Spectral Unmixing with Multiple Dictionaries

Jérémy E. Cohen and Nicolas Gillis

***Abstract*—Spectral unmixing aims at recovering the spectral signatures of materials, called endmembers, mixed in a hyperspectral or multispectral image, along with their abundances. A typical assumption is that the image contains one pure pixel per endmember, in which case spectral unmixing reduces to identifying these pixels. Many fully automated methods have been proposed in recent years, but little work has been done to allow users to select areas where pure pixels are present manually or using a segmentation algorithm. Additionally, in a non-blind approach, several spectral libraries may be available rather than a single one, with a fixed number (or an upper or lower bound) of endmembers to chose from each. In this paper, we propose a multiple-dictionary constrained low-rank matrix approximation model that address these two problems. We propose an algorithm to compute this model, dubbed M2PALS, and its performance is discussed on both synthetic and real hyperspectral images.**

***Index Terms*—spectral unmixing, hyperspectral imaging, nonnegative matrix factorization, endmember extraction algorithms.**

I. Introduction

Developments of the remote sensing technology have given birth to efficient hyperspectral sensors producing high spectral resolution images, called hyperspectral images (HSIs). Researchers now have at their disposal a large quantity of HSIs of very different landscapes on the surface of the Earth, and even of the surface of Mars through the CRISM mission. One of the several uses for these HSIs is to determine and monitor the chemical composition of the scene being studied. Indeed, each material has a recognizable spectral signature that is captured by high resolution spectral sensors. Various types of similar materials can be identified, which allows, for instance, the monitoring of the evolution of the deforestation in the Amazonian forest [1], or of the melting of Alpine snow [2].

However, because of the high spectral resolution requirements, HSIs lack in spatial resolution. Therefore, there may be several materials in a single pixel, resulting in a mixture of their spectral signatures. Spectral unmixing is interested in finding these spectral signatures given the HSI. Spectral unmixing has been a central topic for signal processing research over the last ten to twenty years; see, e.g., the surveys [3]–[5]. In particular, two approaches to spectral unmixing can be distinguished: the blind approach where no *a priori* information is available other than known properties of spectra and abundances (such as non-negativity), and the non-blind approach where spectral libraries containing reference spectra are available. A list of well-known algorithms for both approaches is given in Section II.

a) Contributions: In this paper, we introduce a new model based on sparse coding that lies between blind and non-blind approaches, designed to offer new possibilities for dealing with several spectral libraries for a single HSI. For instance, it enables a simple user-control interface for choosing pure pixels areas in an HSI, based either on manual selection, or semi-supervised or unsupervised segmentation. The proposed model also allows for using multiple spectral libraries in the non-blind approach. Spectral signatures are then chosen from a set of known libraries, with the number of atoms to be picked in each dictionary (or an upper or lower bound).

b) Outline: Section II introduces the basic spectral unmixing problem cast in terms of sparse coding, along with some state-of-the-art methods to tackle it. Section III introduces the proposed multiple dictionary matrix factorization model and a fast algorithm dubbed M2PALS to tackle it featuring an assignment problem. Section IV shows the efficiency of the proposed model and algorithm on synthetic data built from the Urban HSI compared to state-of-the-art methods, and an illustration on the Urban HSI where pure-pixel areas are either manually selected or automatically selected using off-the-shelf segmentation methods.

c) Notations: Matrices are represented by case letters M, index sets by calligraphic letters $\mathcal{K}$, the ith column of matrix M is denoted $M(:,i)$ and $M(:,\mathcal{K})$ denotes the subset of columns indexed by $\mathcal{K}$. We also denote $\#\mathcal{K}$ the cardinality of $\mathcal{K}$, and M^T the transpose of M.

II. A combinatorial formulation of sparse coding

First, before introducing the new multiple dictionary matrix factorization model, let us recall the single dictionary formulation. Let M be a m-by-n data matrix. In this work, we assume the following:

1) The matrix M admits an approximate low-rank factorization $M \approx AB^T$ of size r, that is, A and B have r columns.
2) The noise is Gaussian. Missing data, stripes or impulsive noise are ignored although common in hyperspectral imaging. The factorization therefore can be written as $X = AB^T + N$ where N is the realization of a random variable following a white Gaussian distribution.
3) Columns of factor A are a subset of columns of the known m-by-d dictionary matrix D.

