

Spatial Low-Rank Tensor Factorization and Unmixing of Hyperspectral Images

William Navas-Auger, *Senior Member, IEEE*, and Vidya Manian, *Member, IEEE*

Abstract— This work presents a method for hyperspectral image unmixing based on non-negative tensor factorization. While traditional approaches may process spectral information without regard for spatial structures in the dataset, tensor factorization preserves the spectral-spatial relationship which we intend to exploit. We used a rank- $(L, L, 1)$ decomposition which approximates the original tensor as a sum of R components. Each component is a tensor resulting from the multiplication of a low-rank spatial representation and a spectral vector. Our approach uses the spatial factors, to identify high abundance areas where pure pixels (endmembers) may lie. Unmixing is done by applying Fully Constrained Least Squares such that abundance maps are produced for each inferred endmember. Results of this method are compared against other approaches based on non-negative matrix and tensor factorizations. We observed a significant reduction of spectral angle distance for extracted endmembers and equal or better RMSE for abundance maps as compared with existing benchmarks.

Index Terms— Spatial Low-rank tensor decomposition, remote sensing, hyperspectral image unmixing.

I. INTRODUCTION

ONE pervasive problem in remote sensing is the identification of materials based on their spectral signature [1]. When a pixel is recorded by the sensor, it can gather reflected radiation from more than one material or substance. This happens because there may be an insufficient spatial resolution for the sensor to capture individual materials or the substances in question are mixed uniformly. In either case, we can infer that mixed pixels have spectra that are some combination of the individual substances. The underlying assumption is that for a given scene with thousands of pixels there are a few material types such as water, vegetation, soil, concrete, different types of sediments, and minerals that have constant spectral properties.

II. BACKGROUND

Hyperspectral images are 3-dimensional data cubes with two spatial dimensions and one spectral dimension with hundreds of bands. To apply traditional signal processing algorithms, multidimensional arrays are unfolded; usually along the spectral dimension. In the resulting data set, every pixel is considered an independent sample of the material. This treatment ignores spatial relationships amongst neighboring pixels that could be exploited. Previous work using tensor factorizations for classification or feature extraction is

discussed in [2]. Tensor decompositions have also been used for blind signal unmixing as demonstrated in [3].

A. Hyperspectral Unmixing

Hyperspectral unmixing (HU) refers to the process of separating the spectral components of a hyperspectral image (HSI) as a matrix factorization where source signals, namely endmembers, are summed in proportion to their abundance plus a noise term to approximate the mixed pixels. This is referred to as the linear mixing model (LMM) expressed as [4]:

$$\mathbf{Y} = \mathbf{S}\mathbf{A} + \mathbf{W} \quad (1)$$

where $\mathbf{Y} \in \mathbb{R}^{M \times N}$, $\mathbf{S} \in \mathbb{R}^{M \times R}$, and $\mathbf{A} \in \mathbb{R}^{R \times N}$. \mathbf{Y} represents the HSI with M spectral bands and N pixels. The columns of \mathbf{S} represent the “pure” materials or endmembers, and \mathbf{A} is the fractional abundance matrix indicating the proportion in which endmembers contribute to every pixel. The term $\mathbf{W} \in \mathbb{R}^N$ accounts for noise. The LMM has two important constraints: all elements of \mathbf{Y} , \mathbf{S} and \mathbf{A} are non-negative (ANC), and abundances sum-to-one (ASC). Solving for \mathbf{S} and \mathbf{A} cannot be done analytically but can be approximated by numerical methods. Furthermore, identifying the number of endmembers, R , is a problem in itself. However, there are algorithms to estimate R as well such as HySime [5].

