

Plant Leaves Recognition Combined PCA with AdaBoost.M1

Hui Chen^a, Haodong Zhu^{b,*}, and Xufeng Chai^c

^aEngineering Training Centre, Zhengzhou University of Light Industry, Zhengzhou, 450002, China

^bSchool of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou, 450002, China

^cSida Foreign Language Primary School, Henan Experimental High School, Zhengzhou, 450000, China

Abstract

In order to improve the overall performance of plant leaves recognition, this paper proposed a novel method combining PCA with AdaBoost.M1 to recognize plant leaves. The proposed method firstly carries out the image preprocessing, which includes the image gray processing, the image binarization, and the edge extraction; extracts the 13 features of plant leaf with the characteristics of rotation invariance, proportion invariance, and translation invariance; subsequently employs PCA to reduce the dimensions of these feature parameters; and finally adopts the AdaBoost.M1 classifier to train and recognize the reduced-dimension plant leaf images. Simulation experiment results indicate that the proposed method is able to improve the overall performance effectively of plant leaves recognition.

Keywords: plant leaves recognition; performance improvement; PCA; AdaBoost.M1; image processing

(Submitted on November 8, 2018; Revised on December 10, 2018; Accepted on January 12, 2019)

© 2019 Totem Publisher, Inc. All rights reserved.

1. Introduction

As the most abundant life form on earth, plants are the essential environmental resources for humans. They play an important role in maintaining ecological balance, improving climate, and so on. In recent years, with the economic development of human society, competition between man and nature is increasing. Plant species are falling sharply, and a large number of plant species are on the verge of extinction. Therefore, the task of recognizing plants to maintain plant species diversity is becoming more and more urgent.

Traditional plant leaves recognition relies mainly on manual operations and requires a large amount of personnel and time. The recognition results are susceptible to subjective factors of technicians. With the development of information processing technologies, it has gradually become a new tendency to solve the problems of plant leaves recognition by means of digital image processing technology and machine learning technology. Many research works focus on plant leaves recognition by means of new information processing technologies and have obtained abundant research results [1]. For example, Lee [2] adopted deep learning to improve plant recognition performance, Zhang [3] used the genetic algorithm and correlation-based feature selection method to identify apple leaf disease, and Zhu [4] combined the cosine theorem with the K-means algorithm to classify plant leaves. Although these methods have certain recognition effects, they also have the shortcomings of lower timeliness and lower recognition rate, and their overall performance of plant leaves recognition is not high.

In order to improve the overall performance of plant leaves recognition, this paper considers the shape features and the geometric features of plant leaves as recognition features, employs the PCA algorithm to reduce the dimensions of the characteristic parameters, and adopts the AdaBoost.M1 algorithm to train and recognize plant leaves.

2. Preprocessing and Initial Feature Extraction of Plant Leaf Images

The color, shape, texture, and other basic features of plant leaves can be used as an important basis for plant leaves

* Corresponding author.

E-mail address: zhuhaodong80@163.com

recognition. However, the extraction of these features is easily influenced by the photographic distances, illumination, shadows, and other external factors when the plant leaf images are obtained. Therefore, we need to preprocess the collected plant leaf images to facilitate subsequent feature extraction.

2.1. Image Preprocessing

The plant leaf images used in this paper are from the plant leaf image database formed by the Institute of Intelligent Machinery of Chinese Academy of Sciences (<http://www.intelengine.cn/data>). The plant leaf image database contains 16846 plant leaf images of 220 kinds of plants. Here, we use the leaf image of one kind of plant as an example, and the preprocessing processes are shown from Figures 1 to 4: (1) Convert the original color image to grayscale image. (2) Adopt the OTUS algorithm to generate the binarization image. (3) Use the Laplace operator extract the edge of the binarization image.



