A Comparison of Pattern Classification Approaches for Structural Damage Identification

Zhou Qingqing[1], Zhou Qifeng[1], Ning Yongpeng[1], Yang Fan[1] , Lei Jiayan[2],Tian Run[1],Liu Minda[1]
[1]Department of Automation
[2]Department of Civil Engineering
Xiamen University
Xiamen, China
{dzqq0815@126.com, {zhouqf, zhouqf}@xmu.edu.cn
{961331014, 7704707, 37683667, 1198083772}@qq.com

Journal of Digital Information Management

ABSTRACT: *A structural damage identification approach based on wavelet packet decomposition (WPD) and random forests (RF) was proposed and compared with other pattern classification approachs. The main procedure involves extracting energy features from vibration acceleration data through wavelet packet decomposition and then using these features as input for a RF classifier. The experiment was carried on an 8-storey steel shear building model in the case verification. The results show that the proposed method is effective for structure damage detection and can achieve a higher accuracy than other commonly used methods.*

Categories and Subject Descriptors:
I.5 {Pattern Recognition]: Models; **G.1.2 [Approximation]:** Wavelets and fractals

General Terms: Structural damage detection, Pattern Matching, Wavelet Analysis

Keywords: Wavelet Packet Decomposition, Pattern Matching, Extraction algorithms, Sensor identificaiton

Received: 7 November 2011, Revised 29 December 2011, Accepted 7 January 2012

1. Introduction

The performance of civil structures and mechanical systems deteriorates during their service life. Due to aging, damages, and other harmful effects, the structural stiffness of construction will decreases. By instrumenting structures with a vibration sensor system, structural health monitoring (SHM) aims to provide reliable and economical approaches to monitor the performance of structural systems in an early stage so as to facilitate the decisions on structure maintenance, repair and rehabilitation [1]. That may greatly improve social and economic benefits, and reduce the probability of accident happened as well.

In recent years, many experts and scholars have deeply researched the structural damage identification in buildings and bridges. Great efforts on the development of vibration-based damage identification methods have been made. And many effective methods of pattern recognition is applied in the field of damage identification, and achieved good results, such as neural networks, support vector machine (SVM), principal component analysis (PCA), independent component analysis (ICA) etc. Wavelet packet decomposition (WPD) has attracted increasing attention to fault diagnosis and structural damage detection. WPD is a wavelet transform where the signal is passed through more filters than discrete wavelet transform (DWT) [2]. In DWT, each level is calculated by passing only the previous approximation coefficients through a high and low pass filters. While, in WPD, both the detail and approximation coefficients are decomposed. This paper proposes a damage identification method based on the integration of Wavelet packet decomposition and random forest. Comparing with other signal processing and pattern classification methods, the proposed method can achieve better performance.

The rest of the paper is organized as follows: Section 2, 3 gives the background materials of damage features extraction algorithm and random forest. The framework of damage identification model is presented in Section 4. Section 5 carries out experiments on an eight–story shearing steel frame structure to compare the computational efficiency and classification performance of PCA_RF, WPD_RF and WPD_SVM. Finally, the

conclusion and some remarks are given in Section 6.

2. Features Extract from WPD

According to the characteristics of wavelet packet analysis, it can be applied in the field of structural damage identification; the extraction procedure of damage features is as follows [3]:

Step 1: Do N -layer wavelet packet decomposition for the acceleration signal, and extract the signal feature of all frequency components from low to high frequency.

Step 2: Recostruct wavelet packet decomposition coefficients, and extract signals S_{Nj} of each band range;

Step 3: According to the formula:

$E_{Nj} = \int \left| S_{Nj} \right|^2 dt = \sum_{k=1}^{n} \left| x_{jk} \right|^2$ calculate the total energy of each band signal, x_{jk} where means the amplitude of the kth discrete points of reconstruction signal S_{Nj}; N is the number of discrete points of reconstructing signal S_{Nj};

Step 4: Normalize E_{Nj}, and then get final damage feature vectors $T = \left[E'_{N1}, E'_{N2}, \cdots E'_{N2^N} \right]$.

A comparison of energy of structural vibration response signals between damaged and undamaged structures in some special frequency bands will exhibit some remarkable differences. This is because the structural damage will suppress or enhance certain components of vibration response signal [4]. Thus, the energy of structural vibration response signals with different frequency components contains ample information of structural damage. The energy variation of one or several frequency components can indicate a special structural damage status.

