

Article

Correlated Spatio-Temporal Data Collection in Wireless Sensor Networks Based on Low Rank Matrix Approximation and Optimized Node Sampling

Xinglin Piao ¹, Yongli Hu ¹, Yanfeng Sun ^{1,*}, Baocai Yin ¹ and Junbin Gao ²

¹ Beijing Key Laboratory of Multimedia and Intelligent Software Technology, College of Metropolitan Transportation, Beijing University of Technology, Pingleyuan 100, Chaoyang District, Beijing 100124, China; E-Mails: piaoxinglin1987@gmail.com (X.P.); huyongli@bjut.edu.cn (Y.H.); ybc@bjut.edu.cn (B.Y.)

² School of Computing and Mathematics, Charles Sturt University, Bathurst, NSW 2795, Australia; E-Mail: jbgao@csu.edu.au

* Author to whom correspondence should be addressed; E-Mail: yfsun@bjut.edu.cn; Tel.: +86-10-6739-6568 (ext. 2108); Fax: +86-10-6739-6568 (ext. 2115).

External Editor: Leonhard M. Reindl

Received: 16 September 2014; in revised form: 21 November 2014 / Accepted: 26 November 2014 / Published: 5 December 2014

Abstract: The emerging low rank matrix approximation (LRMA) method provides an energy efficient scheme for data collection in wireless sensor networks (WSNs) by randomly sampling a subset of sensor nodes for data sensing. However, the existing LRMA based methods generally underutilize the spatial or temporal correlation of the sensing data, resulting in uneven energy consumption and thus shortening the network lifetime. In this paper, we propose a correlated spatio-temporal data collection method for WSNs based on LRMA. In the proposed method, both the temporal consistence and the spatial correlation of the sensing data are simultaneously integrated under a new LRMA model. Moreover, the network energy consumption issue is considered in the node sampling procedure. We use Gini index to measure both the spatial distribution of the selected nodes and the evenness of the network energy status, then formulate and resolve an optimization problem to achieve optimized node sampling. The proposed method is evaluated on both the simulated and real wireless networks and compared with state-of-the-art methods. The experimental results show the proposed method efficiently reduces the energy consumption of network and prolongs the network lifetime with high data recovery accuracy and good stability.

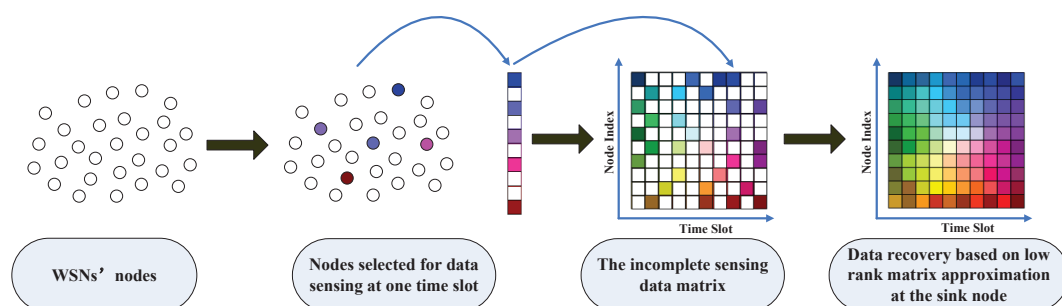
Keywords: wireless sensor networks; data collection; low rank matrix approximation.

1. Introduction

Data collection is a fundamental application of wireless sensor networks (WSNs). For example, in the environmental monitoring scenario, the physical quantities, such as temperature, humidity and light illumination, *etc.* are usually sensed and transmitted from sensor nodes to sink nodes through multi-hop routing [1]. Since the sensor nodes usually have limited computing ability and power supply, a primary goal of data collection is to obtain the sensing data at required accuracy with least energy consumption and thus to prolong the network lifetime.

Since the physical quantities describing natural phenomena are generally locally consistent and WSNs nodes are often deployed redundantly to get complete environmental monitoring data, the sensing data often have high spatial or temporal correlation [2]. Thus many researchers utilize the spatial and temporal correlation to reduce the network energy consumption by delivering part of sensing data or their aggregation, such as distributed source coding techniques [3–5], in-network collaborative wavelet transform methods [6,7] and clustered data aggregation methods [8–11]. Although these methods can save certain data communication cost, they would bring extra computational overheads for sensor nodes and so demand considerable sensing cost. Recently the low rank matrix approximation (LRMA) method [12,13] provides a new scheme for data collection in WSNs [14–16]. The main idea of the LRMA based method is that the sensing data of WSNs are regarded as a low rank matrix and recovered from a subset sensing data. As demonstrated in Figure 1, the data collection consists of two stages: the data sensing stage and the data recovering stage. At the data sensing stage, instead of using all sensor nodes, a subset of sensor nodes which are the shaded ones in the second subfigure is randomly selected to sense physical quantity and deliver the sensing data to the sink node at each data collection round. At the data recovering stage, the sink node receives these incomplete sensing data over some data collection rounds shown in the third subfigure in which the shaded entries represent the valid sensing data and the white entries are unknown. And then we could use them to recover the complete data by LRMA method. As the LRMA based data collection methods reduce either the data communication cost or the data sensing cost, high energy efficiency can be achieved even with an increased computation cost at the sink node.

Figure 1. The correlated spatio-temporal data collection method based on LRMA and optimized node sampling.



However, the current LRMA based data collection methods [14–16] also suffer from certain inefficiency both at the data sensing stage and the data recovering stage. At the data sensing stage, a random sampling scheme for the selection of sensing nodes is adopted in the current methods [14–16]. According to the LRMA theory [13], the uniformly random sampling method is an optimized sampling way to get good matrix complement result. However, the ideal random sampling method will result in uneven energy distribution of the network as the many-to-one data collection pattern, and thus shorten the network lifetime. At the data recovering stage, most of the current methods do not explore and utilize the spatial or temporal correlation of the sensing data. This could be the contributing factor to the certain setback of these methods, such as low data recovery accuracy, and lacking robustness to noises and outliers. Therefore, to improve the efficiency of the current LRMA based data collection methods, we always face two challenges. On the one hand, we should find out a proper way to formulate the relation between two important components in the data sensing procedure, the node sampling scheme and the scheme of estimating energy status of the network. Generally, the former has direct effect on the data recovery accuracy, and the later can determine the lifetime of the network. On the other hand, the main obstacle is how to represent the spatial and temporal relations of the incomplete node sensing data and further integrate these constraints to the current LRMA data recovery model. In this paper, we present a new data collection method based on LRMA, which improves the performance of the current method both at the data sensing and data recovering stages. At the data sensing stage, we propose to use Gini index to measure the spatial distributions of the selected sensing nodes and the evenness of energy status of the network. Under this measurement, the selection of sensing nodes is formulated as an optimization problem with the constraint of the network energy status. At the recovering stage, observing that the sensing data from a sensor generally has temporal consistency and the sensing data from different nodes has spatial correlation excepting for some noise and outliers. We give the representations of the spatial and temporal relationship of the node sensing data and add these constraints into the basic LRMA model and construct a correlated spatio-temporal constrained LRMA model (ST-LRMA). Then the ST-LRMA model is solved by an iteration method.

The proposed ST-LRMA based data collection method is tested on both simulated and real wireless networks compared with other relevant methods. The experimental results demonstrate the proposed method achieves better data recovery accuracy and improves energy efficiency to prolong network lifetime. Additionally, it is robust to noise and outliers.

The main contributions of this paper are:

- (1) We propose a correlated spatio-temporal constrained low rank matrix approximation model, a novel data collection scheme in WSNs;
- (2) We propose an optimized node sampling model for reducing energy consumption of WSNs, in which Gini Index is adopt and used as a tool to improve the current uniformly random sampling method.

The rest of the paper is organized as follows. In Section 2, we summarize the related work. Section 3 introduces the basic LRMA model based on random sampling. Section 4 presents the proposed ST-LRMA model. Section 5 elaborates the new optimized node sampling method. We will show the experimental results of our proposed methods compared with state-of-the-art methods, such as basic

LRMA model [14], T-LRMA model [15,16] and S-LRMA model in Section 6. Section 7 concludes the paper with a discussion on future research.

2. Related Works

In many data collection scenarios in WSNs, such as environmental monitoring [17], to fully sense physical quantities efficiently, it is necessary to transmit all sensing data to the sink node correctly and accurately. Doing so would consume too much energy in the networks. However, it is unnecessary to get the entire sensing data, given that there exists high correlation in the sensing data. Hence, the best strategy is to explore energy efficient data collection methods which approximately get the sensing data with a given accuracy.

The conventional approximate data collection methods include data aggregation techniques, distributed source coding and collaborative in-network compression methods. Data aggregation techniques [8–11] formulate the network sensor nodes into a hierarchical structure like trees or clusters. Generally the sensing nodes send the sensing data to the middle aggregator nodes. These aggregators collect data from multiple sensor nodes and aggregate the sensing data by aggregation functions, and then send the aggregated results to upper aggregators or sink nodes. Only a subset of nodes is involved in the communication with an aggregator and thus the overall energy consumption of the network during data collection is reduced. Distributed source coding schemes are based on the Slepian-Wolf coding theory [3–5]. In these methods, the sensor nodes are modeled as correlated sources of information and the dependency between the measured data are exploited to obtain data compression. Collaborative in-network compression methods can be viewed as an extension of traditional transforms in signal processing to sensor networks with irregular topology. Nodes can communicate with their neighbors and spatial correlation is exploited by collaborative transforms such as the distributed wavelet transform [6] and the graph wavelet transform [7].

