Research on the Application of Machine Vision in Tea Autonomous Picking

Jun Tian¹, Honglin Zhu¹, Wenjiang Liang², Jingshan Chen¹, Feijuan Wen¹ and Zhang Long¹, *

¹School of Engineering, Southwest Petroleum University, Nanchong 637100, China. ²School of Information, Southwest Petroleum University, Nanchong 637100, China.

*Corresponding author: longzhang@swpu.edu.cn

Abstract. As a non-destructive, real-time, fast, objective, and economical detection method, machine vision technology has been widely used in target recognition and positioning in various fields in recent years. Selective picking of tea buds is an important prerequisite for high-quality and efficient tea production. Autonomous picking of tea based on machine vision has become a research hotspot at home and abroad. At present, many scholars have studied the visual recognition and positioning methods in the process of intelligent tea picking, but the effect has not yet reached the actual picking requirements. In order to give everyone a comprehensive understanding of the existing research methods, this article systematically reviews the tea bud segmentation and bud picking point positioning methods based on machine vision technology, and prospects the application prospects and directions of machine vision technology in tea picking.

Keywords: Machine Vision; Tea Picking; Visual Recognition and Localization; Image Processing; Image Segmentation.

1. Introduction
With the rapid development of the tea industry, the requirements for high-quality tea production are getting higher and higher [1]. Tea harvesting is the prerequisite for tea production. Similar to the harvesting time of other crops, tea harvesting is also seasonal and asynchronous. The traditional manual tea picking method is extremely inefficient and consumes farming time, and the quality of tea varies from person to person, and some novices often fail to meet the requirements [2]. Although the mechanical tea picking method improves work efficiency, due to lack of autonomy, the young shoots and old leaves are cut across the board, and the quality of tea is not guaranteed, which destroys the integrity of the shoots [3, 4]. At the same time, it also damages the body of tea trees and reduces the output of tea in the coming year. With the rise of machine vision technology, a new method has been provided to solve the problem of selective picking with quality and quantity during the tea picking process [5-7].

The key to selective tea picking lies in the identification and positioning of tea buds. To ensure the integrity of the bud identification and the accurate positioning of the picking points are the prerequisites for intelligent picking. Domestic and foreign scholars have carried out a large number of related technical researches on this issue and achieved certain results [8-10], but there is still a certain distance
from practical application. In order to facilitate researchers to understand the current research progress on the identification and positioning of tea buds, this article reviews and analyzes the existing research results. This paper summarizes the methods and characteristics of tea buds segmentation and picking point location, and analyzes the technical difficulties of tea buds identification and location, which provides a reference for tea selective picking research.

2. Image Segmentation Method Of Tea Buds

2.1. Image Segmentation of Tea Buds based on Threshold Method

Threshold segmentation is a common algorithm for directly segmenting an image [11], and the key is how to choose an appropriate threshold. The methods of threshold selection mainly include: the method of selecting the threshold using the principle of maximum correlation, the method based on the stable state of the image topology, the gray-level co-occurrence matrix method, the maximum entropy method and the peak-valley value analysis method. In most cases, the choice of threshold will use a combination of two or more methods. The grayscale interval is usually selected by the grayscale histogram of the image. There are only two parts of the target and the background in an image, and the grayscale histogram presents obvious double peaks. At this time, the valley is set as the threshold value to $T$, then the segmentation expression can be written as:

$$f(i, j) = \begin{cases} 
255, & f(i, j) < T \\
0, & f(i, j) \geq T 
\end{cases}$$

The image processed by formula (1) is divided into black and white parts, that is, the target is separated from the background. The global threshold value only depends on the image gray value, and only related to the nature of each image pixel. The local threshold depends on the gray value of the image and some local characteristics of the neighborhood of the point, that is, related to the characteristics of the local area. The dynamic threshold or the adaptive threshold not only depends on the gray value of the image and some local characteristics of the neighborhood of the point, but also on the spatial coordinates, that is, the obtained threshold is related to the coordinates.