Figure 1. The original image



Figure 2. The grayscale image



Figure 3. The binarization image

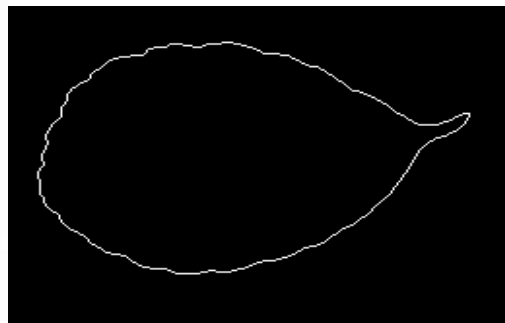


Figure 4. The profile image

2.2. Initial Feature Extraction

After acquiring the profile images, we can get the regional shape features of the plant leaf [5], such as its area, its perimeter, its length of the minimum external moment, and its width of the minimum external moment. However, these features are not steady and often change with the changes of the plant leaf, and they cannot objectively describe the type of the plant leaf. Therefore, based on regional shape features, we further calculate the relevant geometric shape features [6] with the characteristics of rotation invariance, proportion invariance, and translation invariance, such as eccentricity ratio, shape parameter, sphericity ratio, circular degree, width degree, and aspect ratio. In addition, the image moment also has the characteristics of rotation invariance, proportion invariance, and translation invariance; here, we also use seven Hu invariant moments [7] as the features of plant leaf.

2.2.1. Hu Invariant Moments

The concept of moments comes from the field of mechanical design, proposed by Hu [8]. In image processing, moments can be used to describe the features of images, which play an important role in image processing and pattern recognition. Using the nonlinear combination of geometric moments, Hu gave seven invariant moments with translation invariance, rotation invariance, and scale invariance. Invariant moments can be used as a stable feature to describe the target area in image processing. Therefore, invariant moments are also used as an identification feature to identify plant leaves.

In the process of plant leaf recognition, we add seven Hu invariant moments to the recognition features. There are many calculation methods for Hu invariant moments. The invariant moments used in this paper are the improved moment

algorithm proposed by Chen. The specific algorithm is described as follows:

First, Chen's $p+q$ stage improvement moments are defined as follows:

$$M_{pq} = \int_c x^p y^q |ds| \quad (1)$$

In Equation (1), $p, q = 1, 2, 3, \dots, n$, \int_c is the integral of the line along the closed contour, and $ds = \sqrt{(dx)^2 + (dy)^2}$.

In practical application, we use Equation (2) to approximate M_{pq} :

$$M_{pq} = \sum_{(x,y) \in C} x^p y^q \quad (2)$$

Similarly, the centre distance μ_{pq} can be calculated approximately by Equations (3) and (4) as follows:

$$\mu_{pq} = \sum_{(x,y) \in C} (x - \bar{x})^p (y - \bar{y})^q \quad (3)$$

$$\bar{x} = \frac{M_{10}}{M_{00}}, \quad \bar{y} = \frac{M_{01}}{M_{00}} \quad (4)$$

In order to obtain scale invariance, we regularize μ_{pq} as follows:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad (5)$$

In Equation (5), the regularization factor γ can be calculated as $\gamma = p + q + 1$.

Using η_{pq} , we can obtain seven Hu invariant moments by means of Equations (6) to (12).

$$\phi_1 = \eta_{20} + \eta_{02} \quad (6)$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \quad (7)$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (\eta_{03} - 3\eta_{21})^2 \quad (8)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{03} + \eta_{21})^2 \quad (9)$$

$$\begin{aligned} \phi_5 = & (3\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ & + [(3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \end{aligned} \quad (10)$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (11)$$

$$e = \sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i) \quad (12)$$

2.2.2. Tamura Texture Features

As one of the three underlying features of the image, texture features not only have nothing to do with the colours or

brightness of the image, but also contain the arrangement and order information of the surface organization structure of the object in the image, which reflects the connection between the contextual content in the image. They can respond effectively to the repeated information in the object. Therefore, texture features are an important visual feature that has been widely used in image classification and retrieval.

