Target extraction of fruit picking robot vision system

Junyan Zhang
Jiangsu Union Technical Institute Suzhou Industrial Park Branch, Suzhou, China

Corresponding author and e-mail: Junyan Zhang, 108645502@qq.com

Abstract. Real-time and accurate target recognition and extraction of crop fruit images in the field is the key technology for picking robot vision system, and the essence of target extraction is image segmentation. When most of the water (vegetable) fruit is in the picking period, the surface color and the background color are quite different, while the surface colors of the same variety are similar, which are reflected in the different distribution characteristics of the fruit surface color and the background color in the color space. According to this characteristic, an image segmentation algorithm based on color space reference table for fruit target extraction of fruit picking robot vision system is proposed. The algorithm first establishes a color space reference table from the fruit sample image, and then uses a method similar to "convolution" to segment the image according to the color space reference table. Compared with other existing methods, the method is based on color information processing, which can remove the background more cleanly. Without segmentation processing and no complicated operation on the background, the method is beneficial to real-time image processing of the robot.

1. Introduction
The rapid development of fruit farming has increased the market demand for orchard machinery. The labor used in the picking operation accounts for 33% to 50% of the labor used in the whole production process. The picking operation is complicated and the seasonality is strong. If manual picking is used, it is not only inefficient, labor-intensive, but also easy to cause fruit damage, which directly affects the quality of the fruit. The use of the picking robot not only improves the picking efficiency, but also reduces the damage rate, saves labor costs, and improves the economic benefits of the fruit growers. Therefore, it is of great significance to improve the mechanization degree of the picking operation. With the decline of agricultural practitioners and the increasing aging trend, the development and utilization of picking robots has enormous economic benefits and broad market prospects, which meet the needs of social development. The normal operation of any kind of fruit picking robot depends on the correct identification of the working object. Therefore, to realize the fruit harvesting of the fruit picking robot, the key is to identify the fruit from the fruit tree and determine the exact spatial position of the fruit so as to be the robot. The movement provides parameters to complete the fruit picking. Therefore, the performance of the visual system will directly affect the fruit picking effect. The performance of the vision system should be evaluated in terms of both accuracy and speed. There are many factors that affect the accuracy and speed of the visual system, such as unstable illumination, wind, fruit and leaves, and fruit growth. At present, the method of relying solely on machine vision to identify fruits has not yet reached commercialization, and it is necessary to improve the algorithm, reduce the use cost, shorten the picking time, improve the machine recognition rate and picking efficiency as soon as possible.
2. Problems and analysis of the visual system of fruit picking robot

Through the analysis of the current situation of fruit picking robots at home and abroad in the previous section, it is not difficult to find that the research results on visual direction are relatively rare in our country. Due to the increase in the output of the domestic fruit industry, the corresponding agricultural product harvesting robots are also urgently needed to be developed. There are many common problems in the process of picking robots to consider from the following aspects:

First of all, in the domestic research on fruit picking robots, the recognition rate of picking is not very high. The identification of fruits is generally studied from several aspects: color component difference, gray value threshold, fruit geometric characteristics and so on. However, in the actual picking environment, there is interference of various information, so that the recognition rate is not high. If the identification is based on the way of fruit characteristics, it is necessary to extract a relatively complete boundary, but usually encounters the overlap of fruits or the occlusion of fruits and leaves. At this time, the identification of fruits is more difficult. The fruit is different in color and growth under different growth environments. These differences will make the standard of image segmentation ununified, resulting in reduced adaptability of the picking robot.

In order to meet the requirements of actual picking: to reduce the labor intensity of workers and improve the efficiency of picking, it is necessary to improve the picking speed. This requires not only the high demands on the mechanical system, but also the speed of image processing. This will result in a higher real-time performance in actual work. In addition, in the actual picking environment, the problem of illumination change will be encountered, which brings a lot of inconvenience to the visual system in image processing, and a certain compensation method is needed in image processing.

The research in this paper is mainly aimed at the problems encountered by apple picking robots at work, mainly from the following aspects: (1) Identification and localization of multiple fruits and identification and localization of multiple overlapping fruits are encountered during the picking process. In the case of the arrival, the fruit growth posture and the occlusion often make it impossible to accurately extract the fruit and locate the fruit at the time of picking. This also reduces the picking efficiency of the fruit picking robot. (2) The accuracy of the binocular vision system is also a consideration. When locating the spatial position of the fruit, you need to know the internal and external parameters of the camera. However, when the camera captures images, the lens distortion problem often occurs, which results in a large error in the internal and external parameters of the camera, which indirectly leads to a large positioning error on the fruit. This requires consideration of the camera calibration, improving the distortion problem and improving the calibration accuracy. (3) Exploring accurate image matching algorithms can reduce the matching error rate, which is especially important for picking work.

