



Multifeature Hyperspectral Unmixing Based on Tensor Decomposition

Mohamad Jouni, Mauro Dalla Mura, Lucas Drumetz, Pierre Comon

► To cite this version:

Mohamad Jouni, Mauro Dalla Mura, Lucas Drumetz, Pierre Comon. Multifeature Hyperspectral Unmixing Based on Tensor Decomposition. 2021. hal-03480890v2

HAL Id: hal-03480890

<https://hal.science/hal-03480890v2>

Preprint submitted on 30 Dec 2021 (v2), last revised 18 Sep 2023 (v7)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multifeature Hyperspectral Unmixing Based on Tensor Decomposition

Mohamad Jouni, *Member, IEEE*, Mauro Dalla Mura, *Member, IEEE*, Lucas Drumetz, *Member, IEEE*,
and Pierre Comon, *Fellow, IEEE*

Abstract—Hyperspectral unmixing is an active area of blind source separation. It refers to the representation of mixed pixels (samples) as a set of pure materials (sources), weighted by their abundances. Since spectral features alone are often insufficient, it is common to rely on other features of the scene as additional knowledge. In this paper, the hyperspectral image (HSI) is represented as a high-order tensor with additional features in a multimodal, multifeature framework, spanning modes of pixels, spectral features, and additional features, where matrix models become insufficient. This requires the use of tensor models, and particularly the Canonical Polyadic Decomposition, which is blind and straightforward for unmixing, and maintains the physical properties of the data. So far, this model has been applied in preliminary and specific applications, and still lacks a general framework for unmixing including the interpretation of the results. In this paper, we propose a methodological framework for multifeature unmixing based on the Alternating Optimization Alternating Direction Method of Multipliers algorithm and incorporating Abundance Sum-to-one Constraint (AO-ADMM-ASC), with in-depth mathematical, physical and graphical interpretations and links to the Extended Linear Mixing Model. Moreover, we propose to incorporate Mathematical Morphology as spatial features in multifeature unmixing and revise the work of patch features in order to demonstrate the interest of the proposed framework. Experiments on real HSI data sets show the efficiency of AO-ADMM-ASC and allows an in-depth interpretation of the model based on the quality of the features and the variation of the imposed rank.

Index Terms—Blind Source Separation, Hyperspectral Unmixing, Tensor Decomposition, Extended Linear Mixing Model.

I. INTRODUCTION

HYPERSPECTRAL IMAGERY refers to the acquisition of images of a scene over a wide and almost continuous spectrum. A hyperspectral image (HSI) contains pixels that can cover areas of pure or mixed materials and amounts to a high spectral feature diversity [1], [2], thus allowing to perform a blind source separation (BSS) [3]–[5] on the observed spectral signatures to blindly extract those of pure materials (*sources*), also called *endmembers* (EM), and their per-pixel (*per-sample*) abundances. This case of BSS is known as hyperspectral unmixing (HU), which is an active research topic with several applications like remote sensing, chemometrics, biomedical

imagery, etc [6]–[15], which allows to understand and quantify the physical components of a scene.

A significant part of BSS and HU research relies on matrix factorization that is constrained to properly model the context of the problem. Consequently, an observed data matrix $M \in \mathbb{R}^{I \times J}$ (with I samples and J features) is decomposed by minimizing the generic cost function [16]:

$$\underset{A, B}{\operatorname{argmin}} \|M - AB^T\|_F^2 + r(A) + r(B) \quad (1)$$

where the columns of $B \in \mathbb{R}^{J \times R}$ represent the R estimated source signals, the rows of $A \in \mathbb{R}^{I \times R}$ represent their per-pixel abundances, and $r(\cdot)$ encodes the imposed constraints and/or regularizations to enforce desirable properties on the solutions. In the case of HU, a classical approach is Nonnegative Matrix Factorization (NMF) which relies on a linear mixing model (usually referred to as LMM) of the observed HSI matrix (see Fig. 1), such that A and B are element-wise *nonnegative*, which applies in most domains of BSS other than HU where the factor matrices are associated with some physical variables, and the rows of A are subjected to the *Abundance Sum-to-one Constraint* (ASC, i.e. $\sum_{r=1}^R a_{ir} = 1 \forall i \in \{1, \dots, I\}$), which applies to domains where the coefficients of the decomposition are proportions.

When only few materials concur in the mixture for each pixel, sparsity is imposed on the abundances [17]. Finally, in order to account to the spectral variability (SV) of the sources (e.g., variations in the EMs due to illumination changes or topographic effects) in real data, the Extended Linear Mixing Model (ELMM) was proposed to extend the LMM to account to said SVs. ELMM is an active topic that has seen a lot of progress recently [11]–[14].

In some scenarios, the HSI is treated as a data cube [18]–[23], and in several others, the HSI does not come alone as we might have additional modalities such as when we have a time series [24], when combined with other modalities (e.g., multi-sensor analysis, LiDAR fusion) [25]–[28], or when we extract some spatial features from the image (such as in classification problems [29]–[32]). Such scenarios have recently also concerned other areas of BSS such as multi-channel signal processing [33]–[35] and multidimensional biomedical signal and image processing [36]–[40]. In these scenarios the data are represented natively as *tensors*¹ [41], where usually

M. Jouni, M. Dalla Mura, and P. Comon are with Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, GIPSA-lab, 38000 Grenoble, France (e-mail: mohamad.jouni@gipsa-lab.fr; mauro.dalla-mura@gipsa-lab.fr; pierre.comon@gipsa-lab.fr). M. Dalla Mura is also with the Institut Universitaire de France (IUF) and with Tokyo Tech World Research Hub Initiative (WRHI), School of Computing, Tokyo Institute of Technology, Tokyo, Japan.

L. Drumetz is with IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238, Brest, France (e-mail: lucas.drumetz@imt-atlantique.fr).

Manuscript received ...; revised ...

¹A tensor can be represented as a multidimensional array. The order of a tensor refers to the number of its array's indices, which is also the number of its *modes*. For example, a tensor of dimensions $I \times J \times K$ is said to have three *modes*, and is called a *third-order tensor*. Data sets with order 3 or above are described as *high-order tensors*.

the challenge is the proper modeling of a joint factorization of multivariate representations without losing the multimodal structure, and hence its interpretation in terms of BSS.

Among these scenarios, we focus on the case of HU where the HSI is augmented into a tensor with different feature diversities represented as new modes, which we refer to as *multifeature HU*, answering questions such as: How can we jointly perform a constrained factorization in such settings? and how can we interpret the extracted factors? An analysis of the literature shows that there are works that perform NMF with additional constraints [42]–[44], and others that consider the case of multimodal inputs with coupled NMF [45], [46], but this is different from considering data as tensors in our case.

As the native structure of our data is a tensor, we consider the problem in terms of tensor decomposition (TD) [41], which is the natural framework for processing multimodal data in the signal and image processing community [47]–[49]. There are many types of TD approaches such as Tucker Decomposition, Block Term Decomposition, Canonical Polyadic Decomposition (CPD), etc [50]. However, we choose to use CPD since it is a natural model for source separation and allows for interpretability of the extracted components, which are core aspects of this work. In the third-order case, as illustrated in Fig. 2, CPD decomposes a data tensor $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$ into a diagonal core tensor $\mathbf{\Lambda} \in \mathbb{R}^{R \times R \times R}$ and 3 factor matrices $\{\mathbf{A} \in \mathbb{R}^{I \times R}, \mathbf{B} \in \mathbb{R}^{J \times R}, \mathbf{C} \in \mathbb{R}^{K \times R}\}$, each representing one of the 3 modes² of \mathcal{T} . It also extends NMF to high-order structures and can adopt all of its features, especially that of imposing constraints. Consequently, CPD is computed by solving the cost function minimization [41]:

$$\underset{\mathbf{A}, \mathbf{B}, \mathbf{C}}{\operatorname{argmin}} \|\mathcal{T} - \mathbf{\Lambda} \bullet_1 \mathbf{A} \bullet_2 \mathbf{B} \bullet_3 \mathbf{C}\|_F^2 + r(\mathbf{A}, \mathbf{B}, \mathbf{C}) \quad (2)$$

where \bullet_d denotes the tensor contraction operator along the d -th mode, and $r(\cdot)$ represents the imposed constraints and regularizations.

A. Related Works and Contributions

In the context of HU, CPD has been used with multitemporal/angular HSIs [24] as well as with HSIs having an additional diversity of extracted neighborhood patches [14] (see Fig. 7 for a 5×5 patch-HSI tensor). Moreover, some works [29]–[31] jointly considered HSIs with spatial features extracted by Mathematical Morphology (MM) filters [51]–[54] in the framework of multifeature scene classification (see Fig. 8). These works show that CPD is a suitable approach for joint decomposition. However, they present some limitations.

From an algorithmic perspective, the nonnegative constraint is merely implemented in [55] by projection onto the nonnegative orthant, which exhibits some computational issues [48]. In [14], nonnegative ALS is used where ASC is also naively implemented by projecting the abundances on the unit simplex, contrary to the common practice in the matrix case [17] where ASC is embedded in the updates. In [24], the nonnegative CPD is computed using the ProCo-ALS algorithm, which

is considerably fast [47] but not so flexible with additional constraints. Finally, in [29]–[31], an alternative algorithm is proposed based on AO-ADMM [49] with compression and nonnegative constraints, which is flexible and stable with large datasets, but has not yet addressed multifeature HU which requires further modeling (i.e., sparsity, ASC).

In terms of interpretability, [24] faced a certain challenge in interpreting the third-mode factors, perhaps due to the naive employment of CPD. In [14], a link was established between CPD and ELMM, but was not deeply investigated in that it was restricted to the case of patches and tested only with synthetic data, with yet another important challenge in interpreting the factors, which poses an ambiguity on its performance and the meaning of the extracted features. Finally, [14] and [29]–[31] perform TD with spatial features, however, the spatial features used can allow limited flexibility (e.g., patches), or the interpretation of the factor matrices was not addressed knowing that incorporating them with CPD showed improvement in supervised classification (e.g., MM). Here, we wish to consider such operations in a BSS framework from the lens of multifeature HU with in-depth interpretability.

This presents us with two main challenges such as (1) specifically tuning AO-ADMM to incorporate ASC, which is challenging due to the multilinear structure of CPD, particularly in modeling the samples as a convex combination of the spectral sources in a multimodal setting, and (2) exploring the meaning of the extracted features in these conditions. To our knowledge, imposing ASC in CPD or AO-ADMM as a natural extension of NMF in [17] has not been done. Furthermore, we are interested in finding a general framework for multimodal HU favouring the interpretation of its results under any third-mode diversity. That said, we mention our contributions to jointly deal with these limitations in the following.