Based on the psychological sensory analysis of textures by people, Tamura et al. proposed a method of quantifying texture features with six parameters [9-11]. Among these six parameters, roughness degree, contrast degree, and direction degree are the most important. In this paper, these three parameters are added to the recognition features. The calculation of these three parameters in this paper is as follows:

As the most basic feature in texture, the roughness degree represents the graininess in the texture. When the window in the texture mode is larger, it will make people feel that the image looks rougher. The following are the specific calculation steps for roughness degree:

Firstly, the size of the active window is set to $2^k \times 2^k$. The average gray value of the pixels in the window of the target image is calculated as follows:

$$A_k(x, y) = \sum_{i=x-2^{k-1}}^{x+2^{k-1}} \sum_{j=y-2^{k-1}}^{y+2^{k-1}} f(i, j) / 2^{2k} \quad (13)$$

In Equation (13), $k=0,1,\dots,5$ and $f(i, j)$ is the gray value of the pixel at coordinate (i, j) .

Secondly, in the horizontal and vertical directions, the difference of average gray level between disjoint windows is calculated separately. The calculation formula is shown as Equation (14):

$$\begin{aligned} E_{k,h}(x, y) &= |A_k(x+2^{k-1}, y) - A_k(x-2^{k-1}, y)| \\ E_{k,v}(x, y) &= |A_k(x, y+2^{k-1}) - A_k(x, y-2^{k-1})| \end{aligned} \quad (14)$$

The optimum size formula is $S_{best}(i, j) = 2^k$. If the current K value is taken, the difference E can reach the maximum. $S_{best}(i, j) = 2^k$ is the best size.

Finally, the roughness degree can be obtained by calculating the average of the best windows for each pixel location in the image:

$$F_{crs} = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n S_{best}(i, j) \quad (15)$$

In Equation (15), m represents the length of the image and n represents the width of the image.

The contrast degree is obtained by counting the distribution of the gray values of the target area. Its definition formula is $\alpha_4 = \mu_4 / \sigma^4$. Here, μ_4 represents the fourth moment and σ^2 represents the variance. It can be calculated by the following Equation (16):

$$F_{con} = \frac{\sigma}{\alpha_4^{1/4}} \quad (16)$$

F_{con} gives a global measure of the contrast degree of the whole image or region.

Different texture images have different directivity. For the texture of an image, its distribution or concentration in some directions can be expressed by direction degree. The calculation steps are as follows:

First, the gradient vector of the current pixel is calculated. Equation (17) can be used to express its modulus and direction:

$$\begin{aligned} |\Delta G| &= (|\Delta_H| + |\Delta_V|)/2 \\ \theta &= \tan^{-1}(\Delta_V/\Delta_H) + \pi/2 \end{aligned}$$

(17)

In Equation (17), Δ_H and Δ_V are obtained by convoluting the image with the following two 3×3 operators:

$$\begin{bmatrix} -1 & 0 & 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & -1 & -1 & -1 \end{bmatrix}$$

Secondly, the histogram of θ is obtained by Equation (18):

$$H_D(k) = N_\theta(k) / \sum_{i=0}^{n-1} N_\theta(i)$$

(18)

Among them, n is the quantization level of the directional angle and t is the threshold. $N_\theta(k)$ is the number of pixels when $|\Delta G| \geq t$, $(2k-1)\pi/2n \leq \theta \leq (2k+1)\pi/2n$. When the histogram H_D is relatively flat, the directionality of the current image is not very prominent. If H_D contains obvious peaks, the orientation of texture in the image is strong.

Finally, Equation (19) is used to calculate the direction degree:

$$F_{dir} = \sum_p^{n_p} \sum_{\phi \in w_p} (\phi - \phi_p)^2 H_D(\phi)$$

(19)

n_p represents the number of peaks in histogram H_D . p represents the peak of histogram H_D . For any peak p , w_p represents how many regions can reach p . ϕ_p is the peak centre of the largest histogram value in w_p .

2.2.3. Initial Feature Extraction Results

Here, we use the leaf images of four kinds of plants as an example. The extraction results of feature parameters are shown in Table 1.

