

Research Article

Hybrid Feature Extraction of Pipeline Microstates Based on Φ -OTDR Sensing System

Yanzhu Hu, Zhen Meng , Xinbo Ai , Han Li, Yu Hu , and Huiyang Zhao

Beijing University of Posts and Telecommunications, No. 10 Xi Tu Cheng Road, Beijing 100876, China

Correspondence should be addressed to Zhen Meng; zmeng@bupt.edu.cn and Xinbo Ai; axb@bupt.edu.cn

Received 12 March 2019; Revised 17 July 2019; Accepted 4 August 2019; Published 22 September 2019

Academic Editor: Manuel Pineda-Sanchez

Copyright © 2019 Yanzhu Hu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a general integration method which can effectively describe the characteristics of pipeline leakage and help distinguish multiple pipeline microstates. Since the rapid development of Φ -OTDR in recent years, this technology has been applied to more and more fields, such as fiber optic safety monitoring, seismic monitoring, and structural health monitoring. Among them, Φ -OTDR has the characteristic of continuous full-scale monitoring in pipeline monitoring, but there are few researches on pipeline state characteristics at present. In this paper, based on the analysis of the pipeline state with Φ -OTDR technology, a method of extracting multiple microstates of pipelines is proposed. This method combined with the peak-to-average power ratio, short-term interval zero crossing, and fractal characteristics in the frequency domain can effectively characterize the microstate of pipes and provide support for identification of more microstates of pipelines. These features reflect the common characteristics of leaks in gas pipelines and liquid pipelines. Meanwhile, their combination features can represent the small differences in pipeline states. The experimental results show that the method can effectively characterize the microstate information of the pipeline, and the recognition rate of the hybrid feature under two kinds of pipeline leakage and multipressure conditions reaches 91% and 83%.

1. Introduction

Pipeline is an important infrastructure in modern society. Because of environmental corrosion, aging, and man-made damage, pipelines are prone to leak. Pipeline leakage is an important security issue, which directly affects the normal operation of life and production. Because huge losses to normal life and production are caused by pipeline leakage, monitoring of the pipeline is imperative [1–4]. In the aspect of pipeline monitoring, point sensors are easily disturbed by humidity, low temperature, electromagnetic radiation, and other factors, which makes it difficult to identify the pipeline states accurately. Φ -OTDR is a kind of distributed optical fiber sensing technology which collects the disturbance signals surrounding the optical fiber. Based on the advantages of high sensitivity and fast response, Φ -OTDR can overcome the interference of humidity, low temperature, and electromagnetic radiation and will be the trend of pipeline monitoring [5–6]. Because of its high accuracy and high spatial resolution, Φ -OTDR is widely used in fields of fiber optic safety monitoring, perimeter security, submarine cable safety, and structural health monitoring [7].

At present, the wavelet has been used to analyze pipeline leakage states through single threshold comparison [8]. Qu et al. proposed an “energy-pattern” method based on the wavelet and support vector machine (SVM), which can recognize whether any abnormal event is taking place [1]. Wang used the fractal box dimension and improved approximate entropy to distinguish leakage and interference signals. However, only the anomaly in the pipe is studied, and recognition results are greatly influenced by the noise [9]. Meanwhile, the development of optical fiber perimeter security is fast. Huang et al. proposed a high-resolution scheme combining empirical mode decomposition with kurtosis, whose average accuracy reached 89% [6]. After that, a high-resolution scheme combining empirical mode decomposition with hybrid features was proposed. In such methods, EMD works in an iterative mode and consumes excessive computation [10]. However, it requires a lot of research to be done in monitoring and representing the states of pipelines by using Φ -OTDR.

The pipeline leakage signal, which is a kind of nonlinear and nonstationary signal, is affected by multiple factors. The

signal based on the Φ -OTDR sensing system is extremely sensitive to the monitoring events with significant changes in magnitude, which requires features to adapt to strong noise interference and high hardware cost in signal processing, while there is fading in most Φ -OTDR sensing systems, resulting that features in the time domain find it difficult to represent the effective information. Moreover, the amount of monitoring data by the Φ -OTDR sensing system is huge, and long time of operation is not suitable for engineering application. From the above researches, the sole feature extraction is hard to represent the enough effective information of pipeline leakage, which is not enough to support multistate or even microstate recognition [11]. This paper proposes a method of extraction of pipeline microstates based on the Φ -OTDR sensing system. The method combines the peak-to-average power ratio (PAR), short-term interval zero-crossing rate (STI-ZCR), and fractal characteristics in the frequency domain (FF) [12, 13]. The peak-to-average power ratio can reflect the overall linear information of the signal over a period, which responds to signal distortion effectively and does not cause any hardware processing cost. Zero crossing reflects the sensitivity of the signal energy to the amplitude and describes the frequency-domain information of signals from the time domain. Moreover, we extend the range of frequency-domain information by improving this feature, while retaining the convenience of calculation. As the fading of Φ -OTDR, the direct fractal feature of the time domain is not enough to characterize the sufficient information of the target event. But the leakage signals based on Φ -OTDR have self-similarity at different scales [14, 15]. Therefore, the fractal characteristics in the frequency domain are chosen to reflect the subtle characteristic of the signal frequency. The hybrid features consist of the peak-to-average power ratio, short-term interval zero crossing, and fractal characteristics in the frequency domain, which not only reflect leakage states of pipelines but also represent more detailed information for multiple microstates of pipelines. Although these features cannot directly obtain the hiding pipeline state information, machine learning methods, such as support vector machine (SVM), C5.0, and random-forest algorithm, can be used to recognize the microstates of pipelines [16, 17]. The experimental results prove this method based on hybrid features which can effectively improve the pipeline leakage event recognition accuracy with the classifier of random forest, whose average accuracy is above 91% on two kinds of pipeline leakage and 83% on four microstates.

This paper is organized as follows: Section 1 introduces the feature extraction research on pipelines based on Φ -OTDR. Section 2 introduces the related work to extract features of the pipeline. Section 3 describes the method of hybrid feature extraction and the classifier algorithm. Section 4 presents the method application and details of experiments. Finally, Section 5 summarizes the main results and concludes the whole article.

