RESEARCH ARTICLE

Spatial positioning and recognition method of tobacco leaf blossom tops based on binocular visual communication

Weihua Qin[1], Chaofei Yang[1], Xiangguo Cheng[1], Shengli Chen[1,*], Lianhao Li[2], Chenhui Zhu[2]

[1]Sanmenxia Branch of Henan Tobacco Company, Sanmenxia, Henan, China. [2]Henan Agricultural University, Zhengzhou, Henan, China

Received: November 23, 2023; **accepted:** January 21, 2024.

During the growth of tobacco, the appearance of tobacco leaf flowers marks the key reproductive stage of the growth cycle. Management measures during this period will directly affect the energy distribution of the plant and determine the quality and final yield of the tobacco leaves. Therefore, research on the spatial positioning and identification technology of tobacco leaf flower tops not only helps to accurately grasp the growth progress of tobacco, but also provides technical support for the realization of precision agriculture. At present, the monitoring technology of tobacco leaf flowering mainly relies on traditional manual observation and recording. Although this method is intuitive, it has the problems of low efficiency, strong subjectivity, and inability to conduct large-scale real-time monitoring, which seriously affects the collection of tobacco leaf flowering information accurately. With the development of precision agriculture, some automation technologies have been introduced into the collection of tobacco leaf flowering information. This study was based on binocular visual communication technology to identify the spatial positioning of tobacco leaf flower tops. A binocular visual communication system was designed, and a spatial positioning recognition method for tobacco leaf flower tops and an optimization technology for Convolutional Neural Networks (CNN) oriented spatial recognition were proposed. The results of spatial positioning recognition for tobacco leaf flower tops were analyzed and showed that the binocular visual communication system could effectively collect crop status information. The image segmentation algorithm in the visual system could effectively identify tobacco flower tops. The system had strong information analysis ability and could accurately predict the quality and yield of tobacco flower tops based on images. When the algorithm constructed in this study processed images of different areas of tobacco leaf flower tops, the centroid deviation distance was between 0.30 and 1.60 m, which was better than other algorithms. Reduced centroid deviation distance appeared after Fuzzy C-Means (FCM) processing. Compared with the illustration grid, the accuracy of the Visual Geometry Group (VGG) grid was improved, and when the ratio of the style image impact factor and the content image impact factor was 5, the accuracy rate increased the fastest. The method proposed in this study improved not only the accuracy of image processing, but also the accuracy and efficiency of tobacco leaf flower top information collection, showing better performance than the existing technology. The application of this technology would improve the scientific and technological content of agricultural production and reduce labor costs.

Keywords: visual communication; spatial positioning; image segmentation; tobacco leaf tops; clustering algorithm; recognition.

***Corresponding author:** Shengli Chen, Henan Tobacco Company Sanmenxia City Company, Sanmenxia 472000, Henan, China. Email: shengli202311@163.com.

Introduction

Tobacco leaf is the main raw material for manufacturing tobacco products. The technology and quality control during tobacco leaf production and processing are directly related to the quality and yield of tobacco products. The positioning and identification of tobacco leaf flower tops play a crucial role in the growth, harvesting, and primary processing of tobacco leaves [1, 2]. Traditionally, the positioning and identification of tobacco leaf flower tops mainly relies on manual work, which is not only labor-intensive and inefficient, but also prone to misjudgments and omissions, resulting in unstable tobacco leaf quality and reduced production efficiency. The shape and color of tobacco leaf blossoms have their own specific characteristics. However, in the actual production environment, due to the influence of light, background, and other environmental factors, the image recognition of tobacco leaf blossoms still faces many challenges [3]. In addition, tobacco leaf flower tops will change in shape and color during the growth process, which also brings unpredictable difficulties to their spatial positioning and identification. With the rapid development of computer vision technology, image processing and pattern recognition technology are increasingly used in agriculture, industry, and daily life [4]. Binocular vision, as an important branch of computer vision, can obtain three-dimensional information of objects by simulating the human binocular vision mechanism, and can accurately position objects in space. Machine vision technology refers to empowering machines with visual judgment and visual analysis capabilities through image processing algorithms and image acquisition terminals, thereby prompting machines to complete complex operations [5]. Humans receive external information mainly through vision. Allowing robots to have human machine vision functions can greatly improve the machine's ability to receive external information. The emergence of machine vision technology has solved the high cost and low efficiency of manual operations and has high practical significance for the automation and modernization of agricultural production.

Many scholars have conducted certain research on improving the picking efficiency of different crops. Dong *et al*. proposed an extension research method to optimize the classification and capture of robotic arm targets in extended visualization programs. The study required that the designed sorting robot must be able to effectively detect all goods in different storage environments. Based on the regression method of Single Shot MultiBox Detector (SSD) target detection, the three-dimensional target was reconstructed through the default unit. The results showed that the robot's accuracy in sorting goods increased by 79%, which contributed to the accuracy and efficiency of imports [6]. Hsieh *et al*. proposed to use Region Convolutional Neural Network (RCNN) and binocular imaging technology to spatially identify the maturity and position of tomato fruits. An image capture and target detection model was constructed, and then the three-dimensional position of the planted fruit was compared with the actual position. The correlation between the fruit sizes was calculated. The results demonstrated that the correct identification rate of ripe fruits was as high as over 95% [7]. Sverdlichenko *et al.* proposed a strawberry picking robot that combined You Only Look Once (YOLO) and RCNN to achieve low-price picking of strawberries. Multiple images of occluded strawberries were collected, and different models were used to identify strawberry picking. A stereo camera was used to position the strawberry in three dimensions. The results showed that the method constructed in the experiment could accurately capture the location of strawberries [8]. To improve the efficiency of sugarcane planting, Wang *et al*. proposed a bud positioning method based on computer binocular vision using grayscale horizontal projection to determine the position of the sugarcane stem segment and used color space conversion to determine the plane position of the bud. The method proposed had a matching accuracy of 98% for seed sprouts, which provided a certain

basis for resisting damage to sprouts during automatic sugarcane planting [9].