- We propose a methodological framework for dealing with multifeature HU based on AO-ADMM by Huang [49], and expand it to incorporate ASC with joint nonnegativity and sparsity, such that the case of NMF [17] becomes a special case. The proposed multifeature HU is a general algorithm that can be applied in other domains of BSS where convex combinations of sources apply.
- We establish a unified framework for the interpretability of multifeature HU. In particular, the link between ELMM and CPD [14] is expanded by providing in-depth physical and graphical insights for better interpretability of the CPD model and its factors.
- We propose to include MM as spatial features to perform a spectral-spatial HU and demonstrate the aforementioned points. We also revise [14] and provide detailed interpretations on the cases of patches and MM, which has not been addressed in any of the previous works [14], [29]–[31]. This analysis also shows that MM is better suited since it embeds physically meaningful features (scale and brightness of objects) into HU unlike patches.

The remainder is organized as follows. In section II, we introduce some background. In section III, we detail the proposed framework. In section IV, we present our experiments and results. Finally, we draw out some conclusions in V.

²Pixels form only one mode, even if images are often seen as 2D objects.

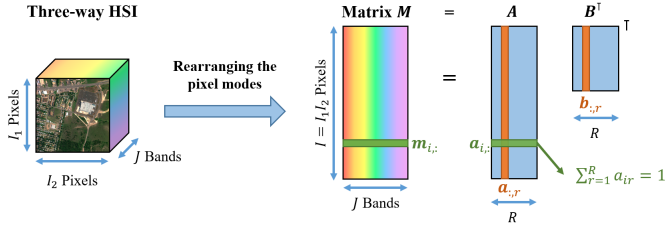
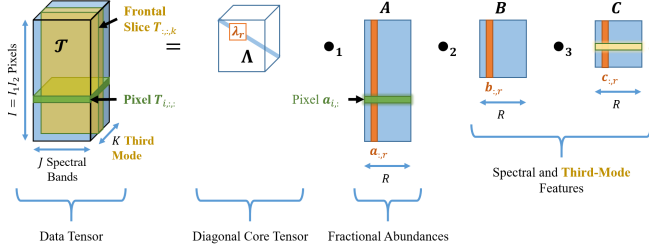


Fig. 1: Classical matrix-based HU using NMF (i.e., LMM)

Fig. 2: Third-order CPD. One pixel $T_{i, :, :}$ (horizontal green slab) represents a matrix of features and is associated with one row of A . Similarly, one slice $T_{:, :, k}$ (frontal yellow slab) represents a matricized HSI and is associated with one row of C .

II. BACKGROUND

In this section, we briefly review the existing notions in the literature upon which we base our algorithm and general interpretation of the multimodal HU framework. First, we explain how ASC is applied in NMF [17] as the proposed framework extends this for TD. Then, we give a brief account on ELMM, including graphical and visual interpretations, which will be the basis for the proposed interpretation. Finally, we discuss the link between CPD and ELMM preliminarily presented in [14].

A. NMF with ASC, Nonnegativity, and Sparsity

In the NMF case [17], when sparsity (ℓ_1 norm) and ASC are imposed on the abundances, (1) becomes:

$$\begin{aligned} \underset{A, B}{\operatorname{argmin}} \quad & \frac{1}{2} \|M - AB^T\|_F^2 + \alpha \|A\|_1 \\ \text{s.t. } \quad & A \succeq 0, B \succeq 0, \sum_{r=1}^R a_{i,r} = 1 \quad |\forall i \in \{1, \dots, I\} \end{aligned} \quad (3)$$

where $\alpha > 0$, and \succeq denotes element-wise nonnegativity. A simple strategy to embed ASC goes by stacking a row vector in B and a column vector in M such that [17]:

$$\tilde{M} = [M \mid \delta \mathbf{1}_{I \times 1}], \quad \tilde{B} = \begin{bmatrix} B \\ \delta \mathbf{1}_{1 \times R} \end{bmatrix},$$

where δ is a constant that is usually set as the mean of M , and the last row of \tilde{B} is reset to δ after each iteration. This operation ensures that ASC is embedded in NMF since $\forall i \in \{1, \dots, I\}$ we have $m_{i, J+1} = \sum_{r=1}^R a_{i,r} b_{J+1, r} = \sum_{r=1}^R a_{i,r} \delta = \delta$, corresponding to $\sum_{r=1}^R a_{i,r} = 1$. Then, (3) becomes:

$$\underset{A, \tilde{B}}{\operatorname{argmin}} \quad \frac{1}{2} \|\tilde{M} - A\tilde{B}^T\|_F^2 + \alpha \|A\|_1 \quad \text{s.t. } A \succeq 0, \tilde{B} \succeq 0 \quad (4)$$

There are many algorithms proposed in the literature that deal with sparse NMF and ASC, which are out of the scope of this work [42], [43]. In our case, we extend it within the AO-ADMM framework for CPD, in which NMF becomes a special case for order-2 tensors, which we propose in Section III-A by extending this simple strategy of embedding ASC to tensors, which we refer to as AO-ADMM-ASC.

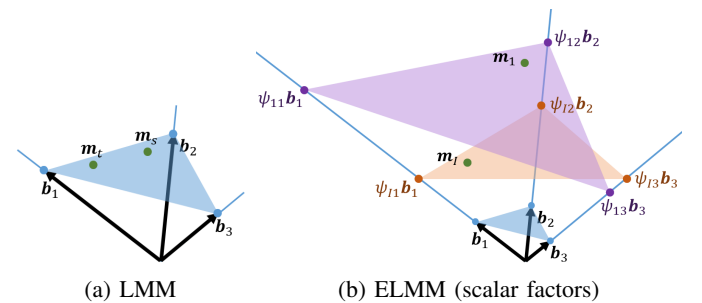
B. ELMM

While LMM is seen as a direct approach for HU, it cannot model SVs represented by nonlinear effects or illumination conditions. One way to account to said effects is through ELMM [12], which in general assumes additional degrees of freedom that account to said SVs *at the pixel level* by introducing a *pixel-dependent* SV function $f_i : \mathbb{R}^J \rightarrow \mathbb{R}^J$, which maps each EM b_r to a new spectral signature $b_r^{(i)}$ that best reflects the targeted SVs:

$$m_i = \sum_{r=1}^R a_{ir} f_i(b_r) = \sum_{r=1}^R a_{ir} b_r^{(i)}. \quad (5)$$

For example, in the case of different illumination conditions, this can be represented as a scaling factor for each pixel on the EMs. In the following, we present the parts that are at the basis of the interpretability of our proposed framework.

When ASC is imposed in LMM, all the pixels will lie on the convex hull of the set of estimated EMs (the columns of B , where in this case, we note that unless the EMs are not affinely independent, which is unlikely, and $R \leq J + 1$, the convex hull is a simplex), and the fractional abundances in each row of A define the coordinates of each pixel in the convex hull, which is illustrated in Fig. 3a. With the introduction of ELMM, the pixels will not lie on the same simplex anymore as each pixel is mapped to a new set of EMs, which is illustrated in Fig. 3b in the case where the SV is modeled by a scaling factor ψ_{ir} that is pixel- and EM-dependent such that $b_r^{(i)} = \psi_{ir} b_r$ [12]. Accordingly, in the case of scaling factors, the new spectral signatures $b_r^{(i)}$ are allowed to move only along the directions

Fig. 3: Graphical comparison between LMM and ELMM (scalar factors) in the case of three spectral signatures $\{b_1, b_2, b_3\}$ and two example pixels $\{m_s, m_t\}$. The relative coordinates in the simplices are conserved.

of \mathbf{b}_r , then the following are equivalent:

$$\mathbf{M} = (\mathbf{A} \boxtimes \boldsymbol{\Psi}) \mathbf{B}^\top \quad (6)$$

$$\mathbf{m}_i = \sum_{r=1}^R a_{ir} \psi_{ir} \mathbf{b}_r = \mathbf{a}_{i,:} \text{diag}\{\boldsymbol{\Psi}_{i,:}\} \mathbf{B}^\top = \mathbf{a}_{i,:} \boldsymbol{\Psi}_{(i)} \mathbf{B}^\top \quad (7)$$

where $\boldsymbol{\Psi} \in \mathbb{R}^{I \times R}$ is the matrix of scaling factors, and \boxtimes is the Hadamard product. These mathematical, graphical, and visual relationships are the key to elaborate and interpret the case of CPD in Section III-B.

C. CPD and ELMM

In the work of [14], a representation of CPD in terms of ELMM has been presented, which will be reported here. CPD decomposes a third-order tensor \mathcal{T} such that:

$$\mathcal{T} = \mathbf{\Lambda} \bullet_1 \mathbf{A} \bullet_2 \mathbf{B} \bullet_3 \mathbf{C} \iff \mathcal{T} = \sum_{r=1}^R \mathbf{a}_{:,r} \otimes \mathbf{b}_{:,r} \otimes \mathbf{c}_{:,r} \quad (8)$$

$$\iff t_{ijk} = \sum_{r=1}^R a_{ir} b_{jr} c_{kr} \quad (9)$$

where \otimes designates the outer product, t_{ijk} is a (scalar) entry of \mathcal{T} , R is the number of estimated sources, and $\mathbf{\Lambda} \in \mathbb{R}^{R \times R \times R}$ is a diagonal tensor containing scaling factors (by absorbing the ℓ_2 -norms of the columns of the factor matrices) in order to fix the scaling indeterminacy [41], [48]. In (8) and (9), and for the convenience of the following expressions, we suppress the expression of $\mathbf{\Lambda}$ by absorbing its scaling factors in the columns of \mathbf{C} . As shown in [14], assuming that $\mathbf{T}_{:,:,k}$ designates the k -th frontal slice of \mathcal{T} , one can write:

$$\mathbf{T}_{:,:,k} = \mathbf{A} \text{diag}\{\mathbf{c}_{k,:}\} \mathbf{B}^\top = \mathbf{A} \boldsymbol{\Psi}_{(k)} \mathbf{B}^\top = \mathbf{A} \tilde{\mathbf{f}}_k(\mathbf{B})^\top \quad (10)$$

$$\iff \mathbf{t}_{i,:,k} = \sum_{r=1}^R a_{ir} (\mathbf{b}_r c_{kr}) = \sum_{r=1}^R a_{ir} \mathbf{f}_k(\mathbf{b}_r) \quad (11)$$

where $\mathbf{t}_{i,:,k}$ designates one pixel of $\mathbf{T}_{:,:,k}$. From the perspective of each frontal slice, CPD resembles a regularized coupled NMF of the frontal slices where \mathbf{A} is a common factor:

$$\argmin_{\mathbf{A}, \mathbf{B}^{(k)}} \|\mathbf{T}_{:,:,k} - \mathbf{A} \mathbf{B}^{(k)\top}\|_F^2 \text{ s.t. } \mathbf{A} \succeq 0, \mathbf{B}^{(k)} \succeq 0 \quad (12)$$

where $\mathbf{B}^{(k)} = \tilde{\mathbf{f}}_k(\mathbf{B})|_{\forall k \in \{1, \dots, K\}}$, which is similar to the case of ELMM. The latter was used in [14] as a way to impose a spatial smoothing (12) on the abundances within a patch of pixels. This link between CPD and ELMM was preliminary presented and restricted to an application of patches with only simulated data, and the meaning of the SVs was not explored. In Section III-B, we propose a generalized in-depth interpretation of multi-feature HU, and in Section III-C we present how to incorporate spatial features (e.g., patches and MM) in such a setting.