Table 1. The results of feature extraction

Leaf type \ Feature parameters	Osmanthus fragrans leaf	Papaya leaf	Ginkgo leaf	Red maple leaf
Eccentricity ratio	3.557	1.869	6.489	1.308
Shape parameter	0.420	0.508	0.280	0.550
Sphericity ratio	0.233	0.357	0.135	0.476
Circular degree	0.063	0.133	0.037	0.215
Width degree	0.646	0.581	0.701	0.610
Aspect ratio	3.870	2.372	6.844	1.449
Hu-1	3.108213e-001	1.939770e-001	5.318941e-001	1.717795e-001
Hu-2	7.037294e-002	1.158463e-002	2.572693e-001	2.033385e-003
Hu-3	3.088964e-004	2.419828e-005	7.248066e-004	2.004446e-004
Hu-4	3.917290e-005	9.865724e-007	6.190927e-004	1.910993e-005
Hu-5	3.543746e-010	2.065179e-012	4.146983e-007	1.182654e-009
Hu-6	-3.021417e-006	-6.483320e-008	3.123133e-004	8.559427e-007
Hu-7	4.294488e-009	4.355636e-012	-3.128954e-009	-1.343189e-011
Roughness degree	7.193	16.902	23.495	25.788
Contrast degree	0.331	0.761	0.854	0.498
Direction degree	0.058	0.095	0.076	0.013

3. Feature Dimension Reduction by PCA

PCA (Principal Component Analysis) is a multivariate statistical analysis method that can analyse the main influencing factors from multiple objects. As a typical feature dimension reduction method, its aim is to reduce the redundant information and reserve the highlight meaningful information. The obtained principal components C_i ($i = 1, \dots, n$) should have the following properties:

- (1) The principal components are not related to each other, that is, for any i and j , the related coefficient $Corr(C_i, C_j) = 0$, $i \neq j$, $i = 1, \dots, n$, $j = 1, \dots, n$.
- (2) The vector constituted by the combined coefficient $(\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{in})$ should be a unit vector.
- (3) The variance of the principal components is decreasing in turn, that is: $Var(C_1) \geq Var(C_2) \geq \dots \geq Var(C_n)$.
- (4) The total variance of the principal components is equal to the total variance of the original data: $Var(C_1) + \dots + Var(C_n) = Var(x_1) + \dots + Var(x_p) = P$, which indicates the principal component is a linear combination of the original variables, neither increasing nor decreasing the amount of information.

Considering the numbers of plant leaf images and the numbers of features are extensive in this paper, this causes the numbers and the dimensions of feature vectors in the feature space to be relatively higher. Therefore, we need to reduce the dimensions of the original feature data, and the processes are described based on PCA as follows:

- (1) Normalize the original plant leaf feature vector sets to get the covariance matrix of these vector sets:

$S = \frac{1}{m} \sum_{i=1}^m (x_i - \mu)(x_i - \mu)^T$. Here, m represents the total number of plant leaf images, the dimension of each plant leaf feature vector is n , x_i is the feature vector of the i^{th} training sample, and $\mu = \frac{1}{m} \sum_{i=1}^m x_i$ represents the mean vector of the training sample set.

- (2) Arrange the values of these features in descending order: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r$, and their corresponding feature vectors η_1, \dots, η_r ($r \leq n$) form a feature subspace G .
- (3) Project each plant leaf image x_i into this subspace G , $\partial_i = G_i(x_i - \eta)$, $i = 1, 2, \dots, r$, and ∂_i represents the location of the sample x_i in the subspace, which is the main feature of x_i and can replace x_i input to the AdaBoost.M1 algorithm.

4. Plant Leaves Recognition by AdaBoost.M1 Algorithm

Freund and Schipare proposed the AdaBoost algorithm, which was originally designed for the two-classification problems in 1995 [12]. The AdaBoost.M1 is the generalization of the AdaBoost algorithm to deal with the multi-classification problems [13], and its process is as follows:

- (1) Given sample set $S = \{(x_1, y_1), \dots, (x_n, y_n)\}$, where $x_i \in X$, $y_i \in Y = \{1, 2, \dots, k\}$, and $i = 1, 2, \dots, n$, k represents the number of classes, n represents the number of samples, the weak classifier is $h_t : X \rightarrow Y$, $t = 1, 2, \dots, T$ and the number of weak classifiers is T .
- (2) Initialize the sample weights: $D_t(i) = 1/n$, $i = 1, 2, \dots, n$.
- (3) For $t = 1, 2, \dots, T$.

