http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (23): 7574-7579, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.7574.7579

Classification of Lettuce Nitrogen Levels Based on Image Feature Extraction and Optimization

[1,3]Sun Jun, [1]Jiang Shuying, [2]Mao Hanping, [2]Zhang Xiaodong, [2]Zhu Wenjing and [1]Wang Yan
[1]School of Electrical and Information Engineering of Jiangsu University,
212013, Zhenjiang, People Republic of China
[2]Laboratory Venlo of Modern Agricultural Equipment, Jiangsu University,
212013, Zhenjiang, People Republic of China
[3]Key Laboratory of Agri-informatics, Ministry of Agriculture, 100081, Beijing, People Republic of China

Abstract: The feature extraction and optimization of lettuce leaf image are the important premise of classification recognition of lettuce nitrogen levels. The lettuce samples of different nitrogen levels were cultivated in soilless cultivation using nitrogen nutrition of different concentrations. When the lettuce leaf images were collected, image features have been extracted, including texture features, shape features and color features. Because of the redundancy of characteristic values, there were influences in the accuracy and efficiency of image recognition. Genetic algorithm was used to optimize 11 eigenvalues and the Principal Component Analysis (PCA) dimension reduction method was used to choose 12 principal component feature values whose cumulative contribution rate reached 98.24%. Later, the Support Vector Machine (SVM) was used as classifier. The 90 samples were chosen as training samples and the remaining 30 samples were chosen as the test samples. The result shows that, the prediction accuracy of SVM classifier based on genetic algorithm feature optimization reaches 93.33% and that based on PCA features optimization reaches 76.67%. So the genetic algorithm feature optimization is more suitable for lettuce leaf image feature optimization.

Key words: Feature selection, data dimensionality reduction, genetic algorithm, PCA

INTRODUCTION

Nitrogen level is very important during the growth period of lettuce. Insufficient or excessive nitrogen content can make cell change in blade, which directly affect the color, shape and texture characteristics of lettuce leaves. To extract and analyze these characteristics effectively is the key to forecast N element level in lettuce leaves. Because the number of features reflecting color, shape and texture of lettuce leaf is large, so the optimal feature selection is important. Principal Component Analysis (PCA) is a feature extraction method used commonly. Kapalavayi and Sharma, A. applied feature selection method based on hierarchy to text classification and achieved good results (Kapalavayi *et al*., 2006; Sharma and Kuh, 2008). At present, with the development of genetics, the literature (Guan *et al*., 2010) proposed feature selection based on genetic algorithm and simulated annealing algorithm method and apply it to feature selection in text, in order to reduce the dimension of feature vector. Considering the quickness and the effectiveness of genetic algorithm, in this study, genetic algorithm is applied to optimal selection of lettuce characteristics and is compared with PCA method, to select a method which is more suitable for image characteristics optimization of lettuce leaves. At present, the report about using genetic algorithm in plant leaves image features optimization is rare.

MATERIALS AND METHODS

Planting and collecting of samples: In order to obtain the samples in different nitrogen levels, Japan yamazaki formula is used in this study. Perlite bag, nutrient solution with distilled water and the automatic irrigation system for irrigation were used to cultivate the samples of lettuce. Nitrogen levels of nutrient solution include normal nitrogen level, deficient nitrogen level and excessive nitrogen level.

The special requirements of fieldwork should be taken fully consider when selecting image acquisition device. The below problems should be considerd, such as whether or not portable about the device, whether or not using a dc power supply equipment, whether or not the image precision meets the experiment requirement, etc. The camera model is determined as Canon of

Corresponding Author: Sun Jun, School of Electrical and Information Engineering of Jiangsu University, 212013, Zhenjiang, China