With the continuous development of various computer technologies, binocular vision technology has gradually become a hotspot. For improving the accuracy of industrial robots in grabbing items, Wan *et al*. designed a large object grabbing means for industrial robots with binocular vision guidance. The three round holes of the object were used as fine features to determine the position and posture of the object. The designed system had high robustness and the positioning accuracy reached -1 mm [10]. To improve the accuracy of industrial robot positioning, a new robot technology based on binocular vision technology was developed to analyze the principle of position positioning, capture the characteristics of the target using URF, and introduce the back propagation neural network to locate the target position. The method was efficient and effective in feature extraction speed [11]. To monitor the residential safety of the elderly, Zhu *et al*. proposed a method based on binocular vision positioning. Image data was used to construct an imaging model, and feature extraction was completed to estimate the camera movement and status of the elderly. The proposed method could achieve precise positioning of indoor activities of the elderly and effectively improve the ability of real-time monitoring [12]. To improve the accurate capture of substation features by live working robots during their work, Jian *et al*. proposed an outdoor three-dimensional reconstruction method that combined multi-dimensional laser and binocular vision. Multi-line lasers were used to capture artificial features, and a natural light filtering method based on frame difference was used to reduce the interference of natural light on the laser. The results showed that the proposed method could effectively achieve three-dimensional reconstruction of the target [13].

Binocular vision technology has been successfully applied in many fields, such as robot navigation, three-dimensional reconstruction, and virtual reality. However, in the agricultural field, especially in tobacco leaf production, the application of binocular vision technology is still in its infancy, which provides new inspiration for the automatic positioning and identification of tobacco leaf flower tops. Some scholars believe that the two-dimensional maximum entropy image segmentation algorithm improved by the Genetic Algorithm (GA) processing the initial image. Compared with the one-dimensional threshold segmentation algorithm, it considers not only the one-dimensional characteristics of the image grayscale histogram, but also the spatial neighborhood information of the pixel space, and therefore, reduces the computational complexity and improves the computing efficiency. Other scholars pointed out that CNN could have relatively ideal results in different data sets. It not only has good generalization performance, but also can greatly reduce labor costs and improve detection accuracy through automatic learning. With the advent of the artificial intelligence era, computer vision regarding the spatial position and identification of tobacco leaf flower tops has gradually become a breakthrough point in the study of traditional agriculture. This study proposed a method for spatial positioning and identification of tobacco leaf blossom tops based on binocular visual communication and used CNN to optimize the recognition of tobacco leaf blossom tops to provide technical support for the automated production and intelligent management of tobacco leaves.

Materials and Methods

Binocular visual communication system design

The tobacco leaf blossom tops information collection robot consists of two main components: software design and hardware design. The software aspect primarily encompasses the design of the human-computer interaction interface, image processing algorithms, and installation of hardware drivers. The hardware component involves the utilization of a binocular camera, along with the design of

the infrared emission module, information processing system, and image acquisition [14]. The overall framework design was illustrated in Figure 1.

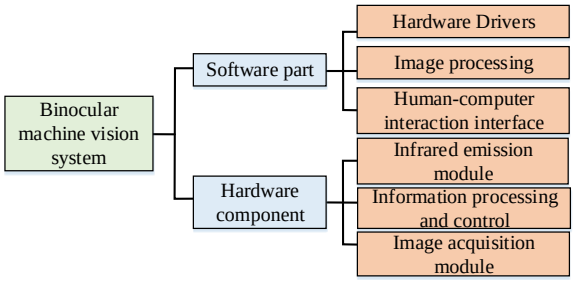


Figure 1. Overall framework design of binocular machine vision system.

The workflow of the tobacco leaf blossom tops information collection robot mainly includes two steps. The first step involves capturing the image signal of the target object, which is achieved through the utilization of technologies such as infrared emission and image acquisition. Subsequently, the collected image signal is converted into a digital image and transmitted to the information processing system *via* a designated transmission path. Within the processing system, the image signal undergoes further processing. Finally, the system completes tasks such as object identification, data collection, and output prediction. The collected data is then presented to users through a user-friendly human-computer interaction interface [15]. Tobacco leaf blossom tops information collection robots were required for collecting clear and complete images. The most common collecting chips currently were Charge Coupled Device (CCD) chips and Complementary Metal Oxide Semiconductor (CMOS) chips. Compared with CMOS, CCD has lower application cost and poor image quality. CMOS is significantly better than CCD in image processing and noise reduction and anti-interference. In particular, the latest CMOS chip has a lower price. Therefore, CMOS clearly exceeds CCD in terms of overall performance. Due to its superior image signal-to-noise ratio and resolution, CMOS has emerged as the preferred sensor chip for image acquisition, gradually replacing other comparable products. This study utilized CMOS chips as the image acquisition tool. The corresponding hardware framework of the system was illustrated in Figure 2.

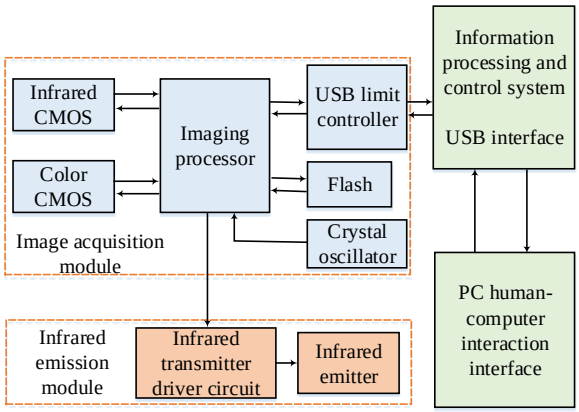


Figure 2. Machine vision system hardware framework.