2. Preview Works of Signals Collection

Figure 1 shows the schematic diagram of the pipeline leak experiment. The Φ -OTDR sensing system mainly consists of a laser, acousto-optic modulator, erbium-doped optical fiber amplifier, photodetector, analog-to-digital converter, data

acquisition card, and personal computer. We have two exactly same pipes, and we make a leak hole on one of them which is shown as the upper pipeline in Figure 1. Through the Φ -OTDR sensing system, the leakage data and the noise with different types of inputs are collected. Also, we control the pipe output by the control valves and pressure gauge and collect the pipeline leakage data at different states through the Φ -OTDR sensing system.

The domain morphologies of the leakage signals are shown in Figure 2. The pipeline leakage data are obtained through the Φ -OTDR system. The figure is a pair of signals collected under the same conditions: gas leakage signals and gas noise signals.

From the above analysis, we can find the following three aspects:

- (a) The pipeline leakage signal is a continuous and nonstationary signal. There is no obvious difference between the leakage signal and the noise signal, which needs feature extraction to obtain the effective information hidden in the signal.
- (b) The pipeline data acquisition environment is complex, and single feature finds it difficult to fully characterize pipeline leakage information and distinguish the tiny states of pipelines.
- (c) Because of the correlation fading effect of Φ -OTDR, time-domain morphological features cannot effectively represent pipeline information to recognition. This requires that feature extraction should consider both computational efficiency and conventional time-domain features.

Through large researches of pipeline detection based on Φ -OTDR and feature engineering, the hybrid features of the peak-to-average power ratio, short-term interval zero crossing, and fractal characteristics in the frequency domain can effectively reflect the accurate information of pipeline leakage and help identify the leakage, types, and multiple microstates of pipelines.

3. The Proposed Method

Because the data of pipeline leakage based on Φ -OTDR are nonstationary and because of its related fading effect, the conventional time-domain characteristics of leakage data cannot be used. In order to solve this problem and support the pattern recognition for multiple microstates of pipelines, this paper also takes into account the energy characteristics, linear characteristics, and local self-similarity of the signal and chooses the peak-to-average power ratio, short-term interval zero crossing, and frequency fractal to reflect the common features and details of the pipeline leakage [18, 19]. The overall block diagram of the method is shown in Figure 3.

This paper only discusses the identification of point data and the analysis models of point data which could be applied to the monitoring and identification of the whole fiber optic cable. So, we select several stable points on the two experimental pipelines as the dataset and perform data

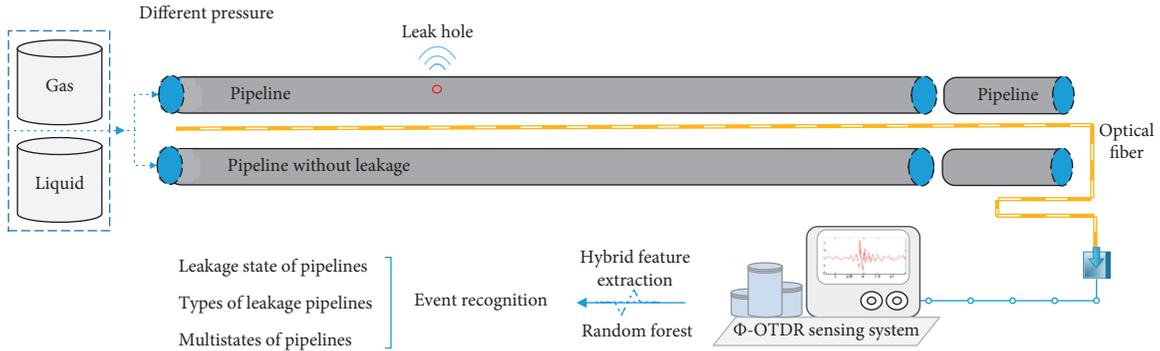


FIGURE 1: Schematic diagram of the pipeline leak experiment.

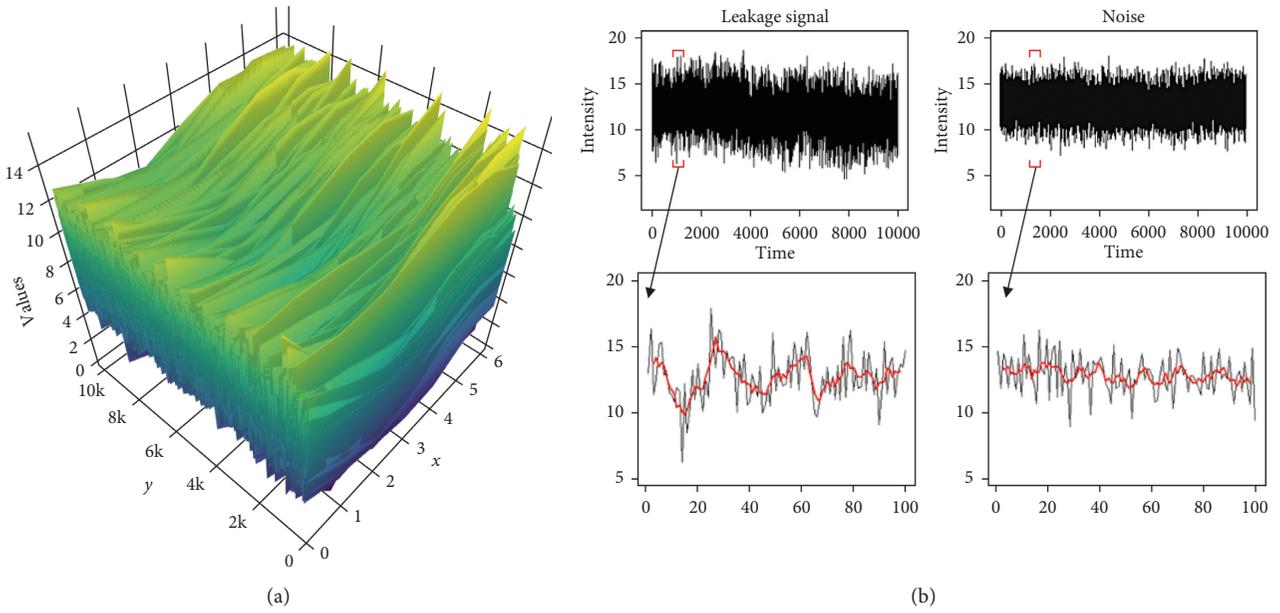


FIGURE 2: Leakage signals of gas pipelines. (a) 3D leakage signal of gas pipelines. (b) Time-domain characteristics of the leakage signal.

