



Article

Missing and Corrupted Data Recovery in Wireless Sensor Networks Based on Weighted Robust Principal Component Analysis

Jingfei He , Yunpei Li , Xiaoyue Zhang and Jianwei Li

Tianjin Key Laboratory of Electronic Materials and Devices, School of Electronics and Information Engineering, Hebei University of Technology, Tianjin 300401, China; liyunpei_hebut@163.com (Y.L.); zxy_zhangxiaoyue@163.com (X.Z.); ljw__0917@163.com (J.L.)

* Correspondence: hejingfei@hebut.edu.cn

Abstract: Although wireless sensor networks (WSNs) have been widely used, the existence of data loss and corruption caused by poor network conditions, sensor bandwidth, and node failure during transmission greatly affects the credibility of monitoring data. To solve this problem, this paper proposes a weighted robust principal component analysis method to recover the corrupted and missing data in WSNs. By decomposing the original data into a low-rank normal data matrix and a sparse abnormal matrix, the proposed method can identify the abnormal data and avoid the influence of corruption on the reconstruction of normal data. In addition, the low-rankness is constrained by weighted nuclear norm minimization instead of the nuclear norm minimization to preserve the major data components and ensure credible reconstruction data. An alternating direction method of multipliers algorithm is further developed to solve the resultant optimization problem. Experimental results demonstrate that the proposed method outperforms many state-of-the-art methods in terms of recovery accuracy in real WSNs.

Keywords: wireless sensor networks; missing and corrupted data recovery; weighted nuclear norm; robust principal component analysis



Citation: He, J.; Li, Y.; Zhang, X.; Li, J. Missing and Corrupted Data Recovery in Wireless Sensor Networks Based on Weighted Robust Principal Component Analysis. *Sensors* **2022**, *22*, 1992. <https://doi.org/10.3390/s22051992>

Academic Editors: Alvaro Araujo Pinto and Hacene Fouchal

Received: 6 January 2022

Accepted: 1 March 2022

Published: 3 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Wireless sensor networks (WSNs) contain a group of spatially distributed sensor nodes that are capable of communicating wirelessly and collecting data from the surrounding environments [1,2]. Recently, WSNs have been widely applied in different domains, such as environmental monitoring [3], military management [4], and health care [5]. Typically, the main task of WSNs is to collect sensing data from all sensor nodes to a certain sink and then perform further analysis based on the monitoring data, and the collected data are usually composed of readings sensed by multiple nodes in consecutive time slots. However, due to the poor environments and energy constraints in WSNs, data loss and corruption are inevitable in practical applications. Therefore, it is important to reconstruct the real data from partially collected data with corruption.

Recently, various reconstruction methods have been proposed for data recovery in WSNs. Based on data interpolation techniques, a K nearest neighbor (KNN)-based method [6] was proposed to simply utilize the values of the nearest neighbors to estimate the missing values. The Delaunay triangulation (DT) [7] utilizes the vertices as their global errors to reconstruct virtual triangles for data interpolation. Based on compressed sensing (CS) [8], the distributed compressed sensing (DCS) method [9,10] was proposed to exploit the sparsity of the data under various transform domains.

Since many signals in various applications are always distributed into two-dimensional data (i.e., matrix form) and exhibit second-order sparsity (i.e., the low-rankness), matrix completion (MC) [11] has emerged as a novel technology and has been applied to many fields, such as image inpainting [12], magnetic resonance imaging [13], and recommendation systems [14]. The matrix completion aims at recovering the missing entries of a

low-rank matrix from the incompleteness observations, which can be formulated as a rank minimization problem. In general, solving this problem is NP-hard, since the rank function is non-convex. Fortunately, the nuclear norm, the sum of all singular values of the matrix, is the convex approximation of the rank function and can be used as an alternative [11].

Since the readings collected from N nodes during M time slots in WSNs can also be distributed into a matrix exhibiting low-rankness, the matrix completion-based methods have been proposed to utilize the correlation of WSNs data. An efficient data collection approach (EDCA) [15] and spatiotemporal compressive data collection (STCDG) method [16] were firstly proposed to recover the WSNs data by exploiting the spatiotemporal correlation in the form of low-rankness. Recently, several methods jointly utilizing low-rank and spatiotemporal sparsity feature [17,18] were proposed. Considering that the missing of row of the data matrix due to a broken node will greatly degrade the recovery accuracy, the matrix completion method [19] was proposed to utilize the interpolation technique for WSNs data recovery. In addition, in order to address the needs of real-time reconstruction of data in practical applications, the sliding window-based reconstruction approach [20,21] was proposed to achieve real-time data recovery.

However, the reconstruction performance of these methods will greatly degenerate when corruption exists in the sampled data. Direct constraint of the low-rankness cannot avoid the impact of corruption on the reconstruction of normal data. A two-phase MC-based data recovery scheme (MC-Two-Phase) [22] was proposed to recover the normal data without the influence of corruption by detecting the corruption with the principal component analysis (PCA) [23] before reconstruction. Although PCA can be utilized to detect faults corrupted by small noise, it has the problem of poor robustness. To overcome the limitations of PCA, the robust principal component analysis (RPCA) method [24–26] have been proposed in recent years. The RPCA method improves the robustness since it only emphasized that the noise is sparse regardless of the strength of the noise. However, it is unreasonable to treat all singular values equally in the traditional RPCA algorithm, since different singular values may contain signal information with different important levels.

To solve the above problem, we propose a weighted robust principal component analysis (WRPCA) method for the reconstruction of WSN data with corruption. The main contributions of this paper are the following:

Firstly, based on RPCA, the original data with outliers are decomposed into a sum of a low-rank normal data matrix and a sparse abnormal matrix to avoid the influence of outliers during reconstruction.