3. Random Forests method

Random forests (RF) are a machine learning algorithm for classification or regression developed by Leo Breiman [5-8]. RF is a combination of tree predictors such that each tree depends on the values of a random vector sampled independently with the same distribution for all trees in the forest [6]. So, random forests method is to generate a large number of trees and let them vote for the most popular class, each decision tree is constructed by recursively splitting each internal node into left and right nodes until the stop criterion is met. Random forest uses both Bagging (bootstrap aggregation), a successful approach for combining unstable learners, and random variable selection for tree building. The idea is maintaining the low-bias trees while reducing their correlation with each other.

Definition: A random forest is a classifier consisting of a collection of tree-structured classifiers $\{h(x, \theta_k), k=1,...\}$, where the $\{\theta_k\}$are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input $\boldsymbol{x}$.

Random forests method has following advantages:

(1) It is a fast algorithm (can be faster than growing/pruning a single tree) and easily parallelized.

(2) Comparable to SVM and Neural Networks, RF can reduce variance and get low estimating bias and has good accuracy without over-fitting.

(3)RF handles high dimensional data more effectively.

(4) It can give internal unbiased estimate of test set error as trees are added to ensemble.

The algorithm has been shown to have desirable properties, such as the rather stable, robust with noise and convergence of generalization errors.

4. Damage identification based on WPD and RF

In this paper, we design a damage identification model based on WPD and RF. WPD can extract the energy in different frequency bands, while the different damages have different energy distribution, so the extracted energy can reflect the type of damage and then they were used as the input vectors of random forests to identify the damage. Compared with other commonly used pattern classification approaches RF has more stable classification performance and much faster classification speed.

4.1 Experimental model

To illustrate the effectiveness of proposed method, four methods were tested on an 8-storey steel shear building model. It is shown in Figure.1, the plane of steel framework in each layer is 350mm×250mm, and the storey is 200mm high. Structural damage was happened when the stiffness reduced between two layers. In this experiment, stiffness reduction was simulated by changing the weight of steel column pieces. Excitation point is located in the third layer, and there is a PCB sensor in each frame. Setting the sampling frequency is 128Hz, and the sampling time is 40s.

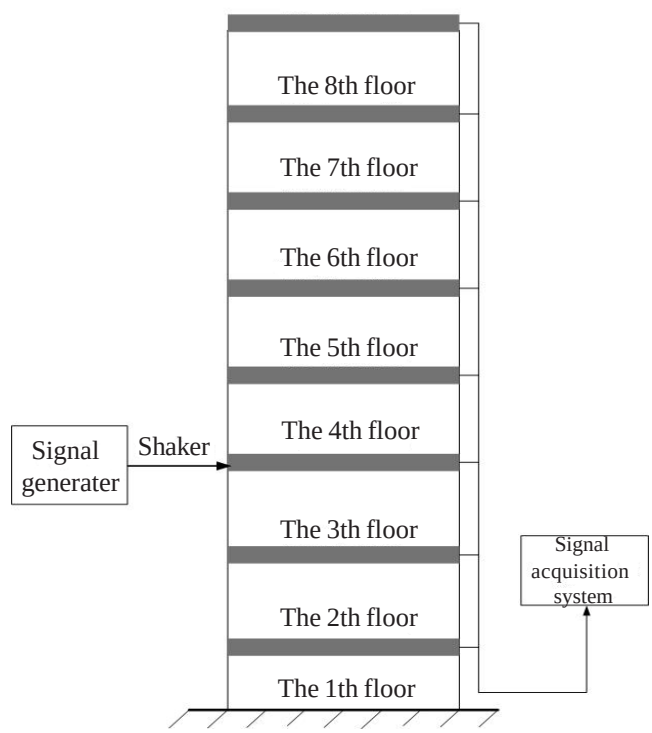

Figure 1. An 8-storey steel shear building model

4.2 Experiment data collection

Experiment simulates the various common structural damages: (1) there is un-damage in all eight layers; (2) the 3rd layer occurs 25% damage; (3) the 3rd layer occurs 50% damage; (4) the 5th layer occurs 25% damage; (5) the 5th layer occurs 50% damage. Ten samples are collected for each case. We defined class label as follows:

Patterns of Damage	Class label
Un-damage	1
25% damage in the 3rd layer	2
50% damage in the 3rd layer	3
25% damage in the 5th layer	4
50% damage in the 5th layer	5

Table 1. Damage cases and Class labels

4.3 Experimental process

In the experiment, WPD and RF are combined to build a classification model by using 5 data sets. Wavelet packet decomposition was used to decompose the signals original. These samples are decomposed into a set of 16-dimensional feature vectors that can be used as the input vectors of damage identification. Figure.2 to Figure.6 shows the band energy of each class label.

