An Adaptive Compressed Sensing Algorithm of Optical Fiber Pipeline Pre-warning Data

Hongjie Wan[1], Haojiang Deng[2], Xiaoming Xie [1], Qiaoning Yang [1] and Dan Su[1]

[1]*Information Engineering Dept, Beijing University of Chemical Technology, China*
[2]*National Network New Media Engineering Research Center, Institute of Acoustics, Chinese Academy of Sciences, China*
{wanhj, xmxie, yangqn}@mail.buct.edu.cn; denghj@dsp.ac.cn; sudanjean@163.com

Abstract

For distributed optical fiber pipeline pre-warning system, the sampling rate used is very high and thus huge data will be generated, which makes it difficult to transfer and store. Compressive sensing is a new compressed sampling method in the field of signal processing which compresses and samples the signal simultaneously. In this paper, an adaptive compressive sensing method is presented for compression and reconstruction of distributed optical fiber pipeline data. First, partial reconstruction based detection method is used to detect whether a hazardous event happened, then different compression ratios are taken for different classes of signal thereby increasing the compression ratio. In signal reconstruction phase, a sparsity determination algorithm is used to determine the sparsity of different segment of the signal, and then wavelet tree combined with CoSamp algorithm is adopted to reconstruct the signal. The adaptive compression algorithm improves the compression ratio and the sparsity determination in reconstruction phase can determine the sparsity of each segment when the signal varies without prior knowledge of the sparsity of the signal. Experimental results show that, the proposed algorithm can obtain higher reconstruction accuracy at a relatively high compression ratio. Furthermore, location simulation shows that the reconstructed signal by the proposed method is effective for danger signal positioning.

Keywords: *Data compression, adaptive Compressive sensing, optical fiber pipeline, Orthogonal Matching Pursuit (OMP), Tree based CoSamp (TCMP)*

1. Introduction

For the transportation of oil and natural gas, pipeline transport is a safe and convenient method of transportation. On the route of the pipeline transportation, due to natural disasters, human illegal mining as well as the aging of the pipe itself and other reasons, the pipeline transportation security is threatened. The distributed optical fiber pipeline pre-warning system detects the vibration signal around the pipeline using distributed optical fiber and gives advance warning of threat signal to guarantee the safety of the pipeline. However, the data sampling rate used in the method is much high which results in a flood of monitoring data [1]. This brings a lot of inconvenience for data transfer and storage. In order to reduce the amount of data in the pipeline monitoring process, there have been many articles and some progresses reported. For example, Jintao etc. detected the importance of the data blocks based on the difference and the dynamic range threshold, and then used a combination of wavelet transform and Huffman coding method for lossless compression of important data [2]. Xu Quansheng etc. proposed a method of pipeline leak detection data compression algorithm based on predictor and Rice coding method [3]. Tongyun proposed a segment adaptive data compression algorithm [4]. However, these methods are based on the Nyquist Sampling

Theorem. Under these methods, the signal is first sampled and then compressed. Unlikely, the compressed sensing based method compresses and samples the signal simultaneously using the sparsity of the signal under a transformation.

In compressed sensing, the data sampling rate can be substantially less than the Nyquist sampling rate for simultaneously sampling and compression. The Signal is measured by a matrix, and the measurement matrix is usually a random matrix. The most often used random matrices are random Gaussian matrix, sub-Gaussian matrix and Bernoulli matrix [5]. Also, there are some other matrix construction methods that is applicable to the compressed sensing [6]. The signal reconstruction methods are mainly divided into greedy iterative method and method based on dynamic programming. The reconstruction accuracy depends on the number of measurements and this will affect the signal compression ratio. For distributed optical fiber pipeline data, in the long running process, sometimes there exists threatening signal, sometimes not. The normal operation signal is mainly the noise data, and the data of this type does not require very high precision. Therefore, this article first detects the observed data to determine whether a threatening signal exists, if there exists, more measurements will be taken, otherwise fewer measurements. This method compresses the signal adaptively. In the signal reconstruction stage, the tree-based recovery algorithm [7] is used for its higher accuracy to piece wise and smooth signal. In the recovery algorithm, tree based algorithm and CoSamp [8] algorithm are combined to reconstruct the signal. But this method requires pre-given signal sparsity which is more difficult for the actual signal. And different signals in different time periods will have different sparse degrees, so an adaptive method is adopted to determine the sparsity of the signal before signal recovery. The above method presents adaptive processing in the compression and reconstruction aspects thus ensuring high reconstruction accuracy under higher compression ratio.

The structure of the paper is organized as follows. The second section describes the principle of compressed sampling and reconstruction. The third part describes the tree based recovery algorithm and the proposed adaptive compressed sensing algorithm. In the fourth part, positioning method of the danger signal is given. And in the fifth part, the proposed algorithm is carried out on the data collected at the scene of the optical fiber pre-warning system and the experimental results are given. The final part concludes this paper.