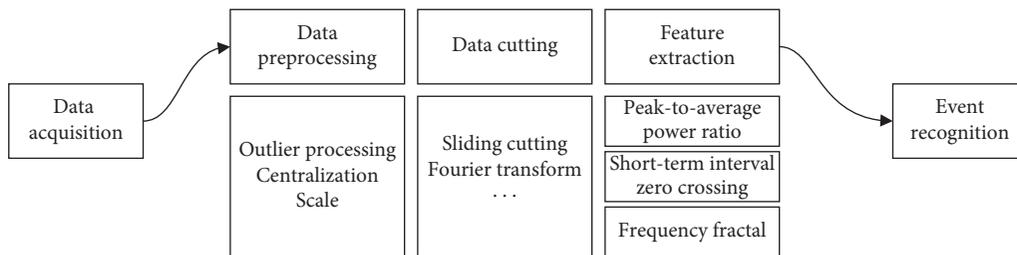


FIGURE 3: Diagram of the hybrid feature extraction method.

preprocessing. In order to evaluate the hybrid features, we have introduced the classifier of random forest.

3.1. Data Preprocessing

3.1.1. Outlier Processing. In the original signal, there is some single point impact noise generated by the equipment itself. In order to reduce the impact of single point on feature calculation, we replace these outlier points by the window mean values. Steps are as follows.

The first step is to find all the offset points by analyzing the global distribution of the data and calculating the maximum and minimum values:

$$\begin{aligned}
 \text{minimum} &: Q1 - k(Q3 - Q1), \\
 \text{maximum} &: Q3 + k(Q3 - Q1), \\
 \text{outlier} &: P < \text{minimum} \text{ or } P > \text{maximum},
 \end{aligned} \tag{1}$$

where $Q3$ is the third quartile, $Q1$ is the first quartile, and $k=3$.

The second step is to replace these points by the window mean values. The window length is usually set according to the actual requirement.

3.1.2. Centralization. Centralization is to subtract the mean from the raw data, and the processed data fluctuate near zero. This operation is to facilitate the calculation of the improved peak-to-average power ratio and short-term interval zero-crossing rate.

$$\text{sgn}'(x) = \begin{cases} \text{sgn}(x) - a, & \text{sgn}(x) \geq 0, \\ \text{sgn}(x) - (-a), & \text{sgn}(x) < 0. \end{cases} \quad (2)$$

In general, a takes the mean value of $\text{sgn}(x)$.

3.1.3. Standardization and Normalization. In order to reduce the influence of signal amplitude difference and focus on the essence of the leakage of pipelines, the acquisition data should be standardized and normalized firstly. This step can eliminate the errors in the acquisition of optical time-domain reflectometry and ensure the subsequent computation speed. The formulas are as follows:

$$\begin{aligned} c_{\text{Standardization}}(x) &= \frac{x - \mu}{\sigma}, \\ c_{\text{Normalization}}(x) &= \frac{c(x) - \min[c(x)]}{\max[c(x)] - \min[c(x)]}. \end{aligned} \quad (3)$$

3.2. Data Cutting. The collected pipeline leakage signal is a time series of amplitude. In order to facilitate feature extraction, the time series needs to be processed, mainly using the method of cutting the window. Because the pipeline leakage state is continuous and transformed gradually, in order not to omit the pipeline leakage information, we use the sliding step method to extract the features. The steps of cutting the window are as follows:

- (a) Window length N and sliding step length S are set. We find the best effect usually at $S=1/2N$. The window length is limited by the sampling frequency, and the window must contain more than 2 peaks and valleys.
- (b) The peak-to-average power ratio and short-term interval zero-crossing rate are calculated by sliding windows one by one. More details are given in the next section.
- (c) Fourier transform is used to obtain the frequency-domain amplitude sequence of the original signal and then get the fractal characteristics by the sliding window.

3.3. Feature Extraction. The pipeline leakage signal, which is a nonlinear, nonstationary signal with the fading phenomenon, is affected by multiple-factors coupling. Hybrid features consist of peak-to-average power ratio in the field of signal processing, short-term interval zero-crossing rate in the field of speech processing, and frequency fractal in the

field of image processing. However, the leakage signal and the nonleakage signal have an intrinsic relationship under the intensity characterization but are not directly related to the strength. Therefore, the relative transformation characteristics of the pipeline signals over a time period should be paid more attention [12, 20]. These three features are based on the time-frequency analysis of the pipeline leakage signal. This combination covers the common features and local detailed information of pipelines. Therefore, the hybrid features are a combination of common and individual characteristics that can accurately reflect the leakage [21].

3.3.1. Peak-to-Average Power Ratio. PAR is a kind of feature extraction method in the field of signal processing, which reflects the overall linear characteristics of signals. The large PAR represents a larger relative peak and the linear range of the signal [22]. When it comes to the field of feature analysis, PAR will not cause any hardware processing costs, which also avoids the impact of the fading. In the case of strong noise interference, the abnormal value of an extremely strong noise signal can be processed, the strongest part of noise can be removed, and the stable part of the signal peak value can be selected. By this way, the contrast of features can be increased. PAR is regarded as the reciprocal estimation of signal-to-noise ratio. According to the characteristics of the leakage signal, its definition is modified as follows:

Step 1. Get the absolute value $|\text{sgn}(x)|$.

Step 2. Take the mean value $|x|_{\text{peak}}$ of over 90% quantiles in the signal $|\text{sgn}(x)|$.

Step 3. Take the mean value x_{rms} of $|\text{sgn}(x)|$.

