On-line blind unmixing for hyperspectral pushbroom imaging systems

Ludivine Nus, Sebastian Miron, David Brie

ON-LINE BLIND UNMIXING FOR HYPERSPECTRAL PUSHBROOM IMAGING SYSTEMS

Ludivine Nus, Sebastian Miron, David Brie

CRAN, Université de Lorraine, CNRS, Vandœuvre-lès-Nancy, France
ludivine.nus@univ-lorraine.fr, sebastian.miron@univ-lorraine.fr, david.brie@univ-lorraine.fr

ABSTRACT

In this paper, the on-line hyperspectral image blind unmixing is addressed. Inspired by the Incremental Non-negative Matrix Factorization (INMF) method [2], we propose an on-line NMF which is adapted to the acquisition scheme of a pushbroom imager. Because of the non-uniqueness of the NMF model, a minimum volume constraint on the endmembers is added allowing to reduce the set of admissible solutions. This results in a stable algorithm yielding results similar to those of standard off-line NMF methods, but drastically reducing the computation time. The algorithm is applied to wood hyperspectral images showing that such a technique is effective for the on-line prediction of wood piece rendering after finishing.

Index Terms— Hyperspectral imaging, Pushbroom imager, On-line Non-negative Matrix Factorization, Minimum volume constraint.

1. INTRODUCTION

This paper aims at developing real-time hyperspectral image blind unmixing methods, which are required in industrial applications for controlling and sorting input materials. One of the most employed technical solutions makes use of a hyperspectral pushbroom imager installed on the production line. The main feature of pushbroom imaging systems is that the hyperspectral data cube is acquired slice by slice, sequentially in time (which in fact is representing the so-called along track spatial dimension OY). Each slice is an image of dimensions $n_x \times n_\lambda$, where n_x denotes the across-track spatial dimension (one line of the scene) and n_λ the spectral dimension (wavelengths). The hyperspectral image is created by moving the sensor in the spatial direction OY (Fig.1) or by moving the objects directly on the conveyor belt.

1.1. On-line NMF

A relevant method dedicated to this type of applications is *on-line* Non-negative Matrix Factorization (NMF), which
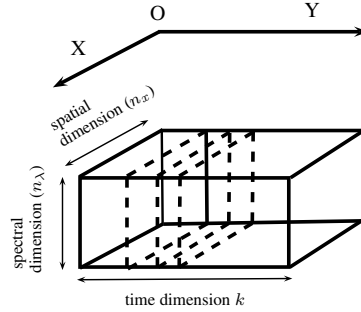
Fig. 1: The structure of the data acquired with a pushbroom imager

is an adaptive version of the classical NMF [6]. For a non-negative matrix $\mathbf{X} \in \mathbb{R}_+^{n_\lambda \times n_x}$, the NMF consists in finding two non-negative matrices $\mathbf{S} \in \mathbb{R}_+^{n_\lambda \times r}$ and $\mathbf{A} \in \mathbb{R}_+^{r \times n_x}$ with $r \leq \min(n_\lambda, n_x)$ such that [6]: $\mathbf{X} \approx \mathbf{SA}$. In hyperspectral imagery, the n_x columns of $\mathbf{X}$ represent the data samples recorded at n_λ wavelengths. $\mathbf{S}$ is a matrix containing the r normalized endmembers and $\mathbf{A}$ is a matrix containing on its columns the mixing coefficients (the abundances) for the recorded samples. The endmembers and the abundances are the parameters of interest to estimate. In general, the NMF method suffers from non-uniqueness of the solution [10]. To regularize the factorization, additional constraints must be used, such as sparsity [5], minimum volume constraint [8, 13], minimum distance endmembers constraint [16], *etc.*. To tackle large-scale and streaming data, several on-line NMF algorithms have been proposed aiming at updating both factor matrices $\mathbf{S}$ and $\mathbf{A}$ as the sample size of $\mathbf{X}$ increases. The work presented in [3] assumes that the endmember matrices between two successive samples are related by a linear transformation, which facilitates the parameters update. The on-line Incremental NMF (INMF) [2] considers that the endmember matrices evolve slowly between two consecutive acquisitions, which allows to alleviate the computational overhead. This is now the most widely used assumption adopted in on-line NMF algorithms. In [7], an on-line Itakura-Saito divergence-based NMF is proposed. An on-line NMF with volume constraint