Although the above mentioned conventional data collection methods are considered to be energy efficient and widely used in WSNs applications, they have their respective advantages in different aspects and the energy efficiency of these methods is relative to specific application scenarios. For example, the aggregation methods generally depend on a particular routing protocol, and the source coding methods typically require exact knowledge of the correlation between the measurements of different sensor nodes. It is highly desired to investigate the generic data collection method with high energy efficiency. The typical representatives are the recently proposed compressive sensing (CS) based methods [18–22] and the LRMA based methods [14–16].

The CS based methods provide two features of universal sampling and decentralized encoding, which make it a new paradigm for data collection in WSNs. In a single-hop network, a universal compressive wireless sensing method is proposed to deliver the sensing data by synchronized amplitude-modulated analog transmissions to the fusion center [23]. The sparsity of the sensing data and the CS decoding algorithms are discussed for data collection application in [24]. The first complete scheme to apply compressive sensing theory to data collection in large scale sensor networks is presented in [18]. Then an adaptive compressive sensing based data collection method is proposed in [19]. By maximizing information gain per energy cost during each measurement, the method adaptively collects data in an

energy efficient way. However, during each measurement, the projected vector is obtained by solving an NP-hard optimization problem, which brings in relatively high computational and communication overhead. Energy efficiency of applying CS to data collection was investigated in [20], aiming at minimizing energy consumption through combining compressed aggregation with routing protocol. The optimization problem is proved to be of NP-completeness and both optimized and nearly-optimized solutions are given. It is proved that CS based data collection methods reduce the communication cost with high efficiency, but the data decoding usually involves high computation and the energy efficiency generally predominates in large scale WSNs. In addition, during the process of data collection, it generally demands complete sampling for all nodes in each round. Once there exist data missing, error or outliers, the result of data reconstruction would be greatly compromised.

As an efficient data recovery method, the LRMA model is widely used in many scenarios, and especially the fast algorithm for the solution of low rank minimal optimization was proposed in [25]. A CS method combined with low rank approximation was proposed to interpolate the Internet traffic matrices, in which the spatio-temporal properties of the matrices were considered [26]. The low rank completion method was adopted to reconstruct the fingerprint from a small subset of training samples [27]. The spatial correlation structure was combined with low rank completion for Received Signal Strength Indication (RSSI) recovery [28]. The basic LRMA model is first introduced to reduce energy consumption in data collection in [14]. Compared with the globally random sampling of the CS based data collection method which realizes energy efficiency by reducing the communication cost, the random sampling scheme of the LRMA based method reduces both the communication and sensing cost at nodes, which is the main virtue of the LRMA based method. However, in the basic LRMA model, the spatial or temporal correlation of the sensing data is not exploited. To obtain better data recovery results, a temporally constrained LRMA model for data collection was proposed in [15]. The similar idea was shared by our previous work [16]. However, the spatial properties of the sensing data are not well explored in any currently existing methods. In this paper, we consider both the temporal and spatial correlation of the sensing data simultaneously to propose an improved LRMA based data collection method. Furthermore, the issues of optimizing sensing node selection and balancing network energy distribution are also discussed.

3. The Basic LRMA Based Data Collection Method

In this section, we formulate the basic LRMA model for data collection in WSNs from the low rank theory [13]. For a wireless network with N sensor nodes, the T rounds data collection will form a matrix of measurements $\mathbf{X} \in R^{N \times T}$. As shown in Figure 1, at each data collection round, the LRMA based method randomly selects a subset of sensor nodes for data sensing. Thus the measurement matrix is generally incomplete, *i.e.*, only the positions corresponding to the selected nodes offer valid sensing data. Here we adopt a mask operator $\mathcal{A}(\cdot)$ [13] to represent the subset sampling procedure:

$$\mathcal{A}(\mathbf{X}) = \mathbf{M}. \quad (1)$$

where \mathbf{M} is an incomplete matrix with only a sparse set of elements with valid sensing values at relevant positions. For the sake of clarity, the operator $\mathcal{A}(\cdot)$ can be specified as an element-wise matrix product as follows [13]:

$$\mathcal{A}(\mathbf{X}) = \mathbf{Q} \odot \mathbf{X}. \quad (2)$$

where \odot denotes the Hadamard product of two matrices *i.e.*, $\mathbf{M}(i, j) = \mathbf{Q}(i, j)\mathbf{X}(i, j)$. \mathbf{Q} is a $N \times T$ matrix defined as by the following form [14]:

$$\mathbf{Q}(i, j) = \begin{cases} 1, & \text{if the } i\text{th node has valid sensing data at the } j\text{th time slot.} \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

Generally, the measurements from one node are continuous in time domain, which shows temporal correlation. On the other hand, the sensing data from adjacent nodes have similar values or the sensing data of a node can be represented by the measurements of its neighboring nodes, which shows spatial correlation. These spatio-temporal correlations bring the low rank property of the measurement matrix \mathbf{X} . According to the low rank matrix completion theory [12,13], it is highly possible to recover a low-rank matrix from a subset of its entries, generally a random subset with enough elements. Thus, we can recover \mathbf{X} by solving the following optimization problem:

$$\begin{aligned} \mathbf{X}^* = \underset{\mathbf{X}}{\operatorname{argmin}} \quad & \operatorname{rank}(\mathbf{X}), \\ \text{subject to} \quad & \mathcal{A}(\mathbf{X}) = \mathbf{M}. \end{aligned} \quad (4)$$

This is the basic LRMA (B-LRMA) model [12,13] for data collection. In the representation of B-LRMA model, there is no constraint on the rows or columns continuity of \mathbf{X} . That is, the intrinsic properties of the measurements, the spatial and temporal correlations, are not represented in this model. So to get better recovery result of \mathbf{X} , we propose a correlated spatio-temporal LRMA model. It will be described in detail in the next section.

4. The ST-LRMA Based Data Collection Method

The sequential measurements from a sensor, a row of the measurement matrix, are correlative in time domain and generally behave consistently and smoothly except for some noise and outliers, so it is natural to require the recovered matrix to preserve this property. Based on this observation, we share the similar idea of [15,16] and introduce a temporal constraint into the B-LRMA model to build a temporal constrained LRMA model (T-LRMA) for maintaining the consistency and smoothness in row direction, the adjacent columns of the recovered measurements. We formulate the T-LRMA model as follows [15,16]:

$$\begin{aligned} \mathbf{X}^* = \underset{\mathbf{X}}{\operatorname{argmin}} \quad & \operatorname{rank}(\mathbf{X}) + \lambda_1 \|\mathbf{X}\mathbf{D}\|_F^2, \\ \text{subject to} \quad & \mathcal{A}(\mathbf{X}) = \mathbf{M}. \end{aligned} \quad (5)$$

where λ_1 is a tunable parameter, and \mathbf{D} is a matrix with the following form [15,16]:

$$\mathbf{D} = \begin{pmatrix} -1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & -1 \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}_{T \times (T-1)}. \quad (6)$$

$\|\mathbf{XD}\|_F^2$ represents Frobenius norm of difference matrix of \mathbf{X} in row direction, and thus enforces the consistency of the columns of the recovered \mathbf{X} .

In practice, the sequential consistency is a ubiquitous property for many physical data and its sparse representation is well studied by many researchers. A fused LASSO method was proposed for problems with features that can be ordered in sequence [29]. In this method, the sparsity of regression coefficients and the smoothness constraint of the successive coefficients are considered simultaneously. A spatial subspace clustering method (SpatSC), was proposed for classifying the drill pipe spectral data [30]. In this method, each individual regression coefficient vector of the prototypic data is demanded to be sparse and considered to be similar to neighboring. An ordered subspace clustering (OSC) method was proposed to segment data drawn from a sequentially ordered union of subspaces was proposed in [31]. In this method, each sequential nature of sequential data is incorporated by its neighbor penalty term to enforce similarity. The idea of preserving sequential consistency is shared by these methods and our method, but we apply the sequential constraint on the recovered data instead of the sparse coefficients.

Except for the temporal correlation of measurements from one node, there also exists spatial correlation among the measurements from different sensor nodes. The fundamental argument for this point is that the physical quantities from natural environment generally have smooth distribution in spatial domain, for example, the temperature distribution in a monitoring area. Another reason for favoring the spatial correlation is that the sensor nodes are generally deployed redundantly, sometimes densely, in order to obtain complete sensing data, which brings the spatial correlation among the measurements, especially for adjacent nodes. The closer nodes usually have similar measurements. To describe this property of the measurements, we incorporate the spatial constraints into the B-LRMA model and propose a spatial constrained LRMA model (S-LRMA) as follows:

$$\begin{aligned} \mathbf{X}^* = \underset{\mathbf{X}}{\operatorname{argmin}} \quad & \operatorname{rank}(\mathbf{X}) + \lambda_2 \|\mathbf{SX}\|_F^2, \\ \text{subject to} \quad & \mathcal{A}(\mathbf{X}) = \mathbf{M}. \end{aligned} \quad (7)$$

where λ_2 is a tunable parameter, and \mathbf{S} is a matrix representing the relation among the measurements from different nodes, the rows of the measurement matrix \mathbf{X} .

