http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (14): 2851-2856, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.2851.2856

Optical Fiber Pipeline Data Compression based on Segment Sequential Compressed Sensing

[1]HongJie Wan, [2]XueGang Chen, [1]WeiGuo Lin and [3]HaoJiang Deng
[1]College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China
[2]Department of Computer science, Xiangnan University, China
[3]National Network New Media Engineering Research Center, Institute of Acoustics,
Chinese Academy of Sciences, Beijing, China

Abstract: In distributed optical fiber pipeline pre-warning system, the sampling rate is very high for threatening event location, so vast data will be generated. Huge amount of data is inconvenient for transfer or storage. Because compressive sensing is a widely used methods for sampling and compressing data in the same time in recent years, this study adopts the compressive sensing approach to reduce the data quantity. In compressive sensing, the sparsity of each segment is important for signal recovery and it controls the measurement number needed for certain recovery accuracy of the recovered signal. The sparsity should be known in advance to determine the measurement number, but it is difficult to achieve. This is specially exemplified in optical fiber pipeline compressive sensing as the optical fiber pipeline data is longtime running data and the sparsity of every segment varies with time. In this study, the sequential approach joint with linear prediction is used to fix the measurement number of each segment. This approach further reduces the amount of data on the basis of compressive sensing. Simulation is carried out on the actual optical fiber pipeline pre-warning data and the experimental results show that the reconstruction SNR could exceed 26 dB.

Key words: Segment compressive sensing, optical fiber pre-warning, pipeline safety, sequential compressed sensing

INTRODUCTION

The distributed optical fiber pipeline pre-warning system is presented by Yan *et al.* (2005) and has been studied and used thereafter. It is stated in his PhD thesis that the sampling rate of the optical fiber pipeline pre-warning system should be very high to guarantee the location accuracy of the dangerous event (Yan, 2006). The system is a real time and long time running system, but the high Nyquist sampling frequency results huge amount of data to be transfer and storage which is inconvenient for the system. For the system of this type, there have been many research focused on the compression of the data achieved. However, all these researches are based on the Nyquist sampling theorem and few researches has been developed for the compression of the optical fiber pipeline pre-warning data.

The compressive sensing is a recently develop-ped technique and has been widely used in many areas including image compression, pattern recognition, medical imaging, wireless communi-cation, astronomy, bio sensing information theory and other related areas. In compressive sensing, the sampling process and the compression process is carried out in the same time. Therefore, compared with the compression methods based on the Nyquist sampling theory, it is not necessary to sample and store the vast amount of data firstly and then compress in this framework. However, it should be noted that the compressive sensing approach is based on the characteristic of the Nyquist sampled data. Above all, for the advantage of the compressive sensing, it is adopted to compress the optical fiber pipeline pre-warning data in this study. As mentioned early in this study, the optical fiber pipeline pre-warning system is a long time running system and the sparsity of every segment varies with time. The sparsity determines the measurement number need for every segment as the recovery accuracy is controlled by the measurement number under many cases. However, to know the sparsity of each segment in advance is very difficult. For this reason, a sequential approach is taken in this study to avoid this problem. This approach receives the measurement samples sequential and stops receiving when a certain criteria is met. With this approach, the sparsity should not be given in advance thus the recovery accuracy can be guaranteed. In detail, Orthogonal Matching Pursuit (OMP)

Corresponding Author: XueGang Chen, Department of Computer science, Xiangnan University, China Tel: 130-1612-8080

(Tropp and Gilbert, 2007) based sequential method is firstly used to estimate the sparsity of the signal, then the measurement number for each segment of the long time signal is slightly changed. The linear prediction method is utilized to predict the measurement number. In this way, it is not necessary to start from one observed value for each segment, thus reducing the computational complexity.

The study is organized as follows. In the second section, the principle of compressive sensing is briefly introduced. Then, the sequential receiving approach is introduced and the segment compressive sensing using sequential methods for optical fiber pre-warning data is presented. Simulation is carried out on the real sampled signal and results are given to support the method presented. At last, the study is concluded in the fifth section.

