Measurement 43 (2010) 513–519

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

A SVM-based pipeline leakage detection and pre-warning system

Zhigang Qu *, Hao Feng, Zhoumo Zeng, Jingchang Zhuge, Shijiu Jin

State Key Laboratory of Precision Measuring Technology & Instruments, Tianjin University, Tianjin 300072, PR China

ARTICLE INFO

Article history:
Received 15 August 2008
Received in revised form 7 September 2009
Accepted 7 December 2009
Available online 4 January 2010

Keywords:
Pipeline leakage
Pre-warning
Distributed optical fiber
Support vector machine
Wavelet packet

ABSTRACT

A SVM-based pipeline leakage detection and pre-warning system is presented in this paper. In the system an optical cable is laid in parallel with a pipeline in the same ditch and three single mode optical fibers inside constitute the distributed vibration sensor. The sensor is based on Mach–Zehnder optical fiber interferometer and can detect the vibration signals along a pipeline in real time. Then the eigenvectors of vibration signals are extracted by "energy-pattern" method based on wavelet packet decomposition. Subsequently the vibration signals are recognized by support vector machine (SVM) through the features so that it can judge whether any abnormal event is taking place. If any abnormal event is found along a pipeline, the location is thus calculated. A series of trials in situ have been done, showing that the system is of good accuracy and real time performance both in recognition and locating.

1. Introduction

Nowadays, the prevalent methods to judge a pipeline leakage are evaluation of parameters such as pressure in pipeline, flow rate, and temperature obtained by common pipeline leakage detecting devices to infer whether a leakage has occurred [1]. These methods are liable to be influenced by the quality of the material transported and other factors [2]. Most importantly, warning of leakage exclusively comes after the occurred leakage, which may cause loss of life and property along with environmental pollution.

Aiming to monitor and locate the possible abnormal events (e.g. manual digging above a pipeline and illegal constructions, etc., which might cause a pipeline leakage) along pipeline before a leakage takes place, a new pipeline leakage detection and pre-warning system based on distributed optical fiber is constructed [3]. Aside from detecting and locating the leakage points, the system can also perform pre-warning monitoring and locating when the abnormal events are taking place.

How to judge whether abnormal events along a pipeline are taking place is one of the key techniques in the system. In this system the vibration signals caused by leakage or other abnormal events along a pipeline can be detected by the distributed optical fiber sensor in real time. Subsequently the features of detected signals can be extracted by "energy-pattern" method based on the wavelet packet decomposition. Finally SVM is employed as the classifier in the system, which can recognize the abnormal events along a pipeline effectively and in real time.

2. System principle

2.1. Measurement principle

As shown in Fig. 1, an optical cable is laid alongside the pipeline. In order to obtain the vibration signals along the pipeline [3], three single mode optical fibers in the optical cable are used to build the distributed vibration sensor on the basis of Mach–Zehnder optical fiber interferometer principle. Among the three fibers in the cable, two of them are sensing fibers; the third one is applied in signal transportation. The light waves in the two sensing optical fibers are converged into interfering signals, which are transmitted to a photodiode that transforms optical signals into electric signals. The signals are then processed by an amplifying and filtering circuit. Further process and

* Corresponding author.
E-mail address: qzgwin@gmail.com (Z. Qu).

