Research Article | Vol. 28, No. 3 / 3 February 2020 / Optics Express 2699
Optics EXPRESS

Segregating the true perturbation position from ghost energy points region in ϕ-OTDR systems

MUHAMMAD ADEEL,[1] JAVIER TEJEDOR,[2] JAVIER MACIAS-GUARASA,[3] CHAO SHANG,[4,5,6,*] AND CHAO LU[1,5]

[1] *Photonics Research Centre, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China*
[2] *Fundación Universitaria San Pablo CEU, Madrid, Spain*
[3] *Universidad de Alcalá, Madrid, Spain*
[4] *Photonics Research Centre, Department of Electrical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China*
[5] *The Hong Kong Polytechnic University, Shenzhen Research Institute, 518057 Shenzhen, China*
[6] *Key Laboratory of Luminescence and Optical Information, Ministry of Education, Institute of Optical Information, School of Science, Beijing Jiaotong University, Beijing 100044, China*
[*] *cshang@bjtu.edu.cn*

Abstract: ϕ-OTDR perturbation detection applications demand optimal precision of the perturbation location. Strategies for improving both signal-to-noise (SNR) and precision of the perturbation location in a laboratory environment may fail when applying to a very long fiber under test (FUT) in real-field environments. With this deployment, meaningful energy points representing the response of a certain perturbation can be located at random locations of the fiber other than the original location of the perturbation. These random locations are referred to as the ghost energy points that confuse the system to mistakenly consider the location of these points as the original perturbation location. We present in this paper a novel space-time scanning (ST-scan) method that segregates the ghost energy point locations from those of the real perturbation so that the original perturbation location estimation is improved.

1. Introduction

Multi-event recognition applications with distributed acoustic sensing in ϕ-OTDR systems have gained much attention in the last decade [1–3]. This was possible due to the low cost and relatively high sensitivity nature of ϕ-OTDR systems, which involve a wide range of applications including warning systems [4], damage tolerance systems [5], and intrusion detection systems [1,2]. Literature reveals that the perturbation detection system improvement usually comes from signal-to-noise (SNR) improvements [4,6–14], and more robust algorithms for perturbation location [6,9,15]. However, in those previous works, the experimental framework commonly chosen for evaluation has been restricted to the laboratory environment. With this, it is certainly difficult to mimic a real-field setup for keeping a very large gauge-length (ρ), as compared to the spatial resolution (σ) due to space restrictions. Therefore, the aforementioned strategies for perturbation location detection will fail when used in a real-field environment with $\sigma \ll \rho$, since those did not address this configuration. To the best of our knowledge, the exact location detection of the applied perturbation with the condition $\sigma \ll \rho$ in real-field environments has never been investigated so far [1–3,16–19].

In a real-field environment with a very long fiber under test (FUT), one of the major concerns of any application relates to obtaining a better precision in the disturbance location estimation, which is straightforward by keeping σ very small. In these applications, the distributed acoustic sensing is a second major concern which demands a large value for ρ. With these rules in mind,

#378365 https://doi.org/10.1364/OE.378365
Journal © 2020 Received 23 Sep 2019; revised 5 Dec 2019; accepted 21 Dec 2019; published 21 Jan 2020

Optics EXPRESS

a laboratory setup may not be feasible as the condition $\sigma \ll \rho$ does not restrain the effect of the real perturbation to a single position.

Let us consider the perturbation region to define the whole range along the length ρ representing the response of the applied perturbation. Unlike the condition $\sigma \geq \rho$, the response of the fiber along the whole perturbation region is discontinued at random fiber locations. This discontinuity is referred here to as the zero energy points, whereas there are spatial locations that represent the response of the applied perturbation and can be referred to as useful spatial locations with high energy points. By imposing the condition $\sigma \ll \rho$, the probability of getting high energy points decreases to a great extent, which in turn has a severe implication when used in both perturbation detection and recognition applications. For example, in a real-field environment, if a selected spatial point is considered for analysing the initial raw data, one may get no useful data at low or zero energy points with a change in the position of the perturbation source. Besides getting a very small number of energy points, a second drawback of the condition $\sigma \ll \rho$ is that the influence of the perturbation may lead to high energy replicas appearing along time across the entire length ρ for a single source of perturbation, leading to what we will refer to as ghost energy points. The reason is that the length ρ represents the stretched fiber without any discontinuity and any displacement at any position within ρ causes the spread of this effect along the whole length ρ. A simple differential signal makes it difficult to determine the exact location of the source causing perturbation within the entire length ρ due to inherent unwanted effects in these signals, such as insignificant response for a smooth change in the signal, and also due to noise effects.

There are numerous advantages in finding the useful spatial locations with high energy points for $\sigma \ll \rho$. These include extracting useful data in multi-event perturbation recognition applications, obtaining the exact perturbation location, and determining the range where the ghost energy points appear within the limit of the entire length ρ. This paper focuses on finding the true location of the source causing perturbation by exploiting the signal propagation time delay (τ) from the perturbation source to a given point in the fiber within the length ρ from the ghost energy points, and also determines the range where the ghost energy points will be found along the fiber. To do so, the ST-scan method is proposed to conduct a dual scan process in the space (S) and time (T) domains, in order to determine the exact location of the perturbation and the entire range affected by this perturbation from the ghost energy point locations. In the space domain scanning (S-scan), high energy points in the space domain across a certain number of time data sets are obtained, each corresponding to the entire range of the FUT. In the time domain scanning (T-scan), the active region of perturbation (i.e., the region in the time domain that represents the perturbation) is estimated. This can only be carried out on high energy points, which means that the zero energy points cannot provide any information regarding the applied perturbation in the time domain.

Inspired from the simplicity and ease of the cost of installation of direct detected ϕ-OTDR systems in very long FUT applications like pipeline or border security [1,6,20–22], our proposed work has been implemented with a direct detected ϕ-OTDR system. The matched filtering (MF) technique presented in [23] has been exploited within the ST-scan process to alleviate the effect of all possible noise interferences that are the main hurdles in detecting true energy points in the space domain and the impact of the applied perturbation along the time domain.

Experiments were conducted from real field activity carried out by an excavator hitting the ground in certain positions of a 45km-long FUT lying beneath the ground next to a live pipeline (from the dataset described in [17]). Results have shown that the ST-scan method has been proved to be useful for finding the exact location of the source causing the perturbation and determining the range of the fiber affected by this perturbation from the ghost energy points within the length ρ.

The rest of the paper is organized as follows: Next section describes the algorithms and the methodology. Then, the experimental setup, results, and discussion are presented, and the last section provides some final conclusions on the reported work.

2. Algorithmic and methodological proposal

This section describes the problems related to the condition $\sigma \ll \rho$, the algorithmic proposal (including the MF technique), and the methodology of the ST-scan method to solve the associated problems.

