Research Article

Real-Time Data Recovery in Wireless Sensor Networks Using Spatiotemporal Correlation Based on Sparse Representation

Jingfei He and Yatong Zhou

The School of Electronics and Information Engineering and Key Laboratory of Electronic Materials and Devices of Tianjin, Hebei University of Technology, Tianjin 300401, China

Correspondence should be addressed to Jingfei He; hejingfei@hebut.edu.cn

Received 24 February 2019; Revised 15 April 2019; Accepted 8 May 2019; Published 20 May 2019

Academic Editor: Antonio Guerrieri

Copyright © 2019 Jingfei He and Yatong Zhou. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to data loss and sparse sampling methods utilized in WSNs to reduce energy consumption, reconstructing the raw sensed data from partial data is an indispensable operation. In this paper, a real-time data recovery method is proposed using the spatiotemporal correlation among WSN data. Specifically, by introducing the historical data, joint low-rank constraint and temporal stability are utilized to further exploit the data spatiotemporal correlation. Furthermore, an algorithm based on the alternating direction method of multipliers is described to solve the resultant optimization problem efficiently. The simulation results show that the proposed method outperforms the state-of-the-art methods for different types of signal in the network.

1. Introduction

Wireless sensor networks (WSNs) are developed to sense various environmental information and widely used in numerous applications including military surveillance, environmental monitoring, and health care monitoring. However, due to hardware, wireless transmission, and energy constraint in WSNs, only partial raw sensed data can be gathered in the sink. As a result of the existence of data missing, data recovery in WSNs achieves more attention in recent years.

Affected by the hardware and wireless transmission in WSNs, the network faces the missing of raw sensed data due to data loss. Furthermore, to reduce the energy consumption in WSNs, methods based on compressed sensing (CS) [1] have been proposed. Since the dense sampling CS based methods [2, 3] using dense sensing matrix require the participation of all sensor nodes and still have a large amount of data to be transmitted, sparse sampling CS based methods using sparse sensing matrix were proposed. Among them, to reduce the amount of transmitted data, there are several sparse sampling methods using a small number of sensor nodes to participate in data collection, which also yields data missing in the network. These data missing problems caused by data loss or

utilizing sparse sampling method lead to the requirement of data recovery in WSNs.

Actually, the data in WSNs generally have strong spatiotemporal correlation due to dense distribution of nodes and the slow change of environmental information, which makes it possible to recover the whole data from partial samples. A number of data reconstruction methods were proposed for data missing. In the case of only a little data missing, the traditional data interpolation methods such as K-Nearest Neighbors (KNN) [4] and Multichannel Singular Spectrum Analysis (MSSA) [5] can be used. Recently, compressed sensing theory, as a powerful signal processing technology, has been widely utilized to recover the missing data by enforcing the sparsity constraints. By designing a sparsest measurement matrix with only one nonzero entry in each row, method [6] based on CS was proposed to recover the missing data in WSNs. ScoRel [7] framework was proposed using joint CS and principal component analysis for effective data reconstruction. Besides, as the second-order sparse representation of CS, matrix completion (MC) [8] has emerged recently. Several methods [9–13] based on MC were proposed by enforcing the low-rank constraint on the WSNs data to take advantage of the spatiotemporal correlation. Specifically, an efficient data collection approach

(EDCA) [9] was proposed using a scalable power-saving sampling model to reduce energy consumption. EDCA uses the matrix factorization method for nuclear norm minimization to recover the missing data. To further improve the data recovery accuracy, Spatiotemporal Compressive Data Collection (STCDG) [10] was proposed taking advantage of the low-rank feature and short-term stability feature in sensor data. Considering the data loss in WSNs, an environmental space time improved compressive sensing (ESTI-CS) [11] algorithm with a multiattribute assistant component was proposed. The low-rank structure, spatial similarity, temporal stability, and multiattribute correlation are exploited in ESTI-CS. Afterwards, the method utilizing both the low-rank and temporal sparsity feature was also proposed [13]. These existing approaches have achieved impressive results on data reconstruction.