0263-2241/$ - see front matter
doi:10.1016/j.measurement.2009.12.022

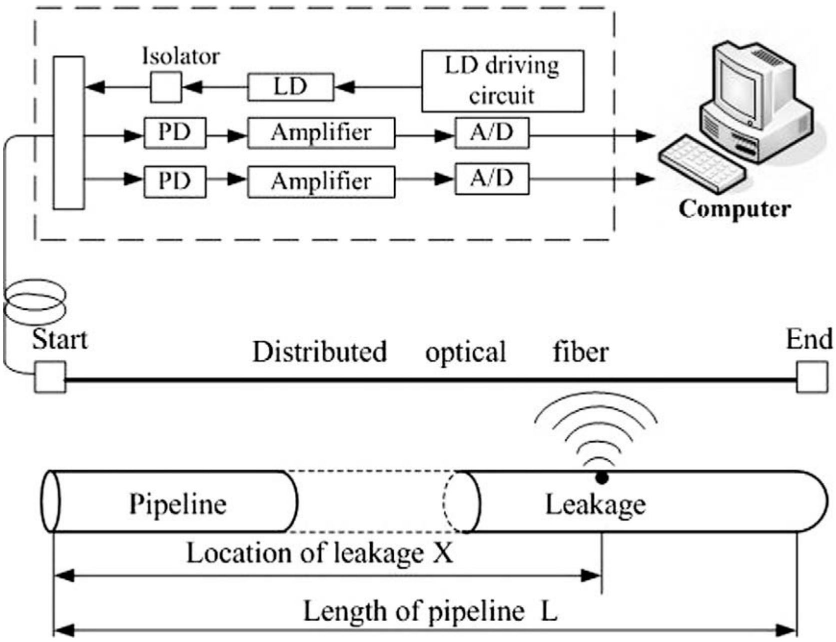

Fig. 1. System measuring principle.

analysis are needed after transmitting the signals to a computer through an A/D converter.

Due to the different permutation positions of the two sensing optical fibers in the optical cable, different stress and strain are generated on the two sensing optical fibers of the optical cable when the optical fibers are activated by vibrations. Accordingly, the two beams of coherent light wave undergo phase change respectively [4]. Light intensity of the two coherent light wave beams is presented as

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos[\Delta s(t) + \Delta\varphi], \tag{1}$$

where $\Delta s(t)$ is the difference of modulated phase of the two beams of coherent light wave, $\Delta\varphi$ is the difference of their initial phases, I_1 and I_2 are the squares of their amplitudes of vibration in the optical field. Suppose I_0 is the total light intensity transmitted into the two sensing optical fibers, and α is the mixed efficiency of the two coherent light waves, then

$$I(t) = I_0\{1 + \alpha\cos[\Delta s(t) + \Delta\varphi]\}. \tag{2}$$

Only if the light intensity of alternating current is taken into account, the above equation can be simplified as

$$I(t) = I_0\alpha\cos[\Delta s(t) + \Delta\varphi]. \tag{3}$$

The light intensity signals are transformed into electric current signals through photoelectric detector; therefore, the alternating current quantity of light current is

$$i(t) = KI_0\alpha\cos[\Delta s(t) + \Delta\varphi], \tag{4}$$

where K is the photovoltaic conversion ratio.

In Eq. (4), when $\Delta\varphi$ is a constant $\pi/2$, the slopes of photocurrent and detected phase have reached maximum value, which leads to the highest measuring sensitivity. $\Delta s(t)$ being a variable, a detected signal is the function of the difference of modulated phase of the two beams of coherent light wave $\Delta s(t)$. Therefore, by detecting any change of interfering optical signals in real time, the vibration signals along a pipeline can be detected through the distributed optical fiber sensor. In this way, the real time monitoring and pre-warning of pipeline leakage can be achieved.

2.2. Abnormal events locating principle

Locating is another key technique in pipeline monitoring. Once the abnormal events are found along the pipeline, the locations should be calculated with minimized error.

In this system, two beams of light waves simultaneously transmit in opposite directions (named Direction1 and Direction2) in both sensing optical fibers. As shown in Fig. 2, the two optical fibers in the cable generate stress and strain once a vibration signal caused by abnormal events (such as pipeline leakage or manual digging) appears along the pipeline. The stress and strain results in phase modulation of light waves at the point. Having created phase modulation, the light waves transmit to the two ends of the sensor along the optical fiber and interfere respectively at both ends. The arriving time of the two interference light waves at the PD1 and PD2 is different, thus the point where the abnormal event occurs can be accurately obtained in

$$X = L - \frac{v(t_2 - t_1)}{2}, \tag{5}$$

where X is the distance from the point to the start of the sensor (m), L is the length of the pipeline (m), t_1 and t_2 are time when the two photoelectric detectors at both ends detect the pipeline leakage signal respectively ($t_2 > t_1$), v is the transmission velocity of a light wave in the optical fiber