B. Non Negative Matrix Factorizations

Matrix factorizations such as Principal Component Analysis and Singular Value Decomposition produce orthogonal components helpful in dimensionality reduction but do not satisfy the ANC and ASC imposed by the LMM. Also, the orthogonal components they produce do not easily translate to physical phenomena. Non-negative matrix factorizations (NMF) can be used to separate signals into their constituent parts [6]. The non-negativity constraint also makes the decomposition easy to interpret as they relate to parts of the original data. Several algorithms exist to compute non-negative factors. Some of the most common ones are Alternating Least Squares (ALS/HALS) and Multiplicative Update (MU) [7]. These methods approximate a solution for \mathbf{S} and \mathbf{A} by minimizing the cost function:

$$C(\mathbf{S}, \mathbf{A}) = \|\mathbf{Y} - \mathbf{S}\mathbf{A}\|_F \quad (2)$$

subject to $\mathbf{S} \geq 0$, and $\mathbf{A} \geq 0$. Both algorithms are initialized with random inputs and iteratively update \mathbf{S} and \mathbf{A} until the solution converges. Improvements have been made by adding regularization terms to the cost function that promotes sparsity

The authors would like to thank the Laboratory for Applied Remote Sensing, Imaging, and Photonics (LARSIP), in the Department of Electrical and Computer Engineering, at the University of Puerto Rico, for providing the

resources for completion of this paper. Email: william.navas@upr.edu, vidya.manian@upr.edu

of the abundance matrix under the assumption that pixels are a mixture of few endmembers. The cost function with regularization takes the form:

$$C(\mathbf{S}, \mathbf{A}) = \|\mathbf{Y} - \mathbf{SA}\|_F + \lambda \|\mathbf{A}\|_p \quad (3)$$

Where $\|\cdot\|_p$ denotes the ℓ_p -norm of the abundance matrix as shown on eq. 4.

$$\|\mathbf{A}\|_p = (\sum_{r,n=1}^{R,N} |\mathbf{a}_r(n)|^p)^{1/p} \quad (4)$$

Commonly used norms are the ℓ_1 -norm, also known as the Manhattan Distance; and the ℓ_2 -norm which is the same as the Euclidean Distance. The use of other norms with p in the range (0,1) is discussed in [8], where the authors show that the $\ell_{1/2}$ -norm produces improved results as compared to other unmixing algorithms.

C. Tensor notation and definitions

Using notation from Kolda [9], we will use letters with bold script font to denote tensors, bold non-script font denote matrices, and bold lowercase indicate vectors.

Let $\mathbf{Y} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ be a tensor, where N is the number of dimensions (or modes) and I_n is the size on the n^{th} dimension. The number of dimensions is also referred to as the order of the tensor. An HSI cube is a third-order tensor with two spatial dimensions of size (I, J) and one spectral dimension of size K , such that $\mathbf{Y} \in \mathbb{R}^{I \times J \times K}$. An element of \mathbf{Y} is referenced as y_{ijk} .

Definition 1: A n -mode *fiber* is a column vector whose elements are obtained by fixing all tensor indices but the n^{th} one. For a third-order tensor, $\mathbf{Y} \in \mathbb{R}^{I \times J \times K}$, a fiber are is referenced as $\mathbf{y}_{:jk}$, $\mathbf{y}_{i:k}$, and $\mathbf{y}_{ij:}$, where the colon indicates all elements along that dimension.

Definition 2: A *slab* or *slice* is a matrix obtained by fixing all but two indices of a tensor. For a third-order tensor with indices (i, j, k) , slabs are denoted as matrices, $\mathbf{Y}_{i:}$, $\mathbf{Y}_{j:}$, and $\mathbf{Y}_{::k}$.

Definition 3: The n -mode *matrization* or unfolding of a tensor $\mathbf{Y} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, is the process of reordering the tensor elements into a matrix $\mathbf{Y}_{(n)}$, whose columns are the mode- n fibers of \mathbf{Y} ; such that $\mathbf{Y}_{(n)} \in \mathbb{R}^{I_n \times I_1 I_2 \dots I_{n-1} I_{n+1} \dots I_{N-1} I_N}$.

