Journal of GeoEngineering, Vol. 13, No. 2, pp. 69-77, June 2018
http://dx.doi.org/10.6310/jog.201806_13(2).3

PILE DEFECT IDENTIFICATION BASED ON MULTI-HIGHER ORDER MOMENT FEATURE

Weixin Kang [1], Yuchen Zhao [2*], and Lingxi Liu [3]

ABSTRACT

This article is mainly devoted to pile defect identification and classification. The main objective of the work is to improve the pile features. Accordingly, a novel feature based on higher order statistical moments is proposed in this paper. The stress wave reflected signals are possessed by wavelet packet transform method. Three higher order statistical parameters—variance, skewness, and kurtosis are calculated in each wavelet packet band of the decomposed signals. The sliding window method is proposed to extract characteristics in every time interval. In particular, the Principal Component Analysis (PCA) is used to reduce dimension of the merged feature and eliminate the relevance among them. Then the features are fed into Support Vector Machine (SVM). Compared with three other existing features and the 42-dimensional feature, the multi-higher order moment feature achieves the highest classification accuracy which reaches 98%. The simulation results demonstrate that the proposed feature can be used as a suitable tool for pile defect detection. It is simple and effective.

Key words: Pile foundation, defect identification, higher order moment, feature extraction, PCA, classification.

1. INTRODUCTION

There is no denying that a firm pile foundation is the essential base of buildings. Pile foundation supports the building and guarantees the users' safety. However, in recent years, the defective pile rate has already reached 20% in China. In order to improve the quality of project, these defective piles should be identified and classified.

At present, the low strain integrity testing method is commonly used in pile quality detection. Several researchers have investigated the stress wave reflected signals by pattern recognition, mathematical theories, computer technology and other methods. During the whole process, feature extraction is a key step. A good and sensitive feature affects the sequent defect diagnosis as well as the discrimination accuracy directly. It not only can compress data but also remain main information of original signals. In previous studies, most of the features were power spectral density (Cai *et al*. 2002), wavelet coefficients (Pan *et al*. 2006), and variance (Kang *et al.* 2010). There is only one feature about entropy (Kang *et al.* 2015) is proposed in recent years. Apparently, these features are barren and not enough to reveal characteristics of original signals. Therefore there are a lot of useful information that need to be mined in the stress wave reflected signals. A novel feature has yet to be found.

To address this issue, the purpose of this paper is extracting more appropriate features. Due to the fact that the reflected signals are of nonlinearity and non-Guassianness, we extracted higher order statistics, which are very useful in problems where either non-Gaussianity, or nonlinearities, and have good ability to overcome the noise. Then, the non-Guassianness features are available, and a lot of useful information in the reflected signals can be extracted (Mendel and Jerry 1991).

Manuscript received October 12, 2017; revised February 28, 2018; accepted March 1, 2018.
[1] Professor, Harbin Engineering University, Harbin, China.
[2] Graduate student (corresponding author), Harbin Engineering University, Harbin, China (e-mail: trustzhaoyuchen@163.com).
[3] Graduate student, Harbin Engineering University, Harbin, China.

First, the wavelet packet is applied to process the stress wave reflected signals. Then, the higher order moment features at different wavelet bands are extracted by the sliding window method. Since the number of feature vectors are excessive, PCA is used to reduce dimension of the features. The combination of variance, skewness, and kurtosis are fed into the SVM for classification.

An outline of this paper is as follows: Section 2 proposes algorithms including the wavelet packet transform and higher order moment algorithm. Section 3 describes the procedure of feature extraction, dimension reduction by PCA and discusses defect classification by SVM. Section 4 gives experimental results, the performance comparisons of conventional features, and the proposed feature. Conclusion is presented at last.

2. METHODOLOGY

2.1 Wavelet Packet Transform

Wavelet is called "numerical microscrope" in signal and image processing (Morlet *et al.* 1982). In recent years, the wavelet analysis has been applied in many fields, such as signal denoising, image compression, feature extraction, and so on. The representation of a set of time-dependent data on a wavelet basis leads to a special structure of raw data that is localized in the frequency and time domains at the same time (Li and Chen 2014). Because of these advantages, the pile reflected signals which are extremely nonlinear and non-stationary can be processed by wavelet. As the extension of wavelet theory, the wavelet packet transform (WPT) inherited wavelet bases, and it incorporates linear combinations of usual wavelet functions as its bases (Ogden 1997; Percival and Walden 2000). The wavelet packet transform not only decomposes the low-frequency band but also decomposes the high-frequency band so that the detail information in high-frequency band will not be omitted. Its algorithms are as follows.

Assume that U represents a set, $U_n(j)$ is the closure of $U_n(t)$, $U_{2n}(j)$ is the closure of $U_{2n}(t)$, and $U_n(t)$ satisfies the relations as

follows:

$$U_{2n}(t) = \sqrt{2}\sum_{k\in z} h(k)\, U_n(2t-k) \tag{1}$$

$$U_{2n+1}(t) = \sqrt{2}\sum_{k\in z} g(k)\, U_n(2t-k) \tag{2}$$

The scaling function relates to the low pass filters with $h(k)$, and the wavelet function relates to the high pass filters with $g(k)$ (Scholkopf and Smola 2001). When $n = 0$, $U_0(t) = \phi(t)$; when $n = 1$, $U_1(t) = \psi(t)$. They are orthogonal scaling functions, and wavelet functions, respectively.

$$\phi(t) = \sqrt{2}\sum_{k\in z} h(k)\, \phi(t)(2t-k) \tag{3}$$

$$\psi(t) = \sqrt{2}\sum_{k\in z} g(k)\, \psi(t)(2t-k) \tag{4}$$

2.2 Higher Order Statistics

In reality, many processes are not Gaussian, such as the pile reflected signals in this study. Higher order statistics are good at dealing with non-Gaussian or possible nonlinear processes (Mendel and Jerry 1991). We used these moments partially because they are more intuitive than cumulants (Gelb and Andrew 2010). In this paper, we extracted higher order moments at different wavelet packet bands, including variance, skewness, and kurtosis. These higher order moments can reveal the distribution of each wavelet packet band. It can extract and seek out hidden information contain in the signals using higher order statistics parameters which are excellent in many respects, particularly in feature extraction for classification (Karaye *et al.* 2014).