III. PROPOSED MULTI-FEATURE HU

In this section, we present the proposed multi-feature HU framework based on CPD. First, we talk about the implementation of AO-ADMM-ASC. Then, we interpret the ELMM model based on TD and propose to include spatial features as third-mode.

A. AO-ADMM-ASC with Nonnegativity and Sparsity

In CPD, after imposing nonnegativity on the factor matrices, and sparsity and ASC on the abundances, (2) becomes:

$$\begin{aligned} \argmin_{\mathbf{A}, \mathbf{B}, \mathbf{C}} \|\mathcal{T} - \mathbf{\Lambda} \bullet_1 \mathbf{A} \bullet_2 \mathbf{B} \bullet_3 \mathbf{C}\|_F^2 + \alpha \|\mathbf{A}\|_1 \\ \text{s.t. } \mathbf{A} \succeq 0, \mathbf{B} \succeq 0, \mathbf{C} \succeq 0, \sum_{r=1}^R a_{ir} = 1 \quad |\forall i \in \{1, \dots, I\}| \end{aligned} \quad (13)$$

In (13), since \mathbf{A} adopts the sum-to-one constraint on its rows, it is hence enough to absorb only the norms of the columns of \mathbf{B} and \mathbf{C} into $\mathbf{\Lambda}$, which we suppress again in the following expressions for the sake of convenience. In order to solve (13), we adopt AO-ADMM [49], where the factor matrices are updated in an alternating way and where each update of a factor matrix is optimized as an ADMM subproblem. In order to model the problem as close as possible to LMM, we propose to embed ASC in the model, where the key point becomes to extend the concept from NMF to CPD by stacking a row vector in \mathbf{B} and a lateral slice in \mathcal{T} (i.e., $\mathbf{T}_{:,J+1,:} \in \mathbb{R}^{I \times 1 \times K}$) to ensure that $\sum_{r=1}^R a_{ir} = 1 \quad |\forall i \in \{1, \dots, I\}|$.

First, we introduce a row vector in \mathbf{B} such that $b_{J+1,r} = \delta c_{K,r}^{-1} \quad |\forall r \in \{1, \dots, R\}|$, which corresponds to constructing the slice $\mathbf{T}_{:,J+1,:}$ where $t_{i,J+1,k} = \delta \sum_{r=1}^R a_{ir} c_{K,r}^{-1} c_{k,r} \quad |\forall k \in \{1, \dots, K\}|$ and $\mathbf{t}_{:,J+1,K} = \delta \mathbf{1}_{I \times 1}$. This verifies that $t_{i,J+1,K} = \delta = \sum_{r=1}^R a_{ir} b_{J+1,r} c_{K,r} = \delta \sum_{r=1}^R a_{ir} c_{K,r}^{-1} c_{K,r}$, corresponding to $\sum_{r=1}^R a_{ir} = 1$. Finally, \mathcal{T} and \mathbf{B} have to be updated at the end of each full AO-ADMM iteration as follows:

$$\tilde{\mathcal{T}} = [\mathcal{T} \mid \mathbf{T}_{:,J+1,:}], \quad \tilde{\mathbf{B}} = \begin{bmatrix} \mathbf{B} \\ \mathbf{b}_{J+1,:} \end{bmatrix}, \quad (14)$$

where δ is usually assigned the mean of \mathcal{T} . Then, (3) becomes:

$$\begin{aligned} \argmin_{\mathbf{A}, \tilde{\mathbf{B}}, \mathbf{C}} \|\tilde{\mathcal{T}} - \mathbf{\Lambda} \bullet_1 \mathbf{A} \bullet_2 \tilde{\mathbf{B}} \bullet_3 \mathbf{C}\|_F^2 + \alpha \|\mathbf{A}\|_1 \\ \text{s.t. } \mathbf{A} \succeq 0, \tilde{\mathbf{B}} \succeq 0, \mathbf{C} \succeq 0 \end{aligned} \quad (15)$$

At this stage, solving (15) with AO-ADMM becomes simple. We demonstrate the solution for one ADMM update of \mathbf{A} . As for $\tilde{\mathbf{B}}$ and \mathbf{C} , they follow the same procedure as that of \mathbf{A} but where $\alpha = 0$ since sparsity is not imposed on them. Supposing that $\tilde{\mathbf{T}}_{(1)}$ represents the first-mode unfolding of $\tilde{\mathcal{T}}$, we can write the sub-problem of \mathbf{A} as follows:

$$\begin{aligned} \mathbf{A} = \argmin_{\mathbf{A}} \frac{1}{2} \|\tilde{\mathbf{T}}_{(1)} - \tilde{\mathbf{W}}_A \mathbf{A}^\top\|_F^2 + \alpha \|\mathbf{A}\|_1 \\ \text{s.t. } \mathbf{A} \succeq 0 \end{aligned} \quad (16)$$

where $\tilde{\mathbf{W}}_A = \tilde{\mathbf{B}} \odot \mathbf{C}$ represents the Khatri-Rao product [41]. By introducing the splitting variable $\bar{\mathbf{A}} = \mathbf{A}^\top$, expression (16) becomes:

$$\begin{aligned} \argmin_{\mathbf{A}, \bar{\mathbf{A}}} \frac{1}{2} \|\tilde{\mathbf{T}}_{(1)} - \tilde{\mathbf{W}}_A \bar{\mathbf{A}}\|_F^2 + \alpha \|\mathbf{A}\|_1 \\ \text{s.t. } \bar{\mathbf{A}} = \mathbf{A}^\top \text{ and } \mathbf{A} \succeq 0 \end{aligned} \quad (17)$$

Adopting ADMM for (17), the updates of $\bar{\mathbf{A}}$ and \mathbf{A} become:

$$\begin{aligned} \bar{\mathbf{A}} &\leftarrow (\tilde{\mathbf{W}}_A^\top \tilde{\mathbf{W}}_A + \rho \mathbf{I})^{-1} (\tilde{\mathbf{W}}_A^\top \tilde{\mathbf{T}}_{(1)} + \rho(\mathbf{A} + \mathbf{U}_A)^\top) \\ \mathbf{A} &\leftarrow \max(0, \bar{\mathbf{A}}^\top - \mathbf{U}_A - \frac{\alpha}{\rho}) \\ \mathbf{U}_A &\leftarrow \mathbf{U}_A + \mathbf{A} - \bar{\mathbf{A}}^\top \end{aligned} \quad (18)$$

where U_A is called the dual variable. Finally, for order-2 tensors, this model becomes equivalent to solving NMF (4). The implementation of AO-ADMM-ASC is summarized in Algorithm 1.

Algorithm 1 AO-ADMM-ASC: The set $\{H_1, H_2, H_3\}$ replaces $\{A, B, C\}$ from Section III-A. In (18), $\alpha = 0$ for the updates of B and C since sparsity is not imposed on them.

Require: $\mathcal{T}, H_1, H_2, H_3, U_1, U_2, U_3, \alpha$

Initialize H_1, H_2, H_3 ;

Initialize U_1, U_2, U_3 to zero;

repeat

Set \tilde{H}_2 and $\tilde{\mathcal{T}}$ based on (14);

Set $H_2 \leftarrow \tilde{H}_2$ and $\mathcal{T} \leftarrow \tilde{\mathcal{T}}$;

for $\forall d \in \{1, 2, 3\}$ **do**

$W_d = \odot_{j \neq d} H_j$;

$\rho = \text{trace}(W_d^T W_d) / R$; [49]

Update H_d with (18) as an ADMM sub-problem;

end for

Normalize the columns of H_2 and H_3 (ℓ_2 -norm) into Λ such that $\lambda_{rrr} = \|\mathbf{h}_{:,r}^{(2)}\|_2 \|\mathbf{h}_{:,r}^{(3)}\|_2 \quad \forall r \in \{1, \dots, R\}$

until Termination criterion (e.g., number of iterations)

return H_1, H_2, H_3, Λ

B. Interpretation of ELMM Based on TD

Here, we build upon what has been presented in Sections II-B and II-C as methodological, physical, and graphical basis for the multimodal HU interpretation. For that, we first draw the analogies across the aforementioned expressions (in Sections II and III) in terms of links between the frontal slices of the tensor and the meaning of the SV function in multimodal HU, which helps to elaborate the interpretation of the tensor-based model at the basis of that of the matrix-based one. Then we visualize the expressions in order to interpret multimodal HU in terms of graphical representations as subspaces while commenting on the physical role of the extracted factors and the input rank.

First, we note that expressions (10) and (11) are analogous to the ELMM expression (7) with the major difference that in CPD, ELMM scaling factors are *dependent on frontal slices* instead of pixel dependent. Second, we visualize (6) and (7) in Fig. 4, and (10) and (11) in Fig. 5. Looking at (10), the frontal slices $T_{::,k}$ and the physical meaning that they represent have a direct influence on the SV function f_k and the interpretation of the SVs, which is simply reflected as scaling factors in each row of C , i.e. $c_{k,:}$ (or $\Psi_{(k)} = \text{diag}\{c_{k,:}\}$). As a result, with A and B fixed, each frontal slice $T_{::,k}$ is inherently factorized into a set of R scaled EMs $\{c_{kr} \mathbf{b}_r\}_{r \in \{1, \dots, R\}}$ where the SVs are absorbed in $c_{k,:}$ and the EMs $\{\mathbf{b}_r\}$ are independent of the frontal slices. This also means that R is a major parameter that represents the degrees of freedom especially the scaling factors c_{kr} , which then simultaneously represent the *factorized evolution of sources across the third mode* as $c_{:,r}$ and the *per-slice modeling of the SV* as $c_{k,:}$. Intuitively, one can say that the effects of the transformations (e.g., scale, illumination) [31] or the natural evolution (e.g., time) [24] of a scene held by

$T_{::,k}$ on the sources become observed through the scaling factors of the TD. Moreover, we point out the following:

- In CPD, there are as many SV functions (and simplices) as the frontal slices of the tensor, which is significantly lower than the number of pixels ($K \lll I$).
- In CPD, one row of R scaling factors in C corresponds to a full frontal slice and is shared by all the pixel rows of A , while in classical ELMM, each row of scaling factors in Ψ corresponds to one pixel of M and interacts with only one row of A . This clearly appears when we compare (5) and (7) to (10) and (11), and Fig. 4 to Fig. 5.
- On a graphical representation, when ASC is imposed, CPD suggests that each simplex contains I pixels as illustrated in Fig. 6a, such that the relative coordinates of the pixels inside each convex hull are the same since each row of C interacts with all of A .

