

Karhunen-Loeve Transform and Sparse Representation Based Plant Leaf Disease Recognition

Jie Tian^{1,2}, Qiu-xia Hu², Xiao-yi Ma^{*1,3}

¹College of Mechanical and Electronic Engineering, Northwest A&F University, Shaanxi, China

²College of Information Engineering, Northwest A&F University, Shaanxi, China

³College of Water Resources and Architectural Engineering, Northwest A&F University, Shaanxi, China

*Corresponding author, e-mail: xiaoyimasl@yahoo.com.cn

Abstract

To improve the classification accuracy rate of apple leaf disease images and solve the problem of dimension redundancy in feature extraction, Karhunen-Loeve (K-L) transform and sparse representation are applied to apple leaf disease recognition. Firstly 9 color features and 8 texture features of disease leaf images are extracted and taken as feature vectors after dimensionality reduction by the K-L transform. Then, for each of apple mosaic virus, apple rust and apple alternaria leaf spot, 40 apple leaf images are selected as the training samples, whose feature vectors are made up of the dictionary of the sparse representation, respectively. Each testing sample is classified into the class with the minimal residual. The identifying results using the proposed method are analyzed and compared with those of the Support Vector Machine (SVM) and original sparse representation method. The average classification accuracy rate of the proposed method is 94.18%, which confirms its good robustness. In addition, the proposed method not only improves the plant leaf disease classification accuracy but also solves the redundancy problem of the extracted features.

Keywords: plant disease, disease recognition, K-L transform, sparse representation

Copyright © 2014 Institute of Advanced Engineering and Science. All rights reserved.

1. Introduction

Plant diseases cause major production and economic losses in agriculture. The spread of plant diseases drives the development of machine learning methods to identify these diseases [1-2]. It is necessary to help farmers or producers identify early symptoms of plant diseases by analyzing the digital images. Thus the applied artificial neural networks and SVM (support vector machine, SVM) techniques [1-2], [4-5] for image recognition are applied in plant diseases. Compared to neural network, SVM find not the local minimum value, and SVM has been better applied in disease recognition [4]. However, in plant disease identification study, we usually rely on artificial selection of SVM parameters on classification of the samples. As a result, the selection of SVM model parameters is random, and the application of SVM is not easy to find suitable parameters.

In recent years, the technology of sparse representation has achieved significant success in computer vision and is an ℓ^1 minimization-based optimization method. It is applied to classification and has better classification performance than SVM [6-8]. But these classification methods based on sparse representation exist two limitations [6-7]: One limitation is the dimension redundancy of feature extraction. Constructing the sparse dictionary for the classification, using all the features will increase the calculation of the dictionary. The other limitation is not removing the correlation among the characteristics. Generally speaking, all kinds of features especially color features have a lot of relevancies that may decline the classifier performance.

This paper presents a K-L transform-based sparse representation plant disease classification method. Firstly color features of the first-order moment, second-order moment and third-order moment, texture features of co-occurrence matrix are acquired for the apple leaf diseases [9]. The sparse representation classification method is used in [6-7], without dimensionality reduction. In this paper the K-L transform-based feature selection is performed to reduce the redundancy of the extracted features and the correlation among the features about the images. And the extracted features are used as inputs to sparse representation model. The

results of this study demonstrate that the K-L transform -based sparse representation classification method has higher performance than the state-of-the-art methods.

2. Proposed K-L Transform y Plant Leaf Disease Recognition Based on Sparse Representation

This study implements a machine vision system for the automatic classification of the visual symptoms of plant leaf diseases, through analyzing the colored images. Diseased regions are pre-processed based on wavelet filtering, and image disease features are extracted.

After the image features are extracted, the sparse representation classification method is used to identify the plant diseases. And the key idea of sparse representation classification method is a judicious choice of dictionary representing the test sample as a sparse linear combination of the training samples themselves. The K-L transform-based feature extraction reduces feature dimension redundancy so as to reduce the running time of the dictionary. At the same time, it removes the correlation among the characteristics and improves the apple leaf diseases recognition rate.

