Weighted Tensor Robust Principal Component Analysis to Reduce Noise in Wireless Sensor Networks

1st Hanh Hong-Phuc Vo
Department of Information Communication Convergence Technology
Soongsil University
Seoul, Korea
hanhvo@soongsil.ac.kr

2nd Myungsik Yoo
School of Electronics Engineering
Soongsil University
Seoul, Korea
myoo@ssu.ac.kr

***Abstract*—This paper introduces a solution for reducing noise data in wireless sensor networks employing a weighted tensor robust principal component analysis (WTRPCA) approach. The proposed method effectively separates the data into normal and abnormal tensors, facilitating identifying corruption and preserving key data components. WTRPCA is specifically designed to optimize this reduced process, and experimental results demonstrate its superior accuracy compared to existing approaches.**

***Index Terms*—WSNs, Tensor robust PCA, reduce noise, noise, weight**

I. INTRODUCTION

Wireless sensor networks (WSNs) generate a significant amount of complex data. However, traditional methods that use vectors or matrices struggle to handle this data effectively. They fail to maintain the underlying structure and correlations in the data, leading to a dimension problem. In contrast, tensors, which are multidimensional arrays, provide a better solution for large-scale and diverse WSN data. They can preserve linear and multilinear relationships in the data and offer a more compact representation. Tensor decomposition algorithms are more efficient than vector decomposition methods when it comes to extracting information from high-dimensional data, thanks to their natural compatibility. By applying tensor decomposition techniques like Tucker, CP, higher-order SVD, t-SVD, and TT decomposition, it is possible to eliminate redundancies caused by temporal, spatial, and multi-attribute correlations in WSN data without compromising its inherent structure. This decomposition is crucial for representing data and developing algorithms based on tensors, especially given the vast amount of data and unknown parameters involved. Accurately reconstructing acquired data at the sink node in WSN data processing poses a major challenge, as harsh environmental conditions can introduce corruption, such as outliers, noise, or missing data. Traditional tensor RPCA algorithms [1] may require adjustments because they treat all singular values equally, disregarding the varying importance levels of different signal information. The primary contribution of this study is as follows:

1) Using TRPCA (Tensor Robust Principal Component Analysis), it is possible to separate data that has been corrupted with noise into two distinct components. The first component is a low-rank tensor that represents the normal data, while the second is a sparse noise tensor that represents the noisy data.
2) To preserve the principal components of WSN data and enforce the low-rank structure, the Weighted tensor nuclear norm (WTNN) is utilized as a constraint.

II. METHOD

A. Basic of RPCA

RPCA (Robust Principal Component Analysis) is an improvement over PCA with strong reduction guarantees. It decomposes a matrix $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$ into a low-rank matrix $\mathbf{L}$ and a sparse matrix $\mathbf{S}$. The goal is to recover $\mathbf{L}$ and $\mathbf{S}$ by minimizing the nuclear norm of $\mathbf{L}$ and the ℓ_1-norm of $\mathbf{S}$, subject to the constraint $\mathbf{X} = \mathbf{L} + \mathbf{E}$. The following convex optimization model is used to recover $\mathbf{L}$ and $\mathbf{S}$:

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1 \text{ , s.t. } X = L + S, \tag{1}$$

With $\|\cdot\|_*$ indicates the nuclear norm of a matrix, $\|\cdot\|_1$ the ℓ_1-norm, and λ is a positive weight-adjusting parameter.

B. TRPCA

TRPCA (Tensor Robust Principal Component Analysis) extends RPCA to handle multi-dimensional data while preserving the tensor structure. It has found successful applications in diverse domains such as background subtraction, image processing, and image reconstruction, specifically addressing noise corruption challenges in WSNs. In Figure 1, TRPCA is a method designed to restore low-rank tensors that sparse noise has corrupted. The expanded form of tensor-based RPCA can be defined as follows:

$$\min_{\mathcal{L},\mathcal{S}} ||\mathcal{L}||_\star + \lambda ||\mathcal{S}||_1, \quad \text{s.t.} \quad \mathcal{X} = \mathcal{L} + \mathcal{S}, \tag{2}$$

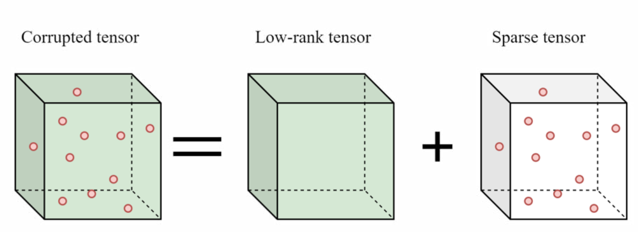


Fig. 1: Low-rank and Sparse components decomposition from noisy observation of TRPCA.

