

# Data Loss and Reconstruction in Wireless Sensor Networks

Linghe Kong, *Member, IEEE*, Mingyuan Xia, Xiao-Yang Liu, Guangshuo Chen, Yu Gu, *Member, IEEE*, Min-You Wu, *Senior Member, IEEE*, and Xue Liu, *Member, IEEE*

**Abstract**—Reconstructing the environment by sensory data is a fundamental operation for understanding the physical world in depth. A lot of basic scientific work (e.g., nature discovery, organic evolution) heavily relies on the accuracy of environment reconstruction. However, data loss in wireless sensor networks is common and has its special patterns due to noise, collision, unreliable link, and unexpected damage, which greatly reduces the reconstruction accuracy. Existing interpolation methods do not consider these patterns and thus fail to provide a satisfactory accuracy when the missing data rate becomes large. To address this problem, this paper proposes a novel approach based on compressive sensing to reconstruct the massive missing data. Firstly, we analyze the real sensory data from Intel Indoor, GreenOrbs, and Ocean Sense projects. They all exhibit the features of low-rank structure, spatial similarity, temporal stability and multi-attribute correlation. Motivated by these observations, we then develop an *environmental space time improved compressive sensing* (ESTI-CS) algorithm with a *multi-attribute assistant* (MAA) component for data reconstruction. Finally, extensive simulation results on real sensory datasets show that the proposed approach significantly outperforms existing solutions in terms of reconstruction accuracy.

**Index Terms**—Wireless sensor networks, data loss and reconstruction, compressive sensing

## 1 INTRODUCTION

PEOPLE investigate the environment in order to understand our physical world. Recently, wireless sensor networks (WSNs) [15], [20] are widely adopted to gather sensory data and reconstruct the environment in the cyber space [11]. There are plenty of environment monitoring applications under the water [30], in the forest [23], and on the volcano [28]. An *environment matrix* (EM) is a common way to represent a dynamic environment. An EM is an  $n \times t$  matrix that records data from  $n$  sensors over  $t$  time intervals. Environment reconstruction [13] attempts to obtain the full and accurate EM from raw sensory data.

### 1.1 Motivation

A great deal of basic scientific work heavily depends on the accuracy of environment reconstruction. For example, scientists reveal the nature of ocean currents from accurate underwater temperature data [30], understand the demand for plant evolution based on the light condition in the forest [23], and discover the eruption omen by monitoring the shake of the volcano [28].

However, since data gathering is largely affected by hardware and wireless conditions, a raw dataset usually has notable missing data. Furthermore, missing data become larger as WSNs grow in scale [3]. Consequently, data loss becomes the key challenge against accurate reconstruction. It is urgent and important to design effective methods to recover incomplete EMs.

### 1.2 Existing Approaches and Limitations

The missing value problem is fundamental in dataset field. Lots of work has been contributed such as K-Nearest Neighbors (KNN) [7], Delaunay Triangulation (DT) [13], and Multi-channel Singular Spectrum Analysis (MSSA) [32]. These methods are often used when there are only a few missing values, but cannot be applied when the missing data grow.

Compressive Sensing (CS) [5], [8] is a powerful and generic technique for estimating missing data. CS can recover an entire dataset from only a small fraction of data as long as these data contain sparse/low-rank features. So far, CS has been utilized to reconstruct network traffic [31], refine localization [24] and improve urban traffic sensing [19]. However, since a WSN has unique data loss patterns, directly applying CS on EM interpolation cannot gain satisfactory accuracy.

### 1.3 Our Work and Contribution

Our work is fourfold: Firstly, we analyze real environmental data from Intel Indoor [1], GreenOrbs [23], and OceanSense [30] projects. We confirm the massive data loss in general applications and mine the specific data loss patterns in WSNs. Then we reveal four features in environmental datasets:

1. *Low-rank structure.* A complete EM can be represented by a few principle data, which underpins the applicability of CS.

- L. Kong is with Shanghai Jiao Tong University, Shanghai, China, and also with Singapore University of Technology and Design, Singapore. E-mail: linghe.kong@sjtu.edu.cn; linghe\_kong@sutd.edu.sg.
- M. Xia and X. Liu are with McGill University, Quebec, Canada. E-mail: {mingyuan.xia@mail, xueliu@cs}.mcgill.ca.
- X.-Y. Liu, G. Chen, and M.-Y. Wu are with Shanghai Jiao Tong University, Shanghai, China. E-mail: {yanglet, chengs, mwu}@sjtu.edu.cn.
- Y. Gu is with Singapore University of Technology and Design, Singapore. E-mail: jasongu@sutd.edu.sg.

Manuscript received 12 July 2013; revised 2 Oct. 2013; accepted 8 Oct. 2013. Date of publication 20 Oct. 2013; date of current version 15 Oct. 2014.

Recommended for acceptance by W.-Z. Song.

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.

Digital Object Identifier no. 10.1109/TPDS.2013.269

2. *Time stability.* The sensory values of one certain node are usually similar at adjacent time slots.
3. *Space similarity.* The sensory values of neighbor nodes are similar.
4. *Multi-attribute correlation.* Multiple environmental attributes have strong correlation in some cases. For example, the change tendency of temperature and light are similar in the OceanSense project [30].

Secondly, motivated by the these observations, we design a novel *environmental space time improved compressive sensing* (ESTI-CS) algorithm for estimating the missing data. ESTI-CS embeds customized features into the baseline CS to deal with the specific data loss patterns, which computes the minimal low-rank approximations of the incomplete EM and refines the interpolation with spatio-temporal features.

Thirdly, when multiple attributes from the same dataset have strong correlation, we design a *multi-attribute assistant MAA* component to leverage this feature for better reconstruction accuracy.

Fourthly, we evaluate the effectiveness of our approach based on trace-driven simulations. We demonstrate that ESTI-CS can outperform existing approaches such as CS, KNN, DT, and MSSA when the raw data contain diverse real loss patterns. Typically, ESTI-CS can achieve an effective environment reconstruction with less than 20 percent error when there are 90 percent missing data in the collected data. In addition, MAA further enhances the performance of ESTI-CS in extensive simulations.

## 2 RELATED WORK

The missing value problem is common in datasets [3]. A great deal of existing work has devoted to interpolating the missing data. K-Nearest-Neighbor (KNN) [7] is a classical local interpolation method. KNN simply utilizes the values of the nearest K neighbors to estimate the missing one. It is frequently used in many low-fidelity estimation cases. Delaunay Triangulation (DT) [13] is a typical global refinement method, which treats the gathered data as vertices. DT takes advantage of these vertices and their global errors to build virtual triangles for data interpolation. It is widely adopted in computer vision for surface rendering. Multi-channel Singular Spectrum Analysis (MSSA) [32] is a data adaptive and nonparametric method based on the embedded lag-covariance matrix. MSSA is often used in geographic data recovery.

Despite much progress in the area of data interpolation, existing methods are suitable for only few missing values, but perform poorly when the loss rate grows high, which is common in WSN cases.

Compressive sensing (CS) is an advanced method to recover the whole condition with just a few sampled data [6], [8]. CS-based methods have been developed for network traffic estimation [31], road traffic interpolation [19], and localization in mobile networks [24]. CS has also witnessed wide applications in WSNs, e.g., recovering signals under noisy background [2], balancing load via compressive data gathering [22]. However, the study of CS for environment reconstruction in WSNs is still vacant.

Existing CS-based interpolation methods cannot be directly applied for accurate environment reconstruction due to two reasons: 1) CS-based methods require the dataset to have inherent structures. Features that are extracted from network traces [31] or road traffic [19] are not applicable for WSN data. 2) CS theory performs well when the missing values follow the Gaussian or pure random distribution [18], [27]. However, as shown in Section 4.2, the loss patterns of WSNs do not satisfy these prerequisites.

To address the above challenges, an effective environment reconstruction approach in WSN is required to deal with the massive data loss problem as well as to study the WSN-specific loss patterns.

## 3 PROBLEM FORMULATION

### 3.1 Environmental Data Reconstruction

Rebuilding the virtual environment (such as the dynamic temperature) in cyber space based on the sensory data is called environment reconstruction.

In environment reconstruction systems, sensor nodes are scattered in the given area. Suppose totally  $n$  sensor nodes are deployed. The monitoring period includes  $t$  time slots. Each sensor node reports its sensory data once per time slot through wireless transmission.  $x(i, j)$  denotes the sensory data of node  $i$  at time slot  $j$ , where  $i = 1, 2 \dots n$  and  $j = 1, 2 \dots t$ .