2. Theory of Compressed Sensing

Compressed sensing is a new signal sampling and compression method which is based on signal sparse feature under a certain transformation. Suppose x is a segment of signal with length N, and $\mathbf{\Psi} = \{\psi_1, \psi_2, \cdots, \psi_N\}$ is an orthogonal basis, then the expansion of x under the orthogonal basis is

$$x = \Psi\theta = \sum_{n=1}^{N} \psi_n \theta_n \tag{1}$$

Where $\theta = \{\theta_1, \theta_2, \cdots, \theta_N\}$ are the expansion coefficients under the orthogonal matrix $\mathbf{\Psi}$. If only K ($K \ll N$) values in θ are more important than others and the unimportant values are near zero, then x is said to be K sparse and $\mathbf{\Psi}$ can be called as the sparse matrix of x. Generally, the sparse nature of the signal under the transform matrix should be known in advance.

On the basis of sparse feature of the signal, the procedure of the compressed sensing method can be displayed as in Figure 1. As can be seen from the Figure, the signal processing procedure of compressed sensing can be divided into two stages including compressed observation and signal reconstruction.

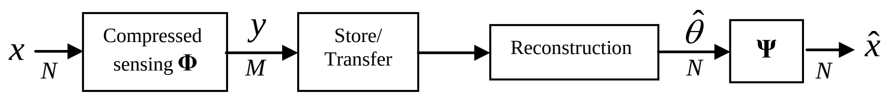


Figure 1. The Compression and Reconstruction Procedure of Compressed Sensing

In compressed observation stage, an observation matrix with the size $M \times N$ ($M << N$) is taken to observe the signal. This observation process can be described as

$$y = \mathbf{\Phi} x = \mathbf{\Phi \Psi} \theta = \mathbf{A} \theta \tag{2}$$

Where **A** can be called as the CS matrix.

It is clear that the observed signal y is an M dimensional vector and thus the signal is compressed during observation process. When the observation matrix $\mathbf{\Phi}$ and transformation matrix $\mathbf{\Psi}$ are strongly incoherent, the original signal x can be recovered with a high probability from y if $M > K \log N$ [9]. When the entries of the observation matrix are i.i.d. Gaussian distribution, Bernoulli distribution or sub-Gaussian distribution variables, the above conditions can be easily achieved.

In Signal reconstruction phase, the reconstruction problem is to solve x from (2). Based on the sparse features of x under the transformation matrix $\mathbf{\Psi}$, the solving method using linear programming (LP) can be formulated as

$$\hat{\theta} = \arg\min_{\theta} \|\theta\|_1, \textit{subject to } y = \mathbf{A}\theta \tag{3}$$

Where $\|\cdot\|_1$ is the ℓ_1 norm. When noise is added during observing process or the sparsity level used is not equal to the actual sparsity level, (3) can be rewritten as

$$\hat{\theta} = \arg\min_{\theta} \|\theta\|_1, \textit{subject to } \|y - \mathbf{A}\theta\|_2 \le \varepsilon \tag{4}$$

For (3) and (4), the recovered signal can be $\hat{x} = \mathbf{\Psi}\hat{\theta}$. A variety of algorithms, such as BP, BPDN, LASSO, etc. based on linear programming method are used to solve the problem shown in (4). Furthermore, the greedy iterative based approach such as matching pursuit [10], orthogonal matching pursuit [11], iterative hard threshold [12] and compressive sampling matching pursuit (CoSamp) [8] may also be used for recovering the signal. The greedy iterative algorithm is fast than the LP based algorithm and the CoSamp method can achieve similar performance as LP based algorithm [13].

3. Adaptive Compressed Sensing for Pipeline Data

3.1. The proposed Adaptive Compression and Reconstruction Process

In optical fiber pipeline pre-warning system, long-running signal is divided into signal having threatening information, and signal of normal operation. For these two types of signals, the required reconstruction accuracy is also different. Since the signal duration time is very long, so the signal compressed sampling is conducted by segment.

Different segments have different sparse degrees and it is very important to determine the sparsity signal of each segment because it can affect the reconstruction accuracy. Based on the above two points, this paper presents an adaptive compression and reconstruction algorithm, the process is shown in Figure 2.

In the compression stage, this paper uses the OMP partial reconstruction method to detect whether there is threatening signal, then takes a different number of measurements for the different categories of data (For example, *M*1 and *M*2 measurements are taken for the normal operation and the threatening segments respectively), thus higher compression ratio is obtained with guaranteed accuracy. In the reconstruction phase, based on the piecewise smooth characteristics of the signal, the CoSamp algorithm and CSSA [14] algorithm is combined for recovery using wavelet tree decomposition. Because the sparsity of each segment of signal is different, the sparsity K is determined firstly using the sparsity determination algorithm, then the CoSamp algorithm and CSSA algorithm is combined for recovery to achieve higher recovery accuracy.