Definition 4: The n -mode *product* is the dot-product of the n -mode fibers of a tensor \mathbf{X} by the columns of a matrix \mathbf{A} where $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_n \times \dots \times I_{N-1} \times I_N}$ and $\mathbf{A} \in \mathbb{R}^{J \times I_n}$. The operation yields a new tensor $\mathbf{Y} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times J \times \dots \times I_{N-1} \times I_N}$. The n -mode product is expressed as $\mathbf{Y} = \mathbf{X} \times_n \mathbf{A}$ and it is equivalent to the matrix multiplication of \mathbf{A} by the n -mode matrization of \mathbf{X} , $\mathbf{X}_{(n)}$ as shown on eq. 6.

$$\mathbf{Y} = \mathbf{X} \times_n \mathbf{A} \Leftrightarrow \mathbf{Y}_{(n)} = \mathbf{A} \mathbf{X}_{(n)} \quad (6)$$

D. Tensor Factorizations

The Tucker decomposition [10], is an N -dimensional analog to singular value decomposition (SDV) where a tensor, $\mathbf{Y} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is decomposed as shown on eq. 7.

$$\mathbf{Y} = \mathbf{G} \times_1 \mathbf{F}^{(1)} \times_2 \mathbf{F}^{(2)} \dots \times_N \mathbf{F}^{(N)} \quad (7)$$

$\mathbf{G} \in \mathbb{R}^{R_1 \times R_2 \times \dots \times R_N}$ and factor matrices $\mathbf{F}^{(n)} \in \mathbb{R}^{I_n \times R_n}$. Low-rank representation is achieved by reducing the dimensions of the core tensor \mathbf{G} such that $R_n < I_n$.

Canonical Polyadic decomposition (CPD) is a special case of the Tucker decomposition where the core tensor \mathbf{G} of \mathbf{t} has all dimensions of the same size and it is also diagonal. CPD represents a tensor as a sum of rank-1 terms. As opposed to Tucker decomposition, CPD is free from rotational ambiguity under mild conditions [11]. The tensor rank is defined in terms of the CPD as the minimum number of rank-1 components needed to exactly reconstruct the original tensor. The CPD can be written as shown on eq (6).

$$\mathbf{Y} = \sum_{r=1}^R (f_r^{(1)} \circ f_r^{(2)} \dots f_r^{(N)}) \quad (6)$$

Where \circ indicates the outer product, and $f_r^{(n)} \in \mathbb{R}^{I_n}$ is a factor for each dimension I_n . We will call $f_r^{(1)}$ and $f_r^{(2)}$ the spatial factors while $f_r^{(3)}$ is the spectral factor. If we let

$$\mathbf{E}_r = f_r^{(1)} \circ f_r^{(2)} \quad (7)$$

and

$$\mathbf{Y}_r = \mathbf{E}_r \circ f_r^{(3)} \quad (8)$$

then, it is evident that spatial information represented by the rank-1 matrix \mathbf{E}_r encodes the magnitude of the r^{th} spectral factor $f_r^{(3)}$, at a given location.

This structure would be analogous to the abundance if ANC and ASC constraints are applied. However, while CPD provides an intuitive relationship between spatial and spectral content for each component, the limitation on \mathbf{E}_r being of rank-1; makes it insufficient to capture complex shapes under a single component. Many components with similar spectra would have to be clustered to produce shapes that capture the abundance of materials in the form of an abundance map.

Figure 1 shows a CPD with eight components ($R=8$). The original HSI has four endmembers. Components Tree-1 and Tree-2 have very similar spectral but the optimization generates 2 components. Road-2, Road-2, and Road-3 show a similar issue where the road is split into 3 components. Constraining the optimization to $R=4$ will produce the required number of components but there aren't enough degrees of freedom to represent spatial features accurately.

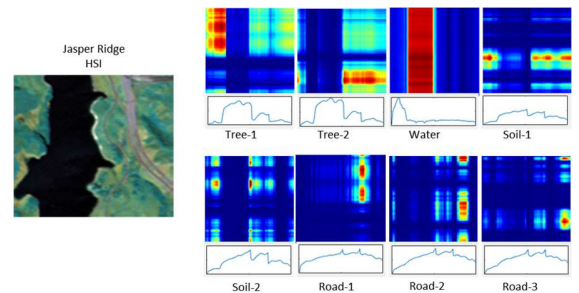


Figure 1. Components of a CPD computed for the Jasper Ridge HSI using CPD with $R=8$. The false color plots correspond to the magnitude of \mathbf{E}_r (red is larger), and the plot below each slab is the spectral factor, $f_r^{(3)}$.