In summary, having a third mode in HSI produces scaling factors in ELMM that absorb the SVs based on the physical meaning of the frontal slices along the third mode (e.g., time, patches, MM), which balances the extracted factors in A and B independently of said SVs. Moreover, the imposed value of R represents the number of extracted sources and scaling factors and has a major effect on the results and the SV interpretation. As R decreases, we tend towards having fewer degrees of freedom, where CPD tends towards extracting the EMs while applying a regularization on the observed pixels influenced by the physical meaning of the information across the third mode. As R increases, we tend towards having more degrees of freedom, where CPD tends towards extracting factors with multi-feature separability of the sources. We note that R should not be too high in order to avoid over-fitting.

C. Spatial Features (Patches/MM)

In this section, we consider including spatial features as a third mode in multifeature HU through two examples, revisiting patches with additional insights and introducing MM, which have not been explored yet and consequently help demonstrate the interpretability of the model especially in terms of physical significance and the variation of the rank. We also refer to Fig. 5 and 6 for illustration.

1) *Patches*: Here, we add that, in fact, a tensor built with patches has an inherently low-rank structure as that of the NMF of the original HSI, represented by $M = T_{::,1}$, since the information contained along the third mode of the tensor is almost essentially the same; that is, the frontal slices in Fig. 5 are just *spatially-shifted versions* of the original image $T_{::,1}$, and the values of these shifts correspond to a *small spatial kernel* (e.g., 3×3 or 5×5). However, what is different in the tensor case is that when this *shifting* information is stacked along the third mode and CPD is imposed with such a low rank, the model applies an implicit smoothing of the pixel tubes belonging to the same patch (which represents a horizontal slice in \mathcal{T}) since the frontal slices are jointly decomposed (through CPD) sharing A and B but with the degrees of freedom (the rank) of a single one of them. Therefore, a main advantage over NMF is that one expects to extract the same sources with a patch-local smoothing of

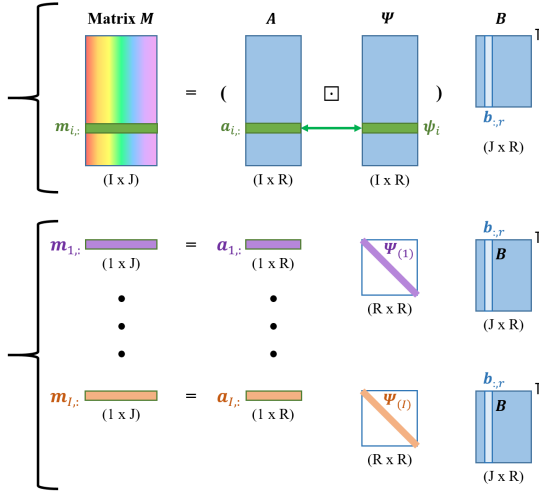


Fig. 4: Our visualization of equations (6) (Hadamard product) and (7) (matrix product). The color code of the bottom part follows that of Fig. 3b. We have $\Psi_{(i)} = \text{diag}\{\Psi_{i,:}\}$.

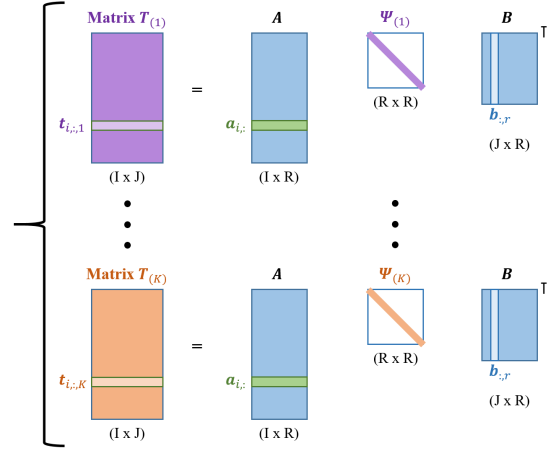


Fig. 5: Our visualization of equation (10). The color code is made analogous to that of Fig. 2 and follows that of Fig. 6a. We have $\Psi_{(k)} = \text{diag}\{c_{k,:}\}$.

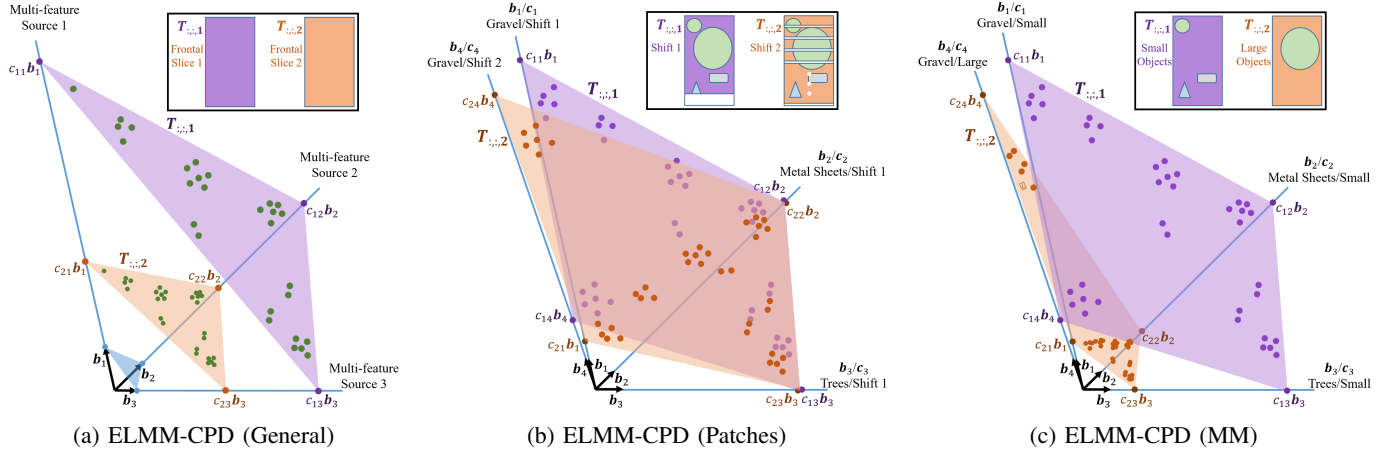


Fig. 6: Graphical representations of (a) CPD with rank $R = 3$ in the case of three spectral signatures $\{b_1, b_2, b_3\}$ and two frontal slices $\{T_{::,1}, T_{::,K}\}$, and (b) patch-CPD and (c) MM-CPD with rank $R = 4$ in the case of four spectral signatures $\{b_1/c_1, \dots, b_4/c_4\}$ (two of which have different third-diversity patterns) and two frontal slices $\{T_{::,1}, T_{::,2}\}$. The relative coordinates of the pixels in the convex hulls must be the same since A is common for all the frontal slices.

the SVs of the estimated EMs, where the SVs are balanced out in the form of scaling factors stored in the rows of C . An important note here is that the scaling factors stored in C may not have a significant physical meaning.

This brings us to another problem, which occurs when R increases. Then, the sources and abundances are expected to replicate since the third mode is based on *shifting* information, implying *redundancy*. In this case, the scaling factors in C only indicate whether an estimated EM in B corresponds to a certain spatial shifting or another. This point is roughly illustrated in Fig. 6b (inspired by Fig. 10) where we have three spectral sources: Gravel, Metal Sheets, and Trees, but CPD is carried out with $R = 4$. Here, the convex hull of $T_{::,1}$ gives a high scaling factor at b_1 and a low factor at b_4 , while that of $T_{::,2}$ gives the opposite with almost the same quantity. This is due to the fact that the materials are present with almost the same quantity in both frontal slices. In other words, there

may be a redundancy problem if some components account for the same material with patches, which does happen in practice. This problem does not appear in the case where the third mode represents a physical meaning of the components such as MM.

2) *Mathematical Morphology*: While using patches is efficient, it still ignores the physical properties of connected pixels, and the SVs are regularized indifferently among pixels belonging to different types of materials. On the other hand, morphological features take into account physical properties such as scale and brightness of objects and promote dealing with SV among pixels sharing these properties. Through MM, we also emphasize the role of incorporating spatial diversities that add physical significance to the objects of the scene. In this sense, one expects that a MM-tensor has an inherently more complicated structure than that of a patch-tensor since the frontal slices in Fig. 5 contain more context on the materials related to their sizes and brightness levels. As such,

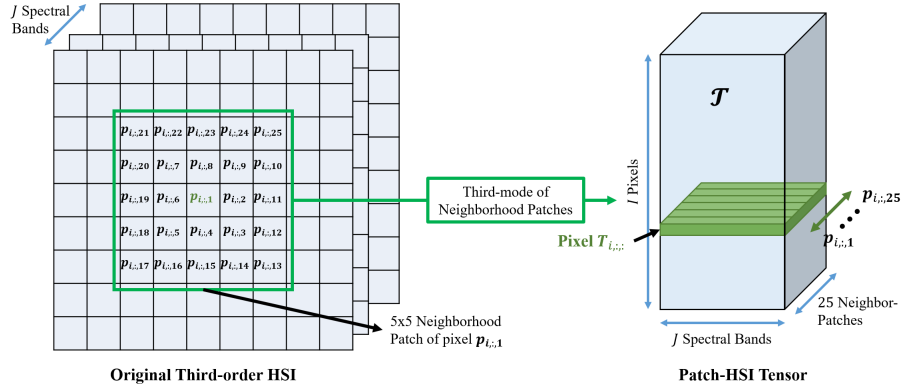


Fig. 7: An illustration of constructing a 5×5 Patch-HSI tensors based on [14].