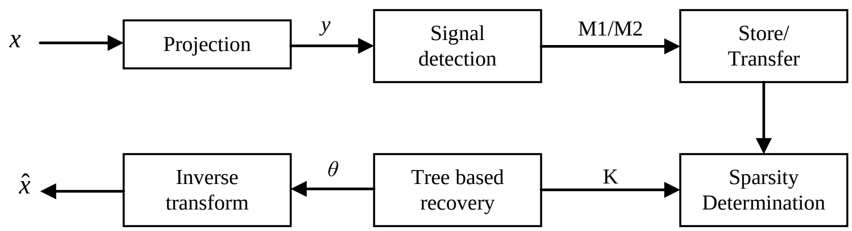


Figure 2. The Process of the proposed Adaptive Compression and Reconstruction Algorithm

3.2. Signal Analysis by Wavelet Tree

The object of signal analysis is to obtain the characteristics in time domain or frequency domain. Fourier analysis is widely used in various fields, but it can not be used to analyze the signal in the frequency domain and time domain at the same time. The wavelet transform overcome the shortcomings of the Fourier analysis, it has the ability to characterize the local signal characteristics in both time and frequency domain. The basis functions of Fourier transform are sine and cosine functions of various frequency components. Similarly, the wavelet transform take the dilation and translation of the mother wavelet as the basis for signal decomposition. Continuous wavelet transform is defined as follows

$$W(a,\tau) = \frac{1}{\sqrt{a}} \int x(t)\psi^*(\frac{t-\tau}{a})dt \tag{5}$$

The greater the similarity between the wavelet function and signal the greater the wavelet coefficients, so wavelet function with great signal similarity to the signal should be chosen for decomposition. In addition, the high similarity also makes relatively sparse decomposition coefficients, which makes it more suitable for compressed sensing.

For discrete signals, the wavelet transform has a tree structure [15]. Assuming that the signal x is a vector with length $N = 2^L$, L is a positive integer, the scaling function ϕ and the wavelet function ψ are as follows

$$\phi_{l,j}(t) = 2^{l/2}\phi(2^l t - j), \psi_{l,j}(t) = 2^{l/2}(2^l t - j) \tag{6}$$

Then the wavelet transform can be written as

$$x = \alpha\phi + \sum_{l=0}^{L-1}\sum_{j=0}^{2^l-1} c_{l,j}\psi_{l,j} \tag{7}$$

Where $\alpha = \langle x, \phi_{0,0} \rangle$ and $c_{l,j} = \langle x, \psi_{l,j} \rangle$ .In above transform, $c_{0,0}$ is the wavelet coefficients at the coarsest scale and α is the corresponding single scaling coefficient. The above transform can be written as matrix form $x = \boldsymbol{\Psi}\theta$, where $\boldsymbol{\Psi}$ is the matrix composed by the scaling function and wavelet function, $\theta = \left[\alpha, c_{0,0}, c_{1,0}, c_{1,1}, \cdots, c_{L-1,2^l-1}\right]^T$ is a vector composed by the scaling and wavelet coefficients. The wavelet functions have nesting structure which means the support of $\psi_{l,j}$ includes the support of $\psi_{l+1,2j}$ and $\psi_{l+1,2j+1}$. Therefore, the corresponding wavelet coefficients have a parent-child relationship and they can form a binary tree relationship which can be shown in Figure 3.

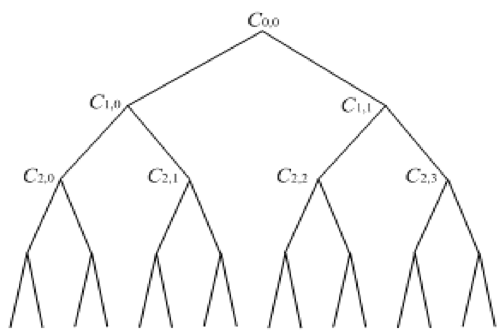


Figure 3. Tree Structure of One Dimensional Wavelet Coefficients

For smooth or piece wise smooth signal, Wavelet coefficients with large magnitude represent the signal discontinuous region, and small wavelet coefficients represent the smooth region. Because the nested relationship of the support, a sub tree with relatively large coefficients will be generated in the wavelet tree. Further, the wavelet coefficients gradually reduce while the scale increases. Based on these characteristics, in the time of signal approximation, large coefficients on the sub-tree are used to approximate the signal transform coefficients and other coefficients are set to zero. While for reconstruction in compressed sensing, the signal can be recovered accurately by finding the sub-tree contains large coefficients.

3.3. Signal Detection using OMP

For measured signal by compressed sensing, the reconstruction quality is related to the number of measured values. The more measured values, the higher the reconstruction accuracy will be. During optical fiber pre-warning pipeline monitoring process, most of the signal is the normal operation signal which will be noise signal. The reconstruction accuracy need not be very high for this type of the signal, so fewer measurements will do and for signal containing threatening event more measurements are required. Therefore, this paper firstly classifies the signal using compressed sensing detection methods and then takes a different number of measurement signals for different categories.

The compressed signal detection or classification is based on the following hypothesis:

H_0： $x = s + n$ ； H_1： $x = n$

Where s indicates the pure vibration signal when threatening event happens and n is assumed to be Gaussian noise. For the optical fiber pre-warning data, the segments containing threatening event will have sparse coefficients under the transformation matrix $\mathbf{\Psi}$. In contrast, the normal operation segments will be noise and the transformed coefficients will be much little. As a result of this, the key coefficient will be different between the two classes of signal. In fact, the different can be formulated as $\|\theta_s\|_\infty > \|\theta_n\|_\infty$, where $\|\theta_s\|_\infty$ and $\|\theta_n\|_\infty$ denote the key coefficient of the threatening segment and noise segment respectively. Based on these considerations, the hypothesis testing problem can be reformulated to check if the maximum reconstruction coefficient exceeds a certain threshold λ, that is $\|\theta\|_\infty > \lambda$. Using this method, the threatening event can be detected.