In the above methods of data reconstruction for data loss or data missing caused by sparse sampling, an environment matrix (EM) recording data from n sensors over t time slots is considered to represent the dynamic environment. The data reconstruction algorithms in the above methods recover the EM from partial collected data. However, data from t time slots are required to construct the EM, which means these low-rank based methods cannot be used in the real-time requirement WSNs. For the real-time data recovery for data missing in WSNs, a few works were proposed so far. In [14], a DCT-Regularized Partial Matrix Completion (DCT-RPMC) algorithm was proposed by introducing historical data to the original data matrix. Both low-rank and band-limited features of sensor data are utilized in DCT-RPMC. Besides, a sequential CS method with sliding window processing (Seq-Prog-CS) method [15] was proposed to reconstruct spatially and temporally correlated sensor data streams via Kronecker sparsifying base. The methods [16] using low-rank feature and sparsity feature were also proposed under the compressive data gathering scheme. The work in [16] recovers the sensor data from measurements obtained by dense sensing matrix, which is not designed for the data missing problem.

To further improve the data recovery accuracy in the real-time requirement WSNs, in this paper, we propose a real-time data recovery method to further exploit the data spatiotemporal correlation. The historical data is introduced to enforce joint low-rank constraint and temporal stability constraint. Besides, we also present an algorithm based on the alternating direction method of multipliers (ADMM) [17] to solve the proposed method.

The rest of this paper is organized as follows: Section 2 introduces the basic concepts of matrix completion. Section 3 presents the low-rank feature and temporal stability of raw sensed data. Section 4 describes the problem formulation and the proposed real-time data recovery method. Section 5 shows some representative simulation results, which is followed by the conclusion of the paper in Section 6.

2. Basics on Matrix Completion in WSNs

Matrix completion is the process of recovering a matrix from a partial sampling of its entries, which is theoretically

impossible without any additional information of the matrix. However, the matrix we wish to recover in many practical problems of interest usually has low-rank or approximately low-rank property. Therefore, the unknown matrix $\mathbf{M} \in \mathbb{R}^{n \times m}$ can be recovered by solving the optimization problem:

$$\begin{aligned} & \text{minimize} \quad \text{rank}(\mathbf{X}) \\ & \text{subject to} \quad \mathbf{X}_{ij} = \mathbf{M}_{ij} \quad (i, j) \in \Pi, \end{aligned} \quad (1)$$

where $\mathbf{X}$ is the decision variable and Π is a subset of the complete set of entries $[n] \times [m]$ (here and in the sequel, $[n]$ denotes the list $\{1, \dots, n\}$). The $\text{rank}(\mathbf{X})$ represents the rank of the matrix $\mathbf{X}$, and $\mathbf{M}_{ij}$ is the sampled set of entries in $\mathbf{M}$. However, solving this optimization problem is often impractical because it is NP-hard. Considering the rank minimization problem simply counts the number of nonzero singular values like the ℓ_0 minimization problem in compressed sensing, one can use an alternative which minimizes the nuclear norm for the matrix [18].

$$\begin{aligned} & \text{minimize} \quad \|\mathbf{X}\|_* \\ & \text{subject to} \quad \mathbf{X}_{ij} = \mathbf{M}_{ij} \quad (i, j) \in \Pi. \end{aligned} \quad (2)$$

The nuclear norm for the matrix completion problem is equal to the sum of the singular values of a matrix. It is the best convex surrogate to the rank minimization, which is the analogue to the convex ℓ_1 norm surrogating the ℓ_0 norm in the CS theory.

In WSNs with n sensors sensing the environmental information every τ time, during $t \times \tau$ time, total $n \times t$ sensor data should be gathered in the sink. These data can be organized into an environment matrix $\mathbf{X}$

$$\mathbf{X} = \begin{bmatrix} f(N_1, T_1) & f(N_1, T_2) & \cdots & f(N_1, T_t) \\ f(N_2, T_1) & f(N_2, T_2) & \cdots & f(N_2, T_t) \\ \vdots & \vdots & \ddots & \vdots \\ f(N_n, T_1) & f(N_n, T_2) & \cdots & f(N_n, T_t) \end{bmatrix} \in \mathbb{R}^{n \times t}, \quad (3)$$

where $f(N_i, T_j)$ denotes the sensor data from node N_i in the j th time slot. Due to the dense distribution of nodes and the slow change of environmental information, the matrix $\mathbf{X}$ is approximately low-rank. Once the data missing occurs, the matrix completion techniques can be utilized as a powerful and effective method in WSNs [9–13].

3. Empirical Study on Environmental Data

In this section, the selected two real environmental data types are introduced, and the low-rank feature and temporal stability of the environmental data are verified.

3.1. Real Environmental Data. To verify the low-rank feature and temporal stability of sensor data in WSNs, two datasets are selected from two real WSNs. The two real environmental data types are GreenOrbs data and Berkeley data.

