Research on Classification of Fiber Intrusion Signal Based on Supported Vector Machines

Jie Zhu
Department of Electronic Engineering
Shanghai Jiao Tong University
Shanghai, China, 200240
zhujie@sjtu.edu.cn

***Abstract*—The widely use of optical fiber gives rise to the need of its protection and intrusion detection. An optical fiber system in which the optical path satisfies the structure of sagnac loop can easily form a distributed fiber senor. With the photo-elastic effect, when intrusion happens, there will be optical signals created in the fiber, and with optical and electrical methods, one can get the intrusion related signals for analysis. After obtaining the signals, we use differential phrase demodulation method to demodulate the signals. With the demodulated signals, the feature vector of the signals can be extracted through time-frequency analysis. Then, supported vector machine (SVM) is used to classify 3 different types of intrusion. For better accuracy of classification, we use wavelet de-noising to do the noise elimination. Field experiments showed that the system is reliable and of good accuracy.**

Keywords-fiber sensor; differential phrase demodulation; supported vector machine; Time-Frequency analysis; wavelet de-noising.

I. INTRODUCTION

Optical fiber now is widely used in our life, such as cable television network, telephone switching network, Internet, and so on. It permits transmission over longer distances and at higher bandwidths than other forms of communication. As it is so widely used, it is vulnerable to different kinds of intrusion. But, with photo-elastic effect, optical fiber can serve as distributed fiber sensor, and this creates convenience for fiber intrusion detection and recognition. By processing the signal given by the "fiber sensor", we can get the information of the intrusion, and then, we are able to classify intrusion with machine learning method. In this paper, *Supported Vector Machine* (SVM) [1] is chosen for classification.

SVM became popular some years ago for solving problems in classification, regression, and novelty detection. Compared with the method of neural network [2] and other machine learning methods, SVM requires less training samples, and this feature meets our requirement in this case.

In order to improve the performance of classification, we focus on the choice of feature vector and noise elimination. In Section IV, 17 features are chosen to form the feature vector after comparison test. In Section V, wavelet de-noising is applied to further enhance accuracy rate of classification

II. METHOD OF SIGNAL ACQUISITION

To perform the fiber intrusion signal recognition, firstly, we should make our optical path satisfy the structure of Sagnac loop [3] just by adding a feedback module and an interference/multiplexing module to our commonly available fiber system. Then, the distributed fiber sensor system is formed as follows:

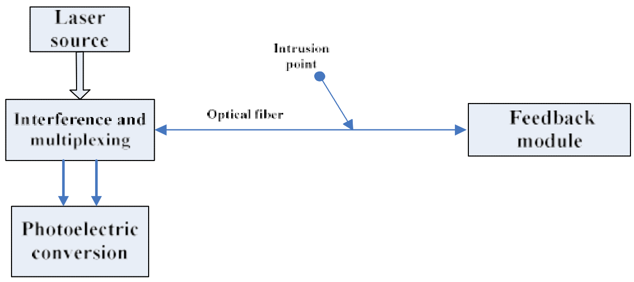

Figure 1. Distributed fiber sensor system.

Once intrusion happens, after photoelectric conversion, we can get the pair of coherent electrical signals $x(t), y(t)$ in the form of (1):

$$\begin{cases} x(t) = A\cos(\Delta\varphi(t) + \alpha) - A\cos\alpha \\ y(t) = B\cos(\Delta\varphi(t) + \beta) - B\cos\beta \end{cases} \tag{1}$$

Here, $\Delta\varphi(t)$is the signal that contains the intrusion information. A and B are amplitude coefficients. The phrases α, β are produced by the interference module [4].

Conduct differential, multiplication, and integration on $x(t)$ and $y(t)$, we get $\Delta\varphi(t)$ with amplitude coefficient $AB\sin(\alpha-\beta)$as follows:

$$\begin{aligned} &AB\sin(\alpha-\beta)\,\Delta\varphi(T) \\ &= \int x'(t) * y(t) - x(t) * y'(t) + \\ &A\cos\alpha * y(t) - B\cos\beta * x(t) \end{aligned} \tag{2}$$

$\Delta\varphi(T)$ is the signal that contains intrusion information, which we desire to obtain.