The tobacco leaf blossom tops information collection robot vision system covered multiple modules. Among the signal acquisition tools used in the image processing module, color CMOS and infrared CMOS were employed. Color CMOS was primarily responsible for capturing color images, while infrared CMOS was utilized for collecting depth image signals. Once the image signals were acquired, the imaging processor converted them into digital images that could be recognized by the computer. The information control system was responsible for processing these digital images, employing both wireless and wired methods. The processed image results were transmitted to the computer terminal *via* the information control system and were then presented through a user-friendly human-computer interaction interface. The specific devices used in the image acquisition and processing process include infrared CMOS cameras, color CMOS cameras, infrared transmitters, and tobacco leaf blossom tops information collection robots [16] with 1,280 (H) × 960 (V) resolution, 57° (H) × 43° (V) shooting angle, and a 15 fps frame rate. The image

resolution of the infrared CMOS camera was 320 (H) × 240 (V), the depth detection domain was 0.8-4 m, the frame rate was 30 fps. The tobacco leaf blossom tops information collection robot served as the carrying platform for this design, and the image processing of the entire process was completed on the computer. For efficient integration into the information collection robot, image processing necessitated a compact computer with high processing speed. The primary role of the vision system was to identify the target object and gather pertinent information. The software component operated within the Linux system, utilizing the programming function OpenCV and OpenGL (Figure 3).

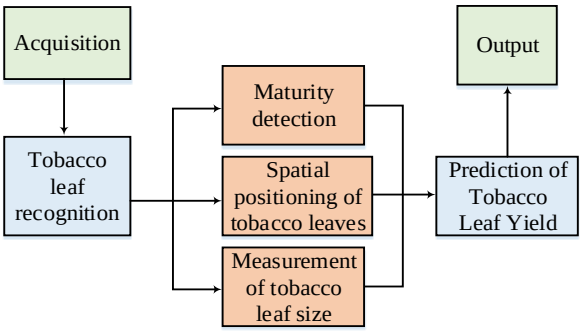


Figure 3. Machine vision system software framework.

The human-computer interaction interface and computing involved a series of 7 steps including system and interface initialization, data communication method determination, image display, CD image conversion, object status detection, data printing, and program closure. Upon launching the vision system and interactive interface, the user selected the desired data connection method (wired or wireless) and configured the corresponding IP address. Subsequently, the computer terminal displayed the relevant image information with color images being converted into depth images [17]. Once the image conversion was complete, the visual system collected the necessary data. The collected data was then transmitted back to the computer for user-friendly viewing, and the data results were printed to establish the data analysis outcomes. The program was closed to conclude the operations.

Spatial positioning and identification method of tobacco leaf blossom tops

The one-dimensional OTSU employs enumeration to identify the gray level of the image being processed, aiming to find the image segmentation threshold when the objective function value is maximized. However, the one-dimensional version is susceptible to noise points and fails to effectively filter out noise interference [18]. The two-dimensional OTSU was hired to perform threshold segmentation on the original image, enabling the determination of the desired range. For the exhibition of mathematical expression, it was assumed that an image *I* existed with size of *M* × *N*. The gray level and the average gray value of its neighborhood were denoted as *I*(*x*, *y*) and *g*(*x*, *y*) as depicted in formula (1).

$$g(x,y)=\frac{1}{9}\sum_{m=-1}^{1}\sum_{n=-1}^{1}l(x+m,y+n) \quad (1)$$

Formula (1) was for the average gray value.

$$p_{lg}=\frac{f_{\lg}}{M\times N} \quad (2)$$

In formula (2), p_{lg} was the joint probability density. l represented simultaneous occurrence frequency in the target $f_{\lg}$. g was the average gray value of the neighborhood. The two-dimensional grayscale histogram could be divided into four distinct regions that were determined by the grayscale values of the neighborhood center point (*s*) and the neighborhood average grayscale (*t*). Regions 1 and 2, positioned diagonally, represented the target area and the background area, respectively. Regions 3 and 4, located far from the histogram, corresponded to noise areas. The probability distribution (p_t) of the target area was calculated using formula (3), while the

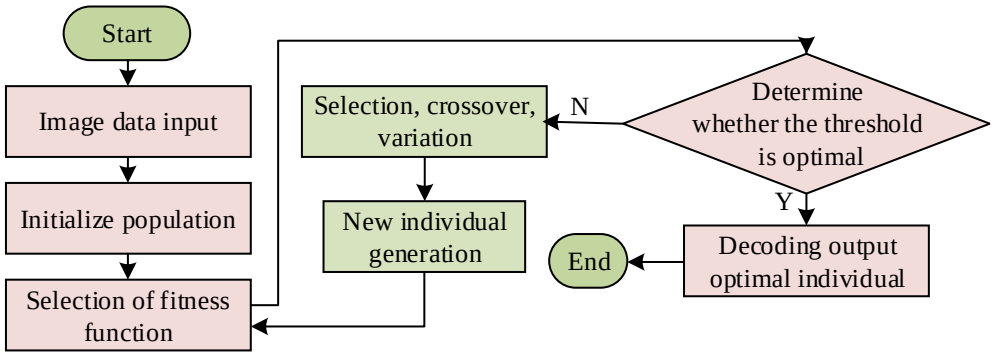


Figure 4. Two-dimensional threshold image segmentation with GA.

probability distribution (p_b) of the background area was calculated using formula (4).

$$p_t = \sum_{l=0}^{s} \sum_{g=0}^{t} p_{\lg} \quad (3)$$

$$p_b = \sum_{l=s+1}^{L-1} \sum_{g=t+1}^{L-1} p_{\lg} \quad (4)$$

The mean vector u_A was possible to be derived through formulas (3) and (4) as shown in formula (5).

$$u_A = (u_{Ai}, u_{Aj})^T = [\sum_{l=0}^{L-1} \sum_{g=0}^{L-1} lp_{\lg}, \sum_{l=0}^{L-1} \sum_{g=0}^{L-1} gp_{\lg}]^T \quad (5)$$

Given that noise points were featured with a low distribution probability, this study disregarded such noise points. Consequently, the objective function calculation formula for the OTSU was assumed as formula (6).