on the abundances is proposed in [17] and an on-line NMF with sparse and smoothing constraints is discussed in [15]. Some on-line variants of the NMF, *e.g.* [4,14], use Stochastic Gradient Descent (SGD) as a strategy to reduce computational complexity. Instead of taking into account all available samples until the considered time instant, SGD-based methods randomly generate a small batch of samples from the subspace spanned by the previous samples, at each iteration.

1.2. Contributions

In this paper, we propose an on-line blind unmixing algorithm for a pushbroom imager, inspired from INMF [2] which presents several interesting features: it supports large datasets, it is easy to implement and computationally efficient. The proposed approach is specifically devised for the acquisition scheme of the pushbroom hyperspectral imaging system: a new data sample is a matrix (represented by a dotted line on Fig.1). This is different from most on-line NMF algorithms that process each new sample in a vector form. In order to reduce the set of admissible solutions, we integrate a minimum volume simplex (MVS) constraint, resulting in the proposed *on-line MVS-NMF*. Unlike [17] that imposes it on the abundances $\mathbf{A}$, in our application it makes sense to impose this constraint on the endmember matrix $\mathbf{S}$, since this will result in abundances having a maximum dissimilarity. Applying the MVS constraint on $\mathbf{S}$ is not trivial because $\mathbf{S}$ is generally a non-square matrix and therefore it does not have a determinant. In this paper, we propose a solution to tackle this problem and provide a mathematical interpretation of it. The remainder of the paper is organized as follows: section 2 is devoted to the derivation of the proposed algorithm. Experiments conducted on numerical simulations as well as on real data are reported in section 3. Conclusions are drawn in section 4.

2. AN ON-LINE NMF ALGORITHM FOR PUSHBROOM HYPERSPECTRAL DATA

2.1. Notations

The principle of the proposed on-line NMF is to alternatively update the endmember and abundance matrices when a new sample arrives at time instant $(k + 1)$. One way to handle the problem is to unfold the hyperspectral image as shown in Fig.2. $\tilde{\mathbf{X}}^{(1)} = \mathbf{X}^{(1)}$ is the first slice of the hyperspectral image and $\tilde{\mathbf{X}}^{(k)}$, the kth slice. The entire data set at time instant $(k + 1)$, *i.e.*, $\mathbf{X}^{(k+1)}$, can be represented as the concatenation of the first k samples with the new incoming sample *i.e.*, $\mathbf{X}^{(k+1)} = \left[\mathbf{X}^{(k)} \ \tilde{\mathbf{X}}^{(k+1)} \right]$. Similarly, we define $\mathbf{S}^{(k+1)} = \left[\mathbf{S}^{(k)} \ \tilde{\mathbf{S}}^{(k+1)} \right]$ and $\mathbf{A}^{(k+1)} = \left[\mathbf{A}^{(k)} \ \tilde{\mathbf{A}}^{(k+1)} \right]$.
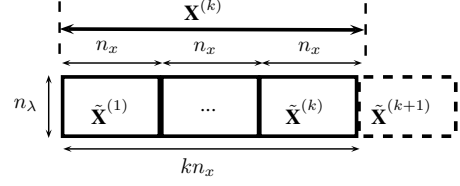

Fig. 2: Unfolded pushbroom hyperspectral image

2.2. Cost function

Let $\mathcal{J}^{(k)}$ denotes the cost function corresponding to the first k samples:

$$\mathcal{J}^{(k)}\left(\mathbf{S}^{(k)}, \mathbf{A}^{(k)}\right) = \sum_{\ell=1}^{k} \left\| \tilde{\mathbf{X}}^{(\ell)} - \tilde{\mathbf{S}}^{(\ell)} \tilde{\mathbf{A}}^{(\ell)} \right\|^2.$$