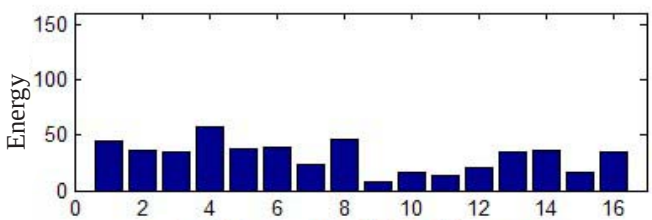

Figure 2. The band energy of class label 1

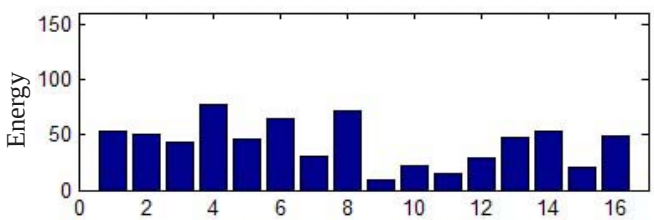

Figure 3. The band energy of class label 2

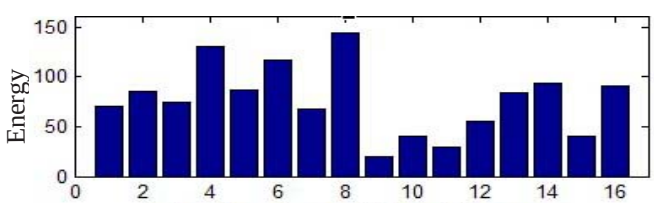

Figure 4. The band energy of class label 3

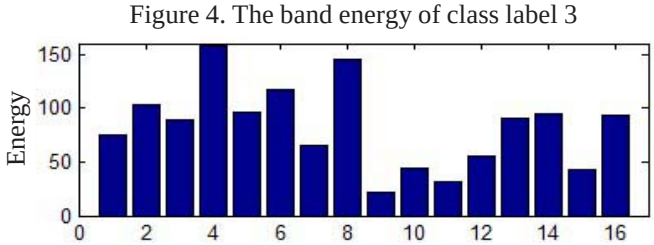

Figure 5. The band energy of class label 4

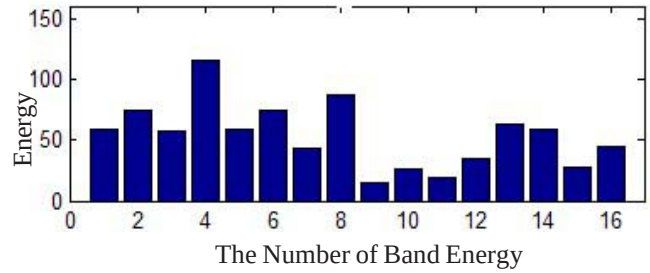

Figure 6. The band energy of class label 5

The above five figures show that, the different damaged types have different band energy, these differences of energy can be as the features to distingue distinguish the different damaged types. Integrating WPD and RF, the recognition accuracy rate has been significantly improved, reached more than 90%, the detail results are displayed in the table2.

5. Comparative experiments

In this section, we are going to do three experiments to compare some pattern classification methods in structural damage detection.

Method1. In this experiment, we will use five data sets above-mentioned to construct a RF model, The different damage cases have different outputs from sensor, so the outputs from the sensor can distinguish damage and un-damage. Figure.7 shows the different waveforms between damaged sensor data (25% damage in the 3rd layer) and undamaged sensor data from the first PCB sensor.