Due to the above inference, the matching pursuit algorithm was taken for signal detection in [16], where partial reconstruction replaced the completely reconstruction thereby reducing computation complexity. OMP algorithm is an improved signal recovery algorithm based on matching pursuit. In OMP, the best projection on the selected support domain is solved each iteration. So it has excellent features including fast convergence, easy to implement, and better performance compared with MP [11]. So, this paper takes the OMP algorithm for signal detection.

Let t be the iteration control variable and r_t be the residual at t-th iteration, then given the CS matrix $\mathbf{A}$ and the measurement vector y, the OMP based detection method is as follows

1) Initialize: $\hat{\theta}_0 = 0$, $r_0 = y$, $\mathbf{\Omega_0} = \varnothing$, $t = 1$;
2) Form signal estimate from residual, $c_t \leftarrow \mathbf{A}^T r_{t-1}$;
3) Update support using the largest entry of the signal estimate, $\mathbf{\Omega}_t \leftarrow \mathbf{\Omega}_{t-1} \bigcup \operatorname{supp}\left(\mathrm{T}_1(c_t)\right)$;
4) Update signal estimate, $\hat{\theta}_t\big|_{\mathbf{\Omega}_t} \leftarrow \mathbf{A}^+_{\mathbf{\Omega}_t} y$, $\hat{\theta}_t\big|_{\mathbf{\Omega}^c_t} \leftarrow 0$;
5) Update measurement residual, $r_t \leftarrow y - \mathbf{A}\hat{\theta}_t$;
6) $t = t + 1$;
7) Repeat 2) to 6) until halting criterion is met.

If $\|\hat{\theta}_t\|_\infty > \lambda$, then there is threatening event in the signal, otherwise there is not. The parameter λ is the threshold for deciding whether there is threatening event in the measurement. In the above algorithm, $\mathbf{\Omega}_t \subset \{1, 2, \cdots, N\}$ is the indices set of the t-th iteration; Λ^C_t is the complementary set of Λ_t in the set $\{1, 2, \cdots, N\}$; $\theta|_{\mathbf{\Omega}}$ denotes the values of θ restricted by the indices set $\mathbf{\Omega}$; $\mathbf{A}^+_{\mathbf{\Omega}_t} = \left(\mathbf{A}^T_{\mathbf{\Omega}_t}\mathbf{A}_{\mathbf{\Omega}_t}\right)^{-1}\mathbf{A}^T_{\mathbf{\Omega}_t}$ is the pseudo-inverse of $\mathbf{A}_{\mathbf{\Omega}_t}$ .In addition, $\mathrm{T}_1(c_t)$ is a function which set all entries of c_t to be zero except the entry with largest magnitude.

3.4. Tree based Recovery

There are many algorithms for compressed sensing signal recovery. As the wavelet transform has a tree structure, and piecewise smooth and smooth signal has sub tree characteristics on the wavelet tree, therefore, the compressed signal recovery algorithm just needs to find the sub-tree which can best approximate the original signal transform then the signal can be reconstructed. In fact, the problem can be formulated as the following optimization problem

$$\tilde{\Gamma} = \arg\max_{\|\Gamma\|_0 = k} \sum_{i \in \Gamma} |c_i|^2 \tag{8}$$

Where Γ denotes the wavelet tree set of size k, i represents the order number of wavelet coefficients which is got by sorting the combinations of scale and position (l, j). The number of wavelet tree with size k is much less than the number of k dimensional combinations in the N-dimensional space, so this recovery algorithm can greatly reduce the search space.

The condensing sort and select algorithm (CSSA) is taken to select the K largest coefficients as the sparse estimate. The CSSA algorithm can find the K largest coefficients, but K must be given in advance. While signal sparsity K also need to be given in advance for CoSamp and SP, this paper adopts the combination of CoSamp and CSSA for signal recovery [15]. The algorithm is as follows

Input: CS matrix **A**, measurement vector y, sparsity K

1) Initialize: $\hat{\theta}_0 = 0$, $r_0 = y$, $t = 1$;
2) Compute the estimation of the signal from the residual of last iteration, $c_t \leftarrow \mathbf{A}^T r_{t-1}$;
3) Prune signal support by CSSA algorithm, $\mathbf{\Omega}_1 \leftarrow \mathrm{supp}(\mathrm{S}(c_t, K))$, where S denotes the CSSA algorithm;
4) Merge support, $\mathbf{\Omega}_2 \leftarrow \mathbf{\Omega}_1 \cup \mathrm{supp}(\hat{\theta}_{t-1})$;
5) Form new signal estimate using the updated support in step 4, $b|_{\mathbf{\Omega}_2} \leftarrow \mathbf{A}^+_{\mathbf{\Omega}_2} y, b|_{\mathbf{\Omega}_2^c} \leftarrow 0$;
6) Prune signal estimate by CSSA algorithm, $\hat{\theta}_t \leftarrow \mathrm{S}(b, K)$;
7) Update residual, $r_t \leftarrow y - \mathbf{A}\hat{\theta}_t$;
8) $t = t + 1$
9) Repeat 2) to 8) until halting criterion is met

Output: $\hat{\theta} \leftarrow \hat{\theta}_t$

The signal can be estimated by $\hat{x} = \mathbf{\Psi}\hat{\theta}$ when $\hat{\theta}$ is got. In the above algorithm, each iteration of CoSamp require operation of order $O(MN)$ complexity, where M is the measurement number and N is the length of original signal. The first step of CSSA is to sort all the wavelet coefficients, so its complexity is $O(NlogN)$. So the combination of CSSA with CoSamp will increase the computation complexity while increasing the recovery accuracy. In addition, there is a shortcoming that the sparsity of the signal should be known in advance which is hard to implement.

