

A New Algorithm for Feature Extraction of Plant Diseases and Insect Pests Signal Based on Optimal Wavelet Packet and Non-negative Matrix Factorization

Changcheng Li^{1, a}, Shuyun Cai^{2, b*} and Laiwu Yin^{2, c}

¹College of Electrical and Information Engineering, Jilin Agricultural Science and Technology College, Jilin, 132101, China

²College of mathematics and statistics, Beihua University, Jilin, 132101, China

³Educational technology and Information Center Jilin Agricultural Science and Technology College, Jilin, 132101, China

^ajilinlcc@sina.com, ^b756401998@qq.com, ^cyinlaiwu@163.com

Keywords: Plant diseases and insect pests, feature extraction, optimal wavelet packet basis, non-negative matrix factorization, algorithm optimization

Abstract. Plant diseases and insect pests have similar symptoms, but it is difficult to distinguish between professional and technical personnel to identify plant diseases and insect pests. In order to accurate extraction of plant diseases and insect pests, physiological and pathological characteristics of signal, puts forward a based on lifting wavelet transform feature extraction algorithm optimization scheme, for the study of plant diseases and insect pests damage signal showing the effect of the prior farmers identify any disease, choose the correct method of governance, quickly make the right decision, improve farmers plant diseases and insect pests, harm signal feature extraction and recognition level. The simulation results show that this algorithm can be used to optimize the stability and convergence, and can be used as an ideal plant disease and insect pests signal feature extraction optimization algorithm, which can effectively identify the different plant diseases and insect pests.

Introduction

It is generally believed that the key technology issues of Agricultural Internet of things for radio frequency induction, in fact, the final solution is the Agricultural Internet of things is the identification of plant potential. The compression algorithm is proposed on the basis of the optimal wavelet packet after analyzing the features of data redundancy and visual redundancy existed in the plant pathology images [1]. The method is presented to use the Gabor wavelet for testing the verge of plant root image [2]. The spectrum noise falls into the scope of multiplicative noise, and the method of combining the logarithmic transformation and wavelet transform is put forward on the basis of the theoretical deduction [3]. The wavelet soft-threshold de-noising method and the fast Fourier transform (FFT) are used to study the basic features and variation of the electrical signal of aloe leaves under different temperatures [4]. The experiment system of the virtual instrument for collecting the plant electrical signals is designed by using the technology of wavelet soft-threshold de-noising method [5]. The continuous Morlet wavelet transform is applied to analyze the variations between the different temperatures, speeds, and humidity data in and out of the corn canopy. The results show that the wavelet transform presents the periodic components included in the variations of micrometeorological factors in and out of the plant canopy [6]. In order to discriminate and recognize the phytoplanktons from different categories, the second-order low-frequency component of Daubechies wavelet is used to extract the discrete spectrum features of ten phytoplanktons [7]. The method of using the multi-scale wavelet transform is presented to test the verge of plant root image [8]. By using non negative matrix factorization based best wavelet packet basis of plant diseases and insect pests damage potential signal registration and recognition, image spatial domain to find a parent subspace W, which form the subspace based potential signal pixels are coincided with the algorithm implementation of plant diseases and insect pests damage potential of feature extraction.

According to the feature extraction and classification of the plant diseases and insect pests potential signal, it can achieve the normal effect of the plant diseases and insect pests, and realize the auxiliary diagnosis of plant diseases and insect pests.

The Principal Theory of Wavelet Packet

The width of time-frequency window of the basic function of wavelet transform decreases with the decrease of the scale of J , and the width of frequency-domain window will increase accordingly, that is, the domain window of the correspondent wavelet basis function increases with the decrease of the scale.