Block Term Decomposition (BTD) is a generalization of CPD. It allows a tensor decomposition to be written as a sum of low-rank terms different than rank-1 [12]. Rank- $(L, L, 1)$ in particular, is a 3-way decomposition specifying rank- L , with L

> 1, for the first two dimensions and rank-1 for the third. The rank- $(L, L, 1)$ can be expressed as:

$$\mathbf{Y} = \sum_{r=1}^R \mathbf{F}_{r::}^{(1)} \mathbf{F}_{r::}^{(2)T} \circ f_{r::}^{(3)} \quad (7)$$

where $\mathbf{F}^{(1)} \in \mathbb{R}^{R \times I_1 \times L}$, $\mathbf{F}^{(2)} \in \mathbb{R}^{R \times I_2 \times L}$, and $f^{(3)} \in \mathbb{R}^{R \times I_3}$.

The motivation for using rank- $(L, L, 1)$ decomposition stems from the characteristics of an HSI. Endmembers on a scene are in the order of one to ten; while spatial dimension sizes may be in the hundreds. This disparity can be addressed with rank- $(L, L, 1)$ in such a way that fewer components are required to achieve a good spatial approximation.

III. METHODOLOGY

We propose using information from the spatial factors $\mathbf{F}^{(1)}$ and $\mathbf{F}^{(2)}$ as shown on eq. (8) to extract endmembers which will later be used to generate abundance maps.

$$\mathbf{E}_r = \mathbf{F}_r^{(1)} \mathbf{F}_r^{(2)T} \quad (8)$$

These factors multiplied result in a spatial-low rank representation of the abundance of one particular spectral component. Hence the name Spatial Low-Rank NTF Unmixing (SLR-NTF). Having estimated endmembers, abundance maps are computed using the fully constrained least-squares (FCLS) method introduced by [13].

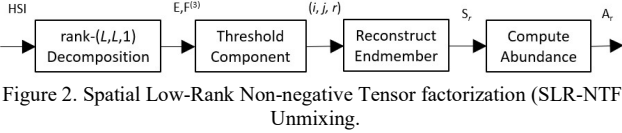


Figure 2. Spatial Low-Rank Non-negative Tensor factorization (SLR-NTF) Unmixing.

Parameter R is set as the number of expected endmembers. We can also write equation (7) in CPD form as a sum of LR components of rank-1 vectors. Hence, the resulting tensor is shown to be rank- LR . shown in eq. (9).

$$\mathbf{Y} = \sum_{r=1}^R \sum_{l=1}^L (\mathbf{F}_{r:l}^{(1)} \circ \mathbf{F}_{r:l}^{(2)} \circ f_{r:l}^{(3)}) \quad (9)$$

BTD is guaranteed to be essentially unique for a tensor with tensor rank $LR \leq \min(I_1, I_2)$. However, HSIs are not guaranteed to have a tensor rank less than the size of their spatial dimensions. However, since the rank is bounded by the total number of linearly independent components, we chose L proportional to the minimum size of the spatial dimensions and inversely proportional to R . Additionally, we weight L by the ratio of spatial to spectral size. This weight, $\min(I_1, I_2)/I_3$, adjusts L , increasing it when the spatial size is large relative to the spectral size and or reducing it when the opposite is true. An increase of spatial size relative to the spectral size would presumably increase the tensor rank assuming the spectral rank remains the same and is significantly smaller than the HIS dimension. The calculation of L is shown in equation (10).

$$L = \left\lceil \frac{\min(I_1, I_2)^2}{R I_3} \right\rceil \quad (10)$$

The selection of an optimal value of L based on signal unmixing continues to be a topic of research.

The rank- $(L, L, 1)$ decomposition was implemented as a non-linear least-squares optimization in Tensorflow 2.4 and the Adam [14] algorithm. Adam is a stochastic gradient descent (SDG) method proven to work well with neural nets and also found to be fast on this application. We used Glorot Gaussian

initialization [15] which resulted in faster convergence as compared to random uniformly distributed values in the range [0,1]. Glorot initialization scales values on a narrow range that is inversely proportional to the factor sizes.