It is critical to design a proper \mathbf{S} to obtain good data recovery results. Generally, the spatial correlation of the sensing data depends on the spatial deployment of the nodes. Therefore, it is a straight forward way to use the spatial relation of the nodes to model the spatial correlation of the sensing data. In the ideal case, such as the simulated network in Section 6, the measurement of a node can be represented as the linear combination of its neighbors according to its adjacent relation. Thus, we can construct \mathbf{S} from

the combination coefficients directly. But in practice, the accurate positions of sensor nodes are usually not available. Additionally, in a complex sensing scenario, the variation of the physical quantities is not smoothly distributed in the spatial domain and the measurements are often interfered by the complicated environment. Hence, in most cases, a feasible way inferring the spatial correlation of the sensing data is to utilize the sensing data themselves.

Although the traditional clustering methods, such as KNN, can be used to obtain the relation of a set of data in Euclid Space, it is difficult to get good result for high-dimension sensing data. Given that manifold learning methods [32–34] have good performance in learning the hidden structure in high dimension data, we adopt the manifold clustering method, namely sparse manifold clustering (SMC) method [35], to get the spatial correlation of sensing data. Conveniently, we implement the SMC model for sensing data spatial clustering as follows.

Similar to the ideal case in the simulated network, the measurement of a node can be represented as the linear combination of the measurements of correlated nodes, *i.e.*, for a row $\mathbf{X}_i, i = 1, \dots, N$ of \mathbf{X} , there exist K most correlated rows $\mathbf{X}_{j_k}, k = 1, \dots, K$ s.t. $\mathbf{X}_i = \sum_{k=1}^K w_{j_k}^i \mathbf{X}_{j_k}$, where $j_k \neq i$. So the problem of constructing \mathbf{S} is to find the suitable weights $w_{j_k}^i, k = 1, \dots, K$ for each $\mathbf{X}_i, i = 1, \dots, N$, which represent the intrinsic structure among the high dimension signals, the rows of the sensing data. According to SMC, for each $\mathbf{X}_i, i = 1, \dots, N$, we firstly construct a new signal matrix \mathbf{Y}^i from the measurement matrix \mathbf{X} as the following form:

$$\mathbf{Y}^i = [\mathbf{Y}_1^i, \dots, \mathbf{Y}_{i-1}^i, \mathbf{Y}_{i+1}^i, \dots, \mathbf{Y}_N^i]. \quad (8)$$

where $\mathbf{Y}_j^i = \frac{\mathbf{x}_j^T - \mathbf{x}_i^T}{\|\mathbf{x}_j^T - \mathbf{x}_i^T\|_2}, j = 1, \dots, N$ and $j \neq i$. Then we construct a diagonal distance matrix \mathbf{Z}^i with elements $\frac{\|\mathbf{x}_j^T - \mathbf{x}_i^T\|_2}{\sum_{j \neq i} \|\mathbf{x}_j^T - \mathbf{x}_i^T\|_2}, j = 1, \dots, N$ and $j \neq i$. Therefore, we can get an optimized sparse coefficient \mathbf{c}_i from the following optimization problem:

$$\min \quad \alpha \|\mathbf{Z}_i \mathbf{c}_i\|_1 + \frac{1}{2} \|\mathbf{Y}_i \mathbf{c}_i\|_2^2, \quad \text{s.t.} \quad \mathbf{1}^T \mathbf{c}_i = 1. \quad (9)$$

From the optimized \mathbf{c}_i , we can get the spatial correlation weights $w_{j_k}^i, k = 1, \dots, K$ by omitting some small elements less than the given threshold ε :

$$w_{j_k}^i = \frac{c_{ij_k} / \|\mathbf{x}_{j_k}^T - \mathbf{x}_i^T\|_2}{\sum_{\substack{j \neq i \\ |c_{ij}| > \varepsilon}} (c_{ij} / \|\mathbf{x}_j^T - \mathbf{x}_i^T\|_2)}. \quad (10)$$

Based on the spatial correlation weights $w_{j_k}^i, k = 1, \dots, K$, the spatial correlation matrix \mathbf{S} can be formulated by the following element representation:

$$\mathbf{S}(i, j) = \begin{cases} 1, & \text{if } i = j. \\ -w_{j_k}^i, & \text{if } j \neq i \text{ and } j = j_k^i, k = 1, \dots, K. \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

One should note that \mathbf{S} constructed in such a way is based on the complete measurement matrix \mathbf{X} . However, at the very beginning, we could only obtain the incomplete sensing data. Therefore, prior to recovering the sensing data, we could estimate the unknown measurements by some simple interpolation algorithms or the B-LRMA. From the estimated complete sensing data, we could construct

the spatial correlation matrix \mathbf{S} . According to our experimental results, the T-LRMA data recovery model in Equation (5) has better estimated accuracy than other algorithms. So, we use the optimized result of the T-LRMA model as the estimation of the unknown sensing data to construct \mathbf{S} .

Having determined the temporal and spatial constraints \mathbf{D} and \mathbf{S} , we integrate the low rank approximation models in Equations (5) and (7) into a combined model to form a novel correlated spatio-temporal constrained model, the ST-LRMA model:

$$\begin{aligned} \mathbf{X}^* = \underset{\mathbf{X}}{\operatorname{argmin}} \quad & \operatorname{rank}(\mathbf{X}) + \lambda_1 \|\mathbf{X}\mathbf{D}\|_F^2 + \lambda_2 \|\mathbf{S}\mathbf{X}\|_F^2, \\ \text{subject to} \quad & \mathcal{A}(\mathbf{X}) = \mathbf{M}. \end{aligned} \quad (12)$$

In general, a rank minimization problem like Equation (12) is NP-hard [12]. The common practice in solving Equation (12) is to replace the rank function with the so-called matrix nuclear norm $\|\cdot\|_*$ which is defined as the sum of singular values of \mathbf{X} . Thus Equation (12) can be converted into a convex optimization problem [12]. Leveraging the computation and memory resource in the sink node and the data recovering delay for large scale WSNs (N nodes), we adopt the approximate algorithm in [12], which is considered to have good efficiency with a relative small memory request. This method seeks for a solution with a fixed low rank r . The process can be described as follows. Suppose that the rank of targeted solution \mathbf{X} is r . Then \mathbf{X} admits a skinny SVD $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where \mathbf{U} is an $N \times r$ unitary matrix, \mathbf{V} is a $T \times r$ unitary matrix and $\mathbf{\Sigma}$ is an $r \times r$ diagonal matrix containing all the singular values σ_k , $k = 1, 2, \dots, r$, which are arranged in a decreasing order. We further write as $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \mathbf{L}\mathbf{R}^T$, where $\mathbf{L} = \mathbf{U}\mathbf{\Sigma}^{1/2}$ and $\mathbf{R} = \mathbf{V}\mathbf{\Sigma}^{1/2}$. So the model in Equation (12) can be replaced by the following constrained minimization model:

$$\begin{aligned} (\mathbf{L}^*, \mathbf{R}^*) = \underset{\mathbf{L}, \mathbf{R}}{\operatorname{argmin}} \quad & \operatorname{rank}(\mathbf{L}\mathbf{R}^T) + \lambda_1 \|\mathbf{L}\mathbf{R}^T\mathbf{D}\|_F^2 + \lambda_2 \|\mathbf{S}\mathbf{L}\mathbf{R}^T\|_F^2, \\ \text{subject to} \quad & \mathcal{A}(\mathbf{L}\mathbf{R}^T) = \mathbf{M}. \end{aligned} \quad (13)$$

The dimension of \mathbf{L} and \mathbf{R} are $N \times r$ and $T \times r$, respectively. In practice, we can specify an estimate to the rank r of \mathbf{X} , which could be larger than the actual rank of the targeted \mathbf{X} . From the lemma in [12], if the restricted isometry property holds on $\mathcal{A}(\mathbf{X})$ and $\operatorname{rank}(\mathbf{X}) < \operatorname{rank}(\mathbf{L}\mathbf{R}^T)$, then Equation (13) is equivalent to the following model:

$$\begin{aligned} (\mathbf{L}^*, \mathbf{R}^*) = \underset{\mathbf{L}, \mathbf{R}}{\operatorname{argmin}} \quad & \|\mathbf{L}\|_F^2 + \|\mathbf{R}^T\|_F^2 + \lambda_1 \|\mathbf{L}\mathbf{R}^T\mathbf{D}\|_F^2 + \lambda_2 \|\mathbf{S}\mathbf{L}\mathbf{R}^T\|_F^2, \\ \text{subject to} \quad & \mathcal{A}(\mathbf{L}\mathbf{R}^T) = \mathbf{M}. \end{aligned} \quad (14)$$

By the method of Lagrange multipliers, the constrained optimization problem Equation (14) can be revised as the following unconstrained model:

$$(\mathbf{L}^*, \mathbf{R}^*) = \underset{\mathbf{L}, \mathbf{R}}{\operatorname{argmin}} \|\mathbf{L}\|_F^2 + \|\mathbf{R}^T\|_F^2 + \lambda_1 \|\mathbf{L}\mathbf{R}^T\mathbf{D}\|_F^2 + \lambda_2 \|\mathbf{S}\mathbf{L}\mathbf{R}^T\|_F^2 + \lambda_3 \|\mathcal{A}(\mathbf{L}\mathbf{R}^T) - \mathbf{M}\|_F^2. \quad (15)$$

where $\|\mathcal{A}(\mathbf{L}\mathbf{R}^T) - \mathbf{M}\|_F^2$ represents the reconstruction error at the sampling subset \mathbf{M} with an adjusting weight λ_3 . This is the final ST-LRMA model for data collection.