3. Color space representation system

In the theory of chromatics, the spaces representing colors are RGB, HSV, YCbC, and L* a* b*. This article focuses on the RGB and L* a* b* spaces involved. The RGB color space is based on the principle of three primary colors and is one of the most basic color representation models. It corresponds to a cube in the Cartesian coordinate system, and R, G, and B represent three coordinate axes, representing red, green, and blue. When R, G, and B both take 0 (that is, the origin of the coordinate), it means black; when both take the maximum value, it means white, as shown in Figure 1. The three components are all between 0 and 255. The L* a* b* model is one of the two color representation systems recommended by the International Commission on Illumination for the uniform evaluation of chromatic aberrations. The advantage is that the luminance components are separated, and the spatial distance can be used to represent the difference between the two colors. L* a* b* medium; L* means brightness, its value is 0 to 100; a* means change from red to green, a* is positive for red, negative is green, and value is (-60, 60); b* means yellow to blue change, b* is positive, yellow means negative, blue is the range, and the range is (-60, 60), as shown in Figure 1. The conversion formula for L* a* b* space and RGB space is

L* = 116(0. 299R + 0. 587G + 0. 114B) 1/3 - 16
a* = 500 [(1. 006(0. 607R + 0. 174G + 0. 048B)] 1/3 - (0. 299R + 0. 587G + 0. 114B) 1/3 ]
b* = 200 [(0. 299R + 0. 587G + 0. 114B) 1/3 - 0. 846(0. 066G + 1. 05B) 1/3 ]
In the L* a* b* color space, a* and b* are luminance-independent components, so the effect of light on the segmentation result in the natural environment can be weakened. These two components can be used to construct the apple sample color space.

4. The role of mathematical morphology algorithm in the process of establishing color space reference table

The fifth step of the color space reference table creation step applies a mathematical morphology operation, which will be described below. The ubiquity of noise in the image determines that the noise obtained from the fourth step also contains the noise of the sample image, which also appears as noise in the result (away from the isolated point of the target region); meanwhile, due to the finite nature of the sample, The sample color distribution area of the result obtained in step 4 is a sparse area in which there are numerous adjacent islands and small holes in the binary image. The purpose of step 5 is to eliminate the noise, fill the small holes and connect the adjacent islands to form a complete area. Expansion and corrosion are the two basic operations of mathematical morphology, let A be the image set and B be the structural element.

The original definition of expansion. The set obtained by expanding B with B is the set of origin positions of B when the displacement of B (the image of B) intersects with at least one non-zero element of A. Improved expansion algorithm. Since there may be a lot of noise in the sample, if the expansion operation actually enlarges the noise according to the original definition of the expansion, the structural element B is taken as a window of 3 × 3, and the origin of the structural element is at the center (Fig. 2). The expansion algorithm is changed to: The set obtained by expanding B with B is the set of origin positions of B when the displacement of B intersects with at least 3 non-zero elements of A. Since the noise is generally far from the target area, there are few non-zero elements around.

The original definition of corrosion. The set obtained by etching A with B is the set of origin positions of B when B is completely included in A. Improved corrosion algorithm. Due to the finite nature of the sample, the result obtained from step 4 is a sparse area. If the corrosion operation is performed according to the original definition of corrosion, all the distribution points may be corroded. Therefore, the same as the expansion, the structural element B is 3 × 3 window, the origin of the structural element is at the center. Change the Corrosion Algorithm to: Corrosion with B. The resulting set is the set of origin positions of B when B is displaced by a maximum of 3 zero elements. Also because the noise is generally far from the target area, most of the surrounding is zero elements, so the noise is eroded (filtered), and useful points inside the target area are preserved.
After the expansion and corrosion algorithms are improved, the internal expansion of the target area plays a leading role, and the external corrosion of the target area plays a leading role, and the boundary of the area acts inversely; therefore, a reasonable combination of expansion and corrosion operations can obtain a tightly distributed complete target area. At the same time, noise is eliminated and the boundaries of the region are preserved.