**Definition 1 Environment Matrix (EM).** *Is a mathematical method to describe the dynamic environment. An EM is defined by  $X = (x(i, j))_{n \times t}$ .*

A complete EM represents that every data in the matrix are successfully collected, i.e., no data loss.

**Definition 2 Binary Index Matrix (BIM).** *Is an  $n \times t$  matrix, which indicates if the data points at the corresponding positions in an EM are missing. BIM is defined as*

$$B = (b(i, j))_{n \times t} = \begin{cases} 0 & \text{if } x(i, j) \text{ is missing,} \\ 1 & \text{otherwise.} \end{cases} \quad (1)$$

**Definition 3 Sensory Matrix (SM).** *Is an  $n \times t$  matrix, which records the raw data collected from WSNs. Due to the presence of missing data, elements of SM are either  $x(i, j)$  gathered by WSNs or zero (missing data point).*

Thereby, an SM is an incomplete EM. An SM is denoted by  $S$  and can be presented by<sup>1</sup>

$$S = B \circ X. \quad (2)$$

### 3.2 Problem Statement

Data reconstruction is to rebuild the real environment (EM) based on the gathered sensory data (SM).

1. In this paper,  $AB$  presents the matrix production of  $A$  and  $B$ .  $A \circ B$  presents the element-wise production of  $A$  and  $B$ .

TABLE 1  
Selected Datasets for Features Analysis

Data Name	Matrix Size	Time Interval
Intel Indoor	49 nodes $\times$ 149 intervals	90 seconds
GreenOrbs	281 nodes $\times$ 170 intervals	10 minutes
Ocean Sense	10 nodes $\times$ 42 intervals	30 minutes

**Definition 4 Reconstructed Matrix (RM).** *Is generated by interpolating the missing values in an SM to approximate EM. RM is denoted by  $\hat{X} = (\hat{x}(i, j))_{n \times t}$ .*

### 3.2.1 Problem: Environment Reconstruction in Sensor Network (ERSN)

Given an SM  $S$ , ERSN problem is to find an optimal RM  $\hat{X}$  that approximates the original EM  $X$  as closely as possible. i.e.,

$$\begin{aligned} \text{Objective : } & \min \|X - \hat{X}\|_F, \\ \text{Subject to : } & S, \end{aligned} \quad (3)$$

where  $\|\cdot\|_F$  is the Frobenius norm used to measure the error between matrix  $X$  and  $\hat{X}$ . For calculating, take  $X$  as an example,  $\|X\|_F = \sqrt{\sum_{i,j} (x(i, j))^2}$ .

In ERSN problem, the objective is to minimize the absolute error. In order to measure the reconstruction error in different scenarios among different methods, we further define the following metric.

**Definition 5 Error Ratio (ER).** *Is the metric for measuring the reconstruction error after interpolation*

$$\epsilon = \frac{\sqrt{\sum_{i,j:b(i,j)=0} (x(i, j) - \hat{x}(i, j))^2}}{\sqrt{\sum_{i,j:b(i,j)=0} (x(i, j))^2}}. \quad (4)$$

Note that the condition  $b(i, j) = 0$  in Eq. (4) indicates that only errors on the missing data are counted.

## 4 DATA LOSS IN SENSOR NETWORKS

In this section, we analyze the data loss in real WSN datasets. The three datasets are from Intel indoor [1], GreenOrbs [23], and OceanSense [30] projects.

### 4.1 Massive Data Loss

Through statistics analysis, we verify that the significant data loss exists in all of these original datasets.

We investigate totally 54 nodes and 84,600 time slots (one month) data from the Intel Indoor dataset, where 23 percent data points are missing. The GreenOrbs dataset also observes 35 percent data loss. And this loss is even larger in OceanSense, which is about 64 percent for 20 nodes and 5,040 time slots (one week). We find that the data loss is common and significant in real WSNs.

### 4.2 Data Loss Pattern

Traditional work usually assumes that the data loss follows a random distribution [19], [32]. However, this claim is not correct in WSN applications. In terms of the nature of WSNs, we synthesize several typical data loss patterns.

#### 4.2.1 Pattern 1 Element Random Loss (ERL)

This is the simplest loss pattern. Data in the matrix are dropped independently and randomly. Missing data points are randomly distributed in a SM. The noise and collision [12] in WSNs are the root causes of this pattern.

#### 4.2.2 Pattern 2 Block Random Loss (BRL)

Data from adjacent nodes in adjacent time slots are dropped together. Congestion [10] always leads to data loss on high-density sensor nodes during a period of time.

#### 4.2.3 Pattern 3 Element Frequent Loss in Row (EFLR)

Unreliable links [29] are common phenomenon in real wireless scenarios. When the link quality is not good, sensory data are prone to loss due to the intermittent transmission. In EFLR, elements in some particular rows have a higher missing probability.

#### 4.2.4 Pattern 4 Successive Elements Loss in Row (SELR)

A certain node starts losing from a particular time slot. This type of loss occurs when some sensor nodes are damaged or run out of energy [26].

#### 4.2.5 Pattern 5 Combinational Loss (CL)

In real world, data loss is a combination of loss patterns above.

## 5 ENVIRONMENTAL DATA MINING

### 5.1 Ground Truth

In order to discover the environmental features, the complete datasets are needed as the ground truth. However, EMs from the three original datasets cannot be directly utilized since they all have considerable data loss. To generate applicable EMs, we perform preprocessing on the raw datasets, which selects the small but complete subsets from these three datasets. The size and time interval of selected matrices are shown in Table 1. As a result, six EMs are generated from preprocessing: indoor temperature, indoor light, forest temperature, forest light, ocean temperature, and ocean light.

### 5.2 Low-Rank Structure Discovery

Environmental data of different locations over different times are not independent. There exists inherent structure or redundancy. We mine these features in above selected datasets by Singular Value Decomposition (SVD), which is an effective non-parametric technique for revealing the hidden structure [16].

Any  $n \times t$  matrix  $X$  can be decomposed into three matrices according to SVD:

$$X = U\Sigma V^T = \sum_{i=1}^{\min(n,t)} \sigma_i u_i v_i^T, \quad (5)$$

where  $V^T$  is the transpose of  $V$ ,  $U$  is an  $n \times n$  unitary matrix (i.e.,  $UU^T = U^T U = I_{n \times n}$ ),  $V$  is a  $t \times t$  unitary matrix (i.e.,  $VV^T = V^T V = I_{t \times t}$ ), and  $\Sigma$  is an  $n \times t$  diagonal matrix constraining the singular values  $\sigma_i$  of  $X$ . Typically, the singular values in  $\Sigma$  are sorted, i.e.,  $\sigma_i \geq \sigma_{i+1}$ ,  $i = 1, 2, \dots$ ,

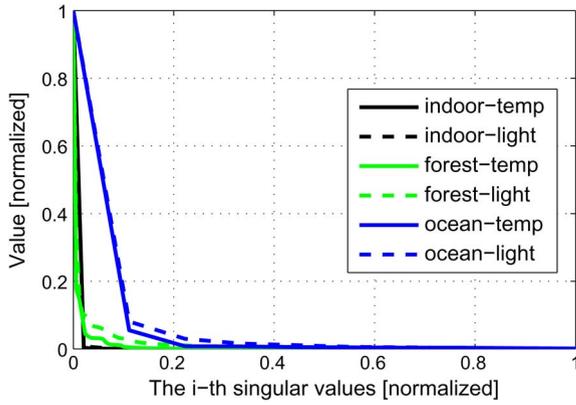


Fig. 1. Low-rank feature.

$\min(n, t)$ , where  $\min(n, t)$  is the number of singular values. The rank of a matrix, denoted by  $r$ , is equal to the number of its non-zero singular values. If  $r \ll \min(n, t)$ , the matrix is considered as low-rank.

In Eq. (5), the singular value  $\sigma_i$  also indicates the energy of the  $i$ -th principal component. The total energy is equal to the sum of all singular value  $\sum_{i=1}^{\min(n,t)} \sigma_i$ . According to PCA, a low-rank matrix [31] exhibits that its first  $r$  singular values occupy the total or near-total energy  $\sum_{i=1}^r \sigma_i \approx \sum_{i=1}^{\min(n,t)} \sigma_i$ .

In Fig. 1, we illustrate the distribution of singular values in 6 EMs. The X-axis presents the  $i$ -th singular values, which is normalized by  $\min(n, t)$  because the scales of 6 EMs are different. The Y-axis presents the values of the sum of first  $i$ -th singular value, which is normalized by  $\max(\sigma_i)$  due to the same reason. This figure suggests that the energy is always contributed by the top several singular values in real environments. For example, the top 5 percent singular values contribute all energy in Indoor-Temp. The universal existence of  $\sum_{i=1}^r \sigma_i \approx \sum_{i=1}^{\min(n,t)} \sigma_i$  and  $r \ll \min(n, t)$  reveals that EMs exhibit obvious low-rank structures. Low-rank features [19] serve for the prerequisite for using compressive sensing.