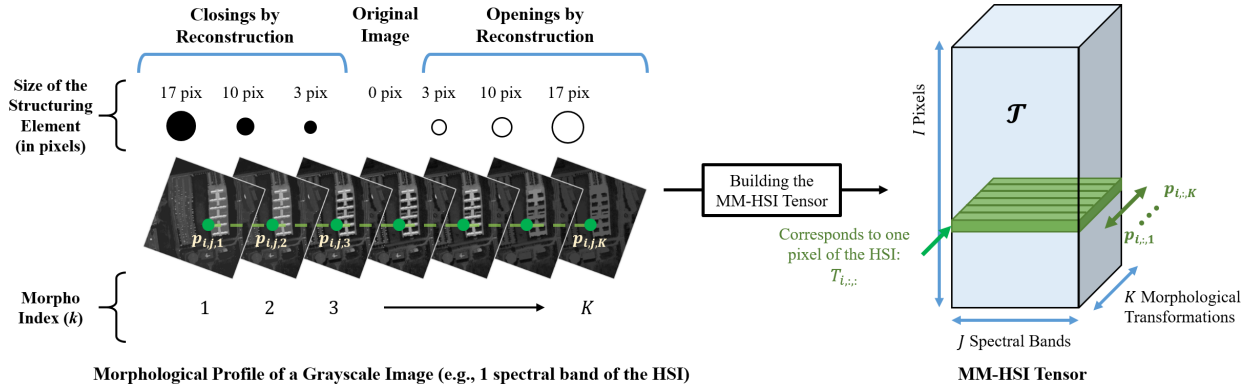


Fig. 8: Example of a sequential morphological filtering of a grayscale image (corresponding to one spectral band of the HSI of Pavia University) with Openings and Closings by Reconstruction using successive sizes of the Structuring Element, which is a disk in this case. The stacking of the transformations, with the original image corresponding to 0 size being placed in the middle, is referred to as the Morphological Profile (MP) of the image. The yellow spots mark the same pixel positioning in each of the transformations. Then, to create the MM-HSI tensor [31], the pixel positioning modes are reordered into lexicographic order along the first mode, and the yellow spots are stored along the third mode, in order to create the tensor.

imposing a low rank promotes spectral smoothing of the SVs based on a morphological regularization of the abundances, while imposing a sufficiently higher rank promotes having a more distinctive spectral-morphological separation of the materials unlike patch-tensors. In this case, since each frontal slice is seen as a characteristic of brightness (which is particularly relevant for ELMM because scaling factors can be directly linked to brightness) and spatial scale (i.e., size of objects in the scene), the scaling factors in \mathbf{C} indicate the quantitative correspondence of an extracted material to the aforementioned physical properties per frontal slice. This point is roughly illustrated in Fig. 6c (inspired by Fig. 11) where $\mathbf{T}_{:,1}$ and $\mathbf{T}_{:,2}$ characterize small and large objects respectively. Here, the convex hull of $\mathbf{T}_{:,1}$ gives high scaling factors for \mathbf{b}_1 , \mathbf{b}_2 , and \mathbf{b}_3 (corresponding to small objects), and a low scaling factor at \mathbf{b}_4 (corresponds to large objects), while that of $\mathbf{T}_{:,2}$ gives the opposite.

IV. EXPERIMENTS AND RESULTS

In this section, we discuss the experiments and results of multimodal HU on real HSIs in terms of AO-ADMM-ASC (compared to Naive ASC [14]), extracted factors, ELMM interpretability, and qualitative comparisons between patch-

and MM-tensors with low and high values of the rank. In each experiment, among 30 random initializations of the factor matrices, the result with the minimum root mean squared error (RMSE) is chosen. The estimated EMs of \mathbf{B} are identified based on their minimum spectral angular distance (SAD), in degrees, with respect to the reference.

The maps and plots shown in the experiments represent the *columns* of the factor matrices. Above each abundance map (AM), we show the material that corresponds to it with its minimum SAD value. We recall that a set of similarly indexed columns, e.g., $\{\mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_1\}$, represent the abundance, spectral source signal, and third-mode source pattern (e.g., morphological print, shifting print, temporal evolution) of one extracted material respectively. Since \mathbf{C} plays a crucial role in the interpretability of ELMM and multifeature HU, we also highlight the relevance of its row components. Each fixed index k in the plot corresponds to a row of \mathbf{C} and thus to a frontal slice in \mathcal{T} , and the vertical grouping of points at said index, as indicated in Fig. 10c, 11c, 12c, and 13c, represents the scalars in that row. This also means that when projected on Fig. 6, in patch-CPD, $k=1$ contains the SV scaling factors that balance the convex hull of the original HSI and are responsible for its reconstruction, while in MM-CPD, the

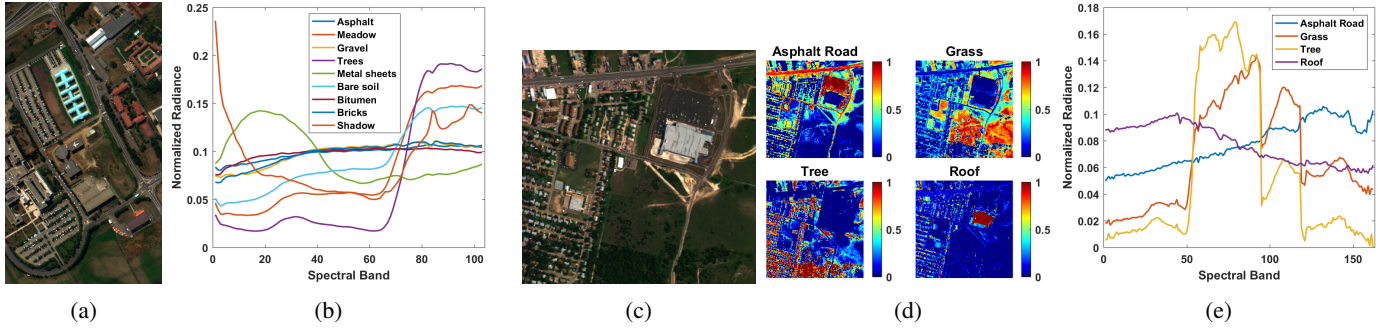


Fig. 9: 9a) Pavia in false colors. 9b) Pavia's spectral reference, which was extracted from the mean spectrum of each class of its spatial GT. 9c) Urban in false colors. 9d) Urban's spatial GT. 9e) Urban's spectral reference.

middle index does that. This will eventually show how MM accounts to physical effects in the scene, while patches does not.

In each case analysis, we look into the components of \mathbf{A} and \mathbf{B} first, which visually and spectrally identify the materials, then we explain their correspondence to those of \mathbf{C} , where we are interested in the significance of the third-mode patterns then their relevance to the original HSI. We note that quantitative validation of the AMs and EMs is usually not evident, especially in the case of the Urban HSI in Fig. 9d where the spatial ground-truth (GT) is not a real GT and is not usable for quantitative comparison. Moreover, there is neither a quantitative nor a qualitative reference for third-mode patterns in the literature, so here we provide an in-depth qualitative analysis.

We want to consider HSIs which show objects with spatial features of different scale and brightness levels, for which urban areas are good candidates. For that, we choose two real HSIs³: *Pavia University* and *Urban*. Fig. 9 shows the two HSIs in false colors with their spectral references. It is worth noting for the sake of interpretation that the MM-tensors are built following the Extended Morphological Profile (EMP) technique used in [31] with Openings and Closings by Reconstruction (ObR and CbR), which correspond to bright and dark objects respectively, with varying sizes of the Structuring Element (SE), which in turns corresponds to the scales of objects⁴. In the following experiments, 4 sizes of the SE are used, corresponding to 8 ObR and CbR transformations, then the dimension of the third mode is $K = 9$ where the original image (corresponding to scale 0) is placed in the middle in the order shown in Fig. ?? for a grayscale image. The patch-tensors are built following [14] as shown in Fig. 7 with 3×3 patches. This means that the dimension of the third mode is $K = 3 \times 3 = 9$ where the original image is placed at $k = 1$. We run our experiments with Intel® Core™ i7-1185G7, 32GB RAM 3200MHz LPDDR4.

A. Results Discussion - Pavia University

In this section, we present the experiments of the HSI of Pavia. We start by comparing AO-ADMM-ASC and Naive

ASC. After that, we focus on CPD and the ELMM analysis of the factors while interpreting the cases of patches and MM. For the MM-tensor, our SEs are disks with the successive radii: $\{2, 7, 12, 17\}$ pixels. Both patch- and MM-tensors then have $K=9$ frontal slices and dimensions $207400 \times 103 \times 9$. Finally, we find that $R=4$ and $R=8$ are the best for low and high values of the rank respectively. In Fig. 9b, some classes have very similar spectral signatures, so, in the following, sometimes we refer to *Trees* and *Meadows* as *vegetation*, and to *Asphalt*, *Bitumen*, *Gravel*, and *Bricks* as *roads* or *roofs*. *Bare Soil* can belong to either of both groups.

1) *AO-ADMM-ASC*: Here, we compare the RMSE results of MM-CPD between AO-ADMM-ASC and Naive ASC [14]. The results are shown in Table I, where we see that with AO-ADMM-ASC we gain in RMSE, which corresponds to a better estimation of the factors with respect to the observed tensor with a small difference in the execution time.

Algorithm	R	RMSE %	Time (s)
Naive ASC [14]	8	7.07	231
AO-ADMM-ASC	8	6.34	384

TABLE I: Pavia. The results of AO-ADMM-ASC and Naive ASC in terms of RMSE and execution time: R is the rank. The results of the minimum RMSE are shown.

