When Sensor Meets Tensor: Filling Missing Sensor Values Through a Tensor Approach

Wenjie Ruan
School of Computer Science
The University of Adelaide
Adelaide, Australia
wenjie.ruan@adelaide.edu.au

Peipei Xu
School of Electronic Engineering, UESTC
Chengdu, China
peipei.xu6@gmail.com

Quan Z. Sheng
School of Computer Science
The University of Adelaide
Adelaide, Australia
michael.sheng@adelaide.edu.au

Nguyen Khoi Tran, Nickolas J.G. Falkner
School of Computer Science
The University of Adelaide
nguyen.tran@adelaide.edu.au

Xue Li
School of IT and EE
The University of Queensland
Brisbane, Australia
xueli@itee.uq.edu.au

Wei Emma Zhang
School of Computer Science
The University of Adelaide
Adelaide, Australia
wei.zhang01@adelaide.edu.au

ABSTRACT

In the era of the *Internet of Things*, enormous number of sensors have been deployed in different locations, generating massive time-series sensory data with geo-tags. However, such sensory readings are easily missing due to various reasons such as the hardware malfunction, connection errors, and data corruption. This paper focuses on this challenge—how to accurately yet efficiently recover the missing values for corrupted time-series sensor data with geo-stamps. In this paper, we formulate the time-series sensor data as a 3-order *tensor* that naturally preserves sensors' temporal and spatial dependencies. Then we exploit its low-rank and sparse-noise structures by drawing upon recent advances in Robust Principal Component Analysis (RPCA) and tensor completion theory. The main novelty of this paper lies in that, we design a highly efficient optimization method that combines the alternating direction method of multipliers and accelerated proximal gradient to recover the data tensor. Besides testing our method using the synthetic data, we also design a real-world testbed by passive RFID (Radio-Frequency IDentification) sensors. The results demonstrate the effectiveness and accuracy of our approach.

1. INTRODUCTION

With rapid development of the IoT (Internet of Things) and constant growth of accessibility of sensors, the amount of sensory data increases exponentially [3]. Such sensor data normally are collected as time-series in an on-line and parallel manner [12]. However, in practice, sensor readings are usually missing due to the unexpected hardware failures (*e.g.,* power outages) or communication conflicts [6]. The missing data will not only greatly decrease the real-time monitoring performance, but also compromise the accuracy of back-end data analysis such as data predication, inference and visualization. Besides the data loss, the observed sensory data are also easily affected by the environment, making the accurate data recovery even more challenging [6].

CIKM'16 , October 24-28, 2016, Indianapolis, IN, USA

DOI: http://dx.doi.org/10.1145/2983323.2983900

The core problem of the missing value recovery lies on how to model the relationship between the known entries and the unknown ones. The most widely used and straightforward technique is filtering or regression algorithms that estimates the missing values according to their local temporal/spatial interdependence, such as Median Filter, Exponential Moving Average [7], Kalman Filtering [5], or regression methods with various complexities [10]. However, such intuitive approaches unavoidably ignore the global correlations of data (*e.g.,* in some circumstances, the missing value may depend on the entries that are far away), leading to inaccurate estimation in some circumstances.

To solve this issue, some researchers treat the sensory data as a matrix and propose various matrix completion/recovery methods to estimate the missing values by capturing their inherent low-rank structure [2, 8]. Those methods usually model time-dependent sensory data as a matrix $S = [s_1, s_2, ..., s_T]$ where vector $s_k \in \mathbb{R}^N$ represents the N-dimension sensor data at timestamp k, and only partial values in S can be observed, represented by $S_{i,j}$ where $(i, j) \in \Omega$ is the subset of known entries. So matrix completion based methods are, in principle, to recover the unknown entries by solving $\min_X \{\text{rank}(X) | \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(S)\}$ [8], where $X \in \mathbb{R}^{N \times T}$ indicates recovered data matrix and $\mathcal{P}_\Omega$ is the project operator that means only entries in Ω are observed. Because it is a *NP-hard* problem (untraceable), many researchers have proposed various convex relaxation approaches trying to solve the problem [2, 8]. Also, several robust low-rank matrix completion methods are recently proposed to deal with the case that the observed data matrix S are corrupted by noise [1]. Though promising, matrix completion based models regard geo/time-dependent sensor data as a 2-D matrix that still takes advantage of the temporal information.

