits harvesting.

Comparison with Related Works

The most important novelty and contribution of the work is to set up the multi-source vision system to combine the advantages of RGB camera and TOF camera and overcome each other's disadvantages.

On the one hand, for example, traditional vision system based on RGB cameras is easy to be affected by the illumination condition changes. In the study of a vision system for plants/weed classification (Jasiński *et al*., 2018), the quality of captured image is good in the sun but obviously worse in shadow, this can be seen in Fig. 17. To improve the image quality, the study has to select the appropriate lighting to reduce the impact of atmospheric conditions. Compared with the work, the vision system in this study has stronger anti-interference ability due to the TOF camera. As described in section 3, images captured from TOF camera are almost unaffected by lighting conditions, which results in a high accuracy of fruit recognition and localization in the case of poor lighting conditions (cloudy, shadow, etc.).

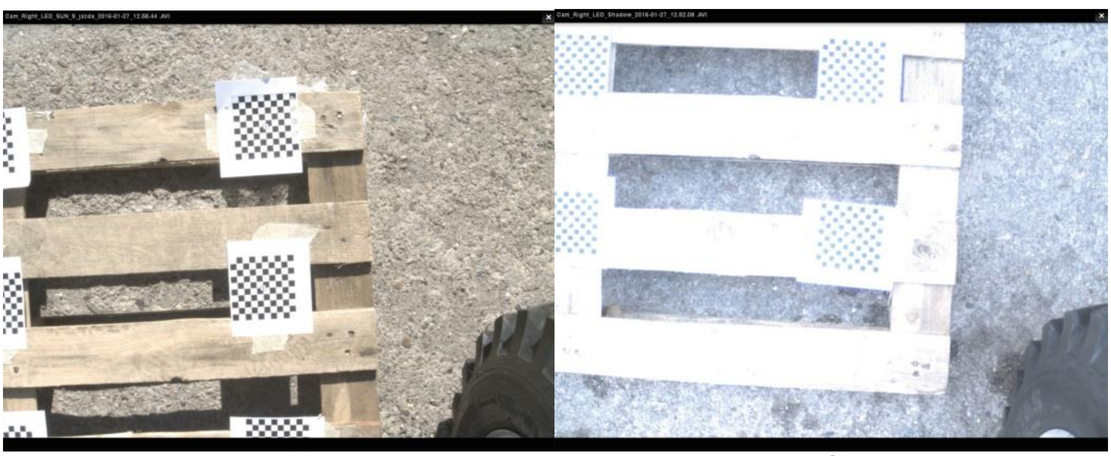
On the other hand, although the TOF camera has been widely used in object recognition and location, the images captured by TOF camera do not have the color information of the objects (Conde, 2020). Because of this deficiency, TOF camera could not be used for maturity classification of fruits. This study solved this problem by registering the images captured by TOF cameras with by RGB cameras. Through the images registration, the vision system can find the location and color information at the same time.

Limitations

We have to point out that there are several limitations to our vision system which could lead to some reliability problems.