The third order central moment is skewness. Skewness is a statistic which describes the asymmetricity of data distribution. When the data is right-skewed distributed, the skewness is greater than zero, and when the data is left-skewed distributed, the skewness is less than zero.

Peakness of a distribution is measured by the fourth order central moment (Kaur *et al.* 2016). Generally, the greater kurtosis is, the more data concentrate, and the distribution curve shows a peak shape. The smaller kurtosis is, the more data disperse, and the distribution curve is flat.

For a signal x, the variance, skewness, and kurtosis are the following expressions. Among them, μ is the mean, σ is the standard deviation, and E represents to calculate the mean.

$$\text{variance} = E\,[(x-\mu)^2] \tag{5}$$

$$\text{skewness} = \frac{E\,[(x-\mu)^3]}{\sigma^3} \tag{6}$$

$$\text{kurtosis} = \frac{E\,[(x-\mu)^4]}{\sigma^4} \tag{7}$$

3. SIMULATION AND EXPERIMENT

In this study, we did simulations by using Matlab, including stress wave signals acquisition, feature extraction, recognition and classification and so on. The stress wave reflected signal data were obtained by finite difference numerical simulation method. This method is based on one-dimensional wave theory of pile-soil system. The wave equation is established, combined with the boundary conditions and initial conditions. Then, obtain results by finite differential method, and curves of velocity response on the pile top is obtained. Moreover, it has been verified that the data obtained by this method are effective and accurate (Zhou 2010). The variables in the simulation were pile length, pile diameter, elastic wave speed, elastic modulus of the pile, and defect location. The duration of simulation of each sample was 14.57 ms.

First, the signals were decomposed by wavelet packet to the third level, as a result, each signal was processed into 8 independent frequency bands, as shown in Fig. 1. These decomposed signals in different frequency bands have different characteristics, and each of them is very useful for fault diagnosis. So, we analyzed the signals in these bands. In this study, the sampling frequency of stress wave reflected signal is f_s = 70281 Hz. When the signal is decomposed by wavelet packet to the n-th level, the bandwidth f_{BW} of the n-th level can be expressed as Eq. (8).

$$f_{BW} = \frac{f_s}{2^{n+1}} \tag{8}$$

so that the frequency range of different bands in the n-th level is as follows:

$$mf_{BW} \sim (m+1)\, f_{BW}\ ,\ \ m = 0, 1, 2, \ldots, 2^n - 1 \tag{9}$$

where m is the ordinal of frequency band in the n-th level.

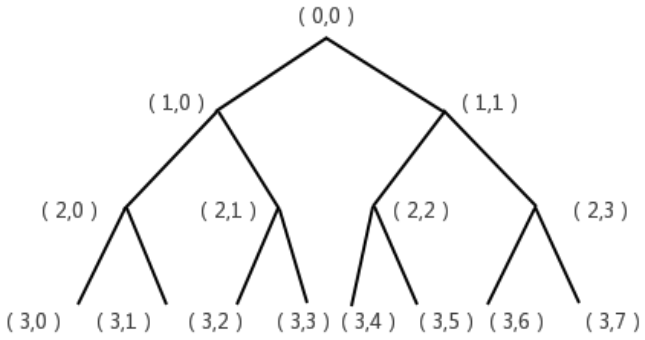


Fig. 1 3-level wavelet packet tree

Wavelet packet decompositions of an integrated pile and a shrinking pile are shown in Figs. 2 and 3. The distribution of 8 different wavelet packet bands are given as follows. Their dimensions are shown in Figs. 4 and 5; where L represents the pile length, and D represents the pile diameter. There is a shrinking defect at L_2 meters from the pile top in Fig. 5; the diameter of defect is D_2 = 0.6 m, and the defect length is L_3 = 0.5 m.

It is known from Fig. 3 that there is a peak at about Point 54 in the first frequency band, which should be caused by shrinking defects. However, Fig. 2 of integrated pile is flat at that point. Whichever the band shows, their distribution and symmetry are different. Variance represents the degree of dispersion. Skewness can describe the asymmetricity of data distribution. Kurtosis

measures sharp parts of a distribution. Apparently, the higher order moments are different, too. Therefore, the features have abilities to distinguish different types of piles and describe properties of original signals. Figure 6 shows the flow diagram of the pile defect identification procedure.

