

Design of intelligent recognition and picking fruit system for 3D virtual robot

HANG XU¹, JIULE TIAN²

Abstract. Recognition system design used for intelligent fruit-picking robots is designed in the Thesis. Two questions proposed initially are solved in the Thesis: (1) Output of smell sensor is optimized by random forest and resolution ratio is enhanced. (2) Algorithm removing original image shadow with nature image is proposed. In overall design, recognition of farm scene is realized by controlling smell sensor and camera with soft hardware; the best effective camera distance of fruit recognition is confirmed to be 50cm-70cm by a number of outdoor tests; picking action is evaluated by combining the results of “gas phase” between smell sensor and camera. The recognition system provides more abundant pattern information for realizing fruit recognition and a technological mean with innovativeness for automation and mechanization of fruit picking.

Key words. Agricultural automation, Robots, Intelligent picking, Image classification, Object recognition.

1. Introduction

China is an agricultural country. Vegetable and fruit industry is an important part of agriculture. There is a vast area of mountainous and hilly land in China and most areas plant fruiters. Picking fruit approximately accounts for more than 50% of the whole amount of production operation. At present, fruit picking mainly employs hand picking with fruit scissors. Hand picking can fairly cope with fruiters that are lower than 2 meters from the ground, but it has great labor strength and low efficiency. As for fruiters that are higher than 2 meters, ladder and other stepstools shall be assisted generally, which is an extremely unsafe method. It can be seen that picking method that relies on manpower has slow picking speed, low efficiency and

¹Department of Mechanical & Automotive Engineering, Zhuhai College of Jilin University, Zhuhai City Guangzhou Province, ZipCode: 519041, China

²Department of Logistics & Information Management, Zhuhai College of Jilin University, Zhuhai City Guangzhou Province, ZipCode: 519041, China

certain dangerousness. In terms of fruit quality, falling pedicle and broken condition may appear due to uneven stress. It will influence appearance of fruit and it is not beneficial to storage, processing and sale of fruit thereby reducing economic income of orchard workers. Application of intelligent picking robots is one of important ways to solve the above questions. In view of the above reasons, a picking recognition system of intelligent robots is designed in the Thesis which is aimed at providing a technological mean with innovativeness for automation and mechanization of fruit picking. The scheme proposed in the Thesis uses cameral to collect image for fruits, uses smell sensor to collect smell for fruits, combines color phase and gas phase for data fusion and makes it as realization of object target recognition.

2. Systematic design

Functionally, the system is mainly composed of data collection, data preprocessing, single chip microcomputer, picking actuator, upper computer and data communication. It is shown in Fig. 1.

The requirements of data processing for lower computer and single chip microcomputer for data communication are high in systematic design. STC89LE58AD is adopted in the Thesis. There is AD converter with 8 channels and 8 bits in the interior of STC single chip microcomputer for the 89 series and it is distributed in 8-bit of P1 port. When the clock is under 40 MHz, an AD conversion can be finished every 17 machine cycles.

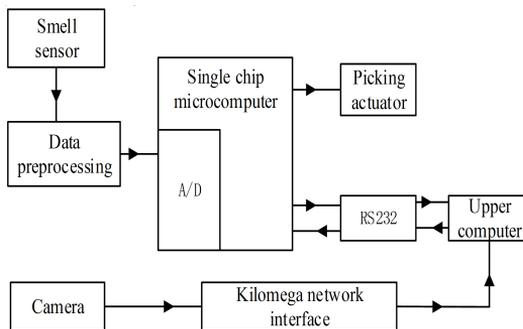


Fig. 1. Recognition system structure figure

Smell sensor also called electronic nose simulates human olfactory system. The structure is divided into three levels: ① Gas sensor array that is equivalent to primary olfactory sensory neurons; ② Signal preprocessing unit; ③ Pattern recognition unit that is equivalent to human brain. It is shown in Fig. 2. TGS2630, TGS825 of TGS series and MQ138, MQ-9 gas sensor of MQ series are selected for smell sensor in the design.