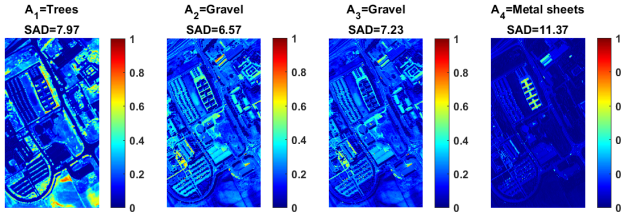
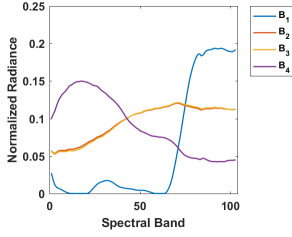
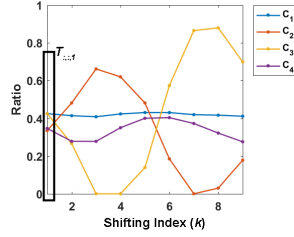
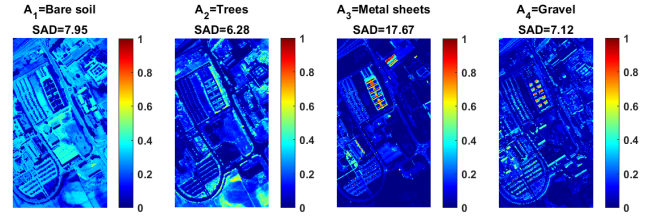
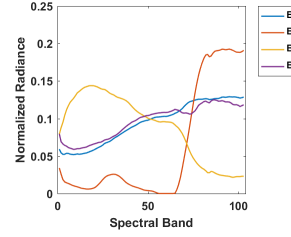
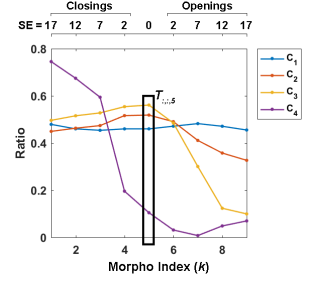
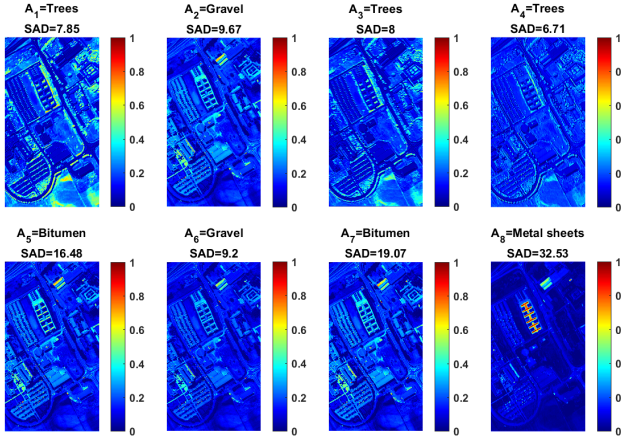
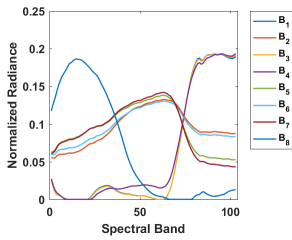
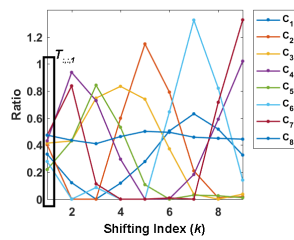
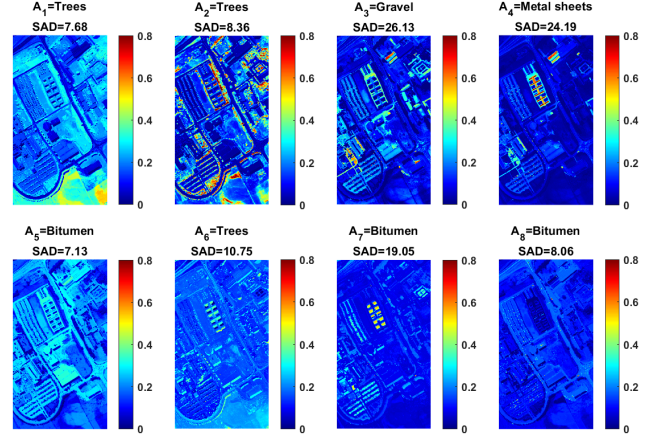
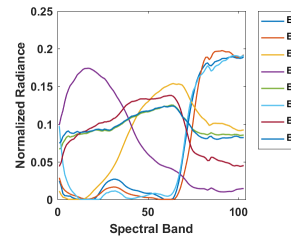
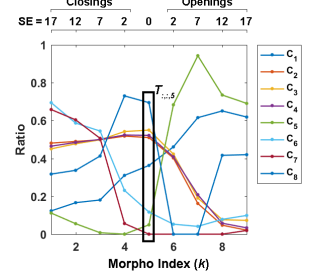
2) *Low rank, ELMM and SV*: Here, we are interested in the property of CPD dealing with SV. Since patch-tensors have as an inherently low structure as that of NMF, we start by considering $R=4$ for patches and MM. As explained in Section III-B, such a low rank highlights the SV and *spatial regularization* aspects of multi-feature HU. We compile the discussion into two stages: (a) Fig. 10 representing patch-CPD, and (b) Fig. 11 representing MM-CPD.

2a) In Fig. 10, \mathbf{B}_1 and \mathbf{B}_4 are identified as *Trees* and *Metal Sheets* respectively, which reflects the areas highlighted in \mathbf{A}_1 and \mathbf{A}_4 , while \mathbf{B}_2 and \mathbf{B}_3 coincide and are both identified as *Gravel*, which reflects the areas highlighted in \mathbf{A}_2 and \mathbf{A}_3 and where we already start to see *replicated* components. In fact, \mathbf{A}_2 and \mathbf{A}_3 are *slightly-shifted* versions of each other, which is explained better in the interpretation of \mathbf{C}_2 and \mathbf{C}_3 .

Looking at Fig. 10c, which is the most interesting, one might intuitively expect to see horizontal curves since, quantitatively, the collective SV is supposedly *constant* in patches.

³The data sets with detailed information are available on the website: <http://lesun.weebly.com/hyperspectral-data-set.html>

⁴This process of constructing MM-tensors is detailed in Section 3 of [31].

(a) Components of A (b) Components of B (c) Components of C Fig. 10: Pavia. CPD results of the Patch-tensor for $R=4$ (a) Components of A (b) Components of B (c) Components of C Fig. 11: Pavia. CPD results of the MM-tensor for $R=4$ (a) Components of A (b) Components of B (c) Components of C Fig. 12: Pavia. CPD results of the Patch-tensor for $R=8$ (a) Components of A (b) Components of B (c) Components of C Fig. 13: Pavia. CPD results of the MM-tensor for $R=8$

However, while C_1 and C_4 look almost straight, C_2 and C_3 are not. We notice here that where C_2 is high, C_3 is low and vice versa. In part, this means that A_2 represents the shifts where $k=\{2, 3, 4, 5\}$, while A_3 represents those where $k=\{6, 7, 8, 9\}$. In another part, the two columns fluctuate in a way that *maintains* a constant SV and *balances out* their quantitative presence across the frontal slices. As for $k=1$, which represents the original HSI, we notice that the scaling factors are almost equal, which means that the spectral vectors

of the convex hull are *equally present* in the HSI, all of which shows that patch-tensors do not account to physical spatial effects.

2b) In Fig. 11, B_1 , B_2 , and B_3 are identified as *Bare Soil*, *Trees*, and *Metal Sheets* respectively, which reflects the areas highlighted in A_1 , A_2 , and A_3 , all of which is similar to those obtained by patches. As for $\{A_4, B_4\}$, while B_4 looks very similar to B_1 , unlike patches, we notice that A_4 highlights interesting shadow areas (dark features), which

clearly reflects the *morphological awareness* incorporated into CPD with MM. The latter becomes more interesting with the interpretation of C .

Looking at Fig. 11c, we observe three main patterns that can be associated to the chosen morphological parameters. First, C_4 corresponds to dark features (reflected by the shadows in A_4) as it has higher values when k corresponds to CbR, then continues decreasing towards ObR. Second, C_2 and C_3 correspond to small features as they have higher values around the middle ($k=5$) where the SEs are small, which is visually reflected through the small objects highlighted in A_2 (trees) and A_3 (metal sheets and vehicles). Third, C_1 is rather steady, which means that the spatial features shown in A_1 are general.

As for $k=5$, which represents the original HSI, we notice that C_2 and C_3 have the highest scaling factors since they correspond to relatively bright objects of the scene, C_1 has a slightly lower factor since it corresponds to darker objects like asphalt roads, building roofs, parking lots, and bare soil areas, and C_4 has the lowest factors since it corresponds to dark shadows. These relationships show the column- and row-wise significance of C and how multi-feature HU can balance out the SVs and the same time reconstruct the original HSI.

3) *High rank, ELMM and multi-feature separability*: Here, we are interested in multi-feature HU when we have more degrees of freedom, where we go more in-depth into the factors of patches and MM for $R=8$. As explained in Section III-B, such a higher rank demonstrates the multi-feature separability of MM and how patches only replicates its components. We compile the discussion into two stages: (a) Fig. 12 representing patch-CPD, and (b) Fig. 13 representing MM-CPD.

3a) In Fig. 12, we end up with more replicas of the same A and B components obtained in Fig. 10, the sets of replicas being columns $\{1, 3, 4\}$ detected as *Trees*, and $\{2, 5, 6, 7\}$ detected as *Gravel* and *Bitumen*. As for the plot of C , the same remarks of Fig. 10c about balancing the constant SVs in patches apply on Fig. 12c, but since there are many replicas, the figure becomes hard to read. Finally, we notice again that for $k=1$, the scaling factors are almost equal.

3b) In Fig. 13, we notice that multi-feature HU is done based on spectral and morphological properties, where we observe three column sets: $\{1, 2, 6\}$, $\{3, 5, 7, 8\}$, and $\{4\}$. B_1 , B_2 , and B_6 are identified as vegetation, which respectively reflects the areas highlighted in A_1 (big vegetation areas like meadow), A_2 (small vegetation areas like trees), and A_6 (dark shadows on vegetation areas). Therefore, while B_1 , B_2 , and B_6 are collinear, unlike, patches, their AMs highlight interesting features accounting to the scale and brightness of vegetation objects, which we discuss more in-depth with the interpretation of C and which applies to the other components as well. B_3 , B_5 , B_7 , and B_8 are identified as *Bitumen* and *Gravel*, which respectively reflects the areas highlighted in A_3 (small or short areas of roads and roofs), A_5 (big connected areas of roads, roofs, and parking lots), A_7 (dark shadows on parking lots and buildings), and A_8 (tiny bright vehicles). Finally, B_4 is identified as *Metal Sheets*, which are small.

Looking at Fig. 13c, we observe four main patterns that can be associated to the chosen morphological parameters: First, C_6 and C_7 correspond to dark features (as observed in A_6 and

A_7) as they have higher values when k corresponds to CbR, then continue decreasing towards ObR. Second, C_2 , C_3 , and C_4 correspond to small features (as observed in A_2 , A_3 , and A_4) as they have higher values around $k=5$ where the SEs are small. Third, C_1 and C_5 correspond to big features (as observed in A_1 and A_5) as they have higher values when k corresponds to big SEs with ObR. Fourth, C_8 corresponds to the tiny vehicles as it is the highest when k corresponds to the smallest SE. Finally, we talk about the original HSI.

As for $k=5$, we notice that C_2 , C_3 , C_4 , and C_8 have the highest scaling factors (relatively bright objects including the vehicles), C_1 and C_5 have lower factors (darker objects like asphalt roads, building roofs, parking lots, bare soil, and meadow areas), and C_6 and C_7 have the lowest factors (dark shadowy features). These relationships showcase the separability of multi-feature HU when the third-mode has a significant physical meaning and when the rank is set to be sufficiently high, which can also be interpreted in terms of ELMM and balancing the SV factors.

B. Results - Urban

In this section, we present the experiments of the Urban HSI following the same order of Pavia. Since we have the same observations, and in order to avoid repetition, we briefly go over the results. For the MM-tensor, our SEs are disks with the successive radii: $\{1, 4, 7, 10\}$ pixels. Both patch- and MM-tensors then have $K=9$ frontal slices and dimensions $94249 \times 162 \times 9$. Finally, we also choose $R=4$ and $R=8$ for the rank.

1) *AO-ADMM-ASC*: Table II shows the RMSE results of MM-CPD between AO-ADMM-ASC and Naive ASC [14], where again with AO-ADMM-ASC we gain in RMSE with a small difference in the execution time.