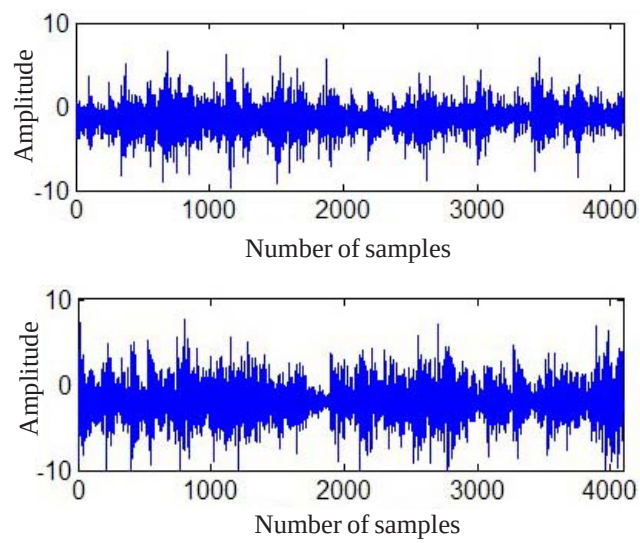

Figure 7. Un-damaged sensor data and damaged sensor data

From the two figures, we observed that there are obvious differences between the undamaged data and damaged data. So the damaged and undamaged sensor data (total of five types) were directly input into RF, the size of each sensor data is 2048×8 (2048 examples with 8 features, which were selected from sample point 1025 to 3072 of the original signal), nTree=1000. The results show that the recognition accuracy rate is only 63%.

Method 2. In the second experiment, we aim to analyze the accuracy of damage detection on combining principal

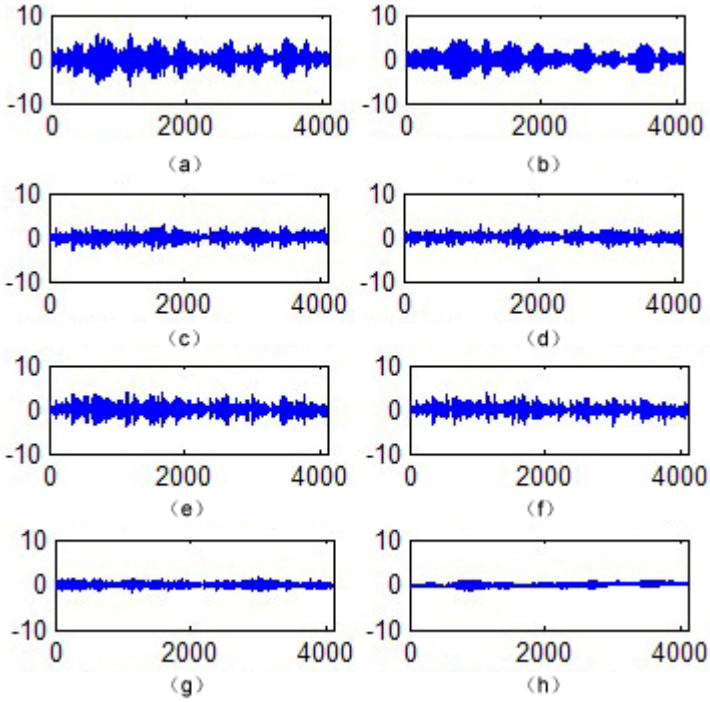

Figure 8. Principal Components of un-damage

component analysis and random forests. Firstly, the five classes of samples respectively worked as input to PCA. Figure.8 shows the signals after PCA, (a)-(h) respectively represents the first to eighth erincipal component, and Figure.9 shows the contribution rate of each signal after PCA.