CCD camera is the module of key acquiring image for mainly existing machine vision. JAI works to provide abundant product selection from line scanning to surface scanning, from simulation to digit, from Camera Link to GigE Vision for machine vision users. JAI CM-030GE /CB-030GE digital cameras are adopted in

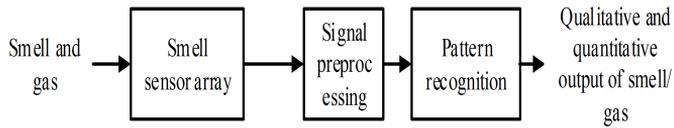


Fig. 2. Function structure figure of smell sensor

the Thesis. The camera uses Sony IC424 CCD sensor with size of 1/3" and effective pixels of 656*494. Its maximum frame rate can reach 90fps under the pattern of maximum resolution ratio and consecutive collection. GigE Vision interface is used to communicate with outside world.

Design thought in the Thesis: shoot objects in fruiters with a camera. Camera transfers the images to upper computer by kilomega network interface. Upper computer receives images and processes images as algorithm. Then learn and judge ultimate production of color phase result and gas phase result, send the information to single chip microcomputer by RS232 and single chip microcomputer controls picking of picking mechanical structure.

3. Smell sensor system based on random forest

3.1. *Random forest*

Random forest can select the category with more votes as judging result by collecting votes of each category by child nodes of many trees. Regression problem will be solved by calculating average value of values in all child nodes for the forest. Elementary subsystem at the time of establishing random forest is also decision tree and it will decide continuously until clean data. For example, object recognition tree in the Thesis can have the following characteristics: color, fruity flavour and so on. Each node of the tree can randomly select subsets from these characteristics to determine how to split the data best. In order to improve robustness, random forest inspects division with out of bag method. As for any given nodes, training is implemented in a subset of data with random choice and substitution. The data that are not chosen are called OOB data which will be used for estimating performance of division. OOB data are 1/3 of all data generally. Scale of random subset is radication of feature number generally.

Growth process in random forest algorithm can be described as:

(1) Assuming that sample set number of original training is sum, select features and establish feature space to create a classification tree;

(2) Random selection feature splits internal nodes of classification and regression tree: assuming that there are X features in total, assign a positive integer, extract Y features randomly from X features as candidate features in each internal node, choose the best division mode from Y features to split nodes and keep the value of Y unchanged in growth process of the whole forest.

(3) Branches of each tree shall not be trimmed away and grow furthest until clean data.

3.2. Systematic design of smell sensor based on random forest

Schematic diagram of smell sensor system based on random forest is shown in Fig. 3. Take a sample of mixed gas with different concentration by multiple sensors, take these outputs as input of random forest and each component concentration of corresponding mixed gas as output of the random forest until output error of forest reaches accuracy. At this time, each tree in the forest stores mapping relation from sensor array signal to gas concentration. It is similar to inverse mapping of sensing process and can be used to recognize unknown patterns of arrays.

As shown in Fig. 3, assuming that mixed gas contains N gases and corresponding concentration of each gas is $\{c_0, c_1, \dots, c_{N-1}\}$, output of sensor for each gas in sensor array is $\{y_0, y_1, \dots, y_m\}$, output of gas sensor array can be expressed as:

$$\begin{cases} y_0 = f(c_0, c_1, \dots, c_{N-1}) \\ y_1 = f(c_0, c_1, \dots, c_{N-1}) \\ \dots \\ y_m = f(c_0, c_1, \dots, c_{N-1}) \end{cases}$$

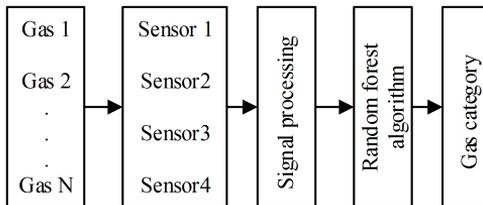


Fig. 3. Schematic diagram of smell sensor system based on random forest

There is a test that recognizes oranges and lemons in design of the Thesis. Take a tree in the forest as example to present decision tree now. It can be seen from Fig. 4 that the decision tree is so complex that error in the training set is 0% (it can be seen that E is 100% or P is 100% from the bottom of Fig. 4) and decision tree can learn and train better, but it also learns the sound and error (see Fig. 4 in Chapter 5). Therefore, cutting of trees must be restrained to reach balance of complexity and performance (such as control of iterations for improving contrast accuracy).