III. SUPPORTED VECTOR MACHINE

SVM became popular some years ago for solving problems in classification, regression, and novelty detection. It was originally designed for binary classification. In this paper, we use one-against-one method [5] to implement our multiclass classification problem. An SVM learns the decision boundary between two classes by mapping the training sample vectors onto a higher dimensional space, and then, determining an optimal separating hyper-plane [6]-[7], as shown in Figure 2.

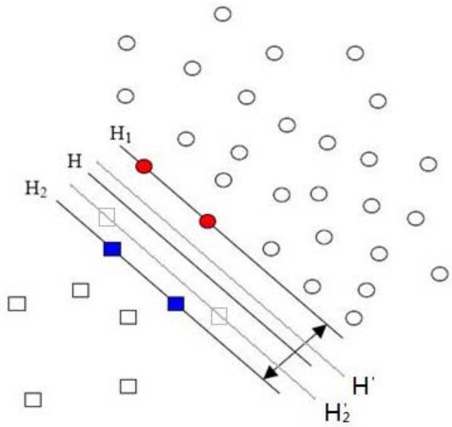

Figure 2. Optimal separation of two linear separable classes by hyper-plane

In Figure 2, "○" represents the class one and the "□" represents the class two. A good choice for classification is the hyper-plane that leaves the maximum margin between the two classes, where the margin is defined as the sum of the distances of the hyper-plane from closest point of the two classes, like the H_1 and H_2 in the above Figure.

Considering the training feature vectors of two classes,

$$(x_i, y_i), x_i \in R^n, y_i \in \{-1, +1\} \quad (3)$$

SVM algorithms will find a pair of parallel optimal hyper-planes, defined as follows:

$$\begin{aligned} H_1 &: y = \boldsymbol{\omega} \cdot \mathbf{x} - b = +1 \\ H_2 &: y = \boldsymbol{\omega} \cdot \mathbf{x} - b = -1 \end{aligned} \quad (4)$$

to separate the two classes, so that the margin, i.e. the distance between two hyper-planes, is the largest. This is the sum of the shortest distance $2/\| \boldsymbol{\omega} \|$ from the hyper-plane to the closest positive and negative examples. The training vectors on the hyper-lanes are called support vectors [8]. The hyper-planes are located by solving the optimization problem:

$$\min \| \boldsymbol{\omega} \|^2 + C \sum \xi_i \quad (5)$$

subject to

$$\begin{aligned} \boldsymbol{\omega} \cdot \mathbf{x} - b &\geq +1 - \xi_i \\ \boldsymbol{\omega} \cdot \mathbf{x} - b &\leq -1 + \xi_i \end{aligned} \quad (6)$$

If $\xi_i = 0$, the two classes are linearly separable and there are no data points between H_1 and H_2. If $\xi_i > 0$, the two classes are not linearly separable; for the data violating the maximum margin condition, a penalty controlled by $C > 0$ is given to balance margin maximization and classification errors. Using Wolfe duality theory [9], the problem can be transformed to the following dual problem:

$$\max \sum_i^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \boldsymbol{X}_i \cdot \boldsymbol{X}_j \quad (7)$$

subject to

$$\sum_i^N \alpha_i y_i = 0 \quad (8)$$

$$0 \leq \alpha_i \leq C \quad (9)$$

Therefore:

$$\boldsymbol{\omega} = \sum_i^N \alpha_i y_i \mathbf{X}_i \quad (10)$$

In the case where a linear boundary is inappropriate, the SVM can map the input vector into a high dimensional space through function $\psi(x)$, where it can construct a linear hyper-plane in the high dimensional space, then kernel function can be expressed as

$$k(\boldsymbol{X}_i \boldsymbol{X}_j) = \boldsymbol{\psi}(\boldsymbol{X}_i) \cdot \boldsymbol{\psi}(\boldsymbol{X}_j) \quad (11)$$

That is, the dot product in that high dimensional space is equivalent to a kernel function of the current space [10].