2.1. Associated problems

We first start discussing the condition $\sigma \geq \rho$, and then we will present the details for $\sigma \ll \rho$. The first condition is certainly more beneficial for perturbation recognition applications as it provides freedom to choose any point within the perturbation region for further data analysis. A perturbation region in this context refers to the region that shows any signal of perturbation along the fiber due to a single source of perturbation. By following the condition $\sigma \geq \rho$ there is no chance of getting ghost energy points as the energy points do not propagate along the perturbation region.

On the other hand, the condition $\sigma \ll \rho$ is better for perturbation localization applications as the better spatial resolution provides a better perturbation location estimate. This practice is normally preferred in the distributed sensing, where in the quasi-distributed sensing the fiber is stretched at regular intervals. Hence, it is quite difficult to deploy the fully distributed approach and very long distance applications in laboratory environments. For example, as the distance ρ was set equal to 80 times σ in our experiments, if we deploy the same ratio to attain $\sigma \ll \rho$ in a laboratory environment, we must severely decrease the pulse width, which eventually increases the speed requirements of the associated optical components. Even assuming a $10ns$ pulse width to get 1m spatial resolution, $\rho = 80m$ is a quite large requirement for a laboratory environment. For this purpose, the real field data were investigated to detect all the potential problems in a fully distributed DAS system. The effect of an applied perturbation is felt all along the distributed gauge length ρ and this effect is observed to propagate along the distributed distance and time. This unusual effect along a very long distance was termed as ghost energy points in this work. If, for example, the duration of an applied impact of an object to hit the ground is $2ms$, the origin of the ghost energy points in the spatial domain starts after $2ms$ and these energy points propagate till the end of a distance ρ, or until the SNR of these energy points becomes lower than the noise level. The origin of the ghost energy points along the fiber refers to the propagation of the effect of the perturbation along time and across the whole distributed sensing distance. As the pulse repetition rate (PRR) is high enough ($1085Hz$ in our case) as compared to the propagation speed of the effect of the perturbation ($\approx 183m/s$), a single location of the ghost energy point is recorded in many incoming traces. Hence, the phase of the Rayleigh backscattered signals at each ghost energy point changes in a similar way as in an actual energy point (the one generated at the perturbation location).

Assuming the condition $\sigma \ll \rho$ has three main drawbacks: (1) Appearance of spatial points with null energy, referred to as zero energy points in the context of this paper, which are located at random locations all along the whole length ρ (this problem increases with a further increase in ρ); (2) Due to the delay in signal propagation (τ), it is difficult to find the exact location of the source causing the perturbation as the influence of the main source of perturbation normally spreads along the length ρ, creating ghost energy points along that range; (3) Within the τ delay, it is hard to segregate true perturbation locations of the different sources causing perturbation from each other that may appear within the length ρ.

The condition $\sigma \ll \rho$ means that the ratio ρ/σ must be high enough so that separating the ghost energy points from the high energy points is possible. However, there is no exact ratio

Optics EXPRESS

of ρ/σ, but it should be taken into account that the impact of the ghost energy points is more significant for very large ratios of ρ/σ. For example, the propagation speed of the ghost energy points along the distributed gauge length is so fast that with a small ratio (i.e., $\rho/\sigma = 5$), this propagation can be hardly determined for a given number of time data samples. This is due to the duration of the impact of the applied hitting on the ground (with the FUT being around 45km long), that might be long enough for the mentioned ratio. Therefore, as a rule of thumb, for a ratio of more than $\rho/\sigma = 50$ (with the FUT around 50km), the effect of the ghost energy points is generally observed, so that the exact location of the perturbation source can be segregated from a ghost energy point. In our experiments, we have set the ratio $\rho/\sigma = 80$ for the 45km-length FUT, so that the high energy points can be distinguished from the ghost energy points.

2.2. Naming conventions/terminology

We define some basic terminologies (refer to Fig. 1 for graphical details) before different scanning methods are introduced: Let $\boldsymbol{\gamma}_t$ represent each t^{th} time data vector in the time domain, and let the differential signal be defined as $\boldsymbol{\delta}_t[s] = \boldsymbol{\gamma}_t[s] - \boldsymbol{\gamma}_{t+1}[s]$, where s represents the index of each sample in the spatial domain. The differential signals play a very important role in detecting any perturbation by eliminating the common unwanted signals that normally exist at the respective spatial location of each acquired time data vector [24]. A space data set defines a collection of space data vectors along the time domain for several consecutive spatial locations, which is used in the MF technique [23]. When processing differential signals, each processing window is defined to be a collection of all space data sets within two time data sets (as the differential signals process two time data sets in a single operation). When applying the MF technique, the processing window is comprised of β time data vectors with different number of space data vectors depending on the processing approach being used.

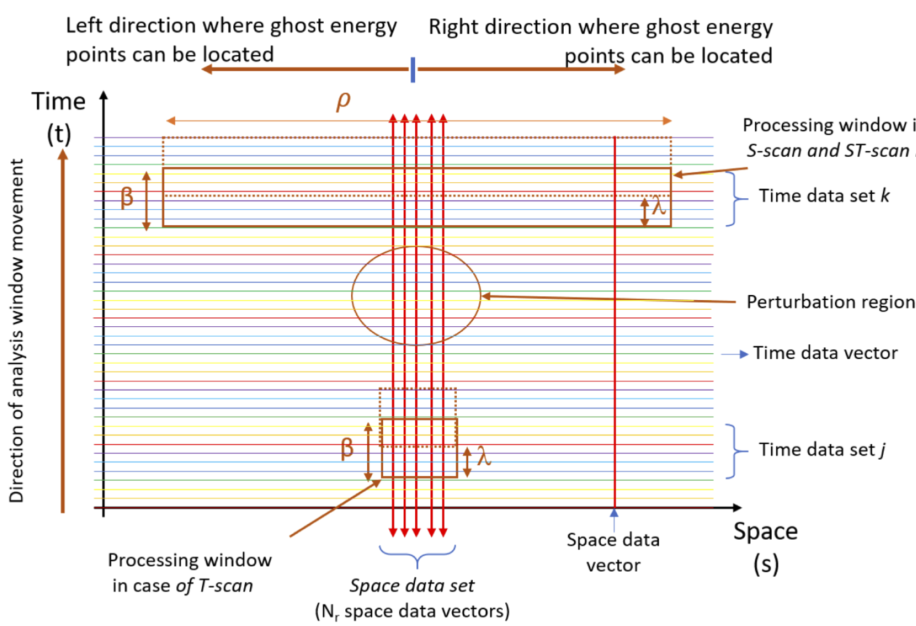


Fig. 1. Name convention definitions.