Recently, tensor completion/recovery has been emerged as an active research topic since Liu *et al.* first proposed a tensor completion method for estimating missing visual data [9]. Many pioneering similar works appear, such as low-rank tensor completion via convex optimization [15], tensor completion based on Tucker rank [13], *etc.* However, most of the related works are mainly attempted in computer vision and machine learning community, and they are mostly based on an assumption that the known entries in the data tensor are accurate without noise [13, 15]. This gap strongly motivates us to answer the question—*can we recover the missing values from noisy sensory data using tensor completion theory?* To the best of our knowledge, this paper is one of the first few works to do so. Comparing to the matrix formulation that only preserves the time-series information, tensor-based model naturally captures both temporal and spatial interdependent relationships of sensory readings.

Nevertheless, applying this high-level idea into practical geo-sensory data requires addressing several challenges. Similar to the robust matrix recovery formulation [1], *robust tensor completion* can be formulated to solve the problem[1] $\min_{\mathcal{X},\mathcal{E}}\{\text{rank}(\mathcal{X})+\lambda\|\mathcal{E}\|_0 \mid \mathcal{P}_\Omega(\mathcal{X}+\mathcal{E})=\mathcal{P}_\Omega(\mathcal{S})\}$. As a result, the first challenge is that, the above optimization problem is NP-hard and thus untraceable [9]. For the matrix version, we can replace rank($\mathcal{X}$) and L_0-norm by its tightest convex surrogate (*i.e.,* trace norm) and L_1-norm accordingly, making it traceable (convex optimization problem). But how to define the rank's convex surrogate for a tensor is a challenging issue. Secondly, even we can define an effective convex surrogate and make the problem traceable, how to efficiently solve the convex optimization problem while guaranteeing its convergence needs careful design.

To address above issues, we generalize the idea of adopting trace norm in matrix completion into the tensor, replacing the rank regularization term by the sum of the trace norm of each tensor-mode unfolding (see details in Sec. 3). Furthermore, to optimize the objective function, we first apply a variable-splitting trick by introducing auxiliary tensor variables to decouple the interdependency of different tensor-modes, then we design an efficient yet convergence-guaranteed optimization method by drawing upon recent advances of ADMM (Alternating Direction Method of Multipliers) [15] and accelerate proximal gradient[14] (see details in Sec. 3 and Algorithm 1). In a nutshell, our main contributions are summarized as follows:

- We develop a robust tensor completion method to recover the missing values for geo-dependent noisy time-series sensor data. To the best of our knowledge, this paper is a very first few attempts to do so.

- We design an efficient ADMM based optimization scheme to solve the robust tensor completion problem with a guarantee of convergence to optimal solution.

2. PROBLEM FORMULATION AND NOTATIONS

First, we mathematically define our target problem. Assuming that we have $M \times N$ sensors deployed in different spatial areas and collect (noisy) sensor readings[2] for T timestamps, we then can formulate it as a 3-order tensor $\mathcal{S} \in \mathbb{R}^{M\times N\times T}$ and $\mathcal{S} = \mathcal{X} + \mathcal{E}$ where X represent the true sensor readings (without noise) and $\mathcal{E}$ means the added noise. We use the projection operator $\mathcal{P}_\Omega(\mathcal{S}) : \mathbb{R}^{M\times N\times T} \to \mathbb{R}^K$ that indicates the K observed sensor readings $S_{i,j,k}$ where the index $(i,j,k) \in \Omega$,mapping a tensor to a vector. Formally, this paper therefore aims to solve the following *Corrupted Sensor Value Recovery* problem:

PROBLEM 1 (CORRUPTED SENSOR VALUE RECOVERY). *To accurately recover the true sensor readings $\mathcal{X}$ and additive noise $\mathcal{E}$ given a partially observed data tensor $\mathcal{S}_\Omega$.*

Throughout this paper, we represent scalars, vectors and matrices by lowercase letters *e.g.,* a, bold lowercase letters such as $\mathbf{a}$, and upper letters A. Tensors of d-order/dimension are written by calligraphic letters like $\mathcal{A} \in \mathbb{R}^{I_1\times I_2\times\cdots\times I_d}$, whose elements are represented by $a_{i_1\cdots i_k\cdots i_d} \in \mathbb{R}$ and $1 \le i_k \le I_k, 1 \le k \le d$. Thus a vector can be seen as a 1-order tensor and a matrix can be seen as a 2-order tensor.

The unfolding (matricization or flattening) is the operation that transforms a tensor into a matrix. The mode-k unfolding of $\mathcal{A}$ is denoted by $unfold(\mathcal{A},k) = A_{(k)} \in \mathbb{R}^{I_k \times \Pi_{i\neq k} I_i}$, *i.e.,* the row of the matrix $A_{(k)}$ are determined by the k-th component of the tensor $\mathcal{A}$, whereas all the remaining components form its column. The opposite operation "*folding*" is defined as $fold(A_{(k)},k) = \mathcal{A}$. The inner product of two tensors with identical size is computed by $<\mathcal{A},\mathcal{B}> := \sum_{i_1,i_2,\cdots,i_d} a_{i_1 i_2\cdots i_d} b_{i_1 i_2 \cdots i_d}$. The Frobenius norm of $\mathcal{A}$ is defined as $\|\mathcal{A}\|_F := (\sum_{i_1,i_2,\cdots,i_d} |a_{i_1 i_2\cdots i_d}|^2)^{\frac{1}{2}}$. The multiplication of a tensor $\mathcal{A} \in \mathbb{R}^{I_1\times I_2\times\cdots\times I_d}$ with a matrix $U \in \mathbb{R}^{J\times I_k}$ in mode-k is denoted by $\mathcal{A} \times_k U \in \mathbb{R}^{I_1\times\cdots\times I_{k-1}\times J\times I_{k+1}\times\cdots\times I_d}$.

3. ROBUST TENSOR COMPLETION

Being similar to matrix completion, Problem 1 can be formulated as solving a low-rank minimization problem.

$$\min_{\mathcal{X},\mathcal{E}} \ \text{rank}_{Tucker}(\mathcal{X}) + \lambda||\mathcal{E}||_0 \quad \text{s.t. } \mathcal{P}_\Omega(\mathcal{X}+\mathcal{E}) = \mathcal{P}_\Omega(\mathcal{S}) \qquad (1)$$

where $\text{rank}_{Tucker}(\mathcal{S})$ is the Tucker-rank of a tensor[4]. Similar to matrix completion, this problem is NP-hard. Thus, to make it tractable, we replace Tucker rank by its *convex surrogate* and use L_1-norm instead of L_0-norm as an optimization problem: $\min_{\mathcal{X},\mathcal{E}}\{\text{ConSurro}(\mathcal{X}) + \lambda||\mathcal{E}||_1 \mid \text{s.t. } \mathcal{P}_\Omega(\mathcal{X}+\mathcal{E}) = \mathcal{P}_\Omega(\mathcal{S})\}$.