Secondly, the low-rankness of WSNs data is revealed by the variation of singular values of two real datasets collected from the Inter Berkeley Research lab and GreenOrbs.

Thirdly, the weighted nuclear norm is introduced to constrain the low-rankness and preserve the principal components of WSNs data.

The rest of this paper is organized as follows. Section 2 presents the basics of RPCA. Section 3 describes the proposed method and the reconstruction method. Section 4 shows the result of computer experiments and analysis, which is followed by the conclusion of the paper in Section 5.

2. Basics of RPCA

Although PCA can be used to detect corruptions, it is sensitive to gross noise and outliers. The performance and applicability of PCA are limited due to the lack of robustness to gross corruptions in real-life scenarios. As an improvement of PCA, RPCA can handle grossly corrupted data well. Suppose that data matrix \mathbf{X} can be viewed as consisting of the two components: a low-rank matrix \mathbf{L} and a sparse matrix \mathbf{S} :

$$\mathbf{X} = \mathbf{L} + \mathbf{S}. \quad (1)$$

The low-rank matrix \mathbf{L} and sparse matrix \mathbf{S} can be obtained by solving the following problem:

$$\begin{aligned} \min_{\mathbf{L}, \mathbf{S}} \text{rank}(\mathbf{L}) + \lambda \|\mathbf{S}\|_0 \\ \text{s.t. } \mathbf{X} = \mathbf{L} + \mathbf{S}, \end{aligned} \quad (2)$$

where $\text{rank}(\cdot)$ denotes the rank of the matrix, $\|\cdot\|_0$ is the l_0 norm, and λ is the balance parameter.

Equation (2) is non-convex and NP-hard, which is difficult to solve. Typically, the matrix nuclear norm, the convex approximation of the rank function, can be used as an alternative. Therefore, the above problem can be cast as the following convex optimization problem:

$$\begin{aligned} \min_{\mathbf{L}, \mathbf{S}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \\ \text{s.t. } \mathbf{X} = \mathbf{L} + \mathbf{S}, \end{aligned} \quad (3)$$

where $\|\mathbf{L}\|_* = \sum_i \sigma_i(\mathbf{L})$ denotes the nuclear norm of matrix \mathbf{L} , $\sigma_i(\mathbf{L})$ is the i -th singular value of matrix \mathbf{L} , and $\|\cdot\|_1$ is the l_1 norm. The main goal of (3) is to reconstruct low-rank normal data \mathbf{L} from the corrupted observation data \mathbf{X} .

RPCA has been successfully applied in different domains, including image processing [27], multimedia [28], document analysis [29], etc. The nuclear norm minimization utilized in (3) shrinks all the singular values equally [30], ignoring that different singular values may have different importance.

Actually, the real data sensed in the monitoring area always exhibit low-rankness, and the unavoidable corrupted data are sparsely distributed in the sensed data matrix. Based on RPCA, we propose a weighted robust principal component analysis method to recover the missing data in WSNs with the data corruption.

3. The Proposed Method

3.1. Problem Formulation and Signal Feature

Consider a WSN consisting of one sink and N sensor nodes, and the sensor nodes sense the environmental information and send the signal to the sink in each time slot. During M time slots, $N \times M$ readings are gathered in the sink and can be organized into a matrix $\mathbf{X} \in \mathbb{R}^{N \times M}$.

However, due to hardware and network conditions, data loss and corruption may occur in the network. Mathematically, only partial data $\mathbf{d} = \Omega(\mathbf{X})$ can be successfully collected in the sink, and the original data \mathbf{X} contain the corrupted data. Here, $\Omega(\cdot)$ is the random sampling operator. That is, under the sampling ratio ρ_s , for a matrix $\mathbf{X} \in \mathbb{R}^{N \times M}$, there are $\mathbf{d} = \Omega(\mathbf{X}) \in \mathbb{R}^{D \times 1}$ entries that are sampled from the whole data randomly, where $D = \lfloor \rho_s NM + \frac{1}{2} \rfloor$. It is worth noting that the sampled partial data \mathbf{d} also contains the sampled corruption data. It is necessary to reconstruct the uncorrupted whole data from the sampled partial data under the sampling ratio ρ_s .

The data sensed in a certain area during a consecutive time are always redundant and highly correlated and can be distributed into a matrix (uncorrupted matrix \mathbf{L}) exhibiting low-rankness. Since the outliers in real WSN are uniformly and randomly distributed, the sparsely distributed corrupted data can be denoted by the matrix \mathbf{S} . Therefore, the whole data \mathbf{X} can be regarded as a combination of the uncorrupted data matrix \mathbf{L} and the corrupted matrix \mathbf{S} .

In order to verify that the uncorrupted data \mathbf{L} in WSNs is low-rank, two datasets from the Inter Berkeley Research lab [31] and GreenOrbs [32] were used as testing data. Since data loss and corruption exist in both two datasets, two small but completed subset data without corruption are selected as the ground truth for our verification experiment. Specifically, the selected Inter Berkeley Research lab subset data including temperature and humidity data were measured by 49 sensor nodes during 138 time slots, and the selected GreenOrbs subset data were measured by 130 sensor nodes during 129 time slots. As shown

in Figure 1, the singular values of the two attribute data matrix illustrate the low-rankness for both two datasets.