Step 4. Obtain the peak-to-average power ratio.

$$\text{par} = \frac{|x|_{\text{peak}}}{x_{\text{rms}}}. \quad (4)$$

3.3.2. Short-Term Interval Zero-Crossing Rate. In the field of speech signal processing, short-term average zero-crossing rates are commonly used for endpoint detection of unvoiced and voiced speech. The zero-crossing rate is the ratio of the sign change of a signal [23]. If it is a sinusoidal signal, its average zero-crossing rate is twice the signal frequency divided by the sampling frequency, and the sampling frequency is fixed. Therefore, the zero-crossing rate can describe the frequency information of the signal from the time domain. When the pipeline continues to leak, the leakage signal is like the unvoiced signal, and the noise signal in the pipeline leakage is like the voiced signal. The pipeline leakage signal is a relatively stable continuous signal when it is stably leaked. In view of the unique characteristics of the pipeline leakage signal and the short-time zero-crossing rate calculation characteristics in the field of speech processing, we improve the short-time zero-crossing rate, retain the convenience of short-time zero-crossing rate calculation, and extend the frequency range information characteristics of the signal from the time domain, which is called the short-term interval zero-crossing rate. This paper calculates the ratio of symbol change of the signal through the interval $[-a, a]$. The improved calculation steps are as follows:

Step 1. Center the original signal by formula (2) and obtain $\text{sgn}'(x)$.

Step 2. Get the absolute value $|\text{sgn}'(x)|$.

Step 3. Set $a = 0.25$. Take the signal greater than $1 - a$ of $|\text{sgn}'(x)|$ and $\text{sgn}''(x)$.

Step 4. Seek the short-term interval zero-crossing rate of $\text{sgn}''(x)$.

$$\text{zcr} = \frac{1}{N} \sum_{m=0}^{N-1} \|\text{sgn}''[x(m+1)] - \text{sgn}''[x(m)]\|. \quad (5)$$

3.3.3. Frequency Fractal. Fractal is a universal characteristic of complex things in nature. The fractal dimension is used to measure the self-similarity of curves in fractal theory in general. The degree of irregularity of nonstationary random vibration signals can be described by the fractal dimension. There are many methods of calculating the fractal dimension, such as the Hausdorff dimension, capacity dimension, and information dimension. Leakage signals and nonleakage signals have an intrinsic relationship under the intensity characterization but are not directly related to the strength; that is, the pipeline leakage signal has a certain self-similarity at different intensities. We extract the frequency fractal to represent the local information in the frequency domain, abbreviated as FF. The information dimension can reflect the inhomogeneity of the distribution of the set to be tested, but the calculation is relatively complex. This method overcomes the problem that it is difficult to quantitatively describe the feature in the time domain based on the Φ -OTDR sensing system and extracts the leakage signal and the noise signal information in the microscopic analysis angle. Based on the analysis and comparison, the fractal dimension is calculated based on the idea of “covering” [24, 25].

$$D = \lim_{x \rightarrow 0} \frac{\lg N(x)}{\lg(1/x)}, \quad (6)$$

where x is the closed interval and $N(x)$ represents the number of sets covering the target set with a subset. Its definition is modified as follows:

Step 1. Get frequency-domain data through the Fourier transform of raw data.

Step 2. Acquire the frequency-domain amplitude sequence.

Step 3. Calculate the fractal by the sliding window.

3.4. Random-Forest Classifier. Random forest is a kind of integrated machine learning method. It uses the random resampling technique bootstrap and the random node splitting technique to build decision trees, whose parameters are independent and identically distributed vectors. Under the given independent variable, each decision tree classification model has one vote to select the optimal classification results. RF has the ability to analyze complex interactions and is robust to the data with noise and missing values [21, 26]. At the same time, it has a fast learning speed.

The implementation process of the random-forest algorithm is as follows:

Step 1. N training sets are extracted from the original datasets with the bootstrap sampling method, and a classification and regression tree is built for each training set. The size of each training set is about 2/3 that of the original dataset.

Step 2. In each node of the tree, m features are randomly selected from all the n characteristics ($m \leq n$). After computing the amount of information contained in each feature, one of the m features is selected to split the nodes.

Step 3. There is maximum growth of each tree without any pruning.

Step 4. N decision trees constitute a random forest. When new data enter random forests, the results of all decision trees are gathered and voting results are used to determine the classification result.

4. Experiments

4.1. Experimental Environment and Parameters. To verify our method of hybrid feature extraction, we designed an experimental environment. The experiment scenes are shown in Figure 4. In the upper left corner, there is a pressure gauge, and in the upper right corner, there is the Φ -OTDR acquisition system. The laser source is a 1550 nm distributed feedback laser. In order to ensure universality and stability of the detection, instead of being around the pipe, the fiber optic cable (400 m long with single mode) is parallel to the pipeline in the lower picture. The white part in the picture is a 2-core optical cable with single mode, and the silver section is the pipelines. The diameter of the leak hole on the top pipeline is 4 mm, while the pipe diameter is 19 mm.

The length, diameter, and other parameters of the pipelines are all the same. Firstly, the pipe pressure was stabilized at 0.4 MPa, and we choose the gas or liquid as the input for the two pipelines and take turns to collect the leakage data from the upper pipeline and the noise from the lower pipeline through the Φ -OTDR sensing system. Then changing the type of input, the above steps are repeated. Finally, we change the pressure to get more microstates of pipelines, such as 0.5–0.4 MPa, 0.4–0.3 MPa, 0.3–0.2 MPa, and 0.2–0.1 MPa. The sampling rate of the Φ -OTDR system was set as $f_s = 10$ kHz, and the recording duration of each trial was set as 10 s.

After the data acquisition and the preprocessing, we extract the hybrid features and build the random-forest model, which helps us recognize the leakage states and the type of the pipeline. The length of the cutting window is 1000, and the length of the sliding step is 500. Each type of features belongs to more than 700 groups in the following sections.

4.2. Leakage State Characterization. In this section, we extract the hybrid features to recognize the leakage state of pipelines. With the gas leakage signal, we obtain the hybrid features, as shown in Figure 5.