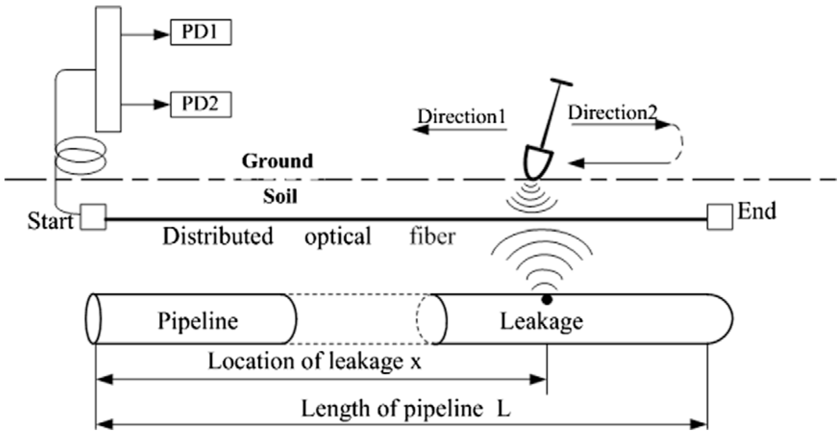

Fig. 2. Locating principle.

(m/s), in which v = c/n, c is the velocity of light in a vacuum (3×10^8 m/s), n is the index of optical fiber.

The interference light waves detected by PD1 and PD2 are strongly correlated [5]. As a result, the time difference of the two light waves can be accurately obtained by means of a correlation algorithm. The two output electronic signals by the two PDs can be demonstrated as follows respectively:

$$\begin{cases} x_1(t) = S(t) + n_1(t) \\ x_2(t) = \alpha S(t + \tau_0) + n_2(t), \end{cases} \tag{6}$$

where $S(t)$ is a signal of vibration, $n_1(t)$ and $n_2(t)$ stand for noise, α is the proportion factor, τ_0 is the time delay between the two detected signals. The time difference between the two signals can be obtained by means of relevant counting operations on the discrete values of the two interference signals; thus, the position of the accident is determined according to Eq. (5).

3. Eigenvector extraction of detected signals

Wavelet analysis is a kind of time-frequency analysis method especially suitable for non-stationary signals, which is a landmark in signal analysis field in recent years [6]. Wavelet packet analysis is a more accurate approach to decomposition than wavelet analysis. The decomposition of wavelet packet analysis on each level is conducted not only on the parts of low frequency, but also on those of high frequency. In this way can resolution be enhanced. Therefore, wavelet packet offers more application advantages [7].

The "energy-pattern" method based on wavelet packet decomposition is employed to extract the eigenvector of the vibration signals along the pipelines. Suppose the sampling frequency of the signals is $2f$, and then if j layer wavelet packet decomposition is carried out on the signals, 2^j frequency bands of equal width can thus be formed. The frequency width of every interval is $f/2^j$. After the decomposition of wavelet packet, the coefficient of j layer wavelet packet is $C^m_{j,k}, k = 0, 1, \ldots, 2^j - 1$ and m is the location in the wavelet packet space.

According to Parseval energy integral equation, we have

$$\int_{+\infty}^{-\infty} |f(x)|^2 dx = \sum |C_{j,k}|^2, \tag{7}$$

which demonstrates that the coefficients of wavelet packet have energy quantity.

Suppose $E_{j,k}$ is the signal energy of the frequency band at the node k of j layer, and then we have

$$E_{j,k} = \sum_m |C^m_{j,k}|^2. \tag{8}$$

And energy $E_{j,k}$ is normalized when

$$E = \sum_{k=0}^{2^j-1} E_{j,k} \quad (k = 0, 1, \ldots, 2^j - 1). \tag{9}$$

Then, we have

$$E'_{j,k} = E_{j,k}/E, \tag{10}$$

where E'_{jk} is the energy obtained after normalization.

Some frequency bands of detected vibration signals are chosen for monitoring, where the changes of normalized energy are most obvious. Then the normalized energy of the selected frequency bands constitutes the eigenvector for the successive input into a classifier.

4. Recognition method based on SVM

The traditional learning machines are mainly based on empirical risk minimization (ERM) principle [8] and the empirical risk will converge at expected risk when the number of training data converges at infinity. Actually the ERM is built on the assumption that the number of the training data converges at infinity [9]. In practical projects the data for training are normally limited so that there are problems in the process of traditional learning machines, for example the existence of many local minima solutions, suffering from "curse of dimensionality" problems for high dimensional input space, etc.