To solve problem Equation (15), we adopt an alternative iteration algorithm for \mathbf{L} and \mathbf{R} . Firstly, \mathbf{L} is initialized randomly. Then optimize \mathbf{R} by a linear least square method with fixing \mathbf{L} . Note that the operator $\mathcal{A}(\cdot)$ confines the equation group onto the subset of the valid entries in \mathbf{M} , so we use a subset

of the equations in the linear least square aggression instead of the complete equation group. After \mathbf{R} has been updated, we fix \mathbf{R} and alternatively optimize \mathbf{L} with respect to the new objective function. Repeat the above alternative iteration procedure until meeting the convergence condition or exceeding the maximal iteration number.

5. Optimized Node Sampling for Energy Efficient Data Collection

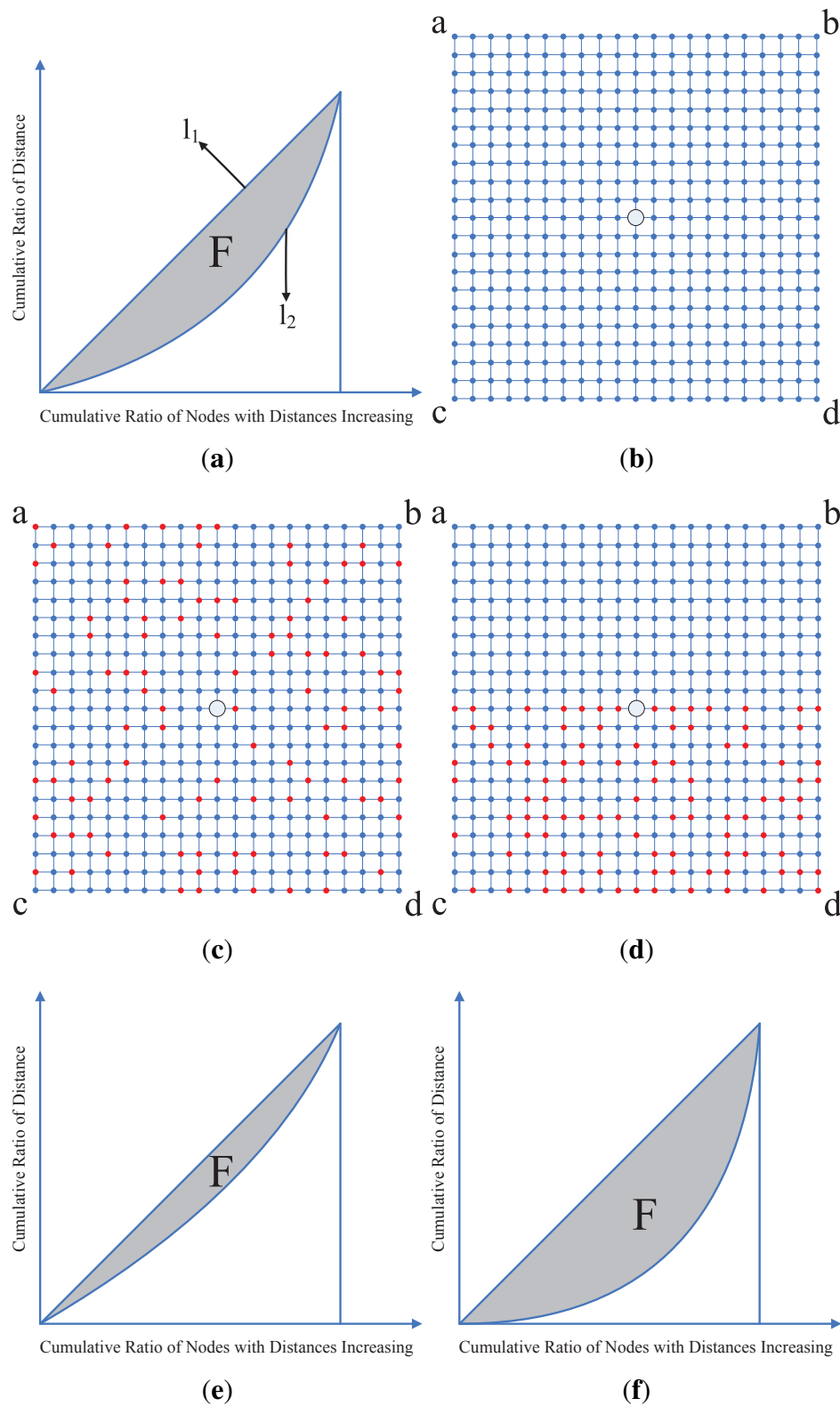
The LRMA based data collection depends on the random sampling measurements \mathbf{M} . From the low rank completion theory [13], to get good recovery accuracy, the elements in \mathbf{M} should better have a uniform distribution. Therefore, the selected nodes for sensing data should have uniform distribution in the deployed area, which is also the request for completely sensing the physical phenomena. However, as the many-to-one data collection pattern, this uniform node sampling method will bring unbalance energy distribution of the network and shorten its lifetime. To realize an energy efficient node selection at each data collection round, we propose an optimized node sampling method considering both the uniform random sampling and the network energy balance. In this method, we use Gini Index to measure the evenness of a random node sampling scheme and the network energy status and then find the optimized node sampling solution to meet the two-fold requests.

Gini index is often used for evaluating data inequality according to frequency distributions. There is a demonstration of Gini Index shown in Figure 2a, where l_1 represents the perfect equal cumulative ratio, l_2 represents the actual cumulative ratio, and the area of shaded part F represents the Gini Index. If l_2 got closer to l_1 , the area of F , the Gini Index, will get smaller. In our method, given the sampling nodes number n (determined by the sampling rate) of one data collection round, the node sampling scheme can be denoted by a binary vector $\mathbf{q} = (q_1, \dots, q_N)^T$, which is a column of \mathbf{Q} in Equation (2), where $q_i = 1$ representing the i th node is selected for sensing data and $\sum_{i=1}^N q_i = n$. For each unselected node in the network, we denote the distance of the node to its nearest selected node by d_j . Then we sort the distances and get an ascending sequence $d_1 \leq d_2 \leq \dots \leq d_{N-n}$. From the sequence, we define the evenness measurement for the desired random sampling scheme by Gini Index in the following form:

$$G_1(\mathbf{q}) = 1 - \frac{1}{N-n} \left(2 \sum_{j=1}^{N-n} d'_j - 1 \right). \quad (16)$$

where $d'_j = \frac{\sum_{k=1}^j d_k}{\sum_{i=1}^{N-n} d_i}$ is the cumulative ratio of distances. By minimization of $G_1(\mathbf{q})$, we obtain the optimized \mathbf{q} representing the best spatial uniform node selection scheme. As the distances among nodes are unknown for a practical network, we use the spatial correlation weights in Equation (10) to replace the distances d_j and compute the $G_1(q)$ in Equation (16). We give a concrete example. As shown in Figure 2. Figure 2b shows a simulated grid network with 21×21 node, and Figure 2c,d show two node sampling schemes in the simulated grid network, where the red nodes represent the nodes selected for sensing data. Figure 2e and Figure 2f show their Gini Index of Figure 2c and Figure 2d respectively. Apparently, the area of F in Figure 2e is much smaller than the one in Figure 2f as the distribution of the selected nodes in Figure 2c is more uniform than it in Figure 2d. So from the measurement of Gini Index, we could obtain the optimized nodes sampling scheme.

Figure 2. Two nodes sampling schemes in simulative networks and their Gini Index. (a) A demonstration of Gini Index; (b) A grid simulated network; (c) A uniformly sampling scheme in simulated networks; (d) A non-uniform sampling scheme in simulated networks; (e) The Gini Index of the scheme in Figure 2c; (f) The Gini Index of the scheme in Figure 2d.



Implementing the data collection round given the node sampling scheme \mathbf{q} , the energy consumption in the data sensing and delivering will result in a new energy status for the network. We denote the remaining energy of each node by e_i , $i = 1, \dots, N$. They also are sorted as an ascending sequence $e_1 \leq e_2 \leq \dots \leq e_N$. Similarly, we define the measurements for the network energy balance by Gini Index in the following form:

$$G_2(\mathbf{q}) = 1 - \frac{1}{N} \left(2 \sum_{i=1}^N e'_i - 1 \right). \quad (17)$$

where $e'_i = \frac{\sum_{k=1}^i e_k}{\sum_{j=1}^N e_j}$.

