Unsupervised Learning Method for Events Identification in φ-OTDR

Jie Zhang
Beijing Technology and Business University
Xiaoting Zhao
Beijing Jiaotong University
Yiming Zhao
Xi'an University of Posts and Telecommunications
Xiang Zhong
Hefei University of Technology
Yidan Wang
Beijing Jiaotong University
Fanchao Meng
Beijing Jiaotong University
Jinmin Ding
Beijing Jiaotong University
Yingli Niu
Beijing Jiaotong University
Xinghua Zhang
Beijing Jiaotong University
Liang Dong
Beijing Haidongqing Electrical and Mechanical Equipment Co.
Sheng Liang (shliang@bjtu.edu.cn)
Beijing Jiaotong University https://orcid.org/0000-0002-8251-5758

Research Article

Keywords: Fiber-optic distributed vibration sensor, phase-sensitive optical time domain reflectometry (φ-OTDR), unsupervised learning, clustering, event identification.

Posted Date: December 3rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1030534/v1

Abstract

In this paper, an unsupervised-learning method for events-identification in φ-OTDR fiber-optic distributed vibration sensor is proposed. The different vibration-events including blowing, raining, direct and indirect hitting, and noise-induced false vibration are clustered by the *k*-means algorithm. The equivalent classification accuracy of 99.4% has been obtained, compared with the actual classes of vibration-events in the experiment. With the cluster-number of 3, the maximal *Calinski-Harabaz* index and *Silhouette* coefficient are obtained as 2653 and 0.7206, respectively. It is found that our clustering method is effective for the events-identification of φ-OTDR without any prior labels, which provides an interesting application of unsupervised-learning in self-classification of vibration-events for φ-OTDR.

I. Introduction

The phase-sensitive optical time-domain reflectometry (φ-OTDR) based fiber-optic distributed vibration sensor can detect and locate the multi-point vibration on long distance sensing optical fiber. The existing research has been engaged in improving multiple performance indicators in the multi-point vibration detection, including dynamic and spatial resolutions, detection distance and accuracy, etc. [1]. So, it has been widely used in the fields of vibration and acoustic wave detection [2], intrusion and pipeline monitoring [3], perimeter security [4–6], structural health monitoring [7, 8], seismic wave monitoring [9, 10], the railway monitoring [11–13], real-time vehicle detection, classification, speed estimation [14] and pipeline leak detection [15]. In practical applications, φ-OTDR is utilized in monitoring different vibration-events, therefore, it is necessary to classify the vibration-events detected to reduce the false alarm rate and obtain more abundant vibration information in different applications.

In addition to the traditional signal-processing methods, deep learning has also been introduced to the vibration-event identification, and a lot of work has been proposed. By using multi-branch long short-time memory modeling combined with convolution neural network (MLSTM-CNN), φ-OTDR can identify four kinds of real disturbance events, including the watering, climbing, hitting, and pressing, and a false disturbance event. The identification rates of the five events are 94.38%, 94.40%, 94.76%, 95.25% and 99.71%, respectively. Average identification rate can reach 95.7%, and nuisance alarm rate (*NAR*) is 4.3% [16]. Wang et al.[17] used random forests (RF) to identify the watering, tapping, squeezing, and non-disturbing events. The experimental identification rates have reached93.79%, 97.36%, 97.06% and 98.12%, respectively, with an average identification rate of 96.58% [17]. The current work can also identify real disturbance events, including working, digging, large vehicle driving, a two-point leakage of a pipeline, and a three-point leakage by CNN [18]. In addition, Generative Adversarial Net (GAN) has also been used to identify noises, walking and vehicles events [19]. Overall, the current event-identification methods for φ-OTDR are summarized in Table 1.

However, these works are mainly based on supervised learning, which requires to collect a large number of samples with manual labelling, limiting the research efficiency and improvement of an intelligent φ-

OTDR. For special applications in complex environments, it may be difficult to implement the field observations of vibration-events, and it is impossible to obtain enough accurate labels, which is the bottleneck problem of the event-identification for φ-OTDR by the supervised learning.

Table 1
Current event-identification methods for φ-OTDR.

Method	Vibration-events	Features	Learning mode
[5]	false-disturbance, walking, wind, shaking, and vibrator	ratio of low frequency energy to total energy, total energy, peak to average ratio	supervised-learning
[14]	cars, SUVs, and trucks	wavelet, short-term energy, short-term zero-crossing rate	supervised-learning
[16]	watering, climbing, knocking, pressing, false-disturbance	short-term characteristics	supervised-learning
[17]	watering, tapping, squeezing, and non-disturbance	peak value, minimum value, peak-to-peak value, root mean square, variance	supervised-learning
[18]	walking, digging, large vehicle, and leakage	spatial-temporal image	supervised-learning
[19]	noises, walking, and vehicle	image of power and phase	unsupervised-learning
[20]	watering, climbing, hitting, squeezing, and false-disturbance	30 and 10 features in time and frequency domain, respectively	supervised-learning
[21]	running, and walking	autonomously extracts feature vectors	supervised-learning
[22]	fence climbing, wall breaking, and wire crossing	space-time image	supervised-learning
[23]	digging with hammer, and pickaxe	variance, skewness, and kurtosis	supervised-learning
[24]	background noises, manual/mechanical excavation, traffic, and noises of a factory	pulse intensity, average amplitude, energy, spectrum variance, information entropy, spectral density amplitude	supervised-learning
[25]	air hammer, electric drill, metal hammer, wood saw, and pile driver	time-frequency characteristics spectrogram	supervised-learning
[26]	gas pipeline leakage, gas noises, liquid pipeline leakage, and liquid noises	peak-to-average power ratio, short-term interval zero crossing	supervised-learning
[27]	digging, walking, driving, and shaking	spectrogram	supervised-learning
[28]	background, walking, jumping, beating with shovel, and digging	temporal-spatial data matrix	supervised-learning