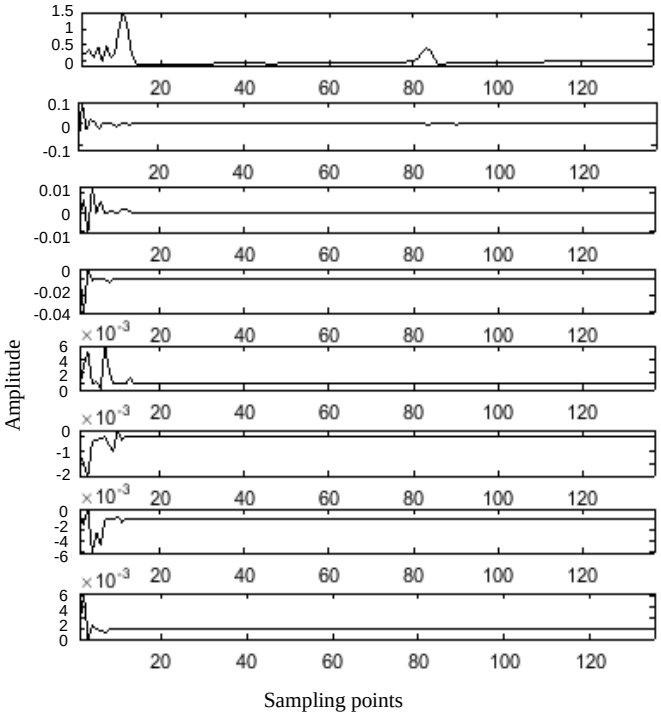


Fig. 2 Wavelet packet decomposition of integrated pile

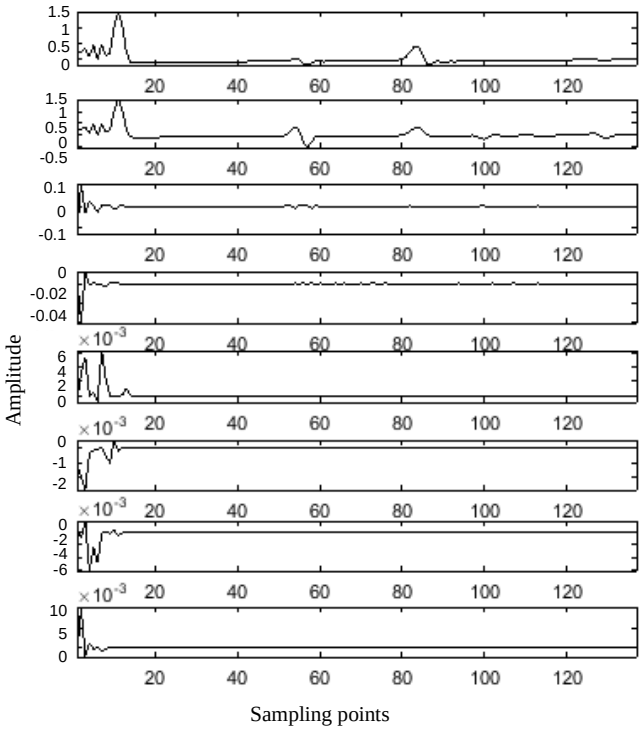


Fig. 3 Wavelet packet decomposition of shrinking pile

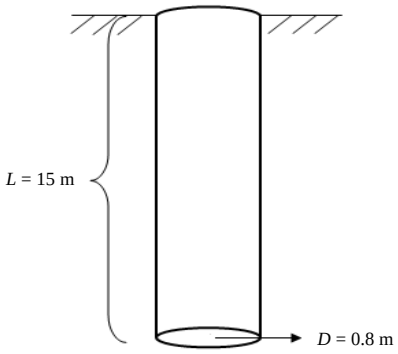


Fig. 4 A integrated pile diagram

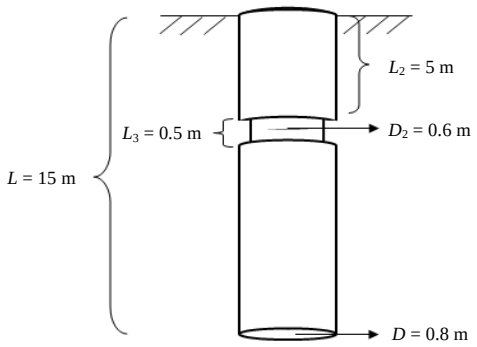


Fig. 5 A shrinking pile diagram

Fig. 6 Diagram of the pile defect identification

3.1 The Sliding Window Method

In the feature extraction stage, we applied Db7 wavelet, which has orthogonality and is compactly supported (Liu *et al.* 2010). The pile reflected signals were decomposed to the third level, so that there were 8 independent frequency bands. If we only extract one feature vector in each frequency band, then the characteristics in time-domain will not be obtained. Apparently, it will waste valuable information provided by signals. To solve this problem, this study proposed the sliding window method. The sliding window can pass through the whole time-domain of signals, and it catches nuance in every time interval. This method can integrate information in both time and frequency domains, and it gives a more detailed description of pile reflected signals.

This approach depends on the window width and step length. In general, the window width is neither too wide nor too narrow. A wide window will hold redundant information, while a narrow window may omit useful characteristics. In this paper, the sample data were obtained by the finite difference numerical simulation method. Each data had 1024 sampling points. After decomposition by wavelet packet, there were 139 sampling points in each band. It is proved through test that choosing 10 sampling points as the window width and 10 sampling points as the step length (the last window held less than 10 sampling points, then added zero for supplement) can achieve favorable effect. So that the window moved 14 steps. The formulation is shown as follows:

$$H = \begin{bmatrix} H_{11} & H_{12} & \cdots & H_{1m} \\ H_{21} & H_{22} & \cdots & H_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ H_{2^n1} & H_{2^n2} & \cdots & H_{2^nm} \end{bmatrix} \tag{8}$$

Here, H is a feature vector matrix. Each row of the matrix represents different frequency band. There were 8 bands in this study. Each column of the matrix covers the sampling points in different bands at the same time; m stands for the number of sampling points; n stands for the decomposition level. The diagram of sliding window is shown in Fig. 7. As it is shown below, the window crosses all the bands and is gradually sliding with time. When the window takes the last step, there will be 14 feature vectors.

3.2 Feature Dimension Reduction

Compared with the conventional feature extraction method, the study obtained more features through the sliding window method. As is known to all, the dimension of the features vector is influential in the performance of the classification system (Simas Filho and Seixas 2016; Pal and Foody 2010; Kesharaju and Nagarajah 2015). On one hand, it can reveal more information when the number of feature vectors increased. On the other hand, too many features will cause redundant data and the training time of classifier will extend. In order to reduce dimension of the features and eliminate the relevance among them, PCA was applied. The analyses are as follows.

Assume that an m-dimensional input vector x_i = (x_1, x_2, x_3, …, x_n). Its sample dimension is n, mean is zero. (If the mean is not equal to 0, do the following calculation to make the mean equal to 0: $x_i' = x_i - E(x_i)$) A linear transformation is expressed as：

$$y_i = U^T x_i \tag{9}$$

where U represents an $m \times m$ orthogonal matrix. The ith column of U is the covariance matrix C of x_i.

$$C = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T \tag{10}$$

Then the problem comes to solve the following eigenvalue problem:

$$\lambda_i u_i = C_i u_i , \quad i = 1, \ldots, m . \tag{11}$$

where λ_i is eigenvalue of covariance matrix, u_i is the eigenvector. Then arrange the eigenvalues from the highest to lowest in order to obtain the corresponding principal component.

$$y_i = u_i^T x_i , \quad i = 1, \ldots, m . \tag{12}$$

3.3 Feature Extraction

The procedures of higher order moment feature extraction are as follows.