2.2. Tea Image Segmentation Combining Color and Region Growth

In order to reduce the influence of the image brightness change on the image segmentation, the image is converted from the RGB color space to the HSI color space [12,13], and the H and S parameters in the HSI model are selected as the characteristics of the buds image, and the seed region is selected. After determining the seed region, based on the characteristics of the color similarity of tea buds, regional growth is performed. Combined with the feature that the regional growth algorithm is basically the same in the same target region and its neighboring block features, multiple sub-blocks of the image are scanned and merged, and the neighbors between the regions performs region growth and region merging according to reasonable merging and stop merging rules, and after morphological processing, small points and holes are removed, so as to achieve rapid and accurate segmentation of tea bud images.

The implementation process of tea image segmentation method combining color and region growth includes: converting the original RGB color image into HSI space, selecting H and S parameters for preliminary seed, then performing seed region growth based on color similarity and region adjacency, and combining the color distance and the edge distance to grow and merge the region, and finally the tea buds of the images acquired from different angles are segmented and compared. The process of completing the tea bud segmentation algorithm is shown in Figure 1.
2.3. Tea Bud Segmentation based on Deep Learning

Aiming at the problems of traditional machine vision-based tea bud detection methods such as poor robustness and low accuracy of manual feature extraction, the deep learning-based target detection YOLO algorithm was applied to the image detection of tea buds under complex backgrounds [14]. And the YOLO network architecture is improved from multi-scale detection. The related technical methods of tea bud segmentation based on deep learning are as follows:

(1) Image pre-segmentation

The input image is pre-segmented using the method based on RGB color information to extract the approximate area of the shoots, and to suppress the background areas such as old leaves and branches in the image. The input RGB image is converted into a single channel gray image by extracting the ultra-green features of tea [15]. The image of the bud area is obtained finally by using the OTSU algorithm to separate the bud area from the background area (old leaves, soil, stem), and mask the background area [16].

(2) YOLO algorithm

The YOLO target detection framework is a single-step detection algorithm that treats the target position frame and classification prediction as a regression problem. Compared with the two-step framework such as Raster-CNN, FPN, etc., in terms of equalization accuracy and speed, the YOLO algorithm can achieve real-time target detection with the same accuracy [17]. The YOLO algorithm unifies the feature extraction network, location box prediction, and category prediction into one framework, which can realize end-to-end training.

In the YOLO framework, the position box, the classification regression network and the image feature extraction network are directly fused into a model. The size of the feature response graph output by the model is $S \times S$ and the dimension is $S \times S$, where $S$ is the number of grids. The frame effect diagram of the YOLO algorithm is shown in Figure 2. In the training stage, it can be equivalent to dividing the input tea bud image into $S \times S$ grids, and then each grid is responsible for predicting $B$ different scales coordinates of the detection frame, the confidence of the detection frame and the probability of each category.

(3) Improved feature extraction network architecture

The feature extraction network of the YOLO series algorithm mainly includes Darknet-9 of YOLOv2. Darknet consists of 19 convolutional layers and 5 pooling layers. The convolution kernel size is 3×3 and

![Figure 1. The processing process of tea bud segmentation algorithm.](image)

![Figure 2. YOLO frame.](image)
1×1, and the step size is 1. For the complex background problem of tea bud image recognition, considering the accuracy of target recognition, based on the Darknet-9 classification network and the pre-segmentation process, an improved tea bud detection network is proposed. The improved network structure for the detection process of tea buds is shown in Figure 3. The improved Darknet only performs predictive regression on two scales of 13×13 and 26×26. The feature extraction layer is composed of multiple residual network layers, and the size of the convolution kernel in each bottleneck module is composed of 3×3 and 1×1.