Method	Vibration-events	Features	Learning mode
[29]	strong wind, heavy rain, sunny weather, tapping, and climbing	time-domain features and time-frequency-domain features	supervised-learning
[30]	watering, knocking, climbing, pressing, and false disturbance	eigenvectors by empirical mode decomposition	supervised-learning
[31]	rain, wind blowing, treading, slapping, and impacting	recurrent plot	supervised-learning
[32]	false disturbance, traffic flow, excavator, road breaker, and digging	space-time signal matrix	supervised-learning
[33]	walking, and running	spectrum	supervised-learning
[34]	internal/corona/surface partial discharges, and noises	multiscale wavelet decomposition	supervised-learning
[35]	extrinsic intrusions and intrinsic corrosions	time-domain features and frequency-domain features	unsupervised learning
[41]	knocking, kicking, wagging, wagging, noises, crashing, cutting, climbing	image	supervised-learning
This work	blowing, raining, direct and indirect hitting	time-domain features	unsupervised learning

The unsupervised learning opens a new window to break through this problem. Shiloh et al. [19] proposes an unsupervised learning method based on the Generative Adversarial Net (GAN) methodology. The GAN-trained network is designed to generating train-sets for DAS ANNs and its experimental testing, which has significantly improved performance, compared with the classifiers that are trained with the pure simulation data or pure experimental data [19]. The results of supervised learning by CNN are compared with the unsupervised learning by *K*-means and auto-encoder. Auto-encoder and CNN have an identification rate of 95% for pipeline defects. *K*-means and CNN have an identification rate of 84.62% for the walking or running. Some researchers also use machine learning algorithm based on neural network to analyze the external intrusion and internal corrosion data of pipes collected by the φ-OTDR system, by the supervised learning algorithm and unsupervised learning algorithm, respectively [35].

No matter the combination of the supervised learning and unsupervised learning or the application of unsupervised learning, the unsupervised learning plays an increasingly important role in the vibration identification field. According to the unsupervised learning, the relationships within the dataset itself can be found, which are underlying features of the data that can be associated with specific events of the class. Although the supervised learning method can well identify events, it is not a very practical and unified method for analyzing various events. Because the supervised learning needs to label the samples,

when the amount of data involved is very large or needs to be recognized on the spot, it will be a very hard work. A lot of time of manually labeling is saved when the unsupervised learning is applied.

In contrast with those studies, our research is main application of the unsupervised learning clustering algorithm to cluster the vibration events of the φ-OTDR detection signal, including blowing, raining, direct and indirect hitting. The scenes are different, and we recognize more types of vibration events. We only need to use the waveform information in the time domain, without turning to the frequency domain to process the data. At the same time, we use Principal Component Analysis (PCA) dimension reduction method in the data processing process to project the data into a low-dimensional subspace to achieve dimension reduction, so that the data can be clustered better. In addition to comparing the clustering results with the original tags to evaluate the performance of the clustering results, we also use the *Calinski-Harabaz* index (*CH*) and *Silhouette* coefficients to judge our clustering performance. In the end, after our data processing optimization, the final vibration identification accuracy rate is 99.4%, which is higher than the accuracy rates of the existing supervised learning methods. From the perspective of application scopes, our method is more convenient and has a wider range of applications, which lays a foundation for pattern identification via the unsupervised learning method in the later stage.

Through the unsupervised learning, vibration events can be clustered without any input of prior labels, which can be considered as a self-identification. Based on the clustering task to obtain different clusters, only a very limited field observation or calibration of different clusters is needed to realize the practical application pattern identification.

Ii. Principle

φ-OTDR is based on the measurement of Rayleigh scattering that occurs when the light travelling in fiber-optic cable is back-scattered due to imperfections along the fiber. Theoretically, the mechanical vibrations due to the physical activities or events taken place around the cable cause fluctuations in the back-scattered light. Interrogating these fluctuations then enabling to detect and classify such activities. In this context, it has already known that the back-scattered signal sensed in a φ-OTDR has generally quite low signal-to-noise ratio (SNR) which highly affects the vibration detection capability of the systems. Thus, several methods such as multi-dimension comprehensive analysis [7], adaptive temporal/match filtering [9], Wavelet denoising [10–11], polarization diversity detection [12], two-dimensional edge detection [13], difference in time-frequency domain [14], Fourier transforming [15], heterodyning [36], Raman amplification [37], signal-noise separation [38], multi-scale permutation entropy and the zero-crossing rate [39] have been used to improve the threat/event detection performance. The optical fiber distributed vibration sensor we used in this paper is φ-OTDR, and its disturbance system structure diagram is shown in Fig.1.

Iii. Experimental Result And Discussion

3.1 Experimental setup

A Narrow Linewidth Laser with 5kHz frequency width and a central wavelength of 1550.92 nm is used as the light source. An acoustic Optic Modulator (AOM) shift chops the continuous light into probe pulses with a pulse width of 100 ns and a pulse interval of 0.02ms. The sensing fiber is a single mode fiber, with about 2 km, a spatial resolution of 10 m. The vibration sampling rate is 90 Ms/s. The vibration event can be located and classified by welding the traditional communication optical cable with the host.

3.2 Training and Test

The process of the principal component analysis (PCA) and clustering is shown in Fig. 2.

3.2.1 Principal Component Analysis (PCA)

The sample data is divided to find the approximate location of the vibration according to the starting point. The vibration data of the approximate vibration position is selected and saved. As shown in Fig. 3, the signals in time domain between several events are different.

Feature extraction is performed on the vibration data of the intercepted vibration events. The eigenvalues of time domain signals can be divided into dimension and dimensionless as shown in Table 2[20]. 14 features are extracted as the input data of clustering algorithm.

Table 2
The features in time domain

	Dimension	Dimensionless
1	maximum	rectified average
2	minimum	kurtosis
3	average	skewness
4	peak-to-peak	form factor
5	variance	crest factor
6	standard deviation	impulse factor
7	root mean-square	margin factor

PCA is a technique for analyzing and simplifying data sets. PCA is often used to reduce the dimensionality of the data set, while maintaining the feature that contributes the most to the variance in the data set. This is done by keeping the low-order principal components and ignoring the high-order principal components. Such low-level components can often retain the most important aspects of the data. The purpose of the *K*-means clustering algorithm is to divide *n* points into K clusters, making each point belong to the cluster corresponding to the nearest mean (this is the clustering center), and take it as the clustering standard. When all the eigenvalues are taken as the input, the clustering effect is relatively poor, so PCA is performed on the eigenvalues to get a data dimensionality reduction. And the corresponding eigenvalues are arranged from large to small as 3.5795 > 0.5354 > 0.2512 > 0.1979 >

0.0034 (the remaining 9 eigenvalues are too small and have been omitted), so the real input of the clustering algorithm has been obtained.

