Applied Mathematics and Nonlinear Sciences

<https://www.sciendo.com>

Determination of the minimum distance between vibration source and fibre under existing optical vibration signals: a study

Haitao Li^{1†}, Guoqing Zhang², Nawaf Alhebaishi², Hamdy Abdelhameed³¹ Department of Foundation Teaching, Jiaozuo University, Jiaozuo 454000, China² School of Civil Engineering, Henan Polytechnique University, Jiaozuo 454000, China³ Applied Science University, Al Eker, Kingdom of Bahrain Hamdy

Submission Info

Communicated by Juan Luis García Guirao

Received June 17th 2021

Accepted September 24th 2021

Available online December 13th 2021

Abstract

The article analyses why colleges and universities should strengthen innovation and entrepreneurship education based on ‘mass entrepreneurship and innovation.’ First, we conduct a questionnaire survey on the status quo of college students’ innovation and entrepreneurship and use nonlinear methods to construct an evaluation model of innovation and entrepreneurship capabilities to evaluate students’ innovation and entrepreneurship capabilities quantitatively. Finally, we verify the effectiveness of the combined evaluation model through the data on the innovation and entrepreneurship of college students. The research results can provide a new idea for appraisal of college students’ innovation and entrepreneurship ability.

Keywords: nonlinear model, analytic hierarchy process, colleges and universities, innovation and entrepreneurship ability evaluation
AMS 2010 codes: 62J12

1 Introduction

As a distributed vibration monitoring technology, phase-sensitive optical time-domain reflectometer (Φ -OTDR) has the advantages of corrosion resistance, electromagnetic interference resistance, high sensitivity, wide detection range [1], and easy concealment [2] etc. Φ -OTDR has been successfully used in perimeter security [3], protection of important underground infrastructures [4] and health monitoring of large buildings [5]. Φ -OTDR technology is also a new way of protecting the underground cable. Because most of the cables are buried in areas of high population density, the behaviour of stealing and damaging cable will bring great inconvenience to people’s life and also cause great economic losses. Φ -OTDR, which is used directly on the paved communication

[†]Corresponding author.Email address: lht770408@163.com

optical fibre without additional sensing devices, does not affect the communication on fibre and also protect the fibre itself [6]. Therefore, Φ -OTDR is a convenient, high cost-effective cable protection way.

The primary problem of vibration monitoring is the localisation of vibration events, but the premise of localisation is to measure the distance between the vibration source and fibre. In the past, the research on Φ -OTDR technology focused on the improvement of the detection range [7] as well as the sensitivity [8–10]. As the technology is maturing, the distance measurement to vibration signal through Φ -OTDR becomes possible.

However, the distance measurement accuracy on Φ -OTDR is low due to its high sensitivity along with other factors such as the inherent noises on devices, environmental noises, spatial resolution and human activities' interference. Thus, this paper focuses on efficient distance measurement for vibration events.

2 Methodology

This paper presents the method to determine the minimum distance between vibration source and fibre under existing optical vibration signals.

Some communication fibres are distributed in the dedicated pipe chase, nevertheless, some are directly embedded under the ground. Fibres in pipe chase are not directly contacted with the soil layers. Thus the capacity of sensing vibration is weak. In comparison, fibres under the ground have a stronger ability to sense vibration. The research of this paper is only limited to distance measurement between vibration source and fibre under the ground. In practice, evacuation is the main threat to optical fibres. Evacuation can be divided into shovel evacuation, excavator evacuation and tamping machine evacuation, etc. The experiment of this paper concentrated on shovel excavation due to its high frequency.

2.1 Problem description

Vibration signals from excavation conform to the propagation law in the soil layers. The attenuation is in the exponential law [11], which means that the farther the signal is from the vibration source, the more it attenuates.

The spatial resolution of distributed fibres refers to the minimum spatial distance as measuring vibration along with the fiber distribution. One fibre can be divided into different segments. Each segment is equivalent to a sensor, and the length is considered as spatial resolution. For example, when spatial resolution is 1 m, vibration signals are distinguished on two fibre segments with an interval of 1 m.

After thorough consideration of spatial resolution and vibration attenuation law, the propagation process of distributed fibre for vibration can be described in Figure 1. The segment closest to the vibration source receives the strongest signal, vice versa. For convenient description, the fibre segment closest to the vibration source is numbered as 0—the 0th fibre or the datum fibre segment. The adjacent fibre segments are numbered as -1 and 1 —the -1 st fibre and the 1 st fibre, and so on. Fibres are marked by using the subscript subsequence ($-N \dots -1, 0, 1 \dots N$). For convenient description, such a method is called MMDFS.

Fig. 1 Diagram of optical fibre receiving vibration signal.