High magnitude regions on E_r , indicate a strong abundance of endmember r . Reconstructed pixels are selected from the regions where $E_r/\max(E_r)$ exceeds a threshold $\gamma=0.95$. and the average of the selected pixels at those locations becomes the reconstructed endmember as shown on eq. (11).

$$s_r = \sum \frac{\mathbf{Y}(\arg_{i,j}(\frac{E_r}{\max(E_r)} > \gamma), r)}{N} \quad (11)$$

Having a set of endmembers, we compute the abundance of materials through FCLS [13]. This method solves the least-squares inverse problem while applying the non-negativity constraint and the sum-to-one constraint imposed on the abundance map.

Since SDG is sensitive to initial condition and HSIs do not necessarily meet the requirements for uniqueness, ten runs for each input HSI are executed. For each run, the root mean square error (RSME), and spectral angle difference (SAD) are computed, respectively.

$$RSME(a_r, a'_r) = \sqrt{\frac{1}{N} \sum_{n=1}^N (a_r(n) - a'_r(n))^2} \quad (12)$$

where a_r is the reference abundance and a'_r is the computed abundance for the r^{th} endmember.

$$SAD(s, s') = \frac{\langle s, s' \rangle}{\|s\| \|s'\|} \quad (13)$$

where, $\langle \cdot, \cdot \rangle$ denotes the inner product, s is the reference endmember, and s' the computed one.

IV. RESULTS

The HSI datasets including ground truth endmembers and abundances used were obtained from Y. Zhu [16]. We compared results against Vertex Component Analysis (VCA), NMF, NMF with l_1 -norm and $l_{1/2}$ -norm regularization, and matrix-vector NTF (MV-NTF) which is the only tensor-based approach.

For Samson, results are excellent showing a 64% and 62% reduction in SAD and RMSE. For Jasper Ridge, RMSE reduces to about 60% and SAD reduces by 29% compared to NMF-L1. Unmixing on the Urban data set shows a SAD reduction of 42% against MV-NTF which is second best. The Urban dataset lacked RSME measurements.

TABLE I

DIMENSIONS OF EACH HSI AND L PARAMETER SETTING AS PER EQ. (7).

| HSI set | I_1, I_2 | I_3 | R | L |
|-----------|------------|-------|-----|-----|
| Samson#1 | 95, 95 | 156 | 3 | 19 |
| J.Ridge#2 | 100, 100 | 198 | 4 | 12 |
| Urban | 307, 307 | 162 | 4 | 145 |

TABLE II

SPECTRAL ANGLE AND RMSE BENCHMARKS FOR SAMSON#1

| | Spectral Angle Distance (SAD) | | | | | Root Mean Squared Error (RMSE) | | | | |
|-------|-------------------------------|--------|--------|----------|---------|--------------------------------|--------|--------|----------|---------|
| | VCA | NMF | NMF-L1 | NMF-L1/2 | SLR-NTF | VCA | NMF | NMF-L1 | NMF-L1/2 | SLR-NTF |
| Soil | 0.4239 | 0.2793 | 0.178 | 0.2074 | 0.0331 | 0.1504 | 0.1633 | 0.1425 | 0.1719 | 0.0546 |
| Tree | 0.1118 | 0.115 | 0.0542 | 0.0559 | 0.0343 | 0.1483 | 0.171 | 0.1341 | 0.1683 | 0.0388 |
| Water | 0.0662 | 0.0804 | 0.0778 | 0.0731 | 0.0414 | 0.1055 | 0.061 | 0.036 | 0.0395 | 0.0244 |
| Avg. | 0.2006 | 0.1582 | 0.1033 | 0.1121 | 0.0363 | 0.1543 | 0.1318 | 0.1042 | 0.1266 | 0.0393 |

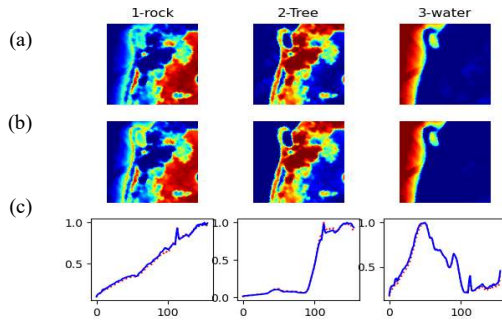


Figure 2. Samson#1 unmixing results. (a) Ground truth abundance. (b) computed abundance. (c) Reconstructed endmembers (solid blue) along with ground truth endmembers (dotted red).