The GreenOrbs data are gathered from GreenOrbs [19]. This sensor network system deployed for forest surveillance

FIGURE 1: The proportion of sum of the top j singular values.

includes 330 nodes sensing temperature, humidity, and light. Since the raw sensor data in GreenOrbs have data loss, the small but complete subsets were selected as the ground truth for GreenOrbs data. The selected matrix consists of the sensor data from 130 sensor nodes over 120 time slots.

The Berkeley data are gathered from WSN deployed in Intel Berkeley Research lab [20]. This system includes 54 sensors sensing temperature, humidity, light, and voltage once every 31 seconds. Same as the GreenOrbs data, the small but complete subsets were selected as the ground truth for Berkeley data. The selected matrix consists of the sensor data from 54 sensor nodes over 120 time slots.

3.2. Low-Rank Feature. The environment matrix consists of sensor data from different nodes over consecutive time slots. In WSN application, since sensor nodes are densely deployed, the sensor data from different nodes in one-time slots are usually similar. Besides, the sensor data from one node over consecutive time slots are relatively stable due to the slow change of environmental information. As a result, the environment matrix containing the redundant data should have low-rank feature. The Singular Value Decomposition (SVD) was used to verify the low-rank feature.

For the environment matrix $\mathbf{X} \in \mathbb{R}^{n \times t}$, let $\sigma_i(\mathbf{X})$ denote the i th largest singular value of $\mathbf{X}$. The nuclear norm of $\mathbf{X}$ can be stated as $\|\mathbf{X}\|_* = \sum_{i=1}^r \sigma_i(\mathbf{X})$, where r is the rank of $\mathbf{X}$. For the purpose of verifying the low-rank feature of the environment matrix in WSNs, function

$$g(j) = \frac{\sum_{i=1}^j \sigma_i}{\sum_{i=1}^r \sigma_i} = \frac{\sum_{i=1}^j \sigma_i}{\|\mathbf{X}\|_*} \quad (4)$$

is used to represent the proportion of the sum of the top j singular values. The proportion of the sum of the top j singular values is presented in Figure 1. B_Temp, B_Hum, B_Light, and B_Voltage denote the temperature, humidity,

FIGURE 2: CDF of normalized $\mathbf{V}_{ij}$ in the environmental matrix.

light, and voltage in Berkeley data, respectively. G_Temp, G_Hum, and G_Light denote the temperature, humidity, and light in GreenOrbs data, respectively. As shown in Figure 1, for Berkeley data, the proportion of the sum of the top 6 singular values is 84.4% for light data and 95.9% for temperature data. For GreenOrbs data, the proportion of the sum of the top 10 singular values is 84.8% for light data, 90.5% for humidity data, and 93.9% for temperature data. Therefore, the environment matrix $\mathbf{X}$ has a good low-rank feature for each type of environmental information in both datasets.

3.3. Temporal Stability. For most environmental information sensed by WSNs, there are no major changes over adjacent sampling time slots. As a result, the sensor data are relatively stable over consecutive time slots; i.e., there exists temporal stability for each row of environment matrix.

To verify the temporal stability, the difference between adjacent time slots denoted by $\mathbf{V}_{ij}$ for the node N_i in the j th time slot is calculated:

$$\mathbf{V}_{ij} = |f(N_i, T_j) - f(N_i, T_{j-1})| \quad (5)$$

The normalized $\mathbf{V}_{ij}$ is then obtained by dividing by the maximal difference between adjacent time slots in the environmental matrix. Figure 2 shows the cumulative distribution function (CDF) of normalized $\mathbf{V}_{ij}$ in the environmental matrix for each type of signals in two datasets. The x-axis presents the normalized difference between adjacent time slots, and the numerical values on the y-axis present the corresponding CDF values. As shown in Figure 2, most of the values of normalized $\mathbf{V}_{ij}$ are really small, especially for the Berkeley data. Since the sampling time interval is only 31 seconds in Berkeley data, some values are not changed between two consecutive time slots. Therefore, the temporal stability in each row of environment matrix can be utilized for data recovery in WSNs.