From Equations (16) and (17), we get a multi-objective 0-1 programming model to search the energy efficient node sampling \mathbf{q} as the following form:

$$\min \quad Z = \{G_1(\mathbf{q}), G_2(\mathbf{q})\}, \quad \text{s.t.} \quad \mathbf{1}^T \mathbf{q} = n. \quad (18)$$

This is a non-linear optimization problem. Considering the computation restriction in WSNs, we use the intelligent optimization method, Particle Swarm Optimization (PSO) [36], to solve this optimization model. Once obtaining the optimized node sampling scheme \mathbf{q} , we can implement the current data collection round and so repeat the procedure getting continuous sensing data for low rank recovery at the sink node.

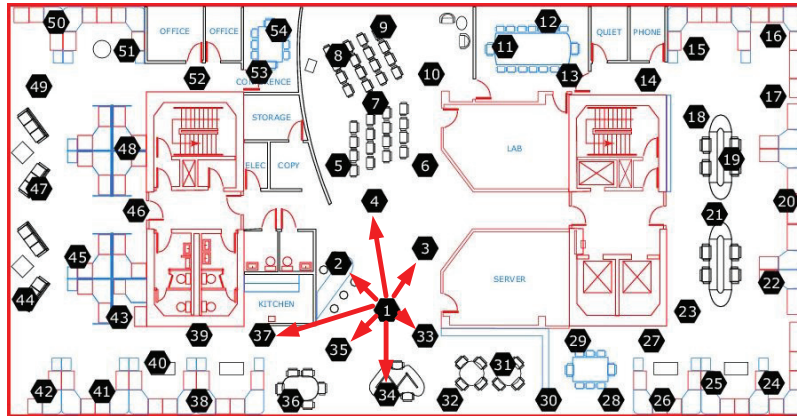
6. Experiments and Results

To evaluate the proposed method, we conduct data collection and data recovery experiments on both a simulated network and a real wireless sensor network and compare our method with other relevant methods. In this section, we first give the experiments setup and then report the experimental results.

6.1. Experiment Environments and Parameter Setting

6.1.1. The Structure of the Simulated and Real Networks

The simulated network is a regular 21×21 grid network, and the center node is set as the sink node, shown in Figure 2a. In the simulated network, the physical quantities can be accurately sensed and compared with the recovered data, so the proposed method can be evaluated confidently. The real data collection scenario is the wireless network of Intel Berkeley Research Lab [37], which has 54 nodes and the 5th node is the sink node, shown in Figure 3. The sensing data of the real network include several physical quantities, such as temperature, humidity and illumination, *etc.* The data were collected over several months with a sensing time slot of half minute.

Figure 3. The wireless network of Intel Berkeley Research Lab [37].

6.1.2. The MAC Protocol and Routing of the Simulated and Real Networks

To the regular grid of the simulated network, the experiments are implemented on a scheduled medium access control (MAC). Specifically, IEEE 802.15.4 is adopted as the MAC protocol. The data transmission rate is set 250 kbps. To the nodes irregularly deployed in the real network, the experiments are constructed with 433 MHz mica2dot nodes which adopt the default CSMA based B-MAC as the MAC layer protocol.

To model the sensing data delivering procedure, we adopt a simple routing method to transport a sensing packet from a node to the sink node, in which a node selects the one-hop adjacent node with maximal remaining energy to transmit the sensing packet. In the simulated network, the one-hop adjacent nodes of a node are the directly linked nodes in the grid. In the real networks, we set a maximal distance for a node to transmit data package to others in one-hop. In our experiments, the upper bound is assigned to 10 m. Under this setting, we obtain the connection of the nodes in the real network. For example, the one-hop links from node 1 are shown as the arrowed lines in Figure 3.

6.1.3. The Energy Consumption Model of the Simulated and Real Networks

To evaluate the energy consumption of the network, we adopt the energy consumption model in [38] in our experiments, in which the node transmitting and reception energy consumption $E_{Tx}(d)$ and E_{Rx} are defined as follows:

$$E_{Tx}(d) = E_l + \epsilon_a d^2. \quad (19)$$

$$E_{Rx} = E_l. \quad (20)$$

where d is the transmitting distance, E_l is the energy consumed of radio electronics, ϵ_a is the power amplifier. In our experiments, all nodes are initialized with $2 \times 10^4 J$ of energy and the parameters $E_l = 50 nJ/bit$, $\epsilon_a = 10 pJ/bit/m^2$. The sensing packet is assumed to be a fixed size of 1000 bits.

Based on the above energy consumption model, given a node sampling rate, the network lifetime is defined as the maximal data collection rounds if the network energy can ensure the sensing data being transferred to the sink node.

6.1.4. Data Recovery Parameters Setup

In the simulated network, to simulate the change of a physical quantity, such as temperature, we use four curves, shown as Figure 4, to represent the sensing data during a period of time corresponding to the four corner nodes in Figure 2a. The sensing data of other nodes are then generated by the bilinear interpolation method. A snap of the simulated measurements of the 21×21 grid is shown in Figure 5a. These clean sensing data are regarded as the ground truth of the physical quantity. However, the real sensing data usually have some noise, so we add Gaussian noise into the clean sensing data to simulate actual measurement data, shown as Figure 5b. To form the measurement matrix \mathbf{X} for data recovery, here \mathbf{X} is set as a square matrix with dimension of $(21 \times 21 - 1) \times 440$, *i.e.*, $N = T = 440$. The parameters in Equations (9), (10) and (15) are empirically set as: $\lambda_1 = \lambda_2 = 0.01$, $\lambda_3 = 10$, $\alpha = 0.1$, $\varepsilon = 10^{-5}$.

In the real network, the temperature sensing data collected on 53 nodes at March 1st 2004 are used as the measurement data which has dimension of 53×2880 . As there exist large difference for the dimensions of the row and column, we segment the measurement data into 48 blocks, where each block has same dimension of 53×60 . One sample of these block is shown in Figure 6a. In the data recovery experiments, each block is assigned as the measurements matrix \mathbf{X} , *i.e.*, $N = 53$, $T = 60$. The parameters in Equations (9), (10) and (15) are set as: $\lambda_1 = \lambda_2 = 0.01$, $\lambda_3 = 10$, $\alpha = 0.1$, $\varepsilon = 10^{-5}$. The whole recovered data can be obtained by stacking all the recovered blocks.

Figure 4. The simulated sensing data of four corner nodes **a**, **b**, **c**, **d** of the simulated network in Figure 2a.

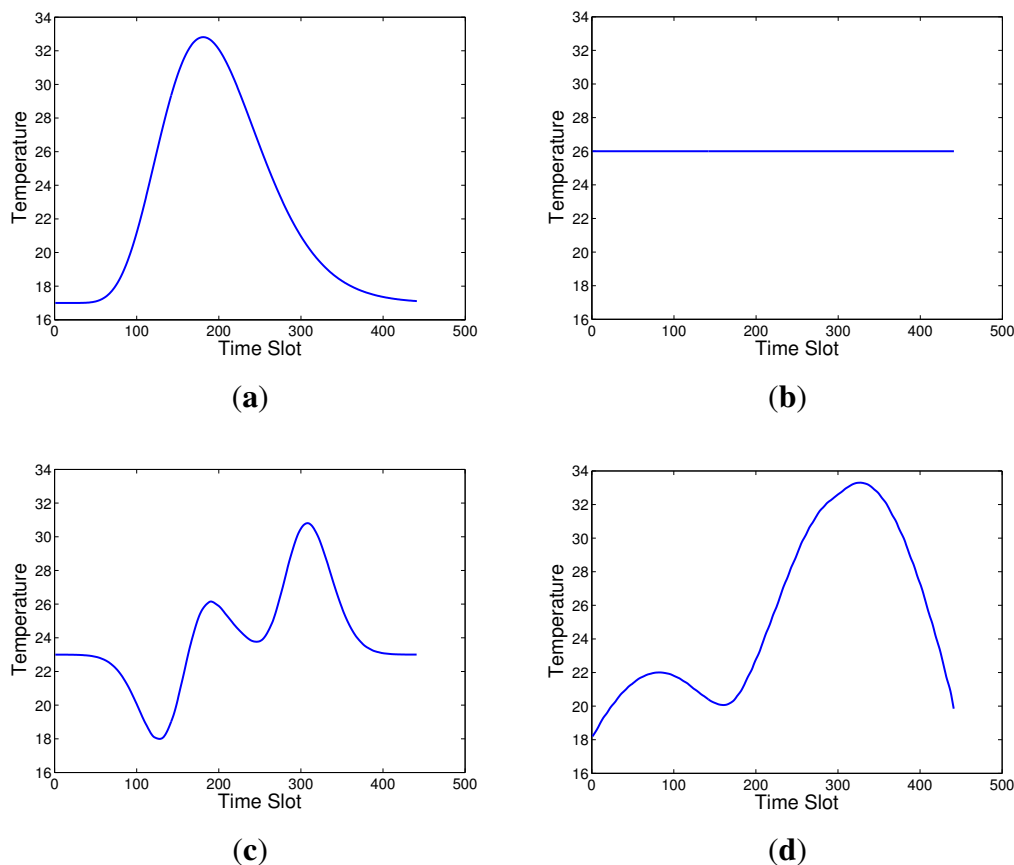


Figure 5. A snap of the sensing data of the simulated network. (a) The clean sensing data; (b) The sensing data with noise.