3.5. Segment Sparsity Determination

Using the algorithm described in Section 3.3, the relatively more precise estimate of the original signal can be obtained. While the OMP algorithm needs not a prior knowledge of the sparsity of the original signal, the CoSamp needs a prior knowledge [8]. Therefore, the algorithm that combines CoSamp and CSSA will need in advance the sparsity of the signal. For the long time running signal, the signal is recovered each segment and the sparsity will vary for different segment. If the different sparsity value other than the real sparsity value is taken for the recovery algorithm, there will cause some error. So the sparsity of each piece of signal needs to be pre-determined. In this paper, an iterative method is used to estimate the sparsity K of the signal [13].This algorithm starts from a small initial sparsity value and the sparsity is gradually increased until certain condition is met. In essence, it searches the supports according to the importance degree gradually and merges the support with the support of the last iteration, and finally updates the support in the current iteration. The algorithm is as follows

Input: CS matrix $\mathbf{A}$, measurement vector y, sparsity step κ

1) Initialize: $\hat{\theta}_0 = 0$, $r_0 = y$, $F_{\mathbf{0}} = \varnothing$, $K = \kappa$, $t = 1$, $j = 1$;
2) Form signal estimate from residual, $c_t \leftarrow \mathbf{A}^T r_{t-1}$;
3) Make support candidate, $\mathbf{\Omega}_1 \leftarrow F_{t-1} \bigcup \mathrm{supp}\left(T_K(c_t)\right)$;
4) Refine support, $\mathbf{\Omega}_2 \leftarrow T_K(\mathbf{A}^{+}_{\mathbf{\Omega}_1} \mathrm{y})$;
5) Update residual, $r \leftarrow y - \mathbf{A}_{\mathbf{\Omega}_2} \mathbf{A}^{+}_{\mathbf{\Omega}_2} y$;
6) If stopping criterion is met, stop the iteration, or go to 7);
7) If $\|r\|_2 \geq \|r_{t-1}\|_2$, then go to 8), or go to 9);
8) Update support size, $j = j + 1$, $K = j \times \kappa$;
9) Update the parameters for next iteration, $F_t = \mathbf{\Omega}_2$, $r_t = r$, $t = t + 1$;
10) Repeat 2) to 9).

Output: sparsity K

In the above algorithm, $T_K(x)$ sets all entries except the K largest entries of x to be zero. The variable t is the iteration number counter; j controls the sparsity increase and κ controls the increase step. The candidate support and the refined support are changing in the whole algorithm, while the update process is similar to that of SP algorithm and CoSamp algorithm in certain iteration.

There are mainly the following ways to halt the iteration: 1) When the norm of the residual $\|r\|_2$ is small than a threshold value ε , but it is hard to determine the threshold because the norm of the measurement will affect it; 2) When the relative change of the recovered signal is small than a threshold ξ , though better than the first method, it is also hard to determine the threshold for this method in some cases. For the easier to implement characteristics of 2), this paper adopts this as the halting condition setting.

The sparsity step parameter κ will also affect the sparsity estimation. The larger step can make the algorithm converge fast but may bring large error. In contrast, little step can make the algorithm converge slowly but may bring little error. To get high accuracy, this paper use little step.

4. Positioning of the Threatening Signal

For distributed pipeline pre-warning system, when found threatening events, the threat incident location should be positioned. During positioning process, the key is to find the time difference Δt of the threat signal passing by the first end and the terminal photo detectors. Then the location of the threatening event is $x = \frac{L - v\Delta t}{2}$. L is the length of the fiber and v is the propagation velocity of the light in the fiber. The commonly used method to get the time difference is the correlation method, which is to get the maximum correlation position of the photoelectric signals at the two ends of the sensor [1]. Assuming that $x(t)$ and $y(t)$ is the photoelectric signal detected at the first end and the terminal respectively, then the correlation of $x(t)$ and $y(t)$ is

$$R(\tau) = \int_{-\infty}^{+\infty} x(t)y(t+\tau)dt \tag{9}$$

The corresponding τ_0 when the correlation function reaches the maximum value is the expected time difference Δt. After the analog signal converted to digital signals, the photoelectric signals of the first end and the terminal can be respectively expressed as $x(n)$ and $y(n)$ (n is integer), then the correlation function of N points can be expressed as

$$R(J) = \frac{1}{N}\sum_{n=0}^{N-1} x(n)y(n+J) \tag{10}$$

Assuming that $J = J_0$ corresponds to the max value of the correlation function R, then the time difference is $\Delta t = J_0 T$, where T is the sampling interval for the digital signal.