3.2.2 Clustering

Through the data preprocessing stage, we obtain a total of 1160 samples data of blowing vibration events, 550 samples data of raining vibration events, 300 samples data of indirect hitting vibration events by knocking on the iron net, 270 samples data of direct hitting vibration events and a total of 2280 samples data.

The location selection of K initialized centroids has a great influence on the final clustering result and running time, so it is necessary to select the appropriate K centroids. If it is just a completely random selection, it may cause the algorithm to converge very slowly. The K-Means++ algorithm [42] is an optimization of the K-Means method of randomly initializing the centroid.

The optimization strategy for initializing the centroid of K-Means is detailed below. =

a. Select a point randomly from the set of input data points as the first clustering center μ_1
b. For each point x_i in the data set, calculate the distance
$D\left(x_i\right) = \arg min \|x_i - \mu_r\|_2^2 r = 1,2,\cdots k_{selected}$
c. Select a new data point as the new cluster center. The principle of selection is as follows: the point with larger D(x) has a higher probability of being selected as the clustering center
d. Repeat b and c until K cluster centroids are selected
e. The K centroids are used as the initial centroids to run the standard K-means algorithm.

Four types of events will be identified so the cluster class K is set to 4. We varied the number of reduced features via PCA from 2 to 6 and performed *K*-means clustering. The analysis results are listed in Table 3. It can be seen from the table that when the number of clusters is 4 and the PCA feature is 5, the accuracy is the highest, which is 87.1%. At the same time, when the number of features is increased, the accuracy is no longer improved.

Table 3
The clustering situation where *K*= 4

Number of features	Number of samples	Number of correct samples	Accuracy
6	2280	1986	87.1%
5	2280	1986	87.1%
4	2280	1780	78.1%
3	2280	1718	75.4%
2	2280	1736	76.1%

In the clustering process, we find that the waveforms of indirect hitting events are similar to those of direct hitting events. Therefore, we classify indirect hitting events and direct hitting events as hitting events. The three vibration events are clustered, as shown in Table 4. When the number of clusters is 3 and the PCA feature is 4, the accuracy is the highest, which can reach up to 89.4%. At the same time, when the number of features is increased, the accuracy is no longer improved.

Table 4
The clustering situation where *K*=3

Number of features	Number of samples	Number of correct samples	Accuracy
6	2280	2039	89.4%
5	2280	2039	89.4%
4	2280	2039	89.4%
3	2280	2023	88.7%
2	2280	1998	87.6%

In the process of clustering, we find that the incorrectly clustered vibration data points show the characteristics of the regionalization. That is, in the process of selecting the sample data, some samples are not the vibration data. So, we manually eliminate these non-vibration sample data points, and get a total of 2044 optimized samples data. PCA dimension reduction is performed on the optimized data, and then the *K*-means clustering algorithm is performed, and the results obtained are shown in Table 5. When the number of clusters is 3 and the number of features is 4, the accuracy can reach up to 99.4%, and when the number of features is increased, the accuracy will not be improved. From Fig. 4, we can also see that the clustering effect is good and Idx(Idx represents the dimension of data, equivalent to the two columns of our samples) is the vector that predicts the cluster index.

Table 5
The clustering results where *K*=3.

Number of features	Number of samples	Number of correct samples	Accuracy
6	2044	2031	99.4%
5	2044	2031	99.4%
4	2044	2031	99.4%
3	2044	2025	99.1%
2	2044	2012	98.4%

Later, confusion matrix is analyzed to evaluate the performance of the algorithm as shown in Fig,5.

3.2.3 Performance measurement

In order to evaluate the performance of the clustering algorithm, the *Silhouette* coefficient S_i and *CH* index are used as evaluation indexes. The calculation of *Silhouette* coefficient is shown as follow:

$$S_i=(b_i-a_i)/max(a_i, b_i) \quad (1)$$

In this formula, *a* represents the average distance between the sample point and the other points in the same cluster, that is, the similarity between the sample points and the other points in the same cluster; *b* represents the average distance between the sample points and the other points in the next closest cluster, namely the similarity between the sample points and the other points in the next closest cluster. What *K*-means pursues is that for each cluster, the difference within the cluster is small, while the difference outside the cluster is large. The silhouette coefficient *S* is the key index to describe the difference inside and outside the cluster. According to the formula, the value range of *S* is [-1, 1], and the closer the value is to 1, the better the clustering performance. On the contrary, the closer the value is to -1, the worse the clustering performance. When the contour coefficient is equal to 0, it means that the clusters overlap. As shown in the Table 6 below, when the number of clusters is 3 and the feature is 4, the silhouette coefficient is the highest 0.7206. The *Silhouette* coefficients of different clusters show that the samples can be split into 3 clusters. All the points in the three clusters have a large *Silhouette* coefficient (0.8 or greater), indicating that the clusters are well separated.

Table 6
The average *Silhouette* coefficient.

K	Number of features	Number of samples	Average *Silhouette* coefficient
2	4	2044	0.6193
3	4	2044	0.7206
4	4	2044	0.5960
5	4	2044	0.6090

The calculation of *CH* index is shown in **(2)**:

$$s(k)=\frac{tr(B_K)}{tr(W_K)}\frac{m-k}{k-1} \quad (2)$$

In this formula, m is the number of samples in training set, and k is the number of classifications. B_k is the covariance matrix between classifications, and W_k is the covariance matrix of data within classifications. That is to say, the smaller the covariance of the data within the classification, the better, and the larger the covariance between the classifications, the better, so the *CH* index score will be high, and the boundary between clusters is obvious. We can know from the article [40] that the higher the CH index, the better the clustering effect. And as shown in the Fig. 6, the maximum *CH* index is 2653 when the number of cluster classes is 3.

3.3 Result and Discussion

A total of 14 input features are obtained by feature extraction from samples, and their dimensions are reduced by PCA. The first 6 features scored after reducing dimension are selected for research. We find that when the number of clusters is 4, and the number of features is 5, the equivalent classification accuracy is as high as 87.1%; when the number of clusters is 3 and the number of features is 4, the accuracy is as high as 89.4%. After we manually remove the non-vibration data in the sample, we get that when the number of clusters is 3 and the feature is 4, the accuracy rate reaches the highest 99.4%. And the average silhouette coefficient and *CH* index can reach 0.7206 and 2653, respectively, which represent the best performance of the clustering task. It is found that the clustering method is effective to be utilized in practical applications. Integrate Z-score normalization, feature extraction, PCA dimensionality reduction, and calculation of the distance between the sample data and the cluster center into one functional section. The real-time data collected by the distributed optical fiber sensor can be used to determine which vibration events the sample belonged to through this functional block. In the subsequent research, we will continue to improve the clustering method to achieve real-time monitoring and alarm systems for vibration events.

