# Improving the Analysis of Hyperspectral Images Using Tensor Decomposition

Laura-Bianca BILIUS[1,2], Stefan Gheorghe PENTIUC[1,2]
[1]Faculty of Electrical Engineering and Computer Science, Stefan cel Mare University of Suceava
[2]MintViz Lab | MANSiD Research Center, Stefan cel Mare University of Suceava
laura.bilius@usm.ro, pentiuc@eed.usv.ro

***Abstract*—With the development of remote sensing hyperspectral image analysis techniques become more important. Known methods and algorithms for pattern recognition and clustering together with others built especially for this type of data are trying to be used for dimensional reduction, spectral unmixing, decomposition, segmentation. In this paper, we approached Parafac decomposition due to the lower computation time for obtaining a model that explains very well the real data. Image segmentation techniques are applied to the abundances map to distinguish materials in an area of interest. These techniques are used to obtain a representation that will facilitate the interpretation of the information contained in the hyperspectral image.**

***Keywords—Tensorial decomposition, Parafac, segmentation, hyperspectral images***

## I. INTRODUCTION

Hyperspectral images (HSI) are becoming a valuable tool, thanks to the possibility to use this information in mapping, environmental monitoring, detecting the size and shape of urbanized areas and, other fields of application. HSI consists of a large number of spectral bands that offer the possibility to analyze, identify, and measure different objects without having direct contact.

Due to the low resolution of hyperspectral images acquired from sensing systems, a pixel is a mixture of several materials, called mixed pixels. The spectral and spatial resolution is important for identifying the materials. Spectral unmixing is a process that decomposes a spectrum into a collection of endmembers and their abundances which extract the pure spectral components [1]. The detection of contributing materials in an area and estimating the corresponding proportion is possible by unmixing the HSI data [2].

Parafac (parallel factor analysis) Decomposition can estimate pure spectra even if in the hyperspectral image is mixed pixels [3]. The link between linear spectral models and tensor decompositions provides an easy interpretation of data, keeping the spectral and spatial structure without any information loss. This paper describes a method to analyze and discover materials in an area of interest using Parafac decomposition and segmentation of abundances map.

This work is supported by the project ANTREPRENORDOC, in the framework of Human Resources Development Operational Programme 2014-2020, financed from the European Social Fund under the contract number 36355/23.05.2019 HRD OP /380/6/13 – SMIS Code: 123847.

## II. RELATED WORK

Over time numerous methods and algorithms have been proposed to decompose hyperspectral images for clustering and analyze different areas. For example, the authors of [4] proposed the use of compression-based nonnegative CP decomposition to analyze HSI, which is an extension of linear spectral unmixing.

Another CANDECOMP/PARAFAC for HSI, which decomposes the tensor allowing the exploitation of spatial and, spectral information, was proposed in [5]. In paper [6], it was proposed a new model for spectral mixture called ALMM (augmented linear model mixture) which considers various spectral variabilities. In paper [7], the authors have proposed an algorithm for hyperspectral unmixing called PISINMF which improves spectral unmixing accuracy.

Also, some studies analyze different spectral models. For example, in the paper [8], the authors studied models such as LMM (linear mixing model), FCLSU (fully constrained least-squares unmixing), SCLSU (scaled constrained least-squares unmixing), ELMM (extended LMM) and RELMM (robust ELMM) in an ecosystem with high spectral variability.

A study that strengthens that linear models for mixed pixels are a good approach is [9], where the results obtained by the authors proved a high precision, reduced computation time, and a simple model. Regarding map segmentation, in the study [10] the authors provided a model based on quadtree and spectral clustering which showed correctness and rapidity. Considering that image segmentation can be used to find objects and boundaries in images [11], we approached this method on abundances map.

## III. CONCEPT AND THEORY

The Hyperspectral images are tensors $H \in \mathbb{R}\, I \times J \times K$ with three dimensions, where I and J are spatial dimensions and K is the spectral dimension. The author of [2] introduces the link between LSMM and tensor decompositions. According to [2], we approached the linear spectral mixture model (LSMM) which represents the spectrum of a pixel with K bands with R endmembers and their abundances.

978-1-7281-6870-8/20/$31.00 ©2020 IEEE

The LSMM for a pixel x is given by [2]:

$$x = Ms + \eta \quad (1)$$

Where x is a K × 1 vector of observed pixel, s is an R × 1 vector of abundances fractions for endmembers, η is the additive noise and M is the spectrum matrix. The mixing model written as a matrix for I × J pixels is described as [2], [12]:

$$X = MS + \Upsilon \quad (2)$$

Where X is the hyperspectral data, S is the matrix of abundances of all pixels on all endmembers and Υ is the additive noise.