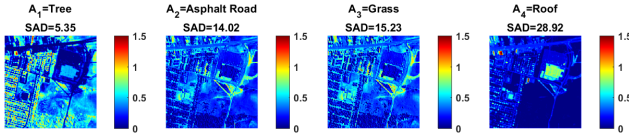
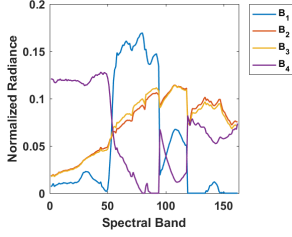
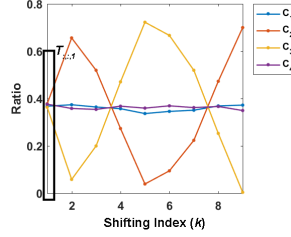
Algorithm	R	RMSE %	Time (s)
Naive ASC [14]	8	7.88	124
AO-ADMM-ASC	8	6.87	251

TABLE II: Urban. The results of AO-ADMM-ASC and Naive ASC in terms of RMSE and execution time: R is the rank. The results of the minimum RMSE are shown.

2) *Low rank, ELMM and SV*: Here, we discuss the results for $R=4$, where (a) Fig. 14 represents patch-CPD (b) Fig. 15 represents MM-CPD. A fast look at the figures shows that we have the same observations as those of Pavia:

2a) In Fig. 14, B_2 and B_3 coincide, A_2 and A_3 are replicas, and they represent *Asphalt+Grass*. Moreover, in Fig. 14c, we see the same patterns and fluctuations that were observed in Fig. 10c related to the constant SV and its quantitative balance in patches, and the scaling factors are equal for $k=1$. On the other hand, $\{A_1, B_1\}$ and $\{A_4, B_4\}$ represent *Tree+Grass* and *Roof* respectively with steady C_1 and C_4 patterns.

2b) In Fig. 15, while B_4 looks similar to B_1 , we notice that C_4 has the same pattern observed in Fig. 11c, which corresponds to dark shadows and is reflected in A_4 , which highlights shadows of buildings and trees that fall on grass areas. As for the other components, they can be interpreted

(a) Components of A (b) Components of B (c) Components of C Fig. 14: Urban. CPD results of the Patch-tensor for $R=4$

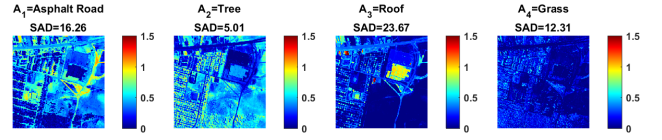
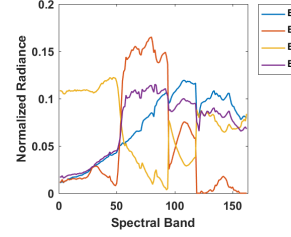
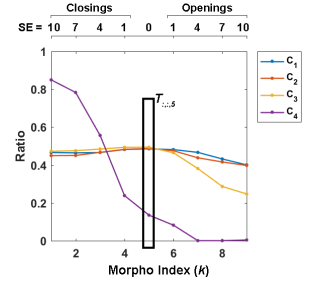
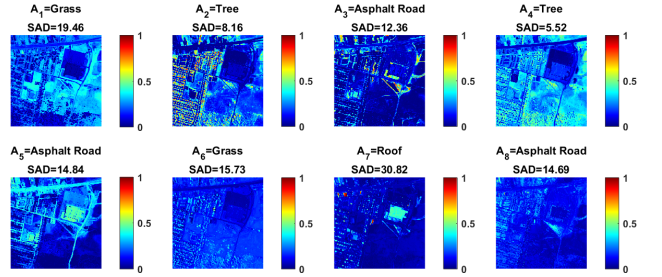
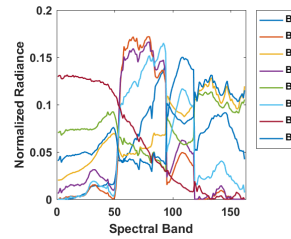
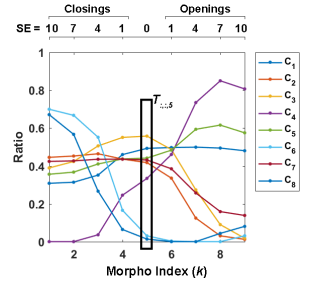
similarly to those in the case of Pavia (including for $k=5$), where $\{A_1, B_1, C_1\}$, $\{A_2, B_2, C_2\}$, and $\{A_3, B_3, C_3\}$ represent *Asphalt+Grass*, *Tree+Grass*, and *Roof* respectively.

3) *High rank, ELMM and multi-feature separability*: Here, we discuss the results for $R=8$, where Fig. 16 represents MM-CPD. We skip the case of patch-CPD in order to avoid repetition, where we simply end up with more *replicas* of the components of figures 14a and 14b. In Fig. 16, we are interested in the features of the AMs that do not appear in patch-CPD as the comments on the spectral and morphological patterns are the same as those of Pavia; where the plots reflect the qualitative features that appear in the respective AMs.

We observe three column sets: $\{1,2,6\}$, $\{3,5,7,8\}$, and $\{4\}$. First, A_1, A_2, A_4 , and A_6 were identified as *Vegetation*: A_1 highlights grass fields, which is close to the *Grass* reference and does not appear in patch-CPD. A_2 and A_4 highlight small and big areas, and together they correspond to the *Tree* reference. A_6 highlights dark shadows (which is reflected in C_6). Second, A_3, A_5 , and A_8 are identified as *Asphalt Road*: A_8 seems to correspond to dark features (looking at C_8). A_3 highlights small roads such as dirt and narrow streets, while A_5 highlights large roads like the main and connected roads, which are clearly highlighted unlike the case of patches. Third, A_7 is identified as *Roof* and highlights both small and large building roofs.

V. CONCLUSION

In this paper, we proposed a methodological framework for multifeature HU based on CPD and the AO-ADMM-ASC algorithm, where the samples (pixels) represent a convex combination of the sources. We also established a unified framework for the interpretability of multifeature HU into “multilinear” subspaces which involved mathematical, physical, and graphical representations of the CPD model with ELMM and SV. Finally, we proposed to include MM as spatial features in a spectral-spatial HU and provided in-depth insights on how patches and MM behave within the aforementioned framework in terms of the interpretability of the factors and the

(a) Components of A (b) Components of B (c) Components of C Fig. 15: Urban. CPD results of the MM-tensor for $R=4$ (a) Components of A (b) Components of B (c) Components of C Fig. 16: Urban. CPD results of the MM-tensor for $R=8$

variation of the input rank, where MM incorporates physically meaningful features into the data tensor.

In the future, we plan to explore the tensor Block Term Decomposition which allows some flexibility with the tensor structure and can be seen as an extension to Spectral Bundles for SV [11], but also comes with many challenges such as the rank and the interpretation of the subspaces. Moreover, areas of BSS other than HU may be explored. Finally, it is worth mentioning that some Deep Learning (DL) approaches are being considered for HU (which still suffer from the increasing and flexible dimensionality of HSIs and the difficulty of finding data sets for training especially in a blind framework). However, by developing our methodological study of tensor-based unmixing and understanding its subspace structure and the interpretation of its results, we can pave the way towards

implementing interpretable DL models based on tensor structures and tensor subspace learning [22], [56], [57], which can be very helpful in blind settings such as BSS and HU. For example, in [20], some relationships between TD and Neural Networks are drawn for supervised scene classification, and in [58], an open question is posed whether Tensor Networks can work as Machine Learning models.