$$J(s,t) = p_t[(u_{ti} - u_{Ai})^2 + (u_{tj} - u_{Aj})^2] + p_b[(u_{bi} - u_{Ai})^2 + (u_{bi} - u_{Ai})^2] \quad (6)$$

where the optimal threshold (s,t) was determined, where the objective function value $J(s,t)$ was maximized. To enhance the two-dimensional threshold image segmentation algorithm proposed earlier, this study incorporated a genetic algorithm. The genetic algorithm employs eight-bit binary coding, representing the threshold pair (s) as a vector $a = [t_1, t_2, t_3, ..., t_8, ..., t_{16}]$. The first eight bits correspond to the binary code, while the last 8 of them represent the second threshold's binary encoding. By utilizing crossover and mutation operators, it optimizes the calculation of the optimal threshold. This approach introduces new data into the calculation database, increasing diversity and enabling the identification of global-scale issues, which is a feasible solution to avoid falling into local optimum [19]. The algorithm flow of the two-dimensional threshold segmentation improved by the genetic algorithm designed in this study was displayed in Figure 4. The threshold was set in an eight-bit binary encoding form, and the number of individuals in the population was selected to be 30 to 50. The corresponding function of the optimal threshold was hired to be the fitness function, that was formula (6) to calculate the individual fitness value under this condition. Under the settled iteration conditions, assumption that if it was satisfied, the optimal solution was considered to be the individual. Otherwise, the genetic algorithm proceeded with selection, crossover, and mutation operations to generate a new offspring population, whose fitness was then recalculated using an appropriate fitness function until the optimal individual was selected for decoding.

Spatial recognition optimization technology for CNN

Assuming that the input image size was 224 × 224, the response matrix would be obtained after each layer of neural network. After completing the tobacco leaf blossom tops image style

extraction, the RGB (Red, Green, Blue) space image was converted into an HSV (Hue, Saturation, Value) color space image. The RGB difference method determines the detecting area. The digital image is composed of RGB gray scales. The calculation that was built for letting RGB images change into HSV color space was formula (7).

$$H=\begin{cases} underfined, if, m=n \\ 60\times\dfrac{G-B}{m-n}, if\,(m=R) and (G\geq B); \\ 60\times\left(\dfrac{G-B}{m-n}+6\right), if\,(m=R) and (G<B); \\ 60\times\left(\dfrac{B-R}{m-n}+2\right), ifm=G \\ 60\times\left(\dfrac{R-G}{m-n}+4\right), ifm=B \end{cases} \quad (7)$$

where m = max(R, G, B), n = min(R, G, B), $V = m$, $S=1-n/m$. The gray image V_1 was processed by histogram equalization to obtain brightness V, and then an image with uniform grayscale was obtained through mapping 1 (V_2). The cumulative distribution function of each gray level probability was displayed in formula (8).

$$S_K=T(\mathrm{rk})=\sum_{\mathrm{j=0}}^{\mathrm{k}}\frac{\mathrm{n_j}}{\mathrm{n}}=\sum_{\mathrm{j=0}}^{\mathrm{k}}\mathrm{Pr(rj)} \quad (8)$$

where V_2 of V_1 total possible gray levels was represented by L. The total number of pixels n in T, the r_k image n_k, was r_k. s_k was the probability of the gray value of the original image. The calculation formula for converting HSV color space into RGB space was shown formula (9).

$$(\mathrm{R},\mathrm{G},\mathrm{B})=\begin{cases} (v,z,x), c_1=0; \\ (y,v,x), c_1=1; \\ (x,v,z), c_1=2; \\ (x,y,z), c_1=3; \\ (z,x,v), c_1=4; \\ (v,x,z), c_1=5; \end{cases} \quad (9)$$

where $C_1=\left[\dfrac{H}{60}\right]\bmod 6$, $C_2=\left[\dfrac{H}{60}\right]-C_1$. After nonlinear filter combination, when the l layer model was defined that it existed with N different filters, then this l layer also had feature map of identical N outputs. The feature output value of F_{ik}^{l}, the k th position in the l th layer was k. The vectorized white noise image and photo image were set to be represented by p and x, respectively. The corresponding feature response matrices of the white noise image and photo image of the first layer were F and P, respectively. Then the average error loss function expression of the response matrix could be expressed as:

$$L_{content}(p,x,l)=\frac{1}{2}\sum_{ik}(F_{ik}^{l}-P_{ik}^{l})^2 \quad (10)$$

The layer vectorized feature map and the inner product between each other were $G_{\mathrm{ij}}^{\mathrm{l}}$ represented by references, and the overall style of the image could be represented by the combined Ram matrix. Since the image format of the input model was an RGB image, the 3 × 3 filter could easily correspond to the 3 × 3 area with a gray level of 0. The filter would not cause changes in the feature response values. The feature response values would only be affected by the neural network interference in primitive values. When the base value was 0, the relationship between the two filters had no value [20]. In order to prevent this situation from happening, formula (11) was used to improve the mutual relationship.

$$G_{\mathrm{ij}}^{\mathrm{l}}=\sum_{k}F_{\mathrm{ik}}^{\mathrm{l'}}F_{\mathrm{jk}}^{\mathrm{l'}}=\sum_{k}(F_{\mathrm{jk}}^{\mathrm{l}}+\frac{1}{F_{\mathrm{jk}}^{\mathrm{l}}+1})(F_{\mathrm{ik}}^{\mathrm{l}}+\frac{1}{F_{\mathrm{ik}}^{\mathrm{l}}+1}) \quad (11)$$

When $F_{\mathrm{jk}}^{\mathrm{l}}$ value was 0, the value $F_{\mathrm{ik}}^{\mathrm{l'}}$ was not affected by the filter. If $F_{\mathrm{jk}}^{\mathrm{l}}$ value was large, then the value $F_{\mathrm{ik}}^{\mathrm{l'}}$ was less affected by the filter. The study used the gradient descent method to obtain the characteristics of tobacco leaf blossom tops images, and used white noise images and Gram matrix to minimize the root mean square

to complete the matching [21]. Assume that the vectorized white noise image and the original tobacco leaf blossom tops image were represented by x and a, respectively, the l corresponding Gram matrices of the white noise image and the original tobacco leaf blossom tops image of the layer were G^l and A^l, respectively. The expressions of the corresponding loss function normalization processing and style loss could be expressed in formula (12).