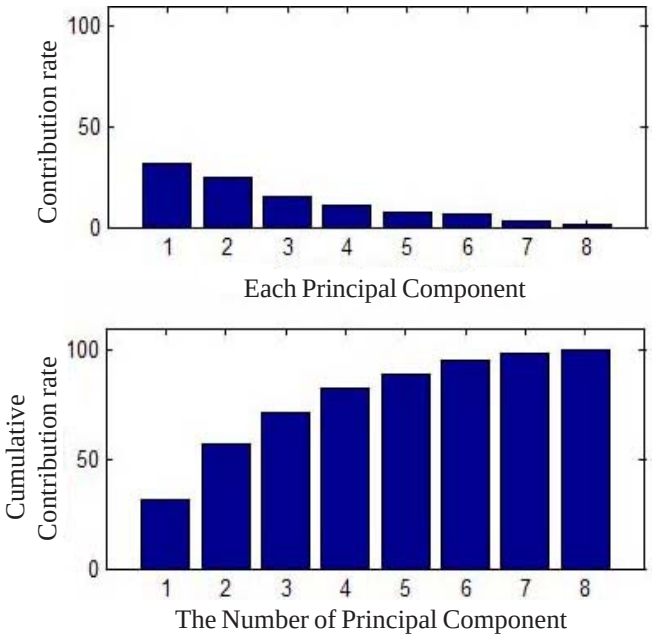

Figure 9. Contribution rate of each Principal Component

From Figure.9, we learned that only six principal components can represent more than 96% original information. Six principal components were selected to replace the eight signals from sensors. Secondly, the processed data (after PCI) were input into RF, the size of each sensor data is 2048×6 (2048 examples with 6 features), nTree=1000. The classification accuracy is measured as Experiment 1. And the result shows that the recognition accuracy rate is better than Method 1.

Method 3. The last experiment, we compared the performance achieved by integrating WPD and RF with that achieved by integrating WPD and SVM.

Firstly, 70% of the each data was randomly selected (which had be as input in experiment1-3) as training data to construct a SVM model, and then apply it to test unseen data (the remaining 30% of the data), exploring that if they are correctly recognized. Results in table2 showed that WPD-RF obtains better classification accuracy than WPD-SVM.

Method of Pattern recognition / Method of Data pre-processing	RF	SVM
The original signal	63.05%	67.74%
PCA to the original signal	65.58%	0.35%
WPD to the original signal	**92.38%**	7.46%

Table 2. The Accuracy Compared RF with SVM

6. Conclusion

The results listed in Table 2 show that the damage identification accuracy obtained by WPD and RF is the best compared with those obtained by direct RF, PCA_RF and WPD_SVM. Both the detail and approximation coefficients of the sensor signals are decomposed in WPD, and the obtained components were input into RF for structural damage classification. Different types and levels of damage can be identified well from the prediction

output. In future, we will continue to analyze wavelet packet decomposition and random forests for detecting the damage location and consequently support the repair decisions.

7. Acknowledgment

This work is supported by the Fundamental Research Funds for the Central Universities (201012106)

References

[1] Song, Huazhu., Zhong, Luo., Han, Bo(2005). Structural Damage Detection by Integrating Independent Component Analysis and Support Vector Machine. Lecture Notes in Computer Science, 670-677.

[2] Yang, Xiaoyuan., Yu, Lei, Xiaozhong, Pan., Jia, Liu, (2009). Universal Image Steganalysis Based on Wavelet Packet Decomposition and Empirical Transition Matrix in Wavelet Domain. International Forum on Computer Science-Technology and Applications, 179-182.

[3] He, Haoxiang., Yan, Weiming., Peng, Lingyun, (2008). Damage Identification of Reinforced Concrete Bridge Based on Support Vector Machine. *Journal of Highway and Transportation Research and Development*, 2 (3) 65-69.

[4] Yam, L. H., Yan, Y. J., Jiang, J. S. (2003). Vibration-based Damage Detection for Composite Structures Using Wavelet Transform and Neural Network Identification, Composite Structures, p. 403-412.

[5] Breiman, L. (1996). Random preditors. *Machine Learning*, 24 (2) 123-140.

[6] Breiman, L. (2001). Random forests. *Machine Learning* 45 (1) 5-32.

[7] Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J. (1984). Classification and regression trees. Cole, Pacific Grove, California, USA: Wadsworth & Brooks.

[8] Dietterich, T. (1998). An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting and randomization. Machine Leaning, p. 1-22.

[9] Huaizhong, Zhang., Hamprecht, Fred., Amann, Anton(2005). Report about VOCs Dataset's Analysis based on Random Forests Method. High-Performance Computing in Asia-Pacific Region.

[10] Zhou, Qifeng., Hong, Wencai., Shao, Guifang., Cai, Weiyou(2009). A New SVM-RFE Approach towards Ranking Problem, *In*:International Conference on Intelligent Computing and Intelligent Systems. 11, 270-273.

[11] Qifeng, Zhou., Hong Wencai., Yang ,Fan., Luo, Linkai, (2010). Feature selection of random forest based proximity matrix difference. Huazhong, *J. Univ. of Sci. & Tech.* (Natural Science Edition), 38 (4) 58-61.