Step 1: Decompose pile signals

The pile reflected signals were transformed by Db7 wavelet packet to the third level. As a result, each signal was processed into 8 independent frequency bands, we extracted features from the frequency bands of the third level.

Step 2: Feature extraction

We used the sliding window method to extract features. We chose 10 sampling points as the window width, and 10 sampling points as the step length (The last window held less than 10 sampling points, then added zero for supplement). So that the window moved 14 steps, as shown in Fig. 7. Then, we calculated the higher order moments (the second central moment, the third central moment, and the fourth central moment) of the 8 frequency bands contained in each window respectively. At last, each kind of higher order moment would have 14 features, the 42-dimensional feature was obtained.

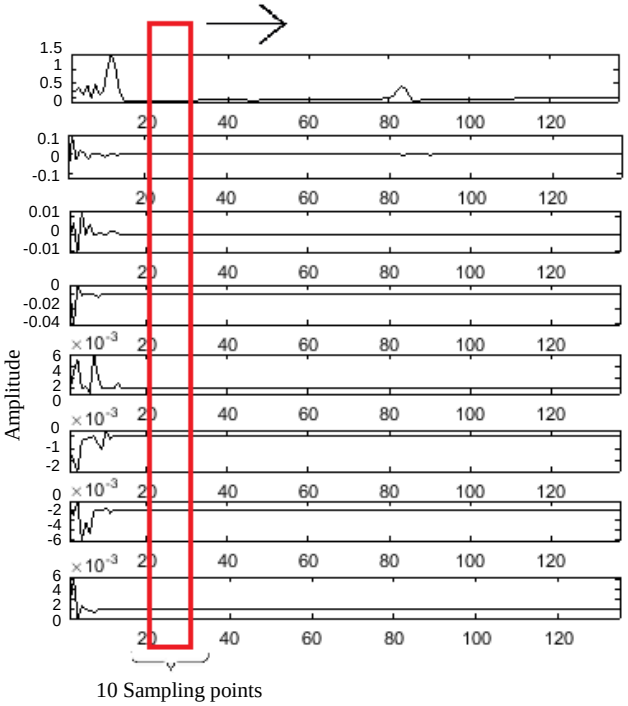


Fig. 7 Diagram of sliding window

A feature example of a shrinking pile is shown as follows. The vector F is consisted of 42 elements. The first 12 elements are the second central moment feature, the next 12 elements are the third central moment feature, and the last 12 elements are the fourth central moment feature.

F = [1.0672, 1.3594, 0.0852, 0.2210, 6.6891e05, 0.0475,
0.0018, 0.0452, 0.0702, 0.0201, 0.0265, 0.0033,
0.0029, 0.0001, 2.497, 3.5944, –0.0564, –0.0370,
–0.0219, –0.0190, –0.0115, –0.0036, 0.0605,
–0.0136, –0.0084, –0.0011, 8.4308e-05, –0.0017,
6.9896, 11.3521, 0.0446, 0.0254, 0.0126, 0.0104,
0.0054, 0.0011, 0.0490, 0.0067, 0.0035, 0.0002,
7.6219e-06, 0.0004] (13)

We plotted the feature in Fig. 8. Different kind of piles have different values of features, the following work of recognition and classification will be processed by SVM.

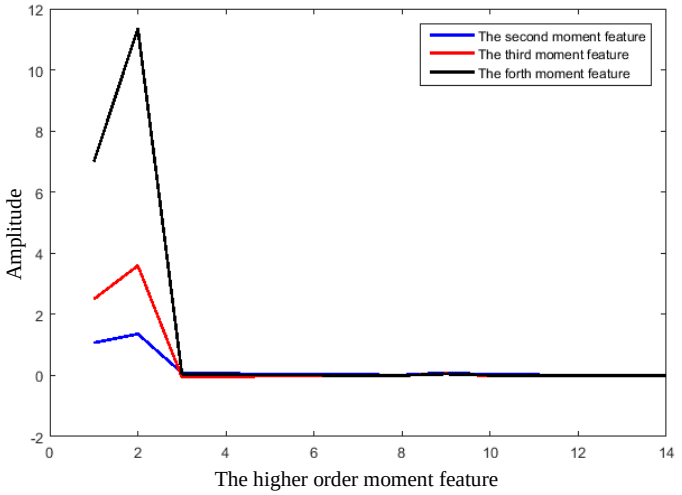


Fig. 8 The higher order moment feature of shrinking pile

For comparison, the proposed feature will be compared to conventional features including power spectrum density, variance and entropy. The feature extraction of power spectrum density, variance, and entropy are as follows.

Step 1: Decompose pile signals

The pile reflected signals were transformed by Db7 wavelet packet to the third level. Then extracted features from the 8 independent frequency bands of the third level.

Step 2: Feature extraction

To obtain the power spectrum density feature, do Fourier transform of coefficients in each frequency bands respectively. In order to extract features by sliding window conveniently, set the points of Fourier transform to 140. Then calculated the the square of amplitude, the sum of the square of the real and imaginary parts. As a result, each signal sample had an 8 × 140 power spectral density matrix. The using of sliding window method was same as the higher order moment feature extraction. At last, there were 14 power spectral density features of each sample.

To obtain the variance feature, the sliding window was set up as that in above feature extraction. Then, calculated the variance in each window. At last, there were 14 variance features of each sample.