SVM is a novel learning method derived from the statistical learning theory, which shows superior performance compared with traditional learning machines. The core idea of SVM is to map the original pattern space into the high dimensional feature space by some non-linear mapping functions so that the optimal separating hyper-plane can be constructed in the feature space. Thus, the non-linear problem in low dimensional space corresponds to the linear problem in the high dimensional space [10].

Binary classification is the basis of a multi-classification problem. Take binary classification [11] as an example to have a brief illustration on the principle of SVM. Suppose $(\mathbf{x}_i, y_i)$ are linearly separable data, in which $\mathbf{x}_i \in R^n, y \in \{-1, 1\}, i = 0, 1, \ldots, l$. As for SVM, the aim of data training is to construct a function to classify the testing data as correctly as possible. The function of hyper-plane for binary classification problem is

$$y_i[(\mathbf{w} \cdot \mathbf{x}_i) + b] - 1 \geqslant 0, i = 1, 2, \ldots, l. \tag{11}$$

Then constructing the optimal separating hyper-plane can be summarized as the problem below,

$$\begin{aligned} &min \quad \Phi(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 = \frac{1}{2}(\mathbf{w} \cdot \mathbf{w}), \\ &s.t. \; y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geqslant 0, i = 1, 2, \ldots, l, \end{aligned} \tag{12}$$

which is a classical convex optimization problem and can be simplified by converting the problem with KKT conditions into the equivalent Lagrange dual problem and be transformed to

$$\begin{aligned} &max \quad W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j), \\ &s.t. \quad \sum_{i=1}^{l} y_i \alpha_i = 0, \alpha_i \geqslant 0, i = 1, 2, ..., l. \end{aligned} \tag{13}$$

As a result, by solving the dual optimization problem, the coefficient α^* can be obtained, which leads to decision function

$$f(\mathbf{x}) = \mathrm{sgn}\left\{\sum_{i=1}^{l} \alpha_i^* y_i (\mathbf{x}_i \cdot \mathbf{x}) + b^*\right\}. \tag{14}$$

As to linearly non-separable problem, a positive slack variable $\xi_i \geqslant 0$ is introduced into the constraints.

For non-linear cases, the input vector is mapped into a high dimensional space, and the optimal hyper-plane can be constructed in the high dimensional space to solve such non-linear problems.

As a result, a classification problem can be transformed into a quadratic programming problem by SVM. Lots of algorithms in optimization theory store the kernel matrix in the memory so that for large datasets there will be many system resources occupied. To solve that problem sequential minimal optimizer [12] is employed in training the dataset in this paper.

Recognizing the abnormal events along pipelines is a typical multi-classification problem. Among several methods [13] that have been used for multi-class classification, "one-against-one" method [14] is more suitable for practical use than others. In this study, the "one-against-one" method is employed in this paper to recognize the typical abnormal events along the pipelines.

5. Field trials

Basically, two types of trials have been done to estimate the performance of the system. One is the abnormal events identification trial in the China Special Equipment Inspection and Research Center pressure vessel experiment site, in which some abnormal events have been done to test the identification performance of the system. The other one is abnormal events location trial along a 35.149 km-long product oil pipeline, which has been done to test the location performance of the system.

5.1. Abnormal events identification trial

In this trial, three kinds of abnormal events are created to test the identification performance of the system and a 150 m-long gas pipeline with a diameter of Φ159 is used. The testing optical cable is laid parallel to the pipeline in the same ditch, which is approximately 1 m below the ground and with 0.5 m space above the pipeline.

Three cases are created respectively, which are gas leakage, manual digging and human walk above the pipeline. When gas leaks from the pipeline in the experiment, the pressure of the gas is 0.8 MPa; the diameters of holes on the surface of the pipeline in the gas leakage experiment are 1–5 mm. Manual digging and human walk are created above the pipeline within 1 m beside the pipeline at both sides.