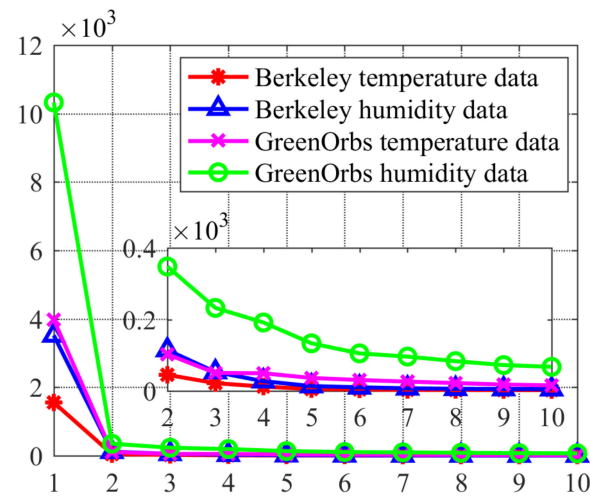


Figure 1. The first 10 singular values of two attribute data matrix for Berkeley and GreenOrbs data.

3.2. Proposed Method

Since the original data matrix \mathbf{X} can be decomposed into a low-rank matrix \mathbf{L} and a sparse matrix \mathbf{S} , the WSNs data recovery problem can be expressed by (3). However, the NNM method adopts the same threshold for each singular value, which is not appropriate because the larger singular values usually represent the major data components of the data and contain more signal information. The larger singular values should be shrunk less to preserve the major data components.

In order to improve the practicality and flexibility of the nuclear norm, the weighted nuclear norm is utilized in the recovery of WSNs data. The weighted nuclear norm of matrix \mathbf{L} is defined as:

$$\|\mathbf{L}\|_{w,*} = \sum_i w_i \sigma_i(\mathbf{L}), \quad (4)$$

where w_i is the weight coefficient, and $\sigma_i(\mathbf{L})$ is the i -th singular value of \mathbf{L} . It is clear that the weighted nuclear norm becomes the conventional nuclear norm when $w_1 = w_2 = \dots = w_n$.

The weighted nuclear norm minimization (WNNM) based low-rank matrix completion problem can be described as:

$$\mathbf{L} = \underset{\mathbf{L}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{L}\|_F^2 + \lambda \|\mathbf{L}\|_{w,*} \quad (5)$$

Gu et al. [30] proved that the problem can be solved by the following singular value thresholding formula:

$$\mathbf{L}_{k+1} = \operatorname{shrink}(\mathbf{Y}_k, w_i) \quad (6)$$

The larger singular values should be given smaller weights to achieve less shrinkage, and the smaller ones should be given greater weights to achieve more shrinkage. The weights should be inversely proportional to singular values. Therefore, in this paper, we set the weight as:

$$w_i = \frac{c \cdot \sqrt{M} \cdot \sigma^2}{\sigma_i(\mathbf{L}) + \varepsilon}, \quad (7)$$

where $c > 0$ is a constant, M is the number of columns in \mathbf{L} , σ^2 is the variance of noise, and ε only needs to be a very small number to avoid dividing by zero.

By introducing the weighted approach, different singular values are shrunk differently with weight w_i , which further preserves the major components of data. Then, a WSNs data

reconstruction method is proposed by applying WNNM in traditional RPCA to recover \mathbf{L} from partial measurement \mathbf{d} . It can be described as:

$$\begin{aligned} \min_{\mathbf{L}, \mathbf{S}} \quad & \|\mathbf{L}\|_{w,*} + \lambda \|\mathbf{S}\|_1 \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{L} + \mathbf{S}, \mathbf{d} = \Omega(\mathbf{X}). \end{aligned} \quad (8)$$

Only partial measurement \mathbf{d} is known as a prior in (8). The original data \mathbf{X} and uncorrupted \mathbf{L} can be reconstructed from the partial data \mathbf{d} . By introducing a quadratic penalty term, (8) can be converted to the following formulation:

$$\begin{aligned} \min_{\mathbf{L}, \mathbf{S}} \quad & \|\mathbf{d} - \Omega(\mathbf{X})\|_2^2 + \mu \|\mathbf{L}\|_{w,*} + \lambda \|\mathbf{S}\|_1 \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{L} + \mathbf{S}, \end{aligned} \quad (9)$$

where μ and λ are the regularization parameters. The proposed method incorporates both the RPCA and WNNM in a single formulate to further preserve the major data components. The recovered $\hat{\mathbf{L}}$ can be obtained as the uncorrupted completed data in WSNs.

3.3. Model Optimization

To solve (9), a reconstructed algorithm based on an alternating direction method of multipliers (ADMM) [33,34] is introduced. The augmented Lagrangian function of (9) can be written as:

$$\begin{aligned} \mathcal{L}(\mathbf{X}, \mathbf{L}, \mathbf{S}, \mathbf{A}) = \quad & \|\mathbf{d} - \Omega(\mathbf{X})\|_2^2 + \mu \|\mathbf{L}\|_{w,*} + \lambda \|\mathbf{S}\|_1 \\ & + \langle \mathbf{A}, \mathbf{L} + \mathbf{S} - \mathbf{X} \rangle + \frac{\alpha}{2} \|\mathbf{L} + \mathbf{S} - \mathbf{X}\|_F^2, \end{aligned} \quad (10)$$

where \mathbf{A} is the Lagrangian multiplier, and α is the penalty parameter. More details of the proposed algorithm are given as follows.

For the \mathbf{X} -subproblem, we update \mathbf{X}_{k+1} as follows:

$$\begin{aligned} \mathbf{X}_{k+1} &= \underset{\mathbf{X}}{\operatorname{argmin}} \mathcal{L}(\mathbf{X}, \mathbf{L}_k, \mathbf{S}_k, \mathbf{A}_k) \\ &= \underset{\mathbf{X}}{\operatorname{argmin}} \|\mathbf{d} - \Omega(\mathbf{X})\|_2^2 + \frac{\alpha}{2} \left\| \mathbf{L}_k + \mathbf{S}_k - \mathbf{X} + \frac{\mathbf{A}_k}{\alpha} \right\|_F^2. \end{aligned} \quad (11)$$

Here, the preconditional conjugate gradient (PCG) algorithm is applied to solve this problem in this paper.