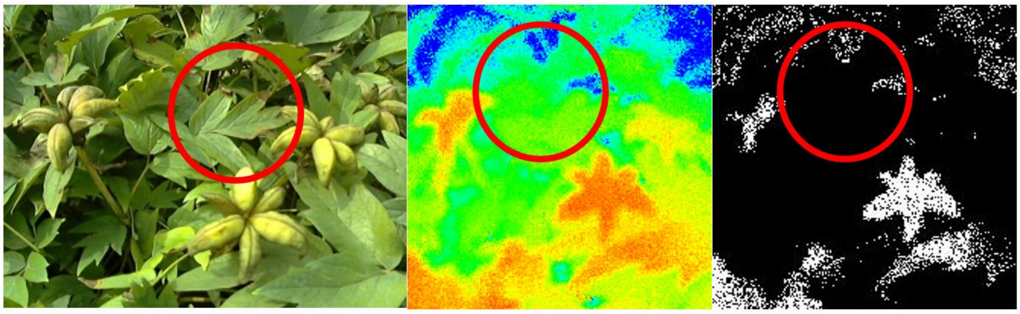
First of all, although we captured images from a natural environment, the growing environment of oil tree peony was too complicated. As an example, in a natural environment, the positions of the fruits are all around the plant, which can cause targets to be missed when the vision system is used to localize all the fruits of the plant. As the features used to recognize the fruits are extracted from depth images, the proposed method is sometimes unable classify the fruits and other objects such as leaves next to the fruits, because they have similar depth information. Figure 18 showing as an example. The fruits in the red circle are showed as the background after image segmentation. As can be seen, there two fruits in the red circle. The latter is covered by the former and the depth of the former is very similar to the leaves around them. That resulted in both the depth image and the segmentation result.

Therefore, we will continue to work to improve the efficiency of the vision system in natural environments. The second limitation concerns the image registration algorithm. As the key stage of the multi-source vision system and the precondition for fruit recognition and maturity classification, the image registration algorithm used in our system may be further optimized to reduce the complexity and improve the efficiency of our method. Moreover, the accuracy of the fruit localization and maturity classification also should be improved by increasing the number of training samples.



(a) (b)

Fig. 17: Pictures captured from the vision system in the related works; (a) is the picture captured in the sun. (b) is the picture captured in the shadow



(a) (b) (c)

Fig. 18: Example of Image segmentation failure; (a) is the RGB image. (b) is the depth image. (c) is the result of image segmentation

Finally, the depth measurement errors of the TOF camera were not well considered in this study. In this study, the depth images captured by the TOF camera were used to detect and locate peony fruit, as described in section 3. Depth measurements obtained by TOF cameras face the occurrence of several systematic and non-systematic errors (Chiabrando *et al*., 2009). As an example, (Foix *et al*., 2011) described in their study that there are five types of systematic errors (depth distortion, integration-time-related error, built-in pixel-related errors, amplitude-related errors and temperature-related errors) and four non-systematic errors (signal-to-noise ratio, multiple light reception, light scattering and motion blurring) that should be taken into account.

Therefore, the main content of further research is to reduce these systematic and non-systematic errors and improve the measurement accuracy.

Conclusion

In this study, a vision system based on TOF and RGB cameras was set up for fruit localization and maturity classification of oil tree peony, which can be applied to robotic tree peony pods harvester. This system can capture 4 kinds of images at the same time (RGB image, depth image, amplitude image and intensity image). The color information of fruit in the RGB image can be used for maturity classification and the location information extracted from the depth image and intensity image can be used for fruit localization. In order to match images captured from two cameras, an image registration method was carried out in this study based on the SURF algorithm and NDLT algorithm. Furthermore, we set up an MLP classifier to recognize and locate the peony fruits and an MLP maturity classifier to classify the maturity of the peony fruits. 90 Groups of tree peony fruits images were captured by this vision system and were used to test the performance of the system. The result shows that the fruit recognition rate of the system is 85.74% and the average calculated location error is 3.53. The recognition rates of maturity classification of 3 different ripeness stages of fruits are 91.94, 85.45 and 100% and the total maturity classification recognition rate is 91.68%. It can be concluded from the experiments that this system achieved a high accuracy of fruit localization and maturity classification and it is capable for a robotic tree peony fruit harvester. Future study would involve improving resolution ratio of the system, recognition rate of fruits with complex growth and efficiency of the system.

Acknowledgement

The fundamental research fund for the central universities (no. 2015zcq-gx-02)

Author's Contributions

Jiaming Liu: Has conceived and designed the experiments, data analysis, manuscript writing and publication.

Dong Zhao: Has reviewed and revised the manuscript.

Tianxing Li: Critical revision of the article.

Jian Zhao: Final approval of article.