(Refer to the supplemental file which is available in the Computer Society Digital Library at <http://doi.ieeecomputersociety.org/10.1109/TPDS.2013.269> for the mining of temporal stability and spatial similarity features.)

### 5.3 Multi-Attribute Correlation

We are aware of the following two facts in real WSN applications. 1) Usually, WSNs gather multiple attributes simultaneously, e.g., a TelosB node [21] senses three environmental attributes: temperature, light and humidity. 2) Multiple attributes have correlations in some applications. For instance, the empirical study [17] reveals that several attributes do have relationship such as relative humidity and dewpoint temperature. Therefore, we propose to mine and exploit such correlations to further optimize the accuracy of environment reconstruction.

#### 5.3.1 Joint Sparse Decomposition

In order to mine the correlations, a *Joint Sparse Decomposition (JSD)* method is proposed to jointly divide multi-attribute EMs into a public sub-matrix  $W$  and multiple private sub-matrices  $\Delta$ . All sub-matrices have the same size of EMs, but their magnitudes are smaller. Suppose two attributes

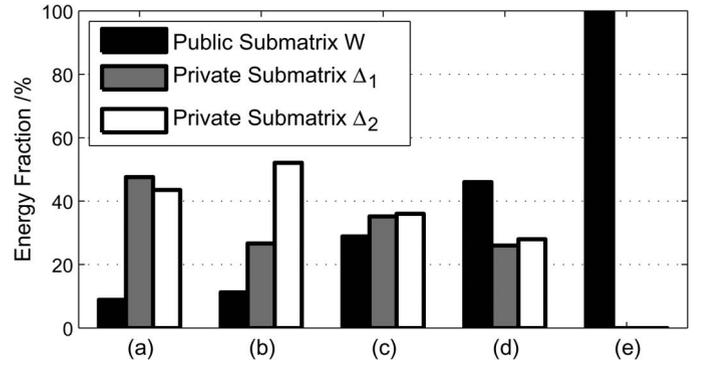


Fig. 2. Correlation analysis by JSD.

$X_1 = (x_1^{(1)}, \dots, x_1^{(t)})$  and  $X_2 = (x_2^{(1)}, \dots, x_2^{(t)})$ , where  $x_k^{(j)}$  presents the  $j$ -th column vector of EM  $X_k$ ,  $j = 1, 2, \dots, t$ . For both column vector  $x_1^{(j)}$  and  $x_2^{(j)}$ , the goal is to split them into:

$$\begin{aligned} x_1^{(j)} &= w^{(j)} + \delta_1^{(j)}, \\ x_2^{(j)} &= w^{(j)} + \delta_2^{(j)}, \\ w^{(j)} &= \Psi v^{(j)}, \end{aligned} \tag{6}$$

where  $w^{(j)}$  is the public sub-vector of  $x_1^{(j)}$  and  $x_2^{(j)}$ , which is the multiplication of a wavelet basis  $\Psi$ [4] and a sparse vector  $v^{(j)}$  satisfying  $w^{(j)} = \Psi v^{(j)}$ . The private sub-vectors are represented by  $\delta_1^{(j)}$  and  $\delta_2^{(j)}$  respectively. According to Compressive Sensing theory [8], [5],  $v^{(j)}$ ,  $\delta_1^{(j)}$ , and  $\delta_2^{(j)}$  are obtained by solving an  $l_1$ -norm minimization problem as the following:

$$\hat{v} = \underset{v}{\operatorname{argmin}} \|v\|_1, \quad \text{s.t. } x = Av \tag{7}$$

where  $\|\cdot\|_1$  is the  $l_1$ -norm,  $v = (v^{(j)T}, \delta_1^{(j)T}, \delta_2^{(j)T})^T$ ,  $x = (x_1^{(j)T}, x_2^{(j)T})^T$  and  $A = (\Psi, I, \mathbf{0}; \Psi, \mathbf{0}, I)$ . It was proved in [8] that solving Eq. (7) is NP-hard, so we adopt the least angle regression method, which was proposed in [9], to obtain  $\hat{v}$ . Then the public sub-vector  $w^{(j)}$ , the private sub-vectors  $\delta_1^{(j)}$ , and  $\delta_2^{(j)}$  can be calculated from  $\hat{v}$ .

Applying JSD onto every column vector,  $X_1$  and  $X_2$  are decomposed as

$$\begin{aligned} X_1 &= W + \Delta_1, \\ X_2 &= W + \Delta_2. \end{aligned} \tag{8}$$

#### 5.3.2 Correlation

The energy fraction of public sub-matrix  $W$  is used to measure the correlation between two attributes, where the total energy is the sum of all singular values of three sub-matrices  $W$ ,  $\Delta_1$ , and  $\Delta_2$ . Fig. 2 shows the energy fraction of sub-matrices after JSD in diverse groups. Group a shows the results of JSD on two irrelevant random matrices. The public sub-matrix  $W$  contains only 7 percent of total energy. Group b shows the results of indoor-temp and indoor-light. The change of indoor-light has mutations by manually switching lights on/off, which leads to low correlation with indoor-temp  $W = 11\%$ . The results of forest-temp and forest-light are shown in group c. Both outdoor light and temperature vary according to the sun. However, due to the influence of tree shade, the correlation is not very strong. So  $W$  contains 29 percent of total energy,

TABLE 2  
Inherited Low-Rank Analysis after JSD

Dataset	Sub-Matrix	Inherited Low-rank
Intel Indoor	$W$	7%
-temp	$\Delta_1$	2%
-light	$\Delta_2$	8%
GreenOrbs	$W$	9%
-temp	$\Delta_1$	9%
-light	$\Delta_2$	12%
OceanSense	$W$	17%
-temp	$\Delta_1$	11%
-light	$\Delta_2$	18%

while the private sub-matrices  $\Delta_1$  and  $\Delta_2$  contain 35 percent and 36 percent respectively. And sensor nodes are fully exposed under the sun in OceanSense. Hence, high correlation between ocean-temp and ocean-light is shown in group d, where  $W = 46\%$ . Two same matrices have definitely highest correlation. When JSD is operated on two same EMs  $X_1 = X_2$ ,  $W$  contains 100 percent energy and  $\Delta_1 = \Delta_2 = 0$  in group e. Fig. 2 validates that JSD can be utilized to measure the correlation between two matrices. In addition, the higher correlation between two EMs, the more energy contains in the public sub-matrix.

### 5.3.3 Inherited Low-Rank

After decomposed from EMs by applying JSD, the sub-matrices  $W$ ,  $\Delta_1$  and  $\Delta_2$  still exhibit the low-rank features.

All sub-matrices are decomposed respectively by SVD method. The same method mentioned in Section 5.2 is adopted to determine the inherited low-rank feature. In Table 2, we show the percentage of singular values that contain 90 percent of the total energy. As shown, for all the dataset, 7 percent to 20 percent of the top singular values can concentrate 90 percent of the total energy. Hence, the inherited low-rank features are exhibited in sub-matrices, which indicates that any of  $W$ ,  $\Delta_1$  and  $\Delta_2$  can be recovered by CS based method.

Correlation and inherited low-rank motivate us to improve ESTI-CS by multi-attribute correlation.

## 6 ENVIRONMENTAL SPACE TIME IMPROVED COMPRESSING SENSING APPROACH

We propose a novel missing data estimation approach to address ERSN problem. The proposed algorithm, namely *environmental space time improved compressive sensing* (ESTI-CS), takes into consideration the spatio-temporal features to optimize the estimation accuracy.

### 6.1 Compressive Sensing Based Approach

Since we have revealed the low-rank structure in most real environment datasets, we propose to use CS method to estimate missing data from the SM.

The goal of solving ERSN problem is to estimate  $\hat{X}$ . According to Eq. (5), any matrix can be decomposed by SVD into  $\sum_{i=1}^{\min(n,t)} \sigma_i u_i v_i^T$ . Through the inverse process, we can also create an  $r$ -rank approximation  $\hat{X}$  by using only the  $r$  largest singular values and abandoning the others:

$$\sum_{i=1}^r \sigma_i u_i v_i^T = \hat{X}. \quad (9)$$

This  $\hat{X}$  is known as the best  $r$ -rank approximation that minimizes the error measured by Frobenius norm. Nevertheless, the optimal  $\hat{X}$  cannot be obtained directly by this way as we do not know matrix  $X$  and the proper rank in advance.