To obtain the entropy feature, the using of sliding window method was same as the above feature extraction. In Eq. (13), the sum of sampling points in each frequency band denotes S_i, where N represents the number of frequency bands, P_i represents the ratios of sampling points in each frequency band to all the frequency bands. The entropy H is given by Eq. (14). There were 14 entropy features of each sample.

$$P_i = S_i / \sum_{i=1}^{N} S_i \tag{14}$$

$$H = -\sum_{i=1}^{N} P_i \ln P_i \tag{15}$$

3.3 Classification

We chose five kinds of pile samples for classification: Integrated pile, shrinking pile, expanding pile, segregated pile and fractured pile. Each kind of piles had 60 sets of samples. A simulated integrated pile example is as Fig. 9(a). The pile length is L = 15 m, pile diameter is D = 0.8 m. The density of pile is 2400 kg/m^3, the longitudinal wave velocity is 3676 m/s, elastic modulus of the pile is decided by Eq. (15). Shear modulus of soil around the pile-soil is 4.05 × 10^7 Pa, density of soil around the pile-soil is 1800 kg/m^3, shear modulus of soil at the bottom of a pile is 8 × 10^7 Pa, density of soil at the bottom of a pile is 2000 kg/m^3, and Poisson's ratio of soil around the pile-soil is 0.28. The excited force is 1 N·s, and duration of it is 1 × 10^{-3}s. A simulated shrinking pile example is as Fig. 9(b). Its parameters are same as the integrated pile. However, there is a shrinking defect at 5 ~ 5.5 meters from the pile top, the diameter of defect is D_2 = 0.6 m. A simulated expanding pile example is as Fig. 9(c). It is formed by the increasing of pile diameter. There is a expanding defect at 5 ~ 5.5 meters from the pile top, the diameter of defect is D_2 = 1 m. A simulated segregated pile example is as Fig. 9(d). The improper mix proportion of concrete makes the internal structure uneven. There is a segregated defect 5 ~ 5.5 meters from the pile top. The wave velocity of segregated defect is 3400 m/s. A fractured pile example is as Fig. 9(e). There are crazes in the pile, which cause fractured defect. There was a fractured defect 5 ~ 5.5 meters from the pile top.

The higher order moment features (the variance, skewness and kurtosis) were extracted. In the next stage, the features were fed into Support Vector Machine (SVM). Support Vector Machines are useful and have advantages in classification learning (Cristianini and Shawe-Taylor 2000; Vapnik 1998). For the purpose of classification, a linear or a non-linear separation surface which can divide the input samples should be obtained (Vishwanathan and Narasimha Murty 2002). The SVM can turn nonlinear problems to linear problems by increasing dimension, and then dealing with linear problems in the higher dimensional space (Kang and Zhao 2017). Besides, it proves advantages in the small statistical samples cases and nonlinear cases, so it is suitable for this subject. The core idea of SVM is the minimization of structural risk (Morlet *et al*. 1982). The basic principles of SVM can be illustrated by the two dimension situation in Fig. 10. We have a set of example data samples, which are referred as feature vectors x_i and corresponding labels y_i, i = 1, …, n, $x \in R^d$, $y_i \in \{+1, -1\}$. Classification in SVM based approaches is founded on the notion of hyperplanes (Tan *et al*. 2014). Assume that H is the classification hyperplane. $H1$, $H2$ pass through the samples, which are closest to the optimal classification hyperplane. $H1$, $H2$ parallel to H, the distance between them is classification margin. The optimal hyperplane satisfies the classification requirements, and it makes the classification margin at its maximum, lets the samples away as far as possible.

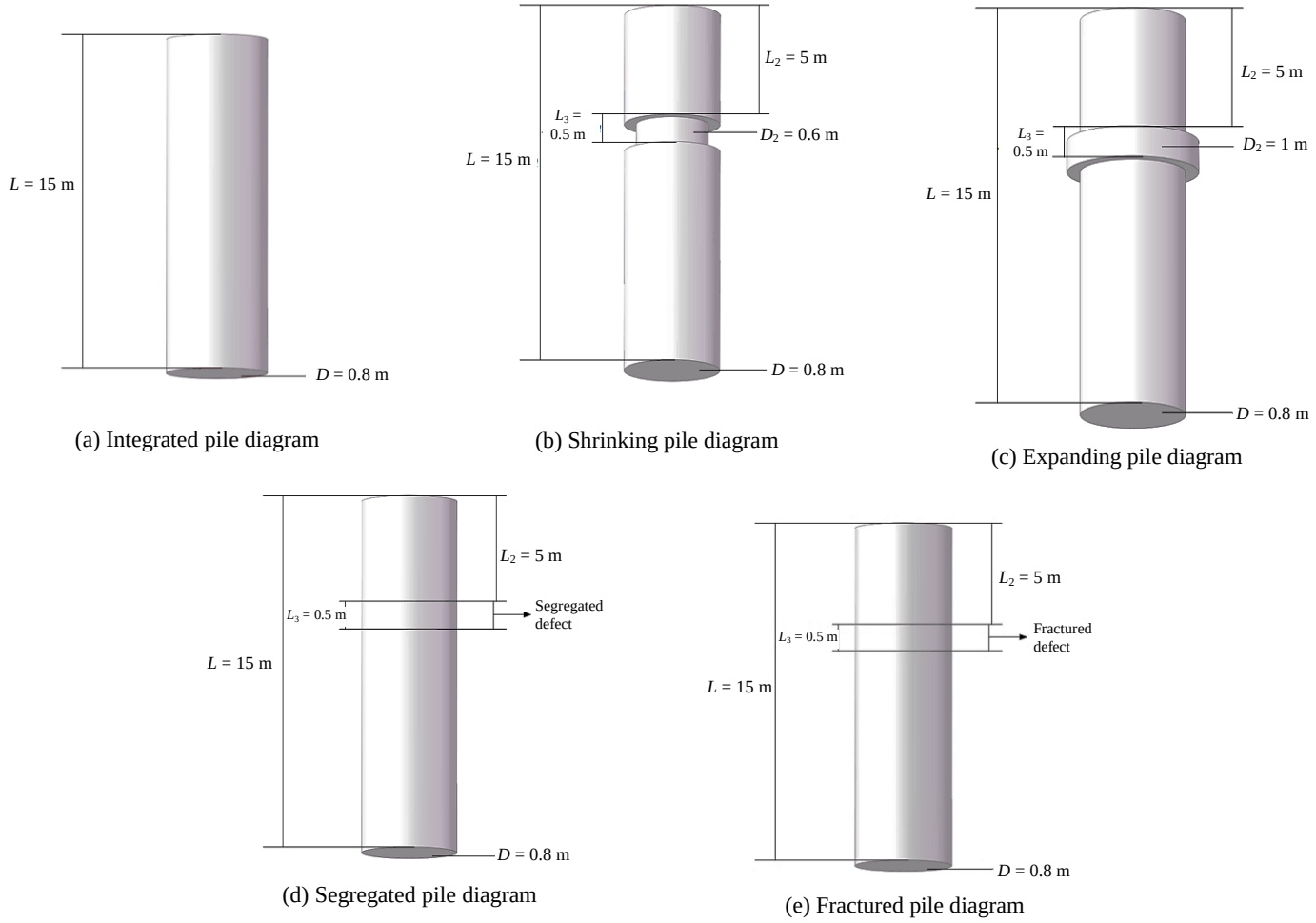
Suppose the optimal hyperplane is presented as follow:

$$w \cdot x + b = 0 \tag{16}$$

The set of example data samples should satisfy the following formula:

$$y_i (w \cdot x + b) \geq 1 \tag{17}$$

where w is weight vector, b is threshold value. The classification margin is $2/\|w\|^2$. In order to maximize it, the problem boils down to minimize $\|w\|^2/2$. According to the Lagrange multiplier, the problem comes to solve:



(a) Integrated pile diagram (b) Shrinking pile diagram (c) Expanding pile diagram

(d) Segregated pile diagram (e) Fractured pile diagram

Fig. 9 Simulated integrated pile examples

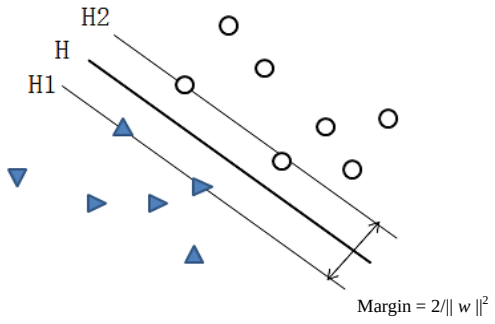


Fig. 10 SVM classifier-linear case

$$f(x)=\mathrm{sgn}(w\cdot x+b^*)=\mathrm{sgn}(\sum_i a_i\, y_i\, x_i\cdot x+b^*) \qquad (18)$$

where a_i is the Lagrange multiplier. However, in the actual situation, the models are often linear inseparable.

Then, the SVM problem turns to solve the convex quadratic programming:

$$\min \frac{1}{2}(w\cdot w)+C\sum_{i=1}^{n}\zeta_i$$

$$\text{s.t. } y_i\left[(w\cdot x_i)+b\right]+\zeta_i\geq 1 \qquad (19)$$

where ζ is slack variable, it represents the deviation degree between actual conditions and ideal linear cases, C is the penalty coefficient, it represents the penalty degree to the misclassification samples.

In nonlinear cases, the problems can be transformed into high dimensional linear problems through a nonlinear transform. The new decision function is:

$$f(x)=\mathrm{sgn}\left(\sum_{i=1}^{n}a_i^*\, y_i\, K(x,\, x_i)+b^*\right) \qquad (20)$$

$K(x_i,\ x_j)$ is a kernel function that satisfies the Mercer condition. As long as the kernel function is appropriate, the classification function can be found in a higher dimensional space. Due to the Support Vector Machine is completely described by training samples and the kernel function, it is important to choose the kernel function. At present, common kernel functions are the following:

- The linear kernel $K(x_i, x_j) = (x_i \cdot x_j)$
- The polynomial kernel $K(x_i, x_j) = (\gamma(x_i \cdot x_j) + a)^b$
- The sigmoid kernel $K(x_i, x_j) = \tanh(\gamma(x_i \cdot x_j) + a$
- The RBF kernel $K(x_i, x_j) = \exp(-\gamma(x_i - x_j)^2)$

For the purposes of classification, we applied LIBSVM (Lin *et al*. 2005). It is a software package created by Chih-Jen Lin, that contains functions about pattern recognition and regression.

The operation is simple, and it is easy to use. Currently, it has been widely adopted by many international famous research institutions. The LIBSVM provides common kernel functions, the RBF kernel function was applied in this paper.

Before inputting features to the classifier, the work of normalization was completed. The values of feature vectors were between 0 to 1. 300 samples including integrated pile, shrinking pile, expanding pile, fractured pile and segregated pile were gotten, each category had 60 data sets. We trained an SVM by half the input data sets, the rest of the samples were test sets. Finally, classification accuracy and the category labels were obtained. The classification categories are reported in Table 1.

4. EXPERIMENTAL RESULTS AND ANALYSIS

At first, for the purpose of comparison, higher order moment feature and conventional features including power spectrum density, variance, entropy were fed into SVM respectively. The experimental results are listed in Table 2. It is obvious that, the proposed feature gets a higher classification accuracy rates than the existing features. Furthermore, the accuracy of each kind of piles can be available. As the outputs of the SVM were predicted category labels, calculated the number of correct predicted labels in each kind of piles, and divided them by the input test sample number of each category. Then the accuracy of each kind of piles can be obtained. The "Average accuracy" is the average accuracy of five kinds of piles. Calculate the number of correct predicted labels in all kinds of piles, and divided them by the total number of input test sample data.

We obtain an average of ≈ 13.33% average accuracy improvement, and the discrimination ability for each category has improved. Particularly, the proposed feature is a good candidate for discriminating integrated, segregated, and fractured types. Consequently, higher order moment feature is sensitive and has a good ability for pile defect recognition.