The vibration source closest to the fibre segment is l_0 , the distance from each segment to the vibration source

is marked as l_i , ($i = -N, \dots, 0, \dots, N$). Δt_i refers to TDOA between the i th fibre and the 0th fibre. Δt_0 , the TDOA of 0th fibre, is equal to 0. Δt_i is >0 . v represents propagation velocity of vibration. d stands for spatial resolution. The relationship between Δt_i and l_i is

$$\Delta t_i = (l_i - l_0)/v \quad (1)$$

$$l_i = \sqrt{l_0^2 + i^2 d^2} \quad (2)$$

Substitute Eq. (2) into (1), get

$$\Delta t_i^2 v^2 + 2\Delta t_i v l_0 = i^2 d^2 \quad (3)$$

The corresponding equations of all fiber segments are simultaneous; then

$$\begin{cases} \Delta t_1^2 v^2 + 2\Delta t_1 v l_0 = 1^2 \Delta d^2 \\ \Delta t_{-1}^2 v^2 + 2\Delta t_{-1} v l_0 = (-1)^2 \Delta d^2 \\ \vdots \\ \Delta t_N^2 v^2 + 2\Delta t_N v l_0 = (-N)^2 \Delta d^2 \end{cases} \quad (4)$$

can be obtained. Equation (4) has multiple equations, but there are only two unknown parameters l_0 and v . Therefore, Eq. (4) is a non-linear overdetermined equation set.

2.2 Principles

The goal of this paper is to substitute TDOA Δt_i into Eq. (4) to solve l_0 . Since Eq. (4) is a non-linear equation set, solving l_0 directly is complicated. Therefore, Eq. (4) is firstly linearised, then the least square principle is used to deduce a solving equation. To ensure the estimation accuracy of Δt_i , the method of TDOA estimation and eliminating outliers are studied.

Distance Estimation between Vibration Source and Fibre Based on the Least Square

(1) Linearise Equation (4)

A variable substitution is made as follows

$$x_1 = v^2, \quad x_2 = 2v l_0 \quad (5)$$

Then Eq. (4) changes into

$$\begin{cases} \Delta t_1^2 x_1 + \Delta t_1 x_2 = 1^2 \Delta d^2 \\ \Delta t_{-1}^2 x_1 + \Delta t_{-1} x_2 = (-1)^2 \Delta d^2 \\ \vdots \\ \Delta t_N^2 x_1 + \Delta t_N x_2 = (-N)^2 \Delta d^2 \end{cases} \quad (6)$$

Let

$$A = \begin{bmatrix} \Delta t_1^2 & \Delta t_1 \\ \vdots & \vdots \\ \Delta t_N^2 & \Delta t_N \end{bmatrix} \quad X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad b = \begin{bmatrix} 1^2 \Delta d^2 \\ \vdots \\ (N)^2 \Delta d^2 \end{bmatrix} \quad (7)$$

Then Eq. (6) changes into

$$AX = b \quad (8)$$

Equation (8) is a standard linear overdetermined equation. The classical least square method could solve the equation, namely the obtained $\bar{v}$ and $\bar{l}_0$ keep the sum of squared residuals (SSR) at both sides of Eq. (8) minimised. As follows

$$\min(b - AX)^T (b - AX) \quad (9)$$

According to the deduction of the least square, the estimation value $\bar{X}$ of X in Eq. (8) is

$$\bar{X} = (A^T A)^{-1} A^T b. \quad (10)$$

Where

$$A^T A = \begin{bmatrix} \sum_i \Delta t_i^4 & \sum_i \Delta t_i^3 \\ \sum_i \Delta t_i^3 & \sum_i \Delta t_i^2 \end{bmatrix}, \quad A^T b = d^2 \begin{bmatrix} \sum_i i^2 \Delta t_i^2 \\ \sum_i i^2 \Delta t_i \end{bmatrix} \quad (11)$$

Let $a = \sum_i \Delta t_i^4$, $b = \sum_i \Delta t_i^3$, $c = \sum_i \Delta t_i^2$, $e = d^2 \sum_i i^2 \Delta t_i^2$, $f = d^2 \sum_i i^2 \Delta t_i$, then

$$A^T A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}, \quad A^T b = \begin{bmatrix} e \\ f \end{bmatrix} \quad (12)$$

Substitute Eq. (12) into (10), get

$$\bar{X} = (A^T A)^{-1} A^T b = \frac{1}{ac - b^2} \begin{bmatrix} ce - bf \\ af - be \end{bmatrix} \quad (13)$$

Therefore,

$$\bar{x}_1 = \frac{ce - bf}{ac - b^2}, \quad \bar{x}_2 = \frac{af - be}{ac - b^2} \quad (14)$$

Equation (14) is substituted into Eq. (5) to obtain the estimation values $\bar{v}$ and $\bar{l}_0$ of v and l_0

$$\bar{v} = \sqrt{\bar{x}_1 \bar{l}_0} = \frac{\bar{x}_2}{2\bar{v}} \quad (15)$$

According to Cauchy inequality,

$$ac - b^2 = \sum_i \Delta t_i^4 \cdot \sum_i \Delta t_i^2 - \sum_i \Delta t_i^3 \cdot \sum_i \Delta t_i^3 > 0 \quad (16)$$

is always $> \text{zero}$. However,

$$ce - bf = d^2 \sum_i \Delta t_i^2 \cdot \sum_i i^2 \Delta t_i^2 - d^2 \sum_i \Delta t_i^3 \cdot \sum_i i^2 \Delta t_i \quad (17)$$

is not always $> \text{zero}$. In other words, $\bar{x}_1$ is possible to be $< \text{zero}$, thus there is the possibility that $\bar{v}$ is a complex number. Such a situation often occurs when Δt_i has a high deviation with the true value.