For the \mathbf{L} -subproblem, we update \mathbf{L}_{k+1} as follows:

$$\begin{aligned} \mathbf{L}_{k+1} &= \underset{\mathbf{L}}{\operatorname{argmin}} \mathcal{L}(\mathbf{X}_{k+1}, \mathbf{L}, \mathbf{S}_k, \mathbf{A}_k) \\ &= \underset{\mathbf{L}}{\operatorname{argmin}} \mu \|\mathbf{L}\|_{w,*} + \frac{\alpha}{2} \left\| \mathbf{L} + \mathbf{S}_k - \mathbf{X}_{k+1} + \frac{\mathbf{A}_k}{\alpha} \right\|_F^2. \end{aligned} \quad (12)$$

In general, the WNNM problem is non-convex. Gu et al. [30] proved that the problem has a fixed point and can be solved by the singular value thresholding formula:

$$\mathbf{L}_{k+1} = \mathbf{U} \mathbf{S}_w(\boldsymbol{\Sigma}) \mathbf{V}^T, \quad (13)$$

where $[\mathbf{U}, \boldsymbol{\Sigma}, \mathbf{V}] = \operatorname{SVD}\left(\mathbf{X}_{k+1} - \mathbf{S}_k - \frac{\mathbf{A}_k}{\alpha}\right)$ is the Singular Value Decomposition (SVD) of $\mathbf{X}_{k+1} - \mathbf{S}_k - \frac{\mathbf{A}_k}{\alpha}$, and $\mathbf{S}_w(\boldsymbol{\Sigma})$ denotes taking singular value thresholding to the diagonal matrix $\boldsymbol{\Sigma}$. Since the threshold can be effected by w_i and μ , here, we set $w_i = \frac{\mu \cdot \sqrt{M} \cdot \sigma^2}{\sigma_i(\mathbf{L}) + \varepsilon}$ to simplify the solution; then, $\mathbf{S}_w(\boldsymbol{\Sigma})_{ii} = \max(\boldsymbol{\Sigma}_{ii} - w_i, 0)$. The initial $\sigma_i(\mathbf{L}_{k+1})$ can be estimated as:

$$\sigma_i(\mathbf{L}_{k+1}) = \sqrt{\max(\boldsymbol{\Sigma}_{ii}^2 - M\sigma^2, 0)}. \quad (14)$$

For the \mathbf{S} -subproblem, we update \mathbf{S}_{k+1} as follows:

$$\begin{aligned}\mathbf{S}_{k+1} &= \underset{\mathbf{S}}{\operatorname{argmin}} \mathcal{L}(\mathbf{X}_{k+1}, \mathbf{L}_{k+1}, \mathbf{S}, \mathbf{A}_k) \\ &= \underset{\mathbf{S}}{\operatorname{argmin}} \lambda \|\mathbf{S}\|_1 + \frac{\alpha}{2} \left\| \mathbf{L}_{k+1} + \mathbf{S} - \mathbf{X}_{k+1} + \frac{\mathbf{A}_k}{\alpha} \right\|_F^2.\end{aligned}\quad (15)$$

We can find the solution via the well-known soft thresholding formula:

$$\mathbf{S}_{k+1} = \operatorname{soft} \left(\mathbf{X}_{k+1} - \mathbf{L}_{k+1} - \frac{\mathbf{A}_k}{\alpha}, \frac{\lambda}{\alpha} \right).\quad (16)$$

For the \mathbf{A} -subproblem, we update \mathbf{A}_{k+1} as follows:

$$\mathbf{A}_{k+1} = \mathbf{A}_k + \alpha(\mathbf{L}_{k+1} + \mathbf{S}_{k+1} - \mathbf{X}_{k+1}).\quad (17)$$

In practical implementation, we initialize \mathbf{X}_0 , \mathbf{L}_0 , \mathbf{S}_0 , and \mathbf{A}_0 as the zeros matrices. Then, (9) can be solved by repeating the above steps until $\|\mathbf{L}_{k+1} - \mathbf{L}_k\|_F / \|\mathbf{L}_k\|_F$ is smaller than a predefined tolerance parameter or the number of iterations reaches the predefined maximum.

The main computational cost of (9) depends on the update of \mathbf{L}_{k+1} , which requires computing the SVD of the $N \times M$ matrix per iteration. The computational complexity per iteration is $O(\min\{NM^2, N^2M\})$.

4. Experiments and Analysis

Most existing WSNs data reconstruction methods (e.g., KNN [6], CS [9], EDCA [15], and methods utilizing both low-rank and sparsity feature [17,18]) have achieved satisfying recovery performances, but they do not consider the case that the WSNs data have outliers. Therefore, the performance of our proposed method is compared with the RPCA method [24] and MC-Two-Phase method [22], which consider outliers during reconstruction.

4.1. Experimental Environments

The two datasets adopted to verify the low-rank property of normal WSNs data were also utilized for the reconstruction experiments. The normal data without corruption can be denoted by \mathbf{L}_{nor} , and let $\mathbf{L}_{\text{nor}_B}$ and $\mathbf{L}_{\text{nor}_G}$ denote the normal matrix for Berkeley data and GreenOrbs data, respectively. The normal data can be regarded as the ground truth WSNs data. In real WSNs, due to data loss and corruption, the measurement \mathbf{d} is partially sampled and contains corruption. To obtain the partial measurement \mathbf{d} from normal matrix \mathbf{L}_{nor} in the experiment, the following steps were performed. Firstly, the partial sampled normal data $\mathbf{d}_{\text{nor}} \in \mathbb{R}^{D \times 1}$ can be obtained by $\mathbf{d}_{\text{nor}} = \Omega(\mathbf{L}_{\text{nor}})$ according to the sampling ratio ρ_s . Then, $\rho_c \times D$ entries in \mathbf{d}_{nor} were randomly selected as the corruption data by adding additional random Gaussian noise with zero mean and variance $\sigma^2 = 20$, where ρ_c is the corruption ratio.