$$E_l = \frac{1}{4N_L^2 M_l^2} \sum_{i,j} (G_{ij}^l - A_{ij}^l)^2 \quad (12)$$

where w_l was the weight value of each layer to the total loss. The gradient could be obtained. To study tobacco leaf blossom tops image style rendering of image content, it was necessary to further set the ratio of content loss and style loss, and then calculate the total loss (formula (13)). The overall loss was calculated by the mean square error of adjacent pixels to ensure the smoothness and visual effect of the image [22].

$$L_{total} = \alpha L_{content} + \beta L_{style} + \gamma L_{noise} \quad (13)$$

where L_{noise}, L_{style}, $L_{content}$, L_{total} were the noise image pixel variable loss value, style image loss value, content image loss value, and overall loss value, respectively. γ, β, α were the first three influencing factors, respectively. If β/α ratio was larger, the noise image would be rendered to a higher degree. On the contrary, the generated image would be closer to the real image. The convolutional layer used to obtain the photo content was CONV4_2. CONV1_1, CONV2_1, CONV3_1, CONV4_1, and CONV5_1 were all convolutional layers used to obtain tobacco leaf blossom tops image style.

Tobacco leaf identification test

To verify the accuracy of the design algorithm in obtaining tobacco leaf blossom tops information, the visual system was employed to measure the quality and diameter of tobacco leaves. The tobacco leaf planting location selected for the experiment was Kunming City, Yunnan Province, China, and the tobacco variety selected was Yunyan 87 that is widely grown in China and is known for its good adaptability and high-quality tobacco leaves. To assess the visual system's performance in the tobacco leaf information collection robot, a control group was established to evaluate the system's effectiveness. The collected tobacco leaf information was presented externally through the human-computer interaction interface, and the output data could be visualized after undergoing relevant processing. A set of 100 tobacco leaf blossom tops images was selected and divided into 10 random groups as sample data. The original images were segmented and subjected to feature pixel processing. The pixel annotation labeling method was used to determine the weight coefficient of the feature data. The segmentation algorithm was used to classify the tobacco leaf blossom tops of the control and image recognition groups.

Results and discussion

Tobacco leaf identification test results

The measurement outcomes of tobacco leaf blossom tops information were presented in Table 1. The results showed that the visual system exhibited high measurement accuracy for tobacco leaf blossom tops, aligning closely with the actual measurement values. The processed image yielded a high true positive rate, indicating that the segmentation algorithm effectively identified tobacco leaf blossom tops (Table 2). This capability met practical operational requirements and established a foundation for subsequent tobacco leaf contour extraction.

Image segmentation results

Three image segmentation algorithms were encompassed in this study including the one and two-dimensional threshold image segmentation algorithm, and an improved version of the later one using GA. The GA parameters included a population size of 40 and 100 iterations. The selection operator was the roulette selection operator, while the crossover operator was two-

Table 1. The measurement parameters of tobacco leaf blossom tops.

Tobacco leaf blossom tops image (set)	Actual diameter (mm)	Measurement diameter (mm)	Relative error (%)	Actual mass (%)	Measurement quality (%)	Relative error (%)
1	17.61	17.43	1.034	2.38	2.42	1.661
2	17.84	17.45	2.237	2.42	2.44	0.824
3	18.95	19.08	0.786	2.84	2.81	1.072
4	14.96	15.34	2.479	1.52	1.54	1.308
5	13.81	14.02	1.499	1.34	1.35	0.752
6	15.37	15.42	0.325	1.62	1.64	1.228
7	13.08	12.87	1.633	1.34	1.28	4.725
8	13.88	12.92	7.437	1.42	1.41	0.71
9	18.21	18.72	2.726	2.62	2.52	3.985
10	17.61	18.03	2.331	2.22	2.32	4.33

Table 2. Tobacco leaf blossom tops identification test results.

Tobacco leaf blossom tops image (set)	True positive	False positive	Consistency
1	97.836	1.999	95.918
2	95.757	1.263	94.564
3	98.939	1.394	96.128
4	96.914	2.028	95.983
5	98.715	1.533	94.128
6	97.195	2.920	97.029
7	97.915	1.416	98.918
8	95.245	1.823	95.190
9	94.923	2.568	94.615
10	96.893	1.818	98.033
Total	96.933	1.876	96.050

point and the mutation operator was the uniform mutation operator. Both the crossover probability and mutation probability were 0.2. A rendering of grayscale image segmentation using different image segmentation algorithms was shown in Figure 5. After being processed by the image segmentation algorithm, the range of tobacco leaf blossom tops recognition in the image was more obvious. However, different algorithms yielded varying recognition area sizes and ranges. The six algorithms' running times for segmenting area 1 were 159 ms, 162 ms, 31,574 ms, 34,327 ms, 180 ms, and 202 ms, respectively. It was evident that the one-dimensional OTSU exhibited image processing that was the highest speed among the objectives. The 3 others had slightly slower processing speeds, but the difference was not significant compared to the fastest objective. The two-dimensional OTSU algorithm and the two-dimensional maximum entropy algorithm (2DME) demonstrated the slowest processing speeds.