2.2.1 TDOA estimation

Accurate TDOA estimation is the key to measure the distance between vibration source and fibre. The classical method for TDOA estimation is GCC. In recent years, GCCSC has been developed to conduct correlation again based on GCC. Under the low SNR, the results from the secondary relation are much better than GCC. We conducted the simulation comparison experiments by aiming at five methods, including cross-correlation (CC), SCOT, ML, GCCSC based on SCOT, and GCCSC based on ML. The findings of the comparative experiments indicate that the secondary relation estimation based on ML weighting has the maximum accuracy.

(1) The TDOA Model

A signal emitting from a remote source and monitored in the presence of noise at two spatially separated receivers can be mathematically modelled as

$$\begin{cases} x_1(n) = s(n) + n_1(n) \\ x_2(n) = As(n-D) + n_2(n) \end{cases} \quad (18)$$

where $s(n)$ is the vibration source signal, $x_1(n)$ and $x_2(n)$ are vibration signals received by two receivers; A represents the attenuation coefficient; D is the TDOA between $x_1(n)$ and $x_2(n)$; $n_1(n)$ and $n_2(n)$ refer to zero-mean additive Gaussian noises, assuming that there is uncorrelation between noises.

(2) GCC

GCC is the improvement on the basic cross-correlation method, and it can be used under the low SNR. GCC's principle is pre-filtering to eliminate noises and disturbance and then carrying out correlation operations.

According to the different weighting methods, GCC can be divided into four types, including Roth, SCOT, PHAT and ML. The ML algorithm can give the big weight to the frequency bands with the high SNR and the small weight to the bands with the small SNR, therefore it is the optimal filter statistically.

The generalised correlation between $x_1(n)$ and $x_2(n)$ is

$$R_{x_1x_2}(\tau) = \int_{-\infty}^{\infty} \psi_g(f) G_{x_1x_2}(f) e^{j2\pi f\tau} df \quad (19)$$

For ML algorithm, $\psi_g(f)$ is given

$$\psi_g(f) = \frac{1}{|G_{x_1x_2}(f)|} \frac{|\gamma_{12}(f)|^2}{[1 - |\gamma_{12}(f)|^2]} \quad (20)$$

where $|\gamma_{12}(f)|^2$ is the coherence function between $x_1(n)$ and $x_2(n)$.

(3) GCCSC

The process of GCCSC is that the signals are autocorrelated and cross-correlated firstly, and then the results of autocorrelation and cross-correlation are used to carry out cross-correlation operation again, so as to improve resolution power and anti-noise performance.

Autocorrelation function of $x_1(n)$ is

$$R_{x_1x_1}(\tau) = R_{ss}(\tau) + R_{n_1s}(\tau) + R_{sn_1}(\tau) + R_{n_1n_2}(\tau). \quad (21)$$

The CC function between $R_{x_1x_2}(\tau)$ and $R_{x_1x_1}(\tau)$ is

$$R_{RR}(\tau) = E[R_{x_1x_1}(n) R_{x_1x_2}(n-\tau)] \quad (22)$$

where $R_{x_1x_2}(n-\tau)$ is calculated by Eq. (19). Ignoring cross-correlation function between noises and signals, Eq. (22) is simplified as

$$R_{RR}(\tau) = R_{RN}(\tau) + R_{RS}(\tau-D). \quad (23)$$

$R_{RN}(\tau)$ is secondary relation for noises, and $R_{RS}(\tau-D)$ is secondary relation for source signals. Considering $R_{RN}(\tau) = 0$, there is maximum as τ in $R_{RR}(\tau)$ is D . D is the estimation value for TDOA.

Improvement

To improve the estimation accuracy, two aspects of the secondary relation are improved: one is limiting the range of delay peak; the other is using three-point interpolation to improve the delay estimate.

2.2.2 Eliminate outliers

The method of calculating Δt_i has been discussed above. However, in practice, affected by noises and interference, the estimated delay value Δt_i often sometimes exceeds the normal range. These points with apparent deviation are named as outliers. Too many outliers can affect the accuracy of estimation, and the removal of outliers can significantly improve the estimation accuracy.