Then the first issue is how to define the convex surrogate of a tensor. For a matrix, the trace norm $\|.\|_*$ is the tightest convex envelop of its rank, used as the convex surrogate. Thus, the idea can be generalized into the high-order tensor, defining its trace norm as the sum of the trace norms [9] of the mode-i unfolding in tensor $\mathcal{X}$, *i.e.,* $\sum_i \|X_{(i)}\|_*$. Equation 1 can be therefore transformed into a convex problem:

$$\begin{aligned} \min_{\mathcal{X},\mathcal{E}} \quad & \sum_{i=1}^{3} \|X_{(i)}\|_* + \lambda||\mathcal{E}||_1 \\ \text{s.t.} \quad & \mathcal{P}_\Omega(\mathcal{X}+\mathcal{E}) = \mathcal{P}_\Omega(\mathcal{S}) \end{aligned} \qquad (2)$$

To solve Eqn. 2, we introduce an Alternating Direction Method of Multipliers [15] that is very efficient in dealing with convex optimization problems by breaking them into smaller pieces, each of which are then easier to handle. However, the trace norm of each mode unfolding $\|X_{(i)}\|_*, (i = 1,2,3)$ shares the same values in data tensor $\mathcal{S}$ and cannot be optimized independently so that existing ADMM cannot directly be applied to our problem. Hence, we split these interdependent terms by introducing auxiliary variables $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3$, so that they can be solved independently. Specifically, we reformulate Eqn. 2 as

$$\begin{aligned} \min_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{X}_3,\mathcal{E}} \quad & \sum_{i=1}^{3} \|X_{i,(i)}\|_* + \lambda\|\mathcal{E}\|_1 \\ \text{s.t.} \quad & \mathcal{P}_\Omega(\mathcal{X}_i+\mathcal{E}) = \mathcal{P}_\Omega(\mathcal{S}), \quad i = 1,2,3. \end{aligned} \qquad (3)$$

[1] $\mathcal{X}, \mathcal{E}, \mathcal{S}$ are the recovered data tensor, additive noise tensor and observed data tensor, $\mathcal{P}_\Omega$ means the known entries of tensor

[2] we assume the additive noises are sufficiently sparse relative to the data tensor S

We hence define its *augmented Lagrangian function* $L_{\mu,\Lambda_i} = \sum_{i=1}^{3}\left(\frac{1}{2\mu}\|\mathcal{P}_\Omega(\mathcal{X}_i+\mathcal{E})-\mathcal{P}_\Omega(\mathcal{S})\|^2 - \langle\Lambda_i, \mathcal{P}_\Omega(\mathcal{X}_i+\mathcal{E})-\mathcal{P}_\Omega(\mathcal{S})\rangle\right) + \sum_{i=1}^{3}\|\mathbf{X}_{\mathbf{i},(\mathbf{i})}\|_* + \lambda\|\mathcal{E}\|_\mathbf{1}$. According to ADMM, we first fix $\mathcal{E}$ to optimize $\mathcal{X}_i$ $(i=1,2,3)$ by solving

$$\min_{\mathcal{X}_1,\mathcal{X}_2,\mathcal{X}_3} \sum_{i=1}^{3}\left(\frac{1}{2\mu}\|\mathcal{P}_\Omega(\mathcal{X}_i+\mathcal{E})-\mathcal{P}_\Omega(\mathcal{S})\|^2 - \langle\Lambda_i, \mathcal{P}_\Omega(\mathcal{X}_i+\mathcal{E})-\mathcal{P}_\Omega(\mathcal{S})\rangle\right.$$

$$\left. + \|X_{i,(i)}\|_*\right) \equiv \sum_{i=1}^{3}\left(\mu\|X_{i,(i)}\|_* + \frac{1}{2}\|\mathcal{P}_\Omega(\mathcal{X}_i)-\mathcal{D}_i\|^2\right) \tag{4}$$