Thus we propose to find  $\hat{X}$  as follows:

$$\begin{aligned} \text{Objective : } & \min(\text{rank}(\hat{X})), \\ \text{Subject to : } & B \circ \hat{X} = S. \end{aligned} \quad (10)$$

We make this assumption according to two reasons. On the one hand, since the reconstructed matrix (RM) is generated from the sensory matrix (SM), it is reasonable to be as close as SM. On the other hand, like the environmental matrix (EM), RM should also have a low-rank structure. Given this, it is still difficult to solve this minimization problem because it is non-convex. To bypass this difficulty, we take advantage of the SVD-like factorization, which rewrites Eq. (5) as

$$\hat{X} = U\Sigma V^T = LR^T, \quad (11)$$

where  $L = U\Sigma^{1/2}$  and  $R = V\Sigma^{1/2}$ . Substituting Eq. (11) to Eq. (10), we can solve the minimization problem according to the compressive sensing theory in [5], [8]. Specifically, if the restricted isometry property holds [25], minimizing the nuclear norm can result to rank minimization exactly for a low-rank matrix. Hereby, we just need to find matrix  $L$  and  $R$  that minimize the summation of their Frobenius norms:

$$\begin{aligned} \text{Objective : } & \min(\|L\|_F^2 + \|R^T\|_F^2) \\ \text{Subject to : } & B \circ (LR^T) = S. \end{aligned} \quad (12)$$

Looking for  $L$  and  $R$  that strictly satisfy Eq. (12) is likely to fail due to two reasons. First, EMs usually approximate low-rank but not exact low-rank. Second, noises in sensory data may lead to the over-fitting problem if strict satisfaction is required. Thus, instead of solving Eq. (12) directly, we solve the following equation using the Lagrange multiplier method:

$$\min(\|B \circ (LR^T) - S\|_F^2 + \lambda(\|L\|_F^2 + \|R^T\|_F^2)), \quad (13)$$

where the Lagrange multiplier  $\lambda$  allows a tunable tradeoff between rank minimization and accuracy fitness. This solution provides the low-rank approximation but not strict satisfaction.

In Eq. (13), 1)  $B$  and  $S$  are known, 2) any  $\|\cdot\|_F^2$  is non-negative, 3) the optimal values approximate 0 by minimizing all non-negative parts. Hence,  $L$  and  $R$  can be estimated in this optimization problem.

### 6.2 Environmental Spatio-Temporal Improvement

ESTI-CS includes two key components: On the other hand, after exploiting the temporal stability and spatial similarity features, we complete ESTI-CS approach by developing Eq. (13) as following:

$$\begin{aligned} \min(\|B \circ (LR^T) - S\|_F^2 + \lambda(\|L\|_F^2 + \|R^T\|_F^2) \\ + \|HLR^T\|_F^2 + \|LR^T T\|_F^2), \end{aligned} \quad (14)$$

where  $\mathbf{H}$  and  $\mathbf{T}$  are the spatial and temporal constraint matrices respectively. Three subjects  $\|\mathbf{H}LR^T\|_F^2$ ,  $\|LR^T\mathbf{T}\|_F^2$ , and  $\|B \circ (LR^T) - S\|_F^2$  are set to be the same order of magnitude, whose coefficients are 1. Otherwise, they may overshadow the others when solving Eq. (14).

### 6.2.1 Temporal Stability Improvement

The temporal constraint matrix  $\mathbf{T}$  captures the temporal stability feature, which outlines that the change between two consecutive time slots is small. Hence, we set  $\mathbf{T} = \text{Toeplitz}(0, 1, -1)_{t \times t}$ . The Toeplitz matrix is defined with central diagonal given by 1, and the first upper diagonal given by  $-1$ , and the others given by 0.

This Toeplitz matrix adds the temporal constraint into the estimation. Importing  $\|LR^T\mathbf{T}\|_F^2$  into Eq. (14) is equal to induct an additional constraint into the original optimization problem. Since the temporal constraint is an inherent feature of environment, this additional constraint can filter more noises and errors in  $LR^T$  estimation.

### 6.2.2 Spatial Similarity Improvement

The spatial constraint matrix  $\mathbf{H}$  captures the spatial similarity feature, which reveals that values among one-hop neighbors nodes are usually similar. Hence, we set  $\mathbf{H}$  to be a row-normalized  $H^*$ , where  $H^* = H + D$ . The matrix  $H$  is a TM-1H, i.e., the one-hop topology matrix mentioned before. And  $D$  is an  $n \times n$  diagonal matrix, which is defined with central diagonal given by  $\text{diag}(d_1, d_2, \dots, d_n)$ , and the others given by 0. In  $D$ ,  $d_i = -\sum H_{(i)}$ .

The spatial similarity constraint is added by the matrix  $\mathbf{H}$ . Computing the result of  $\mathbf{H}X$  is to get the differences between the elements and the average value of their one-hop neighbors in  $X$ . As the same purpose of time improvement part, we introduce the part of minimizing  $\|\mathbf{H}LR^T\|_F^2$  into Eq. (14). It takes advantage of the inherent environment feature as an additional constraint in optimization problem, which leads to a more accurate estimation of  $LR^T$ , i.e.,  $\hat{X}$ .

## 6.3 ESTI-CS Algorithm

We propose an efficient ESTI-CS algorithm to solve the estimation in the optimization problem Eq. (14).

First, we scale the  $\mathbf{T}$  and  $\mathbf{H}$  as all  $\|\cdot\|_{F_2}$  in Eq. (14) having the same order of magnitude. The scaling method is similar to [31]. Then ESTI-CS algorithm solves the optimization in an iterative manner.  $L$  is initialized randomly, so  $R$  can be computed by solving the following contradictory equation:

$$\begin{bmatrix} B \circ (LR^T) \\ \sqrt{\lambda}R^T \end{bmatrix} = \begin{bmatrix} S \\ 0 \end{bmatrix}. \quad (15)$$

This equation can be rewritten as:

$$\begin{bmatrix} \text{Diag}(B_{(i)})LR_{(i)}^T \\ \sqrt{\lambda}R_{(i)}^T \end{bmatrix} = \begin{bmatrix} S_{(i)} \\ 0 \end{bmatrix}, \quad (16)$$

where  $i$  ranges from 1 to  $t$ . This is a combination of multiple standard linear least squares problems. We then have  $R_{(i)}^T = (P_i^T P_i) \setminus (P_i^T Q_i)$ , where  $P_i = [\text{Diag}(B_{(i)})L; \sqrt{\lambda}I_r]$  and

$Q_i = [S_{(i)}; 0_r]$ . Similarly, once  $R^T$  is obtained,  $L$  can be re-computed by fixing  $R^T$ . This mutual re-computing process repeats until the optimal value is reached.

We analyze the complexity of the ESTI-CS algorithm. The key operation is the procedure for computing the inverse matrix, which provides the best approximate solution to the contradictory equation. This procedure is completed by a matrix multiplication [19]. Thus, its time complexity is  $\mathcal{O}(rnt)$ . Since ESTI-CS repeats the procedure for  $\varrho$  times, the total complexity is  $\mathcal{O}(rnt\varrho)$ . From our evaluation experience in Section 8,  $L$  and  $R^T$  usually converge after 5 iterations.

## 7 MULTI-ATTRIBUTE COMPONENT

### 7.1 MAA Overview

Multi-attribute assistant component can be utilized to improve the accuracy of ESTI-CS when correlation exists. Under such scenario, the proposed ERSN problem is extended to  $k$ -ERSN problem: Given  $K$  sensory matrices (SMs)  $S_k$ , where  $k = 1, 2, \dots, K$ , and these SMs have the same size but different values. The goal is to jointly find the corresponding optimal reconstructed matrices (RMs)  $\hat{X}_k$  that approximate the original environmental matrices (EMs)  $X_k$ .

For simplicity, we study the two-attribute situation as an example. Formally, when  $K = 2$ , ERSN problem is formulated as follows: Given  $S_1$  and  $S_2$ , find an optimal solution for  $\hat{X}_1$  and  $\hat{X}_2$ , i.e.,

$$\begin{aligned} \text{Objective:} \quad & \min(\|\hat{X}_1 - X_1\|_F + \|\hat{X}_2 - X_2\|_F), \\ \text{Subject to:} \quad & B_1 \circ \hat{X}_1 = S_1, \\ & B_2 \circ \hat{X}_2 = S_2. \end{aligned} \quad (17)$$

#### 7.1.1 Normalization

Since the magnitudes of attributes are different, it may cause one matrix to overshadow another. In order to overcome this issue,  $X_1$  and  $X_2$  are normalized respectively.