(1) Elimination Indicator

Judging whether a data point is an outlier, it is necessary to select a suitable indicator. In this paper, the Deviation Square (DS) is used to judge whether Δt_i is an outlier. The definition is

$$DS_i = (i^2 d^2 - \Delta t_i^2 \bar{v}^2 - 2\Delta t_i \bar{v} \bar{l}_0)^2, \quad (24)$$

where $\bar{v}$ and $\bar{l}_0$ refer to the estimation values of wave velocity and distance calculated by Eq. (15). Figure 2 shows the geometrical intuitive significance of DS.

Fig. 2 (a) The geometric meaning of DS. (b) DS value of points in (a). In (a), The curve is $f(x) = d^2 x^2$. The solid points are the values of $i^2 d^2$. The hollow points are the values of $\Delta t_i^2 \bar{v}^2 + 2\Delta t_i \bar{v} \bar{l}_0$. The DS is the distance from the hollow point to the solid point with the same i . (b) shows the DS value in i . DS, deviation square.

According to Figure 2(a), when i is -1 , 1 , -6 and 6 , the four DS_i is relatively large. All DS_i in Figure 2(a) are calculated and shown in Figure 2(b). In Figure 2(b), the four DS_i marked with solid points are apparent outliers, and the rest 11 hollow points are more likely to be normal.

In Figure 2(b), the DS value of outliers is higher than normal points. The normal points gather in a specific region with smaller DS values, and the number is more. However, the outliers cluster into one or more classes in the region with larger DS values, and the number is less. Thus DS can be used to distinguish the normal points from outliers effectively.

(2) The Method of Eliminating Outliers

This paper adopts the idea of eliminating outliers by clustering. The primary thought is as follows: First, all points are clustered into multiple classes according to the DS value. Second, the most likely outliers are determined by judging the number of points in the class and the distance between the classes. Third, and after removal of such points. Fourth, the new DS value is calculated. The above steps are repeated until the remaining data can be considered as one class. That is, no apparent outliers.

There are multiple clustering methods, including K-Means, hierarchical clustering, Clustering with Gaussian Mixture Models (GMM) and DBSCAN. However, this paper requires that only one class remain in the end, no matter how many classes there are at the beginning. K-Means and GMM cannot meet the requirement of changeable clustering number, and hierarchical clustering is not easy to determine the dividing layers. As a result, DBSCAN is used in this paper.

DBSCAN has two advantages: (1) High efficiency. It can remove multiple outliers at one time. Compared with other clustering methods, iterations are fewer. (2) The clustering number is reduced with the increase of iterations, and finally into one class.

Before using DBSCAN, it is necessary to set two parameters, one is the cluster radius ε , the other is the minimum number of points (MinPts), sometimes named as cluster density. ε indicates the distance between data in one class. MinPts specifies the minimum number of data contained in one class. How to set the two parameters is described in the next experiment.

2.2.3 The complete method

Considering all the above problems, a complete method to calculate the distance between the vibration source and the optical fibre is described below.

Step 1. The datum fibre segment is determined, and the fibre segments around the datum segment are numbered by MMDFS.

Step 2. The TDOA of the marked fibre segments is estimated by GCCSC, meanwhile, the TDOA value which is negative or is deviated from the normal range is removed, then the initial TDOA vector $D=\{ \}$ is obtained.

Step 3. The estimation values are calculated by Eqs. (8–15).

Step 4. and are substituted into each equation in Eq. (4) to get DS_i . All DS_i form the set.

Step 5. According to the pre-set ε and MinPts, the set S is classified by DBSCAN. If the classification result is more than one class, it turns to step 6; if the classification result is one class, wave velocity and distance are outputted.

Step 6. The classes that have the higher value are considered outliers. The corresponding values in vectors D and S are eliminated. Then, it turns to step 4.

2.3 The experiment

The experiment is divided into two parts, including simulation fibre data and practical fibre data.

2.3.1 The Simulation Experiment

Ricker wavelets were used as the analogue vibration source signal. According to the attenuation coefficient and the distance between the optical fibre and vibration source, the attenuated signal was calculated. Then the attenuated signal was mixed with zero mean Gaussian noise to obtain the simulated vibration signal received by fibre.

2.3.1.1 Ricker wavelet

Ricker wavelet is the typical waveform in vibration simulation. The time-domain expression is shown as follows

$$s(t) = (1 - 2\pi^2 f_M^2 t^2) e^{-\pi^2 f_M^2 t^2} \quad (25)$$

where t is time, and f_M represents the peak frequency. Fourier transform is applied to Equation (25) to obtain the expression of Ricker wavelet in the frequency domain

$$S(f) = \frac{2f^2}{\sqrt{\pi} f_M^3} e^{-\frac{f^2}{f_M^3}} \quad (26)$$

The oscillogram of Ricker wavelet with the center frequency of 50 Hz is shown as follows

Fig. 3 The oscillogram of Ricker wavelet. $f_M = 50$ Hz. (a) The time-domain figure; (b) The frequency domain figure.

