Image Segmentation of Field Rape Based on Template Matching and K-means Clustering

Dujuan Shuai[1,a], Changhua Liu[1,b,*], Xiaoming Wu[2], Hao Li[2] and Fugui Zhang[2]

[1]School of Math and Computer, Wuhan Polytechnic University, Wuhan, 430023, China
[2]Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Wuhan, 430062, China
Email: [a]912987738@qq.com, [b]liuch@whpu.edu.cn

Abstract. Giving that the changing light in the natural condition has negative impacts on the image segmentation of rape fields, the image of rape was processed by template match algorithm and K-means clustering algorithm to extract the rape flowers. In order to achieve the accurately segmentation of the rape flowers, firstly, creating a template library and using the template matching algorithm to locate the target area of the test image. Then, the processed image will be convert to LAB color space, and using K-means clustering algorithm to classify accurately again. Finally, the extracted rapeseed area will be processed by morphology operation. The experimental results indicate that this method can achieve the goal of extracting the rape flowers completely, and it is effectively to remove the negative impacts of the light.

1. Introduction

In China, the rape is an important economic crop, which is the main source of edible vegetable oil(more than 50%), and also an important part of vegetables, green manure, animal feed, bee product development, and rural tourism[1,2]. In addition, the sown area of rape is 1.35 billion hectares, and the planting scope is extensive. Appropriate flowering period plays an important role for the adaptation of rapeseed[1,3]. The study of the traits of flowering rapeseed is also related to the characteristic of rape yield and disease resistance, therefore, how to identify rape flowering period scientifically and effectively has been an important issue in rapeseed industry[4,5].

The traditional identification of rapeseed flowering period mainly depends on the way of artificial, and this method of identification will cause a waste of time, laborious and subjective. It is difficult to achieve high-throughput, accurate flowering phenotype identification for large groups. The development of automatically identification techniques for flowering period of rape can replace human operations with computers, it can significantly increase the throughput of identification, high accuracy and richness of information collection, and it also save a lot of manpower.

In the study of automatic identification of rapeseed flowering period, the most critical and principal task is to achieve accurately segmentation of rapeseed images, and it will have an influence on distinguishing. Until now, scholars at home and abroad have proposed a lot of methods for image segmentation[6,7], mainly for the following three:

First, the segmentation by the color space transforming. The crop image segmentation that is achieved by analyzing color features[8-12]. These methods are adaptable to the change of light in natural and the calculation is simple. The downside is that when there is more noise in the soil, the dislocation rate is higher.

Second, the segmentation by setting the grayscale threshold of images. These methods have a good effect on the image of monotonous color, but it is susceptible to light and weather. When the light and the weather are changes, the errors are not controlled in the segmentation.
Third, the segmentation by such as unsupervised fuzzy clustering[13-17]. These methods can be well segmented in complex and varied environment, but it requires training samples and some human-computer interaction.

This paper mainly focuses on the issue of image segmentation of Daejeon rape flower, which is sensitive to light changes. Firstly, a standard template library of rape flower is established, and using the template to match the rape images that need to be processed, in order to achieve roughly positioning of rapeseed areas. The matched rape images still have a wrong segmentation area due to the influence of light, so the image is converted to the LAB color space[18]. Using of analyzing and reorganizing the LAB space A and B channels to generate a color recombination matrix of a color image, then, the K-means clustering algorithm is used to cluster the colors of images and setting the maximum number of iterations until the cluster center is finally found to accurately segment rape flower. In addition, compared to the image segmentation based on the RGB color feature algorithm, this algorithm analyzes the super green component of the green crop and adjusts the pixel matrix gray value to extract the area of interest of crop. The experiments show that the image Segmentation of field rape based on template matching and K-means clustering is superior to the RGB color feature algorithm. The rape flower is completely extracted. This operation can effectively remove the influence of light, avoid the over-segmentation of the color feature segmentation algorithm. At the same time, the human-computer interaction is less.