The unsupervised learning algorithm proposed in this paper for the automatic classification of vibration events can significantly improve the efficiency of identifying vibration events, but it also has its limitations. For some similar vibration events, that is, the vibration waveform is similar, its identification effect is relatively poor, such as direct hitting and indirect hitting events.

Iv. Conclusion

In this paper, an unsupervised learning method is proposed to identify the vibration events for φ-OTDR. When the number of clusters is 4, the accuracy can reach 87.1%; when the number of clusters is 3, the accuracy rate can reach 89.4%; when we manually eliminate some non-vibration data for the optimization, the accuracy rate can reach 99.4%. The average *CH* index and *Silhouette* coefficient are 2653 and 0.7206, respectively, which represent the best performance of the clustering model. Our unsupervised learning algorithm used for automatic identification of vibration events is effective and highly accurate, and we will continue to expand its applicability in the following research.

Declarations

Acknowledgements

This work is supported by the Fundamental Research Funds for the Central Universities (2019JBM345), the Beijing Natural Science Foundation (4192047).

References

1. Liu, H. Li, T. He, CZ. Fan, ZJ. Yan, DM. Liu, QZ. Sun, “Ultra-high resolution strain sensor network assisted with an LS-SVM based hysteresis model,” Opto-Electronic Advances. Vol. 4, no. 5, pp.

200037, 2021.

2. Wang, Y. Lv, Y. Wang, X. Liu, Q. Bai, H. Zhang, B. Jin, “Adaptability and Anti-Noise Capacity Enhancement for ϕ-OTDR With Deep Learning,” Journal of Lightwave Technology. Vol. 38, no. 23, pp. 6699-6706, 2020.
3. Marcon, M. Soriano-Amat, R. Veronese, A. Garcia-Ruiz, M. Calabrese, L. Costa, M. Fernandez-Ruiz, H. Martins, L. Palmieri, M. Gonzalez-Herraez, “Analysis of Disturbance-Induced “Virtual” Perturbations in Chirped Pulse φ-OTDR,” IEEE Photonics Technology Letters. Vol. 32, no. 3, pp. 158-161, 2019.
4. Zhu, X. Xiao, Q. He, and D. Diao, “Enhancement of SNR and Spatial Resolution in φ-OTDR System by Using Two-Dimensional Edge Detection Method,” Journal of Lightwave Technology. Vol. **31, no.**17, pp. 2851-2856 ,2013
5. Cao, X. Fan, Q. Liu, and Z. He, “Practical Pattern Recognition System for Distributed Optical Fiber Intrusion Monitoring System Based on Phase-Sensitive Coherent OTDR,” Asia Communications and Photonics Conference 2015 OSA Technical Digest pp. 145, 2015.
6. Ma, K. Liu, J. Jiang, Z. Li, P. Li, and T. Liu, “Probabilistic Event Discrimination Algorithm for Fiber Optic Perimeter Security Systems,” Journal of Lightwave Technology. Vol. **36, no.** 11, pp. 2069-2075, 2018.
7. Lu, T. Zhu, L. Chen and X. Bao, “Distributed Vibration Sensor Based on Coherent Detection of Phase-OTDR,” Journal of Lightwave Technology. Vol. **28, no.** 22, pp. 3243-3249, 2010.
8. Yue, B. Zhang, Y. Wu, B. Zhao, J. Li, Z. Ou, and Y. Liu, “Simultaneous and signal-to-noise ratio enhancement extraction of vibration location and frequency information in phase-sensitive optical time domain reflectometry distributed sensing system,” Optical Engineering, Vol. **54, no.** 4, pp. 047101, 2015.
9. Jiang, Z. Wang, Z. Wang, Y. Wu, S. Lin, J. Xiong, Y. Chen, and Y. Rao, “Coherent Kramers-Kronig Receiver for φ-OTDR,” Journal of Lightwave Technology. Vol. **37, no.** 18, pp. 4799-4807, 2019.
10. Liehr, L. A. Jäger, C. Karapanagiotis, S. Münzenberger, and S. Kowarik, “Real-time dynamic strain sensing in optical fibers using artificial neural networks,” Optics Express, Vol. **27, no.** 5, pp. 7405-7425, 2019.
11. Peng, N. Duan, Y. Rao and J. Li, “Real-Time Position and Speed Monitoring of Trains Using Phase-Sensitive OTDR,” IEEE Photonics Technology Letters, Vol. **26**, no. 20, pp. 2055-2057, 2014.
12. Kowarik, M. Hussels, S. Chruscicki, S. Münzenberger, A. Lämmerhirt, P. Pohl and M. Schubert, “Fiber Optic Train Monitoring with Distributed Acoustic Sensing: Conventional and Neural Network Data Analysis,” Sensors, Vol. **20**, no. 2, pp. 450, 2020.
13. Wang, B. Lu, H. Zheng, Q. Ye, Z. Pan, H. Cai, R. Qu, Z. Fang, and H. Zhao, “Novel railway-subgrade vibration monitoring technology using phase-sensitive OTDR,” 2017 25th Optical Fiber Sensors Conference (OFS). IEEE, 2017: 1-4.
14. Liu, J. Ma, T. Xu, W. Yan, L. Ma and X. Zhang, “Vehicle Detection and Classification Using Distributed Fiber Optic Acoustic Sensing,” IEEE Transactions on Vehicular Technology, Vol. 69, no. 2, pp. 1363-1374, Feb. 2020.