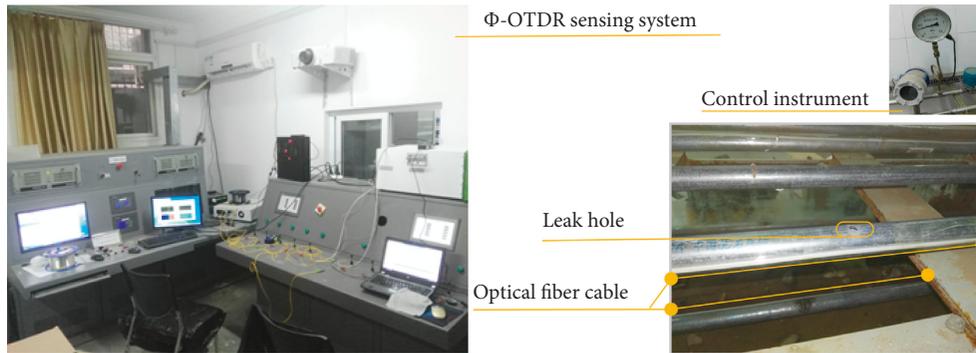


FIGURE 4: Experimental environment.

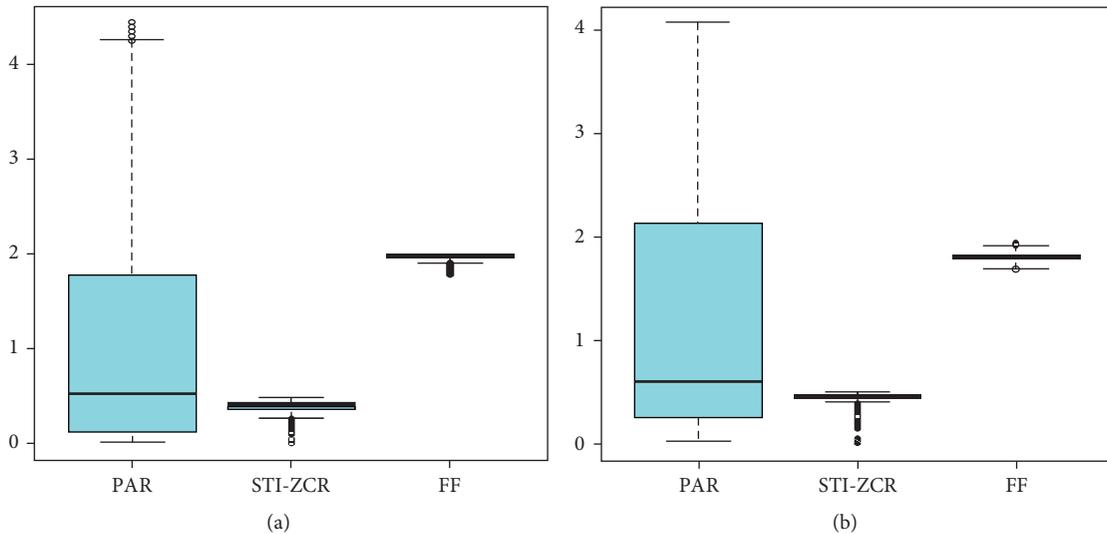


FIGURE 5: Hybrid feature distribution of gas pipelines. (a) Hybrid features of gas leakage pipelines. (b) Hybrid features of gas nonleakage pipelines.

The noise signal has a wider distribution range on PAR and FF. We choose the random-forest classifier to judge the validity of the hybrid features. For comparison, we refer to the SVM method based on the wavelet and the RBF method based on kurtosis [1, 27]. Wavelet is a common signal processing method which has been applied to pipeline monitoring. The RBF method performs great in perimeter security.

In order to evaluate the models more accurately, we choose the other two models for comparison. The results are listed from three aspects of average accuracy, average recall rate, and $F1$ in Table 1.

From the table above, we can find the following:

- The extensive kurtosis used in perimeter security is not suitable for pipeline state recognition.
- When only identifying the leakage of the pipelines, the wavelet feature recognition method has an accuracy close to 80%.
- The accuracy rate of the pipeline state recognition based on the hybrid features reaches 98%. This indicates that the combination features can well reflect pipeline state information.

4.3. Multiple Microstates of Pipeline Representation. The combination of hybrid features and random-forest classifier can well identify whether there is leak in the pipeline. However, in the actual environment, the more the information we get, the easier the arrangement of maintenance work. This is also the significance of the extraction method, which contains energy characteristics, linear characteristics, and frequency-domain local characteristics of pipelines. Here, we also have researched further two-microstate recognition based on this extraction method.

4.3.1. Type Characterization of Leakage Pipelines. For the collected gas leakage data, liquid leakage data, and noise data, we identify the three states based on hybrid features. Among them, noise data are random mixed data of gas noise and liquid noise. After calculation, the feature distribution of these three microstates is as follows (Figure 6).

For comparison, we take C5.0, SVM, and random-forest algorithm into consideration [26]. The recognition results for these three microstates based on hybrid features are shown in Table 2.

The following can be seen from the above results:

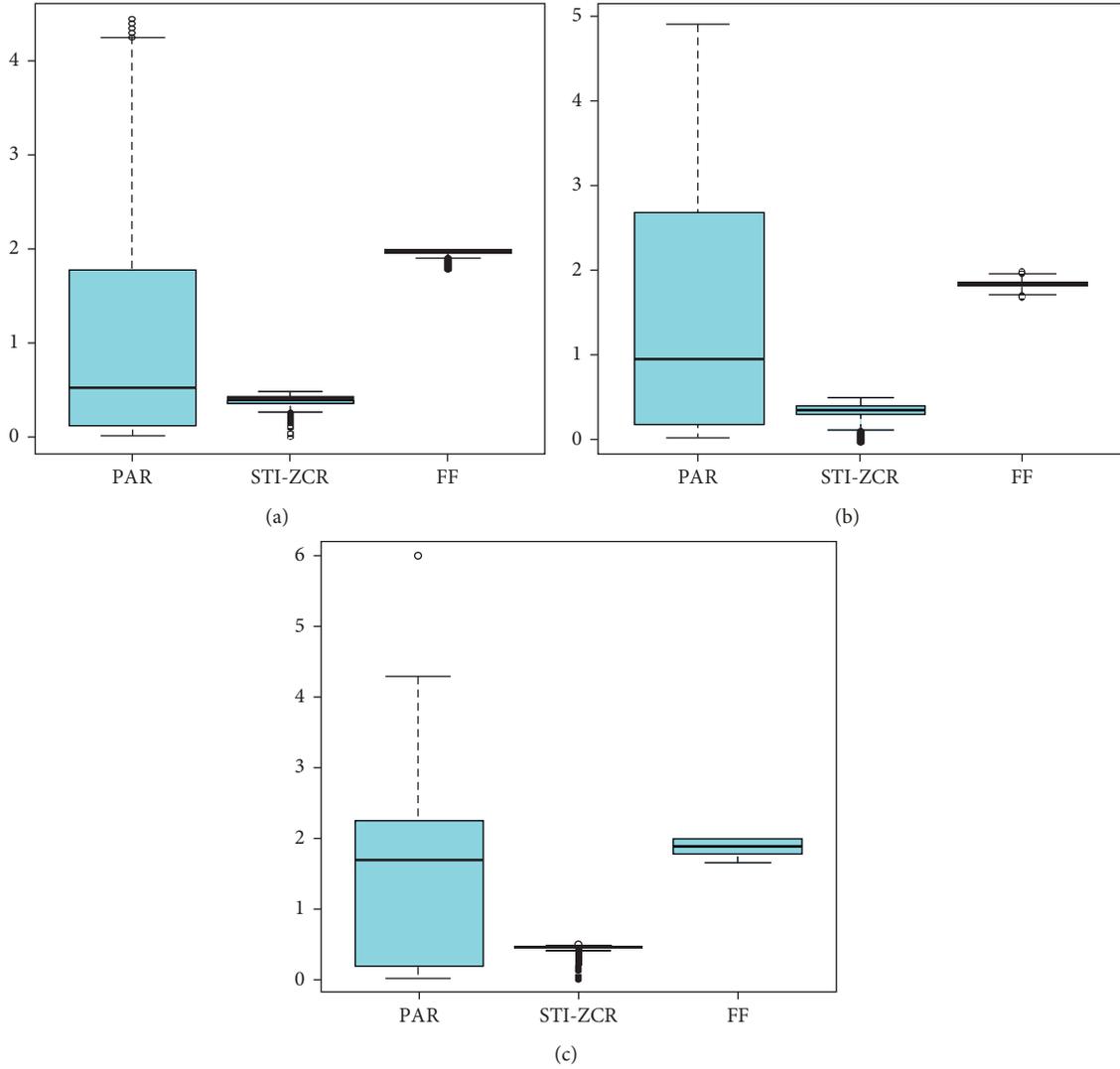


FIGURE 6: Feature distribution of the three microstates. (a) Hybrid features of gas leakage. (b) Hybrid features of liquid leakage. (c) Hybrid features of mixed noise.

- (a) The combination of PAR, STI-ZCR, and FF can effectively improve the recognition rate of microstates of pipelines based on the Φ -OTDR sensing system.
- (b) Compared with C5.0 and SVM, random forest acquires a higher accuracy.

4.3.2. Microstate Characterization of Leakage Pipelines. For more microstates' characterization of leakage pipelines, we control the pressure at 0.5–0.4 MPa, 0.4–0.3 MPa, 0.3–0.2 MPa, and 0.2–0.1 MPa when the gas pipelines are leaking. Regarding these four microstates, the distribution of hybrid features is as follows (Figure 7).

For comparison, we take C5.0, SVM, and random-forest algorithm into consideration. The recognition results for these four states based on hybrid features are shown in Table 3.

When smaller states are recognized, the accuracies with these features decrease obviously. But we still find the

combination of PAR, STI-ZCR, and FF gets the best recognition. Furthermore, random-forest algorithms acquire the optimal recognition results.

5. Discussion and Conclusions

Through the experimental results, we find that the hybrid features proposed take into account the generality of pipeline leakage and also consider the local characteristics of the pipelines based on the Φ -OTDR sensing system. These features proposed in this paper avoid the interference of strong noise and fading effect. From the angle of feature extraction to describe the high-frequency information from the time domain (PAR and STI-ZCR), it helps improving the operational efficiency. This extraction method includes time-frequency information of pipeline states, which can effectively identify pipeline leakage, with accuracy exceeding 98%. In the case of multiple microstates, we can find the FF plays a very significant role in characterizing the microstates