4. Object recognition based on shadow removal

Basis of fruit picking: apart from judging from smell sensor, we can also judge from fruit color. Orchard workers pick fruits according to fruit color in actual picking at present. When picking, orchard workers often have an action that is pushing aside part leaves covering in the fruits to make out maturity of fruits. How will intelligent robots perform the action in the design or must it perform the action? The answer is no. We can “push aside” covered leaves by picture processing. That is shadow removal in picture processing.

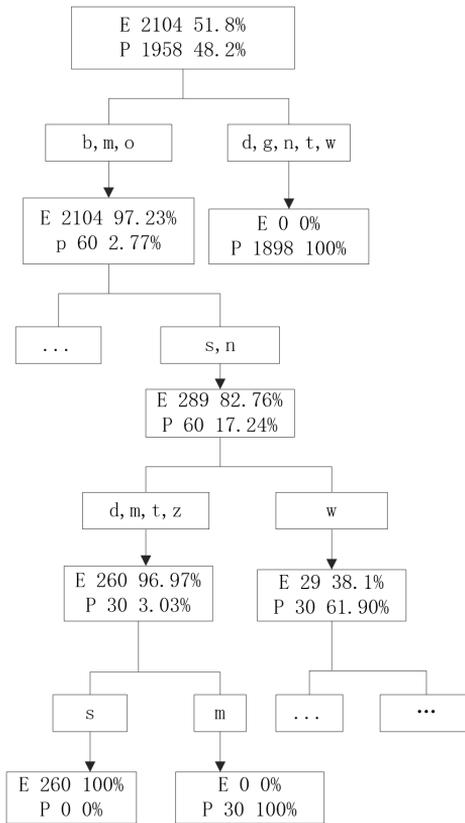


Fig. 4. A determine tree of the forest recognizing oranges (E) and lemons (P)

At present, images collected by most video capture devices are based on *RGB* color space, but there is high correlativity, more redundancy information and large calculation amount among each component in the space. In consideration of increase of computation time as complexity of scene, characteristics method based on *YC_bC_r* color space is chosen in the design. Luminance signal of *YC_bC_r* space and chroma signal *C_b, C_r* are mutual independent. *YC_bC_r* color space is chosen in the Thesis to calculate original shot images and corresponding nature images and remove shadow. *YC_bC_r* decomposition technique has been very mature at present. How to acquire nature images is the key of the Thesis. The design algorithm idea is given below.

Shadow is generated by light source sheltered by objects. Light source is reflected to object surface and image is shot with a camera. Any channel of three channels in the image can be described with the following expression:

$$\rho_m(x, y) = \delta(x, y) \int E(\lambda, x, y) S(\lambda, x, y) Q_m(\lambda) d\lambda, \quad m \in \{r, g, b\}. \quad (1)$$

Where $E(\lambda, x, y)$ is spectral power distribution function, $S(\lambda, x, y)$ is surface re-

flection function, $Q_m(\lambda)$ is frequency sensitivity function of camera, (x, y) is pixel position and λ is wave length.

Assuming that frequency sensitivity function of camera for each component of R, G, B is Dirac delta function in general case,

$$Q_m(\lambda) = q_m\delta(\lambda - \lambda_m). \tag{2}$$

Put equation (2) into equation (1), then (1) can be expressed as:

$$\rho_m(x, y) = \delta(x, y)E(\lambda_m, x, y)s(\lambda_m, x, y)q_m. \tag{3}$$

For convenient description, equation (3) can be simplified into the following equation:

$$\rho_m = \delta E(\lambda_m)s(\lambda_m)q_m. \tag{4}$$

Under the condition of Planck light source, spectral power distribution function of illumination in equation (1) can be expressed into the following equation:

$$E(\lambda, T) = \frac{2\pi hc^2}{\lambda_m^5} \frac{1}{e^{\frac{hc}{\lambda_m kT}} - 1}. \tag{5}$$

Where $h = 6.626 \times 10^{-34}$ joule/s, $c = 3.0 \times 10^8$ meter/s, $k = 1.381 \times 10^{-23}$ joule/Kelvin, color temperature T expresses temperature.