15. Zuo, Y. Zhang, H. Xu, X. Zhu, Z. Zhao, X. Wei and X. Wang, “Pipeline Leak Detection Technology Based on Distributed Optical Fiber Acoustic Sensing System,” IEEE Access, Vol. 8, pp. 30789-30796, 2020.
16. Wang, S. Lou, X. Wang, S. Liang and X. Sheng, “Multi-branch long short-time memory convolution neural network for event identification in fiber-optic distributed disturbance sensor based on φ-OTDR,” Infrared Physics & Technology, Vol. 109, pp. 103414, 2020.
17. Wang, Y. Liu, S. Liang, W. Zhang and S. Lou, “Event identification based on random forest classifier for φ-OTDR fiber-optic distributed disturbance sensor,” Infrared Physics & Technology, Vol. 97, pp. 319-325, 2019.
18. Sun, Q. Li, L. Chen, J. Quan and L. Li, “Pattern recognition based on pulse scanning imaging and convolutional neural network for vibrational events in φ-OTDR,” Optik, Vol. 219, pp. 165205, 2020.
19. Shiloh, A. Eyal and R. Giryes, “Efficient Processing of Distributed Acoustic Sensing Data Using a Deep Learning Approach,” Journal of Lightwave Technology, Vol. 37, no. 18, pp. 4755-4762, 2019.
20. Jia, S. Lou, S. Liang and X. Sheng, “Event Identification by F-ELM Model for φ-OTDR Fiber-Optic Distributed Disturbance Sensor,” IEEE Sensors, Journal. Vol. **20**, no. 3, pp. 1297-1305, 2020.
21. Peng, J. Jian, M. Wang, Q. Wang, T. Boyer, H. Wen, H. Liu, Z. Mao, K. P. Chen, “Big Data Analytics on Fiber-Optical Distributed Acoustic Sensing with Rayleigh Enhancements,” Proc. 2019 IEEE Photonics Conference (IPC), San Antonio, TX, USA, pp. 1-3, 2019.
22. Li, J. Zhang, M. Wang, Y. Zhong, and F. Peng, “Fiber distributed acoustic sensing using convolutional long short-term memory network: a field test on high-speed railway intrusion detection,” Optics. Express Vol. **28**, no. 3, pp. 2925-2938, 2020.
23. A. Abufana, Y. Dalveren, A. Aghnaiya and A. Kara, “Variational Mode Decomposition-Based Threat Classification for Fiber Optic Distributed Acoustic Sensing,” IEEE Access Vol. **8**, pp. 100152-100158, 2020.
24. Wu, X. Liu, Y. Xiao and Y. Rao, “A Dynamic Time Sequence Recognition and Knowledge Mining Method Based on the Hidden Markov Models (HMMs) for Pipeline Safety Monitoring With φ-OTDR,” Journal of Lightwave Technology. Vol. **37**, no. 19, pp. 4991-5000, 2019.
25. Zhang, Y. Li, J. Chen, Y. Song, J. Zhang, M. Wang, “Event detection method comparison for distributed acoustic sensors using φ-OTDR,” Optical Fiber Technology. Vol. **52**, pp. 101980-101986, 2019.
26. Hu, Z. Meng, X. Ai, H. Li, Y. Hu, and H. Zhao, “Hybrid Feature Extraction of Pipeline Microstates Based on φ-OTDR Sensing System,” Journal of Control Science and Engineering, Vol. 2019, no. 1, pp.1-10, 2019.
27. Xu, J. Guan, M. Bao, J. Lu, W. Ye, “Pattern recognition based on time-frequency analysis and convolutional neural networks for vibrational events in φ-OTDR,” Optical Engineering. Vol. **57, no. 1**, pp. 016103, 2018.
28. Shi, Y. Wang, L. Zhao and Z. Fan, “An Event Recognition Method for φ-OTDR Sensing System Based on Deep Learning,” Sensors Vol. **19**, no. 15, pp. 3421, 2019.

29. M. Tabi. Fouda, D. Han, B. An, X. Chen, “Research and Software Design of an Φ-OTDR-Based Optical Fiber Vibration Recognition Algorithm[J],” Journal of Electrical and Computer Engineering, 2020, 2020.
30. Wang, S. Lou, S. Liang, X. Sheng, “Multi-Class Disturbance Events Recognition Based on EMD and XGBoost in φ-OTDR,” IEEE Access Vol. **99**, pp. 63551-63558,2020.
31. Lyu, J. Jiang, B. Li, Z. Huo, J. Yang, “Abnormal events detection based on RP and inception network using distributed optical fiber perimeter system,” Optics and Lasers in Engineering, Vol. **137**, pp. 106377-106384, 2021.
32. Wu, M. Yang, S. Yang, H. Lu, C. Wang, Y. Rao, “A novel DAS signal recognition method based on spatiotemporal information extraction with 1DCNNs-BiLSTM network,” IEEE Access Vol. **8**, pp. 119448-119457, 2020.
33. Wen, Z. Peng, J. Jian, et al., “Artificial intelligent pattern recognition for optical fiber distributed acoustic sensing systems based on phase-OTDR,” 2018 Asia Communications and Photonics Conference (ACP). IEEE (2018), 1-4.
34. Che, H. Wen, X. Li, et al., “Partial discharge recognition based on optical fiber distributed acoustic sensing and a convolutional neural network,” IEEE Access Vol. **7**, pp. 101758-101764, 2019.
35. Peng, J. Jian, H. Wen, A. Gribok, M. Wang, H. Liu, S. Huang, Z. Mao and K.P. Chen, “Distributed fiber sensor and machine learning data analytics for pipeline protection against extrinsic intrusions and intrinsic corrosions,” Optics Express, Vol. 28, no. 19, pp. 27277-27292, 2020.
36. Peng, H. Wu, X. Jia, Y. Rao, Z. Wang, and Z. Peng, “Ultra-long high-sensitivity φ-OTDR for high spatial resolution intrusion detection of pipelines,” Optics Express, Vol. 22, no. 11, pp. 13804-13810, 2014.
37. Liu, J. Ma, W. Yan, W. Liu, X. Zhang, and C. Li, “Traffic flow detection using distributed fiber optic acoustic sensing,” IEEE Access, Vol. 6, pp. 68968-68980, 2018.
38. Liang, X. Zhao, R. Liu, X. Zhang, L. Wang, X. Zhang, Y. Lv, and F. Meng, “Fiber-Optic Auditory Nerve of Ground in the Suburb: For Traffic Flow Monitoring,” IEEE Access, Vol. 7, pp. 166704-166710, 2019.
39. Jia, S. Liang, S. Lou, and X. Sheng, “A k-Nearest Neighbor Algorithm-Based Near Category Support Vector Machine Method for Event Identification of φ-OTDR,” IEEE Sensors Journal, Vol. 19, No. 10, pp. 3683-3689, 2019.
40. Liu, Q. Wang, M. Dong, Z. Zhang, Y. Li, Z. Wang and ZS. Wang, “Application of ++ Algorithm Based on t-SNE Dimension Reduction in Transformer District Clustering,” 2020 Asia Energy and Electrical Engineering Symposium (AEEES). IEEE, 2020: 74-78.
41. Sun, L. Kun, J. Jiang, T. Xu, S. Wang, G. Hairuo, Z. Zhou, X. Kang, Y. Huang and T. Liu, “Optical Fiber Distributed Vibration Sensing Using Grayscale Image and Multi-Class Deep Learning Framework for Multi-Event Recognition,” IEEE Sensor Journal, 2021.
42. Arthur D, Vassilvitskii S. “k-means++: The advantages of careful seeding,” Stanford, 2006.