These assumptions leads to the following low-rank factorization model for M:

$$\begin{aligned} M &= AB^T + N, \mathrm{vec}(N) \sim \mathcal{N}\left(0, \sigma^2 I_{nm}\right), \\ A &= D\left(:, \mathcal{K}\right) = DS \;\text{ where } S \in \{0,1\}^{d\times r}, \end{aligned} \tag{1}$$

for a given noise variance σ^2. In this model, $\mathcal{K}$ is a set of indices of atoms in D corresponding to the columns of the factor matrix A. The matrix S is a sparse selection matrix which has only 1 non-zero entry per column.

The literature on computing S and B is extensive. A first group of methods are the continuous methods, based on applying iterative descent algorithms like ADMM [6] or fast gradient [7] to a relaxed version of (1). In most works, the product of S and B is estimated directly, using another variable $X = SB^T$ that has to be

J.E. Cohen and N. Gillis (corresponding author) are with the Department of Mathematics and Operational Research, Faculté Polytechnique, Université de Mons, Rue de Houdain 9, 7000 Mons, Belgium. E-mails: {jeremy.cohen,nicolas.gillis}@umons.ac.be. The authors acknowledge the support by the F.R.S.-FNRS (incentive grant for scientific research no F.4501.16). NG also acknowledges the support by the ERC (starting grant no 679515).

row sparse, with $M \approx DX$. This constraint on X can be relaxed using the ℓ_1 norm, the mixed norm ℓ_2/ℓ_1 or other convex norms, making the problem convex [7]–[10]. However, it has a very large number of parameters to be optimized (X is a d-by-n matrix) so that continuous methods may be slow and memory consuming. Note that continuous methods allow for adding various constraints like non-negativity.

A second category of sparse coding methods are greedy, or not based on a continuous optimization algorithm:

- A large family of geometric algorithms have been developed in the case where D is the data M itself, which is a common assumption in spectral unmixing referred to as "the pure-pixel assumption". This assumptions implies that the spectra in M form a simplex with r vertices. These geometric algorithms include vertex component analysis (VCA), successive projection algorithm (SPA), successive nonnegative projection algorithm (SNPA) and NFINDR, to cite a few, which aim at finding these vertices by resorting to geometric tools, such as projections [11]–[13]. These methods are usually very fast even for very large data set, but are restricted to the pure-pixel case and may not behave well when the data is grossly corrupted. Also the ℓ_2 reconstruction error may be high since no least square criterion is minimized directly.
- In the general case where the dictionary is not necessarily the data itself, one of the many efficient methods for solving the general sparse coding problem is the Simultaneous Orthogonal Matching Pursuit algorithm [14]. It has been referred to in the spectral unmixing community as self dictionary simultaneous orthogonal matching pursuit (SDSOMP) [15]. Recently, matching pursuit alternating least squares (MPALS) has been proposed that aims at minimizing a ℓ_2 cost function while using a fast greedy approach to find the right atoms in the dictionary [16].

Finally, a last family of methods specific to the spectral unmixing problem uses smart variations on the naive brute force algorithm that computes all possible combinations of atoms to be selected for each pixel. These methods include MESMA, MESLUM, AUTOMCU, or more recently, AMUSES [17]–[20]. A direct consequence of the pixel-wise approach is that the low-rank hypothesis may be violated. Also when the spectral library is large, such methods may be time-consuming. On the other hand, algorithms like MESMA may handle large scenes featuring more than a hundred endmembers. Because these methods may not produce results that satisfy our required assumptions, they will not be used for comparison in this short paper.

In the next section, we introduce another model closely related to (1) which allows multiple dictionaries to be used at once as a set of admissible spectral signatures.