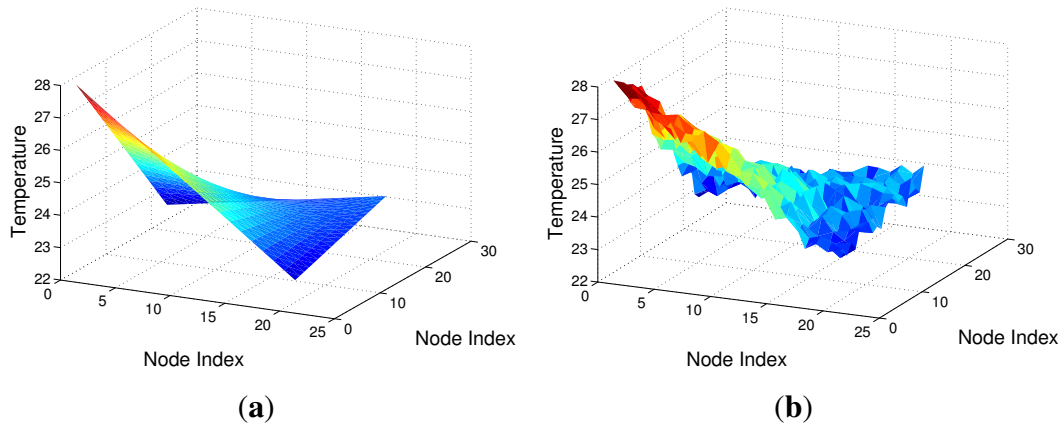
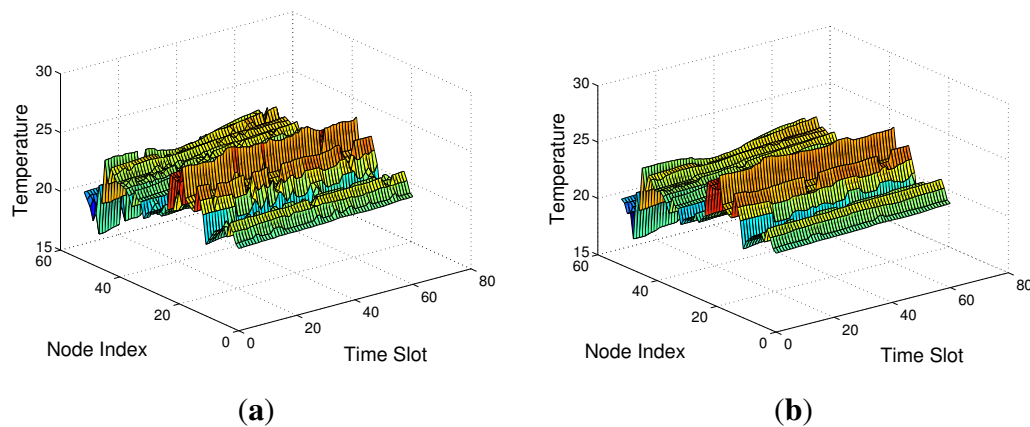


Figure 6. The temperature sensing data from Intel Berkeley Research Lab: (a) The actual sensing data; (b) The filtered sensing data.



To evaluate the accuracy of the recovered measurement, the recovery error is defined as the average relative error between the recovered data and the clean sensing data (denoted by \mathbf{X}^0) at these unknown element with the following representation:

$$err = \frac{\sqrt{\sum_{(i,j) \in \{(i,j) | \mathbf{Q}(i,j)=0\}} (\mathbf{X}^*(i,j) - \mathbf{X}^0(i,j))^2}}{\sqrt{\sum_{(i,j) \in \{(i,j) | \mathbf{Q}(i,j)=0\}} (\mathbf{X}^0(i,j))^2}}. \quad (21)$$

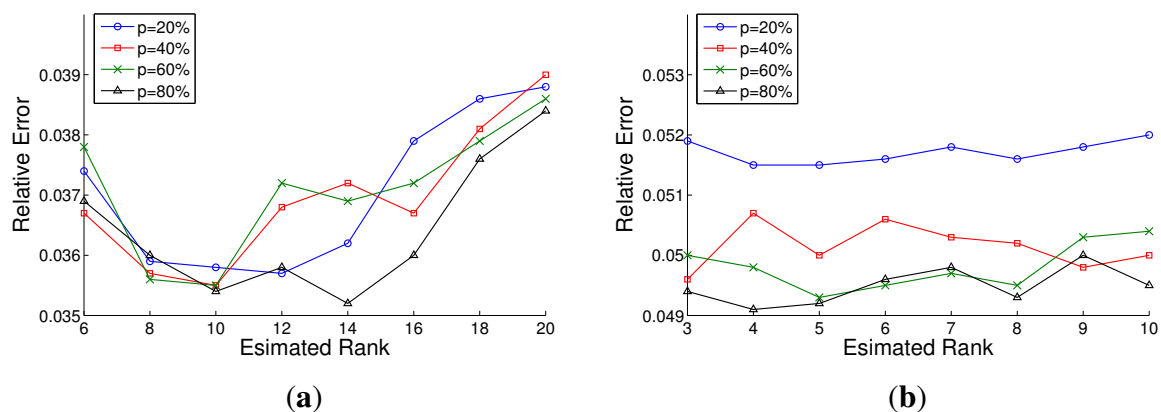
In real networks experiments, we need real clean sensing data to verify the validity of our propose ST-LRMA method. Unfortunately, the actual sampling sensing data are always corrupted and some values are gone missing. The real and accurate temperature data curve should be smooth and continuous, however, the actual temperature sensing data sampled over real networks are always noisy due to one reason or another. Hence, we simply use the Median filter on the temporal direction of the measurements as the clean sensing data to compute the error, *i.e.*, a 1D Median filter with dimension of 5 is used on each

row of \mathbf{X} . The filtered results of the measurements in Figure 6a are shown in Figure 6b. Additionally, for the 48 segmented blocks in the real network, the recovery error is defined as the mean of the recovery errors of the 48 blocks.

To verify the performance of the proposed method on different node sampling rate, defined as the ratio of the number of the valid elements of \mathbf{M} to the number of the complete elements of the measurement \mathbf{X} , the data recovery experiments are conducted with the node sampling rate changing from 5% to 95%.

The rank of the measurements matrix is an important parameter for the LRMA method. However, estimation of the rank from an incomplete matrix is an open problem. To get proper rank estimation, we implement many times data recovery experiments with different rank. In addition, the data recovery results are computed by the above metric. The results of simulative and real networks are shown in Figure 7. According to the results, we choose different rank for the data recovery. In the simulated networks, the appropriate rank is assigned to 10 according to the experimental results shown in Figure 7a. In real network, the appropriate rank is assigned to 5 according to the experimental results shown in Figure 7b.

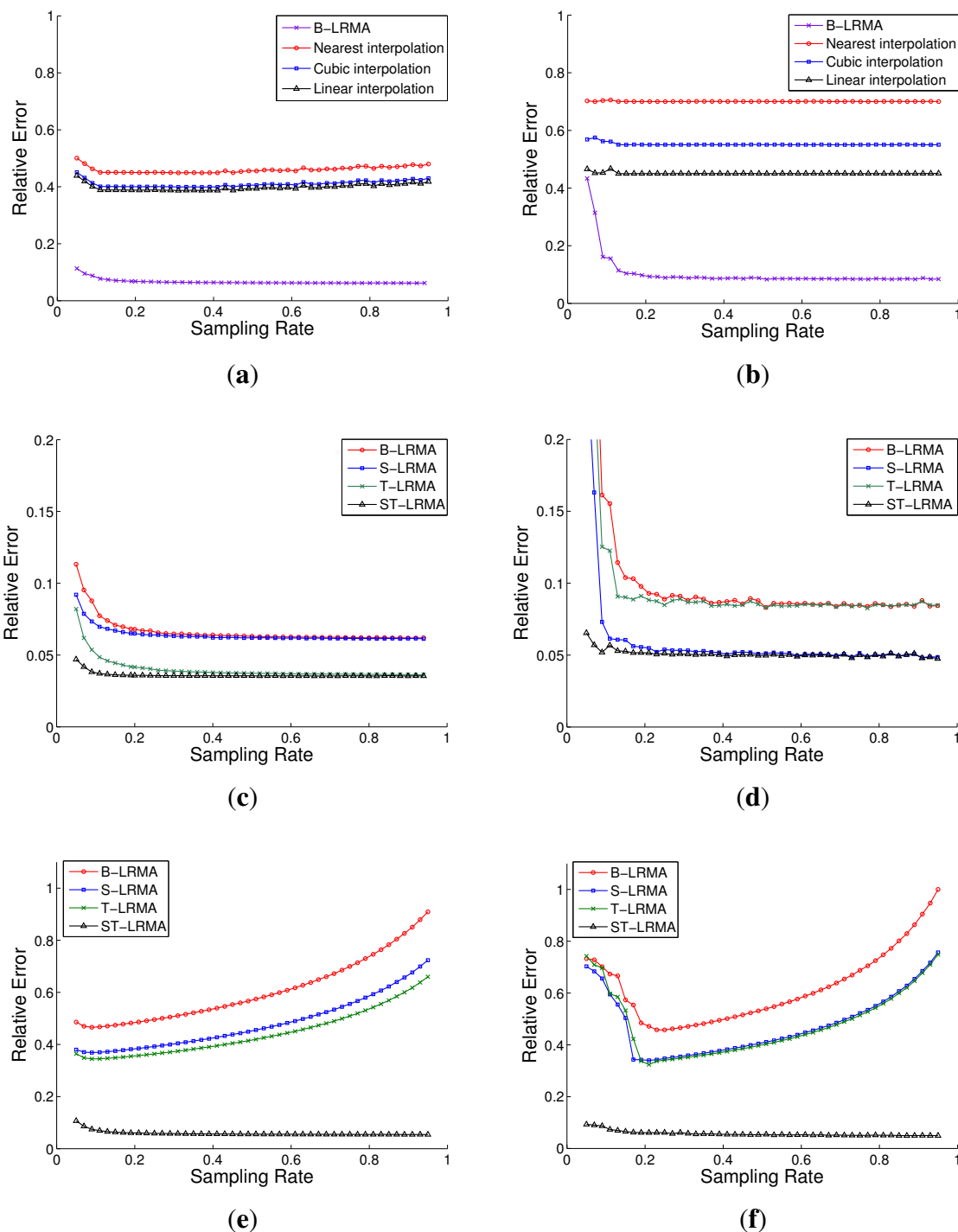
Figure 7. The data recovery results based on different estimated rank and different sampling rate. (a) The results in the simulated network; (b) The results in the real network.