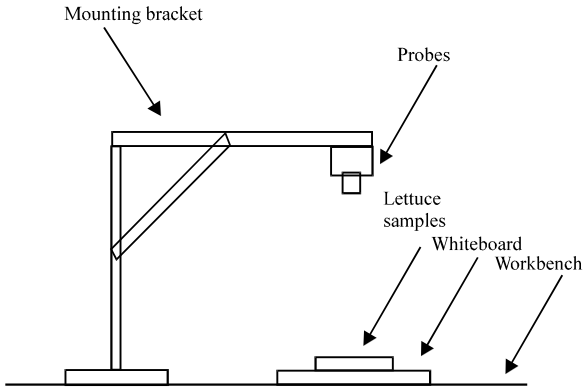

Fig. 1: Image acquisition device

60D, with18 million pixels. And the digital camera was fixed on the bracket, making the camera keep a distance of 20 cm with lettuce leaves, shown in Figure 1. In order to ensure sufficient lighting, the shooting time is between 11 and 13 o'clock. LED is used in order to solve the uneven illumination of lettuce leaf surface. Image gray-scale transformation, image enhancement, image segmentation, expansion and corrosion treatment in the image preprocessing stage, can make the image more clear to select and extract feature easily.

Image segmentation of lettuce leaf: In the image acquisition process of lettuce, The external factors have interference in image sensor,and will directly affect the lettuce analysis of nitrogen level. In order to extract the feature effectively and accurately, first of all to preprocess lettuce leaves, including: image enhancement, segmentation, etc.

Two-dimensional maximum entropy segmentation algorithm: Set image gray scale as L and a_0 (x, y) as the image gray value in the point (x,y), a_2 (x, y) as the average gray level values of the k×k neighborhood in the point of (x, y) as the center whose expression is shown as Eq. 1:

$$a_2(x,y)=\frac{1}{k^2}\sum_{m=-\frac{k}{2}}^{\frac{k}{2}}\sum_{n=-\frac{k}{2}}^{\frac{k}{2}}a_0(x+m,y+n) \tag{1}$$

Among them, $1\le x+m\le M, 1\le y+n\le N, M, N$ as the width and height of the image, K as 3 (Lufang and Leye, 2005). If (x+m, y+n) exceeds the image width and length, the boundary point will be taken instead of a_0 (x+m, y+n).

For a gray image of M*N, if the frequency of binary group (i, j) is f_{ij}, the corresponding joint probability density is:

$$p_{ij}=\frac{f_{ij}}{M\times N} \tag{2}$$

In it, I, j = 0, 1, …… L-1:

$$\sum_{i=0}^{L-1}\sum_{j=0}^{L-1}p_{ij}=1$$

Entropy function expressions of point (i, j) is as follows:

$$\varphi(i,j)=\lg\left(\sum_{i=0}^{s}\sum_{j=0}^{t}P_{ij}\right)+\frac{\sum_{i=0}^{s}\sum_{j=0}^{t}P_{ij}\lg P_{ij}}{\sum_{i=0}^{s}\sum_{j=0}^{t}P_{ij}}+\lg\left(\sum_{i=s+1}^{L-1}\sum_{j=t+1}^{L-1}P_{ij}\right)+\frac{\sum_{i=s+1}^{L-1}\sum_{j=t+1}^{L-1}P_{ij}\lg P_{ij}}{\sum_{i=s+1}^{L-1}\sum_{j=t+1}^{L-1}P_{ij}} \tag{3}$$

The best threshold value (s, t) is when the φ (I, j) taken the maximum value (i, j).

Improve segmentation algorithm: Because two-dimensional maximum entropy segmentation algorithm makes full use of the gray level information and spatial information of images, so it is used widely. Luo *et al.* (2010) improved the traditional maximum entropy segmentation algorithm to solve the question of low precision and poor segmentation results and had achieved better segmentation results (Luo *et al*., 2010). But it assumes that the background region and object area occupied the most of two-dimensional histogram area, ignoring the border area information, so the effect of segmentation is poor in many circumstances. Based on this problem, an improved image segmentation algorithm was applied in this study (Sun *et al*., 2012). Two-dimensional maximum entropy segmentation algorithm will be used to redistrict the calculation area and the minimum fuzzy entropy segmentation is used to redistrict the pixels of image.