Figures

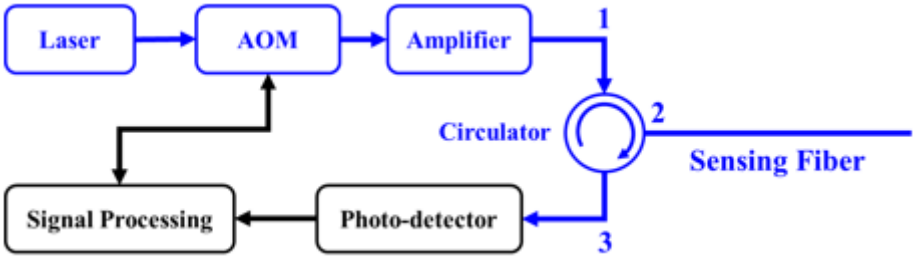

Figure 1

The schematic diagram of φ-OTDR. AOM: acousto-optic modulator.

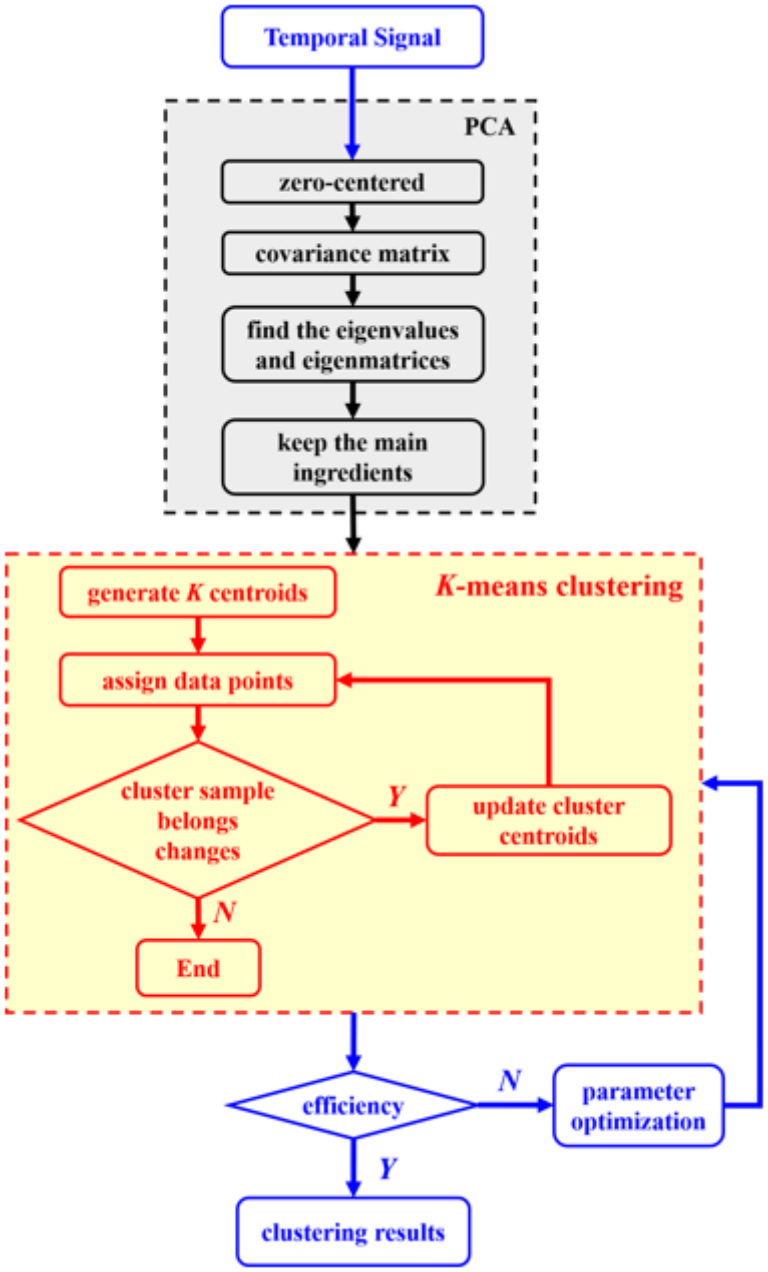

Figure 2

The process of PCA and clustering.