where $\mathcal{D}_i = \mathcal{P}_\Omega(\mathcal{S}) - \mathcal{P}_\Omega(\mathcal{E}) + \mu\Lambda_i$. We define the function $f(\mathcal{X}_i) = \frac{1}{2}\|\mathcal{P}_\Omega(\mathcal{X}_i)-\mathcal{D}_i\|^2$ and calculate the gradient $\nabla f(\mathcal{X}_i) = \mathcal{P}_\Omega^*(\mathcal{P}_\Omega(\mathcal{X}_i)-\mathcal{D}_i)$, where $\mathcal{P}_\Omega^*(\cdot)$ means the adjoint operation of $\mathcal{P}_\Omega(\cdot)$ such as $\mathcal{P}_\Omega^*(\mathcal{S}):\mathbb{R}^K \to \mathbb{R}^{M\times N\times T}$. According to Accelerated Proximal Gradient (APG) method [14], we can independently minimize $\mathcal{X}_i$ through iterative optimization to make the final sum minimal. Specifically, we get optimal $\mathcal{X}_i^{(k+1)}$ given $\mathcal{X}_i^{(k)}$ until it converges by solving

$$\min_{\mathcal{X}_i^{(k+1)}} f(\mathcal{X}_i^{(k)}) + \nabla f(\mathcal{X}_i^{(k)})(\mathcal{X}_i^{(k+1)}-\mathcal{X}_i^{(k)}) + \frac{1}{2\eta}\|\mathcal{X}_i^{(k+1)}-\mathcal{X}_i^{(k)}\|^2$$

$$+ \mu\|X_{i,(i)}^{(k+1)}\|_* = \frac{1}{2\eta}\|\mathcal{X}_i^{(k+1)}-\mathcal{X}_i^{(k)}+\eta\nabla f(\mathcal{X}_i^{(k)})\|^2 + \mu\|X_{i,(i)}^{(k+1)}\|_*$$

$$\propto \frac{1}{2}\|\mathcal{X}_i^{(k+1)}-\mathcal{X}_i^{(k)}+\eta\nabla f(\mathcal{X}_i^{(k)})\|^2 + \eta\mu\|X_{i,(i)}^{(k+1)}\|_* \tag{5}$$

To solve Eqn. 5, we first need to define *singular value thresholding operator* for tensor. For matrix, the singular value threshold operator is defined as $\mathcal{T}_\mu(X) := U\,diag(\bar\sigma)V^\intercal$, where $X = U\,diag(\sigma)V^T$ is the singular value decomposition (SVD) and $\bar\sigma := max(\sigma-\mu, 0)$. Similarly, we define the singular value threshold operator for tensor as $\mathcal{T}_{i,\mu}(\mathcal{X}) := \text{fold}(\mathcal{T}_\mu(X_{(i)}), i)$. We thus can calculate the closed-form solution of Eqn. 5 as follows:

$$\mathcal{X}_i^{(k+1)} = \mathcal{T}_{i,\eta\mu}(\mathcal{X}_i^{(k)} - \eta\nabla f(\mathcal{X}_i^{(k)})) \tag{6}$$

In the next, we will optimize $\mathcal{E}$ when fixed $\mathcal{X}_i$ by solving following problem:

$$\min_{\mathcal{E}} \ \lambda\|\mathcal{E}\|_1 + \sum_{i=1}^{3}\left(\frac{1}{2\mu}\|\mathcal{P}_\Omega(\mathcal{X}_i+\mathcal{E})-\mathcal{P}_\Omega(\mathcal{S})-\mu\Lambda_i\|^2\right)$$

$$\propto \mu\lambda\|\mathcal{E}\|_1 + \frac{1}{2}\sum_{i=1}^{3}\|\mathcal{P}_\Omega(\mathcal{E})-\mathcal{M}_i\|^2 \tag{7}$$

where $\mathcal{M}_i = \mathcal{P}_\Omega(\mathcal{S}) + \mu\Lambda_i - \mathcal{P}_\Omega(\mathcal{X}_i)$.