TABLE III

SPECTRAL ANGLE AND RMSE BENCHMARKS FOR JASPER RIDGE#2

| | Spectral Angle Distance (SAD) | | | | | Root Mean Squared Error (RMSE) | | | | |
|-------|-------------------------------|--------|--------|----------|---------|--------------------------------|--------|--------|----------|---------|
| | VCA | NMF | NMF-L1 | NMF-L1/2 | SLR-NTF | VCA | NMF | NMF-L1 | NMF-L1/2 | SLR-NTF |
| Tree | 0.2565 | 0.2130 | 0.0680 | 0.0409 | 0.0714 | 0.3268 | 0.1402 | 0.0636 | 0.0707 | 0.0587 |
| Water | 0.2474 | 0.2001 | 0.3815 | 0.1682 | 0.2029 | 0.3151 | 0.1106 | 0.0660 | 0.1031 | 0.0387 |
| Soil | 0.3584 | 0.1569 | 0.0898 | 0.0506 | 0.1138 | 0.2936 | 0.2557 | 0.2463 | 0.2679 | 0.0798 |
| Road | 0.5489 | 0.3522 | 0.4118 | 0.3670 | 0.0581 | 0.2829 | 0.2450 | 0.2344 | 0.2737 | 0.0665 |
| Avg. | 0.3528 | 0.2305 | 0.2378 | 0.1567 | 0.1115 | 0.3046 | 0.1879 | 0.1526 | 0.1789 | 0.0609 |

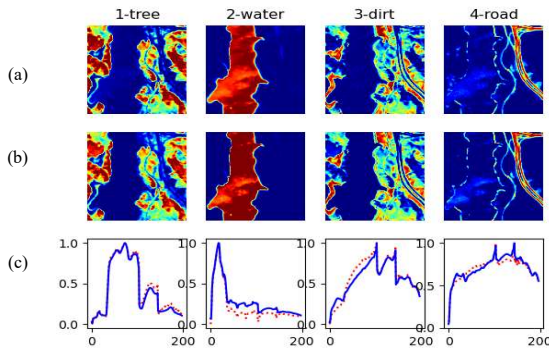


Figure 3. Jasper Ridge#2 unmixing results. (a) Ground truth abundance. (b) Computed abundance. (c) Reconstructed endmembers (solid blue) along with ground truth endmembers (dotted red).

TABLE IV

SPECTRAL ANGLE AND RMSE BENCHMARKS FOR URBAN

| | Spectral Angle Distance (SAD) | | | | | Root Mean Squared Error (RMSE) | | | | |
|---------|-------------------------------|--------|----------|--------|---------|--------------------------------|--------|----------|---------|--|
| | NMF | NMF-L1 | NMF-L1/2 | MV-NTF | SLR-NTF | NMF | NMF-L1 | NMF-L1/2 | SLR-NTF | |
| Asphalt | 0.2114 | 0.1548 | 0.1349 | 0.1638 | 0.0774 | 0.2041 | 0.2279 | 0.3225 | 0.1226 | |
| Grass | 0.3654 | 0.2876 | 0.099 | 0.2268 | 0.2176 | 0.2065 | 0.2248 | 0.3387 | 0.1320 | |
| Tree | 0.1928 | 0.0911 | 0.0969 | 0.1054 | 0.0626 | 0.187 | 0.1736 | 0.2588 | 0.1438 | |
| Roof | 0.737 | 0.7335 | 0.5768 | 0.3707 | 0.2507 | 0.1395 | 0.1861 | 0.1782 | 0.0772 | |
| Avg. | 0.3168 | 0.2269 | 0.3778 | 0.2167 | 0.1521 | 0.1843 | 0.2031 | 0.2746 | 0.1189 | |