4. The Proposed Method

4.1. Problem Formulation. A WSN with n sensor nodes and one sink is considered. The time is divided into equal-sized time slots, and the sensor nodes sense the environmental information and transmit the signal to the sink in each time slot. Let $\mathbf{x} \in \mathbb{R}^{n \times 1}$ denote the raw sensed data in current sampling time slot. However, due to the effect of the hardware and wireless transmission in WSNs, data loss may occur in the network. Besides, to reduce the energy consumption in WSNs, the amount of transmitted data is reduced in sparse sampling based methods. In both cases will result in the data missing problems in WSNs. As a result, only partial data $\mathbf{d} = \Omega(\mathbf{x}) \in \mathbb{R}^{k \times 1}$ are obtained in the sink, where $\Omega : \mathbb{R}^{n \times 1} \rightarrow \mathbb{R}^{k \times 1}$ denotes the sampling operator. That is, k data are obtained, while $n - k$ data are missing. For simplicity, we consider the random data missing case in WSNs. The missing ratio ρ is defined as $\rho = (n - k)/n$.

The aim of data recovery in WSNs is to recover the raw sensed data with high accuracy in the case of data missing. To recover the raw sensed data $\mathbf{x}$ from partial data $\mathbf{d}$ in real-time, the sparsity constraint for $\mathbf{x}$ can be utilized based on CS. Since the ℓ_0 norm minimization problem is NP-hard, ℓ_1 norm minimization can be used as an alternative. Then we have

$$\begin{aligned} & \text{minimize} \quad \|\Psi \mathbf{x}\|_1 \\ & \text{subject to} \quad \mathbf{d} = \Omega(\mathbf{x}), \end{aligned} \quad (6)$$

where $\Psi \in \mathbb{C}^{n \times n}$ is the transform basis. The CS based method cannot achieve satisfying recovery accuracy with large missing ratio. To further improve the recovery accuracy, a real-time data recovery method is proposed by further utilizing the spatiotemporal correlation among WSN data.

4.2. Real-Time Data Recovery Method. To utilize the spatiotemporal correlation among WSN data, the environment matrix should be constructed. We propose to construct the matrix by introducing the historical data denoted by $\mathbf{H} \in \mathbb{R}^{n \times p}$. $\mathbf{H}$ consists of the data collected from last p time slots, and $\mathbf{H}_{i,j}$ denotes the sensor data from node N_i in the last $(p - j + 1)$ th time slot. The environment matrix can be obtained by combining the historical data $\mathbf{H}$ and current data $\mathbf{x}$ as follows:

$$[\mathbf{H} \ \mathbf{x}] = \begin{bmatrix} \mathbf{H}_{1,1} & \cdots & \mathbf{H}_{1,p} & \mathbf{x}_1 \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{H}_{n,1} & \cdots & \mathbf{H}_{n,p} & \mathbf{x}_n \end{bmatrix} \quad (7)$$

As mentioned in Section 3, the generated environment matrix $[\mathbf{H} \ \mathbf{x}]$ has both low-rank and temporal stability features. In this paper, we utilize the joint low-rank constraint and temporal stability to further exploit the data spatiotemporal correlation. To enforce the low-rank constraint, we can minimize the nuclear norm for the generated environment matrix and have

$$\begin{aligned} & \text{minimize} \quad \|[\mathbf{H} \ \mathbf{x}]\|_* \\ & \text{subject to} \quad \mathbf{d} = \Omega(\mathbf{x}). \end{aligned} \quad (8)$$

For the utilization of temporal stability, we propose to use the constraint $\|[\tilde{\mathbf{H}} \ \mathbf{x}]\mathbf{s}\|_2^2$. Here, $\tilde{\mathbf{H}}$ is the last two columns of $\mathbf{H}$; i.e., $\tilde{\mathbf{H}} = [\mathbf{H}_{:,p-1} \ \mathbf{H}_{:,p}]$. $\mathbf{s} = [1 \ -2 \ 1]'$. In order to enforce temporal stability to the raw sensed data $\mathbf{x}$, the historical data collected at adjacent time slots, $\mathbf{H}_{:,p}$ and $\mathbf{H}_{:,p-1}$, can be used. The difference between current time slot and the last historical time slot (i.e., the p th time slot) should be similar to the difference between p th time slot and $(p - 1)$ th time slot. As a result, only the last two columns of historical data $\mathbf{H}$ are needed, while other columns are not needed. The vector $\mathbf{s}$ is used to constrain the similarity between $(\mathbf{x} - \mathbf{H}_{:,p})$ and $(\mathbf{H}_{:,p} - \mathbf{H}_{:,p-1})$. Then we have

$$\begin{aligned} & \text{minimize} \quad \|[\tilde{\mathbf{H}} \ \mathbf{x}]\mathbf{s}\|_2^2 \\ & \text{subject to} \quad \mathbf{d} = \Omega(\mathbf{x}). \end{aligned} \quad (9)$$