III. Multiple-Dictionary Matrix Factorization

In the self-dictionary setting, a user may want some control over the regions of the HSI where the pure pixels are selected from. On the other hand, hand-picking pixels that seem pure may lead to poor results because the pure-pixel property is difficult to assess visually. Moreover, using a segmentation algorithm to compute homogeneous areas does not provide a good input for usual sparse coding method since the segmented regions would be bundled, and coherence lost. Similarly, when working with external libraries, it is reasonable to assume that only a few materials from a specific library should be selected. For example, exactly, at least or at most two spectra related to vegetation among some 20 available vegetation spectra. Lumping the libraries together and running any methods described above would not guarantee such an assignment.

To tackle these issues, we suggest to drop the single dictionary constraint. Using notations from (1), the third working hypothesis is modified to:

3. Columns of matrix A belong to matrices D_i, with d_i the exact number of atoms to be picked in each D_i, $1 \leqslant i \leqslant p$.

This new multi-dictionary model becomes:

$$\begin{aligned} M &= AB^T + N, \, \text{vec}(N) \sim \mathcal{N}\left(0, \sigma I_{nm}\right), \\ A &= \left[D_1\left(:, \mathcal{K}_1\right), \ldots, D_p\left(:, \mathcal{K}_p\right)\right] \Pi, \\ \#\mathcal{K}_i &= d_i \text{ (or } \leqslant d_i) \text{ and } \sum_{i=1}^{p} \#\mathcal{K}i = r, \end{aligned} \tag{2}$$

where $\mathcal{K}_i$ contains the indices of the d_i selected atoms in dictionary D_i for $1 \leqslant i \leqslant p$, and Π is the permutation matrix that matches factors to their corresponding atoms. Note that the model (2) can be adapted to handle different situations; in particular if we have a lower bound on the number of atoms to be selected from each dictionary (instead of the exact value or an upper bound). In that case, we introduce a new dictionary D_{p+1} containing all the other ones and choose $d_{p+1} = r - \sum_{i=1}^{p} d_i$ where d_i is the lower bound for the ith dictionary. Let us illustrate this on a simple example: assume we are given two dictionaries D_1 and D_2 from which we have to select three endmembers. We only know that at least one endmember has to be picked from each dictionary (that is, $d_1 \geqslant 1$ and $d_2 \geqslant 1$). In that case, we introduce a third dictionary $D_3 = [D_1, D_2]$ and impose $d_1 = d_2 = d_3 = 1$ in model (2).

To compute the sets $\mathcal{K}_i$'s and the abundance matrix B, we choose to adapt the MPALS algorithm introduced in [16]. The reason is two-fold: (1) MPALS was shown to perform extremely well, outperforming geometric and continuous approaches in many cases, and (2) it can be easily adapted to the multiple dictionary scenario by solving an assignment problem (see below). The proposed algorithm, named Multiple Matching Pursuit Alternating Least Squares (M2PALS), is an alternating algorithm that computes the second factor B using a least squares update, while using a least squares estimate of A to serve as a proxy for computing the sets and indices $\mathcal{K}_i$. It aims at solving

$$\underset{A=[D_1(:,\mathcal{K}_1),\ldots,D_p(:,\mathcal{K}_p)],\ B}{\operatorname{argmin}} \|M - AB^T\|_F,$$

by first estimating A and B without the dictionary constraint, then projecting A onto the atoms of the dictionaries. As we will see, the projection is a standard assignment problem.

If non-negativity is imposed on A and/or B, M2PALS can be adapted into M2PNALS by using a non-negative least squares solver to update A and B. Typically, these will not be solved up to global optimality for faster computation. In this paper, we will use a block coordinate descent method; see [21]. Other constraints on the abundance maps, such as homogeneity constraints using total variation regularization [22], could be easily incorporated in M2PALS by modifying the estimation of the abundance matrix B.