Equation (2) "can be seen as a matrix factorization problem" according to [13]. In paper [14], the authors affirm that Parafac decomposition is a tensor factorization model. Parafac decomposition returns 3 factor matrices $A \in \mathbb{R}^{I \times N}$, $B \in \mathbb{R}^{J \times N}$ and $C \in \mathbb{R}^{K \times N}$ corresponding to the three dimensions of the hyperspectral image $H \in \mathbb{R}^{I \times J \times K}$ [15]. Mathematically, Parafac decomposes a tensor as a sum of rank-one tensor, where N is the number of components [16]. The Parafac decomposition for a tensor $H \in \mathbb{R}^{I \times J \times K}$ is defined as:

$$X \approx \sum_{n=1}^{N} a_n \circ b_n \circ c_n + E \approx [\![A, B, C]\!] \quad (3)$$

Equation (3) is shown graphically in Fig. 1 for N components, where E is the sum of squares of the residues.

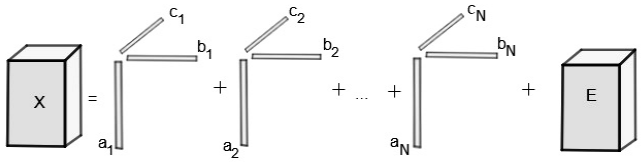


Fig. 1. A graphical representation of a N-component Parafac decomposition of the data X.

Therefore, (3) can be written as [2]:

$$X_{IJ \times K} = (A \odot B)\, C^T \quad (4)$$

Analyzing (2) and (4) we can see that these equations are identical. Based on Linear Spectral Mixture Model (LSMM) we say that a matrix which contains the abundances of all pixels on all endmembers and is defined as follows:

$$M = A * B,\ M \in \mathbb{R}^{I \times J} \quad (5)$$

Fig. 2 shows simulated hyperspectral data that describes better the construction of the abundances map resulted from (5). The matrices A and B are two simulated factor matrices that correspond to the spatial dimension and n is the number of Parafac components.

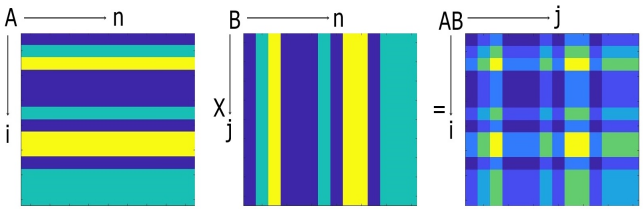


Fig. 2. The construction of abundances map for two simulated factor matrices A and B.

Taking into account that HSI can be approximated by a sum of tensor components, as we can see in (4), A⊙B represents the abundances for each pixel and his corresponding endmember. Such as the spatial information is kept in matrix M, we can call it abundance map [2].

The objective is to use the image segmentation for the abundances map, to obtain information about materials in the interest area. According to [11] and [17], image segmentation is a process of partitioning an image into multiple segments so, in this way, the surface and regions are isolated becoming easier to analyze.

## IV. Results

Starting from tensor decompositions, we approach the possibility to posterize the abundances maps of hyperspectral images to find the materials present in the areas of interest.

We use real hyperspectral data, and the corresponding dimensions are in Table I. We mention that the geometric (spatial) resolution for the Pavia City and University Pavia data sets are 1.3 meters/pixel, for the Salinas data set it is 3.7 meters/pixel and for the Danube Delta it is 30 meters/pixel. Parafac decomposition had a non-negativity constraint for all 3 factor matrices and the number of iterations was 50 for all data sets.

TABLE I. EXECUTION TIME IN SECONDS OF PARAFAC DECOMPOSITION FOR N COMPONENTS

| *Set* | *Dimension* | *50* | *100* | *150* | *200* |
|---|---|---|---|---|---|
| PaviaC | 1096×715×102 | 42.86 | 48.96 | 52.43 | 65.60 |
| PaviaU | 610×340×103 | 15.26 | 15.62 | 18.24 | 20.31 |
| Salinas | 512×217×204 | 13.15 | 17.30 | 19.23 | 20.12 |
| Danube | 181×151×53 | 2.26 | 2.62 | 2.99 | 3.47 |