2.3. Matched filtering technique

An overview of the MF technique is presented here. This algorithm was chosen because of its ability to effectively remove noise effects [23], and hence it can better assist S-scan, T-scan and ST-scan methods. Mathematically, the correlation of adjacent data sets in the space domain for the MF approach is defined as follows:

$$R[n] = 1 - \frac{6\sum_{t=1}^{\beta}\left[\delta_t[n] - \delta_t[n+1]\right]^2}{N_r(N_r^2 - 1)}, \tag{1}$$

where the subscript t varies from 1 to β for a single processing window and N_r represents the entire perturbation range such that $n = 1, 2, \ldots N_r - 1$. During any scanning process, the processing window follows successive shifts in the time domain with magnitude λ as can be seen in Fig. 1, so that a vector $\mathbf{R}_j$ is created for representing the j^{th} processing window.

2.4. S-scan method

The processing window in the S-scan method employs the MF technique and according to Fig. 1, the size of the processing window comprises the entire spatial domain of length ρ. The window is moved along time when the perturbation is applied, and the duration of this window movement must be small enough for obtaining the best resolution of the perturbation location. The width of the processing window in this case is wider enough to cover a major portion of the perturbation region. This window is shifted a few steps with the increment λ and the purpose is to identify the perturbation source location. A drawback of this method is that the system may locate the ghost energy point as the real location of the perturbation if the scan is not started exactly at the time of application of the perturbation. Hence, precise timing information is required as pre-requisite to use this method.

2.5. T-scan method

The processing window in the T-scan method is narrowed down to a length of N_r space data vectors according to Fig. 1. The same MF technique is used to carry out the same operation as that in the S-scan method. In this case, the width of the processing window is narrow enough to cover the whole spatial resolution σ or a region of it. The window is shifted with an increment λ in the time domain all across the data at any of the high energy points with the purpose of determining when the perturbation was applied. In this case, previous information about the high energy point positions is a pre-requisite for this method.

2.6. ST-scan method

Both S-scan and T-scan methods depend on each other for fulfilling the pre-requisite information. Although these techniques might have their own benefits depending on the application, in a realistic application, none of this required information (perturbation timing and high energy point locations) would be known in advance, so that we propose the ST-scan method as the best alternative.

This method aims to find the exact location of a certain source causing the perturbation and employs a combination of the MF technique in the S and T stages of the processing scan, and a simple maximum likelihood estimation (MLE)-based algorithm to detect the exact location of the source causing the perturbation without requiring any pre-requisite information.

In case of ST-scan, the size of the processing window is the same as in the case of S-scan, but it differs in the time duration for which it is shifted. Normally, this duration is made similar to the time duration of the perturbation propagation effects to the extreme points within the length ρ. During the ST-scan process, the processing window follows successive shifts in the time domain

Optics EXPRESS

with magnitude λ as can be seen in Fig. 1, so that a vector $\mathbf{R}_j$ with $N_r - 1$ elements is created for representing each j^{th} processing window.

We can split the process of ST-scan method in two steps: In the first step (Step 1), 2-dimensional data are formed when the MF processing window is moved along the time domain. We define the x-domain of the resultant 2-dimensional matrix to represent the spatial points and the y-domain of the same matrix to define the time domain. In the second step (Step 2), the MLE algorithm is applied and the 2-dimensional matrix is converted into a 1-dimensional vector. If the spatial located points without any perturbation are represented by the zero energy points (ZEP), μ_{ZEP} and σ^2_{ZEP} will represent the average and variance, respectively, of the amplitude of $\mathbf{R}_j$ at the zero energy points. The subset of a single observation $\mathbf{X}_{ML}(j)$ is then obtained after applying the MLE algorithm on a certain j^{th} processing window, as follows:

$$\mathbf{X}_{ML}(j) = \underset{\mu_{ZEP} \in \mathbf{R}}{arg\,max} \left[p \left(\mathbf{R}_j | \mu_{ZEP}, \sigma^2_{ZEP} \right) \right]. \tag{2}$$

The term $p(\mathbf{R}_j|\mu_{ZEP}, \sigma^2_{ZEP})$ is the likelihood of the occurrence of $\mathbf{R}_j$ whether the perturbation is applied or not and it follows a normal distribution ($\mathbf{R}_j \sim \mathcal{N}(\mu_{ZEP}, \sigma^2_{ZEP})$) in the case of the non-perturbation region where all the elements within a single observation represents the ZEP elements only.

Each j^{th} processing window provides non-zero energy point information at a different set of spatially located points n from its subsequent counter-part due to the propagation of the effect of the perturbation with the condition $\sigma \ll \rho$ after the perturbation is applied. Due to this reason, a threshold η can be defined so that $\mathbf{X}_{ML}(j) \neq \mu_{ZEP}$, being n a specific set of spatial points for which $\mathbf{R}_j(n) > \eta$. The number of elements in the x-domain of the 2-dimensional matrix is first shrunk to a random number of elements with the help of the mentioned threshold to represent only those indices of this domain that exceed μ_{ZEP}. Other operations like averaging, maximization, and minimization are then applied to determine a single estimated location for the j^{th} observation. The whole mechanism outputs a 1-dimensional vector where each element represents an estimated spatial location that is affected due to the applied perturbation whereas the index of this vector represents the time domain. Each spatial energy point can be easily mapped with the time index to represent the propagation of these energy points along the whole length ρ.

The conversion of the 2-dimensional matrix to the 1-dimensional vector in the ST-scan process can be visualized in Fig. 2. This figure shows a matrix with values represented by X (those corresponding to $\mathbf{X}_{ML}(j)$ in Eq. (2)). For simplicity, only three values are shown in each observation. However, in a real scenario, the number of boxes with X values randomly varies for each observation. It must be noted that this ST-scan process is not just a scanning process like S-scan or T-scan processes, but a combination of a two-step process, and the node of the V-shape data in the matrix is then exploited as a separation between left and right side data. The data at the left side of the node of V-shape data are termed as 'Lagging end data', whereas the data at the right side of the node of the V-Shape data are termed as 'Leading end data'. Similarly, the fiber at the left side of the real perturbation location is termed as 'Lagging end of the fiber', whereas the fiber at the right side of the real perturbation location is termed as 'Leading end of the fiber'. This matrix is converted to a 1-dimensional vector from *Mean*, *Maximum* and *Minimum* operators, which are applied to the indices of the matrix so that each Y value in Fig. 2 represents the spatial domain data. Instead of a single vector, the same matrix can be divided into two vectors (depending on the left and right side data of the matrix).

To summarize, the ST-scan strategy conveys two objectives: (1) segregating the originally perturbation location from that of the ghost energy points; and (2) finding the entire range of the perturbation caused by the source. These objectives will be shown to be fulfilled in section 3.