① Call the weak classification algorithm and adopt sample weights D_t to get the weak classifier $h_t : X \rightarrow Y$.

② Calculate the weighted error $\varepsilon_t = \sum_{i=1}^n D_t(i)[y_i \neq h_t(x_i)]$.

③ If $\varepsilon_t > 1/2$, then $T = t - 1$, break; otherwise, go to ④.

④ Let $\beta_t = \varepsilon_t / (1 - \varepsilon_t)$, $\alpha_t = \ln(1 / \beta_t)$.

⑤ Let $D_{t+1}(i) = D_t(i) \beta_t^{1 - [h_t(x_i) \neq y_i]}$ and $i = 1, 2, \dots, n$. Here, Z_i is the normalized factor, and the final strong

classifier is $H(x) = \operatorname{argmax}_{y \in \{1, \dots, cn\}} \sum_{t=1}^T \alpha_t [h_t(x) = y]$, where “[]” is defined as follows: for the logical expression π , if π is true, then $[\pi] = 1$; otherwise, $[\pi] = 0$.

The AdaBoost.M1 function used in this paper can be expressed as: function $[trerr, tserr, w] = \text{adaboost.M1}(type, fea_tr, lab_tr, fea_ts, lab_ts, M, cn)$, where “*type*” is the type of weak classifier, *fea_tr* and *lab_tr* represent the training samples and the training sample labels respectively, *fea_ts* and *lab_ts* represent the testing samples and the testing sample labels, *M* is the number of cycles, *cn* means the number of classes, *w* represents the weight of each weak classifier, *trerr* is the train error rate, and *tserr* represents the test error rate.

5. Proposed Plant Leaves Recognition Model and Simulation Experiments

In this paper, we adopt PCA+AdaBoost.M1 to recognize the plant leaves. We firstly preprocess the plant leaf images to get the original features and then employ PCA to extract representative features. Finally, we use the AdaBoost.M1 to recognize the plant leaves. The basic process is shown in Figure 5.

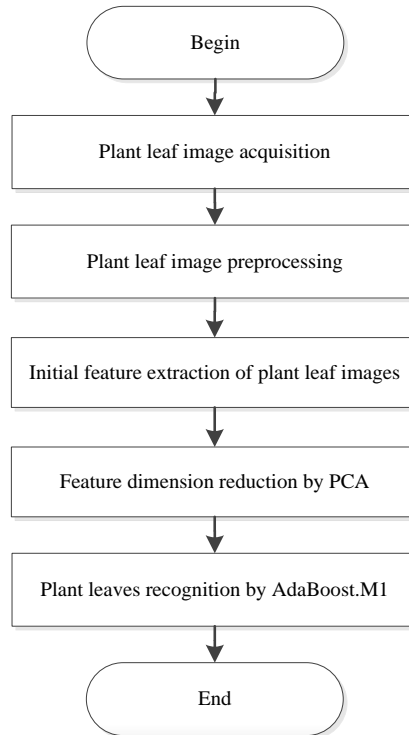


Figure 5. The basic process of the proposed plant leaves recognition method

We select osmanthus fragrans leaf images, papaya leaf images, ginkgo leaf images, and red maple leaf images as the experimental image data set. Each type of plant leaf images contains 60 samples, and the capacity of all samples is 240.

The experiment parameters are set as follows: the number of cycles “*M*” is set to 50, the weak classifier of AdaBoost.M1 is the Bayes classifier, and the number of classes “*cn*” is 4.