IV. CLASSIFICATION OF INTRUSIONS

In field experiment, three types of intrusion are conducted, namely, Type 1: Excavating with shovel near where the cable is buried, Type 2: Beating directly on the optical fiber with hand, Type 3: Beating the fiber cable with shovel. Examples of the time domain waveforms and spectrums of the 3 types of intrusions are shown as follows (Figure 3, Figure 4, and Figure 5):

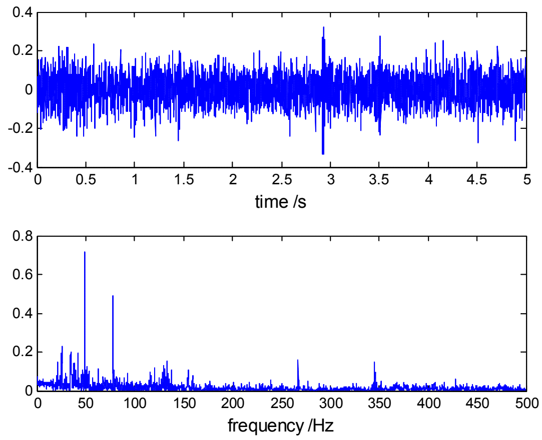

Figure 3. Excavating with shovel (type 1).

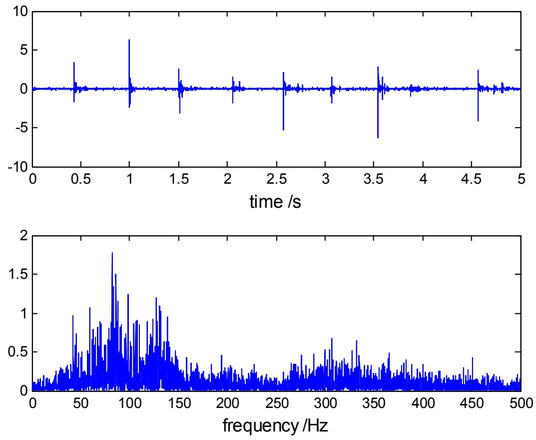

Figure 4. Beating directly on the optical fiber (type 2).

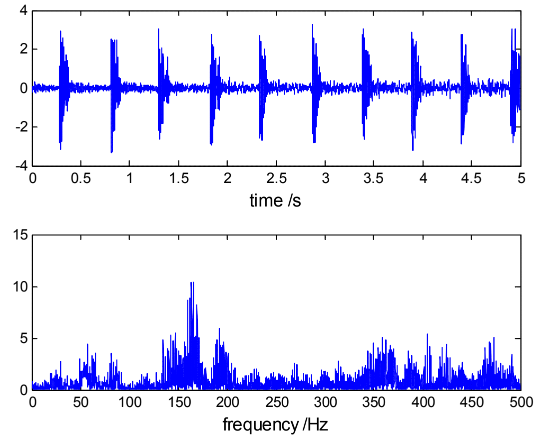

Figure 5. Beating the fiber cable (type 3).

In order to use SVM to do the classification, we must first extract the feature vector. In this paper, we do both time and frequency analysis to get the feature vector suitable for our classification system.

A. Time domain analysis

For every frame of data, in time domain, we compute the following 5 values as components of feature vector.

(1) Maximum value of signal magnitude
(2) Minimum value of signal magnitude
(3) The number of peaks whose height exceed half the maximum magnitude
(4) Average signal magnitude
(5) The ratio of maximum signal magnitude to the average of signal magnitude

B. Power spectrum analysis

For every frame of data, we compute its autocorrelation function, and then, do FFT (Fast Fourier Transformation) to get its power spectrum. We get the following 6 feature values:

(1) The frequency of the highest peak in the power spectrum
(2) The height of the highest peak in the power spectrum
(3) The frequency of the second highest peak in the power spectrum
(4) The height of the second highest peak in the power spectrum
(5) The ratio of the height of the highest and second highest peak in the power spectrum
(6) The number of peaks whose height exceed half the height of the highest peak in the power spectrum

Then, using the energy distribution information we get the following 6 feature values.