Based on the above representation of sparsity, low-rank, and temporal stability constraints, let us combine Equation (6), Equation (8), and Equation (9) into a unified formulation.

$$\begin{aligned} & \text{minimize} \quad \lambda \|\Psi \mathbf{x}\|_1 + \mu \|[\mathbf{H} \ \mathbf{x}]\|_* + \zeta \|[\tilde{\mathbf{H}} \ \mathbf{x}]\mathbf{s}\|_2^2 \\ & \text{subject to} \quad \mathbf{d} = \Omega(\mathbf{x}). \end{aligned} \quad (10)$$

By introducing a quadratic penalty term, the optimization problem Equation (10) can be converted into a corresponding unconstrained formulation as

$$\begin{aligned} \hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \quad & \|\mathbf{d} - \Omega(\mathbf{x})\|_2^2 + \lambda \|\Psi \mathbf{x}\|_1 + \mu \|[\mathbf{H} \ \mathbf{x}]\|_* \\ & + \zeta \|[\tilde{\mathbf{H}} \ \mathbf{x}]\mathbf{s}\|_2^2, \end{aligned} \quad (11)$$

where λ , μ , and ζ are the regularization parameters which control the tradeoff among the optimization targets: presenting sparsity feature of data from current time slot, achieving low-rank of the whole data matrix, maintaining temporal stability, and fitting to the data-fidelity term $\|\mathbf{d} - \Omega(\mathbf{x})\|_2^2$.

4.3. Reconstruction Algorithm. The proposed method in Equation (11) is appealing from a modeling standpoint. To solve the resultant optimization problem, an efficient reconstruction algorithm is described based on ADMM [17]. By adopting variable splitting, the optimization problem in Equation (11) can be converted into the following equivalent constrained optimization problem:

$$\begin{aligned} & \{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{Z}}\} \\ & = \arg \min_{\mathbf{x}, \mathbf{y}, \mathbf{Z}} \|\mathbf{d} - \Omega(\mathbf{x})\|_2^2 + \lambda \|\mathbf{y}\|_1 + \mu \|\mathbf{Z}\|_* \\ & \quad + \zeta \|[\tilde{\mathbf{H}} \ \mathbf{x}]\mathbf{s}\|_2^2 \\ & \text{s.t.} \quad \mathbf{y} = \Psi \mathbf{x}, \\ & \quad \mathbf{Z} = [\mathbf{H} \ \mathbf{x}]. \end{aligned} \quad (12)$$

Then the augmented Lagrangian function for Equation (12) can be expressed as

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \mathbf{y}, \mathbf{Z}, \mathbf{a}, \mathbf{B}) = & \|\mathbf{d} - \Omega(\mathbf{x})\|_2^2 + \lambda \|\mathbf{y}\|_1 \\ & + \langle \mathbf{a}, \mathbf{y} - \Psi \mathbf{x} \rangle + \frac{\alpha}{2} \|\mathbf{y} - \Psi \mathbf{x}\|_2^2 \\ & + \mu \|\mathbf{Z}\|_* + \langle \mathbf{B}, \mathbf{Z} - [\mathbf{H} \ \mathbf{x}] \rangle \\ & + \frac{\beta}{2} \|\mathbf{Z} - [\mathbf{H} \ \mathbf{x}]\|_F^2 + \zeta \|\widetilde{\mathbf{H}} \ \mathbf{x}\|_2^2 \end{aligned} \quad (13)$$

Here, $\mathbf{a} \in \mathbb{C}^{n \times 1}$ and $\mathbf{B} \in \mathbb{C}^{n \times (p+1)}$ are two Lagrangian multipliers. $\alpha > 0$ and $\beta > 0$ are the penalty parameters, and $\|\cdot\|_F$ denotes the Frobenius norm. With the augmented Lagrangian function, i.e., Equation (13), we can use the following alternating direction method to minimize each variate.

$$\mathbf{x}_{k+1} = \arg \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \mathbf{y}_k, \mathbf{Z}_k, \mathbf{a}_k, \mathbf{B}_k) \quad (14)$$

$$\mathbf{y}_{k+1} = \arg \min_{\mathbf{y}} \mathcal{L}(\mathbf{x}_{k+1}, \mathbf{y}, \mathbf{a}_k) \quad (15)$$

$$\mathbf{Z}_{k+1} = \arg \min_{\mathbf{Z}} \mathcal{L}(\mathbf{x}_{k+1}, \mathbf{Z}, \mathbf{B}_k) \quad (16)$$

$$\mathbf{a}_{k+1} = \mathbf{a}_k + \alpha (\mathbf{y}_{k+1} - \Psi \mathbf{x}_{k+1}) \quad (17)$$