Comparison of centroid deviation distances of different segmentation algorithms for tobacco leaf image processing

The comparison of the centroid deviation distance for the segmented tobacco leaf image was illustrated in Figure 6. The centroid deviation refers to the Euclidean distance between the centroid position. When processing area 1 images, the 2DME exhibited the smallest

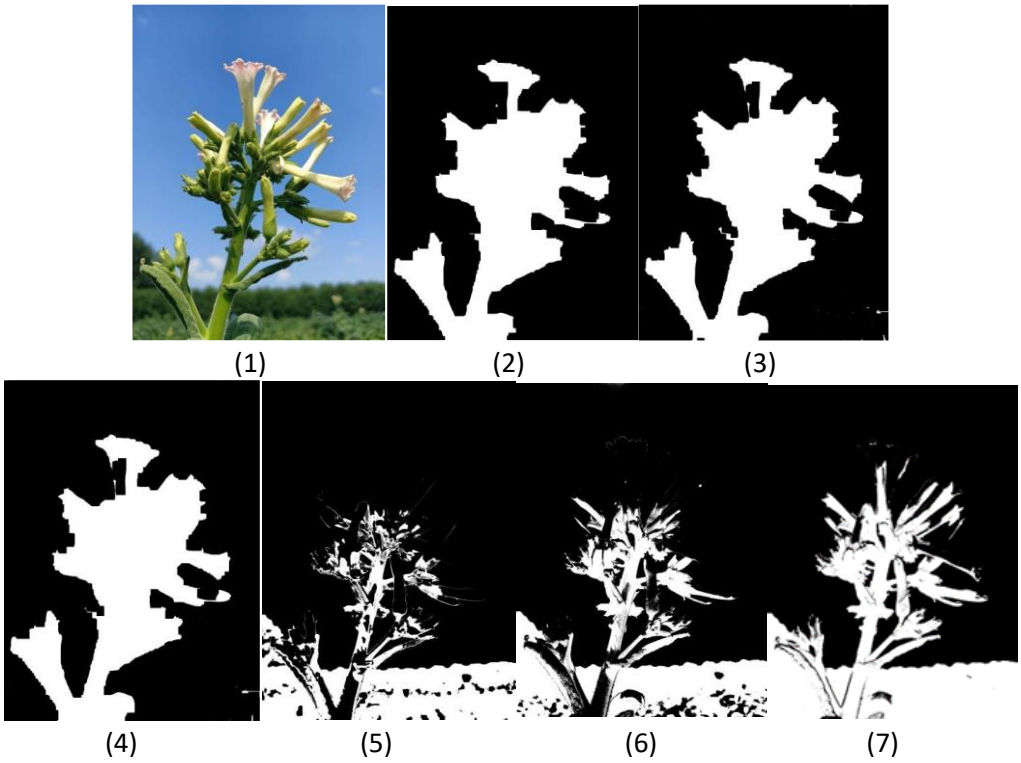


Figure 5. Segmentation outcomes. (1) Initial image, (2) One-OTSU, (3) One-mex, (4) Two-OTSU, (5) Two-mex, (6) GA + Two- OISU, (7) GA + Two-mex.

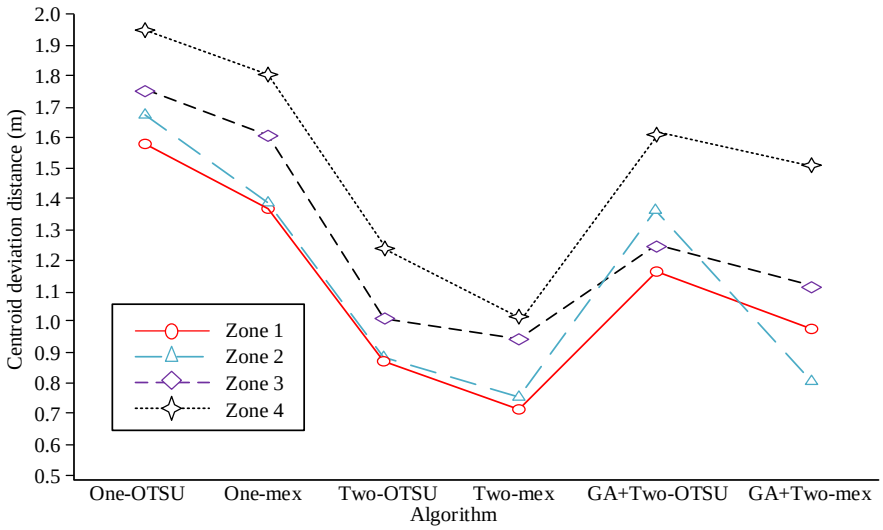
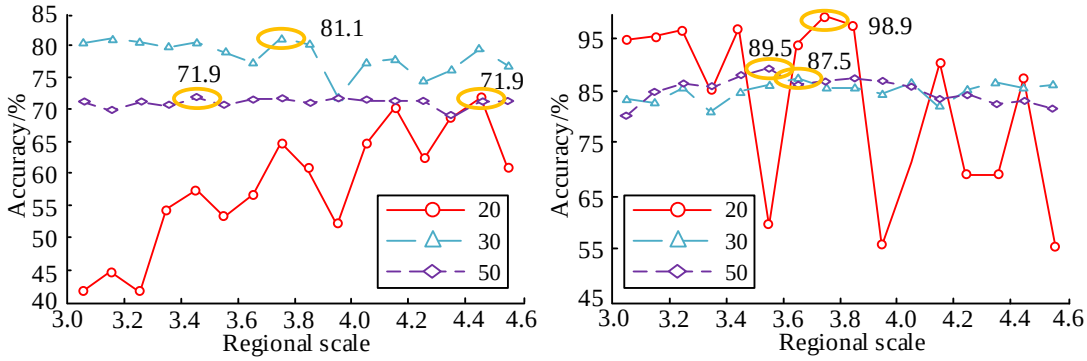


Figure 6. Centroid deviation distance of different segmentations for tobacco leaf blossom tops image.

deviation distance of 0.72 m. For area 2, the GA + 2DME achieved the smallest value of 0.81 m. The 2DME demonstrated the smallest value of 0.95 m when processing area 3 images. Lastly, when processing area 4 images, the 2DME yielded the smallest deviation distance of 1.01 m. 2DME demonstrated the best effect in processing images, followed by the GA + two-dimensional maximum entropy algorithm. However, the 2DME exhibited slow image processing speed.

Table 3. Comparison of centroid deviation distance of tobacco leaf blossom tops image processing.