The parameters λ , μ and α can be chosen according to the characteristics of the signal collected by the sink. In this paper, ε , λ , μ , and α are set to 10^{-6} , 0.05, 3.3, and 0.05, respectively.

The Normalized Mean Absolute Error (NMAE) is used to measure the recovery performance of different methods on missing data and corrupted data:

$$NMAE_{\text{loss}} = \frac{\sum_{i,j \in \Pi_m} |\mathbf{L}_{\text{nor}}(i,j) - \hat{\mathbf{L}}(i,j)|}{\sum_{i,j \in \Pi_m} |\mathbf{L}_{\text{nor}}(i,j)|},\quad (18)$$

$$NMAE_{\text{cor}} = \frac{\sum_{i,j \in \Pi_c} |\mathbf{L}_{\text{nor}}(i,j) - \hat{\mathbf{L}}(i,j)|}{\sum_{i,j \in \Pi_c} |\mathbf{L}_{\text{nor}}(i,j)|},\quad (19)$$

where $\hat{\mathbf{L}}$ is the recovered data, Π_m denotes the missing data set, and Π_c is the corrupted data set. The experimental result is an average of 50 repeated experiments.

4.2. Recovery Performance Comparisons

To compare the proposed method with the existing methods, temperature and humidity data from the Inter Berkeley Research lab and GreenOrbs were utilized for the recovery performance comparisons. With the sampling ratio $\rho_s = 0.1, 0.2, 0.3,$ and 0.4 , the corruption ratio $\rho_c = 0.2, 0.3, 0.4, 0.5,$ and 0.6 , Figures 2 and 3 show the recovery performance of each method for Berkeley temperature and humidity data, while Figures 4 and 5 show the recovery performance of each method for GreenOrbs temperature and humidity data, respectively. As can be seen, the proposed method has a lower NMAE than the comparison methods on both missing and corrupted data in two datasets, especially at low sampling ratio and high corruption ratio.

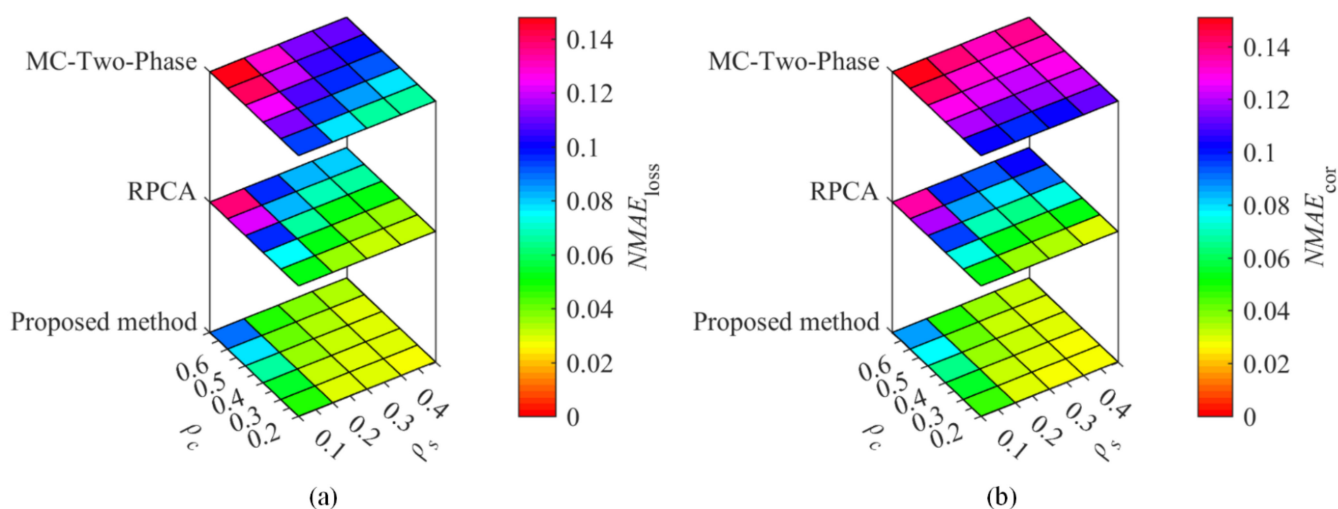


Figure 2. The recovery performance of each method for temperature data sensed in the Inter Berkeley Research lab. The (a) is the NMAE of missing data, (b) the NMAE of corrupted data.