Put equation (5) into equation (4), then (4) can be expressed as:

$$\rho_m = \delta \frac{2\pi hc^2}{\lambda_m^5} \frac{1}{e^{\frac{hc}{\lambda_m kT}} - 1} s(\lambda_m)q_m. \tag{6}$$

Where δ is Lambert projection item. It is dot product of projection surface and illuminating direction. Superficial reflection model is combination of scattering surface and mirror reflective surface. Both extreme situations that are Lambert model and ideal mirror reflection are considered in the Thesis. δ_1 and δ_2 correspond to coefficients of Lambert model and ideal mirror reflection respectively. Equation (6) can be expressed in the following equation again:

$$\rho_m = \rho_{m1} + \rho_{m2} = (\delta_1 + \delta_2 s(\lambda_m)) \frac{2\pi hc^2}{\lambda_m^5} \frac{1}{e^{\frac{hc}{\lambda_m kT}} - 1} q_m. \tag{7}$$

Take logarithm of equation (7),

$$\rho'_m = \log(\rho_m) = \log\left[\frac{(\delta_1 + \delta_2 s(\lambda_m))q_m}{\lambda_m^5}\right] + \log(2\pi hc^2) - \log\left(e^{\frac{hc}{\lambda_m kT}} - 1\right). \tag{8}$$

Make $X_m = \log\left[\frac{(\delta_1 + \delta_2 s(\lambda_m))q_m}{\lambda_m^5}\right]$, $Y_m = \log\left(e^{\frac{hc}{\lambda_m kT}} - 1\right)$, $A_1 = \log(2\pi hc^2)$, $A_2 = \frac{hc}{k}$, $Y_m = \log\left(e^{\frac{A_2}{\lambda_m T}} - 1\right)$. Equation (8) can be simplified into the following

equation:

$$\rho'_m = X_m + A_1 - \log(e^{\frac{A_2}{\lambda_m T}} - 1). \tag{9}$$

Due to $T \in [2500, 10000]^0 K$, $e^{\frac{A_2}{\lambda_m T}} \gg 1$ can be derived, $\log(e^{\frac{A_2}{\lambda_m T}} - 1) \approx \log(e^{\frac{A_2}{\lambda_m T}})$.

Equation (9) can be simplified into:

$$\rho'_m = X_m + A_1 - \frac{A_2}{\lambda_m T}. \tag{10}$$

According to definition of chroma, quantity of RGB forming any special color is called tristimulus value. Such a color can be defined by its tristimulus value coefficient. Chroma of images and projection surface can be calculated as follows:

$$\rho''_m = \frac{\rho'_m}{\rho'_r + \rho'_g + \rho'_b}. \tag{11}$$

$$\rho''_m = \frac{\rho'_m}{\rho'_r + \rho'_g + \rho'_b}, \quad m \in \{r, g, b\}$$

$$X'_m = \frac{X_m}{X_r + X_g + X_b}. \tag{12}$$

$$\sum X'_m = 1 \quad m \in \{r, g, b\}.$$

Transform equation (11) and equation (12) into the following equations:

$$\rho'_m = \rho''_m(\rho'_r + \rho'_g + \rho'_b). \tag{13}$$

$$X_m = X'_m(X_r + X_g + X_b). \tag{14}$$

Put equation (13) and equation (14) into equation (10), the following equation can be acquired:

$$\rho'_m = \rho''_m(\rho'_r + \rho'_g + \rho'_b) = X'_m(X_r + X_g + X_b) + A_1 - \frac{A_2}{\lambda_m T}. \tag{15}$$

Make $m = r, g, b$ respectively, superpose three channels and the following equation can be acquired by final addition:

$$(\rho'_r + \rho'_g + \rho'_b) = (X_r + X_g + X_b) + 3A_1 - \frac{A_2}{\lambda_r T} - \frac{A_2}{\lambda_g T} - \frac{A_2}{\lambda_b T}. \tag{16}$$

Make $\sum \rho' = \rho'_r + \rho'_g + \rho'_b$, $\sum X_m = X_r + X_g + X_b$, equation (16) can be simplified into:

$$\rho''_m \sum \rho' = X'_m \sum X + A_1 - \frac{A_2}{T \lambda_m}. \tag{17}$$

Put equation (14) into equation (17), the following equation can be acquired:

$$T = \frac{[X'_m(\frac{1}{\lambda_r} + \frac{1}{\lambda_g} + \frac{1}{\lambda_b}) - \frac{1}{\lambda_m}]A_2}{\rho'_m - X'_m \sum \rho' + (3X'_m - 1)A_1}. \quad (18)$$