Figure 3 shows that the Ricker wavelet is not only similar to real vibration signals in the time domain but also in the frequency domain. Therefore, the Ricker wavelet is the desirable waveform of simulation vibration signals.

2.3.1.2 Hybrid data

The vibration source data was composed of five Ricker wavelets with a frequency peak of 50 Hz. To improve the data quality, Ricker wavelet amplitude was expanded five times. After attenuation, the vibration source data was mixed with Gaussian noise with zero mean and 0.05 variance. The sampling frequency was set as 10000 Hz, the duration was 3 s, the wave velocity was 10000 m/s, and the minimum distance between the optical fibre and vibration source was 1 m. There were 19 fibre segments receiving vibration signals, and their serial numbers based on MMDFS were $(-9, -8 \dots -1, 0, 1 \dots 8, 9)$.

At the same time, all data was stored as a matrix, with 30,000 rows and 19 columns. The stored data is shown in Figure 4.

Fig. 4 Simulation vibration signals made up of Ricker wavelet and noise. (a) All data is presented in a mesh. (b) Sampling points of No. 10 and No. 1 column of all data.

In Figure 4(b), each column refers to data received by one fibre segment. Among these columns, the 10th column is the datum fibre segment and has the best signal quality with an SNR of -1.05 db. But the first and 19th columns have the worst data quality with an SNR of -2.12 db. Five Ricker wavelets in signals of No. 10 column can be observed obviously, but the Ricker wavelets are extremely close to noises and challenging to distinguish in signals of No. 1 column. The simulation fibre data used in this paper is named as SimuFiberData.

2.3.1.3 TDOA of simulation signals

The experiment used the method of GCCSC based on ML to estimate TDOA. To improve accuracy, three-point interpolation was conducted. Table 1 shows the simulation fiber data for making a comparison on the real TDOA with five methods, including CC, SCOT, ML, GCCSC based on SCOT and GCCSC based on ML. It is worth noting that the unit of TDOA is the sampling number, instead of the second. Mean Square Error and success rate of five methods in Table 1 are shown in Table 2.

Table 1 Comparison of TDOA estimation for simu fibre data.

No.	CC	SCOT	ML	GCCSC based on SCOT	GCCSC based on ML	Real
-9	7	-2	7.3147	7.3945	7.9716	8.055
-8	8	8	8.2805	8.8228	7.4851	7.062
-7	5	5	4.9436	4.3144	8.02	6.071
-6	7	-7	6.9048	6.3663	5.464	5.083
-5	9	-6	6.0667	8.9716	5.7098	4.099
-4	3	-3	2.7389	2.8608	3.211	3.123
-3	5	-1	-1.1754	-1.2043	2.3093	2.162
-2	5	-10	5.2752	5.0414	2.1822	1.236
-1	1	1	0.7914	0.9828	0.8831	0.414
0	0	0	0	0	0	0
1	0	13	-0.016	0.1585	0.3109	0.414
2	2	3	2.8405	2.7737	1.5947	1.236
3	3	-14	3.6201	3.1296	2.7702	2.162
4	1	9	0.9783	1.2096	1.4064	3.123
5	4	-8	3.9291	4.1022	4.1073	4.099
6	8	14	7.8445	7.9643	6.8807	5.083
7	11	-13	11.2104	11.013	6.5797	6.071
8	3	3	11.0275	3.4893	8.4726	7.062
9	6	-10	6.4191	6.4771	7.1539	8.055

CC, cross-correlation.

Table 2 Mean square error and success rate of five methods of Table 1.

	CC	SCOT	ML	GCCSC based on SCOT	GCCSC based on ML
MSE	5.943	17.0042	5.2276	5.6606	0.9272
Success Rate	100%	42.1053%	89.4737%	94.7368%	100%

CC, cross-correlation.

TDOA estimation in Table 1 has two unreasonable situations: (1) TDOA value is negative. For example, the estimation value of No. -9 in SCOT is -2; (2) TDOA value dramatically exceeds the theoretical value. For example, the estimation value of No. 1 in SCOT is 13, but the actual value is 1.236. These belong to

unsuccessful estimation values. Thus the success rate is also an important indicator to estimate TDOA.

The success rate of the five methods in Table 2 shows that GCCSC based on ML has the highest success rate than other methods. The mean square error estimated by GCCSC based on ML is the minimum and it has the highest success rate. For this reason, GCCSC based on ML was selected to do TDOA estimation.

2.3.1.4 Distance estimation

To obtain satisfactory distance estimation results, not only the accurate TDOA method should be selected, but also the TDOA outliers should be removed reasonably. The entire detailed computational process of Simu-FiberData is shown in Table 3.

Table 3 Complete process to simu fibre data.