6.2. The Result of the ST-LRMA Based Data Recovery

The proposed ST-LRMA based method are compared with the conventional interpolation methods (linear, cubic and nearest interpolation), the recently proposed LRMA based methods, B-LRMA [14] and T-LRMA [16], and the LRMA based method with only spatial constraint, denoted by S-LRMA. The results, shown in Figure 8a–d, indicate that the LRMA based methods have dramatic improvement in data recovery accuracy compared the conventional interpolation methods. Moreover, the proposed ST-LRMA method outperforms the other LRMA based methods. Especially in the low sampling rate, the proposed method shows a more impressive superiority. This means the proposed method needs only fairly few measurements to obtain high data recovery accuracy. It can be regarded as an energy efficient data collection method and this can be also proved by the following optimized sampling data collection experiments.

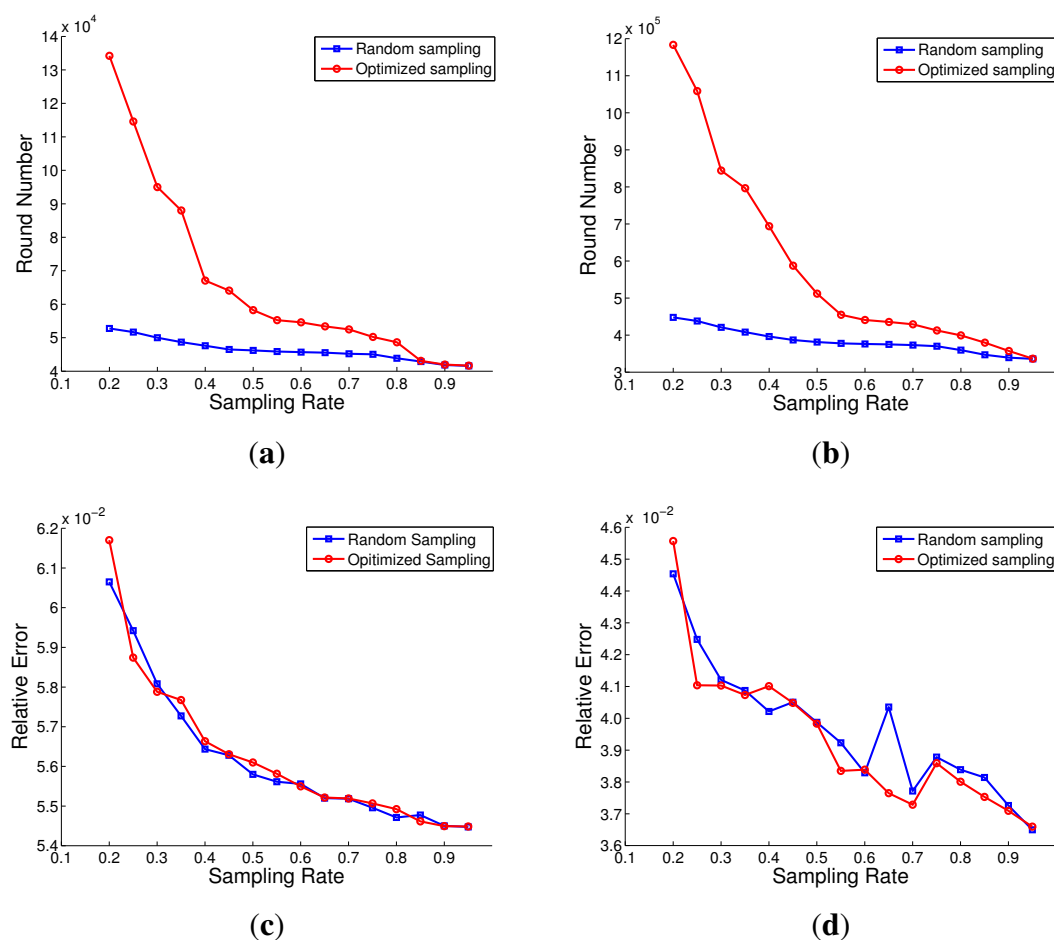
Figure 8. Data recovery results in the simulated and real networks: **(a)** B-LRMA compared with the interpolation methods in the simulated network; **(b)** B-LRMA compared with the interpolation methods in the real network; **(c)** ST-LRMA compared with B-LRMA, S-LRMA and T-LRMA in the simulated network; **(d)** ST-LRMA compared with B-LRMA, S-LRMA and T-LRMA in the real network; **(e)** ST-LRMA compared with B-LRMA, S-LRMA and T-LRMA in the simulated network with the measurement \mathbf{X} missing rows and columns; **(f)** ST-LRMA compared with B-LRMA, S-LRMA and T-LRMA in the real network with the measurement \mathbf{X} missing rows and columns.



According to the low rank completion theory [13], to recover a complete matrix, there should be at least one valid entry at each row and column. However, the data loss often occurs in WSNs. So this condition for the LRMA model may not be guaranteed at some data measurements. To further test the robust of the proposed method, we implement the data recovery experiments with element lost in entire rows and columns of the measurement matrix. Here we randomly remove 10% of rows and columns from the measurement matrices \mathbf{X} both in the simulated and real networks. Then the above data recovery experiments are repeated using the new inputs. The experimental results, shown in Figure 8e,f, show the proposed ST-LRMA has stable recovery results, while B-LRMA, S-LRMA and T-LRMA generate large error and their curves behave un-convergently. It is considered that the improved stability of the proposed method is due to the newly introduced consistency constraints over measurements.

6.3. The Result of the Optimized Sampling for Energy Efficient Data Collection

Figure 9. The optimized node sampling method compared with random sampling method. (a) The network lifetime of the two sampling methods with different sampling rates in the simulated network; (b) The network lifetime of the two sampling methods with different sampling rates in the real network; (c) The data recovery error of the two sampling methods with different sampling rates in the simulated network; (d) The data recovery error of the two sampling methods with different sampling rates in the real network.



To evaluate performance of the optimized node sampling method, we implement data collection in the whole network lifetime with different node sampling rate. The results are compared with the random sampling method, as shown in Figure 9a for the simulated network and Figure 9b for the real network, respectively. It is shown that the optimized node sampling method has better performance than the random sampling method for both the simulated and real networks. Especially in the lower sampling rate, the optimized node sampling method could obtain better network energy distribution and prolong the network lifetime. For example, when the sampling rate is lower than 30%, the network lifetime of the optimized method will be prolonged two times than the network lifetime of the random method.

The data recovery errors are also computed for the two node sampling methods and the results are reported in Figure 9c,d. In the simulated network or the real network, the optimized node sampling method almost has the same performance to the ideal random sampling method which is considered the best way to get good recovery results. From these, we can conclude that the optimized method can not only extend the network lifetime but also keep the accuracy of the data recovery. This is due to the balanced model concerning the node uniform sampling and the energy distribution of network in Section 5.

7. Conclusion

In this paper, we propose an efficient data collection method based on LRMA for WSNs. The contribution of the work lies in two folds. On one hand, a proposed novel ST-LRMA based data collection method introduces the spatial and temporal correlation of the sensing data into the conventional LRMA model. The experiment results indicate that the proposed method has better data recovery accuracy compared with the conventional interpolation methods and the state-of-the-art LRMA based methods. In addition, the proposed method avoids the optimization problem involving empty columns or rows, and could fill the empty columns or rows stably. On the other hand, an optimized nodes sampling method is proposed and integrated with the ST-LRMA based data collection method. In the proposed nodes sampling model, Gini index is adopted to measure the uniformity of the selected nodes and the distribution of the network energy. The data collection experiments were conducted on the simulated and real wireless networks. The results show the proposed method has good performance on network energy efficiency without losing much data recovery accuracy.