Algorithm 1 M2PALS

```
**Input:** Data matrix $M$, dictionaries $D_1, \dots, D_p$, exact (or upper bound on the) number of atoms to pick in each library $d_1, \dots, d_p$, initials factors $A$ and $B$.
**Output:** Factors $A$ and $B$, selected atoms $\mathcal{K}_1, \dots, \mathcal{K}_p$.
**while** the convergence criterion is not met **do**
    Least squares estimate of $A$: $A = \text{argmin}_X \, ||M - XB^T||_F$

    Solve the assignment problem: Find the $\mathcal{K}_i$'s to match $A$ the best using the available dictionaries, that is, solve (3), and update $A = [D_1(:, \mathcal{K}_1), \dots, D_p(:, \mathcal{K}_p)]\Pi$
    Least squares estimate of $B$: $B = \text{argmin}_Y \, ||M - AY^T||_F$
**end while**
```

a) The assignment problem: In M2PALS, after A has been estimated through a least squares update, one has to compute the index sets $\mathcal{K}_i$'s, that is, for each column of A, find the nearest column from one of the dictionaries D_i, while satisfying the constraints that exactly (or at most) d_i columns can be selected in the dictionary D_i. If only one dictionary is provided, each column of A can be processed independently, by picking the nearest column according to some distance criterion. However with multiple dictionaries and the constraint on the fixed number of atoms to be selected in each dictionary, the following assignment problem should be solved: Given $A \in \mathbb{R}^{m \times n}$, find the index sets $\mathcal{K}_i$ ($1 \leqslant i \leqslant p$) with $\#\mathcal{K}_i = d_i$ and $\sum_{i=1}^p d_i = r$, and find a permutation matrix $\Pi \in \{0,1\}^{r \times r}$ such that the matrix $\tilde{A} = [D_1(:, \mathcal{K}_1), \dots, D_p(:, \mathcal{K}_p)]\Pi$ minimizes

$$\sum_{j=1}^{n} \text{dist}\left(A(:,j), \tilde{A}(:,j)\right), \qquad (3)$$

where dist(.,.) is some distance criterion. Note that the permutation matrix Π is necessary since the columns of the input matrix A are not necessarily ordered in the same way as the dictionaries. The problem can be solved as follows. If we assume we know the dictionary from which we pick the atom to approximate each column of A, then the problem is trivial: for each column of A, we pick the column of the given dictionary which is the closest according to dist(.,.). It remains to identify which dictionary is used to approximate each column of A. Defining the distance between the jth column of A and the ith dictionary as

$$d(i,j) = \min_k \, \text{dist}\left(A(:,j), D_i(:,k)\right),$$

problem (3) reduces to an assignment problem where on the left hand side we have the r columns of A, on the right hand side we have d_i copies of dictionary D_i for a total of r nodes (recall $\sum_{i=1}^p d_i = r$), and the weights are given by $d(i,j)$. If d_i are upper bounds on the number of atoms to be selected, then the total number of nodes is larger than r but the assignment problem is solved in a similar fashion. Indeed, recall that an assignment problem is defined as follows: given r persons, l tasks ($l \leqslant r$) and the cost $c_{i,j}$ to assign the ith person to the jth task, assign a different person to each task in order to minimize the total cost. Mathematically, it can be formulated as follows

$$\min_{x_{i,j} \in \{0,1\}, 1 \leqslant i,j \leqslant r} \sum_{i,j} c_{i,j} x_{i,j} \quad \text{s.t.} \quad \sum_i x_{ij} \leqslant 1 \text{ and } \sum_j x_{ij} = 1,$$

where $x_{ij} = 1$ if and only if person i is assigned to task j. An optimal assignment can be obtained using linear programming, or the Hungarian algorithm[1] [23].

Several reasonable distances can be used in the context of spectral unmixing. The Euclidean distance is a simple choice but may be sensitive to scaling of spectra (that can occur for example because of inhomogeneous light intensity within the HSI). Many authors prefer to resort to an angle measurement through the SAM distance, or to remove the average of the spectra before computing an angle measurement, which yields the mean-removed spectral angle (MRSA) distance. These metrics are equally fast to compute, thus the choice is entirely application dependent. Note that various distance measurements may result in various assignment results; see, e.g., the discussion in [24]. In the simulation section, we will use the normalized inner product to asses the closeness between two vectors, that is, $\text{dist}(x,y) = 1 - \frac{x^T y}{||x||_2 ||y||_2}$, which is equivalent to maximize the sum of the normalized inner products between the selected atoms of the dictionaries and the columns of A. If the atoms of the dictionaries are normalized to have unit ℓ_2 norm, it can be shown that the normalized inner product, plugged in the assignment problem, provides the best solution for the index sets $\mathcal{K}_i$ for matching A in the least square sense.