![Figure 3. The process of improved network structure.](image)

3. **Positioning METHOD OF Tea BUDS PICKING POINT**

The identification of tea buds is the prerequisite for intelligent picking of fresh tea, and the extraction of information about the picking position of the buds after identification is also a key technology for intelligent picking.

4. **Two-Dimensional Picking Coordinate Extraction Method based on Tea Image**

Firstly, taking a picture of the tea tree, and then the position information of the tea buds in the horizontal projection plane is determined according to the image [18]. The imaging principle diagram is shown in Figure 4, point C is the actual position of the tea bud, point A is the position of the tea bud on the calibration surface determined by the image, and the actual projection position of the tea bud on the calibration surface is C'. Therefore, the picking point A of the tea buds determined by the image on the calibrated object surface has a deviation AC' from the actual picking point C'. Suppose the depth of the tea buds is CC'=20 mm. For different camera angles of view, the deviation can be calculated according to the formula AC'= CC'*tan(θ/2).

![Figure 4. The diagram of camera imaging principle diagram.](image)

The circumscribed rectangle of a geometric image is a rectangle that completely contains all the points and lines on the figure, and each side is in contact with the figure. There are infinitely many circumscribed rectangles of a figure, and the one with the smallest area is called the smallest circumscribed rectangle [19]. The picking point of the tea bud can be determined by solving the minimum circumscribed rectangle of the outer contour of the tea bud. The common method for solving the minimum circumscribed rectangle is the target rotation method.
A polygon and the angles between its sides and the horizontal axis are shown in Figure 5(a). The pentagon is a convex polygon, so its convex hull is itself. Selecting the 30° side in Figure 5(a) as the start edge of the rotation, rotating the polygon by 30° and find its circumscribed rectangle in the coordinate system direction, the result is shown in Figure 5(b). After reading the images in sequence, the connected regions (each tea bud forms a connected region) are marked, and then the minimum area circumscribed rectangle is solved for each connected region and the center point is marked.

5. Positioning of Tea Buds based on Raster Projection Profile Technology

Grating projection profilometry uses the time-phase method to obtain the phase of the grating field, and uses the discriminant mechanism and morphological filter to remove the noise points, and uses the polynomial approximation method to calculate the height to overcome the nonlinear error [20]. The grating projection method projects a grating field with periodic distribution in space. In the measurement, the phase is used to describe the spatial distribution of the grating field, and the phase is calculated in the fringe image to obtain the three-dimensional coordinates of the point. Since the phase is continuously distributed in space, the grating projection method can directly measure the entire frame through one projection, which is a prominent advantage of the phase method. The measurement system is shown in Figure 6. The work of the whole system can be divided into two steps: the identification of the tea shoots, determining the existence of the tender shoots of the tea and giving the horizontal coordinates; using the raster projection method to measure the actual height of the tender shoots relative to the reference surface.

The system is designed to project four grating fringes with different wavelengths for multiple times, and the fringe number $m_k$ is 1, 4, 16, and 64 respectively. Among them, the main phase field value $\phi_k$ of each fringe is still calculated by the four-step phase shift method, but the phase solution process is as follows:
\[ \varphi_k^u = \begin{cases} \varphi_k^w, & k = 1 \\ \varphi_k^w + 2\pi \text{Round}, & k > 1 \end{cases} \]

In the time phase profile, the fringe patterns of different wavelengths are projected to the measured object, and a fringe pattern sequence of shape modulation is obtained, so that each phase point is independently expanded along the time axis. Compared with the traditional four-step phase shift spatial phase unwrapping method, the time phase unwrapping method does not look for an unfolding path in the two-dimensional phase field, and can effectively avoid the error diffusion caused by the phase solution error point.

6. Stereo Matching and Positioning

As far as binocular stereo vision is concerned, the core of the calculation of three-dimensional information is to know the specific position of the spatial point, that is, to calculate the parallax from the coordinates of the point in the two images [21, 22]. Matching the areas identified by the target detection algorithm in the left and right cameras, determining the corresponding relationship between the image points of the same target in the left and right imaging planes in the three-dimensional space, and completing the search for the correlation degree of the two sets of data, which is the purpose of stereo matching. With the help of stereo matching, the corresponding relationship between the left and right images in the binocular camera is determined, and the picking center point is found.