In this trial, the specification of optical fiber is G.652 and an LD light source, whose wave length is 1550 nm and the power is 1 mW, is employed. A pair of InGaAs photodiode is used as the photoelectric detectors with a 0.1 ns reaction velocity. NI PCI-6132 is used for two channel synchronous sampling. The software for data acquisition and analyzing is realized by LabVIEW, with a view to exhibiting the waveform in real time, online signal analysis and calculations.

The detected signal eigenvector of three cases can be extracted by "energy-pattern" method presented in Section 3. An eigenvector is a vector made up of 8 elements standing for normalized energy in 8 sensitive frequency bands. The signals obtained in the trial and their eigenvectors are shown in Fig. 3a–c. In each figure, the top one is the origin signal of each case while the bottom one is its eigenvector.

Twenty samples from or each case are chosen at random for training the SVM. Then ten samples are chosen randomly to test the trained SVM. During the SVM training process, "one-against-one" method is employed. The multi-class SVM classifier trained is shown in Fig. 4, which is projected to a 2D space based on the first two elements of an eigenvector. In Fig. 4 the samples with circle around are support vectors and two axes are the value of the first two elements in an eigenvector.

The recognition results are shown in Table 1, in which feature1–feature8 are the eight elements of an eigenvector. Three kinds of abnormal events are marked as 1, 2 and 3. In the table there are thirty samples, which are numbered 1–10, 11–20 and 21–30 corresponding to gas leakage, manual digging and walk respectively.

The results of the test exhibit the good performance of the trained SVM expect one error in the test in which sample 12 has been recognized incorrectly as walk, which should be manual digging. The time elapsed is only 0.06s during the whole process including training and test under 1.86 GHz CPU, 1G RAM environment.