REFERENCES

- [1] W.-K. Ma, J. M. Bioucas-Dias, T.-H. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, and C.-Y. Chi, "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 67–81, 2013.
- [2] J. Amigo, *Hyperspectral Imaging*, ser. ISSN. Elsevier Science, 2019.
- [3] P. Comon and C. Jutten, Eds., *Handbook of Blind Source Separation, Independent Component Analysis and Applications*. Oxford UK, Burlington USA: Academic Press, 2010, hal-00460653.
- [4] G. Chabriel, M. Kleinstuber, E. Moreau, H. Shen, P. Tichavsky, and A. Yeredor, "Joint matrices decompositions and blind source separation: A survey of methods, identification, and applications," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 34–43, 2014.
- [5] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE transactions on audio, speech, and language processing*, vol. 14, no. 4, pp. 1462–1469, 2006.
- [6] T. Bajjouk, I. Zarati, L. Drumetz, and M. Dalla Mura, "Spatial characterization of marine vegetation using semisupervised hyperspectral unmixing," in *2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. IEEE, 2019, pp. 1–5.
- [7] T. Imbiriba, R. A. Borsoi, and J. C. M. Bermudez, "A low-rank tensor regularization strategy for hyperspectral unmixing," in *2018 IEEE Statistical Signal Processing Workshop (SSP)*. IEEE, 2018, pp. 373–377.
- [8] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1923–1938, 2018.
- [9] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, "Improved hyperspectral unmixing with endmember variability parametrized using an interpolated scaling tensor," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 2177–2181.
- [10] S. G. Azar, S. Meshgini, S. Beheshti, and T. Y. Rezaei, "Linear mixing model with scaled bundle dictionary for hyperspectral unmixing with spectral variability," *Signal Processing*, p. 108214, 2021.
- [11] R. Borsoi, T. Imbiriba, J. C. Bermudez, C. Richard, J. Chanussot, L. Drumetz, J.-Y. Tourneret, A. Zare, and C. Jutten, "Spectral variability in hyperspectral data unmixing: A comprehensive review," *IEEE Geoscience and Remote Sensing Magazine*, 2021.
- [12] L. Drumetz, M.-A. Veganzones, S. Henrot, R. Phlypo, J. Chanussot, and C. Jutten, "Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3890–3905, 2016.
- [13] L. Drumetz, M. Dalla Mura, G. Tochon, and R. Fablet, "Learning endmember dynamics in multitemporal hyperspectral data using a state-space model formulation," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 2483–2487.
- [14] M. A. Veganzones, J. E. Cohen, R. C. Farias, K. Usevich, L. Drumetz, J. Chanussot, and P. Comon, "Canonical polyadic decomposition of hyperspectral patch tensors," in *2016 24th European Signal Processing Conference (EUSIPCO)*. IEEE, 2016, pp. 2176–2180.
- [15] M. Jouni, "Image analysis based on tensor representations," Ph.D. dissertation, Université Grenoble Alpes, 2021.
- [16] Y.-X. Wang and Y.-J. Zhang, "Nonnegative matrix factorization: A comprehensive review," *IEEE Transactions on knowledge and data engineering*, vol. 25, no. 6, pp. 1336–1353, 2012.
- [17] Z. Yang, G. Zhou, S. Xie, S. Ding, J.-M. Yang, and J. Zhang, "Blind spectral unmixing based on sparse nonnegative matrix factorization," *IEEE Transactions on Image Processing*, vol. 20, no. 4, pp. 1112–1125, 2010.
- [18] F. Xiong, Y. Qian, J. Zhou, and Y. Tang, "Hyperspectral unmixing via total variation regularized nonnegative tensor factorization," *IEEE Transactions on Geoscience and Remote Sensing*, 2018.
- [19] Y. Tan, Z. Li, Y. Xiao, and N. Liu, "Infrared small target detection algorithm based on robust tensor decomposition model within bayesian framework," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 1160–1163.
- [20] K. Makantasis, A. Georgogiannis, A. Voulodimos, I. Georgoulas, A. Doulamis, and N. Doulamis, "Rank-r fnn: A tensor-based learning model for high-order data classification," *IEEE Access*, vol. 9, pp. 58 609–58 620, 2021.
- [21] W. He, Q. Yao, C. Li, N. Yokoya, and Q. Zhao, "Non-local meets global: An integrated paradigm for hyperspectral denoising," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6868–6877.
- [22] B. B. Gatto, E. M. dos Santos, A. L. Koerich, K. Fukui, and W. S. Junior, "Tensor analysis with n-mode generalized difference subspace," *Expert Systems with Applications*, vol. 171, p. 114559, 2021.
- [23] M. Zare, M. S. Helfroush, K. Kazemi, and P. Scheunders, "Hyperspectral and multispectral image fusion using coupled non-negative Tucker tensor decomposition," *Remote Sensing*, vol. 13, no. 15, p. 2930, 2021.
- [24] M. A. Veganzones, J. E. Cohen, R. C. Farias, J. Chanussot, and P. Comon, "Nonnegative tensor cp decomposition of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 2577–2588, 2016.
- [25] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Hyperspectral super-resolution: A coupled tensor factorization approach," *IEEE Transactions on Signal Processing*, vol. 66, no. 24, pp. 6503–6517, 2018.
- [26] C. Prévost, K. Usevich, P. Comon, and D. Brie, "Hyperspectral super-resolution with coupled Tucker approximation: Recoverability and svd-based algorithms," *IEEE Transactions on Signal Processing*, 2020.
- [27] K. Uto, M. D. Mura, and J. Chanussot, "Normal direction and true color estimation of leaves based on tensor decomposition of leaf-scale optical images," in *Whispers 2018 - Poster Session*, 2018.
- [28] Z. Xue, S. Yang, H. Zhang, and P. Du, "Coupled higher-order tensor factorization for hyperspectral and lidar data fusion and classification," *Remote Sensing*, vol. 11, no. 17, p. 1959, 2019.
- [29] M. Jouni, M. Dalla Mura, and P. Comon, "Classification of hyperspectral images as tensors using nonnegative CP decomposition," in *International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing*. Springer, 2019, pp. 189–201.
- [30] —, "Hyperspectral image classification using tensor CP decomposition," in *2019 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2019, Yokohama, Japan, July 28 - August 2, 2019*, 2019, pp. 1164–1167.
- [31] —, "Hyperspectral image classification based on mathematical morphology and tensor decomposition," *Mathematical Morphology - Theory and Applications*, vol. 1, 2019.
- [32] Y. Gu, T. Liu, and J. Li, "Superpixel tensor model for spatial-spectral classification of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, 2019.
- [33] Y. Mitsufuji, N. Takamune, S. Koyama, and H. Saruwatari, "Multi-channel blind source separation based on evanescent-region-aware non-negative tensor factorization in spherical harmonic domain," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 607–617, 2020.
- [34] M. Niknazar, H. Becker, B. Rivet, C. Jutten, and P. Comon, "Blind source separation of underdetermined mixtures of event-related sources," *Signal Processing*, vol. 101, pp. 52–64, 2014.
- [35] J. Virta and K. Nordhausen, "Blind source separation of tensor-valued time series," *Signal Processing*, vol. 141, pp. 204–216, 2017.
- [36] H. Becker, L. Albera, P. Comon, R. Gribonval, F. Wendling, and I. Merlet, "Brain-source imaging: From sparse to tensor models," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 100–112, 2015.
- [37] H. Becker, P. Comon, L. Albera, M. Haardt, and I. Merlet, "Multi-way space-time-wave-vector analysis for eeg source separation," *Signal Processing*, vol. 92, no. 4, pp. 1021–1031, 2012.
- [38] J. Sole-Casals, C. F. Caiafa, Q. Zhao, and A. Cichocki, "Brain-computer interface with corrupted eeg data: a tensor completion approach," *Cognitive Computation*, vol. 10, no. 6, pp. 1062–1074, 2018.
- [39] Z. Zhang, G. I. Allen, H. Zhu, and D. Dunson, "Tensor network factorizations: Relationships between brain structural connectomes and traits," *Neuroimage*, vol. 197, pp. 330–343, 2019.
- [40] P. Mishra, J. M. Roger, D. Jouan-Rimbaud-Bouveresse, A. Biancolillo, F. Marini, A. Nordon, and D. N. Rutledge, "Recent trends in multi-block data analysis in chemometrics for multi-source data integration," *TrAC Trends in Analytical Chemistry*, p. 116206, 2021.
- [41] P. Comon, "Tensors: a brief introduction," *IEEE Sig. Proc. Magazine*, vol. 31, no. 3, pp. 44–53, May 2014, hal-00923279.

- [42] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via $l_{1/2}$ sparsity-constrained nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4282–4297, 2011.
- [43] F. Zhu, Y. Wang, S. Xiang, B. Fan, and C. Pan, "Structured sparse method for hyperspectral unmixing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 88, pp. 101–118, 2014.
- [44] X. Xu, J. Li, S. Li, and A. Plaza, "Generalized morphological component analysis for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 4, pp. 2817–2832, 2019.
- [45] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 528–537, 2011.
- [46] S. Henrot, J. Chanussot, and C. Jutten, "Dynamical spectral unmixing of multitemporal hyperspectral images," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3219–3232, 2016.
- [47] J. Cohen, R. Farias, and P. Comon, "Fast decomposition of large nonnegative tensors," *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 862–866, 2015.
- [48] M. Jouni, M. Dalla Mura, and P. Comon, "Some issues in computing the cp decomposition of nonnegative tensors," in *International Conference on Latent Variable Analysis and Signal Separation*. Springer, 2018, pp. 57–66.
- [49] K. Huang, N. Sidiropoulos, and A. Liavas, "A flexible and efficient algorithmic framework for constrained matrix and tensor factorization," *IEEE Transactions on Signal Processing*, vol. 64, no. 19, pp. 5052–5065, 2016.
- [50] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.
- [51] L. Najman and H. Talbot, *Mathematical morphology: from theory to applications*. John Wiley & Sons, 2013.
- [52] P. R. Marpu, M. Pedergrana, M. Dalla Mura, S. Peeters, J. A. Benediktsson, and L. Bruzzone, "Classification of hyperspectral data using extended attribute profiles based on supervised and unsupervised feature extraction techniques," *International Journal of Image and Data Fusion*, vol. 3, no. 3, pp. 269–298, 2012.
- [53] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *International Journal of Remote Sensing*, vol. 31, no. 22, pp. 5975–5991, 2010.
- [54] M. Dalla Mura, J. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3747–3762, 2010.
- [55] A. Cichocki, R. Zdunek, A. Phan, and S. Amari, *Nonnegative Matrix and Tensor Factorizations*. Chichester: Wiley, 2009.
- [56] B. B. Gatto, K. Fukui, W. S. Júnior, and E. M. dos Santos, "Advances in subspace learning and its applications."
- [57] B. B. Gatto *et al.*, "Pattern-set representations using linear, shallow and tensor subspaces," 2020.
- [58] A. Cichocki, A.-H. Phan, Q. Zhao, N. Lee, I. V. Oseledets, M. Sugiyama, and D. Mandic, "Tensor networks for dimensionality reduction and large-scale optimizations: Part 2 applications and future perspectives," *Foundations and Trends® in Machine Learning*, vol. 9, no. 6, pp. 431–673, 2017.



Mohamad Jouni Mohamad Jouni received the B.Eng. degree in computer and communications engineering from the Lebanese University, Beirut, Lebanon, in 2016, and the M.Sc. and Ph.D. degrees in signal and image processing from the University of Grenoble Alpes, Grenoble, France, in 2017 and 2021 respectively. In 2019, he was a Visiting Researcher at Tokyo Institute of Technology, Tokyo, Japan. Since 2021, he has been a Postdoctoral Researcher at Grenoble Institute of Technology, Grenoble, France. His research interests include multi-

linear algebra, tensor decomposition, artificial intelligence, computational imaging, and hyperspectral image analysis.

PLACE
PHOTO
HERE

Mauro Dalla Mura Biography text here.

PLACE
PHOTO
HERE

Pierre Comon (Fellow, IEEE) received the Graduate degree in 1982, and the Doctorate degree in 1985, both from the University of Grenoble, France. He received the Habilitation to Lead Researches in 1995, from the University of Nice, France. He has been for nearly 13 years in industry, first with Crouzet-Sextant, Valence, France, between 1982 and 1985, and then with Thomson Marconi, Sophia Antipolis, France, between 1988 and 1997. He was with the ISL laboratory, Stanford University, CA, USA, in 1987. He joined in 1997 the Eurecom Institute, Sophia Antipolis, France. He is a Research Director with CNRS since 1998, first with the Laboratory I3S, Sophia Antipolis, France, until 2012, and then with Gipsa-Lab, Grenoble, France. He is currently the Director of Labex Persyval, Grenoble. His research interests include high-order statistics (HOS), blind techniques, statistical signal and array processing, tensor decompositions, multi-way factor analysis, and data science. He was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1995 to 1998, and a member of the French National Committee of Scientific Research from 1995 to 2000. He was the Coordinator of the European Basic Research Working Group on HOS, ATHOS, from 1992 to 1995. Between 1992 and 1998, he was a member of the Technical and Scientific Council of the Thomson Group. Between 2001 and 2004, he acted as a Launching Associate Editor with the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I, in the area of Blind Techniques. He has also been a member of the editorial board of the Elsevier journal Signal Processing from 2006 to 2011, and member of several IEEE Technical Committees. He was in the Editorial Board of the SIAM Journal on Matrix Analysis and Applications from 2011 to 2017. He received several prizes, including the Silver medal of CNRS in 2018. Dr Comon is also a Fellow of Eurasp and SIAM.

PLACE
PHOTO
HERE

Lucas Drumetz (Member, IEEE) received the M.Eng. degree from Grenoble INP in 2013, and the Ph.D degree in 2016 in image and signal processing from Université Grenoble Alpes, for works carried out at GIPSA-lab, Grenoble, France. This work has been awarded the PhD award of the University of Grenoble Alpes in 2017. In 2017, he was a Visiting Assistant Professor at the Department of Mathematics at the University of California, Los Angeles (UCLA). In 2017, he was also a visiting researcher for 10 weeks at the RCAST laboratory at the University of Tokyo, Japan. Since He has been an Associate Professor at IMT Atlantique, in the Mathematical and Electrical Engineering department since 2018. He is part of the OSE (Observations, Signal and Environment) team of UMR CNRS 6285 Lab-STICC. His research interests include inverse problems and machine learning for remote sensing applications, signal and image processing, and optimization techniques.