IV. Application for semi-supervised and unsupervised pure-pixel selection

This section is divided in two parts. The first one is a comparison of M2PNALS (the non-negative version of M2PALS) on synthetic data with state-of-the-art spectral unmixing algorithms, namely SPA [25], SNPA [13], NFINDR [11], Fast Gradient for Nonnegative Sparse Regression (FGNSR) [7], Group Lasso Unit sum Positivity constraints [6], SDSOMP [15] and MPALS [16]. A recent segmentation algorithm, hierarchical rank-one NMF (H2NMF) [26], is also used as a preprocessing tool for M2PNALS to compute a set of dictionaries in an unsupervised manner.

The second one is an example of using M2PNALS to develop a user-friendly pure-pixel selection interface, based on either manual or automatic selection of image areas containing pure pixels.

A. Experiments on synthetic HSIs

For this experiment, we chose $r = 6$ spectra from the Urban HSI[2] (162 spectral bands, images 307 by 307 pixels), which were selected using SPA on the original image. Then the following setting is repeated $N = 100$ times: abundances are drawn from a uniform Dirichlet distribution so that they belong to a simplex with r vertices, and the r pure pixels are added at random positions in the data set. The data is generated by adding a white Gaussian noise with fixed signal to noise ratio (SNR), and then clipping negative entries to zero[3]. The size of the abundance matrix is set to $n = 200$, which is rather small but allows all methods to run in a reasonable computational time. All tests are preformed using Matlab on a laptop Intel CORE i5-3210M CPU @2.5GHz 6GB RAM.

At various SNR levels, for each setting and for each algorithm, the number of wrongly-selected pixels in the synthetic image is

[1] We use the code available at
http://www.mathworks.com/matlabcentral/fileexchange/
20328-munkres-assignment-algorithm.

[2] Available on http://www.erdc.usace.army.mil/

[3] Clipping to zero implies that the SNR is slightly over-estimated, as the noise is no longer Gaussian.

SNR	0	10	20	30	50	Time (s.)
SPA	93.66	79.16	51.83	13.33	0	0.0009
MPANLS	93.50	80.16	47.83	1.66	0	0.0955
SNPA	93.66	79.16	51.33	2.00	0	0.0420
FGNSR	97.00	86.16	58.16	17.66	0	3.7960
GLUP	96.16	82.00	31.83	6.66	0	1.1057
NFINDR	96.00	85.50	43.50	1.66	0	0.0192
SDSOMP	95.83	92.66	94.50	61.50	16.83	0.0154
M2PNALS-01	0	0	0	0	0	0.1068
M2PNALS-10	67.16	59.33	42.16	0.50	0	0.1161
M2PNALS-25	83.33	74.16	51.16	0.50	0	0.1240
M2PNALS-50	88.66	79.50	54.66	0.50	0	0.1332
H2NMF-M2PNALS	93.66	80.33	48.00	3.00	1.50	0.1057

TABLE I

Percentage of miss-selected atoms in synthetic HSIs depending on the SNR (first five columns), and average run times in seconds (the stopping criterion for MPNALS and M2PNALS is that the reconstruction error varies less than $10^{-3}\%$ in two successive iterations which always happened within 50 iterations in this simulation).

stored. Average results are presented in Table IV-A in percentage of miss-selected atoms, along with the average run times. For M2PNALS, additionally, the size of the dictionaries is used as a grid parameter. For instance, M2PNALS-10 means that each dictionary contains 1 pure-pixel plus 9 other randomly chosen spectra from the synthetic HSI (which may be one of the other pure pixels). The M2PNALS-01 is provided for sanity check and is trivially perfect for selecting pure pixels since each dictionary contains exactly one pure pixel.