When the $(k+1)^{th}$ sample, $\tilde{\mathbf{X}}^{(k+1)}$ arrives, the corresponding cost function can be decomposed as follows:

$$\mathcal{J}^{(k+1)}\left(\mathbf{S}^{(k+1)}, \mathbf{A}^{(k+1)}\right) = \sum_{\ell=1}^{k} \left\| \tilde{\mathbf{X}}^{(\ell)} - \tilde{\mathbf{S}}^{(\ell)} \tilde{\mathbf{A}}^{(\ell)} \right\|_F^2 + \left\| \tilde{\mathbf{X}}^{(k+1)} - \tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)} \right\|_F^2. \quad (1)$$

We assume that the endmembers vary only slightly between consecutive samples *i.e.,* $\tilde{\mathbf{S}}^{(k+1)} \approx \tilde{\mathbf{S}}^{(\ell)} \ \forall \ell \leq k$. Thus, at time instant $(k + 1)$, only $\tilde{\mathbf{A}}^{(k+1)}$ needs to be adapted. Thus, the cost function (1) can be expressed as:

$$\mathcal{J}^{(k+1)}\left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)} | \mathbf{A}^{(k)}\right)$$
$$= \mathcal{J}^{(k)}\left(\tilde{\mathbf{S}}^{(k+1)} | \mathbf{A}^{(k)}\right) + \tilde{\mathcal{J}}^{(k+1)}\left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)}\right)$$

with

$$\mathcal{J}^{(k)}\left(\tilde{\mathbf{S}}^{(k+1)} | \mathbf{A}^{(k)}\right) = \sum_{\ell=1}^{k} \left\| \tilde{\mathbf{X}}^{(\ell)} - \tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(\ell)} \right\|_F^2$$

$$\tilde{\mathcal{J}}^{(k+1)}\left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)}\right) = \left\| \tilde{\mathbf{X}}^{(k+1)} - \tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)} \right\|_F^2.$$

In order to add some tracking capacity to the algorithm, the weighting coefficient α $(0 \leq \alpha \leq 1)$ is included to the cost function as formulated by equation (2):

$$\mathcal{J}^{(k+1)}\left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)}\right) = \alpha \mathcal{J}^{(k)} + (1 - \alpha)\tilde{\mathcal{J}}^{(k+1)}. \quad (2)$$

In general, the solution of (2) is not unique. To reduce the set of admissible solution, we regularize the problem by minimizing the volume of the simplex spanned by the endmembers $\tilde{\mathbf{S}}^{(k+1)}$. As $\tilde{\mathbf{S}}^{(k+1)}$ is not a square matrix, we propose to replace $\det\left(\tilde{\mathbf{S}}^{(k+1)}\right)$ (which is the classical simplex volume measure) by $\det\left(\tilde{\mathbf{S}}^{(k+1)T}\tilde{\mathbf{S}}^{(k+1)}\right)$. This is mathematically sound because $\det\left(\tilde{\mathbf{S}}^{(k+1)T}\tilde{\mathbf{S}}^{(k+1)}\right)$ can

be interpreted as the squared volume of $\tilde{\mathbf{S}}^{(k+1)}$ [11, Theorem 7]. Thus, cost function (2) is replaced by:

$$\mathcal{J}^{(k+1)}\left(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)}\right) = \alpha \mathcal{J}^{(k)} + (1-\alpha)\tilde{\mathcal{J}}^{(k+1)}$$
$$+ \mu \ln \det\left(\tilde{\mathbf{S}}^{(k+1)T}\tilde{\mathbf{S}}^{(k+1)}\right). \tag{3}$$

where $\mu > 0$ is the regularization parameter controlling the strength of the MVS constraint. The choice of the logarithm of the determinant rather than the determinant itself is mainly motivated by algorithmic reasons, since it drastically simplify the derivation of the multiplicative update rule. Also, imposing the MVS constraint on the endmembers will result in endmembers having a minimum dissimilarity. By duality [9], the $\tilde{\mathbf{A}}^{(k+1)}$ will correspond to a maximum volume simplex, *i.e.*, abundances having maximal dissimilarity (contrast).