(1) The ratio of the energy of the highest peak to the energy of the whole power spectrum
(2) (2)-(6): The ratio of energy in frequency bands 0-99Hz, 100-199Hz, 200-299Hz, 300-399Hz, 400-499Hz respectively to the energy of the whole power spectrum

With the total 17 features, the feature vector can be constructed and used for SVM classification. We use these features to set the optimal parameters C and G in SVM by 10 fold cross-validation and Grid method. In Figure 6, the blue line has the lowest accuracy, 80%, and the red line have the highest accuracy, 93.5%; so, we could select the optimal parameters in the red line. After selecting the optimal parameters, we could get the final accuracy of classification by SVM, like in Figure 7 and Figure 10.

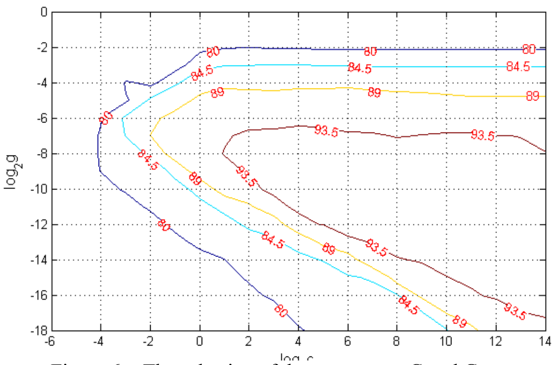

Figure 6. The selection of the parameters C and G

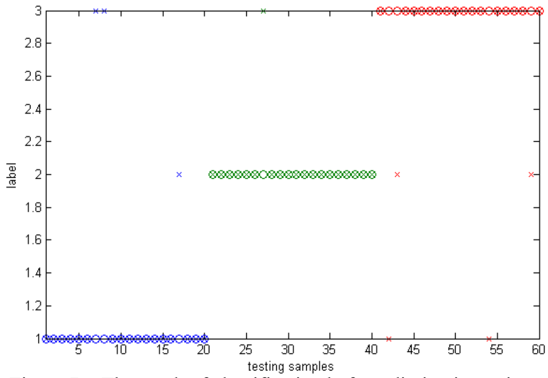

Figure 7. The result of classification before eliminating noise

TABLE I. CLASSIFICATION RESULT

Type 1		Type 2		Type 3	
1	1	2	2	3	3
1	1	3	2	1	3
1	1	2	2	2	3
1	1	2	2	3	1
1	1	2	2	3	3
1	1	2	2	3	3
3	2	3	2	3	3
3	1	2	2	3	3
1	1	2	2	3	2
1	1	2	2	3	3

Figure 7 is result of the classification before eliminating noise, and then, we convert it to be a table as the Table I. In this Figure, "O" is for label of testing samples and "X" is for the prediction label of testing samples. We randomly choose 40 samples from each type of intrusion as training data, and randomly choose 20 samples from the left of each type as testing data. Using one-against-one method to do the multiclass classification, in testing samples of Type 1, there are two misclassifications that mistake Type 1 for Type 3 and one misclassification that mistakes Type 1 for Type 2; in testing samples of Type 2, there are two misclassifications that mistake Type 2 for Type 3; in testing samples of Type 2, there are two misclassifications that mistake Type 3 for Type1 and two misclassifications that mistake Type 3. The accuracy of classification is 85%.

Marking the time domain features as Group 1, power spectrum features as Group 2, and energy distribution features as Group 3, we, respectively, use Group 1, Group 2, Group 3 , Group (1,2), Group (1,3), Group (2,3)'s features to consist the feature vector, and conduct training and testing with the same data set and method as above. The result is shown in Table II:

TABLE II. CLASSIFICATION RESULT

	G1	G2	G3	G1, 2	G1, 3	G2, 3
accuracy(%)	63.3	78.3	76.7	81.6	80	83.3

From the result shown in Table II, one can see that absence of any group will result in decline of the classification accuracy, so, we use 3 groups together to form the feature vector.