As we can see in Fig. 3, we followed the algorithm and we obtain the factor matrices A and B after we applied the Matlab function for Parafac decomposition. We normalized the data for the abundances matrix M because of the extreme values, both positive and negative. Normalization is a process of transforming a relation into a simpler relation structure where no information is lost. In our case, we have obtained values that belong to [0, 1]. The segmentation of the abundances map M was achieved by reducing the number of colors. Specifically, each value in the matrix M is associated with a class that belongs to an interval of *nc* intervals, where *nc* the number of colors we chose.

```
Data: tensor; n – number of Parafac Decomposition
      components; nc – number of classes;
Result: A, B, C – factor matrices; M – abundances
        matrix; P – the segmentation of abundances map.

Function parafac(tensor,n)
    Return A,B,C
end
M=A*B;
N=normalize(M);
[d1,d2,d3]=size(t);
for i ← 1 to d1 do
  for j ← 1 to d2 do
     for k ← 1 to nc-1 do
        if (N(i,j)≥(k-1)/nc) && (N(i,j)<k/nc)then
          P(i,j)=k;
        end
     end
     if N(i,j)≥(nc-1)/nc then
        P(i,j)=nc;
     end
  end
end
```

Fig. 3. Construction of a segmented abundances map

As we can see in Fig. 4c), the graphical representation of the abundances matrix M, and in Fig. 4d), the segmented abundances map using 9 colors, we affirm that visually, the segmentation has provided a result, according to which we can clearly distinguish the areas where different materials are located. The number of classes was chosen after studying the Ground Truth data set (see Fig. 4b)).

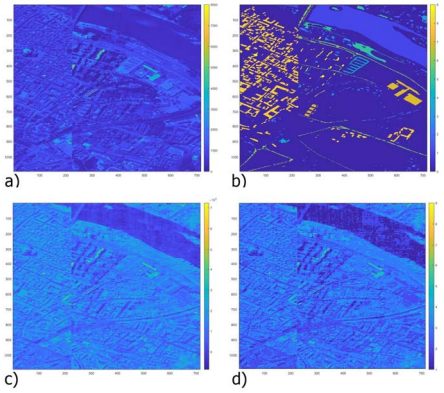


Fig. 4. Sample 25 of 102 for Pavia City dataset (a), Ground Truth (b), Abundances map (c), The segmented abundances map using 9 colors (d).

In Table II we can see the confusion matrix for the Pavia City dataset. We mention that the matrix was constructed using only the elements we find in Ground Truth. The accuracy, in this case, cannot be calculated because we do not have information about the classes obtained after the Parafac decomposition and their correspondence to the Ground-Truth classes. We can see, however, from Table II that there are classes obtained after Parafac decomposition that correspond to several classes from Ground Truth, and visually can be seen in Fig. 4b) and Fig. 4d).

TABLE II. CONFUSION MATRIX FOR PAVIA CITY DATASET

| *Class* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *Total* |
|---|---|---|---|---|---|---|---|---|---|---|
| *1* | 37.60 | 55.36 | 6.89 | 0.14 | 0 | 0 | 0 | 0 | 0 | 65971 |
| *2* | 0 | 0.19 | 4.67 | 16.24 | 29.21 | 23.79 | 17.78 | 6.90 | 1.18 | 7598 |
| *3* | 0 | 0 | 2.42 | 6.37 | 19.93 | 28.77 | 24.75 | 12.33 | 5.40 | 3090 |
| *4* | 0 | 0 | 0.33 | 4.80 | 25.84 | 32.81 | 22.42 | 8.11 | 5.66 | 2685 |
| *5* | 0 | 0 | 0.04 | 5.95 | 30.58 | 29.40 | 20.03 | 10.87 | 3.09 | 6584 |
| *6* | 0.22 | 9.72 | 40.07 | 37.05 | 11.75 | 1.10 | 0.05 | 0.01 | 0 | 9248 |
| *7* | 0 | 3.96 | 31.50 | 42.02 | 19.54 | 2.48 | 0.30 | 0.15 | 0.02 | 7287 |
| *8* | 0.01 | 0.45 | 5.38 | 17.09 | 29.70 | 27.70 | 14.21 | 4.37 | 1.03 | 42826 |
| *9* | 44.07 | 50.89 | 4.89 | 0.13 | 0 | 0 | 0 | 0 | 0 | 2863 |
| *Total* | 26098 | 39378 | 13437 | 15861 | 20777 | 17662 | 10153 | 3727 | 1059 | 148152 |

In Fig. 5 and Fig. 6 we can see the results after applying the same algorithm (see Fig. 1) for Pavia University and Salinas datasets, especially the results of segmented abundances maps for the two datasets.