5. Binocular image matching

The dual-purpose stereo matching is based on the fact that two cameras are located at different visual points to capture the same position and then acquire stereoscopic image pairs. The parallax is obtained from the stereo image pair of the target, and finally the three-dimensional geometric information of the target is obtained. In binocular matching, there are many influencing factors in the shooting environment that make the captured image pairs very different, mainly in the following aspects: noise, grayscale and geometric distortion. In the research of binocular stereo matching algorithm, the researchers have proposed a variety of matching algorithms for these problems, but there is still no stable adaptive algorithm. These algorithms are classified according to different constraints: local algorithms and global algorithms. The focus of the global optimization method is on the calculation of parallax, and most of these algorithms are based on the minimum energy. For local optimization, the main consideration is the matching cost calculation and matching cost. Compared with the former, this method is easier to obtain the disparity value, because it is only necessary to determine the disparity value when the matching value is the smallest in the calculation. In actual research, the matching algorithm is generally considered from several aspects:

First, to determine the primitives that can be adapted to the matching object, according to the difference of the matching objects, it is possible to select from the regional aspect, the phase angle or the linear feature in the experiment. The determination of the primitives is crucial to the effectiveness of the experiment. Secondly, in addition to selecting the appropriate primitives, it is also necessary to determine the matching criteria in the features of the image alignment, such as whether the order between the features is consistent, continuous, and not similar or even unique. Finally, based on the above two points, create a real-time effective matching algorithm structure. Through comparison and research, the SSD algorithm is selected as the matching algorithm, which is suitable for the characteristics of the visual system in terms of sensitivity and speed.

The matching primitive has been mentioned in the previous section. Under normal circumstances, it is unrealistic to require the target to match all the image points in the acquired image pair. First, the image is different from the surrounding environment. However, in the process of matching, you can find as many matching points as possible to match. Based on this, it is very important to select the attribute characteristics of the appropriate target. The purpose of selecting a matching primitive is to obtain a corresponding image feature from the image pair. There are three types of primitives that are often chosen: region, feature, and phase match. These three types are suitable for different situations, so when making a match, analyze the feasibility. The regional primitives have the fewest occurrences in the image, but contain the most content, so they have the characteristics of strong object differentiation and high precision when matching. It is the easiest and easy to operate algorithm for the current matching algorithm. The main purpose is to use gray value similarity to correspond to the relationship between image pairs. In the process of algorithm implementation, one of the image pairs is the research object. And the neighborhood of the gray value to be processed is used as the matching template. A similar function is used to evaluate the similarity of the two, and then another pixel is used to find pixels with similar gray areas. However, it can be seen from the above that the algorithm performs region matching according to the gray value, so it is relatively susceptible to external interference, thereby affecting the matching accuracy and increasing the probability of mismatching. However, this can be pre-processed with image pre-processing techniques, which will greatly reduce noise. In the area matching, we also need to choose the appropriate window. The selection of the window size has a great influence on the matching. If it is too large, it is easy to mismatch. If it is too small, the matching is easy to increase the ambiguity and the accuracy is greatly reduced. Feature-based matching the algorithm can limit the search scope to the features of the target, and then extract the matching features in the image pair to complete the matching work. Since the matching is only for feature attributes during matching, the algorithm is dominant in the matching speed. In addition, the algorithm is not easily disturbed by the outside world, and is relatively stable and has high precision. It is also because of this feature of the algorithm that only the sparse disparity map can be extracted at the end of the matching; and the appropriate features are also selected in the matching to obtain the desired result, the selected features should be highly recognized and the feature should be more dense. Phase matching is a matching using the phase alignment of the image, and matching according to the property that the phases of the corresponding points are the same. The matching algorithm has a good matching effect and has good stability for disturbance factors such as noise and distortion.
6. Fruit target extraction and its experimental study

With the fruit surface color space reference table, it is easy to determine whether a pixel is the target pixel. Here, a method for judging whether a pixel is a target pixel in a statistical sense is used: the number of target pixel points is counted in a 5 × 5 neighborhood of each pixel (whether or not it is a target pixel), exceeding In the case of half, the point is considered to be the target pixel, otherwise it is considered to be a non-target pixel. For the interior of the fruit target area, this is itself a kind of "convolution" filtering; for the outside of the fruit target area, this method can remove the background more effectively. Judging whether a pixel is a target pixel actually obtains a binarized image containing the fruit. The ubiquitous noise and the very rich color of the fruit surface make the judgment based on the reference table model may be wrong, so that the fruit target area of the obtained binary image may have a large number of holes, similar to the 4th when the reference table is established in this paper. The resulting binary image; therefore, the fruit image can also be processed using the improved expansion and erosion algorithms mentioned in Table 3 to fill the holes. After the holes are filled, the area of each fruit can be found by the common area marking algorithm, and the circumscribed rectangles of these areas can be obtained, thereby completing the extraction of the fruit target.