2.1. Images Acquisition of Apple Leaf Diseases

The images set of apple leaves used in this study is obtained from the experimental base of Horticulture College, at the Northwest A & F University of Yang ling, China. We take these pictures by using Canon IXUS850 IS digital camera in natural lighting conditions. The used image format of apple mosaic virus, rust and leaf spot is JPEG. And the images sizes are 256×256 pixel.

2.2. Image Preprocessing of Apple Leaf Diseases

This study can be affected by the image acquisition environment and collecting device as well as other factors, which often reduces the plant disease images' quality. So in order to remove the noise of the images effectively, the study adopts the contrast test on median filtering and wavelet filtering [10]. The results of filtering for apple rust image as an example is shown in Figure 1.

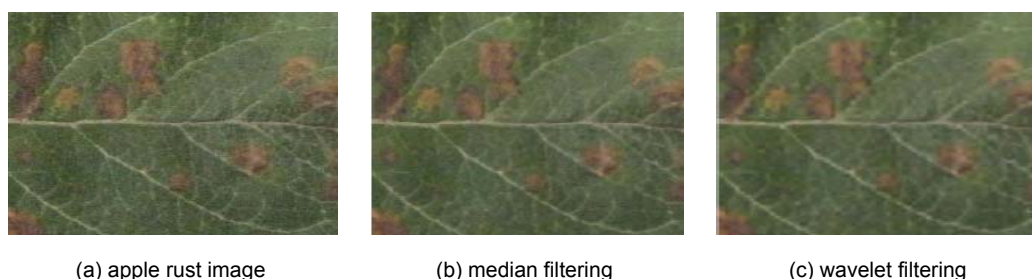


Figure 1. Results of Filtering for Color Images of Apple Rust

2.3. Feature Extraction

Different plant diseases' image color, texture features are different, so the extracted color and texture features of apple diseases leaf image are used as the disease eigenvectors. K-L transform-based feature selection is performed to identify which of these provided most information about the image domain.

The color matrix of first-order, second-order and third-order moments, among other existent techniques, is used to calculate image color. Ultimately, 9 color features are computed according to HSV pixel values. And the extracted texture features of co-occurrence matrix have 4 key characteristics: entropy, energy, moment of inertia and correlation [9]. Finally, 8 texture features are gained.

Let $v_j^{(i)} \in R^m (j = 1, 2, \dots, ni)$ be the sample of m-dimensional color and texture feature. Then denote by matrix $m \times ni$ the data set of the samples matrix of ni disease training images from the i -class, and the number of extracted disease characteristics m equals 17.

2.4. The Sparse Representation of Apple Leaf Disease Images

Given the n_i training samples from the i -th class as columns of a matrix $v^{(i)} = [v_1^{(i)}, \dots, v_{n_i}^{(i)}] \in R^{m \times n_i}$, and $D^{(i)}$ is its corresponding dictionary. A test sample $y \in R^m$ from the same class will approximately lie in the linear span of the training samples associated with $v^{(i)}$:

$$y = v^{(i)} x^{(i)} \quad (1)$$

Where $x^{(i)} = [x_1^{(i)}, x_2^{(i)}, \dots, x_{n_i}^{(i)}]^T \in R^{n_i}$ is a coefficient vector and $x_j^{(i)}$ is the sparse coefficient of y in $D^{(i)}$ for the j -th sample of the i -th class.

We define a new matrix D for the entire training set as the concatenation of the n training samples of K object classes, for $n = n_1 + n_2 + \dots + n_K$:

$$D = [D^{(1)} \ D^{(2)} \ \dots \ D^{(K)}] \quad (2)$$

Then the linear representation of y can be rewritten in terms of all training samples as:

$$y = Dx \quad (3)$$

Where $x = [x^{(1)}; \dots; x^{(i)}; \dots; x^{(K)}] = [0, \dots, 0, x_1^{(i)}, x_2^{(i)}, \dots, x_{n_i}^{(i)}, 0, \dots, 0]^T$ is a coefficient vector whose entries are zero except those associated with the i -th class.