Firstly, we extract 13 features of these plant leaves. The magnitude of each parameter is diversity, which will affect the effectiveness and accuracy of the recognition results; therefore, these features are normalized. Then, these normalized features are reduced in dimensions by PCA, and finally the AdaBoost.M1 is adopted to train and test these plant leaf images that are reduced in dimensions. In order to verify the validity of the proposed method, we also utilize AdaBoost.M1 to train and test the same original data that has not been reduced in dimensions and employ 1-NN classifier to train and test the same original data that has been reduced in dimensions. In order to facilitate comparison, the recognition rate (RT) is set as a percentage, and the running time is set as a positive number greater than 0. Experimental results are shown below in Figure 6.

Since the principal component is a linear combination of original variables, the reduced dimension data does not reduce the information of the original data, so it does not affect the recognition rate of the plant leaf images. As can be seen from Figure 6, compared with the AdaBoost.M1, the recognition rate of AdaBoost.M1+PCA increases by about one percentage point, while the running time decreases by one order of magnitude. This shows that PCA is an effective method to reduce the feature dimensions of plant leaf images. We also adopt PCA+1 NN to identify the same plant leaf images. The recognition rate is 5.2 percentage points lower than that of PCA+AdaBoost.M1, and their running time is on the same order of magnitude.

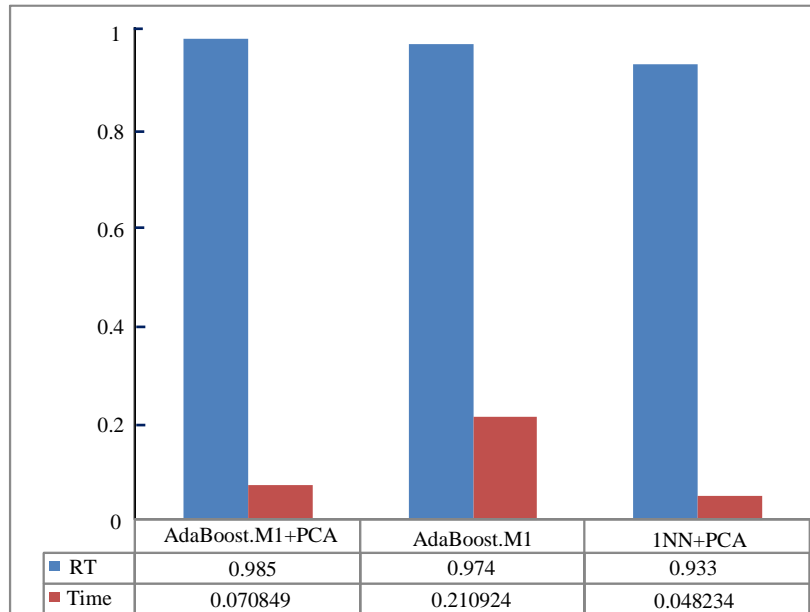


Figure 6. Recognition results of three methods

To sum up, the proposed method can effectively reduce the identification time of plant leaf images on the basis of guarantee recognition rate, and it has better overall recognition performance.

6. Conclusions

This paper proposed a plant leaves recognition method combined PCA with AdaBoost.M1, which uses PCA to reduce the dimensions of plant leaf features and AdaBoost.M1 to train and test the plant leaf images in lower data space. The experiment results show that, compared with the AdaBoost.M1 and PCA+INN, the proposed plant leaves recognition method has a higher recognition rate and better timeliness, and it can provide an effective method for the recognition of plant leaf species.

Acknowledgements

The authors would like to thank the editors and the anonymous reviewers for their helpful comments and suggestions, which have improved the presentation. This work was supported in part by the Key Science Research Project of Colleges and Universities in Henan Province of China (No. 19A520009) and the National Science Foundation of China (No. 81501548.)