#### 7.1.2 Low-Rank Matrix Approximation

Eq. (17) is tied by  $X_1$  and  $X_2$ , so the problem cannot be solved in closure form. However, due to the inherited low-rank feature, this problem can be converted to a rank minimization problem. Thus, the optimal  $X_k$  is evaluated by the problem:

$$\min(\text{rank}(\hat{X}_k)), \quad \text{s.t.} \quad S_k = \hat{X}_k \circ B_k. \quad (18)$$

Still two problems are up against us: 1) the rank calculating operator  $\text{rank}(\cdot)$  is not convex. 2) there is no connection between  $X_1$  and  $X_2$ .

To conquer the difficulty 1), we still utilize SVD-like factorization as  $\hat{X} = LR^T$ . Thus  $\min(\text{rank}(\hat{X}))$  is solvable by looking for  $L$  and  $R$ , which satisfy  $\min(\|L\|_F^2 + \|R^T\|_F^2)$ .

#### 7.1.3 Compressive Sensing-Based Joint Matrix Decomposition

To overcome the difficulty (2), we need to find the correlation between  $X_1$  and  $X_2$ . Through the JSD analysis in Section 5.3,

we separate the approximation  $\hat{X}_1$  and  $\hat{X}_2$  by JSD as

$$\begin{aligned}\hat{X}_1 &= \hat{W} + \hat{\Delta}_1 \\ \hat{X}_2 &= \hat{W} + \hat{\Delta}_2.\end{aligned}\quad (19)$$

Since  $\hat{W}$ ,  $\hat{\Delta}_1$  and  $\hat{\Delta}_2$  inherit the low-rank feature,  $k$ -ERSN problem is reformulated as:

$$\begin{aligned}\text{Objective:} & \quad \min(\|\hat{W}\|_* + \|\hat{\Delta}_1\|_* + \|\hat{\Delta}_2\|_*), \\ \text{Subject to:} & \quad B_1 \circ (\hat{W} + \hat{\Delta}_1) = B_1 \circ X_1, \\ & \quad B_2 \circ (\hat{W} + \hat{\Delta}_2) = B_2 \circ X_2\end{aligned}\quad (20)$$

where  $\|\cdot\|_*$  is the nuclear norm which is defined as the sum of singular values, e.g.,  $\|X\|_* = \sum_{i=1}^r \sigma_i(X)$ .

Furthermore, using SVD-like factorization,  $\|\hat{W}\|_* + \|\hat{\Delta}_1\|_* + \|\hat{\Delta}_2\|_*$  in Eq. (20) is rewritten as:

$$\|L_W\|_F^2 + \|R_W^T\|_F^2 + \|L_1\|_F^2 + \|R_1^T\|_F^2 + \|L_2\|_F^2 + \|R_2^T\|_F^2 \quad (21)$$

where  $L_W, L_1, L_2$  are  $n \times r$  matrices and  $R_W, R_1, R_2$  are  $r \times t$  matrices. Moreover,  $\hat{W} = L_W R_W^T$ ,  $\hat{\Delta}_1 = L_1 R_1^T$  and  $\hat{\Delta}_2 = L_2 R_2^T$ . For short, Eq. (21) is denoted by  $\sum \|L_j\|_F^2 + \sum \|R_j\|_F^2$ , where  $j = 1, 2, W$ .

To avoid overfitting,  $k$ -ERSN problem is rewritten to be a non-stationary optimization problem using the Lagrange multiplier method, i.e.,

$$\begin{aligned}\min & \left( \lambda \left( \sum \|L_j\|_F^2 + \sum \|R_j\|_F^2 \right) \right. \\ & \quad \left. + \|B_1 \circ (L_W R_W^T + L_1 R_1^T) - S_1\|_F^2 \right. \\ & \quad \left. + \|B_2 \circ (L_W R_W^T + L_2 R_2^T) - S_2\|_F^2 \right).\end{aligned}\quad (22)$$

Eq. (22) is the core of MAA component, which is solvable because 1)  $B_1, B_2, S_1$ , and  $S_2$  are known, 2) each  $\|\cdot\|_F^2$  is non-negative, 3) the optimal value can be reached by minimizing all non-negative parts to zero. Hence,  $\hat{X}_1$  and  $\hat{X}_2$  can be estimated by combining Eq. (22) and Eq. (19).

**ESTI-CS with MAA** is to reconstruct several (two as example) environments according to

$$\begin{aligned}\min & \left( \lambda \left( \sum \|L_j\|_F^2 + \sum \|R_j\|_F^2 \right) \right. \\ & \quad \left. + \|B_1 \circ (L_W R_W^T + L_1 R_1^T) - S_1\|_F^2 \right. \\ & \quad \left. + \|B_2 \circ (L_W R_W^T + L_2 R_2^T) - S_2\|_F^2 \right. \\ & \quad \left. + \left\| \mathbb{H}_1(L_1 + L_W)(R_1 + R_W)^T \right\|_F^2 \right. \\ & \quad \left. + \left\| \mathbb{H}_2(L_2 + L_W)(R_2 + R_W)^T \right\|_F^2 \right. \\ & \quad \left. + \left\| (L_1 + L_W)(R_1 + R_W)^T \mathbb{T}_1 \right\|_F^2 \right. \\ & \quad \left. + \left\| (L_2 + L_W)(R_2 + R_W)^T \mathbb{T}_2 \right\|_F^2 \right).\end{aligned}\quad (23)$$

It can be seen that Eq. (23) is the combination of Eq. (22) and Eq. (14). The first three items of Eq. (23) utilize the low-rank feature, which is the fundamental compressive sensing, the fourth and fifth items incorporate the spatial similarity improvement, the last two items merge the temporal stability improvement, and the multi-attribute assistant component is added a lastly into Eq. (23) by  $L_j$  and  $R_j$ , where  $j = 1, 2, W$ .

### 7.1.4 Extension

The MAA component is also suitable for the case of more attributes. For instance, if we obtain  $k$  attributes in one WSN, represented by  $X_1, X_2, \dots, X_k$ . The utilization of MAA is to rewrite Eq. (20) into

$$\|\hat{W}\|_* + \|\hat{\Delta}_1\|_* + \|\hat{\Delta}_2\|_* + \dots + \|\hat{\Delta}_k\|_*.\quad (24)$$

Then  $k$ -ERSN problem can be solved by the similar method of the above two-attribute case.

## 7.2 MAA Algorithm

We only present the core of MAA component in this section, which is to solve Eq. (22). And the realization of spatial-temporal improvement in Eq. (23) is the same method in ESTI-CS, we do not repeat it here.

The algorithm solves the problem in an iterative manner. First,  $L_1, L_2, L_W, R_1$ , and  $R_2$  matrices are initialized randomly. Then,  $R_W$  can be calculated from the initialized matrices by solving the equation

$$\begin{bmatrix} B_1 \circ (L_W R_W^T) \\ B_2 \circ (L_W R_W^T) \\ \sqrt{\lambda} R_W^T \end{bmatrix} = \begin{bmatrix} S_1 - L_1 R_1^T \\ S_2 - L_2 R_2^T \\ 0 \end{bmatrix}.\quad (25)$$

Eq. (25) is solvable using the linear least square method. After  $R_W$  is obtained,  $L_W$  can be computed using the same procedure by fixing  $R_W$ . Similarity, any of  $L_1, L_2, R_1$ , and  $R_2$  is computed by fixing the other three. Using the iterative manner, these four matrices can be obtained one-by-one. The computational complexity of ESTI-CS-MAA is the same as ESTI-CS.

## 8 PERFORMANCE EVALUATION

### 8.1 Methodology

The proposed ESTI-CS approach is compared with existing algorithms for missing data interpolation for environmental reconstruction in WSNs.

#### 8.1.1 Ground Truth

Since the performance evaluation needs complete EMs  $X$  to compute the metric of error ratio (ER), we utilize the datasets as shown in Table 1. Six EMs are adopted: indoor-temp, indoor-light, forest-temp, forest-light, ocean-temp and ocean-light.

#### 8.1.2 Methods

To verify the effectiveness of ESTI-CS, four classic interpolation methods are selected for comparison. They are compressive sensing (CS) [19] (computational complexity  $\mathcal{O}(rntq)$ ), Delaunay Triangulation (DT) [13] (complexity  $\mathcal{O}(nt \log nt)$ ), Multi-channel Singular Spectrum Analysis (MSSA) [32] (complexity  $\mathcal{O}(rnt \log nt + r^2 nt)$ ), and K-Nearest Neighbor (KNN) [7] (complexity  $\mathcal{O}(nt)$ ). The parameter  $K$  in KNN is set to be  $\sum_{i=1}^n H_{(i)}/n$ . The parameter  $M$  in MSSA is set to 32 as suggested by [32].