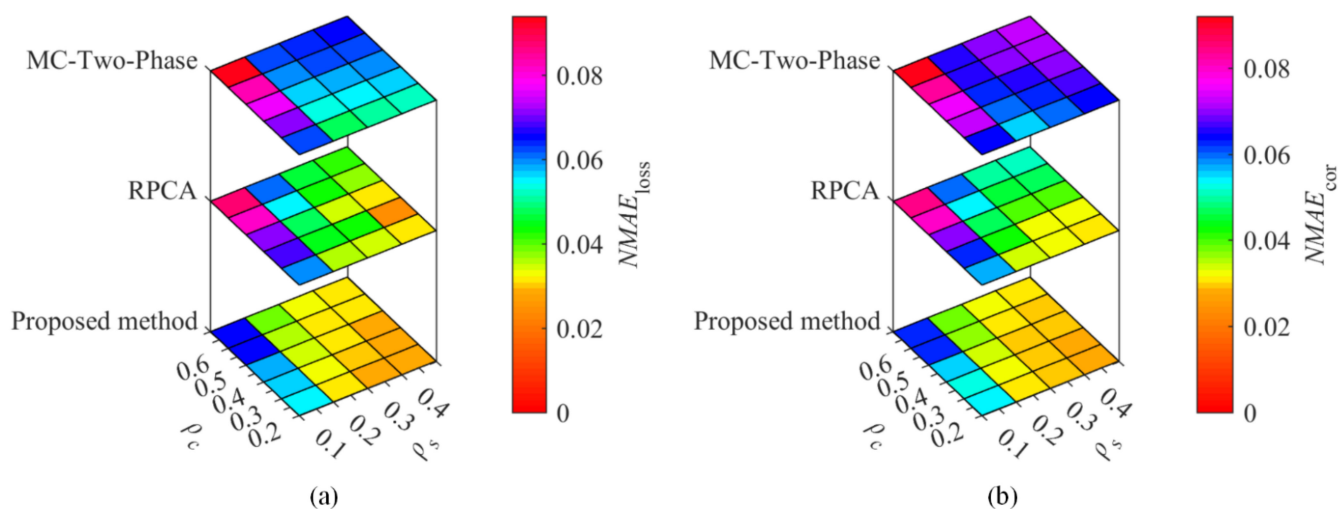


Figure 3. The recovery performance of each method for humidity data sensed in the Inter Berkeley Research lab. The (a) is the NMAE of missing data, (b) the NMAE of corrupted data.

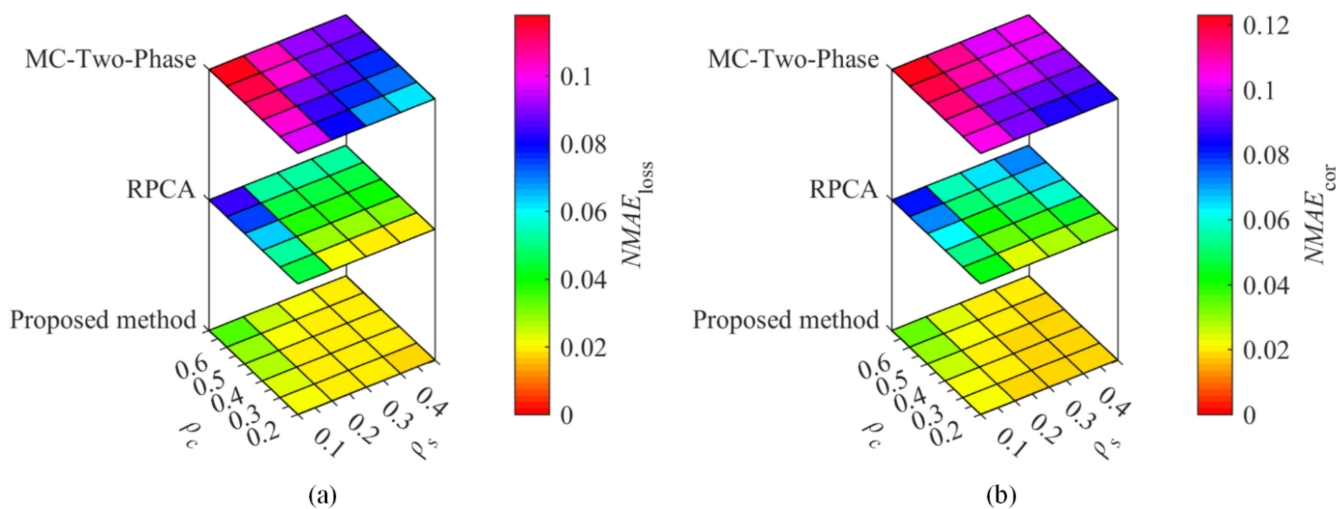


Figure 4. The recovery performance of each method for temperature data sensed in GreenOrbs. The (a) is the NMAE of missing data, (b) the NMAE of corrupted data.

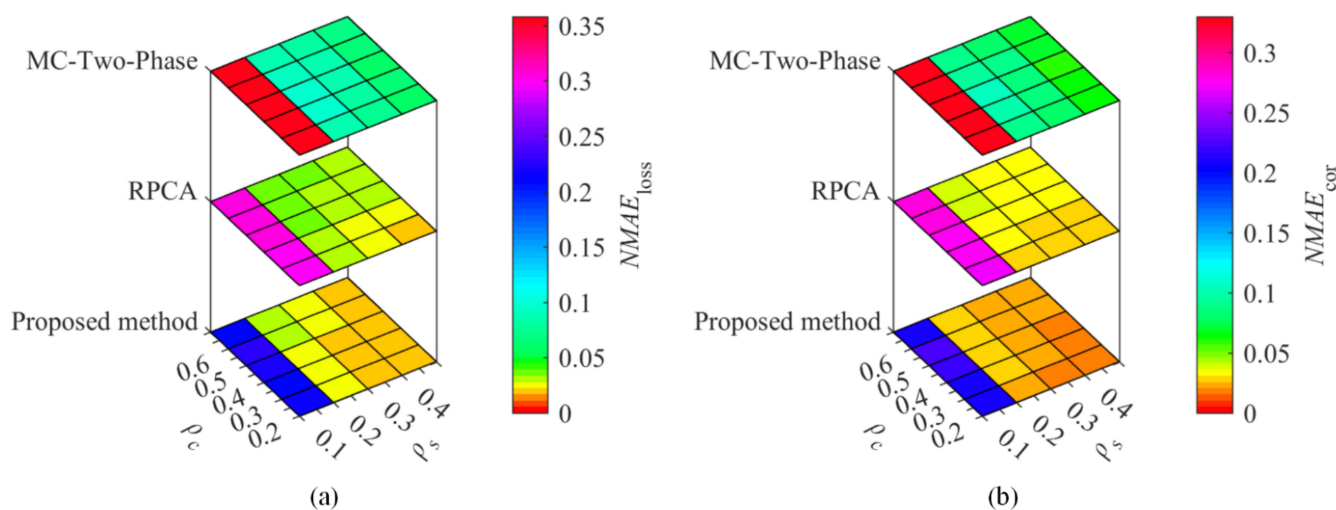


Figure 5. The recovery performance of each method for humidity data sensed in GreenOrbs. The (a) is the NMAE of missing data, (b) the NMAE of corrupted data.