Compressed sensing algorithm can compress the original signal by sampling at a very low speed than the Nyquist frequency through random projection of the original signal. However, the quality of the reconstructed signal will vary with the recovery algorithm and the measurement number. According to the positioning principle described above, for compressed sensing, the quality of the recovered signal will depend on whether it will affect the positioning accuracy. Therefore, the efficiency of the compressed sensing algorithm can be evaluated by comparing the positioning result of the original signal and the recovered signal.

5. Experimental Results

In this paper, on-site shovel digging is carried out to simulate the real mining stolen behavior and the simulation analysis is carried out on the data collected. The buried optic fiber cable is used on-site which is buried 0.5 meter depth and 500 meters long. Digging with a spade in the buried fiber optic cable segment is executed. Such a mining behavior can simulate the real mining potential illegal behavior. The signal generated by the mining behavior is sampled at 4MHz sampling rate. Due to the high sampling rate, large amount of data was generated. So the original signal is down-sampled to 44.1 KHz and only the segment containing the digging event and a segment of normal operation signal is used in this paper. The data used in this paper is shown in Figure 4.

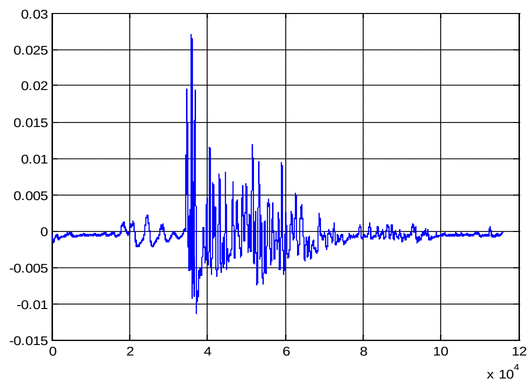


Figure 4. The Mining Signal used in the Paper

The wavelet analysis of the signal is based on the expansion under the waveform basis. Therefore, if the waveform of the signal and the waveform of the basis are more similar, the analysis effect will be better and the transform coefficients will be sparser. Figure 3 showed that the signal used in this paper is piecewise smooth signal, so it will be better to use Daubechies wavelet and this will also be in line with the conditions of the formation of the wavelet tree.

In order to measure the compressed sampled signal's ability to preserve information of the original signal, root mean square error (PRD), signal-to-noise ratio and energy recovery coefficients can be taken as evaluation metrics. To measure the compression capability of the compressed sampled signal, compression rate(R) is taken as evaluation metric [17]. This paper uses the signal-to-noise ratio and compression ratios to evaluate the performance, the calculation method of these two metrics are as follows

$$SNR = 10\log 10 \sum_{n=1}^{N} x^2(n) \Big/ \sum_{n=1}^{N} [x(n) - \tilde{x}(n)]^2 \tag{11}$$

$$R = \frac{\text{size of compressed signal}}{\text{size of original signal}} \tag{12}$$

In (11), x is the original signal and $\tilde{x}$ is the reconstructed signal from the compressed sampled signal.

The first experiment is focused on signal detection. In order to classify the noise and the signal containing threats, OMP based partial reconstruction algorithm is used to estimate the maximum recovered value of the noise segments. The measurement number M varies from 10 to 90 with step 5 and the result is show in Figure 5. The segment length of the original signal is taken as 1024 and 512. The values obtained can be used to estimate the threshold to classify the signal. This article takes 0.1 as the threshold to detect the presence of the threatening signal.

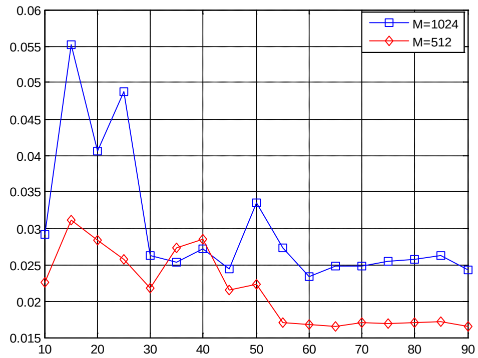


Figure 5. Maximum Recovered Values under Different Measurement Numbers for Noise Signal when the Original Signal Length is 512 or 1024

For the second experiment, as in the adaptive reconstruction process, it is very important to determine the sparsity and the sparsity step is more important for the accuracy of the sparsity determination algorithm and the accuracy of the sparsity will affect the reconstruction quality. Therefore, the second experiment is to evaluate the reconstruction performance under different sparsity steps of the sparsity determination algorithm. For comparison, the original signal length is taken as 512 and 1024, and the corresponding measurement number is 50 and 70 respectively. The result is shown in Figure 6 and the horizontal axis indicates the sparsity step, the vertical axis represents the SNR of the reconstructed signal. As can be seen from the figure, in the case of different sparsity steps, the SNR is different but there is slightly difference when the sparsity step is small. When the original signal length is 1024 and the sparsity step is 10, the SNR is low; when the original signal length is 512 and the sparsity step exceeds 6, the SNR begins to decrease. The low SNR under big sparsity is because the estimated sparsity will be more likely to deviate from the true sparsity when the true sparsity is between the last update and the next update of the sparsity determination algorithm.