No.	True	TDOA	1 st		2 nd	
			i	DSi	i	DSi
1	8.055	7.9716	−9	248.021	0	0
2	7.062	7.4851	−8	28.806	0	0
3	6.071	8.02	−7	286.574	0	0
4	5.083	5.464	−6	1.424	−6	0
5	4.099	5.7098	−5	153.47	0	0
6	3.123	3.211	−4	0.612	−4	1.402
7	2.162	2.3093	−3	0.191	−3	0.03
8	1.236	2.1822	−2	22.245	0	0
9	0.414	0.8831	−1	2.88	−1	1.612
10	0	0	0	0	0	0
11	0.414	0.3109	1	0.032	1	1.28E−01
12	1.236	1.5947	2	2.861	2	1.199
13	2.162	2.7702	3	10.526	3	7.343
14	3.123	1.4064	4	124.818	0	0
15	4.099	4.1073	5	8.236	5	7.932
16	5.083	6.8807	6	221.649	0	0
17	6.071	6.5797	7	3.14	7	0.673
18	7.062	8.4726	8	70.983	0	0
19	8.055	7.1539	9	711.649	0	0
DSMSE				105.4509		2.2577
$\bar{v}$				8508.390		9370.197
$\bar{l}_0$				1.419		0.958

DS, deviation square.

The 1st column in Table 3 is the serial numbering for TDOA. Successively, the 2nd column is the true TDOA; the 3rd column is the TDOA estimation value according to GCCSC based on ML. After the 3rd column, data in two columns are divided into a group. Each group represents one distance estimation. No. 1 column of each group is the serial numbering based on MMDFS. If the serial numbering is 0, it means data at this point have been eliminated from the distance estimation. No. 2 column in each group is the corresponding DS of data involved in TDOA estimation. The antepenultimate row is the mean square error of all DS values after each distance estimation, implying the effect of DS clustering. The penultimate row is the estimation value of wave velocity. The tailender row is the estimation value of the distance between the vibration source and optical fibre.

In the first estimation, wave velocity is 8508.390, the distance is 1.419, both of which belong to the reasonable value range. However, DS has the more significant mean square error and it is 105.451, indicating that the

clustering result is not desirable. The TDOA value involved in estimation probably has outliers, which should be eliminated.

The primary step of eliminating outliers was to set up cluster radius ε and cluster density MinPts, both of which required pre-estimation. Moreover, the scatter diagram of DS can provide intuitional help for pre-estimation. As shown in Figure 5, the hollow points are normal points, but the solid points are suspected outliers. The distribution of the hollow points is relatively concentrated. The distance between two hollow points is about 10. The number of hollow points is about 10. The distribution of solid points is relatively scattered, and the distance between two solid points is >50 . There were six solid points, among which five are within the range of 150–300, and one point is far away from others, therefor the five points were more likely to gather into a cluster. Cluster radius ε could be set slightly larger than the distance between two normal points. Cluster density MinPts was determined by the number of outliers in a cluster. Cluster radius ε was set as 10 and cluster density MinPts was set as 5 in the first time removing outlier.

Fig. 5 The Scatter Diagram of No. 5 and 7 columns of Table 3.

After removing outliers in the first time, the second estimation was conducted to obtain wave velocity as 9370.197 and distance as 0.958, close to the set value 1; DS mean square error was very small, that is 2.2577, indicating that the clustering result was desirable and outliers had been eliminated. Figure 5(b) shows the scatter diagram of all DS values in the No. 7 column in Table 3—the scatter diagram after removing outliers, showing that the DS values of residual points were maintained at 10 and there were no outliers. The whole calculation was over.

2.3.2 The practical fibre data

2.3.2.1 Collection of fibre data

Fibre data were collected from the space in a college in Beijing, China. The space was an area of 800 m² in a rectangle. The south-north length was about 20 m and the south-east length was about 40 m. The length of fibre was 300 m, paved around the site with two circles. Its buried depth was about 10 cm under the ground.

The fibre was wireline cable. NBX-S3000 from the Japanese Neubrex Company was used as the fibre collecting device. The sampling frequency was 10000 Hz. Fibre spatial resolution was 1.03 m, thus the entire fibre could be divided into 290 segments.

The excavation was carried out at 108.83 m of the fibre. A shovel was used to excavate on the ground. The excavation position was distant from the fibre about 1 m. The strongest excavation signal was at 108.83 m.

Excavation signals could be distinguished at 99.59 m and 117.04 m. But it was difficult to distinguish farther excavation signals with the naked eyes. The vibration signals received from 99.59 m to 117.04 m were chosen as the analysis objects. According to MMDFS, the fibre segment at 108.83 m was the datum fibre segment with the serial numbering as 0. The serial numbering at 109.86 m, 117.04 m, 107.8 m and 99.59 m was 1, 8, -1 and -9 , respectively. There were a total of 18 fibre segments, which had the sequence numbered as (-9 , -8 , -7 , -6 , -5 , -4 , -3 , -2 , -1 , 0, 1, 2, 3, 4, 5, 6, 7, 8).

Practical fibre data were greatly affected by the background noises, and the SNR was low. The best SNR obtained from the strongest excavation signal at 108.83 m was 17.72 dB, the worst SNR was 9.49 dB. At 117.04 m, the best SNR was 8.45 dB; the worst SNR was -2.27 dB.