COMPRESSED SENSING

In compressive sensing framework, the signal is sampled and compressed in the same time. In contrast, for the compression method based on the Nyquist sampling theorem, the data should be sampled first and then compressed. It seems that there is no relevance between the compressive sensing and Nyquist sampling theorem. In fact, the compressive sensing framework is in some sense related with the Nyquist sampling theorem. Supposing there is a segment of signal with length N and it is denoted as x. Then, as is known to all that, under a certain set of orthogonal basis $\{\psi_i\}_{i=1}^N$, the signal x can be expanded as:

$$x = \sum_{i=1}^{N} \theta_i \psi_i = ?\ \theta \tag{1}$$

where, $\Psi = [\psi_1, \psi_2, \cdots \psi_N]$ is a transform matrix contains the set of orthogonal basis and θ is the transform coefficient. In signal processing, the coefficient is usually sparse. That means most of the element in θ is zero or approaches zero, only a small portion of the element is important. The number K of the large elements in θ is called the sparsity of the segment under the transform matrix Ψ. In most cases, K is far less than N.

Under the descriptions given above, the signal is projected by a random matrix Φ which is called as the measurement matrix. The projection process can be displayed as follows:

$$y = F\,x = F\ ?\ \theta = A\theta \tag{2}$$

The size of the matrix Φ is M*N and M<< N. The much smaller number M accomplishes the compression of the original signal x. The commonly used random matrix Φ in compressive sensing is the random Gaussian matrix, as the signal projected by this type of random matrix can be recovered accurately under some condition (Eldar, 2012). To satisfy this condition, the measurement number should be the order of:

$$M = O(K\log(N/K)) \tag{3}$$

Random matrixes including Bernoulli matrix and sub-Gaussian matrix can also be used for compressive sampling.

The compressive sampled signal should recover the origin signal accurately. There have been many algorithms developed for recovery of the compressive sampled signal. The recovery problem can be formulated as:

$$\hat{\theta} = \arg\min_{\theta} \|\theta\|_1, \text{subject to } y = A\theta \tag{4}$$

Where:

$$\|\theta\|_1 = \sum_{i=1}^{N} |\theta_i|$$

is 1-norm. This type of formulation is referred to as the l_1 minimization approach (Tsaig and Donoho, 2006). The other basic recovery algorithmis the greedy iterative algorithm, including matching pursuit (MP) (Mallat and Zhang, 1993), Orthogonal Matching Pursuit (OMP)(Tropp and Gilbert, 2007) , iterative hard threshold (Blumensath and Davies, 2009), CoSamp (Needell and Tropp, 2010) and other methods. After getting the coefficient $\hat{c}$, the original signal can be approximated by the orthogonal transformation matrix, that is $\hat{x} = \Psi\hat{c}$.

SEGMENT BASED SEQUENTIAL COMPRESSED SENSING

Sequential compressed sensing: Now it is clear that the high recovery quality requires enough measurements at sampling stage and with Gaussian random matrix, the measurement number should exceeds M = O(Klog(N/K)). It can be seen that this quantity is strongly related with the sparsity K of the original signal. The stopping criteria can avoid the sparsity problem and decide when to stopping receiving measurement (Malioutov *et al*., 2010). Assuming there have M measurements already been received, then According to (2), it can be written as:

$$y_i = a_i'\theta_T, i = 1, \cdots, M_1 \tag{5}$$

where, α_i is a random vector whose elements are independent and identical Gaussian random variables with zero mean and variance 1 and θ_T is the true transform coefficient of the original signal under the transform matrix. After receiving the M_1 measurements, the recovery algorithm is performed on these measurements and a value θ_{M_1} is achieved. Then two classes of signal are considered on when to stop receiving the measurement. The first type is the simplest, that is the exact sparse signal. In regard to this type, the stopping criteria is $\theta_{M_1+1} = \theta_{M_1}$. With this criteria met, the recovered coefficient is the true one and M_1 measurements are enough. Another type is the compressible signal which has many elements close to zero but not zero. With regard to this type, the stopping criteria is a bit complex. First, let $\triangle\theta = \theta_T - \theta_{M_1}$ be the difference between the true solution and the solution got by the M_1 measurements. Then take another L measurements $y_i = a_i'\theta_T, 1 \le i \le L$. After this, the following deviation can be given:

$$\tilde{y}_i = y_i - a_i'\theta_{M_1} = a_i'\triangle\theta, 1 \le i \le L \tag{6}$$