References

1. Z. B. Wang, H. L. Li, Y. Zhu, et al., "Review of Plant Identification based on Image Processing," *Archives of Computational Methods in Engineering*, Vol. 24, No. 3, pp. 637-654, 2017
2. S. H. Lee, C. S. Chan, S. J. Mayo, et al., "How Deep Learning Extracts and Learns Leaf Features for Plant Classification," *Pattern Recognition*, Vol. 71, pp. 1-13, 2017
3. C. L. Zhang, S. W. Zhang, J. C. Yang, et al., "Apple Leaf Disease Identification using Genetic Algorithm and Correlation based Feature Selection Method," *International Journal of Agricultural and Biological Engineering*, Vol. 10, No. 2, pp. 73-84, 2017
4. H. D. Zhu and Z. Shen, "Plant Leaf Identification Method based on Cosine Theorem and K-Means," *Journal of Huazhong Normal University*, Vol. 48, No. 5, pp. 650-655, 2014
5. H. D. Zhu, H. C. Li, D. Wu, et al., "Image Texture Feature Extraction based on Cloud Platform and New ImageClass," *Journal*

- of *Computational Information Systems*, Vol. 12, No. 17, pp. 6311-6321, 2015
6. H. D. Zhu, D. Wu, H. C. Li, et al., "Plant Leaves Extraction Method under Complex Background based on Closed-Form Matting Algorithm," *Journal of Computational Information Systems*, Vol. 11, No. 21, pp. 7633-7640, 2015
 7. H. X. Kan, L. Jin, and F. L. Zhou, "Classification of Medicinal Plant Leaf Image based on Multi-Feature Extraction," *Pattern Recognition and Image Analysis*, Vol. 27, No. 3, pp. 581-587, 2017
 8. M. K. Hu, "Visual Pattern Recognition by Moment Invariants," *IRE Transaction Information Theory*, Vol. 8, No. 2, pp. 179-187, 1962
 9. H. Tamura and S. Mori, "Textural Features Corresponding to Visual Perception," *IEEE Transactions on Systems, Man, & Cybernetics*, Vol. 8, No. 6, pp. 460-473, 1978
 10. X. Q. Li, Q. Shao, and J. J. Wang, "Classification of Tongue Coating using Gabor and Tamura Features on Unbalanced Data Set," in *Proceedings of 2013 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, pp. 108-109, USA, New York, 2013
 11. S. Y. Liao and T. Q. Huang, "Video Copy-Move Forgery Detection and Localization based on Tamura Texture Features," *2013 6th International Congress on Image and Signal Processing (CISP)*, IEEE, pp. 864-868, USA, New York, 2013
 12. C. S. Lv, T. Zhang, and C. Lin, "Face Detection based on Skin Color and AdaBoost Algorithm," in *Proceedings of the 2017 29th Chinese Control and Decision Conference (CCDC)*, pp. 1363-1367, Chongqing, China, May 2017
 13. A. Minz and C. Mahobiya, "MR Image Classification using Adaboost for Brain Tumor Type," in *Proceedings of the 2017 IEEE 7th International Advance Computing Conference (IACC)*, pp. 701-705, Hyderabad, India, Jan 2017

Hui Chen received her B.S. degree from Henan University, Kaifeng, Henan Province, China, in 1992. Since 1993, she has been with the faculty of the Engineering Training Centre, Zhengzhou University of Light Industry, Zhengzhou, Henan Province, China, where she is currently an experimentalist. Her major research interests include Intelligence Information Processing and Data Mining.

Haodong Zhu received his B.S. degree from Lanzhou Jiaotong University in 2004, his M.S. degree from Sichuan University of Science & Engineering in 2008, and his Ph.D. from Graduate University of Chinese Academy of Sciences in 2011. As a postdoctoral researcher, he engaged in image big data processing in the Postdoctoral Mobile Station of Computer Science and Technology at Tongji University from 2014 to 2016. As a visiting scholar, he engaged in micro-blog big data processing at Griffith University from 2017 to 2018. Since 2010, he has been an associate professor and master's tutor in the School of Computer and Communication Engineering at Zhengzhou University of Light Industry. His major research interests include cloud computation, intelligence information processing, computing intelligence, and data mining.

Xufeng Chai graduated from Henan University in 2013. At present, he is a faculty member in the Sida Foreign Language Primary School at Henan Experimental High School. His major research interest is mathematics education.