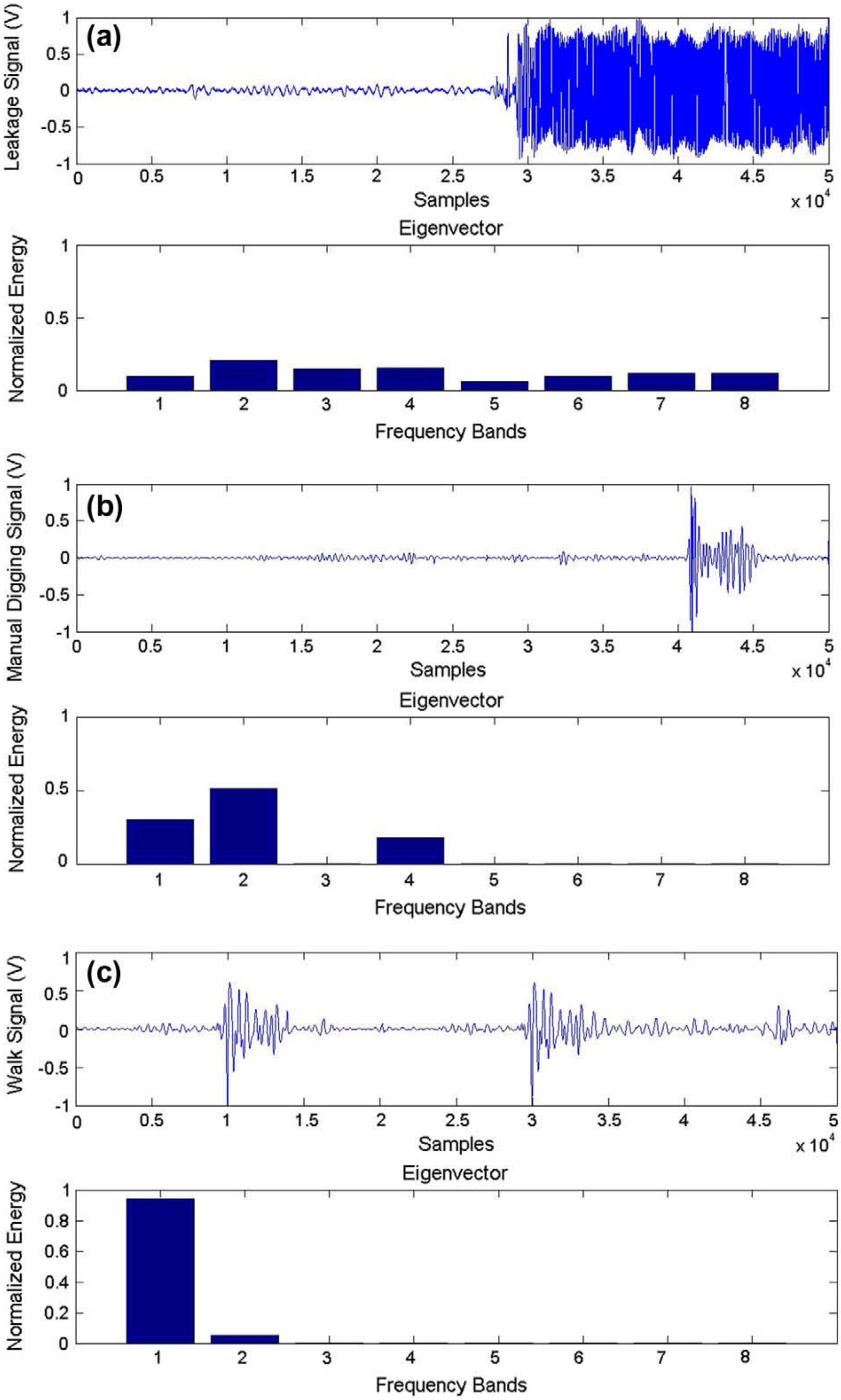

Fig. 3. (a) Waveform and eigenvector of gas leakage signal; (b) waveform and eigenvector of manual digging signal; (c) waveform and eigenvector of human walk signal.

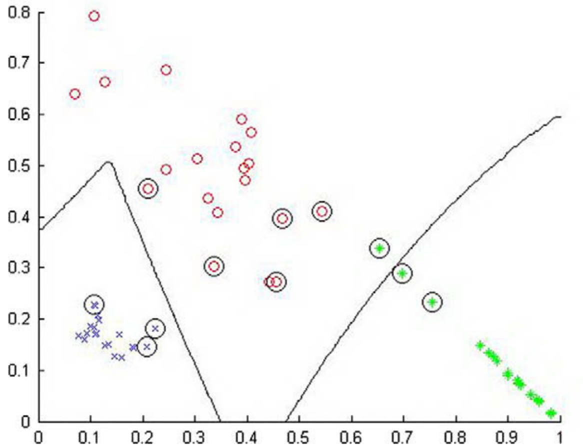

Fig. 4. SVM classifier after training.

5.2. Abnormal events location trial

A series of location trial are carried out along a 35.149 km-long product oil pipeline with a diameter of 426 mm. In the experiments many manual digging were made above the pipeline to test the locating performance of the system as show in Fig. 5.

Several points are chosen for testing the locating performance of this system, which shows the system is of good repeatability and precision with error around ±200 m. For example 1 m from the starting point of the sensor is se-

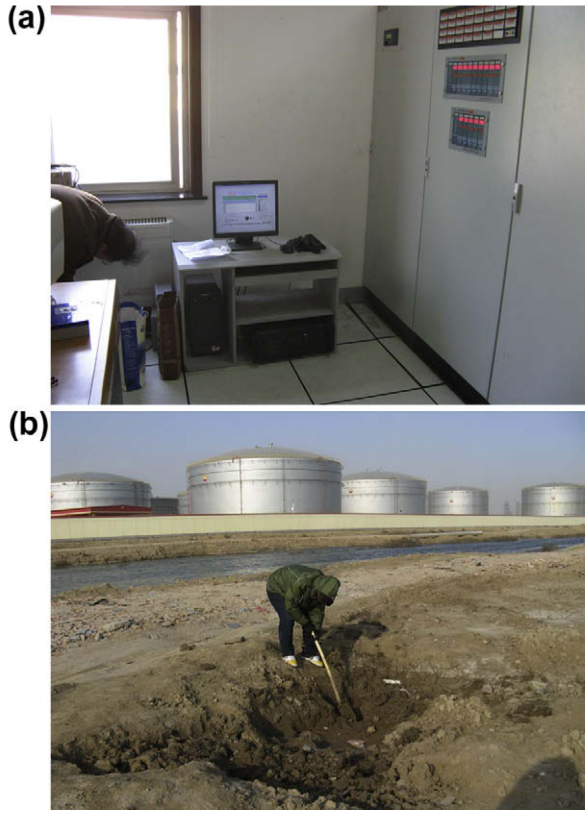

Fig. 5. (a) The system is monitoring the pipeline in the station control room; (b) manual digging in location trial.