2.3. On-line MVS-NMF

The estimation of parameter matrices $\tilde{\mathbf{S}}^{(k+1)}$ and $\tilde{\mathbf{A}}^{(k+1)}$ is acheived by minimizing the cost function (3) using a gradient descent technique, where the step size is chosen in order to obtain multiplicative update rules (similarly to the classical NMF [6]). The update rule for $\tilde{\mathbf{A}}_{rj}^{(k+1)}$ can be written as:

$$\tilde{\mathbf{A}}_{rj}^{(k+1)} \leftarrow \tilde{\mathbf{A}}_{rj}^{(k+1)} \frac{\left(\tilde{\mathbf{S}}^{(k+1)T}\tilde{\mathbf{X}}^{(k+1)}\right)_{rj}}{\left(\tilde{\mathbf{S}}^{(k+1)T}\tilde{\mathbf{S}}^{(k+1)}\tilde{\mathbf{A}}^{(k+1)}\right)_{rj}}.$$

The update of $\tilde{\mathbf{S}}^{(k+1)}$ is more complex. According to [12],

$$\frac{\partial \ln \det\left(\tilde{\mathbf{S}}^{(k+1)T}\tilde{\mathbf{S}}^{(k+1)}\right)}{\partial \tilde{\mathbf{S}}^{(k+1)}} = 2 \times \left(\tilde{\mathbf{S}}^{(k+1)\dagger}\right)^{T}$$

where $\dagger$ stands for the pseudo-inverse of a matrix. The scalar constant 2 above can be absorbed in the regularization parameter. However, the use of the pseudo-inverse of $\tilde{\mathbf{S}}^{(k+1)}$ no longer guarantees the non-negativity of $\tilde{\mathbf{S}}^{(k+1)}$ and precludes the derivation of a multiplicative update rule. In order to avoid this problem, we follow [17] which proposes to use the natural gradient [1] in $\tilde{\mathbf{S}}^{(k+1)}$ update. After some algebraic manipulations, the updating of $\tilde{\mathbf{S}}^{(k+1)}$ results in:

$$\tilde{\mathbf{S}}_{ir}^{(k+1)} \leftarrow \tilde{\mathbf{S}}_{ir}^{(k+1)} \frac{\left(\mathbf{N}^{(k+1)}\tilde{\mathbf{S}}^{(k+1)T}\tilde{\mathbf{S}}^{(k+1)}\right)_{ir}}{\left(\tilde{\mathbf{S}}^{(k+1)}\mathbf{M}^{(k+1)}\tilde{\mathbf{S}}^{(k+1)T}\tilde{\mathbf{S}}^{(k+1)}\right)_{ir} + \mu\tilde{\mathbf{S}}_{ir}^{(k+1)}}$$

where $\mathbf{N}^{(k+1)} = \alpha\mathbf{N}^{(k)} + (1-\alpha)\tilde{\mathbf{X}}^{(k+1)}\tilde{\mathbf{A}}^{(k+1)T}$ and $\mathbf{M}^{(k+1)} = \alpha\mathbf{M}^{(k)} + (1-\alpha)\tilde{\mathbf{A}}^{(k+1)}\tilde{\mathbf{A}}^{(k+1)T}$. The quantity $\tilde{\mathbf{S}}^{(k+1)T}\tilde{\mathbf{S}}^{(k+1)}$ refers to the natural gradient. The Algorithm 1 summarizes the proposed on-line MVS-NMF algorithm (for simplification, the indices $(k+1)$ are omitted). It includes two main loops: the outer loop produces estimates of $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{S}}$ at each time instant. These estimates are iteratively refined in the inner loop using a fixed number of iterations $Niter$.