#### 8.1.3 Procedure

The procedure of simulation is:

1. Generate BIM  $B$  according to four loss patterns.
2. Compute SM  $S$  according to Definition 3:  $S = B \circ X$ .

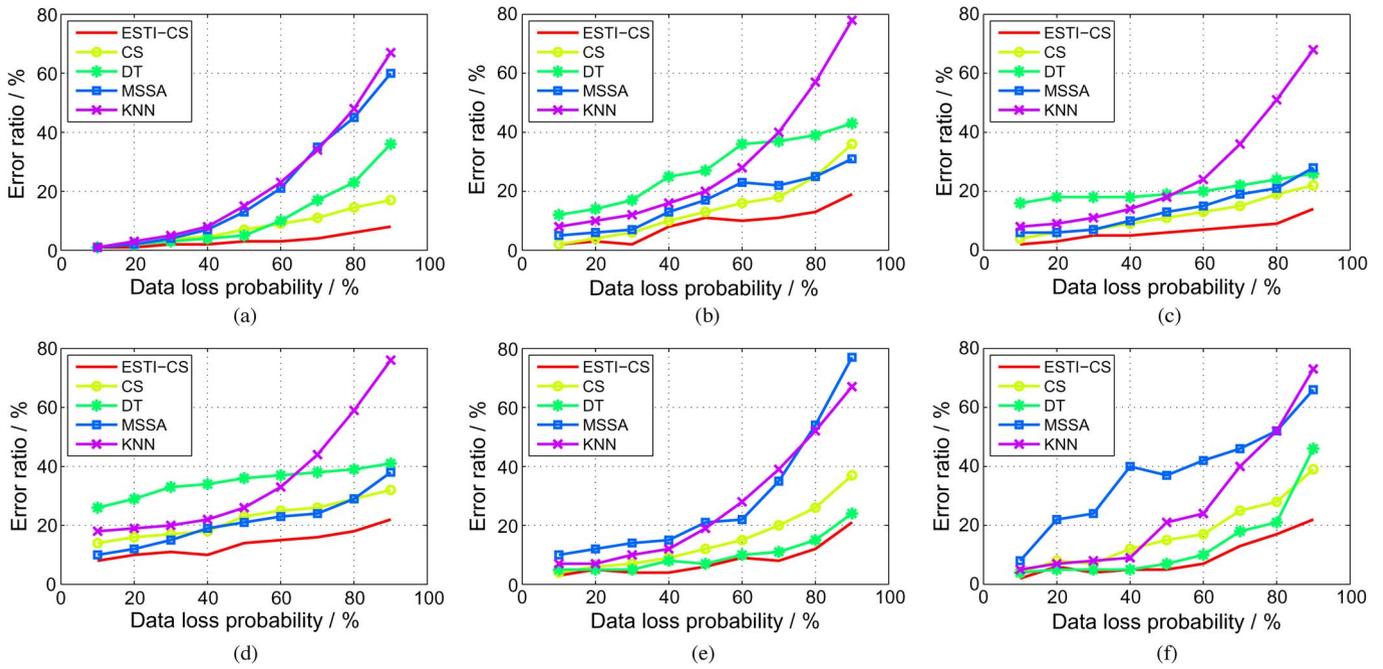


Fig. 3. Error ratio performance of five algorithms in the basic data loss pattern: element random loss. (a) Indoor temp. (b) Indoor light. (c) Forest temp. (d) Forest light. (e) Ocean temp. (f) Ocean light.

3. All interpolation algorithms being tested take SMs as input and generate RMs.
4. The accuracy metric ER is computed between EMs and RMs. And finally, these errors are compared.

### 8.1.4 Series

Three series of experiments are evaluated. The basic experiment measures the performance of different algorithms against typical random loss probability. The second experiment evaluates the performance in diverse data loss patterns. And the third experiment compares the performance of ESTI-CS with MAA and ESTI-CS without MAA.

## 8.2 Comparison on Random Loss Pattern

In the basic comparison, we test the error ratios under diverse algorithms on the element random loss (ERL) pattern only. The data loss rate  $p_{ERL}$  ranges from 10 percent to 90 percent. Fig. 3 shows the results. The X-axis presents the data loss probability, and the Y-axis is the value of ER, which represents the reconstruction accuracy.

In the indoor-temp Fig. 3a, ESTI-CS shows the best performance. Even 90 percent data have been lost, ESTI-CS still can reconstruct the environment with  $ER \leq 10\%$ . While ER of CS is about 19 percent, DT is close to 38 percent, and ER of KNN and MSSA are more than 60 percent. ESTI-CS is much better than other algorithms in this scenario. In the indoor-light Fig. 3b, ESTI-CS still outperforms the others, but the advantage is less significant than that in indoor-temp. The reason is that the indoor temperature change has strong spatio-temporal feature. However, the change of indoor light is largely influenced by the light switch. So the indoor light dataset observes more artificial changes than spatio-temporal stability.

The performance of Forest-Temp and the Forst-Light are similar. The reason is that both the temperature and the

light are mainly effected by the sun due to an outdoor application. These two environment attributes have strong correlation. As shown in Figs. 3c and 3d, ESTI-CS achieves the best environment reconstruction among the five algorithms. CS, MSSA and DT fall behind ESTI-CS a little. KNN is not bad when  $p_{ERL} < 50\%$ , but when  $p_{ERL} > 50\%$ , ER of KNN increases quickly.

In the ocean-temp Fig. 3e, ESTI-CS and DT produce the similar performance. When the data loss is 90 percent, they achieve  $ER < 30\%$ . Meanwhile, the ERs of CS, KNN and MSSA are bigger. In the ocean-light Fig. 3f, the performance of ESTI-CS and DT are similar with the range of loss rate from 10 percent to 80 percent. When the loss rate increases to 90 percent, ER of DT also increases rapidly, and ER of ESTI-CS still keeps within 22 percent. These two figures indicate that ESTI-CS perform better than DT, CS, KNN and MSSA in this outdoor and small-scale WSN scenario.

Overall, ESTI-CS obtain lower interpolation error, which can be used in almost all tested datasets with different loss ratios. KNN and DT produce similar but the poor ER performance, because both of them interpolate with only the space relation among nodes but no time relation consideration. CS and MSSA are better than KNN and DT, but still worse than ESTI-CS. Especially, at the high data loss cases (data loss  $\geq 80$  percent), ESTI-CS exhibits an evident advantage over other algorithms. In all dataset, ESTI-CS can successfully achieve an environment reconstruction with 20 percent error when there are 90 percent data are missing.

## 8.3 Performance on Data Loss Patterns

In Fig. 4, we plot the comparison histograms of five algorithms for reconstructing the environment with different data loss patterns.

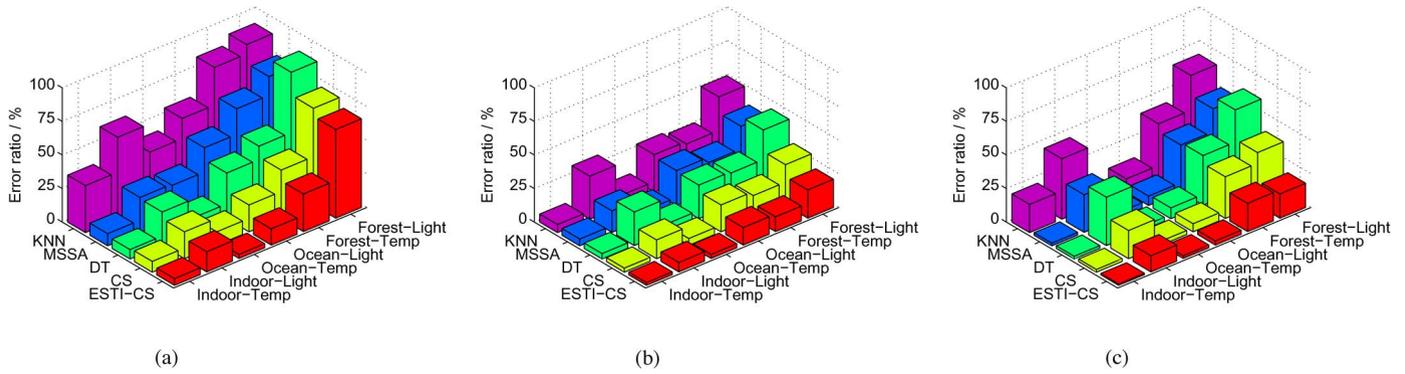


Fig. 4. Error ratio performance in different loss patterns. (a) Block random loss. (b) Element frequent loss in row. (c) Combinational loss.