Ethics

Authors should address any ethical issues that may arise after the publication of this manuscript.

References

Abdel-Aziz, Y. I., Karara, H. M., & Hauck, M. (2015). Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. Photogrammetric Engineering & Remote Sensing, 81(2), 103-107.

Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). Computer vision and image understanding, 110(3), 346-359.

Bishop, C. (1992). Exact calculation of the Hessian matrix for the multilayer perceptron.

Bulanon, D. M., & Kataoka, T. (2010). Fruit detection system and an end effector for robotic harvesting of Fuji apples. Agricultural Engineering International: CIGR Journal, 12(1).

Bulanon, D. M., Burks, T. F., & Alchanatis, V. (2010). A multispectral imaging analysis for enhancing citrus fruit detection. Environmental Control in Biology, 48(2), 81-91.

Chen, M., & Hashimoto, K. (2017). Vision System for Coarsely Estimating Motion Parameters for Unknown Fast Moving Objects in Space. Sensors, 17(12), 2820.

Chiabrando, F., Chiabrando, R., Piatti, D., & Rinaudo, F. (2009). Sensors for 3D imaging: Metric evaluation and calibration of a CCD/CMOS time-of-flight camera. Sensors, 9(12), 10080-10096.

Chiodini, S., Giubilato, R., Pertile, M., & Debei, S. (2020). Retrieving scale on monocular visual odometry using low resolution range sensors. IEEE Transactions on Instrumentation and Measurement.

Conde, M. H. (2020). A Material-Sensing Time-of-Flight Camera. IEEE Sensors Letters, 4(7), 1-4.

De-An, Z., Jidong, L., Wei, J., Ying, Z., & Yu, C. (2011). Design and control of an apple harvesting robot. Biosystems engineering, 110(2), 112-122.

Ding, Y. (2009). Research on the key technology of kiwifruit recognition and location based on machine vision. Northwest A&F University.

Falie, D., & Buzuloiu, V. (2007, July). Noise characteristics of 3D time-of-flight cameras. In 2007 International Symposium on Signals, Circuits and Systems (Vol. 1, pp. 1-4). IEEE.

Foix, S., Alenya, G., & Torras, C. (2011). Lock-in time-of-flight (ToF) cameras: A survey. IEEE Sensors Journal, 11(9), 1917-1926.

Guo, J. (2009). Research on the object recognition based on SIFT and NDLT. Mater thesis, North University of China, China.

Jasiński, M., Mączak, J., Szulim, P., & Radkowski, S. (2018, March). Autonomous Agricultural Robot–Testing of the Vision System for Plants/Weed Classification. In Conference on Automation (pp. 473-482). Springer, Cham.

Kim, T., & Adali, T. (2002). Fully complex multi-layer perceptron network for nonlinear signal processing. Journal of VLSI signal processing systems for signal, image and video technology, 32(1-2), 29-43.

Kollorz, E., Penne, J., Hornegger, J., & Barke, A. (2008). Gesture recognition with a time-of-flight camera. International Journal of Intelligent Systems Technologies and Applications, 5(3-4), 334-343.

Kumar, R. A., & Rajpurohit, V. S. (2019). Wavelet Features for Pomegranate Sorting Using Machine Vision. In Innovations in Computer Science and Engineering (pp. 179-186). Springer, Singapore.

Lee, C., Song, H., Choi, B., & Ho, Y. S. (2011). 3D scene capturing using stereoscopic cameras and a time-of-flight camera. IEEE Transactions on Consumer Electronics, 57(3), 1370-1376.

Li, J., Wang, T., & Zhang, Y. (2011, November). Face detection using surf cascade. In 2011 IEEE international conference on computer vision workshops (ICCV workshops) (pp. 2183-2190). IEEE.

Li, J., Wang, Y., & Wang, Y. (2012). Visual tracking and learning using speeded up robust features. Pattern Recognition Letters, 33(16), 2094-2101.