To solve the Eqn. 7, we introduce *Homogeneous Tensor Array*[4] by stacking a set of component tensors of the same size along the first mode as follows:

$$\bar{\mathcal{X}} := (\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3)^\intercal \in \mathbb{R}^{3M\times N\times T}, \tag{8}$$

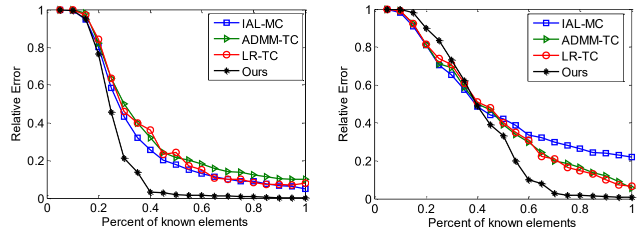
and its corresponding linear operator $\mathcal{C}:\mathbb{R}^{M\times N\times T}\to\mathbb{R}^{3M\times N\times T}$, *i.e.,* $\bar{\mathcal{X}} = \mathcal{C}(\mathcal{X}) \in \mathbb{R}^{3M\times N\times T}$ Then its adjoint operator can be defined as $\mathcal{C}^*:\mathbb{R}^{3M\times N\times T}\to\mathbb{R}^{M\times N\times T}$, such that $\mathcal{X} = \mathcal{C}^*(\bar{\mathcal{X}}) = \sum_{i=1}^{3}\mathcal{X}_i$. As a result, we can rewrite the Eqn. 7 as

$$\min_{\mathcal{E}} \ \mu\lambda\|\mathcal{E}\|_1 + \frac{1}{2}\|\mathcal{C}(\mathcal{P}_\Omega(\mathcal{E}))-\bar{\mathcal{M}}\|^2 \propto \frac{\mu\lambda}{3}\|\mathcal{E}\|_1 + \frac{1}{2}\|\mathcal{P}_\Omega(\mathcal{E}) - \frac{\mathcal{C}^*(\bar{\mathcal{M}})}{3}\|^2$$

$$= \frac{\mu\lambda}{3}\|\mathcal{E}\|_1 + \frac{1}{2}\|\mathcal{P}_\Omega(\mathcal{E}) - \frac{1}{N}\sum_{i=1}^{N}(\mathcal{P}_\Omega(\mathcal{S})+\mu\Lambda_i-\mathcal{P}_\Omega(\mathcal{X}_i))\|^2 \tag{9}$$

where $\bar{\mathcal{M}} = (\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3)^\intercal$, $\mathcal{C}(\mathcal{P}_\Omega(\mathcal{E})) = (\mathcal{P}_\Omega(\mathcal{E}), \mathcal{P}_\Omega(\mathcal{E}), \mathcal{P}_\Omega(\mathcal{E}))^\intercal$.

According to the Iterative Shrinkage Thresholding (IST) algorithm that $\min_\mathbf{y}\{\frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \mu\|\mathbf{y}\|_1\}$ has a closed form solution $\mathbf{y} = \mathcal{S}_\mu(\mathbf{x}) := sign(\mathbf{x})max(|\mathbf{x}|-\mu, 0)$, we can solve Eqn. 9 to get $\mathcal{E} = \mathcal{S}_{\frac{\mu\lambda}{3}}(\frac{1}{3}\sum_{i=1}^{3}(\mathcal{P}_\Omega(\mathcal{S})+\mu\Lambda_i-\mathcal{P}_\Omega(\mathcal{X}_i)))$.



(a) (b)

Figure 1: (a) Relative errors for different observation percentages ($\rho_n = 0.1$, $a = 1$); (b) Relative errors for different observation percentages under noise ($\rho_n = 0.25$, $a = 1$)

Finally, given $\mathcal{X}_i^{(k+1)}$ and $\mathcal{E}_i^{(k+1)}$, we update the Lagrangian multiplier parameter by $\mathbf{\Lambda}_i^{(k+1)} = \mathbf{\Lambda}_i^{(k)} - \frac{1}{\mu}(\mathcal{P}_\Omega(\mathcal{X}_i^{(k+1)}+\mathcal{E}^{(k+1)}) - \mathcal{P}_\Omega(\mathcal{S})$. Note that the above solution is based on the form of 3-order tensor. Extension to higher-order tensor involves fundamentally no new ideas.