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \beta (\mathbf{Z}_{k+1} - [\mathbf{H} \ \mathbf{x}_{k+1}]) \quad (18)$$

The general solutions to Equation (14), Equation (15), and Equation (16) are described as follows. For Equation (14), it can be rewritten as

$$\begin{aligned} \mathbf{x}_{k+1} = \arg \min_{\mathbf{x}} & \|\mathbf{s} - \Phi \mathbf{x}\|_2^2 + \zeta \|\widetilde{\mathbf{H}} \ \mathbf{x}\|_2^2 \\ & + \frac{\alpha}{2} \left\| \mathbf{y}_k - \Psi \mathbf{x} + \frac{\mathbf{a}_k}{\alpha} \right\|_2^2 \\ & + \frac{\beta}{2} \left\| \mathbf{Z}_k - [\mathbf{H} \ \mathbf{x}] + \frac{\mathbf{B}_k}{\beta} \right\|_F^2 \end{aligned} \quad (19)$$

Equation (19) is a standard linear least squares problem, and a number of efficient algorithms can be utilized to solve it. The conjugate gradient algorithm is adopted in this paper. For Equation (15), it can be expressed as

$$\mathbf{y}_{k+1} = \arg \min_{\mathbf{y}} \lambda \|\mathbf{y}\|_1 + \frac{\alpha}{2} \left\| \mathbf{y} - \Psi \mathbf{x}_{k+1} + \frac{\mathbf{a}_k}{\alpha} \right\|_2^2 \quad (20)$$

Then the well-known soft-thresholding formula [21] can be used:

$$\mathbf{y}_{k+1} = \mathcal{S} \left(\Psi \mathbf{x}_{k+1} - \frac{\mathbf{a}_k}{\alpha}, \frac{\lambda}{\alpha} \right), \quad (21)$$

where $\mathcal{S}(q, \tau) := \text{sign}(q) \max(|q| - \tau, 0)$ is a soft-thresholding operator. Similarly, Equation (16) can be rewritten as

$$\mathbf{Z}_{k+1} = \arg \min_{\mathbf{Z}} \mu \|\mathbf{Z}\|_* + \frac{\beta}{2} \left\| \mathbf{Z} - [\mathbf{H} \ \mathbf{x}_{k+1}] + \frac{\mathbf{B}_k}{\beta} \right\|_F^2 \quad (22)$$

The singular value shrinkage operator [22] can be used for Equation (22), and we have

$$\mathbf{Z}_{k+1} = \text{shrink} \left([\mathbf{H} \ \mathbf{x}_{k+1}] - \frac{\mathbf{B}_k}{\beta}, \frac{\mu}{\beta} \right), \quad (23)$$

where $\text{shrink}(\mathbf{A}, \tau)$ denotes applying the soft-thresholding operator (i.e., $\mathcal{S}(q, \tau)$) at level τ to the singular values of matrix $\mathbf{A}$.

Note that $\mathbf{y}_0$, $\mathbf{a}_0$, $\mathbf{Z}_0$, and $\mathbf{B}_0$ can be initialized with zeros vectors or matrices. Besides, the stopping criteria for the solution are set as $\|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2 / \|\mathbf{x}_k\|_2$ being smaller than a predefined tolerance parameter or k exceeding a maximum number of iterations. Based on the above description, the reconstruction algorithm is presented in Algorithm 1.

5. Performance Evaluation

To verify the efficiency of the proposed method, the simulation experiments were designed with two real WSNs data. The performance of the proposed method was compared with the CS method with sparse sensing matrix (denoted as CS method hereafter) [6], joint CS and matrix completion method (CSMC) [16], and Seq-Prog-CS method [15].

5.1. Experimental Environments. The two subsets of Berkeley data and GreenOrbs data used in Section 3 were adopted as the ground truth. For both datasets, one specified column was selected as the data sensed in current time slot, i.e., the raw sensed data $\mathbf{x}$. For each experiment, with the missing ratio ρ set as $\rho = [0.6 \ 0.7 \ 0.8 \ 0.9 \ 0.93 \ 0.96 \ 0.99]$, Ω was generated to randomly retain $[n(1 - \rho)]$ data from $\mathbf{x}$ and discard other data. The CS method, CSMC method, Seq-Prog-CS method, and the proposed method were utilized to recover the current sensed data $\hat{\mathbf{x}}$ from $\mathbf{d} = \Omega(\mathbf{x})$.