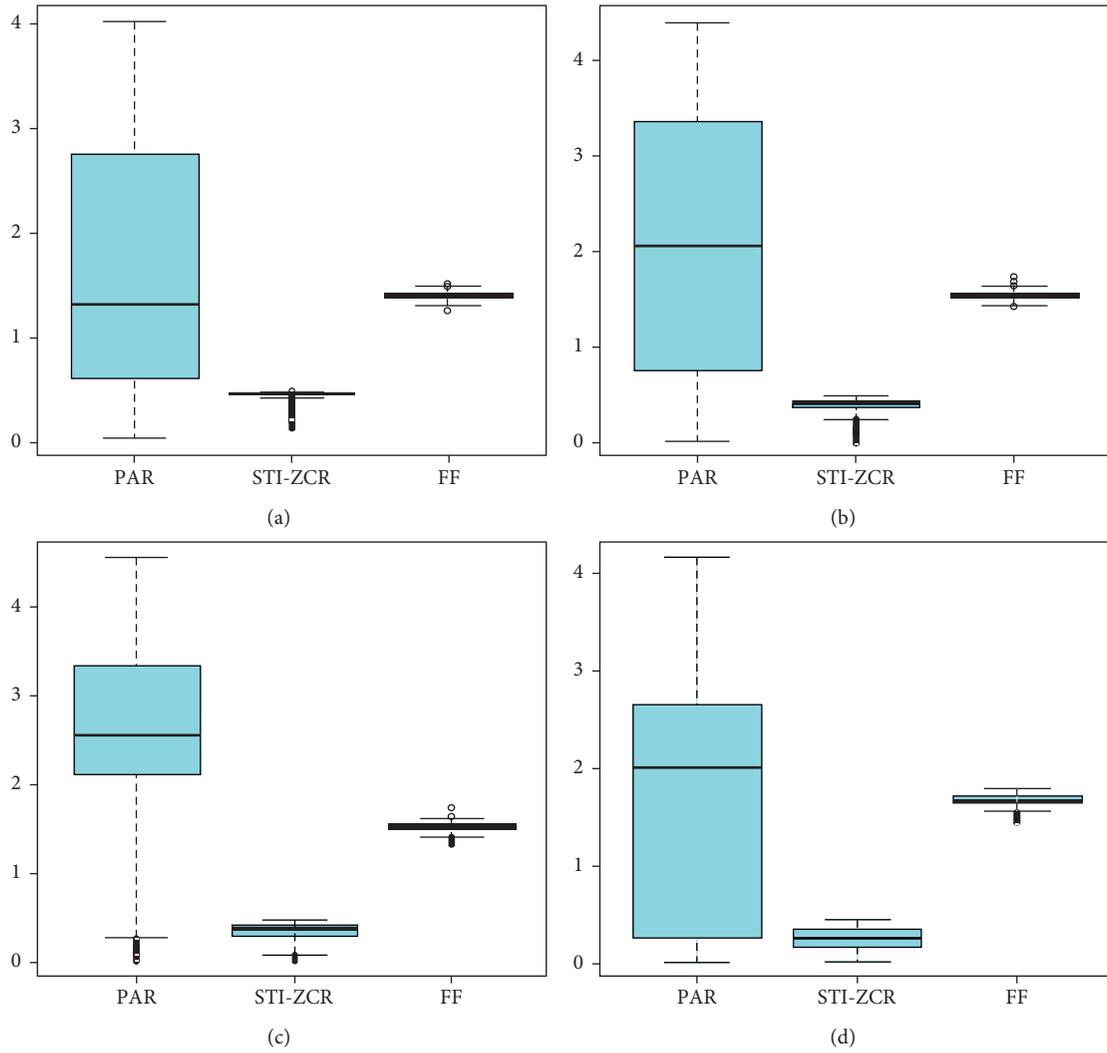


FIGURE 7: Hybrid feature distribution of four microstates. (a) Hybrid features of 0.2–0.1 MPa. (b) Hybrid features of 0.3–0.2 MPa. (c) Hybrid features of 0.4–0.3 MPa. (d) Hybrid features of 0.5–0.4 MPa.

TABLE 1: Comprehensive comparison of hybrid features.

Method	Accuracy (%)	Recall (%)	F1 (%)
Kurtosis with RBF	52.16	49.55	50.82
Wavelet with SVM	79.17	77.97	78.56
Hybrid features with RF (PAR, STI-ZCR, FF)	98.56	98.21	98.38

TABLE 2: Recognition of three microstates with different features.

Algorithms	Features			
	PAR and STI-ZCR (%)	PAR and FF (%)	STI-ZCR and FF (%)	PAR, STI-ZCR, and FF (%)
C5.0	68.31	65.52	70.13	74.92
SVM	73.05	71.32	79.69	86.53
Random forest	80.70	72.57	85.99	91.61

of pipelines based on the Φ -OTDR sensing data. The accuracy rate of pipeline type is more than 91%, and the rate of four various pressure states is also above 83%. Therefore, the combination of hybrid features and RF classifier can be applied to more kinds of microstate identification for

pipeline monitoring based on the Φ -OTDR sensing system. The analysis of pipeline signals from these different aspects provides more possibilities to promote the development of pipeline monitoring based on Φ -OTDR in the engineering practice.

TABLE 3: Recognition of four microstates with different features.

Algorithms	Features			
	PAR and STI-ZCR (%)	PAR and FF (%)	STI-ZCR and FF (%)	PAR, STI-ZCR, and FF (%)
C5.0	55.24	48.92	53.15	59.68
SVM	61.47	58.32	62.75	71.55
Random forest	69.23	60.78	66.22	83.51

It should be mentioned that the current work has its limitations. Although there is rapid development of pipeline monitoring based on Φ -OTDR, the strong noise and fading effect still suppress the feature representation in the time domain. PAR is regarded as a certain estimation of the SNR, and the strong noise in the practice environment requires that feature enhancement methods should be studied [28, 29]. STI-ZCR is an improved feature to express the higher frequency information from the time domain, which avoids the processing costs such as preaggravation. This feature still needs more discussions in the online monitoring field. FF describes the objects from the angle of irregularity and self-similarity, which represents the microscopic leakage signal and extends the range of feature extraction. The features discussed in this paper refer to the fields of signal processing, speech processing, and quantitative description ways like graph theory, which would not obtain the best hybrid features to reflect all the useful information of pipeline signals. More researches referring to other fields will be studied in few years. At the same time, we hope to further expand research on more pipeline leakage scenarios, such as different material pipelines and different embedded environments, and provide support for multiple pipeline microstate monitoring.

Data Availability

The experimental data were obtained by the experimenter on-site using the Φ -OTDR sensing system. There are four categories of data classification: gas pipeline leakage data, gas noise, liquid pipeline leakage data, and liquid noise. Detailed parameters of data acquisition refer to Section 4.1, and data files are stored in the R data format.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was supported by the Beijing Natural Science Foundation under Grant No. 4192042 and the National Science Fund subsidized project under Grant Nos. 61627816 and 61503034.

Supplementary Materials

This research contains a supplementary file, which provides information on four kinds of the sample dataset of the pipeline monitoring data based on the Φ -OTDR sensing

system. The categories of this dataset are gas leakage signal, gas noise, liquid leakage signal, and liquid noise. The length of the sample dataset is 10,000 rows. (*Supplementary Materials*)