Because the traditional two-dimensional maximum entropy segmentation algorithm has not considered the probability distribution of diagonal threshold vector points fully, it leads to the low segmentation accuracy and even the wrong segmentation phenomenon. Therefore, the traditional two-dimensional maximum entropy is improved to make up the shortage in this study. Specific practices are as follows:

First of all, the segmentation area based on the traditional two-dimensional maximum entropy segmentation algorithm is redistricted, making the segmentation area contains more pixels near the threshold. Then, membership function is constructed. It reflects the belonging degree of background and goals

Fig. 2: Original gray lettuce leaves

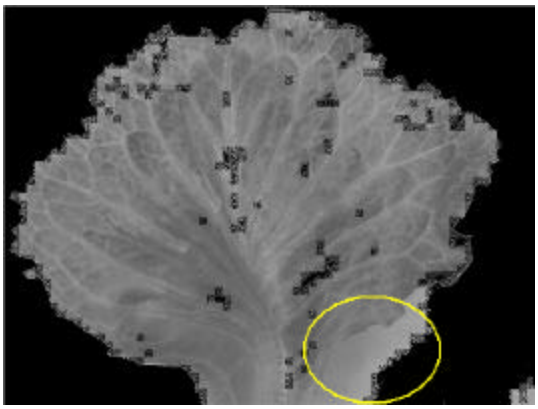

Fig. 3: Segmentation image based on two-dimensional OTSU

and then reclassified the image pixels by using fuzzy minimum entropy algorithm. The original gray image of lettuce leaf is shown as Fig. 2, 3 for the segmentation image based on two-dimensional OTSU is shown as Fig. 3 and the segmentation image based on improved segmentation algorithm image is shown as Fig. 4.

In Figure 3, the part tagged is background being mistaken as lettuce leaf target, however, there is no wrong points phenomenon in Fig. 4.

The region segmented with two-dimensional maximum entropy segmentation algorithm was redistricted and fuzzy minimum entropy is used for subsequent processing, the pixels of segmentation image were reclassified. The results showed that the segmentation effect with the method in my study is improved and the segmentation image is more comprehensive and has relatively complete edge information and strong antinoise ability, etc.

Fig. 4: Segmentation image based on improved segmentation algorithm

Feature extraction of lettuce leaves: The features of lettuce leaves were analyzed and the texture, shape and color features of lettuce leaves were extracted.

Texture feature: In order to reduce the complexity of computation, in this study, we choose consistency, third order moment, smoothness, energy, entropy, average gray scale, standard deviation, variance, roughness, contrast and orientation degree, total of 11 texture eigenvalues (Laddi *et al*., 2013).

Shape feature: In order to ensure that the extracted eigenvalue can not change with image translation, rotation, so seven moment invariants (η_1, η_2, η_3, η_4, η_5, η_6, η_7,) is chosen as recognition features of lettuce leaf nitrogen level (Chen *et al*., 2012).

Color features: Because the components of HIS have very strong independences and the distribution information of image color is mainly concentrated in the low moments, so in this study, the first moment, secondary moment and the third moment of the three components such as H, S, I, a total of 9 characteristic value were computed to express the color distribution of lettuce leaf image.

Because the second moment of S and I value is complex, so it is given up and keep the rest of the 7 characteristic values and its expression is shown as below:

$$\mu_j = \frac{1}{N}\sum_{i=1}^{N} H\left(Q_{i,j}\right) \quad (4)$$

$$\sigma_j = \sum_{i=1}^{N}\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(H\left(Q_{i,j}\right)-\mu_j\right)^2} \quad (5)$$

$$s_j = \sum_{i=1}^{N}\sqrt[3]{\frac{1}{N}\sum_{i=1}^{N}\left(H\left(Q_{i,j}\right)-\mu_j\right)^3} \quad (6)$$

Among them, Qij as the probability of first j color component and gray of the pixel is i. N is the number of pixels in the color image.