Table 1
Test results of recognizing abnormal events by SVM.

Sample	Feature1	Feature2	Feature3	Feature4	Feature5	Feature6	Feature7	Feature8	Recognition result
1	0.0978	0.2036	0.1507	0.1527	0.0607	0.0972	0.1196	0.1178	1
2	0.0972	0.1504	0.1468	0.1474	0.0793	0.1267	0.1105	0.1416	1
3	0.0955	0.2048	0.1518	0.1536	0.0608	0.0975	0.1200	0.1159	1
4	0.0949	0.1477	0.1490	0.1475	0.0800	0.1275	0.1106	0.1428	1
5	0.0698	0.1514	0.1169	0.1508	0.1255	0.0901	0.1279	0.1675	1
6	0.0943	0.1617	0.1504	0.1298	0.0967	0.1157	0.1202	0.1311	1
7	0.1621	0.1314	0.1152	0.1158	0.0952	0.1039	0.1217	0.1548	1
8	0.1191	0.1334	0.1321	0.1520	0.1125	0.1168	0.1206	0.1135	1
9	0.1283	0.1125	0.1379	0.1485	0.0853	0.1215	0.1349	0.1311	1
10	0.0809	0.1500	0.1355	0.1624	0.1052	0.1262	0.1406	0.0993	1
11	0.1065	0.7926	0.0002	0.1003	0.0000	0.0002	0.0000	0.0002	2
12	*0.6318*	*0.3394*	*0.0006*	*0.0278*	*0.0000*	*0.0000*	*0.0002*	*0.0003*	*3*
13	0.3791	0.5361	0.0015	0.0827	0.0000	0.0001	0.0002	0.0002	2
14	0.1285	0.6634	0.0112	0.1954	0.0000	0.0003	0.0006	0.0006	2
15	0.2102	0.4551	0.1184	0.1739	0.0000	0.0005	0.0343	0.0076	2
16	0.5443	0.4102	0.0012	0.0439	0.0000	0.0000	0.0002	0.0002	2
17	0.3261	0.4370	0.0118	0.2177	0.0000	0.0004	0.0026	0.0044	2
18	0.2439	0.4929	0.0404	0.2199	0.0000	0.0006	0.0009	0.0014	2
19	0.1756	0.4528	0.0080	0.3608	0.0000	0.0006	0.0007	0.0015	2
20	0.2327	0.5005	0.0438	0.2073	0.0000	0.0007	0.0104	0.0046	2
21	0.7544	0.2321	0.0003	0.0132	0.0000	0.0000	0.0000	0.0000	3
22	0.8721	0.1270	0.0002	0.0008	0.0000	0.0000	0.0000	0.0000	3
23	0.6954	0.2930	0.0002	0.0113	0.0000	0.0000	0.0000	0.0001	3
24	0.9603	0.0381	0.0001	0.0015	0.0000	0.0000	0.0000	0.0000	3
25	0.9840	0.0138	0.0002	0.0020	0.0000	0.0000	0.0000	0.0000	3
26	0.8633	0.1349	0.0003	0.0014	0.0000	0.0000	0.0000	0.0000	3
27	0.9166	0.0824	0.0001	0.0010	0.0000	0.0000	0.0000	0.0000	3
28	0.9341	0.0636	0.0002	0.0021	0.0000	0.0000	0.0000	0.0000	3
29	0.8476	0.1488	0.0002	0.0033	0.0000	0.0000	0.0000	0.0000	3
30	0.7030	0.2818	0.0006	0.0146	0.0000	0.0000	0.0000	0.0000	3