Our experiments are implemented without considering some network properties, such as the complex routing, the network time duty circle, and the sensing data attributes. The future work is to integrate these network properties into our ST-LRMA based data collection method and apply it on real large scale wireless sensor networks. Another possible future work is to consider the multi-modality sensing data together and to formulate the sensing data as tensors to be recovered by low rank alike constraints.

Acknowledgments

This paper is supported by National Natural Science Foundation of China Grants (No. 61390510, 61133003, 61370119 and 61171169); The Australian Research Council (ARC) through a Discovery Project (DP) grant (DP140102270); Beijing Natural Science Foundation Grants (No. 4132013, KZ201310005006 and PHR(IHLB)).

Author Contributions

This paper presents part of Xinglin Piao Ph.D. study research. Yongli Hu designed the research, drafted the manuscript and contributed to carrying out the experiment and performing the data analysis. Yanfeng Sun, Baocai Yin and Junbin Gao contributed with valuable discussions and scientific advice. All authors read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Li, M.; Liu, Y. Underground coal mine monitoring with wireless sensor networks. *ACM Trans. Sens. Netw.* **2009**, *5*, 1–29.
2. Vuran, M.C.; Akan, O.B.; Akyildiz, I.F. Spatio-temporal correlation: Theory and applications for wireless sensor networks. *Comput. Netw. J.* **2004**, *45*, 245–259.
3. Hua, G.; Chen, C. Correlated data gathering in wireless sensor networks based on distributed source coding. *Int. J. Sens. Netw.* **2008**, *4*, 13–22.
4. Jim, C.; Petrovic, D.; Ramachandran, K. A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks. In Proceedings of Twenty-Second Annual Joint Conference of the IEEE Computer and Communications, San Francisco, CA, USA, 30 March–3 April 2003; pp. 1054–1062.
5. Yuen, K.; Liang, B.; Li, B. A distributed framework for correlated data gathering in sensor networks. *IEEE Trans. Veh. Technol.* **2008**, *57*, 578–593.
6. Ciancio, A.; Patten, S.; Ortega, A.; Krishnamachari, B. Energy efficient data representation and routing for wireless sensor networks based on a distributed wavelet compression algorithm. In Proceedings of the Fifth International Conference on Information Processing in Sensor Networks, Nashville, TN, USA, 19–21 April 2006; pp. 309–316.
7. Crovella, M.; Kolaczyk, E. Graph wavelets for spatial traffic analysis. In Proceedings of Twenty-Second Annual Joint Conference of the IEEE Computer and Communications, San Francisco, CA, USA, 30 March–3 April 2003; pp. 1848–1857.
8. Yoon, S.; Shahabi, C. The clustered aggregation (cag) technique leveraging spatial and temporal correlations in wireless sensor networks. *ACM Trans. Sens. Netw.* **2007**, *3*, 1–39.
9. Liu, C.; Wu, K.; Pei, J. An energy-efficient data collection framework for wireless sensor networks by exploiting spatiotemporal correlation. *IEEE Trans. Parallel Distrib. Syst.* **2007**, *18*, 1010–1023.
10. Gupta, H.; Navda, V.; Das, S.; Chowdhary, V. Efficient gathering of correlated data in sensor networks. *ACM Trans. Sens. Netw.* **2005**, *3*, 31–43.
11. Xu, X.; Li, X.; Wan, P.; Tang, S. Efficient scheduling for periodic aggregation queries in multihop sensor networks. *IEEE/ACM Trans. Netw.* **2012**, *20*, 690–698.
12. Recht, B.; Fazel, M.; Parrilo, P. Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization. *SIAM Rev.* **2010**, *52*, 471–501.

13. Candès, E.; Recht, B. Exact Matrix Completion via Convex Optimization. *Found. Comput. Math.* **2009**, *9*, 717–772.
14. Cheng, J.; Jiang, H.; Ma, X.; Liu, L.; Qian, L.; Tian, C.; Liu, W. Efficient Data Collection with Sampling in WSNs: Making Use of Matrix Completion Techniques. In Proceedings of IEEE Global Telecommunications Conference, Miami, FL, USA, 6–10 December 2010; pp. 1–5.
15. 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.
16. Piao, X.; Hu, Y.; Sun, Y.; Yin, B. Efficient Data Gathering in Wireless Sensor Networks Based on Low Rank Approximation. In Proceedings of IEEE International Conference on and IEEE Cyber, Physical and Social Computing, Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCoM), Beijing, China, 20–23 August 2013; pp. 699–706.
17. Mainwaring, A.; Culler, D.; Polastre, J.; Szewczyk, R.; Anderson, J. Wireless sensor networks for habitat monitoring. In Proceedings of The First ACM International Workshop on Wireless Sensor Networks and Applications, Atlanta, GA, USA, 28 September 2002; pp. 88–97.
18. Luo, C.; Wu, F.; Sun, J.; Chen, C. Compressive data gathering for large-scale wireless sensor networks. In Proceedings of The Fifteenth Annual International Conference on Mobile Computing and Networking, Beijing, China, 20–25 September 2009; pp. 145–156.
19. Chou, C.; Rana, R.; Hu, W. Energy efficient information collection in wireless sensor networks using adaptive compressive sensing. In Proceedings of IEEE Thirty-fourth Conference on Local Computer Networks, Zurich, Switzerland, 20–23 October 2009; pp. 443–450.
20. Xiang, L.; Luo, J.; Vasilakos, A. Compressed data aggregation for energy efficient wireless sensor networks. In Proceedings of Eighth Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), Salt Lake City, UT, USA, 27–30 June 2011; pp. 46–54.
21. Xiang, L.; Luo, J.; Rosenberg, C. Compressed data aggregation: Energy-efficient and high-fidelity data collection. *IEEE/ACM Trans. Netw.* **2013**, *21*, 1722–1735.
22. Wu, X.; Xiong, Y.; Huang, W.; Shen, H.; Li, M. An efficient compressive data gathering routing scheme for large-scale wireless sensor networks. *Comput. Electr. Eng.* **2013**, *39*, 1935–1946.
23. Bajwa, W.; Haupt, J.; Sayeed, A.; Nowak, R. Compressive wireless sensing. In Proceedings of the Fifth International Conference on Information Processing in Sensor Networks, Nashville, TN, USA, 19–21 April 2006; pp. 134–142.
24. Haupt, J.; Bajwa, W.; Rabbat, M.; Nowak, R. Compressed sensing for networked data. *IEEE Signal Process. Mag.* **2008**, *25*, 92–101.
25. Lin, Z.; Ganesh, A.; Wright, J.; Wu, L.; Chen, M.; Ma, Y. Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix. In Proceedings of the Third International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, Aruba, The Netherlands, 13–16 December 2009; pp. 213–216.
26. Zhang, Y.; Roughan, M.; Willinger, W.; Qiu, L. Spatio-temporal compressive sensing and internet traffic matrices. In Proceedings of the Ninth ACM SIGCOM, Barcelona, Spain, 17–21 August 2009; pp. 267–278.

27. Nikitaki, S.; Tsagkatakis, G.; Tsakalides, P. Efficient training for fingerprint based positioning using matrix completion. In Proceedings of the Twentieth European Signal Processing Conference (EUSIPCO), Bucharest, Rumania, 27–31 August 2012; pp. 195–199.
28. Hu, Y.; Zhou, W.; Wen, Z.; Sun, Y.; Yin, B. Efficient Radio Map Construction Based on Low-Rank Approximation for Indoor Positioning. *J. Math. Probl. Eng.* **2013**, *2013*, No. 461089.
29. Tibshirani, R.; Saunders, M.; Rosset, S.; Zhu, J.; Knight, K. Sparsity and smoothness via the fused lasso. *J. Roy. Stat. Soc. B* **2005**, *67*, 91–108.
30. Guo, Y.; Gao, J.; Li, F. Spatial Subspace Clustering for Drill Hope Spectral Data. *J. Appl. Remote Sens.* **2014**, *8*, 1–9.
31. Tierney, S.; Gao, J.; Guo, Y. Subspace Clustering for Sequential Data. In Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; 1019–1026.
32. Roweis, S.; Saul, L. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326.
33. Tenenbaum, J.B.; Silva, V.; Langford, J.C. A global geometric framework for nonlinear dimensionality reduction. *Science* **2000**, *290*, 2319–2323.
34. Belkin, M.; Niyogi, P.; Dietterich, T.G.; Becker, S.; Ghahramani, Z. Laplacian eigenmaps and spectral techniques for embedding and clustering. In Proceedings of Fifteenth Annual Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 3–8 December 2001; pp. 585–591.
35. Elhamifar, E.; Vidal, R. Sparse manifold clustering and embedding. In Proceedings of Twenty-Fifth Annual Conference on Neural Information Processing Systems, Granada, Andalusia, Spain, 12 December 2011; pp. 55–63.
36. Kennedy, J.; Eberhart, R. Particle Swarm Optimization. In Proceedings of IEEE International Conference on Neural Networks, Perth, Australia, 27 November–1 December 1995; pp. 1942–1948.
37. Intel Lab Data. Available online: <http://db.csail.mit.edu/labdata/labdata.html> (accessed on 16 September 2014).
38. Tian, D.; Georganas, N.D. A node scheduling scheme for energy conservation in large wireless sensor networks. *Wirel. Commun. Mob. Comput.* **2003**, *3*, 271–290.