Not too surprisingly, M2PNALS performs better than state-of-the-art methods in the noisy cases. This is natural since M2PNALS makes use of *a priori* information provided by the user on the location of the pure pixels. In the noiseless case, all algorithms perform perfectly except the SDSOMP algorithm, which is a sparse-coding method not adapted to identifying endmembers and rather selects the most representative spectra in the data at each iteration. Also, using H2NMF to segment the synthetic image (here without using any spatial information) and provide candidates D_i deteriorates the results at high SNR but improves on identification at medium SNR. This is natural since segmentation here can be understood as a regularization of the pure pixel selection process. In this simulated experiment, the data lies exactly in a simplex, thus linear unsupervised methods should perform better at high SNR. On the other hand, the segmentation process may include two pure pixels in the same segmented cluster, so that we set $d_i = 2$ for H2NMF-M2PNALS. See the next experiment for a more complete discussion on using segmentation before M2PNALS.

Moreover, note that MPNALS is an algorithm sensitive to initialization although it is able to identify good solutions in most cases regardless of the initialization; see the discussion and experiments in [16]. In this experiment, we chose SNPA for initialization, but better results may be achieved with a different initialization choice. The same comment applies to M2PNALS since it is adapted from MPNALS. Also, it can be checked that M2PNALS is fast compared to continuous algorithms. Actually, its complexity per iteration is linear in m, n and r (assuming we use a first-order method to solve the nonnegative least squares subproblems), which is the same as geometrical algorithms. Therefore, it can be used on large images, as illustrated in the next experiment.

B. Using M2PALS for a user selection of pure-pixel areas

Our motivation for developing M2PNALS is not to outperform state-of-the-art spectral unmixing methods in all scenarios. Rather, M2PNALS is meant to open new possibilities for dealing with spectral unmixing of HSI where a user inputs several spectral libraries.

The flexibility of M2PNALS lies in the variety of ways that the dictionaries can be defined. Of course they can be given by a user as external spectral libraries, but at least three other ways come to mind that only involve the available HSI. Although well-known dictionary learning methods do not apply here since only one dictionary would be learned this way, any segmentation or classification algorithm, which provides a labelled partition of the pixels, can be used to cut the original image into several regions that can be used directly as spectral libraries. Below, we introduce three different strategies for segmentation: a fully supervised, hand selection method, where the user chooses areas that she or he believes contain pure pixels; a semi-supervised state-of-the-art segmentation algorithm [28] based on logistic regression, that makes use of spatial correlations, and for which the training data has been labelled by hand; an unsupervised recent algorithm H2NMF [26] where no spatial information is used to perform segmentation.

These methods are illustrated in this experiment on the Urban HSI using $r = 6$ endmembers, along with the MPALS algorithm, see Figure 1. For the manual selection method, the user is asked to determine the number d_i of pure pixels in each area, whereas d_i are fixed to 1 for the unsupervised segmentations, and to 2 for the semi-supervised one since one of the cluster contained very few atoms. In this example, the centroids of the H2NMF clusters were used to initialize both MPNALS and M2PNALS. The relative reconstruction error of manual, semi-supervised and unsupervised M2PNALS here is respectively 4.28%, 4.12% and 4.20% , while that of MPNALS is 4.05% for a running time of respectively 50, $47s$, $51s$ and $20s$. In all cases, M2PNALS provided a sightly worse reconstruction error, because it is more constrained than MPNALS (which can choose any pixels in the HSI) while the selected areas do not contain the best set of pure pixels (as in the synthetic data experiment). However, segmentation used in conjunction with M2PNALS allowed the user to control in which areas pure pixels were selected, which is not possible with MPNALS nor any geometric pure pixel selection method.

Conclusion

In this paper, we introduced a new sparse coding model that allows for computing spectral unmixing using unknown spectra that belong to several classes of known spectra. This model may be used in post-processing, for instance, after selecting areas containing pure pixels, or after segmentation algorithms that provide homogeneous areas. It may also be used for estimating abundances given an HSI and a collection of libraries, as an alternative to methods such as MESMA. The M2PALS algorithm has been developed to identify the parameters of this multiple dictionary matrix factorization model, and its performance have been demonstrated on a simulated HSI, and on the Urban HSI.