Area	1		2		3		4	
Algorithm	**K-means**	**FCM**	**K-means**	**FCM**	**K-means**	**FCM**	**K-means**	**FCM**
Cluster center	[97.52, 23.51]	[98.27, 20.36]	[98.78, 45.33]	[99.96, 41.28]	[100.58, 73.75]	[99.01, 69.53]	[99.68, 90.42]	[100.05, 85.53]
Center of mass position	[32.35, 4.41]	[33.07, 3.45]	[34.18, 4.62]	[33.28, 3.24]	[34.39, 4.92]	[31.65, 3.05]	-	-
Center of mass deviation distance (m)	0.77	0.56	1.33	0.8	1.67	1.65	-	-
Area (m^2)	98.4	94.2	111.6	111.2	130.8	129.5	252.6	241.3



(a) VGG network structure accuracy (b) Illustration network structure accuracy

Figure 7. Comparison of accuracy results between VGG network and Illustration network.

Therefore, this study introduced GA + two-dimensional maximum entropy algorithm, which demonstrated superior overall performance. By employing the K-means and FCM to process different areas, the results included the cluster center position and centroid deviation distance calculation. As the tobacco leaves mature, the accuracy also became smaller. The K-means calculated the centroid deviation distances for images in areas 1, 2, 3, and 4, resulting in distances of 0.77 m, 1.33 m, 1.67 m, and no data, respectively (Table 3). The FCM clustering algorithm calculated the centroid deviations for images in areas 1, 2, 3, and 4, yielding distances of 0.56 m, 0.80 m, 1.65 m, and no data, respectively. Compared to the K-means, the FCM demonstrated higher accuracy. Although both K-means and FCM did not obtain specific coordinates for the cluster center point when processing the relatively large area 4 image, it displayed no impact on tobacco leaf blossom tops identification area calculation. The FCM achieved clustering accuracy of a higher value, resulting in identification areas of 94.2 m^2, 111.2 m^2, 129.5 m^2, and 241.3 m^2 for the four regions, respectively.

Comparison of accuracy and loss results of different models

Comparing the accuracy at different regional scales, the accuracy results of the VGG grid and the illustration grid were shown in Figures 7(a) and 7(b), respectively. The style picture influence factor and content picture influence factor were set to ratios of 5, 10, and 20. The findings indicated that, as the regional scale and the influence factors increased, the model's accuracy initially showed a significant improvement, followed by a notable decline. For the illustration grid ratios of 5, 10, and 20, the 4.4, 3.7, and 3.4

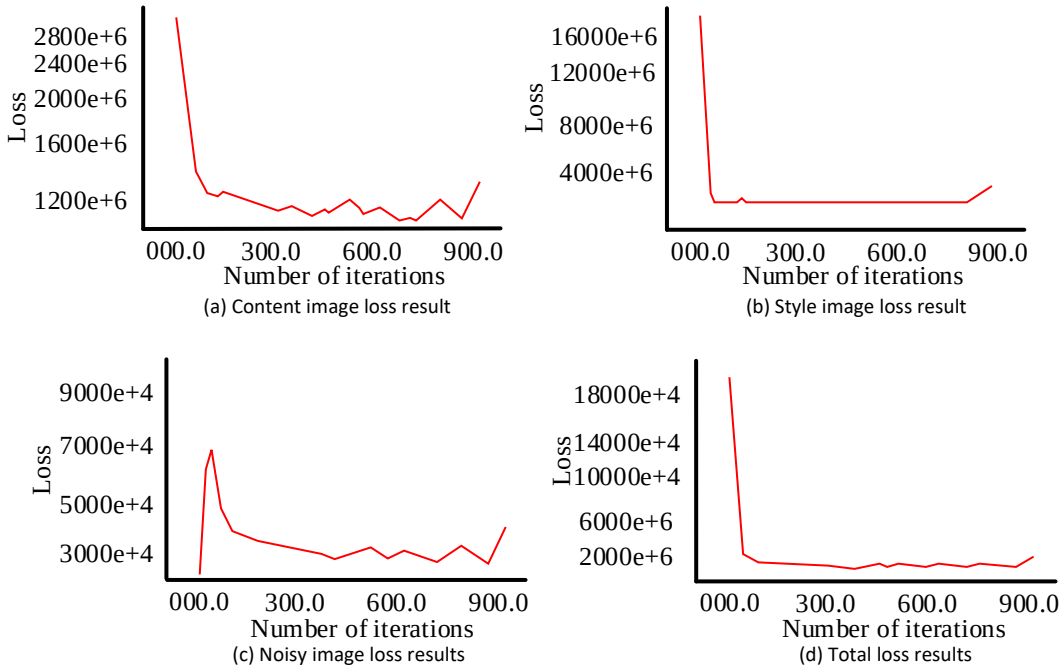


Figure 8. Content images, style images, noise images, and overall loss results.

were the useful values of optimal scales of corresponding regions with 71.9%, 81.1%, and 71.9% accuracy rates, respectively. For the VGG grid ratios of 5, 10, and 20, 3.7, 3.6, and 3.5 were the useful values of optimal scales of corresponding regions with 98.9%, 87.1%, and 89.5% accuracy rate, respectively. When the ratio of style image impact factor and content image impact factor was 10, the accuracy rates of the two structures were similar with a maximum accuracy difference of only 6.0%. This study used the model loss value to judge the performance of tobacco leaf image recognition and Tensorboard to present the changing trend of the loss. The content image, style image, noise image, and the entire loss were displayed in Figure 8. The model loss value evaluated the quality of the generated image from a quantitative perspective, but it could also reflect the subjectivity of the image. Both the style image loss and the overall loss gradually decreased as the number of training times increased, and the loss value quickly reached the convergence value, which was 0 and 2,000e + 6, respectively. At this time, the corresponding number of iterations was approximately 800 and 900, respectively. The noise image loss curve showed a rapid increase at first and then a slow convergence, and the convergence value was repetitive. The peak loss was 6.6e + 4. The loss value was smallest at around 400, 600, 700, and 900 iterations with a value of about 2,800e + 4. The convergence speed of content image loss was slow, and it could reach an optimal convergence value. There was a certain degree of repeatability in the loss value.