As shown in Figure 2b, the $NMAE_{cor}$ of the proposed method, RPCA, and MC-Two-Phase is 0.030, 0.037, and 0.099 when $\rho_s = 0.2$ and $\rho_c = 0.2$, while the values are 0.048, 0.097, and 0.140 when ρ_c is up to 0.6. The results show that the proposed method not only improves the recovery accuracy of WSNs data but also has strong robustness to the gross noise.

Especially, from Figure 4, we can see that as the ρ_c increases from 0.2 to 0.6, the corresponding $NMAE_{loss}$ and $NMAE_{cor}$ of MC-Two-Phase and RPCA dramatically increase, while that has very little change of the proposed method, and even the entire range of change is only less than 0.01.

As shown in Figure 3, the $NMAE_{loss}$ and $NMAE_{cor}$ of the proposed method decrease rapidly with the increase in ρ_s , but there is little change for the MC-Two-Phase. Specifically, in Figure 3b, the $NMAE_{cor}$ of the proposed method, RPCA and MC-Two-Phase is 0.056, 0.069, and 0.076 when $\rho_s = 0.1$ and $\rho_c = 0.4$, while the values are 0.029, 0.040, and 0.069 when ρ_s is up to 0.4.

5. Conclusions

In this paper, we propose a WRPCA method to increase the recovery accuracy of WSNs data with loss and corruption. The original data matrix is treated as a sum of a low-rank normal data matrix and a sparse abnormal matrix to avoid the influence of corruption. In addition, the weighted nuclear norm minimization is utilized to further constrain the low-rankness of the normal data and overcome the problem that the nuclear norm minimization treats all singular values equally. The experimental results show that the proposed method has better recovery performance in both loss and corruption data. In further work, the higher-order low-rankness of multi-attribute data in WSNs can be explored for multi-attribute data reconstruction.

Author Contributions: Conceptualization, J.H. and Y.L.; methodology, J.H. and Y.L.; software, Y.L., X.Z. and J.L.; validation, Y.L., X.Z. and J.L.; writing—original draft preparation, Y.L.; writing—review and editing, J.H. and X.Z.; supervision, J.H.; project administration, J.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (No. 61801164), Key Research and Development Program of Hebei Province (No. F21320301D), Natural Science Foundation of Hebei Province (No. F2019202387), and Hebei Foundation for Returned Overseas Chinese Scholars (No. C20200312).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data in this study are available upon request to the correspondence author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kong, L.; Xia, M.; Liu, X.; Chen, G.; Gu, Y. Data Loss and Reconstruction in Wireless Sensor Networks. *IEEE Trans. Parallel Distrib. Syst.* **2014**, *25*, 2818–2828. [[CrossRef](#)]
2. Rawat, A.; Gupta, A.; Singh, A.; Bhushan, S. Energy conservation and Missing value prediction model in Wireless Sensor Network. In Proceedings of the 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), Ghaziabad, India, 18–19 April 2019; pp. 1–5.
3. Liu, H.; Meng, Z.; Shang, Y. Sensor Nodes Placement for Farmland Environmental Monitoring Applications. In Proceedings of the 2009 5th International Conference on Wireless Communications, Networking and Mobile Computing, Beijing, China, 24–26 September 2009; pp. 1–4.
4. Koskiahde, T.; Kujala, J.; Norolampi, T. A sensor network architecture for military and crisis management. In Proceedings of the 2008 IEEE International Symposium on Precision Clock Synchronization for Measurement, Control and Communication, Ann Arbor, MI, USA, 22–26 September 2008; pp. 110–114.
5. Rabin, M.; Braun, F.; Adler, A.; Wacker, J.; Frerichs, I.; Vogt, B.; Chetelat, O. Wearable Sensors for Frequency-Multiplexed EIT and Multilead ECG Data Acquisition. *IEEE Trans. Biomed. Eng.* **2019**, *66*, 810–820. [[CrossRef](#)] [[PubMed](#)]
6. Jayaraman, P.P.; Zaslavsky, A.; Delsing, J. Cost-Efficient Data Collection Approach Using K-Nearest Neighbors in a 3D Sensor Network. In Proceedings of the 2010 Eleventh International Conference on Mobile Data Management, Kansas City, MO, USA, 23–26 May 2010; pp. 183–188.
7. Kumari, P.; Singh, Y. Delaunay triangulation coverage strategy for wireless sensor networks. In Proceedings of the 2012 International Conference on Computer Communication and Informatics, Coimbatore, India, 10–12 January 2012; pp. 1–5.
8. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306. [[CrossRef](#)]
9. Lv, C.; Wang, Q.; Yan, W.; Shen, Y. Diffusion Wavelet Basis Algorithm for Sparse Representation of Sensory Data in WSNs. *Signal Process.* **2017**, *140*, 12–31. [[CrossRef](#)]
10. Liu, S.; Zhang, Y.D.; Shan, T.; Tao, R. Structure-Aware Bayesian Compressive Sensing for Frequency-Hopping Spectrum Estimation with Missing Observations. *IEEE Trans. Signal Process.* **2018**, *66*, 2153–2166. [[CrossRef](#)]
11. Candès, E.; Recht, B. Exact Matrix Completion via Convex Optimization. *Found. Comput. Math.* **2009**, *9*, 717–772. [[CrossRef](#)]
12. Jin, K.H.; Ye, J.C. Annihilating Filter-Based Low-Rank Hankel Matrix Approach for Image Inpainting. *IEEE Trans. Image Process.* **2015**, *24*, 3498–3511.
13. Tremoulheac, B.; Dikaios, N.; Atkinson, D.; Arridge, S.R. Dynamic MR image reconstruction-separation from undersampled (k,t)-space via low-rank plus sparse prior. *IEEE Trans. Med. Imaging* **2014**, *33*, 1689–1701. [[CrossRef](#)]
14. Ramlathan, A.; Yang, M.; Liu, Q.; Li, M.; Wang, J.; Li, Y. A Survey of Matrix Completion Methods for Recommendation Systems. *Big Data Min. Anal.* **2018**, *1*, 308–323.