(1) feature point extraction and matching based on SURF algorithm

Feature-based matching algorithms mainly include three major features: point features, local features and overall features. The point feature is mainly composed of three parts: SURF, SIFT and Harris [23-25]. The SURF feature has good robustness, can perfectly avoid the interference of external noise and light, and has a faster calculation speed and a strong scale without distortion. In the SURF feature detection algorithm, a Hessian matrix is constructed for second-order differential detection to identify potential points of interest that are invariant to scale and selection. The main method used by this algorithm to calculate the direction of feature points is to count the Harr wavelet features in the circular neighborhood of feature points. The length of the Harr wavelet template is 4 s, that is, in a circular neighborhood with a characteristic point in the center and a radius of \( R \), the horizontal and vertical Harr wavelet characteristics of all points in the \( \pi/3 \) sector are counted, and the calculation results are accumulated. Then rotate the fan shape at intervals of 0.2 radians, and at the same time perform a second statistics on the Harr wavelet eigenvalues in this area, and finally determine that the main direction of this characteristic point is the direction of the fan with the largest value. The schematic diagram is shown in Figure 7.

![Figure 7](image)

Figure 7. Schematic diagram of calculating the main direction of feature points.

(2) Principle of Positioning

When the camera takes an image, any point on the image can find countless mapping points in space. The straight line formed by these mapping points is called the line of sight. If two cameras are used to shoot a target at the same time, there will always be a point on the target that can be used as the
intersection of the two cameras' sight lines. The intersection formed by the image plane and the sight lines of the two cameras is the image forming phase point \[26\].

In the binocular stereo vision system, assuming that the internal parameters of the two cameras are exactly the same, the world coordinate system coincides with the camera coordinate system of the left camera, and the image coordinates of the space point \(P(x, y, z)\) formed on the camera are \(P_i = u_i + v_i\). The distance between the optical centers of the two cameras \(O_iO_j = d\), focal length is \(f\). According to the triangle similarity relationship, the imaging coordinates can be derived.

7. Difficulties In The Identification And Positioning Of Tea Buds

Looking at the current research status of tea visual identification and positioning at home and abroad, although a lot of progress has been made, it still cannot meet the actual picking requirements. There are mainly the following technical difficulties \[27-30\]:

1. There are many varieties of tea, the picking methods are flexible and diverse, and the picking standards are different, which puts forward high requirements for machine vision to identify tea.

2. In the natural environment where tea is grown, the intensity of light is different and the weather is complex and changeable, which has a great impact on the accurate identification of the bud area.

3. The growth posture and growth position of tea are different, and the tea clumps are dense and the distribution density is high, resulting in the young buds being blocked by the stems and old leaves.

4. Due to the light weight of the buds, the movement of the robot causes the tea leaves to swing, which greatly interferes with the accuracy of the visual positioning of the picking point.

In view of the current technical difficulties, four aspects can be considered: (1) By adding sample data, such as taking tea images in different regions and different weathers to improve the universality of the algorithm; (2) Multiple segmentation methods can be integrated to improve the stability of the algorithm to promote its effective role in practical applications \[31\]; (3) There are relatively few researches on the intelligent production of tea based on deep learning, and a variety of deep convolutional neural network tea buds detection models can be established to reduce manual labor influence of feature extraction and external environment on the segmentation and positioning of buds. (4) Aiming at the problem of positioning interference caused by motion, it can be considered to improve the accuracy of positioning by studying dynamic recognition and positioning algorithms \[32\].