It can be deduced that $\tilde{y}_i$ is independent and identical Gaussian variable with zero mean and $\|\triangle\theta\|^2$ variance as the random characteristic of α_i. Based on the central limit theorem, the squared summation of the deviations is $\chi^2(L)$ distributed. Let:

$$\tilde{Y} = \sum_{i=1}^{L} \tilde{y}_i^2$$

then $p(\tilde{Y} / \|\triangle\theta\|^2 > \tilde{Y}_\alpha) = \alpha$ can give us a confidence level of the deviation. With $\tilde{Y} / \|\triangle\theta\|^2 > \tilde{Y}_\alpha$, the deviation can be $\|\triangle\theta\|^2 < \tilde{Y} / \tilde{Y}_\alpha$. Because $E(\tilde{Y}) = L\|\triangle\theta\|^2$,the deviation is $\|\triangle\theta\|^2 = E(\tilde{Y}) / L$. In practice, $\tilde{Y} / L$ is used to estimate the deviation and when the deviation is small enough the stopping criteria is met.

Segment based sequential compressed sensing: In the procedure of sequential compressed sensing, determination of the number of signal measurements is started from a very small value. The actual signal often lasts for a very long time and need to process by segment, the sparsity of the adjacent segment signal usually will not change greatly. Therefore starting from a very small number of measurements each segment to determine if the number of measurements is sufficient is not necessary. In this regard, we first determine a measurement number for the first segment signal. Then, the linear prediction method is used to predict the number of measurements for the next segment. Thirdly, sequential search on the basis of this number is processed by a small-scale. This can reduce computation time and retain the reservation accuracy of the compressed signal to the original signal. In this study, orthogonal matching pursuit is used to the first segment to judge the signal sparsity k. Then a starting measurement number l is calculated by Eq. 3. From this, measurements are taken by sequential compressed sensing discrimination method for the first segment. In fact, the first segment starts with a larger value l. Then determine the sparsity by the convergence situation in OMP algorithm. The OMP algorithm is as follows:

- **Step 1:** Initialize: $r_0 = y, \Lambda_0 = \varnothing, t = 1$
- **Step 2:** Calculate the inner product between all columns α_i of A and the residual vector $r_{t-1}, < a_i^T, r_{t-1} >, 1 \le i \le N$. Obtain the index of α_i corresponding to the maximum inner product:

$$I_t = \underset{i}{\operatorname{argmax}} < a_i^T, r_{t-1} >$$

- **Step 3:** $\Lambda_t \leftarrow \Lambda_{t-1} \cup \{I_t\}$ {update support by residual}
- **Step 4:** $\hat{c}_t|_{\Lambda_t} \leftarrow A_{\Lambda_t}^+ y$, $\hat{c}_t|_{\Lambda_t^c} \leftarrow 0$ {update signal estimate}
- **Step 5:** $r_t \leftarrow y - A\hat{c}_t$ {update measurement residual}
- **Step 6:** if $\|r_t\| \le \varepsilon$, then stop; if not, repeat step 2 to 5

In the above algorithm, t is the iteration number control variable; r_t is the residual of t-th iteration; Λ_t denotes entries set in the t-th iteration, $\Lambda_t \subset \{1 \ldots N\}$; Λ_t^C is complementary set of Λ_t :ε is the a threshold for stopping the iteration; finally:

$$A_{\Lambda_t}^+ = \left(A_{\Lambda_t}^T A_{\Lambda_t}\right)^{-1} A_{\Lambda_t}^T y$$

Optical fiber pre-warning signal is a long duration signal in real life, it needs to be processed by time segment. The change rate of the signal is generally not fast, so that sparse degree of the adjacent segments does not vary greatly. In this study, the method of linear prediction is used to have a forecast of the measurements number for the current segment. Let j represents the segment number of the signal, L(j) denotes the measurements number of the j-th segment, then:

$$M(j) = \sum_{i=1}^{p} a_i M(j-i) \tag{7}$$

In Eq. 7, p is the prediction order and $a_i (1 \le i \le p)$ is prediction coefficient. Then run sequential compressed sensing from T_0 measurements before this predicted number until the stopping criteria is met.