Finally put equation (18) into equation (19), and ultimately get:

$$\rho'_m = X_m + A_1 - \frac{\rho'_m - X'_m \sum \rho' + (3X'_m - 1)A_1}{[\frac{X'_m}{\lambda_m}(\frac{1}{\lambda_r} + \frac{1}{\lambda_g} + \frac{1}{\lambda_b}) - 1]}. \quad (19)$$

In general case, X'_m is aimed at different pixel value. Assuming that X'_m has an initialization value in program design, nature image can be acquired by calculation of equation (19) in this way.

The left picture of Fig. 5 is perspective view of fruiter. It is easy to be found that shadows on the road surface have been removed after algorithm of the Thesis is used. Simultaneously definition of image is increased relatively and leaves on the ground are clearly visible after the shadow is removed.



Fig. 5. Effect picture of shadow removal

5. Test and result analysis

Correct cognition rate and error cognition rate are used to measure object recognition effect in the Thesis. Image processing time is introduced in consideration of real-time requirements of image recognition for picking robots in the process of actual picking. The definitions of three indexes are as follows: (1) correct cognition rate: ratio between quantity that really belongs to fruit pixel in divisional pixels by prediction and quantity of total pixels for fruits. (2) Error cognition rate: ratio between quantity that not belongs to fruit pixel in divisional pixels by prediction and total quantity of pixel that is splitted. (3) Image processing time: the time used for splitting fruit correctly. It is easy to be found in table 1 that the shorter image recognition time is, the better instantaneity of algorithmic will be; the higher value of correct recognition rate is, the more object region can be recognized; the lower value of error recognition rate is, the less error recognition will be. PC configurations used in the test are basic frequency of 2.20 GHz, internal storage of 2GB, operating system of Microsoft Windows XP and programming environment of Visual Studio 2010.

Table 1. Test result of different picture processing under the same test condition

Fig. No.	Fruit pixel	Fruit pixel in decomposition diagram	Non-fruit pixel in decomposition diagram	Correct cognition rate	Error cognition rate	Processing time
6	162032	150932	24078	93.15	14.86	0.0072
7	40508	39710	4197	98.03	10.36	0.0054
8 up	185340	181078	40553	97.70	21.88	0.0065
8 down	158050	154367	27810	97.67	17.59	0.0065
9	162032	123792	51655	76.40	31.88	0.0089

Fig. 6 and Fig. 7 are scene graphs of orange picking, where image a is original image, image b is recognition image acquired by nature image, image c is target locking image. The geometric Fig. drawn by yellow line is target area which is convenient for picking with mechanical hand. Fig. 6 is the result of algorithm recognition when objects are dense. Fig. 7 is the operating recognition result when one object is in foreground and the other is in background. Fig. 8 provides scene graph of lemons picking and processing figures in a similar way.

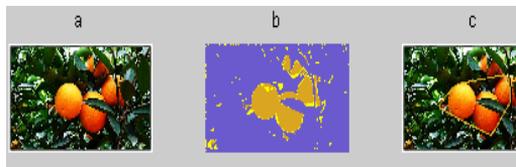


Fig. 6. Operating result when several oranges are crowded together

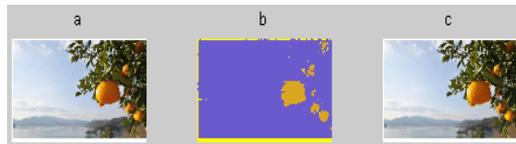


Fig. 7. Operating result when an orange is in the foreground



Fig. 8. Processing effect picture of lemon trees

Without doubt, algorithm in the Thesis still has defects. Drifting objects will be

caused and it is difficult to reach expected objects when fruit color in the image is similar to the color of background. The example of fuzzy object recognition given in Fig. 9 just brings confusion to image processing. Immature lemons will not be picked by robots combining the result of smell sensor in the above section.