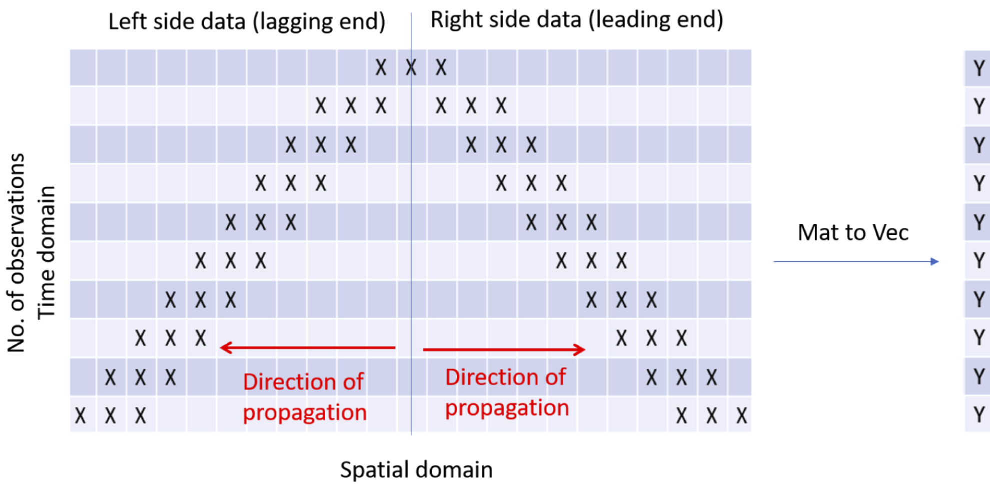


Fig. 2. An overview of the ST-scan process. 'No.' stands for number, 'Mat' for matrix, and 'Vec' for vector.

2.7. Effect of MF on scanning techniques

The MF technique is used on top of differential signals in any scanning processes. From [23], we know that the adjacent correlation of space data sets in this technique provides a better amplitude translation as compared to any other technique (e.g., level crossing, short-time fast Fourier transform, discrete wavelet transform). With $\sigma \ll \rho$, this translation boosts the level at the certain spatial locations within the perturbation region, named as high energy points. By integrating the MF technique within any of the scanning processes, the effect of noise is alleviated at high energy points because the respective elements of all space data sets within the spatial resolution of these points change proportionally. The effect of noise in the respective elements of each adjacently correlated space data set is not proportional and hence degrades the correlation. It implies the non-perturbation region always shows a degradation in the correlation whereas the perturbation region shows a boost in these values [23]. With the correlation operator, the non-consistent noise effects are filtered out, leaving behind the effect due to the high-consistent elements that change proportionally with the applied perturbation.

3. Experimental setup, results, and discussion

3.1. Experimental setup

For our experiments, we have used a direct detected ϕ-OTDR system called FINDAS, described in [20].

To evaluate our proposal, the database described in [17] has been used, specifically the real field recordings of an excavator (a 5 ton Kubota KX161-3) hitting the ground with the shovel. We selected two different locations, located at LOC1 = 23.9km and LOC2 = 27.8km from the FINDAS sensor (which was connected to a 45km-long fiber optic deployed along a live gas pipeline), under different soil and weather conditions.

As reported in [17], in the recording protocol for the database used, the first step was defining a reference meter position that was manually chosen to be the closest to the center of the operation area with good sensitivity, by real time monitoring of the fiber response while a well-defined activity was being carried out. Taking this reference position as the middle position for the recordings, 400 meters were recorded (with a 1 meter readout resolution), 200 meters at each side of the reference position, so that 400 space data vectors were generated for each of the recorded activities (which are split in chunks of 20 seconds for better signal management and storage). The PRR of the injected pulses is 1085Hz. Each sample point from DAQ card with speed of

100MHz represents a distance of 1m which makes a total of 5 resolution points (i.e., $\sigma = 5$m) and the pulse width of each injected pulse is 50*ns*. Therefore, the given configuration satisfies the condition $\sigma \ll \rho$ required for the appearance of the ghost energy points. Specifically, the data used in this paper consist of 4 space data sets (each with 400 space data vectors) in the location named LOC1 and 3 space data sets (each with 400 space data vectors) in the location named LOC2. Each space data vector contains 21700 samples (i.e., 20 seconds times 1085Hz PPR). The actual duration of each impact carried out by the excavator when hitting the ground is around 0.2 seconds. For each j shift of the analysis window using a time displacement of λ, we define $\boldsymbol{\delta}_j$ and $\boldsymbol{R}_j$ as the base and MF representation vectors for each processing window, respectively.

The values for λ and β for all the experiments have been chosen based on preliminary experiments based on the following: The ghost energy points move along the distributed gauge length ρ such that these points appear at random positions along the whole fiber length. Setting the value of λ as low as possible (i.e., $\lambda = 1$), we can provide information regarding the location of all possible ghost energy points, but this increases the processing cost. An optimum configuration for λ is $\lambda \leq \beta$. However, for low ρ values, setting $\lambda > 1$ gives very few observations to the matrix in Fig. 2, which may affect the calculations for determining the exact location of the perturbation. To obtain the best performance with minimal processing delay, we have chosen the most appropriate value of λ in each of the methods ($\lambda = 1$ for S-scan, $\lambda = 10$ for ST-scan and $\lambda = 20$ for T-scan) based on the results from preliminary experiments. Very large values for β are preferred in the absence of ghost energy points as the larger the β, the better the correlation results will be. However, the use of a large value of β makes the result worse, due to the involvement of the sub-parts of two different ghost energy point sets within a single processing window during Step 1 of ST-scan process. Moreover, a larger value of β in T-scan might result in the combination of two adjacent impacts within a single processing window, which may mislead the results by showing a single impact instead of two.

3.2. S-scan results

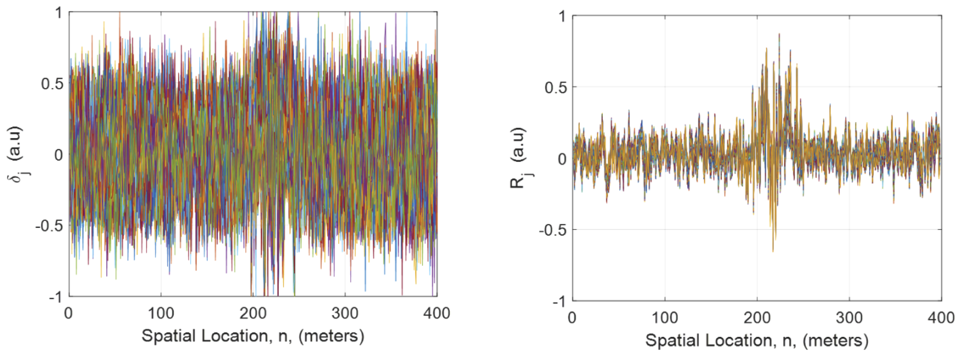
An optimum value of the S-scan process (both window size and window shift) must be less than or equal to the lowest impact hitting duration. It must be noted that the same length of a perturbation region is obtained if the S-scan size is less than or equal to the impact duration. However, the length of a perturbation region is increased due to the addition of the ghost energy points if the size of the S-scan is greater than the duration of the impact. Therefore, we set the time duration at 0.2sec, which allows our system to locate the hitting perturbation at a single location. This means a single processing window with $\beta = 100$ and $\lambda = 1$ creates 100 window shifts that represent 0.2sec. Figure 3 shows the comparison between the differential base signals corresponding to the high energy points and the effect of the MF-based technique applied on them. It can be seen that the MF-based technique is able to separate non-perturbation and perturbation regions better than the differential base signals. This is mainly due to the noise removal effect with this method.