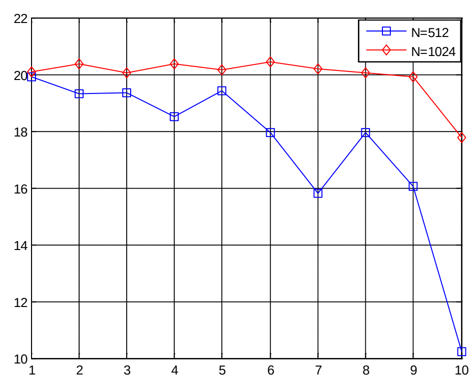


Figure 6. SNR of Adaptive Reconstruction using Sparsity Determination Algorithm as a Function of Sparsity Step under Different Signal Length *N*

Based on these results, a relatively small sparsity step is adopted for sparsity determination in this paper. Also, we take the original signal length as 1024.

In the third experiment, OMP, CoSamp and the proposed algorithm with wavelet tree and sparsity determination are compared on the reconstruction performance. The proposed algorithm can be divided into two versions. The version taking different number of measurements for normal operation signal and threatening signal in the compressed sampling phase can be called as adaptive compression (AC) version. Corre-

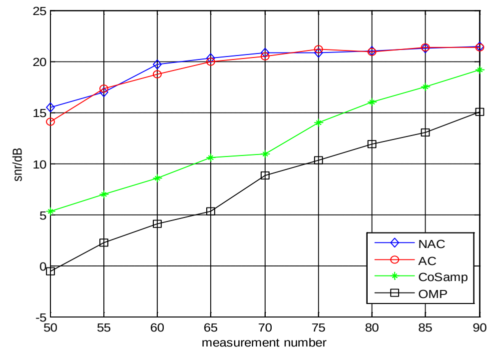


Figure 7. SNR Comparison of OMP, CoSamp, AC Version and NAC Version as the Measurement Number Varies from 50 to 90

spondingly, the version taking the same number of measurements in the compressed sampling phase can be called as non adaptive compression (NAC) version. The original signal length is taken as 1024 in the performance comparison. The measurement number varies from 50 to 90 with a step 5. During the sparsity determination process, the sparsity step is taken as 2. The result is shown in Figure 7. As can be seen from the figure, in the case of non-adaptive compression, the proposed algorithm achieves the highest reconstruction SNR. Under the adaptive compression case, the measurement number of the noise segment is taken 20 less than the non-noise section. The figure showed that there is little difference between the reconstruction SNR of the adaptive and non-adaptive compression version. Also, the proposed algorithm achieves much better performance than the OMP and CoSamp algorithm.

Table 1. Performance Comparison of AC and NAC

Measurement number	NAC			AC		
	CR	**SNR(dB)**	Δt /s	**CR**	**SNR(dB)**	Δt /s
50	0.049	15.48	4.93e-6	0.036	14.08	1.38e-5
55	0.054	16.99	3.94e-6	0.041	17.32	5.92e-6
60	0.059	19.73	1.97e-6	0.046	18.69	3.94e-6
65	0.063	20.27	0.99e-6	0.051	20.00	1.97e-6
70	0.068	20.80	0.99e-6	0.055	20.50	0.99e-6
75	0.073	20.85	0	0.060	21.20	0.99e-6
80	0.078	21.05	0	0.065	20.93	0
85	0.083	21.20	0	0.070	21.33	0
90	0.088	21.44	0	0.075	21.37	0

The fourth experiment is to evaluate the positioning accuracy of the reconstructed signal using the proposed algorithm. The reconstructed signal of the compressed sampled signal using compressed sensing may be not completely consistent with the original. For optical fiber pre-warning system, waveform inconsistency may lead to inconsistent positioning deviation. Therefore, the positioning difference between the reconstructed signal and the original signal is shown in Table 1. Because the signal used is down sampled signal, the positioning differences are expressed in time. The results show that the positioning errors caused by reconstruction signal are very small. When the measurement number exceeds a certain number there is no positioning error. This proves that the proposed algorithm can be applied to the optical fiber pre-warning system. Also, the CR and SNR results at different measurement numbers are listed in Table 1. From the table, it can be seen that when CR is higher than 0.05 the SNR can exceed 20dB for both AC and NAC version.

6. Conclusions

In this paper, an adaptive compression and reconstruction method based on compressed sensing for optical fiber pipeline data is proposed. The method uses OMP based detection method to detect whether a threatening event during the compression stage. In the reconstruction phase, the signal sparsity is firstly determined by the sparsity determination algorithm. Then, because the wavelet tree is suitable to model the smooth or piecewise smooth signal, it is combined with CoSamp algorithm for signal reconstruction. Therefore, the proposed method is adaptive in both the compression and the reconstruction phases. The adaptive compression method can decrease the compression rate and guarantee a high reconstruction SNR; the sparsity adaptive method in the reconstruction stage helps to reconstruct the signal without priory knowledge of the signal sparsity. Experimental results show that the proposed algorithm can obtain a lower compression rate and higher reconstruction SNR. And, from the positioning point of view, simulation results prove the effectiveness of the method for optical fiber pipeline data.