The distribution (see fig.1) of small scale against large frequency window and big scale against narrow frequency window is in accordance with the time-frequency characteristics of signals in the natural world, and suitable for analyzing the signals with random scale. However, many problems, in reality, mainly focus on the interested signals at the some certain phases (points) of time or frequency domain, and only the information at the certain point of time and frequency will be necessary for the extraction, so the maximum possible improvement of the resolution ratio of the time-frequency at the interested frequency point, or of the time at the interested time point, should be the aim. And the fixed distribution of time-frequency window of the orthogonal wavelet transform is not the optimal choice under this circumstance, the main reason is that the multi-resolution decomposition of orthogonal wavelet transform only decomposes the V (scale) space, that is, $V_0 = V_1 \oplus W_1 = V_2 \oplus W_2 \oplus W_j = \dots$, but not further decomposes the W (wavelet) space, and they can be further decomposed with wavelet packet to narrow the frequency spectrum window, which widens with the increase of j in the process of the orthogonal wavelet transform. And the most suitable time-frequency window or optimal basis can be obtained to analyze the signals [9].

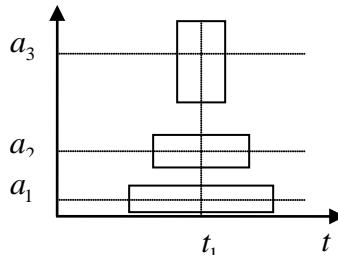


Fig.1 Time-frequency window of wavelet transform

The two-scale relation between the given orthogonal scaling function $\phi(t)$ and wavelet function $\psi(t)$ is:

$$\phi(t) = \sqrt{2} \sum_k h_{0k} \phi(2t - k) \quad (1)$$

The h_{0k} in the expression is the filter coefficients in the analysis of multi-resolution ratio, and the following recursive relations are defined to further expand the two-scale equation:

$$W_{2n}(t) = \sqrt{2} \sum_{k \in Z} h_{0k} W_n(2t - k) \quad (2)$$

If $n=0$, then $W_0(t) = \phi(t)$, $W_1(t) = \psi(t)$. The above defined function set $\{w_n(t)\}_n \in Z$ is the wavelet packet determined by $W_0(t) = \phi(t)$. And the wavelet packet $\{w_n(t)\}_n \in Z$ is a function set containing scaling function $W_0(t)$ and mother wavelet function $W_1(t)$ with certain connection.

The Decomposition of Wavelet Packet Subspace

The decomposition of wavelet packet can be conducted to obtain the subspace composed with scaling functions and the subspace with wavelet functions. According to the expressions (1) and (2), $W_0(t)$ is the scaling function, when $n=0$, and $W_0(t)$ is the wavelet function; when $n=1$, the subspace composed

with scaling functions is $\{V_j\}_{j \in Z}$, and the subspace composed with wavelet functions is $\{W_j\}_{j \in Z}$. The following signs are introduced for the discussion of the space composed with wavelet packets:

$$\begin{cases} U_j^0 = V_j & j \in Z \\ U_j^1 = W_j & j \in Z \end{cases} \quad \square \square \square \quad (3)$$

$V_j = V_{j+1} \oplus W_{j+1}$ can be obtained from the wavelet multi-resolution ratio analysis, which is presented with the expression (3):

$$U_j^0 = U_{j+1}^0 \oplus U_{j+1}^1 \quad j \in Z \quad \square \quad (4)$$

When the expression (4) is expanded to the wavelet packet, and

$$U_j^n = U_{j+1}^{2n} \oplus U_{j+1}^{2n+1} \quad j \in Z, n \in Z^+ \quad \square \quad \square \square \square$$

The expression (5) is the orthogonal decomposition of the subscript n of the expression (3) and expression (4). U_{j+1}^{2n} and U_{j+1}^{2n+1} are the subspaces of U_j^n with U_{j+1}^{2n} against w_{2n} and U_{j+1}^{2n+1} against w_{2n+1} . $L^2(R) = \bigoplus_{j \in Z} W_j$ in the multi-resolution ratio analysis. W_j^n is expressed as U_j^n for the convenience of comparison, and the wavelet packet space can be decomposed as:

$$W_j^n = U_j^n = U_{j+1}^{2n} + U_{j+1}^{2n+1} \quad \square \quad \square \quad (5)$$

From the above expression, the wavelet packet decomposition improves the resolution ratio of the high-frequency components of signals.