In the simulation for BRL pattern, each of the six EMs is set to lose data with the block pattern. The scale and the number of the blocks are random, but the amount of total data loss is 50 percent in this simulation. In Fig. 4a, most algorithms in most EMs perform not well. For example, in forest-light, ER of all algorithms are bigger than 60 percent. The reasons are 1) in the forest, many shadows disturb the spatio-temporal stability. 2) if large blocks of data lose, spatio-temporal optimized estimation is helpless either. These two reasons lead to the result. However, in indoor-temp, ocean-temp, and ocean-light, the environment changes are smoothly, ER of ESTI-CS are less than 5 percent despite 50 percent BRL data loss. Even indoor-light, forest environments, ESTI-CS is still a bit better than the others. In addition, we find that KNN is in big trouble for estimating the missing data in BRL.

In the simulation of EFLR pattern, the rows are randomly selected, the loss frequency in these rows is set  $> 75$  percent, and the totally lose data in matrix is set 50 percent. We find that the results in Fig. 4b are close to the basic comparison, because the data loss in EFLR is similar to ERL pattern. In EFLR, the temporal optimization can contribute a partial effect, but the space optimization still works. So our ESTI-CS still outperforms CS, KNN, DT, and MSSA.

The performance of SELR is similar to EFLR, we show the SELR result in the supplemental document available online.

In the simulation of Combinational Loss pattern, we set 20%ERL + 10%BRL + 10%EFLR + 10%SELR. The results of five algorithms are shown in Fig. 4c. The ER of ESTI-CS is  $\leq 20$  percent in any dataset in the combinational loss pattern.

In summary, ESTI-CS outperforms CS, KNN, DT and MSSA in any data loss pattern.

#### 8.4 Performance on ESTI-CS with MAA

In this series, we evaluate the benefit from MAA component for ESTI-CS. The data loss pattern and loss rate setting are the same as the setting in Section 8.2.

In Fig. 5, we illustrate the ER results of ESTI-CS-MAA algorithm and compare them with ESTI-CS in the case of two attributes under random loss pattern. In every dataset, two attributes temperature and light are reconstructed together by ESTI-CS-MAA. It can be seen that the reconstruction accuracy of ESTI-CS-MAA is universally better than that of ESTI-CS in three real WSN datasets.

Figs. 5a and 5b show ESTI-CS-MAA is slightly better than ESTI-CS in Intel Indoor dataset. In face of 90 percent data loss, ESTI-CS-MAA improves the ER by 2 percent in Indoor-Temp shown in Fig. 5a and 3 percent shown in Indoor-Light in Fig. 5b compared with ESTI-CS.

In GreenOrbs dataset, ESTI-CS-MAA performs better than ESTI-CS. The MAA-enabled algorithm outperforms

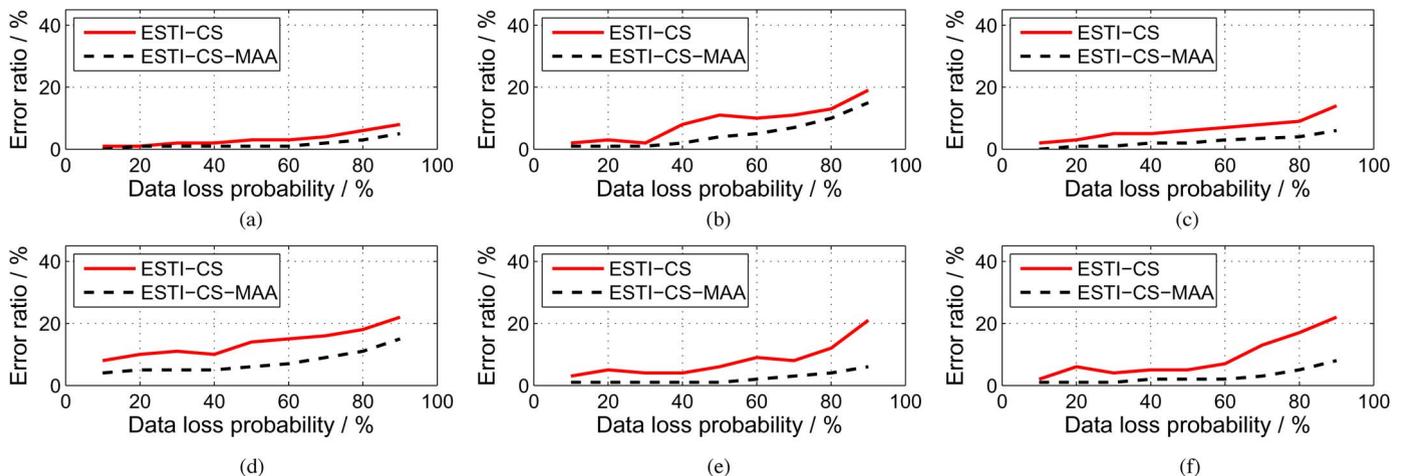


Fig. 5. Error ratio performance of ESTI-CS with MAA in the basic data loss pattern: element random loss. (a) Indoor temp. (b) Indoor light. (c) Forest temp. (d) Forest light. (e) Ocean temp. (f) Ocean light.

baseline ESTI-CS by 7 percent in Forest-Temp shown in Fig. 5c and 6 percent in Forest-Light shown in Fig. 5d.

The results of ESTI-CS-MAA are significantly better than those of ESTI-CS in OceanSense dataset. As shown in Figs. 5e and 5f, 14 percent and 12 percent ER are improved by ESTI-CS-MAA respectively in Ocean-Temp and in Ocean-Light.

Recall the correlation analysis of Fig. 2, the correlation between temperature and light in OceanSense is higher than that in GreenOrbs and even higher than that in Intel Indoor. We summarize that MAA can further improve the ESTI-CS on reconstruction accuracy when multi-attribute correlation exists. Moreover, the higher the correlation is, the more improvement the MAA provides.

## 9 CONCLUSION

In this paper, we studied the data loss and reconstruction problem in WSNs. We verified the massive data loss in real datasets and modeled the special data loss patterns of WSNs. Then, we mined the low-rank, spatial, temporal, and correlation features from WSN datasets. By drawing on these observations, we designed ESTI-CS with MAA algorithm to estimate the missing data. The proposed algorithm combines the benefits of compressive sensing, environmental space-time, and multi-attribute correlation features. Trace-driven experiments illustrated that ESTI-CS outperforms existing interpolation methods.

## ACKNOWLEDGMENT

This research was supported by NSFC under Grants No. 61303202, No. 61073158, No. 61100210, STCSM Project No. 12dz1507400, Doctoral Program Foundation under grant No. 20110073120021, FQRNT grant 131844, Singapore-MIT IDC IDD61000102a, SUTD-ZJU/RES/03/2011, and NRF2012EWT-EIRP002-045. A short version [14] of this paper is published in IEEE INFOCOM'13.