In the above contribution, each sample was extracted 3 kinds of higher order moment features and each kind of feature had 14 feature vectors. So that one sample had 42 feature vectors in all. The 42-dimensional feature is too complex for the classifier's calculating. In view of this, to reduce the amount of input vectors, the 3 kinds of features were fused by PCA, and their dimension was reduced simultaneously. It can be observed from Fig. 11 that, the contribution rate of the first principal component is 42.4%, and the top ten principal components explained 98.8% of the total variance. In general, it can be argued that the principal components remain the original signal's information if the cumulative contribution rate reaches 98%. Therefore, we extracted the top ten principal components as new fusion feature. They were defined as multi-higher order moment feature.

As a comparison, the multi-higher order moment feature (MHF), single higher order moment feature (SHF), and the 42-dimensional feature (42-D) were fed to SVM respectively. Their classification performances and the computational time are shown in Table 3. It can be noted that, the MHF achieves the highest efficiency. The results of classification can be plotted in Fig. 12. The ordinate represents different categories. They are quantized by labels, as shown in Table 1. The abscissa represents test sample sets, and each kind of piles has 30 test sample sets.

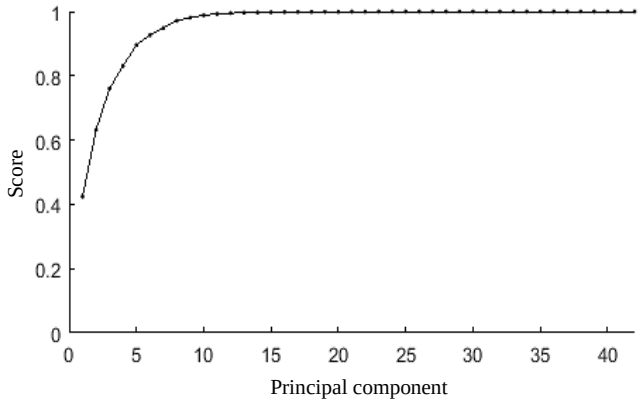


Fig. 11 The score of principal components

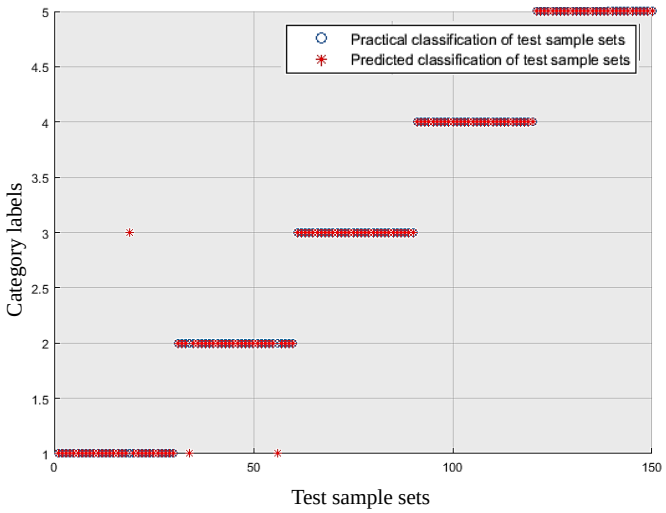


Fig. 12 Practical and predicted classification of test sample sets

Table 1 SVM output variables

Label	1	2	3	4	5
Pile category	Shrinking	Expanding	Integrated	Segregated	Fractured

Table 2 Analysis results of higher order moment feature and existing features

Feature category	Power spectrum density		Variance		Entropy		Higher order moment	
Pile category	Number	Accuracy of each category	Number	Accuracy of each category	Number	Accuracy of each category	Number	Accuracy of each category
Shrinking	60	36.67%	60	10%	60	10%	60	83.33%
Expanding	60	56.67%	60	63.33%	60	56.67%	60	56.67%
Integrated	60	100%	60	76.67%	60	100%	60	100%
Segregated	60	80%	60	100%	60	46.67%	60	100%
Fractured	60	100%	60	23.33%	60	100%	60	100%
Average accuracy	74.67%		54.67%		62.67%		88.00%	

Table 3 Analysis results of MHF, SHF, and 42-D features

Feature category	MHF	Single 2nd-order feature	Single 3rd-order feature	Single 4th-order feature	42-D feature
Average Accuracy (%)	98.00	90.67	87.30	78.67	88.00
Computational Time (s)	5.29	3.66	3.89	4.00	6.05

As the blue circles suggests, they correspond to the different category labels in the ordinate which are practical classification of test sample sets. While the red stars are predicted test sample sets, for example, there is a misclassification sample which should be in label 1, but it is misclassified to label 3. Clearly, it is easily to see misclassification samples in Fig. 12, and we can find that the predicted classification of test sample sets is essentially in agreement with practical classification of test sample sets.

As the combination of SHF, the advantages of MHF are obvious in discrimination, because the fusion feature is more complete and more comprehensive than the single one. Furthermore, compared with the most comprehensive 42-D feature, the classification accuracy increases by 10% as well as the training time reduces considerably.

5. CONCLUSIONS

To sum up, a novel pile defect identification feature, the multi-higher order moment feature has been introduced. In order to process the pile reflected signals, the wavelet packet transform method was applied. Moreover, this paper presented the sliding window method which was able to integrate information in both time and frequency domains and extract more precise features. Then we extracted the higher order moment features (the variance, skewness and kurtosis) in each wavelet packet band. Comparison experiments have accomplished using the SVM. Simulation results show that the higher order moment feature has a higher classification accuracy than other existing features. After reducing dimension by PCA, the accuracy increased to 98%.

We researched on extensions of the low strain integrity testing method including studies of the use of higher order moments in the classification process and overcame its shortcoming in low discrimination accuracy. In virtue of the low-cost, fast speed, and high accuracy advantages, the low strain integrity testing will be a more competitive method in pile quality detection.