The Selection of the Optimal Wavelet Packet Basis

An optimal wavelet packet basis can be selected according to the different features of signals for analysis to present the signal characteristics. The wavelet packet decomposition of signal $f(t)$ is to project $f(t)$ on the wavelet packet basis with the aim of achieving a series of coefficients $d_I^{j,n}$, which are used to describe the features of the signal $f(t)$, and it is better if the difference between the coefficients is larger, because if only a small number of coefficients $d_I^{j,n}$ are large, then these coefficients will represent the features of $f(t)$, and this kind of wavelet packet basis is evidently better. If the difference between this series of coefficients is not large, it will be difficult to find out the features of $f(t)$, and the correspondent basis cannot be the optimal basis. The selection of wavelet packet basis involves two problems, one is what the optimal basis is, and the other is how to select the optimal basis.

In order to describe the properties of the coefficient series $d_I^{j,n}$, the cost function of a sequence needs to be defined first, and the basis which can make the cost function minimal is to be found out among the wavelet packet bases in the wavelet library. The minimal cost represents the ultimate efficiency for a given vector, and the basis is the optimal. The cost function can be defined as any real function M about the sequence, but the additive cost function M is used most to test the concentration ratio. The concentration ratio here means that the M will be large when the coefficient is moderate, otherwise, M will be small when the coefficient difference is large; while the additivity means M is an additive information cost function, both $\lambda(x)$ and $M(x)$ are the smallest.

If cost function M has been selected, the concept of optimal basis is defined as the following:

Set $x = \{x_j\}$ as a vector in the principal space and B is an orthogonal basis selected from the library. If B_x is the coefficient with x set under the basis B , and $M(B_x)$ is the smallest for $x \in V$, then B is the optimal basis. The selection method of the optimal basis will be discussed in the following.

If the orthogonal basis library meets the following requirements. The subset composed with base vectors which equals the set of non-negative integer N has the intervals as the following

$$I_{nK} = [2^K n, 2^K (n+1)] \quad K \in Z, n \in N \quad (6)$$

Every basis of the library corresponds to a non-intersect overlap composed with I_n , k of N .

If $V_{n,k}$ equals $V_{n,k}$, then $V_{n,k} = V_{2n,k+1} \oplus V_{2n+1,k+1}$, and this orthogonal basis library has a binary tree structure.

If the library is a tree, then the optimal basis can be found through the induction of k . $B_{n,k}$ is the basis corresponding to the vector $I_{n,k}$, and $A_{n,k}$ is the optimal basis of span B_n which constrains x . A single basis exists against $k=0$, that is, $I_{n,0}$ is the optimal basis, and $A_{n,0}=B_{n,0}$, then the construction of $A_{n,k}$ with $n \geq 0$ are as the following:

$$A_{n,k} = \begin{cases} B_{n,k} & M(B_{n,k}(x) < (A_{2n,k+1}(x)) + M(A_{2n+1,k+1}(x))) \\ A_{2n,k+1} \oplus A_{2n+1,k+1} & other \end{cases} \quad (7)$$

The expression (13) is the optimal orthogonal basis which generates x for the cost function M .

In the real application, only the frequency-limited decomposition can be implemented on VN. If three-layer decomposition is implemented on VN, the coefficient of signal function $f(x)$ in each subspace can be calculated through the wavelet packet algorithm, then the cost function value of the coefficient on each layer can be generated through $M(x)$, and the following steps are implemented to select the optimal basis.

Signal Characteristics of Plant Diseases and Insect Pests in Non-negative Matrix Factorization

In the higher dimensional space, the distribution of plant pathology image is not very contact, which is not good for the classification, and the calculation complexity is very high. Therefore, the efficient method has to be taken for the dimension reduction of image, that is, features extraction. The subspace analysis is generally applied to the image recognition as an algorithm of features extraction, which is characterized by the low cost of calculation and high capacity of description. The principal idea is to find out a linear or non-linear subspace according to a certain performance target to compress the original image into a low dimension subspace with the aim of making the data in the subspace more contact with better extraction of the feature information contained in data.

The following two steps are used to extract the plant pathology features through the subspace analysis.