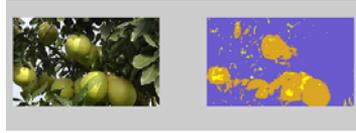


Fig. 9. Examples of unsuccessful lemon recognition

6. Conclusion

In order to improve accuracy of fruit recognition and reduce influence of surrounding environment to recognition, recognition system used for intelligent fruit picking robots is designed in the Thesis. Random forest is adopted to learn optimizing output of smell sensor with the purpose of recognizing smell better; original images collected by camera is used for calculating corresponding nature image to analyze and recognize shadow removal for collected images especially images with shadow in the Thesis. Finally, picking action is evaluated by combining “gas phase” results of smell sensor and camera. The test proves that the recognition system provides comparatively accurate reference for later fruit picking.

Acknowledgement

Premier-Discipline Enhancement Scheme Supported by Zhuhai Government Fund.

References

- [1] FAN X, WANG X, XIAO Y: (2014) *A combined 2D-3D vision system for automatic robot picking*[C]// International Conference on Advanced Mechatronic Systems. IEEE, 2014:513-516.
- [2] SONG K T, WU C H, JIANG S Y: (2017) *CAD-based Pose Estimation Design for Random Bin Picking using a RGB-D Camera*[J]. Journal of Intelligent & Robotic Systems, 2017:1-16.
- [3] HOLZ D, NIEUWENHUISEN M, DROESCHEL D, ET AL.: (2014) *Active Recognition and Manipulation for Mobile Robot Bin Picking*[M]// Gearing up and accelerating cross-fertilization between academic and industrial robotics research in Europe:. Springer International Publishing, 2014:133-153.
- [4] PRETTO A, TONELLO S, MENEGATTI E: (2013) *Flexible 3D localization of planar objects for industrial bin-picking with monocular vision system*[J]. 2013:168-175.
- [5] SHENG-PAN X U, LIU Z J, ZUO Z Q: (2013) *Study on quick picking technology for massive 3D point cloud*[J]. Computer Engineering & Design, 34(8):2764-2768.
- [6] BACHCHE S, OKA K: (2013) *Distinction of Green Sweet Peppers by Using Various Color Space Models and Computation of 3 Dimensional Location Coordinates of Recog-*

- nized Green Sweet Peppers Based on Parallel Stereovision System*[J]. Journal of System Design & Dynamics, 7(7):178-196.
- [7] CAROTENUTO R, CALIANO G, SAVOIA A S: (2013) *3D locating system for Augmented Reality glasses using coded ultrasound*[C]// Ultrasonics Symposium. IEEE, 2013:441-444.
- [8] QIAN K, MA X, DAI X, ET AL.: (2014) *Knowledge-enabled decision making for robotic system utilizing ambient service components*[J]. Journal of Ambient Intelligence & Smart Environments, 6(6):5-19.
- [9] MING Z: (2014) *The Kinematics Simulation Analysis of a Spike-Picking Roller for a New Self-Propelled Maize Harvester Based on ADAMS*[J]. Applied Mechanics & Materials, 496-500:841-844.
- [10] GU B, JI C, WANG H, ET AL.: (2012) *Design and experiment of intelligent mobile fruit picking robot*[J]. Nongye Jixie Xuebao/transactions of the Chinese Society of Agricultural Machinery, 43(6):153-160.
- [11] LIN R, WU N, ZOU J, ET AL.: (2012) *Interactive editing technique and its implementation for virtual tree structure*[J]. Nongye Gongcheng Xuebao/transactions of the Chinese Society of Agricultural Engineering, 28:175-180.
- [12] ZHANG F, LI Z, WANG B, ET AL.: (2014) *Study on recognition and non-destructive picking end-effector of kiwifruit*[C]// Intelligent Control and Automation. IEEE, 2014:2174-2179.
- [13] WEI X, JIA K, LAN J, ET AL.: (2014) *Automatic method of fruit object extraction under complex agricultural background for vision system of fruit picking robot*[J]. Optik - International Journal for Light and Electron Optics, 125(19):5684-5689.
- [14] TEIXIDÓ M, FONT D, PALLEJÀ T, ET AL.: (2012) *An Embedded Real-Time Red Peach Detection System Based on an OV7670 Camera, ARM Cortex-M4 Processor and 3D Look-Up Tables*[J]. Sensors, 12(10):14129-14143.

Received May 7, 2017