## REFERENCES

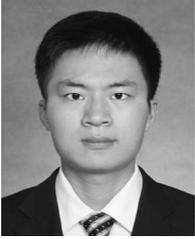
- [1] Intel Indoor Data Trace. [Online]. Available: <http://www.select.cs.cmu.edu/data/labapp3/index.html>
- [2] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive Wireless Sensing," in *Proc. ACM IPSN*, 2006, pp. 134-142.
- [3] M. Balazinska, A. Deshpande, M.J. Franklin, P.B. Gibbons, J. Gray, S. Nath, M. Hansen, M. Liebhold, A. Szalay, and V. Tao, "Data Management in the Worldwide Sensor Web," *IEEE Pervasive Comput.*, vol. 6, no. 2, pp. 30-40, Apr.-June 2007.
- [4] C.S. Burrus, R.A. Gopinath, H. Guo, J.E. Odegard, and I.W. Selesnick, *Introduction to Wavelets and Wavelet Transforms: A Primer*. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.
- [5] E.J. Candés, J. Romberg, and T. Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489-509, Feb. 2006.
- [6] E.J. Candés and T. Tao, "Near-Optimal Signal Recovery from Random Projections: Universal Encoding Strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406-5425, Dec. 2006.
- [7] T. Cover and P. Hart, "Nearest Neighbor Pattern Classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21-27, Jan. 1967.
- [8] D.L. Donoho, "Compressed Sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.
- [9] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least Angle Regression," *Ann. Stat.*, vol. 32, no. 2, pp. 407-499, 2004.
- [10] S. Floyd and V. Jacobson, "Random Early Detection Gateways for Congestion Avoidance," *IEEE/ACM Trans. Netw.*, vol. 1, no. 4, pp. 397-413, Aug. 1993.
- [11] T. He, S. Krishnamurthy, J.A. Stankovic, T. Abdelzaher, L. Luo, R. Stoleru, T. Yan, L. Gu, J. Hui, and B. Krogh, "Energy-Efficient Surveillance System Using Wireless Sensor Networks," in *Proc. ACM MOBISYS*, 2004, pp. 270-283.
- [12] K. Jain, J. Padhye, V.N. Padmanabhan, and L. Qiu, "Impact of Interference on Multi-Hop Wireless Network Performance," in *Proc. ACM MOBICOM*, 2003, pp. 66-80.
- [13] L. Kong, D. Jiang, and M.-Y. Wu, "Optimizing the Spatio-Temporal Distribution of Cyber-Physical Systems for Environment Abstraction," in *Proc. IEEE ICDCS*, 2010, pp. 179-188.
- [14] L. Kong, M. Xia, X.-Y. Liu, M.-Y. Wu, and X. Liu, "Data Loss and Reconstruction in Sensor Networks," in *Proc. IEEE INFOCOM*, 2013, pp. 1654-1662.
- [15] L. Kong, M. Zhao, X.-Y. Liu, J. Lu, Y. Liu, M.-Y. Wu, and W. Shu, "Surface Coverage in Sensor Networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 1, pp. 234-243, Jan. 2014.
- [16] A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E.D. Kolaczyk, and N. Taft, "Structural Analysis of Network Traffic Flows," in *Proc. ACM SIGMETRICS*, 2004, pp. 61-72.
- [17] M.G. Lawrence, "The Relationship Between Relative Humidity and the Dewpoint Temperature in Moist Air: A Simple Conversion and Applications," *Bull. Amer. Meteorol. Soc.*, vol. 86, no. 2, pp. 225-233, Feb. 2005.
- [18] P. Li, T.J. Hastie, and K.W. Church, "Very Sparse Random Projections," in *Proc. ACM SIGKDD*, 2006, pp. 287-296.
- [19] Z. Li, Y. Zhu, H. Zhu, and M. Li, "Compressive Sensing Approach to Urban Traffic Sensing," in *Proc. IEEE ICDCS*, 2011, pp. 889-898.
- [20] X.-Y. Liu, K.-L. Wu, Y. Zhu, L. Kong, and M.-Y. Wu, "Mobility Increases the Surface Coverage of Distributed Sensor Networks," *Comput. Netw.*, vol. 57, no. 11, pp. 2348-2363, Aug. 2013.
- [21] Y. Liu, Y. He, M. Li, J. Wang, K. Liu, L. Mo, W. Dong, Z. Yang, M. Xi, J. Zhao, and X.-Y. Li, "Does Wireless Sensor Network Scale? A Measurement Study on GreenOrbs," in *Proc. IEEE INFOCOM*, 2011, pp. 873-881.
- [22] C. Luo, F. Wu, J. Sun, and C.W. Chen, "Compressive Data Gathering for Large-Scale Wireless Sensor Networks," in *Proc. ACM MOBICOM*, 2009, pp. 145-156.
- [23] L. Mo, Y. He, Y. Liu, J. Zhao, S.-J. Tang, X.-Y. Li, and G. Dai, "Canopy Closure Estimates with GreenOrbs: Sustainable Sensing in the Forest," in *Proc. ACM SENSYS*, 2009, pp. 99-112.
- [24] S. Rallapalli, L. Qiu, Y. Zhang, and Y.-C. Chen, "Exploiting Temporal Stability and Low-Rank Structure for Localization in Mobile Networks," in *Proc. ACM MOBICOM*, 2010, pp. 161-172.
- [25] B. Recht, M. Fazel, and P.A. Parrilo, "Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471-501, Aug. 2010.
- [26] C.-C. Shen, C. Srisathapornphat, and C. Jaikaeo, "Sensor Information Networking Architecture and Applications," *IEEE Pers. Commun.*, vol. 8, no. 4, pp. 52-59, Aug. 2001.
- [27] W. Wang, M. Garofalakis, and K. Ramchandran, "Distributed Sparse Random Projections for Refinable Approximation," in *Proc. ACM IPSN*, 2007, pp. 331-339.
- [28] G. Werner-Allen, K. Lorincz, J. Johnson, J. Lees, and M. Welsh, "Fidelity and Yield in a Volcano Monitoring Sensor Network," in *Proc. USENIX OSDI*, 2006, pp. 381-396.
- [29] A. Woo and D.E. Culler, "A Transmission Control Scheme for Media Access in Sensor Networks," in *Proc. ACM MOBICOM*, 2001, pp. 221-235.
- [30] Z. Yang, M. Li, and Y. Liu, "Sea Depth Measurement with Restricted Floating Sensors," in *Proc. IEEE RTSS*, 2007, pp. 469-478.
- [31] Y. Zhang, M. Roughan, W. Willinger, and L. Qiu, "Spatio-Temporal Compressive Sensing and Internet Traffic Matrices," in *Proc. ACM SIGCOMM*, 2009, pp. 267-278.
- [32] H. Zhu, Y. Zhu, M. Li, and L.M. Ni, "SEER: Metropolitan-Scale Traffic Perception Based on Lossy Sensory Data," in *Proc. IEEE INFOCOM*, 2009, pp. 217-225.



Linghe Kong received the BE degree in automation from Xidian University, in 2005, the Dipl Ing degree in telecommunication with TELECOM SudParis, in 2007, and the PhD degree in computer science with Shanghai Jiao Tong University, in 2012. He is currently a Research Assistant Professor in Shanghai Jiao Tong University and a post-doctoral researcher in Singapore University of Technology and Design. His research interests include wireless sensor networks and mobile computing. He is a member of the IEEE.



**Mingyuan Xia** received the BE degree in computer science and engineer from Shanghai Jiao Tong University, China, in 2011, and is now a PhD candidate in the School of Computer Science at McGill University, Canada. His research interests include wireless sensor networks and operation systems.



**Xiao-Yang Liu** received the BE degree in computer science in Huazhong University of Science and Technology, China, in 2010. He is currently pursuing the PhD degree in Department of Computer Science and Engineer in Shanghai Jiao Tong University. His research interests are in the area of wireless sensor networks.



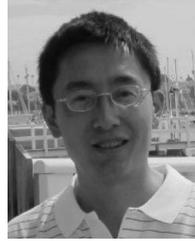
**Guangshuo Chen** received the BE degree in electronic and information engineering from Shanghai Jiao Tong University, China, in 2011, and is now an academic master degree candidate in computer science from Shanghai Jiao Tong University. His research interests include Wireless Sensor Networks (WSNs), Body Area Networks and Vehicular Networks. At present, he focus on the data recovery problem in WSNs.



**Yu (Jason) Gu** is currently an assistant professor in the Pillar of Information System Technology and Design at the Singapore University of Technology and Design. He received the PhD degree from the University of Minnesota, Twin Cities, in 2010. His research includes Networked Embedded Systems, Wireless Sensor Networks, Cyber-Physical Systems, Wireless Networking, Real-time and Embedded Systems, Distributed Systems, Vehicular Ad-Hoc Networks and Stream Computing Systems. He is a member of the IEEE.



**Min-You Wu** received the MS degree from the Graduate School of Academia Sinica, Beijing, China, in 1981 and the PhD degree from Santa Clara University, Santa Clara, CA, in 1984. He is a professor in the Department of Computer Science and Engineering at Shanghai Jiao Tong University and a research professor with the University of New Mexico. He serves as the Chief Scientist at Grid Center of Shanghai Jiao Tong University. His research interests include grid computing, wireless networks, sensor networks, multimedia networking, parallel and distributed systems, and compilers for parallel computers. He is a Senior Member of the IEEE.



**Xue Liu** received the BS degree in mathematics and the MS degree in automatic control both from Tsinghua University, China, and the PhD degree in computer science from the University of Illinois at Urbana-Champaign, in 2006. He is an associate professor in the School of Computer Science at McGill University. He has also worked as the Samuel R. Thompson associate professor in the University of Nebraska-Lincoln and HP Labs in Palo Alto, California. His research interests include computer networks and communications, smart grid, real-time and embedded systems, cyber-physical systems, data centers, and software reliability. He is a member of the IEEE.

▷ For more information on this or any other computing topic, please visit our Digital Library at [www.computer.org/publications/dlib](http://www.computer.org/publications/dlib).