2.3.2.2 TDOA estimation and pre-treatment

Relative to the simulation fibre data, practical fibre data were affected by noises and soil medium, resulting in greater deviation. As estimating TDOA, more unreasonable values than simulation fibre data were generated, sometimes accounting for $>30\%$ of the total data. Therefore, the main task of pre-treatment was to remove unreasonable TDOA.

Fig. 6 The time domain signal at 108 m.

As shown in Figure 6, the fibre data contains four groups of apparent excavation signals circled with dotted lines. In distance estimation, four segments of data are partitioned. The length of each segment ranges from 130 to 200 sampling points. Excavation data in each group regards the No. 0 column as the benchmark to estimate TDOA with other columns and eliminate unreasonable TDOA, remaining 25 effective values.

Because the method adopted for estimating TDOA has strong noise resistance, this paper does not carry out de-noising processing on optical fibre data.

2.3.2.3 Distance estimation

The complete details of distance estimation for excavation data are recorded in Table 4.

Differing from Table 3, because true TDOA is unknown, there is no real value column in Table 4. The rest is the same as Table 4. The original number ranges from 1 to 72. After pre-treatment, only 25 valid TDOA values remained. No. 2 column is the TDOA estimation value. The entire calculation conducted distance estimation five times. Outliers were eliminated four times. The setting of cluster radius ε and cluster density MinPts each time were listed in the last two rows of Table 4. Finally, the distance was 0.951 m, close to 1 m. The experiment proved that the above-mentioned distance estimation method was effective.

Table 4 The overall process of excavation data treatment.

No.	TDOA	1 st		2 nd		3 rd		4 th		5 th	
		i	DSi	i	DSi	i	DSi	i	DSi	i	DSi
11	0.30	1	76.82	1	22.90	1	5.46	1	1.83	1	0.02
12	0.40	2	73.13	2	11.87	2	0.09	2	0.92	2	6.18
13	0.45	3	20.16	3	1.15	3	20.17	3	34.44	0	0
14	0.61	4	1.15	4	34.72	4	103.16	0	0	0	0
15	0.71	5	40.03	5	195.43	0	0	0	0	0	0
16	0.82	6	246.77	6	577.32	0	0	0	0	0	0
17	0.81	7	889.91	0	0	0	0	0	0	0	0
18	0.93	8	1902.98	0	0	0	0	0	0	0	0
20	0.95	-8	1878.44	0	0	0	0	0	0	0	0
21	1.94	-7	504.32	0	0	0	0	0	0	0	0
22	1.61	-6	63.24	-6	268.29	0	0	0	0	0	0
23	1.56	-5	13.22	-5	24.83	-5	101.21	0	0	0	0
24	1.05	-4	81.91	-4	0.01	-4	31.30	-4	62.44	0	0
25	0.28	-3	0.08	-3	16.31	-3	40.42	-3	53.09	0	0
26	0.32	-2	37.20	-2	3.69	-2	0.42	-2	2.83	-2	8.66
27	0.25	-1	52.07	-1	14.75	-1	3.11	-1	0.87	-1	0.01
47	0.53	1	228.79	1	77.38	1	24.15	1	11.17	1	2.14
48	0.81	2	322.49	2	94.42	2	22.01	2	6.90	2	0.06
49	1.36	3	388.30	3	110.21	3	24.98	3	6.87	3	0.04
50	1.81	4	172.13	4	34.59	4	3.03	4	0.06	4	3.18
51	2.55	5	20.72	5	9.16	5	2.03	5	2.11	5	0.13
65	0.63	1	303.47	1	106.37	1	35.55	1	17.58	1	4.40
66	1.13	2	520.48	0	0	0	0	0	0	0	0
67	1.45	3	407.44	3	124.57	3	33.62	3	11.92	3	1.26
68	1.94	4	157.81	4	40.46	4	8.47	4	1.33	4	0.00
DSMSE			336.12		88.42		27.01		14.29		2.17
$\bar{v}$		32894.05i		21501.91i		8321.26i		8756.23		16942.8	
$\bar{l}_0$		-5.50i		-4.884i		-6.976i		4.498		0.951	
ε		100		40		10		10			
MinPts		5		5		5		5			

DS, deviation square; MinPts, minimum number of points.

2.3.2.4 Possibility of failure

However, this method has the possibility of failure. Especially when there are many outliers with the DS values close to each other, they are easy to aggregate into a class, resulting in the deletion of the normal points. In this case, the result has a significant deviation.

3 Result and discussion

When this method is applied to the practical optical fibre data, it is affected by many factors and will produce significant errors. These factors include

3.1 Sampling

Sampling affects the measurement of the time when the vibration reaches the fibre. Sampling occurs at integer multiples of the sampling period. If the time of the vibration signal arriving at the optical fibre is not an integer multiple of the sampling period, the signal arrival time is regarded as the nearest next integer time. For example, the vibration that reaches the fibre is the 5.1 sampling point theoretically, but the actual time collected is in the 6th sampling point.