The training sample are studied to achieve the subspace composed with base vectors and projected in the subspace to obtain the projection coefficients stored in the sample feature library in advance.

The sample to be recognized is projected in the subspace to achieve the features of the sample to be recognized.

This paper puts forward an analysis algorithm of the non-negative matrix factorization for feature extraction in comparisons with the principal component analysis and the linear discriminant analysis, and ten images are selected from the plant pathology image library for the training and testing.

As the pixels of the base vectors FCA and LDA can be positive or negative values, both methods lacks the effects of composing the parts into the whole in the visual sense, and this paper proposes NMF as a new algorithm for feature extraction.

The training sample

$$V = \{x_i\}_{i=1}^n \quad (8)$$

is composed with n plant pathology images (x_i is an m dimension column vector, which is composed with non-negative gray values of a plant pathology image), and it is decomposed into a product of a non-negative $m \times r$ dimension matrix W and a non-negative $r \times n$ dimension matrix H through NMF.

Feature Extraction of Plant Diseases and Insect Pest Potential Signal

This paper intends to use in Changbai Mountain Area of genuine ginseng disease and insect damage potential signal as the object, can sample to get a smaller error from the limited training samples set, ginseng disease and insect pests damage potential signal samples and the number of support vector function is very good solve because of the sample quantity and restrict the correct rate of the question [10]. Were selected 15 cases of ginseng Rhizoctonia, 15 cases of ginseng blight, 15 cases of ginseng

Aphrophora Inmedia uhler and 15 cases of ginseng fruit borer a total of 60 patients with sample data, 60 x 4 sample data set, this group of data normalization, take 40 of them were as training samples, and 20 groups as test samples, training sample set input SVM training, the training model after the completion of the input test set. Simulation experimental results show that the Changbai Mountain Area of genuine ginseng disease and insect pests damage potential normal signal and pest damage potential signal classification accuracy reached in%, ginseng disease and insect pests damage potential signal energy threshold as shown in Fig. 2.

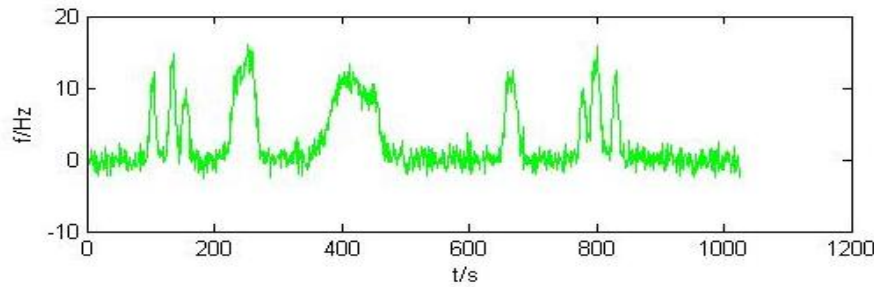


Fig. 2 the energy threshold of the potential signal of ginseng diseases and insect pests

Figure 2 shows that the use of ginseng disease and insect pests signal potential energy threshold extraction time waveform and the method on the signal to noise ratio (SNR) relatively small pest damage potential signal extraction time, threshold can inhibit the influence of background noise, the optimal wavelet packet based and non-negative matrix decomposition algorithm to extract the time limit than the data such as table 1 shows.

Table 1 Optimal extraction time ratio of wavelet packet based algorithm

| Potential signal | Frequency(beats · min ⁻¹) | Cycle(t/s) | Optimal wavelet packet based time entropy |
|-------------------------------|---------------------------------------|------------|---|
| Ginseng normal signal | 76 | 0.7557 | 0.0576 |
| Ginseng blight signal | 73 | 0.6312 | 0.0491 |
| Ginseng blight signal | 71 | 0.6125 | 0.0378 |
| Ginseng Liu froghopper signal | 68 | 0.6218 | 0.0617 |
| Ginseng fruit borer signal | 67 | 0.6913 | 0.0721 |

Extraction of ginseng pest potential signal optimal wavelet packet basis time entropy is used as feature parameters and its potential signal segment after the success of the obtained frequency and cycle characteristic parameters. Through comparison of different ginseng disease and insect pests damage potential energy of the signal and time domain characteristics and improved the best wavelet packet basis and non-negative matrix factorization algorithm of frequency and cycle characteristic parameters.