Further research includes dealing with other types of noise, and the study of the identifiability of the multiple dictionary model (1), which should be stronger than that of the single dictionary model (2); see [16]. Also, the authors are working on adapting

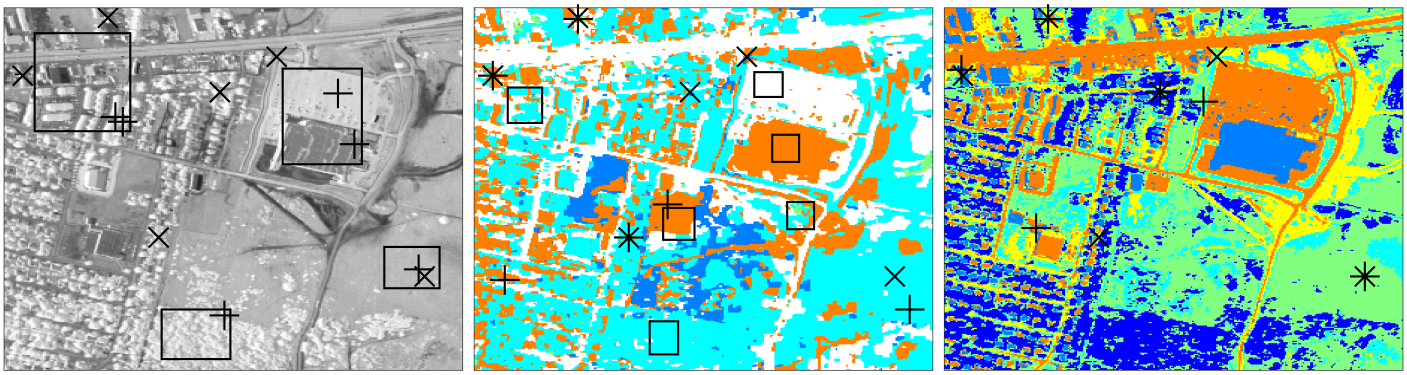

Fig. 1. An example of a manual (left), semi-supervised (middle) and unsupervised selection (right) of pure pixel areas in the Urban HSI. In the manual selection scenario, squares represent two areas containing two pure pixels and two areas containing one. The squares in the semi-supervised segmentation case represent the training data. The pixels identified by M2PNALS are represented with plusses '+' and lie within the selected areas, while pixels selected by fully-blind MPALS are represented with crosses '×'.

M2PALS for the fusion of multispectral and hyperspectral images under the pure-pixel assumption, since the mapping from an HSI to a MSI induces a pixel to area mapping.

Acknowledgments

The authors wish to thank David Kun for sharing the HyperSpectralToolbox, Antonio Plaza for sharing the MLR segmentation code and Rita Ammanouil for sharing the GLUP algorithm. The authors also thank the anonymous reviewers for their insightful comments which helped improve the paper.

References

[1] G. P. Asner, R. E. Martin, C. B. Anderson, and D. E. Knapp, "Quantifying forest canopy traits: Imaging spectroscopy versus field survey," *Remote Sensing of Environment*, vol. 158, pp. 15–27, 2015.

[2] J. Dozier and T. H. Painter, "Multispectral and hyperspectral remote sensing of alpine snow properties," *Annu. Rev. Earth Planet. Sci.*, vol. 32, pp. 465–494, 2004.

[3] J. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, 2012.

[4] A. Zare and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 95–104, 2014.

[5] 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, 2014.

[6] R. Ammanouil, A. Ferrari, C. Richard, and D. Mary, "Glup: Yet another algorithm for blind unmixing of hyperspectral data," *Proc. IEEE WHISPERS*, pp. 1–4, 2014.

[7] N. Gillis and R. Luce, "A Fast Gradient Method for Nonnegative Sparse Regression with Self Dictionary," *IEEE Trans. on Image Processing*, vol. 27, no. 1, pp. 24–37, 2018.