EXPERIMENT RESULTS

To evaluate the method presented above, real data acquired is used. The optical fiber is buried under the ground 0.5 m in depth, then digging near the optical fiber

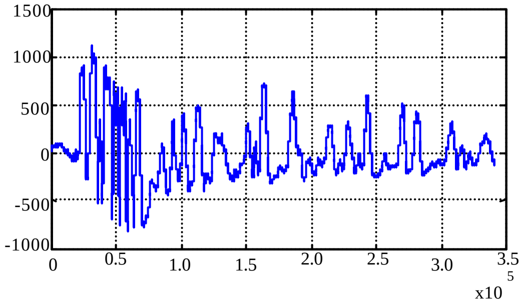

Fig. 1: Shovel digging signal

was operated to simulate the stolen behavior. The sampling rate is 2 MHz which is too high and there will be vast amount of data. For this reason, we only use a fragment of the signal to demonstrate the method and the signal is down sampled to 11.025 KHz firstly. The signal used is as shown in Fig. 1.

It is stated early in section II that the signal is sparse under some transform matrix, here the haar transform is taken to transform the original signal. For simplicity, only the haar transform is studied in this study. Detail wavelet transform orthogonal matrix design can be found in (Ge and Wei, 2007). To assess the recovery quality, SNR is used and it is as follows:

$$\mathrm{SNR} = 10\log10\frac{\sum_{n=1}^{N} x^2(n)}{\sum_{n=1}^{N}[x(n)-\hat{x}(n)]^2} \tag{8}$$

where, x and $\hat{x}$ denote the original and recovered signal, respectively.

For comparison, firstly, signal compression sampling is carried out without using segment based sequential compressed sensing. Each segment of the original signal is 1024 points, measurement numbers from 50 to 200 with a step of 10 are taken for compressing the signal. Then recovery of the signal using OMP algorithm is carried on each measurement number. To give a typical comparison, the recovered signal with 70 and 90 measurements is plotted in Fig. 2. The figure showed that when the measurement number is 70, because there is not enough measurements for some segment, the recovered signal was distorted compared to the original signal which is explicit in the 64th segment in Fig. 2b. In Fig. 2c, the measurement number is 90 and the recovered signal is more smooth and more close to the original signal in Fig. 2a, though it is still not very perfect.

In fact, the SNR of each segment is not the same. All the SNR data is shown by a histogram to display their differences as in Fig. 3. It can be seen that the SNR of each segment is different so the measurement number required for each segment will not be the same. And from Fig. 2 one can draw this conclusion, further more, it is difficult and important to judge the measurement number for each segment. Meanwhile, just from Fig. 2, the fiber optic warning data can be compressed sampled at a compressing ratio about 11:1.

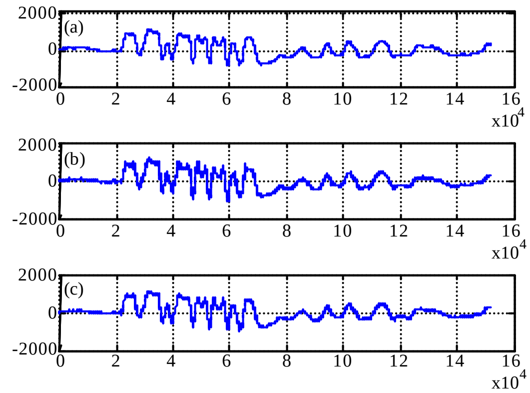

Fig. 2(a-c): (a) The original used (b) the recovered signal when measurement number of each segment equals 70 and (c) The recovered signal when measurement number of each segment equals 90

Table 1: Performance with different t

Evaluation type	Evaluation result				
T	5	10	15	20	25
SNR(dB)	24.96	24.37	22.8	23.99	21.76
L	92	90	82	86	74.9

Next, the segment based CS algorithm had been run on the test signal. Firstly, each segment is running from a very small number, the number is set to be L = 20. The algorithm has been run with T varies from 5 to 25 with the step 5 and the mean SNR and mean measurement number of all segments are displayed in Table 1. Finally the signal is compressed sampled with the best measurement number and the SNR is got for further comparison.