Summary

The traditional Fourier transform of the plant diseases and insect pests damage potential signal diagnosis methods compared and used in this study based on the optimal wavelet packet based and non-negative matrix decomposition algorithm in terms of time complexity, occupied system memory or efficiency is superior to the Fourier transform algorithm. This study presents the optimal wavelet packet basis with try to non-negative matrix factorization algorithm for feature extraction pretreatment effectively filter out the noise and signal potential energy threshold extraction time is to avoid plant diseases and insect pests of potential signal to noise and noise further interference. Simulation experiment results show that the algorithm can effectively identify the normal signal and pest signal of different, than the traditional algorithm improves the accuracy rate of 5.7%, for plant diseases and insect pest damage potential signal diagnosis provided supplementary means.

Acknowledgement

This research is funded by science and technology project of Jilin Province Department of Education. (Project No: Ji Jiao Ke characters [2014] No. 382nd, Ji Jiao Ke characters [2011] No. 152nd, Ji Jiao Ke characters [2016] No. 202nd)

References

- [1] W. Song and Z. Wand, "Application of Wavelet Theory in the Edge Detection of Plant Roots Image," J. Development & Innovation of Machinery & Electrical Products, vol. 23, no. 1, pp. 135-137, January 2010.
- [2] F. Zhang, L. Wang, R. Su, Z. Song, X. Wang and C. Zhu, "Research on Wavelet Analysis in the Characteristics Extracting and Identification of Discrete 3D Fluorescence Spectra of Phytoplankton," J. Chinese Journal of Sensors and Actuators, vol. 20, no. 10, pp. 2143-2150, Oct. 2007.
- [3] Y. Wang, R. Liu and P. Zhou, "Wavelet domain semi-fragile water marking scheme for plant disease image authentication," J. Journal of Zhongkai University of Agriculture and Technology, vol. 21, no. 3, pp. 45-49, May 2008.
- [4] Z. Jiang, Q. Yu, T. Wang and X. Sun, "The Applicationo of Wavelet Transform in Meteorological Study within Plant Canopy," J. Chinese Journal of Agrometeorology, vol. 25, no. 1, pp. 62-65, February 2004.
- [5] W. Chen, S. Weng, L. Wang, C. Chu and H. Liu, "Design of the measurement system for week electrical signal based of virtual instrument and wavelet de-noising," J. Chinese Agricultural Mechanization, no. 5, pp. 94-97, May 2010.
- [6] X. Zhang, N. Yu, G. Xi and X. Meng, "The Analysis on Spectrum Characteristic of Plant Electrical Signal Based on Wavelet De-Noising," J. Journal of Xi'an University of Technology, vol. 27, no. 4, pp. 411-416, June 2011.
- [7] G. Zhou, C. Wang, F. Yang and Y. Li, "Field Collected Plant Spectrum Denoising By Logarithm Transform and Wavelet Transform," J. J. Infrared Millim. Waves, vol. 28, no. 4, pp. 316-320, Aug. 2009.
- [8] W. Song, L. Wang and Y. Cao, "Studying the Edge Detection of Images of Plant Roots Based on Gabor Wavelet Theory," J. Process Automation Instrumentation, vol. 32, no. 3, pp. 24-25, March 2011.
- [9] C. C. Li, L. W. Yin, D. Chen and S. J. Xu, "Lossless Compression of Weak Electrical Signal of Ginseng Molecule Based on Discrete Wavelet Transform and Siesta Program", Advanced Materials Research, Vols.986-987, pp.1950-1953, 2014.
- [10]C. C. Li, L. W. Yin and D. Chen, "Study on Plant Pest Images Identification Based on Lifting Wavelet Transform", Applied Mechanics and Materials, Vols.687-691, pp.3640-3643, NOV. 27 2014.