[8] E. Esser, M. Moller, S. Osher, G. Sapiro, and J. Xin, "A convex model for nonnegative matrix factorization and dimensionality reduction on physical space," *IEEE Trans. on Image Processing*, vol. 21, no. 7, pp. 3239–3252, 2012.

[9] E. Elhamifar, G. Sapiro, and R. Vidal, "See all by looking at a few: Sparse modeling for finding representative objects," in *Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012, pp. 1600–1607.

[10] V. Bittorf, B. Recht, E. Ré, and J. Tropp, "Factoring nonnegative matrices with linear programs," in *Advances in Neural Information Processing Systems (NIPS '12)*, 2012, pp. 1223–1231.

[11] M. E. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Imaging Spectrometry V*, M. R. Descour and S. S. Shen, Eds. SPIE-Intl Soc Optical Eng, 1999.

[12] J. M. Nascimento and J. M. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, 2005.

[13] N. Gillis, "Successive nonnegative projection algorithm for robust nonnegative blind source separation," *SIAM J. Imaging Sci.*, vol. 7, no. 2, pp. 1420–1450, 2014.

[14] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. part i: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.

[15] X. Fu, W.-K. Ma, T.-H. Chan, and J. M. Bioucas-Dias, "Self-dictionary sparse regression for hyperspectral unmixing: Greedy pursuit and pure pixel search are related," *IEEE J. Sel. Top. Signal Process.*, vol. 9, no. 6, pp. 1128–1141, sep 2015.

[16] J. Cohen and N. Gillis, "Dictionary-based tensor canonical polyadic decomposition," *arXiv preprint arXiv:1704.00541*, 2017.

[17] D. A. Roberts, M. Gardner, R. Church, S. Ustin, G. Scheer, and R. Green, "Mapping chaparral in the santa monica mountains using multiple endmember spectral mixture models," *Remote Sensing of Environment*, vol. 65, no. 3, pp. 267–279, 1998.

[18] J.-P. Combe, S. Le Mouelic, C. Sotin, A. Gendrin, J. Mustard, L. Le Deit, P. Launeau, J.-P. Bibring, B. Gondet, Y. Langevin *et al.*, "Analysis of omega/mars express data hyperspectral data using a multiple-endmember linear spectral unmixing model (melsum): Methodology and first results," *Planetary and Space Science*, vol. 56, no. 7, pp. 951–975, 2008.

[19] G. P. Asner, M. M. Bustamante, and A. R. Townsend, "Scale dependence of biophysical structure in deforested areas bordering the tapajos national forest, central amazon," *Remote Sensing of Environment*, vol. 87, no. 4, pp. 507–520, 2003.

[20] J. Degerickx, A. Okujeni, M.-D. Iordache, M. Hermy, S. van der Linden, and B. Somers, "A novel spectral library pruning technique for spectral unmixing of urban land cover," *Remote Sensing*, vol. 9, no. 6, p. 565, 2017.

[21] N. Gillis and F. Glineur, "Accelerated multiplicative updates and hierarchical als algorithms for nonnegative matrix factorization," *Neural computation*, vol. 24, no. 4, pp. 1085–1105, 2012.

[22] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4484–4502, 2012.

[23] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics (NRL)*, vol. 2, no. 1-2, pp. 83–97, 1955.

[24] N. Keshava, "Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 42, no. 7, pp. 1552–1565, 2004.

[25] U. Araújo, B. Saldanha, R. Galvão, T. Yoneyama, H. Chame, and V. Visani, "The successive projections algorithm for variable selection in spectroscopic multicomponent analysis," *Chemometr. Intell. Lab.*, vol. 57, no. 2, pp. 65–73, 2001.

[26] N. Gillis, D. Kuang, and H. Park, "Hierarchical clustering of hyperspectral images using rank-two nonnegative matrix factorization," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 2066–2078, 2015.

[27] Z. Jiang, Z. Lin, and L. S. Davis, "Learning a discriminative dictionary for sparse coding via label consistent k-svd," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 1697–1704.

[28] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 50, no. 3, pp. 809–823, 2012.