By taking the average of all time data vectors in the S-scan results from both the differential base signals and the differential signals processed with the MF technique, Fig. 4 provides a better comparison for both signal sets. The energy point at the spatial location of 224m in Fig. 4(b) has a larger magnitude than that of the nearby energy points, which clearly shows the exact position of the hitting carried out by the excavator. The window is shifted for a very small duration of 0.1sec and the data are considered exactly at the occurrence of the first hit. Figure 4(b) shows the disturbance near the 224m location due to the said reason, and the width of this disturbance would have been wider if the time duration had been longer than 0.1sec.

3.3. T-scan results

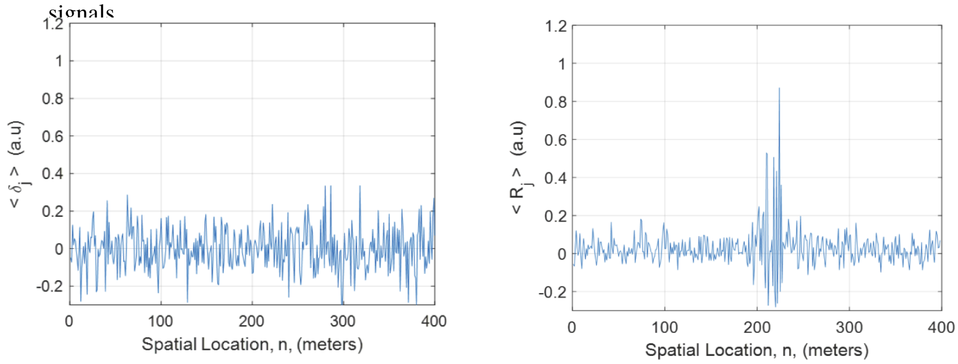
In these experiments, T-scan has been applied to examine the effects of all the hittings on the ground carried out at LOC1. To do so, the window shift should be set to an extent so that it

Optics EXPRESS



(a) (b)

Fig. 3. Relevant signals by considering 200 time data vectors and $\beta = 100$ around the hitting location at 240m for a single impact of excavator hitting before the effect is transmitted to other regions of the fiber at LOC1. (a) $\boldsymbol{\delta}_j$ differential base signals, (b) $\boldsymbol{R}_j$ MF differential signals



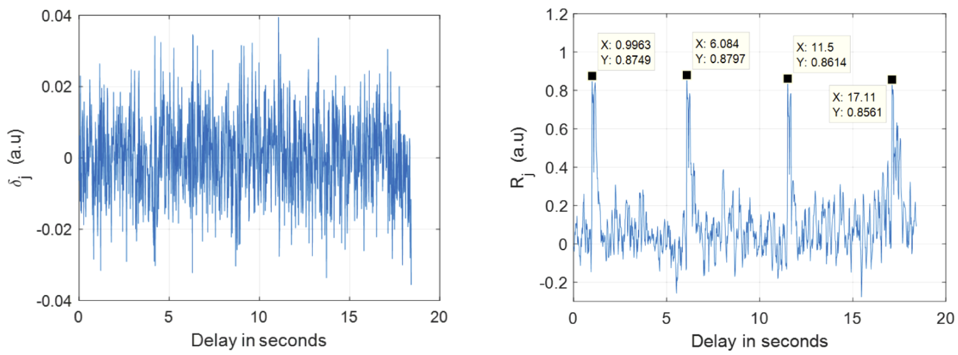
(a) (b)

Fig. 4. Average results with S-scan at the real location of perturbation at LOC1 by considering 200 time data vectors for (a) Differential base signals, (b) MF differential signals.

covers the entire time data set and the window size must be less than or equal to the duration of the impact. The width of the processing window is narrowed down to fit to a certain spatial range of the hitting disturbance which has been set to $N_r = 3$ in our case. The processing window is shifted $\lambda = 20$ along the entire time domain data of 20sec to locate the time instants at which the hitting perturbations were applied. Figure 5 shows the results when using differential base signals and MF differential signals. These show better performance for the MF differential signals, since the impact of each hitting on the ground can be clearly seen in the time domain.

We know that, relative to the PRR, the propagation speed of the applied impact across the length ρ is quite low. This means that the spatial location, if seen from the perspective of a single data trace affected at the reference perturbation location cannot be seen at another spatial location at the same time. After a certain delay, one could see the effects of the perturbation at another location far away from the original perturbation position (i.e., at the so-called ghost energy points). Consider for instance the spatial location at the ghost energy points of 78m and 246m, where the effect of the perturbation from the original position of 224m is received at a certain delay. If the distance of 78m is considered to be the ghost energy point named position-1, and the distance of 246m is considered to be the ghost energy point named position-2, Figs. 6(a) and (b) represents the T-scan results at the energy points of these positions with $\lambda = 20$. The peak that corresponds to the time instant in which the hitting is performed has been considered as the initial time to identify the delay of the ghost energy points. Comparing with the results

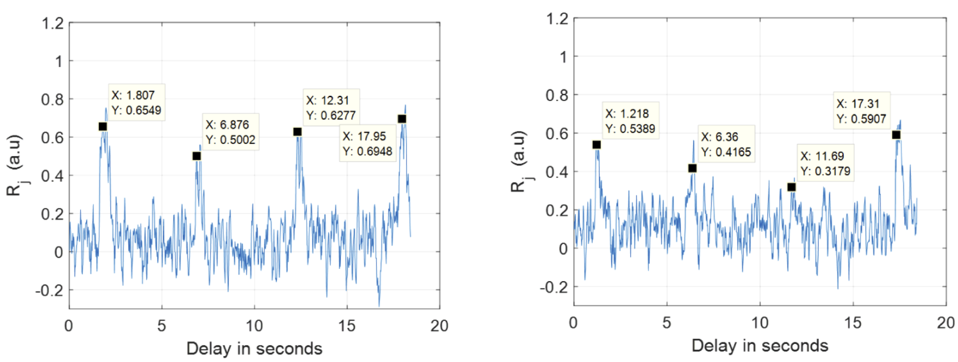
Optics EXPRESS



(a) (b)

Fig. 5. T-scan results at the highest energy point near the real source of perturbation by moving sliding window at LOC1 with λ=20, and β=100 for (a) Differential base signals, (b) MF differential signals.

presented in Fig. 5(b), all the hitting points in the time domain for position-1 and position-2 are delayed from the original position of perturbation with delays around $\tau_1 = 0.8$sec and $\tau_2 = 0.2$sec, respectively. For example, the original signal was detected at 0.9963sec in Fig. 5(b), and at 1.807sec in Fig. 6(a), making a delay $\tau_1 = 1.807 - 0.9963 = 0.8107$; and at 1.218sec in Fig. 6(b), making a delay $\tau_1 = 1.218 - 0.9963 = 0.0.2217$. For the 2$^{\text{nd}}$, 3$^{\text{rd}}$, and 4$^{\text{th}}$ hits, similar delays are observed.