Conclusion

This research mainly used binocular visual communication image segmentation technology to study the spatial positioning and recognition of tobacco leaf blossom tops. The results showed that the visual system effectively identified tobacco leaves, as evidenced by the high true positive rate in the processed tobacco leaf images. The visual system also demonstrated high measurement accuracy for tobacco leaf blossom tops, aligning closely with the actual measured values. Additionally, the image segmentation algorithm proposed in this study

outperformed conventional algorithms like the one-dimensional OTSU algorithm, exhibiting smaller centroid deviation distances and faster running speeds. When comparing the FCM and the k-means cluster analysis outcomes, the FCM achieved higher accuracy. When the ratios of the style image impact factor and the content image impact factor were 5, 10, and 20, the optimal regional scales corresponding to the Illustration grid were 4.4, 3.7, and 3.4, respectively, and the accuracy rates were 71.9%, 81.1%, and 71.9%, respectively. The grid was at ratios of 5, 10, and 20, respectively. The corresponding optimal regional scales of the VGG grid were 3.7, 3.6, and 3.5, and the obtained accuracy rates were 98.9%, 87.1%, and 89.5%, respectively. The accuracy of the VGG grid had improved under the three types of parameter ratios, and when the ratio of the style image influence factor and the content image influence factor was 5, the accuracy rate increased the fastest with an accuracy rate of nearly 20%. This research displayed significance. But problems do exist. Less sample data and insufficient experiment accuracy were major defects. In future work, the sample data of the experiment will be increased.

References

1. Han J. 2020. Study on dynamic target positioning and grabbing based on binocular vision. Acad J Comput Inf Sci. 3(1):78-82.
2. Ji Y, Kumar R, Singh D, Singh M. 2021. Performance analysis of target information recognition system for agricultural robots. Int J Agric Environ Inf Syst. 12(2):49-60.
3. Wei SM, Li Y, Deng K, Lai HC, Tonetti MS, Shi JY. 2022. Does machine-vision-assisted dynamic navigation improve the accuracy of digitally planned prosthetically guided immediate implant placement? A randomized controlled trial. Clin Oral Implants Res. 33(8):804-815.
4. Ortiz-Peregrina S, Ortiz C, Casares-López M, Jiménez JR, Anera RG. 2021. Effects of cannabis on visual function and self-perceived visual quality. Sci Rep. 11(1):1-11.
5. Hu H, Kaizu Y, Zhang H, Xu Y, Imou K, Li M, *et al*. 2022. Recognition and localization of strawberries from 3D binocular cameras for a strawberry picking robot using coupled YOLO/Mask R-CNN. Int J Agric Biol Eng. 15(6):175-179.
6. Dong N, Meng F, Raffik R, Shabaz M, Neware R, Krishnan S, *et al*. 2022. Optimization of target acquisition and sorting for object-finding multi-manipulator based on open MV vision. Nonlinear Eng. 11(1):471-477.
7. Hsieh KW, Huang BY, Hsiao KZ, Tuan YH, Shih FP, Hsieh LC, *et al*. 2021. Fruit maturity and location identification of beef tomato using R-CNN and binocular imaging technology. J Food Meas Charact. 15(6):5170-5180.
8. Sverdlichenko I, Mandelcorn MS, Issashar Leibovitzh G, Mandelcorn ED, Markowitz SN, Tarita-Nistor L. 2022. Binocular visual function and fixational control in patients with macular disease: A review. Ophthalmic Physiol Opt. 42(2):258-271.
9. Wang F, Liu Q, Huang M, Qiao X, Huang Y. 2021. Research on sugarcane seed-bud location based on anisotropic scaling transformation. Appl Eng Agric. 37(6):1119-1130.
10. Wan G, Li F, Zhu W, Wang G. 2020. High-precision six-degree-of-freedom pose measurement and grasping system for large-size object based on binocular vision. Sens Rev. 40(1):71-80.
11. Li L. 2020. Research on target feature extraction and location positioning with machine learning algorithm. J Intell Syst. 30(1):429-437.
12. Zhu L, Zhang Y, Wang Y, Cheng C. 2022. Binocular vision positioning method for safety monitoring of solitary elderly. Comput Mater Contin. 71(1):593-609.
13. Jian X, Chen X, He W, Gong X. 2020. Outdoor 3D reconstruction method based on multi-line laser and binocular vision. IFAC-PapersOnLine. 53(2):9554-9559.
14. Wang L. 2021. Symbol recognition system based on 3D stereo vision. J Intell Fuzzy Syst. 40(4):5985-5994.
15. Gao H, Shen F, Zhang F, Zhang Z. 2022. A high precision and fast alignment method based on binocular vision. Int J Precis Eng Manuf. 23(9):969-984.
16. Xie Q, Hu X, Ren L, Qi L, Sun Z. 2022. A binocular vision application in IoT: realtime trustworthy road condition detection system in passable area. IEEE Trans Ind Inform. 19(1):973-983.
17. Sun Y, Liang L, Sun J, Chen X, Tian R, Chen Y, *et al*. 2022. A binocular vision SSVEP brain–computer interface paradigm for dual-frequency modulation. IEEE Trans Biomed Eng. 70(4):1172-1181.
18. Wang D, Sun H, Lu W, Zhao W, Liu Y, Chai P, *et al*. 2022. A novel binocular vision system for accurate 3-D reconstruction in large-scale scene based on improved calibration and stereo matching methods. Multimed Tools Appl. 81(18):26265-26281.
19. Niu J, Hu Q, Niu Y, Zhang T, Jha SK. 2021. Real-time recognition and location of indoor objects. Comput Mater Contin. 68(2):2221-2229.
20. Yin Z, Ren X, Du Y, Yuan F, He X, Yang F. 2022. Binocular camera calibration based on timing correction. Appl Opt. 61(6):1475-1481.
21. Wang X, Chen T, Wang Y, Zheng D, Chen X, Zhao Z. 2023. The 3D narrow butt weld seam detection system based on the binocular consistency correction. J Intell Manuf. 34(5):2321-2332.
22. Fang Q, Li H, Luo X, Li C, An W. 2020. A semantic and prior-knowledge-aided monocular localization method for construction-related entities. Comput-Aided Civ Infrastruct Eng. 35(9):979-996.