V. WAVELET DE-NOISING

In Section IV, the SVM classification result is not so satisfying and we tried to choose some other time-frequency features to consist the feature vector, but, the improvement is limited. In this situation, de-noising is needed for the improvement of our classification.

In this paper, we adopt the wavelet-based de-noising [11]. Firstly, the wavelet transform performs a correlation analysis; therefore, the output is expected to be maximal when the input signal most resembles the mother wavelet. Secondly, if a signal has its energy concentrated in a small number of wavelet dimensions, its coefficients will be relatively large compared to any other signal or noise that its energy spread over a large number of coefficients. Thirdly, shrinking the wavelet transform will remove the low amplitude noise or undesired signal in the wavelet domain, and an inverse wavelet transform will then retrieve the desired signal with little loss of details.

The noise in the signals $x(t), y(t)$ in (1), that we get after photoelectric conversion, has similar properties as white noise, so, as the most popular method for white noise, de-noising, wavelet de-noising is used in this paper.

We use soft-threshold, choose sym8 in the Symlets family [12], as our wavelet base and do 6 level of decomposition. The choosing of wavelet base and the depth of decomposition level is a compromise between de-noising performance and efficiency [13]. Figure 8 and Figure 9 show the de-noising results of time and frequency domain, respectively.

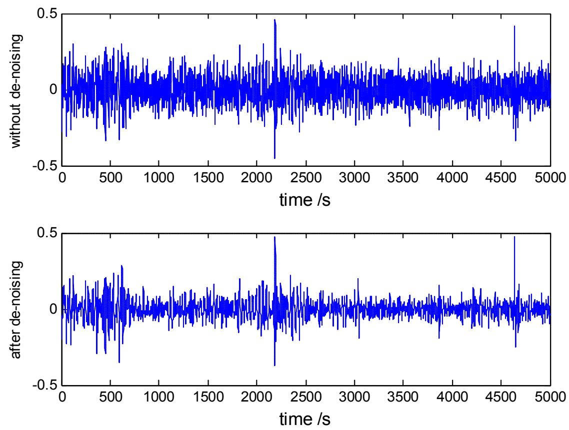

Figure 8. Time domain comparison.

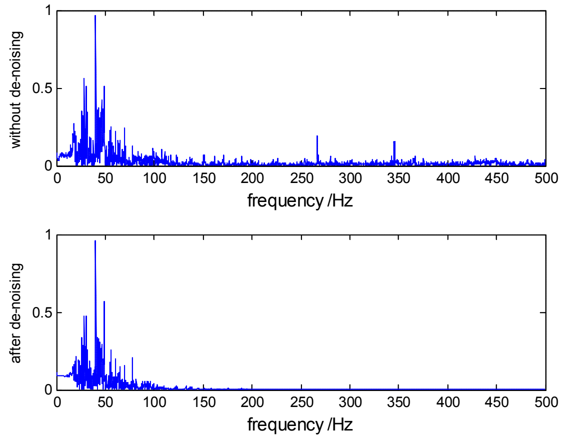

Figure 9. Spectrum comparison.

Using the same feature vector, same training and test data set as in Section IV, we get the following SVM classification result.

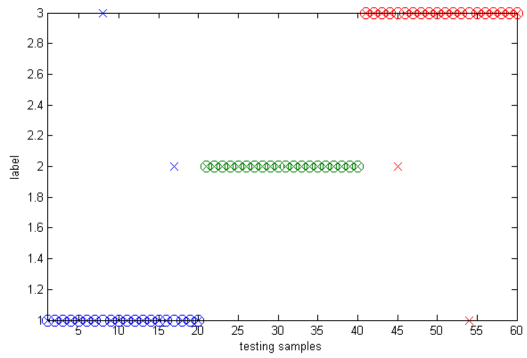

Figure 10. The result of classification after eliminating noise

TABLE III. CLASSIFICATION RESULT WITH DE-NOISING

Type 1		Type 2		Type 3	
1	1	2	2	3	3
1	1	2	2	3	3
1	1	2	2	3	3
1	1	2	2	3	1
1	1	2	2	2	3
1	1	2	2	3	3
1	2	2	2	3	3
3	1	2	2	3	3
1	1	2	2	3	3
1	1	2	2	3	3

Figure 10 is result of the classification after eliminating noise, and then, we convert it to be a table, as the Table III.