(a) (b)

Fig. 6. T-scan results based on MF differential signals at LOC1 with the highest energy point near the spatial location of (a) Position-1=78m, (b) Position-2=246m.

3.4. ST-scan process

The S-scan and T-scan methods worked with time-related and location-related pre-requisites, respectively. In a real-world scenario, we cannot fulfil these pre-requisites and hence it is important to work with the dual scanning process that, combined with the MLE algorithm, will provide the true location of the perturbation without these pre-requisites.

We can split the process of the ST-scan method in two steps.

3.4.1. Step 1

The processing window moves along the time domain with width ρ=400m and length β=100 to create a 2-dimensional matrix. Let the first dimension of this matrix represent the spatially located 399 points after the MF technique is applied. For the second dimension of the matrix, we consider a total of 1000 time data vectors (around 1 second time) and λ = 10, which produce 100 observations. The purpose of attributing 1 second duration is that the propagation of the perturbation up to the extreme length of the perturbation region is around 1 second. The window is then shifted in the time domain, and at each j^{th} shift with an increment λ = 10, the vector $\mathbf{R}_j$ (comprised of 399 spatial points) is calculated. With these parameters, each observation vector (whose length is 399) represents the spatial domain data considering β=1000 time domain samples and 400 spatially located points.

3.4.2. Step 2

The MLE algorithm is applied to convert the 2-dimensional matrix into a 1-dimensional vector, as presented in subsection 2.6. With the MLE, the amplitude of the data for a single observation with the condition η>0.5 contains a few of the high energy points at a specific group of locations n, whereas the rest of the 399 points are attributed a value of zero.

After the MLE algorithm is applied, the number of high energy point positions for a single observation varies randomly, where each observation can be taken as the time information. Instead of a fixed number of 399 spatially located points, we get a varying number of spatial points as there are very few spatial points at a specific time duration that satisfy the MLE-based classification of high energy points. At this stage, the number of spatially located high energy points varies from one observation to another. For the sake of simplicity, we calculate the *Mean*, *Maximum* and *Minimum* values of the positions corresponding to these high energy points above the threshold, as a function of the delay in which they were found. This means that the amplitude of the MF-based results exceeds the threshold for a very few number of indices at a certain j^{th} shift, and as the number of indices varies for each shift, either of the three operations (*Mean*, *Maximum*, *Minimum*) converts a group of a variable number of data samples to a single value using the standard equations to calculate the mean, maximum, and minimum respectively.

3.5. ST-scan results and analysis

The two steps mentioned in the previous subsection showed that the ST-scan method outputs a 1-dimensional vector. The values of this vector represent the spatial domain data and the index of each element is a direct translation of the time domain data. After converting the 2-dimensional matrix into a 1-dimensional vector as a result of the three operators (i.e., *Mean*, *Maximum*, *Minimum*), Fig. 7 provides a rough estimate of the high energy locations versus the delay required to reach these positions from the original location of the applied perturbation, which represent the ghost energy points. For each of these operations (*Mean*, *Maximum*, *Minimum*), we get 100 observations, each with a single value, being this the reason why the 2-dimensional matrix to 1-dimensional vector conversion was carried out.

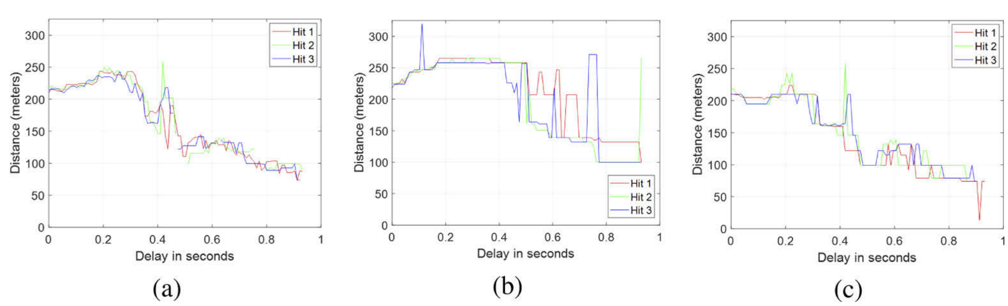


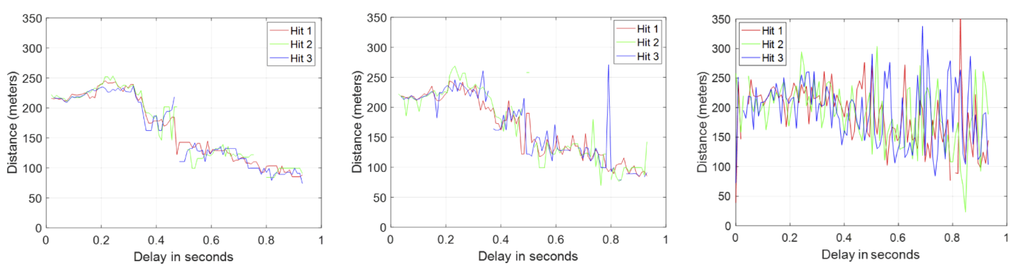
Fig. 7. ST-scan results for different hittings at LOC1 with λ=10 and β=100, showing the (a) Mean, (b) Maximum, (c) Minimum values of the variable vector size of high energy spatial points with the delay for each processing window.

Optics EXPRESS

Figure 7(a) represents the *Mean* position calculated over all the position estimates on each processing window to provide a rough estimate of the influence of the applied impact at a given delay. Therefore, each value in the figure represents the position versus delay of either a ghost energy point or the original location of the applied perturbation (delay=0). We consider the lagging end of the fiber from the original location of the perturbation which means that only this side of the fiber provides a large affected length.

Figure 7(b) shows the *Maximum* position value obtained for each delay, which may give a more accurate estimate of the location of the original perturbation source (224m). We have taken a maximum value to estimate a high accuracy in the original location of the perturbation because in this case, the perturbation propagation is mostly at the lagging end the fiber. *Minimum* values for each observation are presented in Fig. 7(c), which shows the extreme position at the lagging end of the original location of the perturbation in the range where the influence of the applied perturbation is sensed.