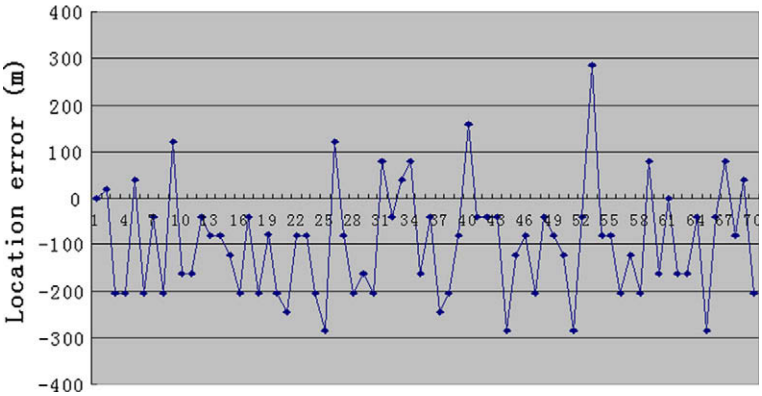

Fig. 6. Locating results for the test point.

lected for a location test point and the location errors (m) for 70 location tests are shown in Fig. 6, in which *x*-axis is test record number and *y*-axis is location error (m).

The system can be designed to monitor a long pipeline up to 50 km effectively with normally ±200 m (or better, regarding the application conditions) location precision in real time.

6. Conclusions

A pipeline leakage detection and pre-warning system is presented in this paper. In addition to the traditional leakage detection, the system can also perform pre-warning monitoring and locating when the abnormal events are taking place along a pipeline. The system can recognize the abnormal events along a pipeline with good recognition correct rate (normally ⩾95%) in real time and locate the abnormal events along a pipeline with good precision (normally ±200 m) so as to prevent potential loss as much as possible. Now the system is successfully operating on four pipelines with a total length of around 150 km.

The pipeline leakage detection and pre-warning system is safe, reliable and corrosion proof with excellent electric insulativity. Furthermore, the performance of the system will not be affected by the property of material transported in a pipeline. Thanks to the above favorable features, this system will be more widely applied in petrochemical industry and other comparable environments with strong electromagnetic interference, flammability, explosibility, strong corrosion and so on.

Acknowledgment

This work is funded by the National Natural Science Foundation of China (No. 60534050).

References

[1] Jie Yang, Gui-zeng Wang, Leak detection and location methods for gas transport pipelines Control and Instruments, Chemical Industry 31 (3) (2004) 1–5.
[2] Shi-dong Zhou, Zhi-min Wu, Ming Wu, Summary of detection technology for petroleum pipeline, Petroleum Engineering Construction 29 (2003) 6–10.
[3] Zhigang Qu, Shijiu Jin, Yan Zhou et al., Study on the distributed optical fiber pipeline leakage pre-warning system and the method of signal analysis, in: Proceedings of the 6th International Pipeline Conference, vol. 3(B), Calgary, Canada, 2006, pp. 677–681.
[4] Yan Zhou, Shi-Jiu Jin, Yun-Chao Zhang, et al., Study on the distributed optical fiber sensing technology for pipeline leakage detection, Journal of Optoelectronics Laser 16 (8) (2005) 933–938.
[5] Yan Zhou, Shijiu Jin, Yunchao Zhang, et al., Distributed optical fiber sensing technology for pipeline leakage detection and location, Acta Petrolei Sinica 27 (2) (2006) 121–124.
[6] I. Daubechies, Ten Lectures on Wavelets, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, USA, 1992.
[7] Guo-An Yang, Bing-Lin Zhong, Research on the extraction method of time-domain symptoms based on wavelet packet decompositions of mechanical fault signal, Journal of Vibration and Shock 20 (2) (2001) 25–28.
[8] V. Vapnik, The Nature of Statistical Learning Theory, Springer Verlag, New York, 1995.
[9] Xuegong Zhang, Introduction to statistical learning theory and support vector machines, Acta Automatica Sinica 26 (1) (2000) 32–42.
[10] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (1995) 273–297.
[11] C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery 2 (1998) 121–167.
[12] Jianmin Li, Bo Zhang, Fuzong Lin, Training algorithms for support vector machines, Journal of Tsinghua University 43 (1) (2003) 120–124.
[13] E. Bredensteiner, K. Bennet Multicategory, Classification by support vector machines, Computational Optimizations and Applications 12 (1999) 53–79.
[14] Chih-Wei Hsu, Chih-Jen Lin, A comparison of methods for multi-class support vector machines, IEEE Transactions on Neural Networks 13 (2) (2002) 415–425.