III. The proposed model

To address the restrictions of using a fixed threshold for each singular value in the reducing noise problem of WSNs, the weighted tensor nuclear norm is employed using tensor singular value decomposition (t-SVD) [4] to recover the missing values. The weighted tensor nuclear norm of the low-rank tensor $\mathcal{L}$ is the totality of weighted singular values in all frontal slices of tensor data, where larger singular values receive less shrinkage to preserve essential data components. [2] proved that WTNNM corporates to TRPCA help improve the performance reconstruction of corrupted data in the computer vision field. The following is a solution to the low-rank optimization problem:

$$\mathcal{L}_{i+1} = \arg\min_{\mathcal{L}} \left(||\mathcal{L}||_{*,\omega} + \frac{\mu_i}{2} \left\| \mathcal{L} - \mathcal{X} + \mathcal{S}_i - \frac{\mathcal{Y}_i}{\mu_i} \right\|_{\mathrm{F}}^2 \right) \quad (3)$$

$$\mathcal{S}_{i+1} = \arg\min_{\mathcal{S}} \left(\lambda ||\mathcal{S}||_1 + \frac{\mu_i}{2} \left\| \mathcal{L}_{i+1} - \mathcal{X} + \mathcal{S} - \frac{\mathcal{Y}_i}{\mu_i} \right\|_{\mathrm{F}}^2 \right) \quad (4)$$

$$\mathcal{Y}_{i+1} = \mathcal{Y}_i + \mu_i(\mathcal{L}_{i+1} + \mathcal{S}_{i+1} - \mathcal{X}) \quad (5)$$

Equation 2 optimization model of the WRTPCA is commonly solved using the Alternating Direction Method of Multipliers (ADMM) method. The ADMM algorithm includes a step size, denoted as μ_k, in updating the dual variable $\mathcal{Y}$ that represents the augmented Lagrange penalty parameter. Algorithm 1 outlines the procedure for utilizing ADMM to separate the low-rank data and sparse noise tensors from the corrupted data tensor.

In the Algorithm 1, $\mathcal{O}$ indicates an $n_1 \times n_2 \times n_3$ with all its elements set to zero. To stop the ADMM algorithm, the convergence criterion is defined as:$\|\mathcal{L}_{i+1} - \mathcal{L}_i\|_F \leq \epsilon$,

$\|\mathcal{S}_{i+1} - \mathcal{S}_i\|_F \leq \epsilon$, $\|\mathcal{X} - \mathcal{L}_{i+1} - \mathcal{S}_{i+1}\|_F \leq \epsilon$. The Frobenius norm of a tensor data refers to the Euclidean norm, also known as the ℓ_2-norm, which is a measure of the magnitude or size of the tensor. It is calculated by sum is taken over all indices (i, j, k) of the tensor, $\|\mathcal{X}\|_F = \sqrt{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \mathcal{X}_{i,j,k}^2}$. The operator $\text{prox}_{l_1,\tau}(\mathcal{X})$ indicates the proximal operator of the ℓ_∞-norm of $\mathcal{X}$, given by $\text{prox}_{l_1,\lambda/\mu_i}(\mathcal{X}) = \max\left((|\mathcal{X}_{ijk}| - \tau), 0\right) \cdot \text{sign}(\mathcal{X}_{ijk})$, applied to all elements $\mathcal{X}_{ijk}$ in the tensor.

In the referenced Algorithm 1, the WTSVT operator is defined (Algorithm 2). In the context of Algorithm 2, $diag$