The term β plays an important role to decide the number of time data vectors within each processing window. Decreasing this value increases the chance to detect the exact location of the perturbation but a drawback of lowering β is that the extreme locations, where the effect of perturbation is felt, cannot be assessed. Figure 8, which is obtained with the *Mean* operator, shows that the initial perturbation location of 224m was detected precisely using all the hitting results by lowering β (i.e., the distance values are centered in the true perturbation location) but at the same time, a smooth graph cannot be displayed with increasing β (i.e., the distance values spread around different locations that of the true perturbation location for large delay values), which represents the ghost energy point locations at the entire perturbation region. The β value optimization falls outside the scope of this paper, hence is proposed for future work.



(a) (b) (c)

Fig. 8. Original and ghost energy point locations for different hittings at LOC1 for λ=10 with (a) β=75, (b) β=50, (c) β=25.

Except for the energy points near zero delay, Figs. 7 and 8 have shown the overall picture of the ghost energy point location in a single graph by considering the perturbation at LOC1 after the matrix in Fig. 2 is converted to a single vector. The y-axis of these figures provides the information regarding the ghost energy points, which occur either at the lagging or leading ends of the fiber. These ends (lagging or leading) correspond to both sides shown in Fig. 2, so that instead of getting a single 1-dimensional vector, two vectors (based on lagging and leading end data) are obtained to prove the usability of the ST-scan method in a further analysis. This analysis provides a clearer picture of the real perturbation location with the symmetry that may occur in those figures from both lagging and leading end data. As the ghost energy points at the leading end of the fiber spread a small number of spatially located points, it is easy to visualize the perturbation locations at the lowest distance location points in Figs. 7 and 8. Therefore, it can be seen a slight raise in the spatial points between 0.2sec and 0.3sec that represents the effect of the perturbation at the leading end of the fiber.

Optics EXPRESS

To visualize the ghost energy points at both lagging and leading ends of the fiber, three different graphs are provided in Fig. 9. This figure follows the *Mean* strategy of Fig. 7 by considering the first hit at LOC1 at either lagging or leading ends of the fiber. By taking the original location of the perturbation in Fig. 9(a), it can be seen that the locations of the ghost energy points are quite close to both sides of the perturbation. Assuming the original perturbation location to be at the leading end of fiber in Fig. 9(b), it can be seen a right shift in the locations of all the ghost energy points at the leading end of fiber. Similar results can be observed in Fig. 9(c), by assuming the location of the perturbation at the lagging end of the fiber. These figures clearly show the original location of the perturbation at 224m of the fiber. In order to further validate our results, we consider the other location where recordings were carried out (i.e., LOC2), which exhibits the ghost energy points distributed at almost the same far extreme points of either side (leading and lagging ends) of the original perturbation location. Figure 10 shows a similar behaviour as that of Fig. 9 and provides the best location estimate at 210m in LOC2. As Figs. 9 and 10 work by plotting the two vectors, each as a result of the *Mean* operator from both lagging and leading end data, it can be clearly seen that the data from both vectors are symmetric with respect to a common position by considering the real location of perturbation. A non-symmetric graphic is obtained if the location of perturbation is assumed to be at either lagging or leading ends of the fiber. Therefore, it can be claimed that no pre-requisite is required in this case and the location of the real perturbation can be determined from the corresponding perfect symmetry of both sides of the data.

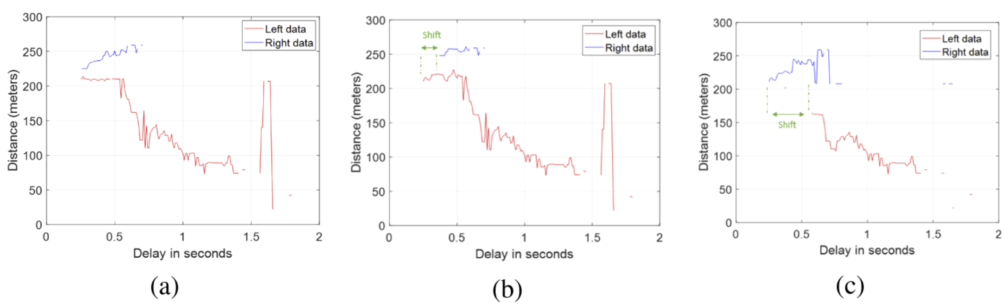


Fig. 9. Original and ghost energy point locations for left data (lagging end data) and right data (leading end data) at LOC1 by assuming the location of perturbation at (a) 224m, (b) 244m, (c) 204m.

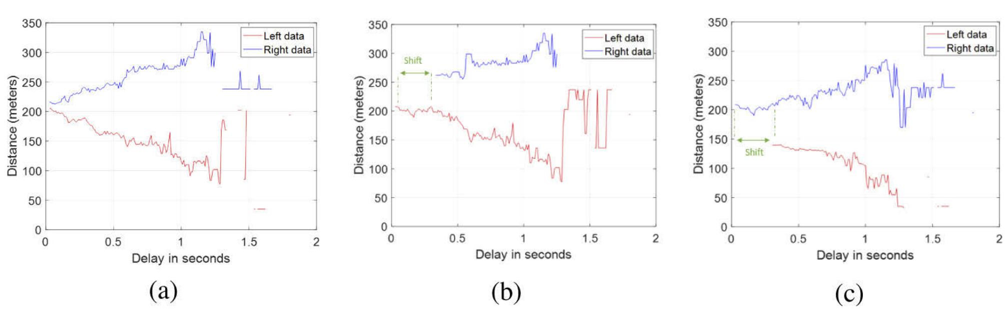


Fig. 10. Original and ghost energy point locations for left data (lagging end data) and right data (leading end data) at LOC2 by assuming the location of perturbation at (a) 210m, (b) 250m, (c) 160m.

4. Conclusions

This paper has presented the ST-scan method that can be used to detect the original location of an applied perturbation nearby a FUT and determine the so-called ghost energy points that emerge in random fiber locations due to the applied perturbation, and may confuse the system when estimating the real perturbation location.

A real-field experiment has been carried out by fulfilling the condition $\sigma \ll \rho$, which is a common practice in pipeline and border security systems. Addressing the issue of slow propagation of a certain perturbation in the time domain, the ghost energy points are established. These energy points represent a certain perturbation at different locations of the fiber within the length ρ. Different scanning techniques like S-scan and T-scan are implemented to see the original location of perturbation and the ghost energy points. However, as these techniques depend on pre-requisite requirements, it is not possible to implement them in a real-time analysis for determining the exact location of perturbation with the condition $\sigma \ll \rho$. Thus, the ST-scan method is proposed for finding the exact location of the perturbation with the condition $\sigma \ll \rho$. With this method, not only the exact location of the perturbation can be determined, but also the ghost energy points can be estimated. Moreover, this method has proved to be useful combined with the MF technique on top of differential signals to remove unwanted noise effects.

Future work will focus on determining the optimum values of λ and β parameters within the ST-scan method to obtain better results.