Data dimensionality reduction based on genetic algorithm: Genetic algorithm (GA) is a searching algorithm based on the principle of natural selection and it simulated the evolutionary mechanism in the nature, having been achieved optimization to certain targets in the artificial recognition system (Tang *et al*., 2011). Specific optimization process is as follows.

The 25 characteristic values were coded using the binary encoding and every chromosome genes corresponds to a feature. If a gene of one chromosome is 1, it means that the characteristic value of gene is selected, otherwise, it is ignored. Because the lettuce characteristic value is 25 dimensions, then the chromosome length L is 25.

Population size was set as 25, the number of iterations as 200 , crossed factor (Pc)as 0.1, variation factor (Pm) as 0.01. Initial population was generated randomly, including 25 individuals, expressed as: (x1, x2, x3 ... x25).

Later, the fitness function value was calculated. Distance criterion between Intra-class and Inter-class was used as fitness function and the value determines the classification capacity of chromosomes c . The expression is as below:

$$F_{fitness}(c) = \frac{tr\left(S_b(c)\right)}{tr\left(S_w(c)\right)} \quad (7)$$

S_b (c) as the distance of the Inter-class, S_w (c) as the distance of the Intra-class.

The evolution of the population includes selection, crossover and mutation operation. When the number of iterations reaches 200, or the absolute value of the fitness function value is less than 0. 001 in 20 consecutive generation, calculation terminated and the optimal solution was output.

Through the experiment, the final result is: (11001010001110010001001 01), namely, the following 11 characteristic values, such as the consistency, third order moment, entropy, standard deviation, orientation degree, η_1, η_2, η_5, I2, S2 and H2, were selected. In characteristic optimization process, the optimal individual fitness function value changes along with the change of the number of iterations trend, as shown in Fig. 4. From the Fig. 5, we can see that when the number of iterations reaches 163 generation, it produces the optimal individual.

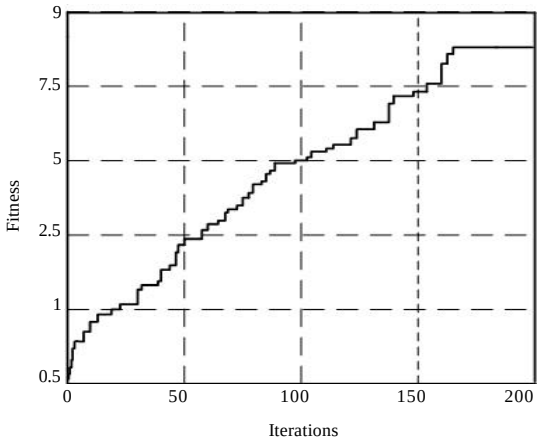

Fig. 5: Best individual fitness value variation tendency along with the number of iterations

Dimensionality reduction based on PCA: The purpose of Principal component analysis (principal component analysis) is to use less variables to explain the vast majority of information in original data and change the original variable value having high correlation into variable values being independent each other (Ning *et al*., 2010). In this study, PCA method was applied to reduce data dimensionality of lettuce leaves' characteristic value. The concrete steps are as below:

- The original data of lettuce characteristics were processed to standardization
- The correlation coefficient matrix among lettuce leaf features was calculated
- Eigenvalue and eigenvector of lettuce leaves were calculated
- p (p = m) principal components were chosen and the contribution rate and the cumulative contribution rate of information were calculated and the corresponding feature vector was calculated vector, which is the principal components wanted

Through the above steps, the 25 eigenvalue of the lettuce leaves images were reduced dimensionality and finally 12 main components whose cumulative contribution rate is 98.24% were chosen.

RESULTS AND DISCUSSION

SVM classifier: The sample is set as $\{x_i, y_i\}, i = 1, 2 \cdots n$, n is the number of training sample set, $x_i \in R, y_i \in \{-1, +1\}$ is classification categories and the SVM decision function is as below (Bazi and Melgani, 2007):

$$f(x) = \text{sign}\left\{\sum_{i=1}^{n} a_i y_i K(x_i \bullet x) + b\right\} \quad (8)$$

The sign is the symbol function, a_i is the Lagrange multiplier.