References

- [1] Z. Qu, H. Feng, Z. Zeng, J. Zhuge, and S. Jin, "A SVM-based pipeline leakage detection and pre-warning system," *Measurement*, vol. 43, no. 4, pp. 513–519, 2010.
- [2] X. Jiang, X. Zhang, H. Yan et al., "Feasibility analysis of subsea pipeline leak detection system," *Petroleum Engineering Construction*, vol. 43, no. 1, pp. 76–79, 2017.
- [3] A. B. Pnev, A. A. Zhironov, K. V. Stepanov, E. T. Nesterov, D. A. Shelestov, and V. E. Karasik, "Mathematical analysis of marine pipeline leakage monitoring system based on coherent OTDR with improved sensor length and sampling frequency," *Journal of Physics: Conference Series*, vol. 584, Article ID 012016, 2015.
- [4] J. Kang and Z. H. Zou, "Time prediction model for pipeline leakage based on grey relational analysis," in *Proceedings of the International Conference on Circuit and Signal Processing & 2010 Second Iita International Joint Conference on Artificial Intelligence*, pp. 2019–2024, Harbin, China, January 2010.
- [5] H. Qu, T. Zheng, L. Pang, and X. Li, "A new detection and recognition method for optical fiber pre-warning system," *Optik*, vol. 137, pp. 209–219, 2017.
- [6] X. Huang, H. Zhang, K. Liu, T. Liu, Y. Wang, and C. Ma, "Hybrid feature extraction based intrusion discrimination in optical fiber perimeter security system," *IEEE Photonics Journal*, vol. 9, no. 1, pp. 1–12, 2017.
- [7] J. Sun, Q. Xiao, J. Wen, and Y. Zhang, "Natural gas pipeline leak aperture identification and location based on local mean decomposition analysis," *Measurement*, vol. 79, pp. 147–157, 2016.
- [8] A. Joshi, L. Udpa, S. Udpa, and A. Tamburrino, "Adaptive wavelets for characterizing magnetic flux leakage signals from pipeline inspection," *IEEE Transactions on Magnetics*, vol. 42, no. 10, pp. 3168–3170, 2006.
- [9] D. Wang, "Identification method for leakage of water supply pipeline with ρ -OTDR based on optical fiber sensing technology," *Journal of Instrumentation*, vol. 38, no. 4, pp. 830–837, 2017.
- [10] C. Guo, Y. Wen, P. Li, and J. Wen, "Adaptive noise cancellation based on EMD in water-supply pipeline leak detection," *Measurement*, vol. 79, pp. 188–197, 2016.
- [11] H. Wu, Y. Qian, W. Zhang, and C. Tang, "Feature extraction and identification in distributed optical-fiber vibration sensing system for oil pipeline safety monitoring," *Photonic Sensors*, vol. 7, no. 4, pp. 305–310, 2017.
- [12] Z. Meng, Y. Hu, and X. Ai, "Features research for pipeline leakage based on ρ -OTDR sensing system," in *Proceedings of the International Conference on Optical Instruments & Technology: Advanced Optical Sensor & Applications*, Madeira, Portugal, January 2018.

- [13] J. C. Yan, C. H. Tian, J. Huang, and F. Albertao, "Incremental dictionary learning for fault detection with applications to oil pipeline leakage detection," *Electronics Letters*, vol. 47, no. 21, pp. 1198-1199, 2011.
- [14] S. Liu, W. Fu, L. He, J. Zhou, and M. Ma, "Distribution of primary additional errors in fractal encoding method," *Multimedia Tools and Applications*, vol. 76, no. 4, pp. 5787-5802, 2017.
- [15] M. Lu, S. Liu, A. Kumar Sangaiah, Y. Zhou, Z. Pan, and Y. Zuo, "Nucleosome positioning with fractal entropy increment of diversity in telemedicine," *IEEE Access*, vol. 6, pp. 33451-33459, 2018.
- [16] L. Breiman, "Random forest," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.
- [17] B. C. Ko, D.-Y. Kim, J.-H. Jung, and J.-Y. Nam, "Three-level cascade of random forests for rapid human detection," *Optical Engineering*, vol. 52, no. 2, Article ID 027204, 2013.
- [18] B. Asgarian, V. Aghaeidoost, and H. R. Shokrgozar, "Damage detection of jacket type offshore platforms using rate of signal energy using wavelet packet transform," *Marine Structures*, vol. 45, pp. 1-21, 2016.
- [19] A. Meng, J. Ge, H. Yin, and S. Chen, "Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm," *Energy Conversion and Management*, vol. 114, pp. 75-88, 2016.
- [20] S. Liu, M. Lu, G. Liu, and Z. Pan, "A novel distance metric: generalized relative entropy," *Entropy*, vol. 19, no. 6, p. 269, 2017.
- [21] A. M. Youssef, H. R. Pourghasemi, Z. S. Pourtaghi, and M. M. Al-Katheeri, "Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah basin, Asir region, Saudi Arabia," *Landslides*, vol. 13, no. 5, pp. 839-856, 2016.
- [22] A. Nayak, S. Lee, and J. W. Sutherland, "Storage trade-offs and optimal load scheduling for cooperative consumers in a microgrid with different load types," *IIEE Transactions*, vol. 51, no. 4, pp. 397-405, 2019.
- [23] J. W. T. Fan, R. S. C. Yeung, and H. S. H. Chung, "Optimized hybrid PWM scheme for mitigating zero-crossing distortion in totem-pole bridgeless PFC," in *Proceedings of the 2018 IEEE Applied Power Electronics Conference and Exposition (APEC)*, IEEE, San Antonio, TX, USA, March 2018.
- [24] P. Gao, Z. Liu, G. Liu, H. Zhao, and X. Xie, "Unified metrics for characterizing the fractal nature of geographic features," *Annals of the American Association of Geographers*, vol. 107, no. 6, pp. 1315-1331, 2017.
- [25] L. Shen, Y. S. Han, and S. Wang, "Research on fractal feature extraction of radar signal based on wavelet transform," in *Advances in Intelligent Systems and Interactive Applications*, pp. 330-336, Springer, Berlin, Germany, 2018.
- [26] T. G. Pavey, N. D. Gilson, S. R. Gomersall, B. Clark, and S. G. Trost, "Field evaluation of a random forest activity classifier for wrist-worn accelerometer data," *Journal of Science & Medicine in Sport*, vol. 20, no. 1, pp. 75-80, 2017.
- [27] K. Liu, M. Tian, T. Liu et al., "A high-efficiency multiple events discrimination method in optical fiber perimeter security system," *Journal of Lightwave Technology*, vol. 33, no. 23, pp. 4885-4890, 2015.
- [28] X. Zhang, Z. Guo, Y. Shan, Z. Sun, S. Fu, and Y. Zhang, "Enhanced Φ -OTDR system for quantitative strain measurement based on ultra-weak fiber Bragg grating array," *Optical Engineering*, vol. 55, no. 5, Article ID 054103, 2016.
- [29] G. Tu, B. Yu, S. Zhen, K. Qian, and X. Zhang, "Enhancement of signal identification and extraction in a Φ -OTDR vibration sensor," *IEEE Photonics Journal*, vol. 9, no. 1, pp. 1-10, 2017.



Hindawi

Submit your manuscripts at
www.hindawi.com