2. Principles of Related Algorithms

2.1 The template matching algorithm

The template matching algorithm is widely used in the field of image recognition and it is one of the most representative algorithms[19,20]. The algorithm idea is to extract a representative region from the image to be processed as a template feature vector, and then the distance is calculated between the image to be processed and the template feature vector, the area is identified based on the calculated minimum distance to determine the position of the region of interest.
Assuming that the sample to be tested is $X = \{a_1, a_2, \ldots, a_n\}^T$, the distance of between the sample data X and the template feature vector $Y_j = \{b_{j1}, b_{j2}, \ldots, b_{jn}\}$ is calculated as

$$d_j = \left(\sum_{i=1}^{n} (a_i - b_{ji})^2 \right)^{\frac{1}{2}} \tag{1}$$

2.2 The K-means clustering algorithm

The K-means clustering algorithm[21] is a mature clustering algorithm with strong self-adaptability and fast operation speed without excessive human-computer interaction. The specific algorithm steps:

Step1: The clustering center is initialized.

Step2: Set the maximum number of iterations to continuously adjust the cluster center.

Step3: If the center of clustering no longer changes or the iteration reaches the maximum number of iterations, stopping iterating, taking the iterative center at this time as the clustering center.

Step4: The data set is divided into K classes. The sum of the Euclidean distances is the smallest between the data in the class and its cluster center, and the Euclidean distance is the largest between classes.

The sample set data processed by the K-means clustering algorithm is multidimensional sample space. Assuming that the sample set is $X = \{x_i \mid x_i \in R^m, i = 1,2,\ldots, n\}$, the point of data x_i as a sample of classification. It is a vector with m representation values, the final clustering goal is to divide the n data points between classes and produces K-classes that is $J = \{j_k | k = 1,2,\ldots, K\}$. Assuming its center of clustering is η_k, then, the Euclidean distance between the data points in the class is contained in each class and the final determined cluster center is:

$$d(x_i, \ \eta_k) = \sqrt{\sum_{l=1}^{m}(x_{il} - \eta_{kl})^2} \qquad (x_i \in j_k) \tag{2}$$

Therefore, the sum of the Euclidean distances of data points from data sets of a class to cluster centers can be expressed as:

$$D(j_k) = \sum_{x_i \in j_k} d(x_i, \ \eta_k) \tag{3}$$

Combining the equations (2) and the equations (3), the Euclidean distance is calculated between each clustered set of data points within its class, and we can know, after dividing the data sample set into sample of K classes, the sum of the total Euclidean distances from the data points in the clustering set to the center of its class can be expressed as:

$$L(x_i) = \sum_{k=1}^{K} D(j_k) = \sum_{k=1}^{K}\sum_{x_i \in j_k} d(x_i, \eta_k) = \sum_{k=1}^{K}\sum_{i=1}^{n} \lambda_i d(x_i, \eta_k) \tag{4}$$

In the equations (3), $\lambda_i = \begin{cases} 1, x_i \in j_k \\ 0, x_i \notin j_k \end{cases}$, this simplified calculation formula, to ensure that the Euclidean distance between all data points and the center of the class clustering is included in the total distance, and it doesn't exist error count. According to the principle of k-means clustering algorithm, to get the minimum $L(x_i)$, The clustering center η_k should be the average of all data points in the classes, constantly getting averages to determine new cluster centers by setting the maximum number of iterations. When the cluster center no longer changes or reaches the maximum number of iterations, at this time the clustering center is finally used as a clustering center.

3. Image Segmentation of Field Rape Based on Template matching and K-means clustering

3.1 The analysis of LAB color space model

In the color space model, it contains a variety of colorful types, various of types and have their own suitable application background. Researchers must analysis each color channel and compare to between them to determine the type of transformation they need. The more common color Spaces are RGB, HSV, HSI, LAB, etc. In this paper, the LAB is selected to process the Daejeon rape RGB image. In the LAB space, the L channel represents the brightness of the color, ranging from [0,100], gradually changing from pure black to pure white; the A channel represents the color change from red to green, with ranging from [+127, -128]; the B channel represents the color change from yellow to blue, with the same range as the A channel. The L channel represents the change in the brightness of the color and therefore has a strong independence during the processing. The impact of lighting cannot be avoided during the collection of Daejeon rape images, the A-channel and B-channel of the LAB color space don’t involve illumination factors, so it is the best choice for the problem of rape flower image segmentation handled by the K-means clustering algorithm.