Acknowledgement

This research is supported by the Fundamental Research Funds for the Central Universities of China (NO.ZZ1226).

References

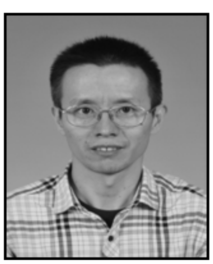
[1] Z. Yan, “Study on the Distributed Optical Fiber Pipeline Safe Detection Technology”, Phd thesis, Tianjin university, **(2006)**.
[2] J. Tao, Q. Peiwen and L. Liang, “Research on MFL Inspection Large Capacity and Hi-Fi Data Compression”, Journal of basic science and engineering, vol. 13, no. 2, **(2005)**, pp. 155-161.
[3] X. Quansheng and L. Sen, “Lossless compression technique of pipeline leakage signal”, Journal of Shenyang University of Technology, vol. 29, no. 6, **(2007)**, pp. 677-681.
[4] T. Yun, H. Songling and Z. Wei, “Oil and gas pipeline EMAT data compression algorithm”, Journal of Tsinghua University (Science & Technology), vol. 29, no. 2, **(2010)**, pp. 1619-1622.
[5] Y. C. Eldar, “Compressed Sensing: Theory and Applications”, Cambridge University Press, **(2012)**.
[6] T. T. Do, L. Gan, N. H. Nguyen and T. D. Tran, “Fast and Efficient Compressive Sensing Using Structurally Random Matrices”, IEEE Transactions on signal processing, vol. 60, no. 1, **(2012)**, pp. 139-154.
[7] M. F. Duarte, M. B. Wakin and R. G. Baraniuk, “Fast reconstruction of piecewise smooth signals from random projections”, Proc. SPARS05, (Rennes, France), **(2005)** November.
[8] D. Needell and J. A. Tropp, “Cosamp: Iterative signal recovery from incomplete and inaccurate samples”, Preprint, **(2008)** March.
[9] E. J. Candes and M. B. Wakin, “An introduction to compressive sampling”, IEEE Signal Processing Magazine, **(2008)**, pp. 21-30.
[10] S. G. Mallat and Z. F. Zhang, “Matching pursuits with time-frequency dictionaries”, IEEE Transactions on Signal Processing, vol. 41, no. 12, **(1993)**, pp. 3397-3415.
[11] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit”, IEEE Transactions on Information Theory, vol. 53, no, 12, **(2007)**, pp. 4655-4666.

[12] T. Blumensath and M. E. Davies, “Iterative hard thresholding for compressed sensing”, Preprint, **(2008)** July.
[13] T. T. Doy, L. Ganz, N. Nguyeny and T. D. Tran, “Sparsity adaptive matching pursuit algorithm for practical compressed sensing”, 42nd Asilomar Conference on Signals, Systems and Computers, **(2008)**, pp. 581-587.
[14] R. G. Baraniuk and D. L. Jones, “A signal-dependent time-frequency representation: Fast algorithm for optimal kernel design”, IEEE Transaction on Signal Processing, vol. 42, **(1994)** January, pp. 134-146.
[15] R. G. Baraniuk, V. Cevher, M. F. Duarte and C. Hegde, “Model-Based Compressive Sensing”, IEEE Transactions on Information Theory, vol. 56, no. 4, pp. 1982-2001, **(2010)**.
[16] M. Duarte, M. Davenport, M. Wakin and R. Baraniuk, “Sparse Signal Detection from Incoherent Projections”, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), **(2006)**, pp. 305-308.
[17] T. Tongkui, W. Hao and S. Huihe, “Process Data Compression Technology: A Survey”, Computers and Applied Chemsitry, vol. 17, no. 3, **(2000)**, pp. 193-197.

Authors

Hongjie Wan he received the B.S. degree in electrical and communication engineering from Xi'an Jiao Tong University in 2000, and the Ph.D. degree in signal processing from the Institute of Acoustics, Chinese Academy of Sciences, in 2005. He is currently working in the Information Engineering Department, College of Information Science and Technology, Beijing University of Chemical Technology. His research interests include statistical signal processing, audio signal processing, pipeline leak detection, pattern recognition and compressed sensing.

Haojiang Deng he received his B.S. degree from Wuhan University in 1993, received his M.S. degree from the 510 Institute of China Academy of Space Technology in 1998, and received Ph.d degree from Institute of Semiconductors of Chinese Academy of Sciences (CAS). He was engaged in the post-doctorate research in Institute of Acoustics (IOA) of CAS till 2003, and began to work in IOA of CAS since 2005. Then he was appointed as a professor in 2007. The main research field of professor Deng covers digital signal processing in audio and video, broadband multimedia communication.



Xiaoming Xie he received his Ph.D. degree in Information and Communication Engineering from Huazhong University of Science and Technology in 1998. From 1998 to 2004, he was a Senior Engineer in Golden Cellular Communications Co., Ltd. From 2004 to now he is an associate professor in the Information Engineering Department, College of Information Science and Technology, Beijing University of Chemical Technology. His research interests include Special mobile communication technology, embedded information processing and Wireless transmission technology.