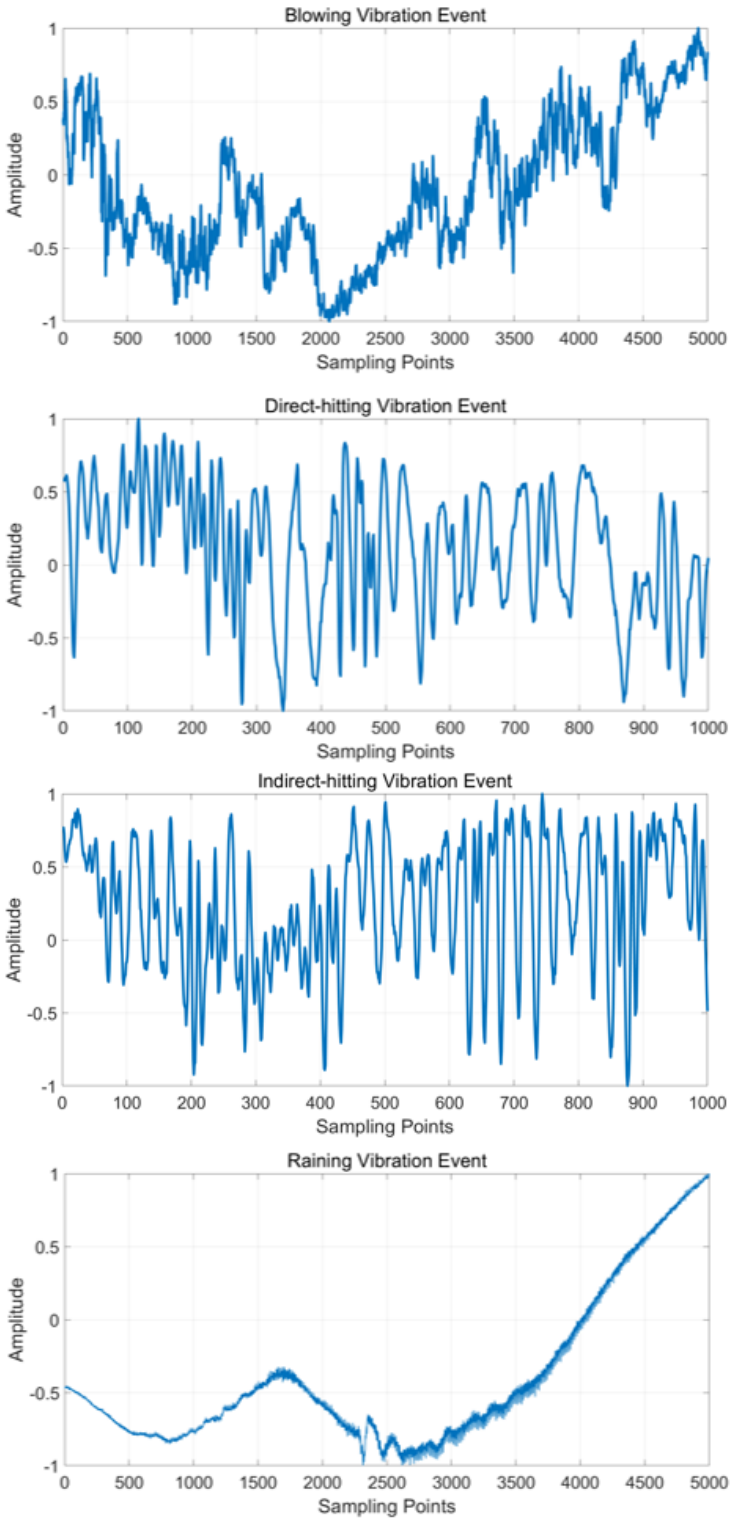

Figure 3

The time-domain waveforms of four vibration events

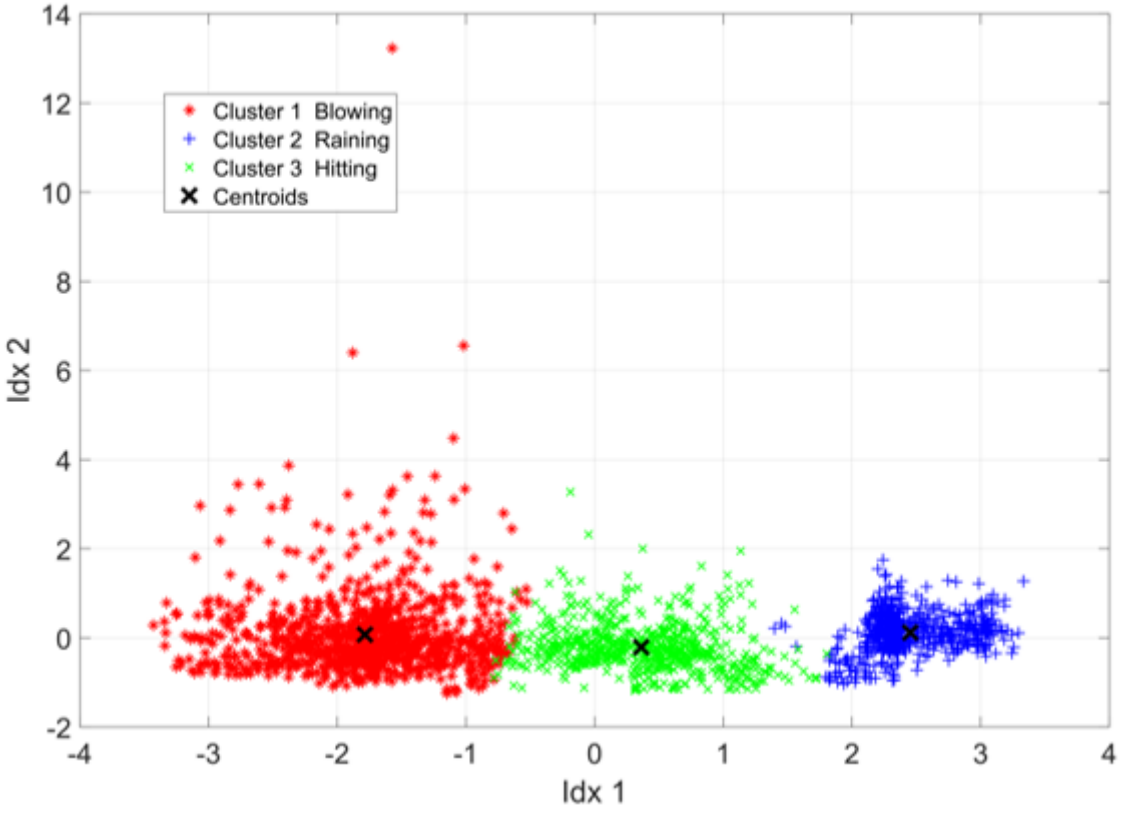

Figure 4

Clustering result graph where K= 3 and the feature is 4.