As the entries of the vector x encode the identity of y , it is tempting to attempt to obtain it by solving the Equation (3) $y = Dx$. This solution is resolved by choosing the minimum ℓ^1 norm:

$$\min \|x\|_1 \quad \text{s.t.} \quad y = Dx \quad (4)$$

Since real plant disease images are noisy, it may not be possible to express the test sample exactly as a sparse superposition of the training samples. The sparse solution x can still be approximately recovered by solving the following ℓ^1 -minimization problem:

$$\min \|x\|_1 \quad \text{s.t.} \quad \|Dx - y\|_2 \leq \varepsilon \quad (5)$$

Where ε is the error tolerance.

Suppose \bar{x} is the approximate solution for x , and α_i is a coefficient vector of \bar{x} associated with the i -th class. And the nonzero entries in $\alpha_i(\bar{x})$ will all be associated with the columns of \bar{x} from a single object class i .

For each class i , let $\alpha_i: R^m \rightarrow R^{n_i}$ be the characteristic function which selects the coefficients associated with the i -th class. Used only the coefficients associated with class i , the given y as $\bar{y} = D\alpha_i(\bar{x})$ can be approximated. Then we denote the residual y from the i -th class training samples as follows:

$$res_i(y) = \|y - D\alpha_i(\bar{x})\|_2 \quad (6)$$

This simple idea of apple leaf disease images classification is that if y belongs to the i -th class training samples, then we classify y based on these approximations by assigning it to class i that minimizes the residual between y and \bar{y} .

2.5. The Sparse Representation of Apple Disease Images Based on K-L Transform

Based on the color and texture features, the K-L transform-based feature extraction are utilized. The mean of samples is given by:

$$mean = \frac{1}{ni} \sum_{j=1}^{ni} v_j^{(i)} \quad (7)$$

A new set of samples named newX is gained by its mean subtracted from each sample value. Then the covariance matrix Σ is calculated as follows:

$$\Sigma = \frac{1}{ni-1} \sum_{j=1}^{ni} newX^T newX \quad (8)$$

The covariance matrix Σ can be decomposed into its corresponding eigenvalues and eigenvectors:

$$\sum U = \lambda U, \quad U = [U_1, U_2, \dots, U_{ni}], \quad \lambda = diag(\lambda_1, \lambda_2, \dots, \lambda_{ni}) \quad (9)$$

Where λ_j are the eigenvalues of Σ and U_j are the corresponding eigenvectors of λ_j .

Let $p_j(1), p_j(2), \dots, p_j(ni)$ be the arrangement of $\lambda_{j,p_i(1)} \geq \lambda_{j,p_i(2)} \geq \dots \geq \lambda_{j,p_i(ni)}$.

Then select the largest ρ eigenvalues and the corresponding converted eigenvectors. Finally the eigenvectors of the apple disease images can be written as:

$$\chi = (v_{j,\rho}^{(i)})^T \in R^\rho, j = 1, \dots, ni. \quad (10)$$

At last the new dictionary of sets of training samples and test samples are gained by $\chi(y)$ and $X(D)$ replacing y and D in the Equation (5) and (6), respectively:

$$\chi(y) = X(D^{(1)})x^{(1)} + X(D^{(2)})x^{(2)} + \dots + X(D^{(K)})x^{(K)} = X(D)x \quad (11)$$

Where $X(D) = [X(D^{(1)}) X(D^{(2)}) \dots X(D^{(K)})]$, $X(D^{(i)}) = [\chi(v_1^i), \dots, \chi(v_{ni}^i)]$ is the dictionary.

2.6. Apple Leaf Disease Recognition Algorithm Based on Sparse Representation

After the K-L transform-based image features are extracted, the new dictionary $X(D)$ is gained and $res_i(\chi(y))$ is calculated to get the classification results. The algorithm of the K-L transform-based based on sparse representation is as follows:

Step 1: Extract the eigenvalues of apple disease images and obtain the corresponding eigenvectors $v_j^{(i)}$.

Step 2: Calculate the dictionary $X(D)$ according to equation (11). Then ρ is initialized and ε is set to be 0.001 for the test sample $\chi(y)$.