In the experiment, we used radial basis kernel function, respectively.

The radial basis kernel function:

$$K(a_i, a_j) = \exp\left\{-\frac{\|a_i - a_j\|^2}{\sigma^2}\right\} \quad (9)$$

When SVM was used to solve nonlinear and high dimensional pattern recognition as well as the small sample problem, it showed the advantages of simple calculation and robustness, so it was introduced into other machine for problems in learning.

Result comparison of PCA SVM and GA-SVM: To examine the characteristics optimization effect of genetic algorithm and the effect of reducing dimensionality of PCA, in the stage of training support vector machine (SVM), 90 samples data of lettuce leaf image were selected, including 30 of normal nitrogen, 30 of deficient nitrogen and 30 of excessive nitrogen. There are 30 testing samples, including 10 of normal nitrogen, 10 of deficient nitrogen and 10 of excessive nitrogen. Through training Support Vector Machine (SVM) using training data set, prediction model was obtained and prediction experiment was made using test set. The chart of forecast effect is shown in Fig. 6, 7 and 8. Figure 6 showed the recognition effect based on genetic algorithm for optimizing feature data, Fig. 7 showed the recognition effect after PCA for reducing the dimension of data, Fig. 8 showed the recognition effect under the condition of not reducing dimensionality. The specific comparison results about prediction time needed and prediction accuracy were shown in Table 1.

Experimental results showed that there were only 2 fault recognition points in the process of identification based on optimization in characteristics using genetic algorithm and there were 6 fault recognition points in the process of identification based on reducing dimensionality using PCA and there were 9 fault recognition points under the condition of without treatment.

The recognition time under without treatment of reducing dimension has longer recognition time in the process of identification and it proved that, it was necessary to make characteristics optimization in predicting data. In addition, compared with PCA method, the method of reducing dimensionality

Table 1: Table of recognition accuracy and recognition time

	Recognition effect based on genetic algorithm	Recognition effect after PCA	Recognition effect under the condition of not reducing dimensionality
Prediction accuracy (%)	93.33	76.67	70.00
Prediction time needed (sec)	108.73	126.38	189.64

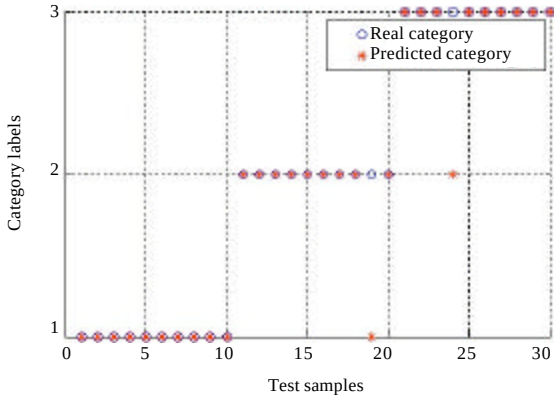

Fig. 6: Prediction results of GA-SVM, real and forecast category picture of test samples

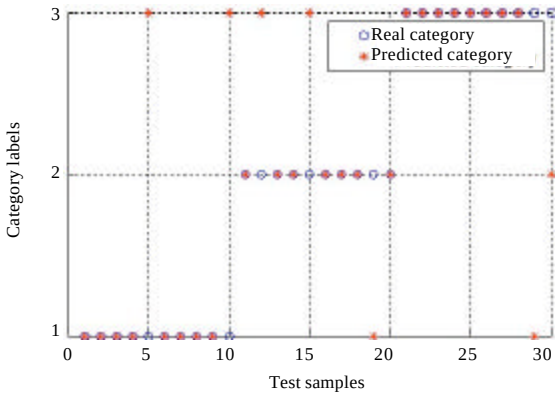

Fig. 7: Prediction results of PCA-SVM, real and forecast category picture of test samples