To measure the recovery performance of methods, the Normalized Mean Absolute Error (NMAE) defined in Equation (24) was used.

$$\text{NMAE} = \frac{\sum_{i \in \Pi} |\mathbf{x}_i - \hat{\mathbf{x}}_i|}{\sum_{i \in \Pi} |\mathbf{x}_i|} \quad (24)$$

Here, Π represents the missing subset of the complete set of entries $[n] \times 1$. That is, only the recovery accuracy of the missing data was calculated. In each experiment, the process of random generation of Ω and data reconstruction were repeated 100 times for all methods. The mean values were presented for the experimental results. Note that the historical data $\mathbf{H}$ is needed in CSMC, Seq-Prog-CS, and the proposed method. We set $p = 10$ for both datasets in the experiments.

5.2. Recovery Performance and Analysis. Figures 3 and 4 show the recovery performances of each method for Berkeley data and GreenOrbs data, respectively.

For the Berkeley data, since the recovery performances of the humidity and voltage data are similar to that of temperature data, the temperature data were selected as a representative. Figure 3 illustrates the recovery performances of each method for temperature and light data in Berkeley

Input:
 (i) The sampling operator Ω ;
 (ii) Regularization parameters λ , μ , and ζ ;
 (iii) Penalty parameters α and β ;
Initialization:
 (i) $\mathbf{y}_0$, $\mathbf{a}_0$, $\mathbf{Z}_0$, and $\mathbf{B}_0$;
 (ii) Iteration number $k = 0$;
while “not meet stopping criteria” **do**
 (1) Update $\mathbf{x}_{k+1}$ by solving Equation (19);
 (2) Update $\mathbf{y}_{k+1}$ with Equation (21);
 (3) Update $\mathbf{Z}_{k+1}$ with Equation (23);
 (4) Update $\mathbf{a}_{k+1}$ and $\mathbf{B}_{k+1}$ with Equation (17) and Equation (18);
 (5) Iteration number $k = k + 1$;
Output: The reconstructed raw sensed data $\hat{\mathbf{x}}$

ALGORITHM 1: The reconstruction algorithm.

FIGURE 3: Recovery performances of each method for temperature and light data in Berkeley data.

data. As shown in Figure 3, the proposed method achieves lower NMAE than the CS, CSMC, and Seq-Prog-CS method for both temperature and light data. With the missing ratio increases, the recovery accuracy of the CS, CSMC, and Seq-Prog-CS method dramatically degrades, while the proposed method still yields satisfying results. The CS method performs the worst by only enforcing the sparsity constraint without historical data. Note that the solution to the proposed method in Equation (11) reduces to the basic CS method in Equation (6) with $\mu, \zeta = 0$. By comparing to the CS method, the benefits of introducing the historical data and utilizing joint low-rank constraint and temporal stability in the proposed method can be more intuitive. Similarly, if the parameters $\zeta = 0$, then the solution to the proposed method in Equation (11) reduces to the CSMC method. It can be seen from Figure 3 that the proposed method performs better than CSMC method, which demonstrates the benefits of enforcing the temporal stability constraint in the proposed method. Although both the CSMC and Seq-Prog-CS method use

FIGURE 4: Recovery performances of each method for temperature and light data in GreenOrbs data.

the historical data like the proposed method, the proposed method with joint low-rank constraint and temporal stability achieves better performance. Even the missing ratio is as high as 0.96, and the NMAE of the proposed method does not exceed 0.01 for temperature data and 0.07 for light data. It is worth noting that the recovery of Berkeley temperature data is better than the recovery of Berkeley light data for each method. The reason is that the light data have looser spatiotemporal correlation, as shown in Figures 1 and 2.

For the GreenOrbs data, the recovery performances of the temperature data and light data were selected to be shown in Figure 4. As can be seen from Figure 4, the proposed method achieves the best recovery accuracy among these three methods for both types of signal in GreenOrbs data. That is, for data missing caused by data loss, the proposed method can achieve higher recovery accuracy than the compared methods. For data missing caused by utilizing sparse sampling to reduce energy consumption in WSNs, to ensure the same data recovery accuracy, the proposed

method requires less sampled data than the compared methods. We can conclude that the proposed method outperforms the CS, CSMC, and Seq-Prog-CS methods for each type of environmental information.