ACKNOWLEDGMENTS

This work is funded by National Natural Science Foundation of China under Grant No. 61371174.

REFERENCES

Cai, Q.Y. and Lin, J.H. (2002). "Pile defect diagnosis based on wavelet analysis and neural networks." *Journal of Vibration and Shock*, **21**(3), 11-17.

Cristianini, N. and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines*. Cambridge University Press.

Dongzi, P. Shengming, C., and Yingdong, T. (2006). "Application of wavelet and neural network to defect diagnosing of piles." *Vibration Measurement & Diagnosis*, **26**(3), 207-302.

Gelb, J. and Andrew, W.O. (2010). "Active sonar clutter classification using higher order moments." *Proceedings of Meetings on Acoustics 159ASA*, ASA, **9**(1), 055001. http://doi.org/10.1121/1.3447961

Hsu, C.W., Chang, C., and Lin, C. (2005). "A practical guide to support vector classification." http://www.csie.ntu.edu.tw/~cjlin-papers/guide.

Kang, W. (2010). *Research on Pile Foundation Defect Detection Technology Based on Wavelet and Support Vector Machine*. Harbin Institute of Technology.

Kang, W.X., Li, J.D., and Liu, Y.M. (2015). "The singularity detection of stress wave signal in one-dimensional components based on quantitative information entropy." *International Journal of Signal Processing, Image Processing and Pattern Recognition*, **8**(10), 203-212. http://doi.org/10.14257/ijsip.2015.8.10.22

Kang, W.X., Zhao, Y., and Liu, L. (2017). "Pile defect detection based on wavelet packet energy ratio and support vector machine." *International Conference on Electronic Measurement and Instruments (ICEMI)*, Yangzhou, China, 92-97. http://doi.org/10.1109/ICEMI.2017.8265724

Karaye, I., Abdullahi, S.S., and Nalan, Ö . (2014). "Analysis of cardiac beats using higher order spectra." *Adaptive Science & Technology (ICAST), 2014 IEEE 6th International Conference on IEEE*, IEEE, 1-8. https://doi.org/10.1109/icastech.2014.7068145

Kaur, S., Harpreet, K., and Debashis, S. (2016). "Image quality measurement through structural similarity based on higher order moments." *India Conference (INDICON), IEEE Annual.*, IEEE, 1-6. https://doi.org/10.1109/indicon.2016.7839044

Kesharaju, M. and Nagarajah, R. (2015). "Feature selection for neural network based defect classification of ceramic components using high frequency ultrasound." *Ultrasonics*, **62**, 271-277. http://doi.org/10.1016/j.ultras.2015.05.027

Li, B., and Chen, X. (2014). "Wavelet-based numerical analysis: A review and classification." *Finite Elements in Analysis and Design*, **81**, 14-31. http://doi.org/10.1016/j.finel.2013.11.001

Liu, Y., Fei, Q., Yinghua, L., and Zhangzhi, C. (2010). "A daubechies wavelet-based method for elastic problems." *Engineering Analysis with Boundary Elements*, **34**(2), 114-121. http://doi.org/10.1016/j.enganabound.2009.08.004

Mendel, J.M. (1991). "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications." *Proceedings of the IEEE*, **79**(3), 278-305. http://doi.org/10.1109/5.75086

Morlet, J., Arens, G., Fourgeau, E., and Giard, D. (1982). "Wave propagation and sampling theory 1, complex signal and scattering in multilayered media." *Geophysics*, **47**(2), 203-221. http://doi.org/10.1190/1.1441328

Ogden, R.T. (2012). *Essential Wavelets for Statistical Applications and Data Analysis*. Springer Science & Business Media.

Pal, M. and Foody, G. (2010). "Feature selection for classification of hyperspectral data by SVM." *IEEE Transactions on Geoscience and Remote Sensing*, **48**(5), 2297-2307. http://doi.org/10.1109/TGRS.2009.2039484

Percival, D.B. and Walden, A.T. (2000). *Wavelet Methods for Time Series for Analysis*. Cambridge University Press. http://doi.org/10.1017/CBO9780511841040

Raiesdana, S., Shamsollahi, M.B., Hashemi, M.R., and Rezazadeh, I. (2007). "Wavelet packet decomposition of a new filter-based on underlying neural activity-for ERP classification." *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, IEEE, 1876-1879. https://doi.org/10.1109/iembs.2007.4352681

Simas Filho, E.F. and Seixas, J.M. (2016). "Unsupervised statistical learning applied to experimental high-energy physics and related

areas." *International Journal of Modern Physics C*, **27**(05), 1630002, 1-16. http://doi.org/10.1142/S0129183116300025

Tan, X., Yu, X.C., and Qin, J.C. (2014). "Multiple kernel SVM classification for hyperspectral images." *Chinese Journal of Scientific Instrument*, **35**, 405-411.

Vapnik, V.N. (1998). *Statistical Learning Theory*. Wiley, New York.

Vishwanathan, S.V.M. and Narasimha Murty, M. (2002). "SSVM: A simple SVM algorithm." *Neural Networks, 2002. IJCNN'02. Proc., 2002 International Joint Conference*, **3**, 2393-2398.

APPENDIX

A detailed example code about wavelet packet decomposition is as follows.

```
clc; clear;
data = load ('signal.dat');
z = zeros (8,139);
wpt1 = wpdec(data, 3, 'db7', 'shannon');
a = wpcoef (wpt1, [3,0]);
b = wpcoef (wpt1, [3,1]);
c = wpcoef (wpt1, [3,2]);
d = wpcoef (wpt1, [3,3]);
e = wpcoef (wpt1, [3,4]);
f = wpcoef (wpt1, [3,5]);
g = wpcoef (wpt1, [3,6]);
h = wpcoef (wpt1, [3,7]);
z(1,:) = a;
z(2,:) = b;
z(3,:) = c;
z(4,:) = d;
z(5,:) = e;
z(6,:) = f;
z(7,:) = g;
z(8,:) = h;
z;
```