When obtaining the sample space, the two-dimensional sample data formed by matrix reshaping using the A-channel and the B-channel is used as a sample set, and then the image is clustered using the K-means clustering algorithm, setting the number of clusters K = 3, finally, the rapeseed flower in Daejeon is divided by the data coordinate mapping.

3.2 Dividing the rape flower

In drone aerial images of Daejeon, the pixels are mainly divided into three categories, they are backgrounds, green rape leaves, and rape flower to be segmented. In natural lighting conditions, the K-means clustering algorithm cannot adapt the problems of segmentation in different illumination intensity, therefore, combined the template matching algorithm and K-means clustering algorithm to accurately divide rape flower. The template matching algorithm can locate the target area in the shortest time with fewer sample in libraries. However, the positioning target contains a part of rape leaf, and it is necessary to further using the K-means clustering algorithm to accurately segment rape. According to the actually situation, the number of clusters is set to three. It is mainly divided into the following seven steps:

Step1: The rape flower template is selected. In order to ensure the effectiveness of positioning using the template matching algorithm, the color index algorithm was used to segment the rape, and the region with well segmentation effect was selected as the standard template library.

Step2: Reading the test image.

Step3: The target regions of test image are selected by the template matching algorithm. At last, calculating the minimum distance between the template feature vector and the test image, the target recognition of the image to be measured is achieved.

Step4: The image color space are converted to obtain a sample set for processing. After processing by the template matching algorithm, the rough extraction of the target region has been finished. Converting it to the LAB colorful space, separating the A channel and B channel to get the recombination matrix, and setting the number of rows and columns of the matrix. At this time, X is the final data set.

Step5: The sample data set X is classified by the K-means clustering. Setting the number of clusters K is 3, the number of iterations is 100, and clustering the sample space X. The K-means clustering algorithm is used to initialize the clustering center, the different categories are marked in labels. Then it takes multiple iterations to obtain the average value and the cluster center is updated. When the clustering center no longer changes or reaches the maximum number of iterations, the clustering is completed.

Step6: The clustering results are displayed and the cluster centers are marked. After the data are classified by the K-means clustering algorithm, the cluster centers of variously types are obtained, the clustering effect and the cluster center point are shown in a two-dimensional sample space.

Step7: Dividing the rape flower. Using the clustering algorithm, the test image is converted into a two-dimensional data set, one of classes which is the pixel of the rape flower. In Step 5, the different categories have been marked, according to the marker type, the pixels is set to black that it isn't indicated the rape flower. The rape flower segmentation has been completed.

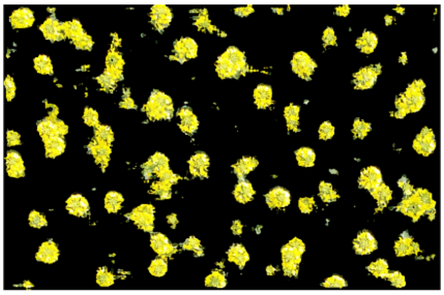
4. Analysis of the experimental results

4.1 The image segmentation experiment of daejeon rape

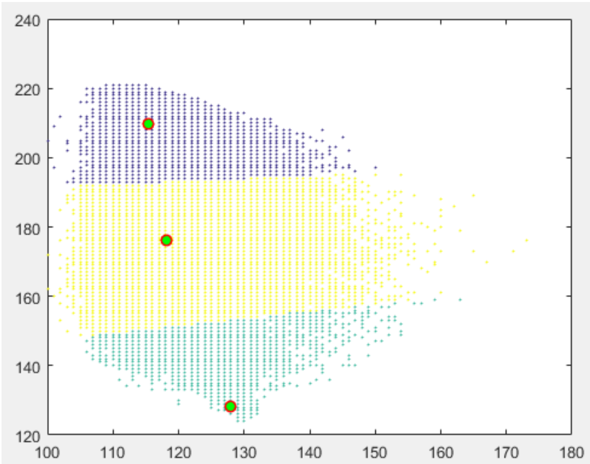
All test images in this article are provided by Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences. The experiment subjects are rape of by large-scale planting, and the way of collection is drone aerial photography. The segmentation experiments are simulated by MatlabR2016b. After the template library is obtained by the template matching algorithm, the target region is located, the K-means clustering algorithm is used to segment the rape flower, and the segmented image is employed to morphological processing.