To reduce searching cost of compressed sensing, the linear prediction of the measurement number is taken. However, get the linear prediction coefficients need further and more computation which is undesirable. So we simply use just two measurement number before the current segment to predict the current measurement number. And the previous one is weighted by a constant $\beta(0<\beta<1)$, another one is weighted by $1-\beta$. Several values of β are taken to predict the measurement number, the final mean measurement number of the signal and the SNR value are shown in Table 2. From Table 2, one can see that

Table 2: Performance with different prediction

Evaluation type	Evaluation result				
ß 0.98	0.8	0.85	0.95	0.9	
SNR(dB)	26.95	26.9	26.6	27	26.99
L 116.2	114	114.4	115.8	115	

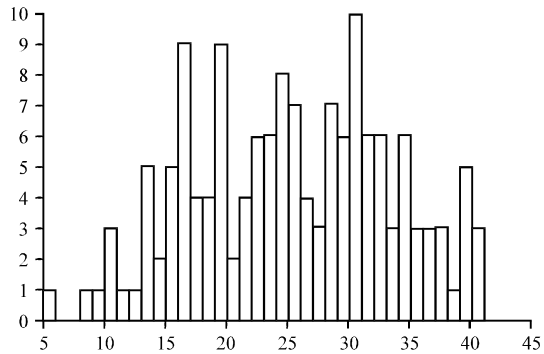

Fig. 3: Histogram composed by SNR of all segments

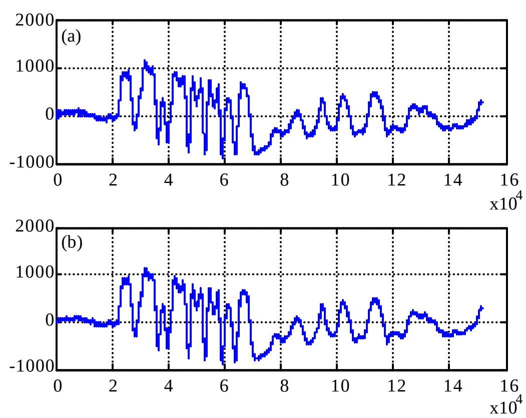

Fig. 4: The figure on the top is recovered using the sequential algorithm. The figure on the bottom is recovered using the segment based sequential algorithm

using more less searching and recovery computation, the compressed sampled signal can be stably recovered.

The recovery signal with the best measurement number and the signal recovered by predicted measurement number is plotted in Fig. 3. The figure show that the proposed algorithm is suitable for long time signal.

CONCLUSIONS

This study presents a segment based sequential compressed sensing algorithm for optical fiber signal compression and reconstruction. In the framework of traditional signal processing, a high speed sampling is required to meet the Nyquist sampling theorem. This will generate vast data. This study use compressed sampling methods which is based on signal sparsity under some kind of transformation. For optical fiber pipeline data, we use haar transformation. But the sparsity of the signal can not know in advance, so sequential compressed sensing is used. The optical fiber signal is long time signal which should be processed by segment. It is not necessary to start from a very small measurement number for each segment in sequential compressed sensing. From a starting measurement number of the first segment, linear prediction is used to predict the measurement number of current segment. Experimental results of using the proposed segment based sequential compressed sensing algorithm showed that the algorithm need not know the signal sparsity in advance and can have the same reconstruction accuracy as sequential compression with less searching.

ACKNOWLEDGEMENT

This research is funded by the Fundamental Research Funds for the Central Universities of China (No. ZZ1226) and in part by the hunan province science and technology plan project of China (No. 2013FJ3032)

REFERENCES

Tsaig, Y. and D.L. Donoho, 2006. Extensions of compressed sensing. Signal Process., 86: 549-571.

Malioutov, D., S. Sanghavi and A. Willsky, 2010. Sequential compressed sensing. IEEE J. Selected Top. Signal Proc., 4: 435-444.

Needell, D. and J.A. Tropp, 2010. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. Commun. ACM, 53: 93-100.

Ge, Z. and S. Wei, 2007. Wavelet Analysis Theory with MATLAB R2007 Implementation. Publishing House of Electronics Industry, China.

Tropp, J.A. and A.C. Gilbert, 2007. Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans. Inform. Theory, 53: 4655-4666.

Mallat, S.G. and Z. Zhang, 1993. Matching pursuits with time-frequency dictionaries. IEEE Trans. Signal Process., 41: 3397-3415.

Blumensath, T. and M.E. Davies, 2009. Iterative hard thresholding for compressed sensing. J. Applied Comput. Harmon. Anal., 27: 265-274.

Eldar, Y.C., 2012. Compressed Sensing: Theory and Applications. Cambridge University Press, China.

Yan, Z., S. Jin and Y. Zhang, 2005. Study on the distributed optical fiber sensing technology for pipeline leakage detection. J. Optoelectronics Laser, 16: 935-938.

Yan, Z., 2006. Study on the distributed optical fiber pipeline safe detection technology. Ph.D. Thesis, Tianjin University, China.