15. Cheng, J.; Jiang, H.; Ma, X.; Liu, L.; Liu, W. Efficient Data Collection with Sampling in WSNs: Making Use of Matrix Completion Techniques. In Proceedings of the 2010 IEEE Global Telecommunications Conference GLOBECOM 2010, Miami, FL, USA, 6–10 December 2010; pp. 1–5.
16. Cheng, J.; Ye, Q.; Jiang, H.; Wang, D.; Wang, C. STCDG: An Efficient Data Gathering Algorithm Based on Matrix Completion for Wireless Sensor Networks. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 850–861. [[CrossRef](#)]
17. He, J.; Sun, G.; Zhang, Y.; Wang, Z. Data Recovery in Wireless Sensor Networks with Joint Matrix Completion and Sparsity Constraints. *IEEE Commun. Lett.* **2015**, *19*, 2230–2233. [[CrossRef](#)]
18. He, J.; Sun, G.; Li, Z.; Zhang, Y. Compressive data gathering with low-rank constraints for Wireless Sensor networks. *Signal Process.* **2017**, *131*, 73–76. [[CrossRef](#)]
19. Kortas, M.; Habachi, O.; Bouallegue, A.; Meghdadi, V.; Ezzedine, T.; Cances, J. Energy Efficient Data Gathering Schema for Wireless Sensor Network: A Matrix Completion Based Approach. In Proceedings of the 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Croatia, 19–21 September 2019; pp. 1–6.
20. He, J.; Zhang, X.; Zhou, Y.; Maibvisira, M. A Subspace Approach to Sparse Sampling Based Data Gathering in Wireless Sensor Networks. *Sensors* **2020**, *20*, 985. [[CrossRef](#)] [[PubMed](#)]
21. Leinonen, M.; Codreanu, M.; Juntti, M. Sequential compressed sensing with progressive signal reconstruction in wireless sensor networks. *IEEE Trans. Wirel. Commun.* **2015**, *14*, 1622–1635. [[CrossRef](#)]
22. Xie, K.; Ning, X.; Wang, X.; Xie, D.; Cao, J.; Xie, G.; Wen, J. Recover Corrupted Data in Sensor Networks: A Matrix Completion Solution. *IEEE Trans. Mob. Comput.* **2017**, *16*, 1434–1448. [[CrossRef](#)]
23. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemometr. Intell. Lab. Syst.* **1987**, *2*, 37–52. [[CrossRef](#)]
24. Candès, E.J.; Li, X.; Ma, Y.; Wright, J. Robust principal component analysis? *J. ACM* **2011**, *58*, 11. [[CrossRef](#)]
25. Li, X.; Ding, S.; Li, Y. Outlier Suppression via Non-Convex Robust PCA for Efficient Localization in Wireless Sensor Networks. *IEEE Sens. J.* **2017**, *17*, 7053–7063. [[CrossRef](#)]
26. Chitradevi, N.; Palanisamy, V.; Baskaran, K.; Nisha, U.B. Outlier aware data aggregation in distributed wireless sensor network using robust principal component analysis. In Proceedings of the 2010 Second International Conference on Computing, Communication and Networking Technologies, Karur, India, 29–31 July 2010; pp. 1–9.
27. Zhang, Z.; Ganesh, A.; Liang, X.; Ma, Y. TILT: Transform Invariant Low-rank Textures. *Int. J. Comput. Vis.* **2012**, *99*, 1–24. [[CrossRef](#)]
28. Zhu, G.; Yan, S.; Ma, Y. Image tag refinement towards low-rank, content-tag prior and error sparsity. In Proceedings of the 18th ACM International Conference on Multimedia, Florence, Italy, 25–29 October 2010; pp. 461–470.
29. Min, K.; Zhang, Z.; Wright, J.; Ma, Y. Decomposing background topics from keywords by Principal Component Pursuit. In Proceedings of the 19th ACM Conference on Information and Knowledge Management, Toronto, ON, Canada, 26–30 October 2010; pp. 269–278.
30. Gu, S.; Zhang, L.; Zuo, W.; Feng, X. Weighted Nuclear Norm Minimization with Application to Image Denoising. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 2862–2869.
31. Sensor Data from Intel Berkeley Research Lab. Available online: <http://db.csail.mit.edu/labdata/labdata.html> (accessed on 25 December 2021).
32. Dong, W.; Liu, Y.; He, Y.; Zhu, T.; Chen, C. Measurement and Analysis on the Packet Delivery Performance in a Large-Scale Sensor Network. *IEEE/ACM Trans. Netw.* **2014**, *22*, 1952–1963. [[CrossRef](#)]
33. Parikh, N.; Boyd, S.P. Proximal Algorithms. *Found. Trends Optim.* **2014**, *1*, 127–239. [[CrossRef](#)]
34. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **2011**, *3*, 1–122. [[CrossRef](#)]