4. EXPERIMENTS

We conducted the experiments in MATLAB R2015b on a laptop with a CORE i7 2.40GHz CPU and 8GB RAM. We used the Tensor Toolbox[3] for tensor operations and decompositions. We first conducted a simulation experiment using synthetic data to compare with the state-of-the-art data recovery methods under different data loss percentages and additive noise ratios. Then we designed a real-world experimental testbed using passive Radio-frequency Identification hardware to evaluate the performance of our method. Similar to other data recovery works [9, 8], we adopted the *relative error* $\frac{\|\mathcal{X}-\mathcal{X}_0\|}{\|\mathcal{X}_0\|}$, where $\mathcal{X}, \mathcal{X}_0$ denote the recovered and original data tensor respectively, to evaluate the recovery performance.

We compared our method with four typical methods: *i) M-AF* means the moving averaging interpolation that is the most widely-used method to fill in missing values in time series data; *ii) IAL-MC* is the inexact augmented Lagrangian method that can recover a data matrix of being arbitrarily corrupted [8], it is a matrix-based robust completion method and greatly motivates our work; *iii) LR-TC* is the earliest yet very effective tensor completion method using block coordinate descent optimization [9]; *iv) ADMM-TC* also utilizes ADMM for solving the tensor completion problem under no noise assumption [15].

4.1 Synthetic Data

Similar to the works in [9, 15], we generated a $50\times50\times20$ data tensor with Tucker rank-$(5,5,5)$. We randomly chose a fraction ρ_n of the tensor entries that were polluted by an additive i.i.d. (*i.e.,* independent and identically distributed) noise following uniform distribution $\mathcal{U}(-a, a)$. Then a fraction ρ_o of the corrupted tensor elements were randomly picked as observed values in $\mathcal{S}_\Omega$. In the experiments, we set μ and η as constants for simplicity.

In Fig. 1, we can see that our method has a significantly higher recovery accuracy than the other three algorithms. In our method, there appears to be a phase transition threshold in ρ_o where the relative error of the algorithm improves drastically. This threshold increases with ρ_n: it is about 40% for $\rho_n = 10\%$, and it increases to 70% for $\rho_n = 25\%$.

[3] http://www.sandia.gov/~tgkolda/TensorToolbox/index-2.6.html

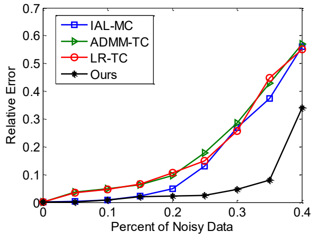


Figure 2: Relative errors for different corruption percentages ($\rho_o = 1$, $a = 1$)

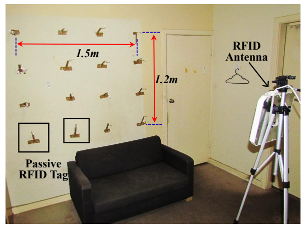


Figure 3: Experimental testbed of RFID sensors

Next, we fixed ρ_o at 100% and varied ρ_n from 1% to 40%. The same set of metrics was reported and plotted in the left graph of Fig 2. Again, our model works better than the other models, which appears to be very susceptible to increasing ρ_n's. There is also a threshold in ρ_n(15% in this case) below, other methods recover the data with higher relative error. The experimental results on the synthetic data suggest that our method can achieve better recovery with partial observations.