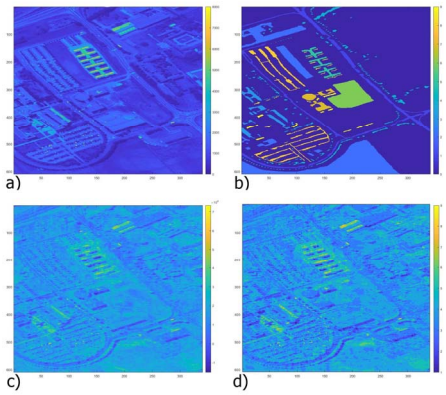


Fig. 5. Sample 25 of 103 for Pavia University dataset (a), Ground Truth (b), Abundances map (c), The segmented abundances map using 9 colors (d).

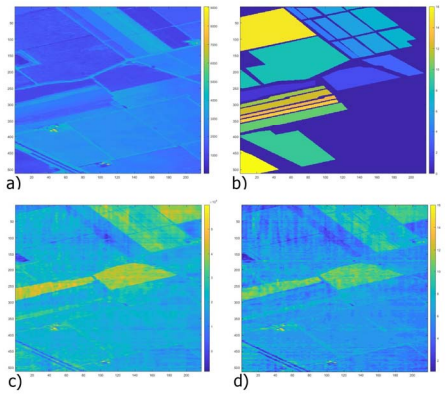


Fig. 6. Sample 25 of 204 for Salinas dataset (a), Ground Truth (b), Abundances map (c), The segmented abundances map using 16 colors (d).

For the Danube Delta dataset, we do not have Ground Truth and information about the materials present in the area of interest. As we can see in Fig. 6, the segmentation result delimits very well the materials. Since we do not have information about the number of materials present in the area of interest, we segmented the abundances map in 4 classes (see Fig. 6c)) and in 7 classes (see Fig. 6d)).

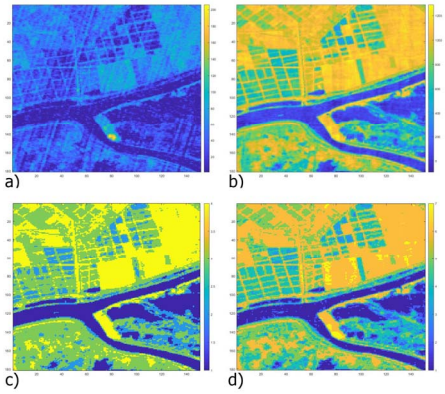


Fig. 7. Sample 25 of 53 for Danube Delta dataset (a), Abundances map (b), The segmented abundances map using 4 colors (c), The segmented abundances map using 7 colors (d).

TABLE III. THE EXPLAINED VARIATION OF PARAFAC DECOMPOSITION FOR N COMPONENTS

| ***Set*** | ***50*** | ***100*** | ***150*** | ***200*** |
|---|---|---|---|---|
| PaviaC | 86.86 | 90.66 | 94.23 | 95.82 |
| PaviaU | 92.57 | 95.47 | 97.68 | 98.50 |
| Salinas | 99.22 | 99.68 | 99.82 | 99.87 |
| Danube | 98.18 | 98.71 | 98.92 | 99.05 |

According to Table I, the execution time of Parafac decomposition for the studied datasets is quite small, given the large size of the data sets. We are also interested about the explained variation of the model. Therefore, Table III confirms that after decomposition we obtained that the discrepancy between the real data and the explained model is satisfactory, in most cases the explained variation is over 95%.

## V. CONLUSIONS

After the Parafac decomposition of tensor representing hyperspectral images, a segmentation algorithm was applied to the abundances map. According to the results for real hyperspectral data, we can affirm that Parafac decomposition offered models very close to real data. The execution time of further processing is much smaller after decomposition since the initial dimensions of the data has been reduced. The segmented images are well defined, the regions being more easily observable for data extraction. The resulting images facilitate the detection of various materials by visual analysis and comparison of the spectra associated with each distinct region. All these processing can be used in the design of a specialized tool for the analysis of hyperspectral images.

## ACKNOWLEDGMENTS

This work is supported by the project *ANTREPRENORDOC,* in the framework of Human Resources Development Operational Programme 2014-2020, financed from the European Social Fund under the contract number 36355/23.05.2019 HRD OP /380/6/13 – SMIS Code: 123847.