6. Conclusion

In this paper, we propose a real-time data recovery method based on sparse representation. By introducing the historical data, the proposed method constrains both low-rank and temporal stability to further utilize the spatiotemporal correlation among WSN data. Furthermore, an efficient reconstruction algorithm is described based on ADMM for the resultant optimization problem. The optimal regularization parameters can be easily selected with the historical data in the network. The experimental results show that the proposed method outperforms the state-of-the-art methods for different types of signal in the network.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by National Natural Science Foundation of China (No. 61801164), Natural Science Foundation of Tianjin City (No. 18JCQNJC01700), Foundation of Hebei Educational Committee (No. QN2018092), and Natural Science Foundation of Hebei Province (No. F2019202387).

References

- [1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] C. Luo, F. Wu, J. Sun, and C. W. Chen, "Compressive data gathering for large-scale wireless sensor networks," in *Proceedings of the 15th Annual ACM International Conference on Mobile Computing and Networking (MobiCom '09)*, pp. 145–156, ACM, September 2009.
- [3] C. Zhao, W. Zhang, Y. Yang, and S. Yao, "Treelet-based clustered compressive data aggregation for wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 9, pp. 4257–4267, 2015.
- [4] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [5] H. Zhu, Y. Zhu, M. Li, and L. M. Ni, "SEER: metropolitan-scale traffic perception based on lossy sensory data," in *Proceedings of the 28th Conference on Computer Communications (IEEE INFOCOM '09)*, pp. 217–225, IEEE, April 2009.
- [6] X. Wu, Y. Xiong, P. Yang, S. Wan, and W. Huang, "Sparsest random scheduling for compressive data gathering in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 10, pp. 5867–5877, 2014.
- [7] G. Quer, R. Masiero, G. Pillonetto, M. Rossi, and M. Zorzi, "Sensing, compression, and recovery for WSNs: sparse signal modeling and monitoring framework," *IEEE Transactions on Wireless Communications*, vol. 11, no. 10, pp. 3447–3461, 2012.
- [8] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.
- [9] J. Cheng, H. Jiang, X. Ma et al., "Efficient data collection with sampling in WSNs: making use of matrix completion techniques," in *Proceedings of the 53rd IEEE Global Communications Conference (GLOBECOM '10)*, pp. 1–5, IEEE, Miami, Fla, USA, December 2010.
- [10] J. Cheng, Q. Ye, H. Jiang, D. Wang, and C. Wang, "STCDG: an efficient data gathering algorithm based on matrix completion for wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 2, pp. 850–861, 2013.
- [11] L. Kong, M. Xia, X.-Y. Liu et al., "Data loss and reconstruction in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 11, pp. 2818–2828, 2014.
- [12] K. Xie, L. Wang, X. Wang, G. Xie, and J. Wen, "Low cost and high accuracy data gathering in WSNs with matrix completion," *IEEE Transactions on Mobile Computing*, vol. 17, no. 7, pp. 1595–1608, 2018.
- [13] J. He, G. Sun, Y. Zhang, and Z. Wang, "Data recovery in wireless sensor networks with joint matrix completion and sparsity constraints," *IEEE Communications Letters*, vol. 19, no. 12, pp. 2230–2233, 2015.
- [14] K. Yi, J. Wan, L. Yao, and T. Bao, "Partial matrix completion algorithm for efficient data gathering in wireless sensor networks," *IEEE Communications Letters*, vol. 19, no. 1, pp. 54–57, 2015.
- [15] M. Leinonen, M. Codreanu, and M. Juntti, "Sequential compressed sensing with progressive signal reconstruction in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 3, pp. 1622–1635, 2015.
- [16] J. He, G. Sun, Z. Li, and Y. Zhang, "Compressive data gathering with low-rank constraints for Wireless Sensor networks," *Signal Processing*, vol. 131, pp. 73–76, 2017.
- [17] N. Parikh and S. P. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.
- [18] E. J. Candès and T. Tao, "The power of convex relaxation: near-optimal matrix completion," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2053–2080, 2010.
- [19] W. Dong, Y. Liu, Y. He, T. Zhu, and C. Chen, "Measurement and analysis on the packet delivery performance in a large-scale sensor network," *IEEE/ACM Transactions on Networking (TON)*, vol. 22, no. 6, pp. 1952–1963, 2014.
- [20] S. Madden, *Intel Berkeley Research Lab Data*, Intel Corporation, USA, 2004, <http://berkeley.intel-research.net/labdata.html>.
- [21] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.
- [22] J. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.