Li, L., Liu, H., Xu, Y., & Zheng, Y. (2020, June). Measurement Linearity and Accuracy Optimization for Time-of-Flight Range Imaging Cameras. In 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) (Vol. 1, pp. 520-524). IEEE.

Li, T., Liu, L., Sun, X., & Yu, Z. (2017). Remote sensing image registration based on the Harris-SURF algorithm. Information & Communication, 11, 9-10.

Liu, J., Zhao, D., & Zhao, J. (2020). Study of the Cutting Mechanism of Oil Tree Peony Stem. Forests, 11(7), 760.

Mao, Y., Han, J., Tian, F., Tang, X., Hu, Y., & Guan, Y. (2017). Chemical composition analysis, sensory and feasibility study of tree peony seed. Journal of food science, 82(2), 553-561.

Matas, J., & Chum, O. (2004). Randomized RANSAC with Td, d test. Image and vision computing, 22(10), 837-842.

Mühlenbein, H. (1990). Limitations of multi-layer perceptron networks-steps towards genetic neural networks. Parallel Computing, 14(3), 249-260.

Muja, M., & Lowe, D. G. (2014). Scalable nearest neighbor algorithms for high dimensional data. IEEE transactions on pattern analysis and machine intelligence, 36(11), 2227-2240.

Park, W. J., Ji, S. W., Kang, S. J., Jung, S. W., & Ko, S. J. (2017). Stereo vision-based high dynamic range imaging using differently-exposed image pair. Sensors, 17(7), 1473.

Rakun, J., Stajnko, D., & Zazula, D. (2011). Detecting fruits in natural scenes by using spatial-frequency based texture analysis and multiview geometry. Computers and Electronics in Agriculture, 76(1), 80-88.

Rossi, F., & Conan-Guez, B. (2005). Functional multi-layer perceptron: a non-linear tool for functional data analysis. Neural networks, 18(1), 45-60.

Schnabel, R., Wahl, R., & Klein, R. (2007, June). Efficient RANSAC for point-cloud shape detection. In Computer graphics forum (Vol. 26, No. 2, pp. 214-226). Oxford, UK: Blackwell Publishing Ltd.

Schwarz, S., Sjöström, M., & Olsson, R. (2014). Multivariate Sensitivity Analysis of Time-of-Flight Sensor Fusion. 3D Research, 5(3), 18.

Sharma, N. K., Rathore, S., & Khan, M. R. (2020, January). A Comparative Analysis on Coordinate Rotation Digital Computer (CORDIC) Algorithm and Its use on Computer Vision Technology. In 2020 First International Conference on Power, Control and Computing Technologies (ICPC2T) (pp. 106-110). IEEE.

Su, J., Ma, C., Liu, C., Gao, C., Nie, R., & Wang, H. (2016). Hypolipidemic activity of peony seed oil rich in α-linolenic, is mediated through inhibition of lipogenesis and upregulation of fatty acid β-oxidation. Journal of food science, 81(4), H1001-H1009..

Sun, X., Li, W., Li, J., Zu, Y., Hse, C. Y., Xie, J., & Zhao, X. (2016). Process optimisation of microwave-assisted extraction of peony (Paeonia suffruticosa Andr.) seed oil using hexane–ethanol mixture and its characterisation. International Journal of Food Science & Technology, 51(12), 2663-2673.

Takahashi, M., Fujii, M., Naemura, M., & Satoh, S. I. (2013). Human gesture recognition system for TV viewing using time-of-flight camera. Multimedia tools and applications, 62(3), 761-783.

Wu, S. (2020). Determination of chlorophyll content in rice based on computer vision. Journal of agricultural mechanization research.

Zhao, Y., Gong, L., Huang, Y., & Liu, C. (2016). A review of key techniques of vision-based control for harvesting robot. Computers and Electronics in Agriculture, 127, 311-323.