Step 3: Solve the ℓ^1 -minimization problem $\|X(D)x - \chi(y)\|_2 \leq \varepsilon$.

Step 4: Compute the residuals $res_i(\chi(y)) = \|\chi(y) - X(D)\alpha_i(\bar{x})\|_2$. And the identity result equals $\arg \min_i res_i(\chi(y))$.

3. Research Method

Select 120 images of apple leaf disease (Fig.2) as the training set, where there are 40 mosaic virus, rust and leaf spot apple leaf images, respectively. Select another 128 image as

the test set at the same time, where mosaic virus, rust and leaf spot include 39, 41 and 48 images, respectively. Experimental computer CPU is Intel (R) Core (TM) 2, 1.00GB memory, and the operating system is Windows XP. Use the Matlab7.0 and VC++6.0 languages to implement apple leaf diseases images classification.

Figure 3 shows an example to classify the apple mosaic virus from the apple leaf diseases by the proposed method. The test sample y is from the class of apple mosaic virus and No. of mosaic virus, rust and leaf spot disease are from 1 to 40, 41 to 80, 81 to 120 in the dictionary of training samples. Then the sparse coefficient x and the test sample residual $res_i(\chi(y))$ are computed.

From Figure 3(a) below, the result shows that y has the maximum coefficient on the disease class of apple mosaic virus and has only one nonzero entry on other disease classes, which signifies x is sparse. According to Figure 3(b), it shows that the apple mosaic virus minimizes the residual, then we classify y by assigning it to the class of the apple mosaic virus.

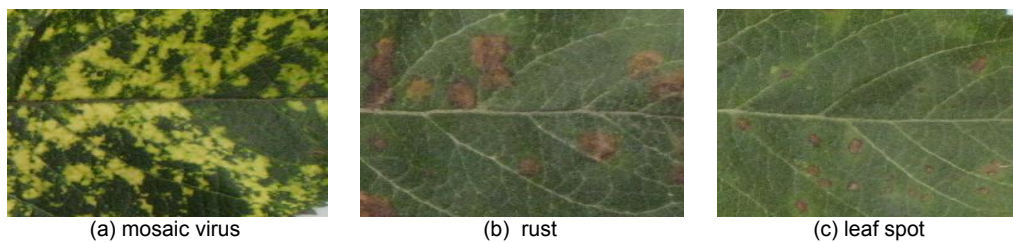


Figure 2. Apple Leaf Disease Images

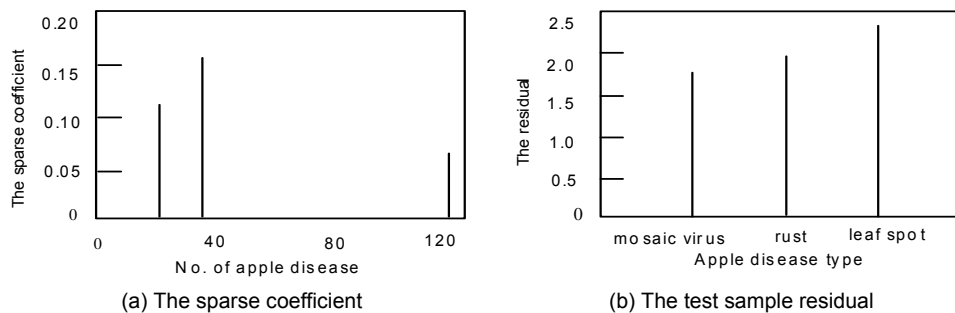


Figure 3. The Example of Apple Leaf Disease Recognition Approach Based on Sparse Representation

4. Results and Analysis

To verify validity of the K-L transform-based sparse representation classification method, we use SVM model as a counterpart algorithm [4]. By using the training samples and test samples in section 3.1, we compare SVM, sparse representation approach and K-L transform-based sparse representation approach in term of apple leaf disease recognition accuracy rate[5], as shown in Table 1.