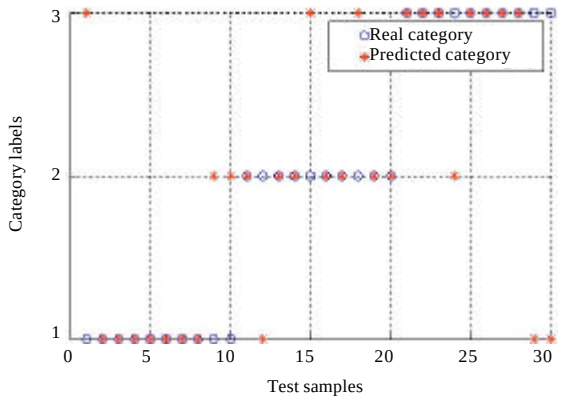

Fig. 8: Prediction results of untreatment, real and forecast category picture of test samples

based on genetic algorithm is more effective with a high classification accuracy and short time needed.

CONCLUSION

In this study, the genetic algorithm and PCA dimensionality reduction method were applied to optimize characteristic in predicting the nitrogen level in lettuce leaves respectively. The results showed that 11 dimension feature vectors were chosen finally after optimization in genetic algorithm, removing 14 dimension redundancy feature vectors, then 12 principal components were chosen after reducing dimension with PCA. Prediction results showed that in the condition under without treatment, the data identification efficiency is low and recognition time is longer. No matter in recognition accuracy or in recognition time, the genetic algorithm is better than PCA method and the PCA has some limitations in optimizing nonlinear data.

ACKNOWLEDGEMENT

This work is supported by National natural science funds projects(No.31101082, No.61075036), A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions(PAPD), the Open Fund Project of KeyLaboratory of Agri-informatics, Ministry of Agriculture (2013007).

REFERENCES

Bazi, Y. and F. Melgani, 2007. Semisupervised PSO-SVM regression for biophysical parameter estimation. IEEE Trans. Geosci. Remote Sensing, 45: 1887-1895.

Chen, B.J., H.Z. Shu and H. Zhang, G. Chen, C. Toumoulin, J.L. Dillenseger and L.M. Luo, 2012. Quaternion zernike moments and their invariants for color image analysis and object recognition. Signal Proces., 92: 308-318.

Guan, H.O., S.H. Xu and F. Tan, 2010. Image segmentation model of plant lesion based on genetic algorithm and fuzzy neural network. Trans. Chinese Soc. Agric. Machinery, 41: 163-167.

Kapalavayi, N., M.S.N. Jayaram and G.Z. Hu, 2006. Hierarchical approach to select feature vectors for classification of text documenta. Proceedings of the International Conference on Computer Systems and Applications, March 8-11, 2006, Sharjah, United Arab Emirates, pp: 1180-1183.

Laddi, A., S. Sharma, A. Kumar and P. Kapur, 2013. Classification of tea grains based upon image texture feature analysis under different illumination conditions. J. Food Eng., 115: 226-231.

Lufang, Z. and G. Leye, 2005. 2-D maximum entropy method in image segmentation based on genetic quantum algorithm. Comput. Applic., 8: 25-25.

Luo, G.H., W.M. Huang and L. Song, 2010. 2-D maximum entropy spermatozoa image segmentation based on canny operator. Proceedings of the International Conference on Intelligent Computing and Integrated Systems, October 22-24, 2010, Guilin, pp: 243-246.

Ning, Y.Y., W.J. Li and X.N. Wang, 2010. Method for vehicle-logo recognition based on principal components analysis and BP neural network. J. Liaoning Normal Univ., 2: 179-184.

Sharma, A. and A. Kuh, 2008. Class document frequency as a learned feature for text categorization. Proceedings of the International Joint Conference on Neural Networks, June 1-8, 2008, Hong-Kong, China, pp: 2988-2993.

Sun, J., Y. Wang, X. Wu, X. Zhang and H. Gao, 2012. A new image segmentation algorithm and its application in lettuce object segmentation. Telkomnika, 10: 557-563.

Tang, B., J.Y. Kong and X.D. Wang, 2011. Feature dimensions reduction and its optimization for steel surface defect based on genetic algorithm. J. Iron Steel Res., 23: 59-62.