Fig.1 shows the blooming scene of rape under the light of natural, and Fig.2 shows the scene with different flowering degrees.

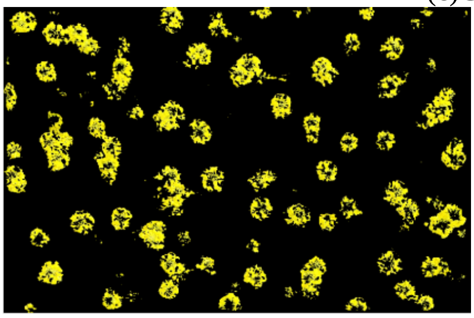
(a)The original image



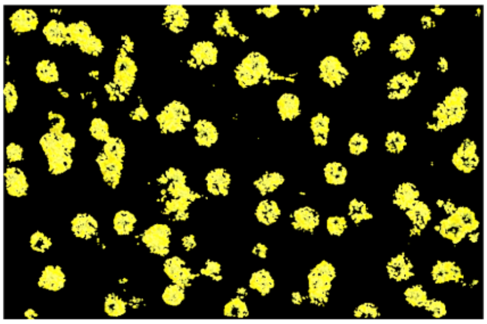
(b)Target area positioning



(c)Clustering results



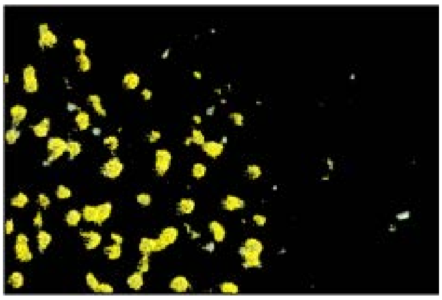
(d)The segmented rape flower



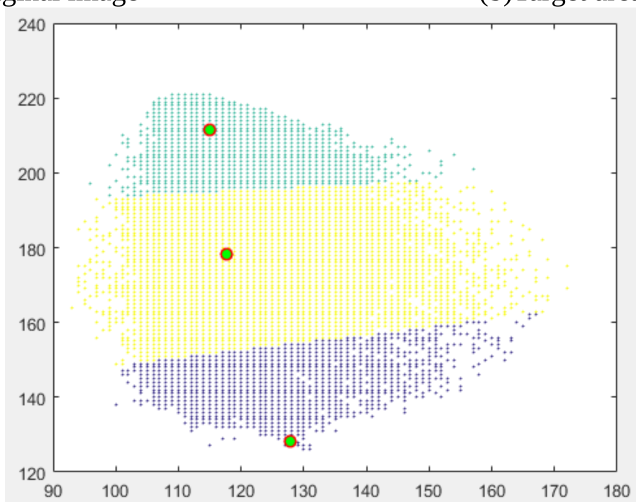
(e)Dilated

Figure 1 The effect of segmentation

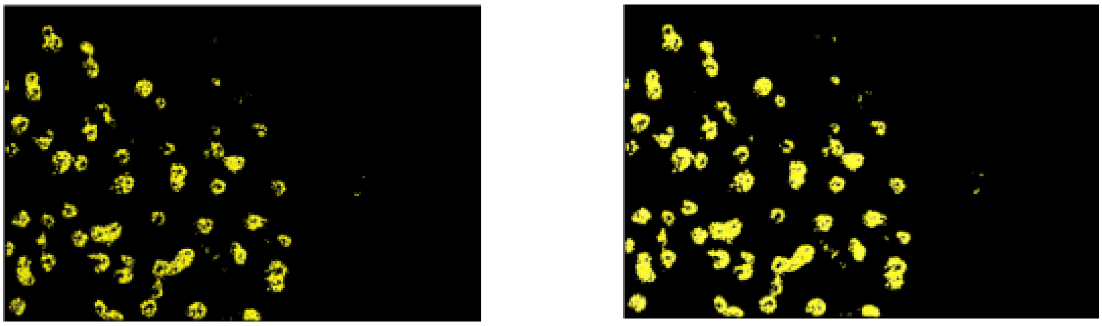
(a)The original image



(b)Target area positioning



(c)Clustering results



(d)The segmented rape flower (e)Dilated

Figure 2 The effect of segmentation

4.2 The analysis of experiment

The experimental results shown in Fig.1 and Fig.2 indicate that the precision of the template matching algorithm is accurately in the recognition of the target region by creating the template library. When the flowering degree of plants is different, the area of target can be accurately positioned. The problem of deviation of target segmented in the processing of images by the K-means clustering algorithm is prevent. As shown in the Fig.(b), there are still leaf residue phenomena in the area after positioning, which is not enough to achieve accurately extraction of rape flower., so the K-means clustering algorithm is used to further classify for the target area positioning image, and setting the area outside the canola is black. The segmented rape flower was dilated by morphology processing to ensure that the shape of flower was complete, and the experimental results could reach the requirements of field rape image segmentation.

5. Conclusion

This article has achieved accurate rapeseed segmentation for large-scale cultivation of rape in the field. It has broad application prospects for realizing high-throughput precision and identification of the flowering period of rape populations. Aerial pictures of daejeon rape was captured under natural lighting conditions, affected by illumination quite, differences in the degree of day lighting due to differences in the flowering of field crops, extracting rapeseed pixels, traditional color indexing algorithms are extremely sensitive to background noise, the specific threshold segmentation method deals with a single range, it isn't suitable for processing rich colors and light intensity image. This article considers that color images can easily cause over-segmentation and missed target extraction and so on, using a combination of template matching algorithm and the K-means clustering algorithm to extract rape area. Summary for the following four points:

(1) Although the traditional method of threshold segmentation reached the effect of the rape feature extraction, the extraction is roughly, and it is sensitive to background noise and illumination changes. The effect of extraction is serious affected by it that whether the threshold is appropriate or not. It can't effectively solve the problem of image segmentation for the processing of aerial image recognition.

(2) To solve these problems such as the poor segmentation adaptability in changes of light intensity, over segmentation of images, and dividing the background of the target region, in this paper, the template matching algorithm was used to establish a standard template library, and the target region of the image to be processed was accurately positioned. The stems and leaves of the rape are separated.

(3) After the target region is selected by the template matching algorithm, there is a phenomenon that the part of leaves are divided into it. At this time, the region is converted to the LAB colorful space, the A channel and B channel are reconstructed. By the K-means clustering, it realized the color segmentation and morphological processing. It prevents the positioning area deviation in the template matching algorithm, the rape flower is accurately extracted.

(4) The experiments show that combined the template matching algorithm and K-means clustering algorithm can effectively avoid the error segmentation because of the light, and the target area is accurately and the calculation speed is fast. It is suitable for solving the problems of image segmentation under the complex background.

6. Acknowledgments
This research was supported by the National Key Program for Research and Development(2016YFD0100202).

7. References
[1] D.H. Fu, L.Y. Jiang. Mason A S, et al. Research progress and strategies for multifunctional rapeseed: A case study of China, J. Journal of Integrative Agriculture, 15(2016) 1673-1684. Doi: 10.1016/S2095-3119(16)61384-9.

[2] H.Z. Wang, Y. Yin. Analysis and strategy for oil crop industry in China, J. Chinese journal of oil crop science, 36(2014) 414-421. Doi: 10.7505/j.issn.1007-9084.2014.03.020

[3] L. Xu, K. Hu, Z. Zhang, et al. Genome-wide association study reveals the genetic architecture of flowering time in rapeseed (Brassica napus L.), J. Dna Research An International Journal for Rapid Publication of Reports on Genes & Genomes, 23 (2015) 43. Doi: 10.1093/dnares/dsv035.