It is assumed that the arrival time of the vibration to the fibre is uniformly distributed, the average arrival time obtained will be 0.5 sampling points later than the actual arrival time during sampling.

The better method to solve the sampling problem is interpolation which can greatly reduce the sampling error.

3.2 Noises

Noises will affect the accuracy of TDOA estimation. With the decrease of SNR, the accuracy of TDOA estimation also decreases sharply. Moreover, when the SNR exceeds 10 dB, the accuracy of TDOA estimation obtained is also higher.

3.3 Spatial resolution

In this paper, a fibre segment is equivalent to a sensor. However, a sensor can be thought of as a point, and a fibre segment as a line segment whose length is equal to the spatial resolution. Therefore, the spatial resolution will cause a $\pm d/2$ error in the estimation of the closest point of the vibration source to the optical fibre. Further reduction of spatial resolution is the key to improve the accuracy of distance estimation.

3.4 Fibre pavement

Fibres in the experiment are paved in the earth ditch with a depth of 10–15 cm. In the pavement process, the fibre is not straight due to the factors such as the fluctuation of the pavement surface and the natural bending of the fibre.

3.5 Uneven soil layers

The propagation speed of the vibration signal in the soil layer is related to the propagation medium. If the elasticity modulus of the propagation medium increases, the velocity will be faster. The elasticity modulus of stone is greater than that of soil, and the elasticity modulus of soil is greater than that of sand. In the experimental site, when the earth ditch was excavated, it was found that there were more stones and harder soil. Therefore, when the vibration signal propagated in the soil layer, the wave velocity in different media was different. It even happened that the non-datum fibre segment received the signal earlier than the datum fibre segment.

4 Conclusions

Safety protection for underground communication fibre is of great significance. When a threatening event to optical fibre security happens, the effective distance measurement between the threat location and the optical fibre is an important basis for further protective actions. OTDR technology is used to make use of the existing communication fibre without adding additional sensors. This is a new way to protect the security of optical fibre by using the fibre itself as the sensor.

The method of distance measurement used in this paper includes four steps: fibre marking, TDOA estimation, outliers' removal, and distance calculation. It is proved that the method is correct and useful in theory.

However, affected by many factors, it is still not a very accurate distance estimation method in practice. To further improve the accuracy of the estimation, certain improvements, such as reducing spatial resolution and

increasing SNR, need further study.

References

- [1] I García. Optical Fiber Sensors for Aircraft Structural Health Monitoring[J]. *Sensors*, 2015, 15(7):15494-15519.
- [2] Kikuchi K. Fundamentals of Coherent Optical Fiber Communications[J]. *Journal of Lightwave Technology*, 2016, 34(1):157-179.
- [3] Ding Z, Du Y, Liu T, et al. Distributed Optical Fiber Current Sensor Based on Magnetostriction in OFDR[J]. *IEEE Photonics Technology Letters*, 2015, 27(19):2055-2058.
- [4] B António, Joan C, Sergi V. A Review of Distributed Optical Fiber Sensors for Civil Engineering Applications[J]. *Sensors*, 2016, 16(5):748.
- [5] Polynkin P, Temyanko V, Mansuripur M, et al. Efficient and scalable side pumping scheme for short high-power optical fiber lasers and amplifiers[J]. *IEEE Photonics Technology Letters*, 2015, 16(9):2024-2026.
- [6] Kerbage C, Windeler R S, Eggleton B J, et al. Tunable devices based on dynamic positioning of micro-fluids in micro-structured optical fiber[J]. *Optics Communications*, 2016, 204(1-6):179-184.
- [7] Li C, Xiao J, Guo T, et al. Interference characteristics in a Fabry–Perot cavity with graphene membrane for optical fiber pressure sensors[J]. *Microsystem Technologies*, 2015, 21(11):2297-2306.
- [8] Teng C X, Yu F D, Jing N, et al. The influence of temperature to a refractive index sensor based on a macro-bending tapered plastic optical fiber[J]. *Optical Fiber Technology*, 2016, 31(sep.):32-35.
- [9] Boaron A, Boso G, D Rusca, et al. Secure quantum key distribution over 421 km of optical fiber[J]. *Physical review letters*, 2018, 121(19):190502.1-190502.4.
- [10] Selvi M, Rajendran L. Application of modified wavelet and homotopy perturbation methods to nonlinear oscillation problems[J]. *Applied Mathematics and Nonlinear Sciences*, 2019, 4(2):351-364.
- [11] Feng J, Meng M, Liu S, et al. Prediction of Chinese Automobile Growing Trend Considering Vehicle Adaptability based on Cui–Lawson Model[J]. *Applied Mathematics and Nonlinear Sciences*, 2020, 5(2):367-376.

This page is intentionally left blank