Algorithm 1 Weighted tensor robust principal component utilizing ADMM

Input: $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with random noise, weight vector ω, $\lambda = \frac{1}{\sqrt{\max(n_1,n_2) \times n_3}}$
Initialize: $\mathcal{L}_0 = \mathcal{O}S_0 = \mathcal{O}$, $W_0 = \mathcal{O}$, $\mu_0 = 1 \times 10^{-1}$, $\mu_{\max} = 1 \times 10^{10}$, $\epsilon = 1 \times 10^{-8}$, $\rho = 1.1$
while not converged **do**
  Update $\mathcal{L}_{i+1} = WTSVT(\mathcal{X} - \mathcal{S}_k - \frac{\mathcal{Y}_i}{\mu_i}, \omega, \mu_i)$
  Update $\mathcal{S}_{i+1} = \text{prox}_{l_1, \frac{\lambda}{\mu_i}}(\mathcal{X} - \mathcal{L}_{i+1} - \frac{\mathcal{Y}_i}{\mu_i})$
  Update $\mathcal{Y}_{i+1} = \mathcal{Y}_i + \mu_i(\mathcal{L}_{i+1} + \mathcal{S}_{i+1} - \mathcal{X})$
  Update $\mu_{i+1} = \min(\rho\mu_i, \mu_{\max})$
  Verify the convergence criteria $\|\mathcal{L}_{i+1} - \mathcal{L}_i\|_F \leq \epsilon$ or $\|\mathcal{S}_{i+1} - \mathcal{S}_i\|_F \leq \epsilon$ or $\|\mathcal{X} - \mathcal{L}_{i+1} - \mathcal{S}_{i+1}\|_F \leq \epsilon$
end while
Output: $\mathcal{L}$, $\mathcal{S}$

Algorithm 2 Weighted tensor singular value thresholding method

Input: Data tensor $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}, \omega, \mu$
$\mathcal{R} = \text{fft}(\mathcal{M}, [], 3)$
for $i = 1$ to n_3 **do**
  [U,S,V] = *SVD*($\mathcal{R}$(:,:,i))
  $diagS = diag(\mathbf{S}) - \frac{\omega}{\mu}$
  ddiagS = $diag(diagS)$
  $\hat{\mathcal{U}}(:,:,i)$ = **U**
  $\hat{\mathcal{S}}(:,:,i)$ = **ddiagS**
  $\hat{\mathcal{V}}(:,:,i)$ = **V**
end for
$U = ifft(\hat{\mathcal{U}}, [], 3)$,$S = ifft(\hat{\mathcal{S}}, [], 3)$, $V = ifft(\hat{\mathcal{V}}, [], 3)$
Output: $\mathcal{U}$, $\mathcal{S}$, $\mathcal{V}$

refers to the diagonal elements of a matrix. SVD represents Singular Value Thresholding, a method that breaks down a matrix into three matrices: a left singular matrix, a diagonal matrix containing singular values, and a right singular matrix. fft represents the Fast Fourier Transform, used to compute a sequence's Discrete Fourier Transform (DFT). $ifft$ is the Inverse Fast Fourier Transform, which reverses the FFT process.

IV. Experiments and results

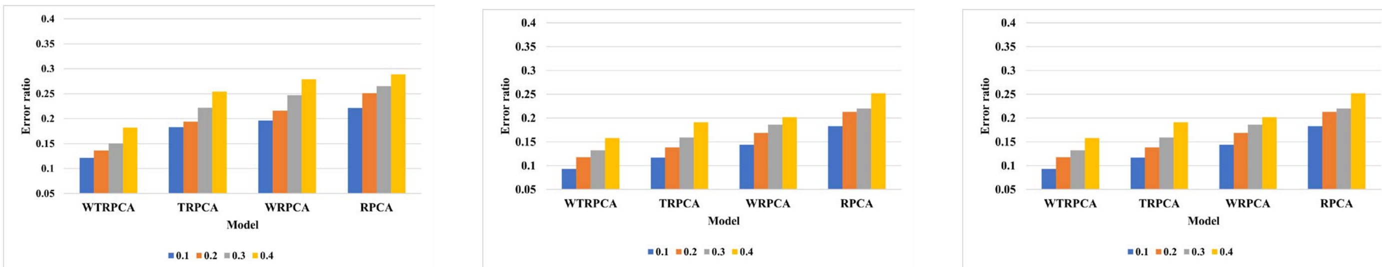
The NDBC-TAO dataset[1] provides the experiment's temperature, humidity, and pressure data, while the Normals Hourly of Climate dataset[2] supplies temperature, dew point, and wind speed measurements. The tensor data has a shape of $n_1 \times n_2 \times n_3$, where n_1 denotes the number of sensor nodes, n_2 represents the number of features, and n_3 describes the number of time slots. Random Gaussian noise with a mean of zero and a variance of $\sigma^2 = 20$ is introduced to both datasets. A fixed weight of $(1, 1.5, 2)$ is also assigned. Furthermore, the

[1]https://tao.ndbc.noaa.gov/tao/data_download
[2]https://www.ncdc.noaa.gov/cdo-web/datasets

noise ratios are 10%, 20%, 30%, and 40%. This study uses The error ratio as a metric to evaluate performance.