[4] Raman H, Raman R, Coombes N, et al. Genome-wide association analyses reveal complex genetic architecture underlying natural variation for flowering time in canola, J. Plant Cell & Environment, 39 (2016) 1228-1239. Doi: 10.1111/pce.12644.

[5] N. Wang, B. Chen, K. Xu, et al. Association Mapping of Flowering Time QTLs and Insight into Their Contributions to Rapeseed Growth Habits, J. Frontiers in Plant Science, 2016. Doi: 10.3389/fpls.2016.00338.

[6] Y.H. Fang, Research on Automatic Identification for Rapeseed Plant Growth Stages Based on Computer Vision, D. Huazhong Agricultural University, 2015.

[7] L.L. Zhou, F. Jiang. Survey on image segmentation methods, J. Computer application research, 3407 (2017) 1921-1928.

[8] B.G. Shen, S.R. Chen, J.J. Yin, H.P. Mao. Image recognition of green weeds in cotton fields based on color feature, J. Journal of Agricultural Engineering, 06 (2009) 163-167.

[9] R.F. Zhai, Y.H. Fang, C.D. Lin, et al. Segmentation of field rapeseed plant image based on Gaussian HI color algorithm, J. Journal of Agricultural Engineering, 32 (2016) 142－147.

[10] Z.B. Zhang, X.W. Luo, Y. Zang, et al. Segmentation algorithm of green crop based on color feature, J. Journal of Agricultural Engineering. 27 (2011) 183-189.

[11] Y.L. Wu, L. Zhao, H.Y. Jiang, X.Q. Guo, F. Huang. Image segmentation method for green crops using improved mean shift, J. Journal of Agricultural Engineering, 30(2014) 161-167.

[12] B. Zhao, Z.H. Song, W.H. Mao, E.R. Mao, X.C. Zhang. Agriculture Extra-green Image Segmentation Based on Particle Swarm Optimization and K-means Clustering, J. Journal of Agricultural Machinery, 40(2009) 166-169.

[13] R. Hettiarachchi, J. F. Peters. Voronoi region-based adaptive unsupervised color image segmentation, J. Pattern Recognition, 65(2017) 119-135.

[14] L.M. Xu, J.D. Lv. Bayberry image segmentation based on homomorphic filtering and K-means clustering algorithm, J. Journal of Agricultural Engineering, 31(2015) 202-208.

[15] X.C. Li, H.K. Liu, F. Wang, C.Y. Bai. The survey of fuzzy clustering method for image segmentation, J. Journal of Image and Graphics, 17(2012) 447-458.

[16] M.X. Zhang, L.C. Jiao, W.P. Ma, J.J. Ma, M.G. Gong. Multi-objective evolutionary fuzzy clustering for image segmentation with MOEA/D, J. Applied Soft Computing, 48(2016) 621-637.

[17] C. Li, D.Q. Feng. Research of K-means Clustering Algorithm for Images Based on Weighted Euclidean Distance, J. J. Zheng zhou Univ.(N at Sci. Ed.), 43(2011) 109-113.

[18] R. Huang, N. Sang, D.P. Luo, Q.L. Tang. Image segmentation via coherent clustering in LAB color space, J. Pattern Recognition Letters, 32(2011) 891-902.

[19] J. Gao, X.W. Li, J. Zhang, B.H. Lu. Image Registration Algorithm Based on Template Matching, J. Xi′an Jiaotong University, 41(2007) 307-311.

[20] J. Li, H. Zhou, J. Zhang, B. Gao. Tracking algorithm by template matching based on particle swarm optimization, J. Journal of Computer Applications, 35(2015) 2656-2660.

[21] X.G. Wang, S.H. Chen. The Ripe Strawberry Image Segmentation Algorithm Based on K－means Clustering, J. Agricultural Mechanization Research, 35(2013) 51-54