In this Figure, "O" is for label of testing samples and "X" is for the prediction label of testing samples. The accuracy of classification reaches 95%. From the comparison of the results of two trials (one with wavelet de-noising and one without), we can see that wavelet de-noising is quite effective for the improvement of classification accuracy.

VI. CONCLUSION AND FUTURE WORK

In this paper, we focused on the improvement of performance of SVM-based fiber intrusion signal recognition system. Three different types of intrusion are experimented, and kinds of time-frequency features are tried to serve as components of feature vector for SVM. Finally, 17 kinds listed in Section IV are chosen. Wavelet de-noising is used to enhance the performance of classification, and the improvement is obvious. Further research is needed to improve the composition of feature vector and find better de-noising methods to make the classification more accurate.

REFERENCES

[1] R.J. Kuo and C.M. Chen, "Evolutionary-Based Support Vector Machine", Industrial Engineering and Engineering Management (IEEM), 2011, pp. 472-475.

[2] E. Byvatov, U. Fechner, J. Sadowski and G. Schneider, "Comparison of Support Vector Machine and Artificial Neural Network Systems for Drug/Nondrug Classification", J. Chem. Inf. Comput. Sci., 2003, vol. 43, no. 6, pp. 1882-1889.

[3] A. Owen, G. Duckworth and J. Worsley, "OptaSense: Fibre Optic Distributed Acoustic Sensing for Border Monitoring", Intelligence and Security Informatics Conference (EISIC), European, 2012, vol. 59, pp. 362-364.

[4] Changyu Shen, Jinlei Chu, Yanfang Lu, et al. "High Sensitive Micro-Displacement Sensor Based on M-Z Interferometer by a Bowknot Type Taper", Photonics Technology Letters, IEEE, 2014, vol. 26, pp. 62-65.

[5] Chih-Wei Hsu and Chih-Jen Lin, "A comparison of methods for multiclass Support Vector Machines," IEEE Trans. Neural Networks, vol. 13(2), 2002, pp. 415- 425.

[6] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," Mach. Learn. , Jan. 2002, vol. 47, no. 1-3, pp. 389-422.

[7] C.-H. Li, H.-H. Ho, Y.-L. Liu, B.-C. Kuo, and J.-S. Taur, "An automatic method for selecting the parameter of the normlized kernel function to support vector machines," *J. Inf. Sci. Eng.*, Jan. 2012, vol. 28, no. 1, pp. 1-15.

[8] A. Pattle and D. S. Chouhan, "SVM kernel function for classification," Advances in Technology and Engineering (ICATE), 2013, pp. 1-9.

[9] P. Wolfe, "A duality thorem for non-linear programming", Q. Appl. Math. 19, 1961, pp. 239-244.

[10] Changxue Ma, Randolph, M.A; Drish, J, "Distributed fiber optic acoustic sensor for leak detection," Proceedings ICASSP-2001, vol. 1, pp. 381-384.

[11] G. Garg, V. Singh, J.R.P. Gupta and A.P. Mittal, "Optimal algorithm for ECG denoising using Discrete Wavelet Transforms", ICCIC, 2010, pp. 1-4.

[12] V. P. Dimri, N. Vedanti and S. Chattopadhyay, "Fractal analysis of aftershock sequence of the Bhuj earthquake: A wavelet-based approach", Current Science, May 2005, vol. 88, no. 10.

[13] Tae Hwan Lee and Byung Cheol Song, "De-noising algorithm using sparse 3D transform-domain collaborative filtering and adaptive soft thresholding", ISCE, 2011, pp. 128-131.