4.2 RFID Sensory Data

Passive RFID tags are one of the most frequently-used sensors due to its cheap price and battery-free character [11]. It is widely used to identify and track objects through remotely accessing the electronically stored data. However, since passive RFID tags are powered by radio signals and deliver the data via the weak backscatter signal, they experience severe RSSI reading loss, especially with a high sampling rate or when tag/reader is moving. As a result, how to accurately recover missing RSSI readings is still a challenge, especially for a large-scale RFID usage. To deal with this practical issue as well as test our method, we designed a testbed consisting geo-tagged 4×4 RFID sensor array (see Fig. 3) and collected overall 115,200 RSSIs (Received Signal Strength Indicators) readings (one hour's data). We formulated it into a data tensor with different sizes to simulate various using scenarios (*e.g.,* 4×4, 20×20, 20×40 and 40×40 sensor arrays).

Table 1: Relative errors for different tag-array size under 20% missing values

Sensor Array Size	4×4	20×20	20×40	40×40
MAF	0.2431	0.2462	0.2505	0.257
Our method	0.2481	0.187	0.126	0.091

Table 1 shows the recovery results of our method and MAF (most frequently used in practical RFID system). For small-scaled deployment (*i.e.,* 4×4 sensor array), our method achieves similar performance to MAF. However, our tensor-based method performs significantly better than MAF in a large-scale deployment (*e.g.,* 20×20 sensor array). The lack of performance improvement in 4×4 sensor array mainly lies in the fact that the low-rank structure only exists in mode-3 of data tensor which conflicts our assumption that requires tensor low-rank in all modes. This problem will leave for our future work.

5. CONCLUSION

In this paper, we propose a method for recovering the missing data by using the robust tensor completion. The proposed method can accurately recover the missing values given partial observed corrupted data. Our method assumes that the data tensor is low rank in all unfolding modes, however this maybe not the case in certain applications. We will study this issue in our future work.

6. REFERENCES

[1] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.

[2] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.

[3] L. Da Xu, W. He, and S. Li. Internet of things in industries: a survey. *Industrial Informatics, IEEE Transactions on*, 10(4):2233–2243, 2014.

[4] D. Goldfarb and Z. Qin. Robust low-rank tensor recovery: Models and algorithms. *SIAM Journal on Matrix Analysis and Applications*, 35(1):225–253, 2014.

[5] M. S. Grewal. *Kalman filtering*. Springer, 2011.

[6] H.-P. Hsieh, S.-D. Lin, and Y. Zheng. Inferring air quality for station location recommendation based on urban big data. In *Proc. of the 21th ACM SIGKDD Intl. Conference on Knowledge Discovery and Data Mining*, pages 437–446, 2015.

[7] A. Lawrance and P. Lewis. An exponential moving-average sequence and point process. *Journal of Applied Probability*, pages 98–113, 1977.

[8] Z. Lin, M. Chen, and Y. Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv:1009.5055*, 2010.

[9] J. Liu, P. Musialski, P. Wonka, and J. Ye. Tensor completion for estimating missing values in visual data. In *Computer Vision, 2009 IEEE 12th Intl. Conference on*, pages 2114–2121, 2009.

[10] A. M. Norcia, M. Clarke, and C. W. Tyler. Digital filtering and robust regression techniques for estimating sensory thresholds from the evoked potential. *IEEE Engineering in Medicine and Biology Magazine*, 4(4):26–32, 1985.

[11] W. Ruan. Unobtrusive human localization and activity recognition for supporting independent living of the elderly. In *Proceedings of 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pages 1–3, 2016.

[12] Y. Sakurai, Y. Matsubara, and C. Faloutsos. Mining and forecasting of big time-series data. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 919–922. ACM, 2015.

[13] M. Signoretto, R. Van de Plas, B. De Moor, and J. A. Suykens. Tensor versus matrix completion: a comparison with application to spectral data. *Signal Processing Letters, IEEE*, 18(7):403–406, 2011.

[14] K.-C. Toh and S. Yun. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pacific Journal of Optimization*, 6(615-640):15, 2010.

[15] R. Tomioka, K. Hayashi, and H. Kashima. Estimation of low-rank tensors via convex optimization. *arXiv preprint arXiv:1010.0789*, 2010.