8. Conclusion

Carrying out research on machine vision technology and realizing the automatic control and refined management of the whole process of the tea industry are of great significance for ensuring the high yield, efficiency, quality, safety and health of the tea industry and achieving sustainable development. Aiming at the problems of selective and autonomous tea picking, this paper reviews the application status of machine vision technology in tea intelligent picking, expounds the related algorithms of tea sprout segmentation and positioning using machine vision, and analyzes the basic principles and technical points. Current researches on tea bud segmentation and positioning are mainly for specific environments or tea types, but there are many types of tea, and the growth environment is greatly affected by regions and weather, which leads to poor algorithm stability and general applicability. At present, the difficulty in the identification and positioning of tea buds lies in the diversity of tea, the influence of environmental background, and the dynamic interference. The solution can be thought from aspects such as intelligent algorithms, dynamic recognition and positioning, etc. With the rapid development of machine vision technology, it has great development space and broad application prospects in tea picking and processing.

Acknowledgments

This work was supported by the Nanchong city and school science and technology strategic cooperation project (Grant No.19SXHZ0041 and No.18SXHZ0049), Southwest petroleum university key project of extracurricular experiment in 2021 (Grant No. NKSZ21006).
References
[1] P. Bassi, V. Kumar, S. Kumar, et al, “Importance and prior considerations for development and utilization of tea bags: a critical review,” Journal of Food Process Engineering, vol. 43, pp. 1-10, 2020.
[2] J. Baeten, K. Donne, S. Bocdrij, et al, “Autonomous fruit picking machine: a robotic apple harvester,” STAR, vol. 42, pp. 531-539, 2008.
[3] N. Kondo, M. Monta, T. Fujiura, “Fruit harvesting robots in Japan,” Advances in Space Research, vol. 18, pp. 181-184, 1996.
[4] Y. Geng, “Research on the application of automation software control system in tea garden mechanical picking,” Advances in Intelligent Systems and Computing, vol. 1017, pp. 1830-1836, 2019.
[5] H. L. Yang, L. Chen, Z. B. Ma, et al, “Computer vision-based high-quality tea automatic plucking robot using Delta parallel manipulator,” Computers and Electronics in Agriculture, vol. 181, pp. 105946, 2021.
[6] Y. Xiong, Y. Y. Ge, L. Grimstad, “An autonomous strawberry-harvesting robot: design, development, iteration, and field evaluation,” Journal of Field Robotics, vol. 37, pp. 202-224, 2020.
[7] Y. T. Chen, S. F. Chen, “Localizing plucking points of tea leaves using deep convolutional neural networks,” Computers and Electronics in Agriculture, vol. 171, pp. 1-7, 2020.
[8] H. K. Xia, B. G. Shi, H. X. Huang, et al, “Application progress of image processing in intelligent picking of tea sprouts,” Anhui Agricultural Science Bulletin, vol. 25, pp. 133-134, 2019.
[9] Y. Y. Song, W. N. Li, B. Li, et al, “Design and test of crawler type intelligent tea picker,” Journal of Agricultural Mechanization Research, vol. 42, pp. 123-127, 2020.
[10] L. Y. Wang, X. Yang, “Trajectory tracking control for robotic manipulators endowed with feedforward and neural networks,” Electric Machines and Control, vol. 17, pp. 113-118, 2013.
[11] C. H. Qian, M. Y. Li, Y. Ren, “Tea sprouts segmentation via improved deep convolutional encoder-decoder network,” IEICE Transactions on Information and Systems, vol. 103, pp. 476-479, 2020.
[12] F. Yang, L. Yang, Y. Tian, et al, “Recognition of the tea sprout based on color and shape features,” Nongye Jixie Xuebao/transactions of the Chinese Society of Agricultural Machinery, vol. 40, pp. 119-123, 2009.
[13] K. H. B. Ghazali, J. Ma, R. Xiao, et al, “An innovative face detection based on YCgCr color space,” Physics Procedia, vol.25, pp. 2116-2124, 2012.
[14] H. L. Yang, L. Chen, M. T. Chen, et al, “Tender tea shoots recognition and positioning for picking robot using improved YOLO-V3 model,” IEEE Access, vol. 7, pp. 180998-181011, 2019.
[15] H. Q. Su, J. Wenc, “A new algorithm based on super-green features for ‘ostu’ method using image segmentation,” World Automation Congress (WAC), 2012.
[16] Ch. V, M. V, M. Sirisha, “Remote sensing image segmentation using OTSU Algorithm,” International Journal of Computer Applications, vol. 178, pp. 46-50, 2019.
[17] J. Yang, Y. Jian, “Cloud platform system for the diagnosis of typical tea plant diseases based on neural network,” Journal of Physics: Conference Series, vol. 1648, pp. 042086, 2020.
[18] K. F. Wang, “Quantitative detection of internal defects based on morphological opening, filling and binarizing operations on wrapped phase of out-of-plane deformation in digital speckle pattern interferometry,” NDT & E international, vol.103, pp. 173-176, 2019.
[19] Y. F. Gong, L. H. Yuan, L. Tang, “Self-adaptive trajectory tracking control of RBF neural network SCARA manipulator,” Machine Tool & Hydraulics, vol. 42, pp. 41-46, 2014.
[20] Y. He, X. L. Li, X. F. Deng, “Discrimination of varieties of tea using near infrared spectroscopy by principal component analysis and BP model,” Journal of Food Engineering, vol. 79, pp. 1238-1242.
[21] O. O. Arjenaki, P. A. Moghaddam, “Online tomato sorting based on shape, maturity, size, and surface defects using machine vision,” Turkish Journal of Agriculture and Forestry, vol. 37,
9