## REFERENCES

[1] Xavier Ceamanos, Silvia Valero, “4 - Processing Hyperspectral Images”, Editor(s): Nicolas Baghdadi, Mehrez Zribi, Optical Remote Sensing of Land Surface, Elsevier, 2016, Pages 163-200, ISBN 9781785481024, https://doi.org/10.1016/B978-1-78548-102-4.50004-1.

[2] Y. Qian, F. Xiong, S. Zeng, J. Zhou and Y. Y. Tang, "Matrix-Vector Nonnegative Tensor Factorization for Blind Unmixing of Hyperspectral Imagery," in IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 3, pp. 1776-1792, March 2017. doi: 10.1109/TGRS.2016.2633279

[3] Miguel de la Guardia Ana Gonzalvez Illueca, “Food Protected Designation of Origin, Volume 60 1st Edition”, Elsevier 2013, Published: 1st July 2013, eBook ISBN: 9780444595720

[4] M. A. Veganzones, J. E. Cohen, R. Cabral Farias, J. Chanussot and P. Comon, "Nonnegative Tensor CP Decomposition of Hyperspectral Data," in IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 5, pp. 2577-2588, May 2016. doi: 10.1109/TGRS.2015.2503737

[5] Fang, Leyuan & nan jun, he & Lin, Hui. (2017). CP tensor-based compression of hyperspectral images. Journal of the Optical Society of America A. 34. 252. 10.1364/JOSAA.34.000252.

[6] D. Hong, N. Yokoya, J. Chanussot and X. X. Zhu, "An Augmented Linear Mixing Model to Address Spectral Variability for Hyperspectral Unmixing," in IEEE Transactions on Image Processing, vol. 28, no. 4, pp. 1923-1938, April 2019.doi: 10.1109/TIP.2018.2878958

[7] Shao, Yang & Lan, Jinhui & Zhang, Yuzhen & Zou, Jinlin. (2018). Spectral Unmixing of Hyperspectral Remote Sensing Imagery via Preserving the Intrinsic Structure Invariant. Sensors. 18. 3528. 10.3390/s18103528.

[8] E. Ibarrola-Ulzurrun, L. Drumetz, J. Marcello, C. Gonzalo-Martín and J. Chanussot, "Hyperspectral Classification Through Unmixing Abundance Maps Addressing Spectral Variability," in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 7, pp. 4775-4788, July 2019.doi: 10.1109/TGRS.2019.2892903

[9] Liu, Jiaodi & Cao, Weibin. (2011). Based on linear spectral mixture model (LSMM) Unmixing remote sensing image. Proceedings of SPIE - The International Society for Optical Engineering. 8009. 10.1117/12.896397.

[10] Tian, Yafu & Wang, ke & Li, Ruifeng & Zhao, Lijun. (2018). A fast incremental map segmentation algorithm based on spectral clustering and quadtree. Advances in Mechanical Engineering. 10. 168781401876129. 10.1177/1687814018761296.

[11] Ying Tan, Chapter 11 - Applications, Editor(s): Ying Tan, Gpu-Based Parallel Implementation of Swarm Intelligence Algorithms, Morgan Kaufmann, 2016, Pages 167-177, ISBN 9780128093627, https://doi.org/10.1016/B978-0-12-809362-7.50011-X.

[12] C. Shi and L. Wang, "Linear Spatial Spectral Mixture Model," in IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 6, pp. 3599-3611, June 2016.doi: 10.1109/TGRS.2016.2520399

[13] Qian, Yuntao & Xiong, Fengchao & Zeng, Shan & Zhou, Jun & Tang, Yuan. (2016). Matrix-Vector Nonnegative Tensor Factorization for Blind Unmixing of Hyperspectral Imagery. IEEE Transactions on Geoscience and Remote Sensing. PP. 1-17. 10.1109/TGRS.2016.2633279.

[14] Yang, Bo & Zamzam, Ahmed & Sidiropoulos, N.D.. (2018). ParaSketch: Parallel Tensor Factorization via Sketching. 10.1137/1.9781611975321.45.

[15] Kolda, Tamara G. and Brett W. Bader. “Tensor Decompositions and Applications.” SIAM Review 51 (2009): 455-500.

[16] Hansen, Samantha & Plantenga, Todd & Kolda, Tamara. (2013). Newton-Based Optimization for Kullback-Leibler Nonnegative Tensor Factorizations. Optimization Methods and Software. 30. 10.1080/10556788.2015.1009977.

[17] Gurusamy, Vairaprakash & Kannan, Subbu & G.Nalini,. (2014). REVIEW ON IMAGE SEGMENTATION TECHNIQUES.