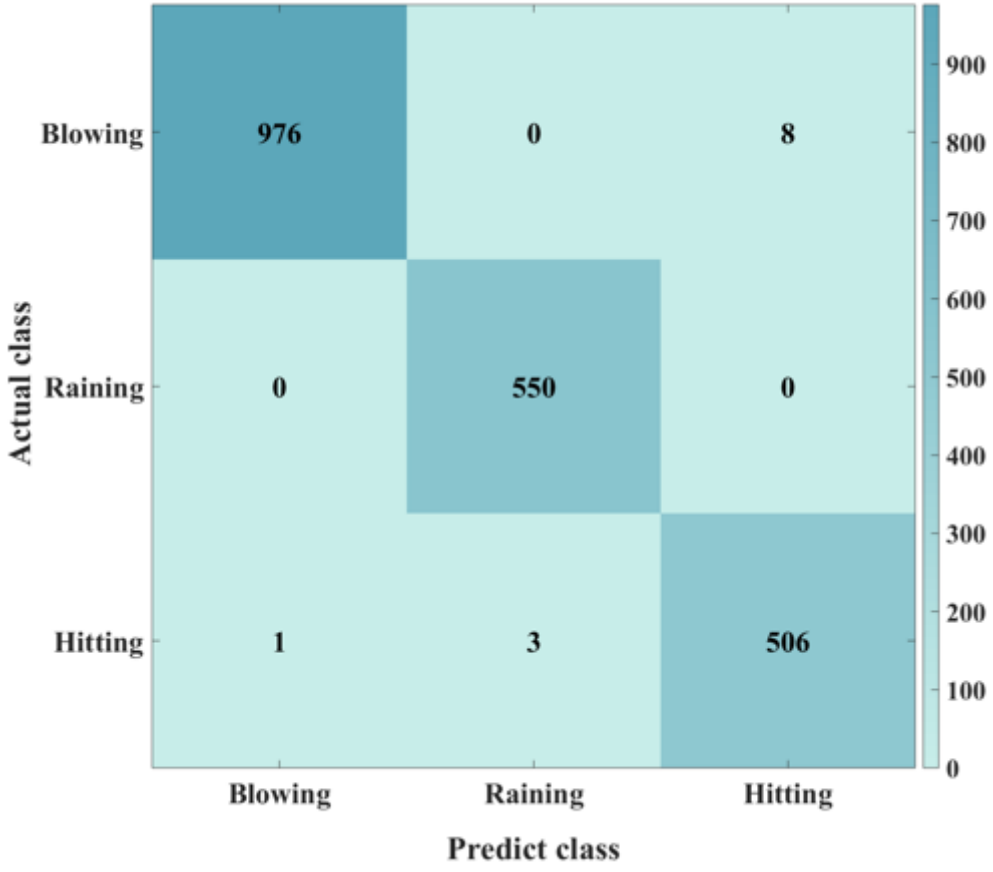

Figure 5

The confusion matrix for classifications of three vibration events.

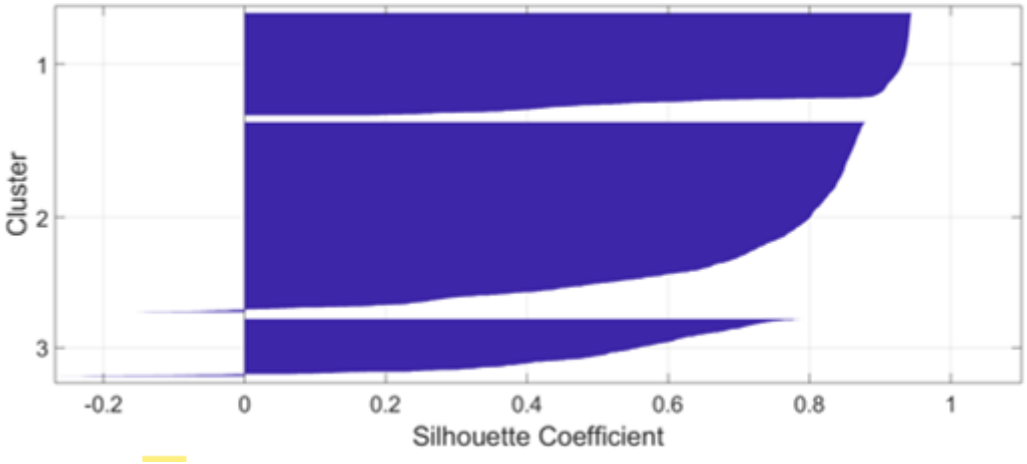

Figure 6

The Silhouette plot where K= 3 and the feature is 4.

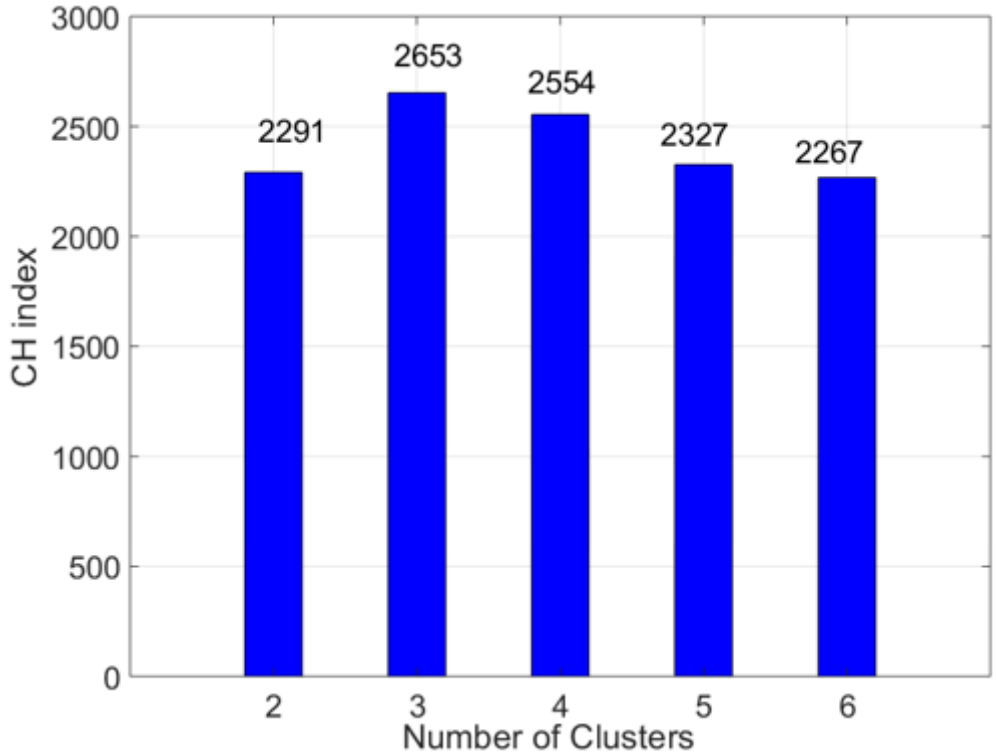

Figure 7

CH index with different K.