Based on the above algorithm, the author selected strawberry, orange and tomato in the H SVCbCr and L*a*b* color spaces for experiments. The results are shown in Tables 1 and 2. Table 1 shows the distribution of fruit surface color samples in the HS, CbCr and a*b* chromaticity spaces, respectively. Table 2 shows the results of the test samples under the HS V, YCbCr and L*a*b* color models.

The background of the fruit picking robot working environment is complex, but basically similar, and the pixels of the background image also have a certain distribution in the color space. In order to better remove the background, the color space reference table of the background is established according to the method of establishing the color space reference table, and the common part of the background color space reference table is subtracted from the fruit color space reference table, so that the segmentation effect is better. The fruit color space reference table. Test results on tomatoes show that this improvement is necessary.

The author has tested the image with a lot of salt and pepper noise and random noise. The results show that the same effect can be achieved without filtering before the target extraction, because the method described in this paper performs noise throughout the image processing. Special treatment, indicating that the method is fast and simple.

Compared with the gray histogram segmentation method, since the method is based on color, more information in the image is utilized, and the background can be removed more cleanly, thereby facilitating further image analysis. Clustering segmentation algorithm and random model-based algorithm are mostly based on more complex mathematical formulas. This method divides the current image by looking up the table. There is no complicated mathematical operation, and the time complexity is linear. Conducive to real-time image processing in robot control. Region-based methods such as region-growth algorithms, segmentation operations are also performed in the background section, while the region-growth algorithm is a recursive algorithm with high time complexity; this method only deals with fruit targets (sometimes including fruit stems), no processing on the background, saving calculation time, which is conducive to efficient operation of the robot.

After obtaining the color distribution area of the fruit surface, the fruit surface color distribution model can be obtained by fitting the boundary of the region. Compared with the reference table-based method, this method saves memory space (the reference table requires 256 × 256 × 2 B space, if the reference table uses run-length encoding, the actual memory is also very small), but if the boundary is complicated, The calculation is also complicated when the image is divided. Although ellipse can be used to fit the boundary to simplify the calculation when segmenting the image, the assumption that the color distribution is close to the ellipse is not reasonable. In particular, some fruits in the picking period are not fully mature, and the surface color may be distributed in multiple separations. Area. Therefore, compared with the distribution model based method, the reference table based method is not only simpler, but also has a lower misclassification rate.

7. Conclusion

Experiments on the images of strawberries, oranges and tomatoes show that the algorithm based on the fruit color space reference table can quickly target fruit in complex backgrounds under the HS V, YCbCr and L*a*b* color models. The extraction effect is good, which proves that this algorithm is feasible and efficient.
References

[1] Zhang Tiezhong, Lin Baolong, Gao Rui. Target Extraction of Fruit Picking Robot Vision System [J]. Journal of China Agricultural University, 2004(2).

[2] Zhao Jinying, Zhang Tiezhong, Yang Li. Target Extraction of Tomato Picking Robot Vision System [J]. Transactions of the Chinese Society of Agricultural Machinery, 2006, 37(10): 200-203.

[3] Zhou Jun, Liu Rui, Zhang Gaoyang. Design of Fruit Picking Robot System Based on Stereo Vision [J]. Transactions of the Chinese Society of Agricultural Machinery, 2006, 21(6):158-162.

[4] Wu Yiman. Design of Target Recognition Vision System for Mobile Picking Robot Based on DSP [J]. Agricultural Mechanization Research, 2018.

[5] Wen Zhiqiang, Cai Zixing. Moving Target Extraction Method Based on Mobile Robot Vision System [J]. Computer Engineering, 2007(24):203-205.

[6] ZHANG Guiying, Zhang Xianjie. Target Extraction of Mobile Robot Vision Control System [J]. Manufacturing Automation, 2012(17): 71-72.