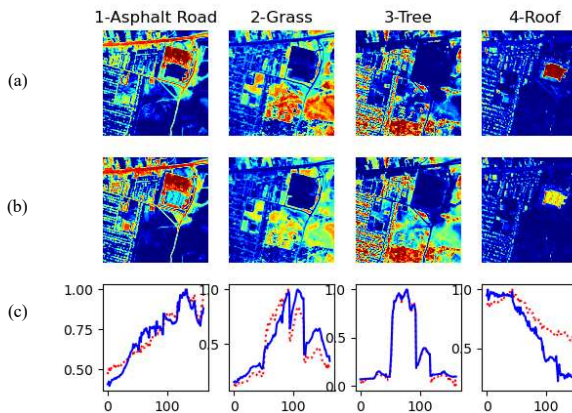


Figure 4. Urban unmixing results. (a) Ground truth abundance. (b) Computed abundance. (c) Reconstructed endmembers (solid blue) along with ground truth endmembers (dotted red).

V. CONCLUSIONS

SLR-NTF unmixing shows improved SAD and RMSE performance over VCA-FCLS, NMF, and MV-NTF and NTF on the benchmark HSI datasets. Careful selection of spatial low-rank L performs better than a fixed proportion of the spatial dimension. Additional improvements may be possible by characterizing its behavior with synthetic images; particularly, the estimation of tensor rank based on image content. The tensor factorization step is inherently more computationally complex than non-tensor approaches, but the advantages for the simultaneous analysis of spatial-spectral structure are evident. The advent of faster hardware and the ubiquity of GPUs are making these techniques practical.

REFERENCES

- [1] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. Wiley, 2003.
- [2] N. D. Sidiropoulos, L. D. Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor Decomposition for Signal Processing and Machine Learning," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551–3582, Jul. 2017.
- [3] Y. Qian, F. Xiong, S. Zeng, J. Zhou, and Y. Y. Tang, "Matrix-Vector Nonnegative Tensor Factorization for Blind Unmixing of Hyperspectral Imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1776–1792, Mar. 2017, doi: 10.1109/TGRS.2016.2633279.
- [4] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
- [5] J. M. P. Nascimento and J. M. Bioucas-Dias, "Hyperspectral signal subspace estimation," in *2007 IEEE International Geoscience and Remote Sensing Symposium*, Jul. 2007, pp. 3225–3228.
- [6] D. Donoho and V. Stodden, "When Does Non-Negative Matrix Factorization Give a Correct Decomposition into Parts?," in *Advances in Neural Information Processing Systems 16*, S. Thrun, L. K. Saul, and B. Schölkopf, Eds. MIT Press, 2004, pp. 1141–1148.
- [7] G. Zhou, A. Cichocki, Q. Zhao, and S. Xie, "Nonnegative Matrix and Tensor Factorizations: An algorithmic perspective," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 54–65, May 2014.
- [8] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral Unmixing via $L_{1/2}$ Sparsity-Constrained Nonnegative Matrix Factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4282–4297, Nov. 2011, doi: 10.1109/TGRS.2011.2144605.
- [9] T. G. Kolda and B. W. Bader, "Tensor Decompositions and Applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
- [10] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, pp. 279–311, 1966.
- [11] N. D. Sidiropoulos and R. Bro, *On the Uniqueness of Multilinear Decomposition of N-way arrays*. 2000.
- [12] L. Lathauwer, "Block Component Analysis, a New Concept for Blind Source Separation," 2012, vol. 7191, pp. 1–8.
- [13] D. C. Heinz and Chein-I-Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, Mar. 2001, doi: 10.1109/36.911111.
- [14] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*. 2017.
- [15] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *J. Mach. Learn. Res. - Proc. Track*, vol. 9, pp. 249–256, 2010.
- [16] F. Zhu, "Spectral Unmixing Datasets with Ground Truths," *CoRR*, vol. abs/1708.05125, 2017, [Online]. Available: <http://arxiv.org/abs/1708.05125>.