[22] J. Miao, T. Z. Huang, X. Zhou, et al, “Image segmentation based on an active contour model of partial image restoration with local cosine fitting energy,” Information Sciences, vol. 447, pp. 52-71, 2018.

[23] S. Borah, M. Bhuyan, “Quality indexing by machine vision during fermentation in black tea,” Proceedings of the SPIE International Society of Optical Engineering, pp. 468-475, 2003.

[24] Z. Tang, Y. C. Su, M. Joo, et al, “A local binary pattern based texture descriptors for classification of tea leaves,” Neurocomputing, vol. 168, pp. 1011-1023, 2015.

[25] R. Michael, S. Jack, “Chan-vee reformulation for selective image segmentation,” Journal of Mathematical Imaging and Vision, vol. 61, pp. 1173-1196, 2019.

[26] X. J. Jin, Y. Chen, Y. Q. Guo, et al, “Tea flushes identification based on machine vision for high-quality tea at harvest,” Applied Mechanics & Materials, vol. 288, pp. 214-218, 2013.

[27] X. Han, Y. Qi, H. Sun, et al, “Research on color image segmentation algorithm based on improved Grab-cut,” IEEE International Conference on Information & Automation. IEEE, 2018.

[28] J. Fan, K. Y. Yaud, A. K. Ekmagarmid, et al, “Automatic image segmentation by integrating color-edge extraction and seeded region growing,” IEEE Transactions on Image Processing, vol. 10, pp. 1454-1466, 2001.

[29] W. Rawat, Z. H. Wang, “Deep convolutional neural networks for image classification: a comprehensive review,” Neural Computation, vol. 29, pp. 2352-2449, 2017.

[30] Ramnarayan R, Saklani N, Verma V. A review on edge detection technique "canny edge detection"[J]. International Journal of Computer Applications, 2019, 178(10):28-30.

[31] Suprijanto, E. Juliastuti, A. Rakhmawati, “Computer vision for black tea grading based on the black tea particles and tea infusion color,” International Journal of Imaging & Robotics, vol. 10, pp. 17-29, 2013.

[32] F. N. Chen, F. Cheng, “Defective kernel detection using a linear colour CCD,” The Imaging Science Journal, vol. 61, pp. 361-368, 2013.