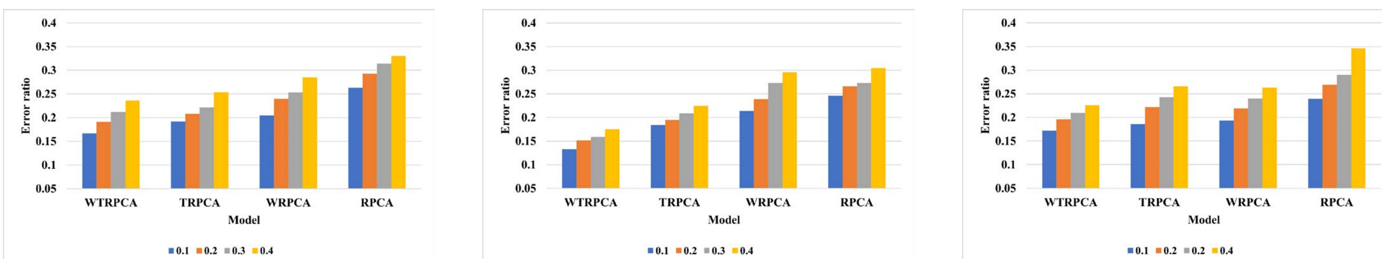
$$\text{Error ratio} = \frac{\sqrt{\sum_{j \in n_2} \left|\left|L_{:j:} - \hat{L}_{:j:}\right|\right|_F^2}}{\sqrt{\sum_{j \in n_2} \left|\left|L_{:j:}\right|\right|_F^2}} \quad (6)$$

The error ratio for each attribute in the dataset is determined by comparing the corresponding lateral slices of the original low-rank tensor $\mathcal{L}$ with the noise-reduced low-rank tensor $\hat{\mathcal{L}}$.



(a) Temp (b) Humidity (c) Pressure

Fig. 2: NDBC-TAO dataset



(a) Temp (b) Dew point (c) Wind speed

Fig. 3: Normals Hourly of Climate dataset

We compare our model WTRPCA with other competing models such as Robust Principal Component Analysis (RPCA) [1], Weighted Robust Principal Component Analysis (WRPCA) [3], Tensor Robust Principal Component Analysis (TRPCA) [5]. The performance of these models is evaluated, and it is observed that WTRPCA outperforms the other models. The illustrations referred to as Figure 2 and Figure 3 offer comparative evaluations of reducing noise performance on two separate datasets: NDBC-TAO and Climate. These comparisons show that the WTRPCA methodology consistently outperforms other techniques at various noise levels.

V. Conclusion

This study introduces a novel approach called WTRPCA to enhance the accuracy of reducing noise data in WSNs. Our proposed method addresses the challenge of corruption by combining a low-rank tensor to represent normal data and a sparse tensor to capture abnormal data patterns. This integration enables effective mitigation of corruption effects in WSNs. To further improve the reduce process, we incorporate the WTNN, which enhances the extraction of correlations within frontal slices while recovering low-rank tensors. WTNN incorporates a weighting mechanism that assigns different importance to singular values based on their significance. Through extensive experiments, we evaluate the performance of our approach using the error ratio metric. The results clearly demonstrate the superiority of our approach in handling corrupted data and achieving enhanced accuracy in recovering the original information within WSNs.

Acknowledgment

This work was supported by Institute of Information communications Technology Planning Evaluation(IIP) grant funded by the Korea government(MSIT)(No.2019-0-00135, Implementation of verification platform for 5G based smart sensor).

References

[1] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):1–37, 2011.

[2] Quanxue Gao, Pu Zhang, Wei Xia, Deyan Xie, Xinbo Gao, and Dacheng Tao. Enhanced tensor rpca and its application. *IEEE transactions on pattern analysis and machine intelligence*, 43(6):2133–2140, 2020.

[3] Jingfei He, Yunpei Li, Xiaoyue Zhang, and Jianwei Li. Missing and corrupted data recovery in wireless sensor networks based on weighted robust principal component analysis. *Sensors*, 22(5):1992, 2022.

[4] Yang Mu, Ping Wang, Liangfu Lu, Xuyun Zhang, and Lianyong Qi. Weighted tensor nuclear norm minimization for tensor completion using tensor-svd. *Pattern Recognition Letters*, 130:4–11, 2020.

[5] G Rajesh and Ashvini Chaturvedi. Data reconstruction in heterogeneous environmental wireless sensor networks using robust tensor principal component analysis. *IEEE Transactions on Signal and Information Processing over Networks*, 7:539–550, 2021.