Funding

Ministerio de Economía y Competitividad (ARTEMISA (TIN2016-80939-R), HEIMDAL-UAH (TIN2016-75982-C2-1-R)); Universidad de Alcalá (ACERCA (CCG2018/EXP-019)); General Research Fund (52168/17E, PolyU 152658/16E); National Natural Science Foundation of China (U1701661).

Acknowledgments

Assistance in acquiring the requisite data was provided by the University of Alcalá, Madrid, Spain.

Disclosures

The authors declare no conflicts of interest.

References

1. J. Tejedor, J. Macias-Guarasa, H. F. Martins, D. Piote, J. Pastor-Graells, S. Martin-Lopez, P. Corredera, and M. Gonzalez-Herraez, "A novel fiber optic based surveillance system for prevention of pipeline integrity threats," Sensors **17**(2), 355 (2017).
2. J. Tejedor, J. Macias-Guarasa, H. F. Martins, J. Pastor-Graells, S. Martín-López, P. C. Guillén, G. D. Pauw, F. D. Smet, W. Postvoll, C. H. Ahlen, and M. González-Herráez, "Real field deployment of a smart fiber-optic surveillance system for pipeline integrity threat detection: Architectural issues and blind field test results," J. Lightwave Technol. **36**(4), 1052–1062 (2018).
3. Q. Sun, H. Feng, X. Yan, and Z. Zeng, "Recognition of a phase-sensitivity OTDR sensing system based on morphologic feature extraction," Sensors **15**(7), 15179–15197 (2015).
4. F. Peng, H. Wu, X.-H. Jia, Y.-J. Rao, Z.-N. Wang, and Z.-P. Peng, "Ultra-long high-sensitivity ϕ-OTDR for high spatial resolution intrusion detection of pipelines," Opt. Express **22**(11), 13804–13810 (2014).
5. Z. Qin, L. Chen, and X. Bao, "Continuous wavelet transform for non-stationary vibration detection with phase-OTDR," Opt. Express **20**(18), 20459–20465 (2012).
6. H. Wu, S. Xiao, X. Li, Z. Wang, J. Xu, and Y. Rao, "Separation and determination of the disturbing signals in phase-sensitive optical time domain reflectometry (ϕ-OTDR)," J. Lightwave Technol. **33**(15), 3156–3162 (2015).
7. Y. Lu, T. Zhu, L. Chen, and X. Bao, "Distributed vibration sensor based on coherent detection of phase-OTDR," J. Lightwave Technol. **28**(22), 3243–3249 (2010).
8. M. Ren, P. Lu, L. Chen, and X. Bao, "Theoretical and experimental analysis of ϕ-OTDR based on polarization diversity detection," IEEE Photonics Technol. Lett. **28**(6), 697–700 (2016).

Optics EXPRESS

9. P. Stajanca, S. Chruscicki, T. Homann, S. Seifert, D. Schmidt, and A. Habib, "Detection of leak-induced pipeline vibrations using fiber-optic distributed acoustic sensing," Sensors **18**(9), 2841–E284118 (2018).
10. X. He, S. Xie, F. Liu, S. Cao, L. Gu, X. Zheng, and M. Zhang, "Multi-event waveform-retrieved distributed optical fiber acoustic sensor using dual-pulse heterodyne phase-sensitive OTDR," Opt. Lett. **42**(3), 442–445 (2017).
11. X.-P. Zhang, W.-Y. Qiao, Z.-H. Sun, Y.-Y. Shan, J. Zeng, and Y.-x. Zhang, "A distributed optical fiber sensing system for synchronous vibration and loss measurement," Optoelectron. Lett. **12**(5), 375–378 (2016).
12. F. Jiang, H. Li, Z. Zhang, Y. Zhang, and X. Zhang, "Localization and discrimination of the perturbation signals in fiber distributed acoustic sensing systems using spatial average kurtosis," Sensors **18**(9), 2839 (2018).
13. Q. Tian, C. Zhao, Y. Zhang, and H. Qu, "Intrusion signal recognition in OFPS under multi-level wavelet decomposition based on RVFL neural network," Optik **146**, 38–50 (2017).
14. T. Zhu, X. Xiao, Q. He, and D. Diao, "Enhancement of SNR and spatial resolution in phase-OTDR system by using two-dimensional edge detection method," J. Lightwave Technol. **31**(17), 2851–2856 (2013).
15. H. Kong, Y. Dong, Q. Zhou, W. Xie, C. Ma, and W. Hu, "Events detection in OTDR data based on a method combining correlation matching with STFT," in *Proc. of Asia Communications and Photonics Conference (ACP)*, (2014), pp. 1–3.
16. F. Peng, N. Duan, Y. Rao, and J. Li, "Real-time position and speed monitoring of trains using phase-sensitive OTDR," IEEE Photonics Technol. Lett. **26**(20), 2055–2057 (2014).
17. J. Tejedor, H. F. Martins, D. Piote, J. Macias-Guarasa, J. Pastor-Graells, S. Martin-Lopez, P. Corredera-Guillén, F. De-Smet, W. Postvoll, and M. González-Herráez, "Toward prevention of pipeline integrity threats using a smart fiber-optic surveillance system," J. Lightwave Technol. **34**(19), 4445–4453 (2016).
18. J. Tejedor, J. Macias-Guarasa, H. F. Martins, S. Martin-Lopez, and M. Gonzalez-Herraez, "A Gaussian Mixture Model-Hidden Markov Model (GMM-HMM)-based fiber optic surveillance system for pipeline integrity threat detection," in *Proc. of International Conference on Optical Fiber Sensors*, (2018), pp. WF36:1–WF36:4.
19. J. Tejedor, J. Macias-Guarasa, H. Martins, S. Martin-Lopez, and M. Gonzalez-Herraez, "A contextual GMM-HMM smart fiber optic surveillance system for pipeline integrity threat detection," J. Lightwave Technol. **37**(18), 4514–4522 (2019).
20. H. F. Martins, S. Martin-Lopez, P. Corredera, M. L. Filograno, O. Frazao, and M. González-Herráez, "Coherent noise reduction in high visibility phase-sensitive optical time domain reflectometer for distributed sensing of ultrasonic waves," J. Lightwave Technol. **31**(23), 3631–3637 (2013).
21. Z. Sha, H. Feng, and Z. Zeng, "Phase demodulation method in phase-sensitive OTDR without coherent detection," Opt. Express **25**(5), 4831–4844 (2017).
22. Z. Qin, H. Chen, and J. Chang, "Detection performance improvement of distributed vibration sensor based on curvelet denoising method," Sensors **17**, 1380 (2017).
23. M. Adeel, C. Shang, K. Zhu, and C. Lu, "Nuisance alarm reduction: Using a correlation based algorithm above differential signals in direct detected phase-OTDR systems," Opt. Express **27**(5), 7685–7698 (2019).
24. J. C. Juarez, E. W. Maier, and H. F. Taylor, "Distributed fiber-optic intrusion sensor system," J. Lightwave Technol. **23**(6), 2081–2087 (2005).