K-L transform has the principal component number p valued 14. Selected candidate features are used as inputs to sparse representation classification model. Table 1 illustrates: (1) When selected 14 features are based on K-L transform, the proposed method has the highest classification accuracy reaching to 92.31%,90.24%,100.00% for mosaic virus, rust and leaf spot apple leaf diseases, respectively. (2) Without candidate features are selected by K-L transform, the proposed method has 27.21% average higher classification accuracy than the sparse representation method. (3) The average classification accuracy of the proposed method and the sparse representation method for plant disease classification are 35.54% and 8.33% higher than the SVM method, respectively.

The proposed classification model based on K-L transform and sparse representation has much higher classification accuracy than that based on sparse representation or SVM. This

may because that K-L transform calculates principal components of the original feature vector inputs and linearly transforms the high dimensionality input vector into a low-dimensionality one.

Furthermore, the plant leaf disease recognition based on sparse representation shows higher classification accuracy than SVM, as discussed in [7].

Table 1. Experimental Results of Apple Disease with the Classification Method Based on SVM and Sparse Representation

disease type	SVM		sparse representation		K-L transform + sparse representation	
	recognition number	recognition accuracy/%	recognition number	recognition accuracy/%	recognition number	recognition accuracy/%
mosaic virus	21	63.64	26	66.67	36	92.31
rust	17	41.46	26	63.41	37	90.24
leaf spot	34	70.83	34	70.83	48	100.00

5. Conclusion

The identification of the symptoms of plant diseases by means of K-L transform-based sparse representation method may support farmers during their daily struggle against disease outbreaks. Considering the early period symptoms of the plant diseases using hyperspectral images and better classification algorithms implementation is our future study.

Extract the color features and co-occurrence matrix texture features of apple diseases images. K-L transform is utilized to reduce the feature redundancy as well as remove the correlation among the characteristics.

In order to reduce the calculation of the dictionary, the K-L transform-based sparse representation classification method is used to identify the plant diseases. The experimental results show that when selected features number p valued 14 are based on K-L transform, the classification accuracy reaches to 92.31%, 90.24%, 100.00% for mosaic virus, rust and leaf spot apple leaf diseases, respectively. The proposed method has higher classification accuracy than the sparse representation-based and SVM.

Acknowledgements

This research was sponsored by National Natural Science Foundation of China (No. 51279167, No. 61001100).

References

- [1] Patil JK, Kumar R. Advances in image processing for detection of plant diseases. *Journal of Advanced Bioinformatics Applications and Research*. 2011; 2(2): 135-141.
- [2] Sankarana S, Mishraa A, Ehsani R, et al. A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture*. 2010; 72: 1-13.
- [3] Pang JL, Zhang CL, Pan W, et al. Automated identification of tomatoes with diseases using artificial neural network trained with genetic algorithms. *Transactions of the Chinese Society of Agricultural Engineering*. 2004; 20(3): 113-116.
- [4] Tian YW, Li TL, Li CH, et al. Method for recognition of grape disease based on support vector machine. *Transactions of the CSAE*. 2007; 23(6): 175-180.
- [5] Huang YB, Lan Y, Thomson SJ, et al. Development of soft computing and applications in agricultural and biological engineering. *Computers and Electronics in Agriculture*. 2010; 71: 107-127.
- [6] Han AT, Guo XH, Liao Z, et al. Classification of agricultural pests based on compressed sensing theory. *Transactions of the Chinese Society of Agricultural Engineering*. 2011; 27(6): 203-207.
- [7] Yang SQ, Ning JF, He DJ. Identification of varieties of rice based on sparse representation. *Transactions of the Chinese Society of Agricultural Engineering*. 2011; 27(3): 191-195.
- [8] Wright J, Yang AY, Ganesh A, et al. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2008; 31(2): 210-227.
- [9] Xu K, Li Y. An image search approach based on local main color feature and texture feature. *Journal of Xi'an Shiyou University (Natural Science Edition)*. 2005; 20(2): 77-79.
- [10] Chen WF. Wavelet Analysis and Its Application in Image Processing. Beijing: Science Press. 2002: 156-160.