Algorithm 1 On-line MVS-NMF

Input : $\tilde{\mathbf{X}}^{(1)}, r, \alpha, \mu, Niter =$ number of iterations
Initialization:
$\mathbf{N} = \text{zeros}(n_\lambda, r), \mathbf{M} = \text{zeros}(r, r),$
$\tilde{\mathbf{A}} = \text{rand}(r, n_x), \tilde{\mathbf{S}} = \text{rand}(n_\lambda, r)$
$\mathbf{A} = [\,]\ \mathbf{S} = [\,].$
Output : $\mathbf{A}$ and $\mathbf{S}$
while New sample available **do**
   $iter = 1$
   while $iter < Niter$ **do**
      $\tilde{\mathbf{A}}_{rj} \leftarrow \tilde{\mathbf{A}}_{rj} \dfrac{\left(\tilde{\mathbf{S}}^T\tilde{\mathbf{X}}\right)_{rj}}{\left(\tilde{\mathbf{S}}^T\tilde{\mathbf{S}}\tilde{\mathbf{A}}\right)_{rj}}$
      $\mathbf{N} \leftarrow \alpha\mathbf{N} + (1-\alpha)\tilde{\mathbf{X}}\tilde{\mathbf{A}}^T$
      $\mathbf{M} \leftarrow \alpha\mathbf{M} + (1-\alpha)\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T$
      $\tilde{\mathbf{S}}_{ir} \leftarrow \tilde{\mathbf{S}}_{ir} \dfrac{\left(\mathbf{N}\tilde{\mathbf{S}}^T\tilde{\mathbf{S}}\right)_{ir}}{\left(\tilde{\mathbf{S}}\mathbf{M}\tilde{\mathbf{S}}^T\tilde{\mathbf{S}}\right)_{ir}} + \mu\tilde{\mathbf{S}}_{ir}$
      $iter \leftarrow iter + 1$
   end while
   $\mathbf{A} \leftarrow [\mathbf{A}, \tilde{\mathbf{A}}]$
   $\mathbf{S} \leftarrow [\mathbf{S}, \tilde{\mathbf{S}}]$
end while

---

3. EXPERIMENTAL RESULTS

3.1. Synthetic data

In this section, a number of experiments are conducted on simulated hyperspectral data. The goal is to assess the influence of parameters $Niter$ and α on the convergence of the on-line MVS-NMF and to compare the results provided by the on-line MVS-NMF to those obtained by an off-line MVS-NMF using the same MVS penalty function. We simulated a hyperspectral image of size $119 \times 40 \times 190$, composed of three endmembers which do not vary over time. Here 119 corresponds to the number of wavelengths and 40×190 to the (spatial $\times$ time) dimensions. Each new time sample is a 119×40 slice of the hyperspectral image. Noise was added up to an $SNR = 20$ dB. To assess the convergence speed of MVC-NMF, the residual error *vs.* the sample number (from 1 to 190) was evaluated for three different values of the parameter $Niter$ (see Algorithm 1). The values of the different parameters were set to $r = 3$, $\alpha = 0.99$, $\mu = 10^{-5}$. The regularization parameter μ was fixed after successive trials. The results are shown on Fig.3.a. As $Niter$ increases, fewer samples are required to converge to the correct solution and if $Niter$ is large enough (about 800 in our example), the convergence is almost instantaneous. At the same time, the asymptotic error decreases as $Niter$ increases. Fig.3.b shows the convergence curves for $\alpha = 0.8,\ 0.9,\ 0.99$ while the other parameters are set to $r = 3$, $Niter = 10$, $\mu = 10^{-5}$. As α decreases, the asymptotic error decreases while the convergence speed does not seem to be much affected. Actually, the two parameters are strongly coupled. Fig.3.c shows the

convergence curves for different couples $(Niter, \alpha)$, chosen such that α^{Niter} stays constant. For all couples, the asymptotic error is the same while the convergence speed increases as $Niter$ increases. However, the computation time increases linearly with $Niter$. In a last experiment, we compared the behavior of the on-line MVS-NMF to that of the off-line MVS-NMF *vs.* the SNR. To evaluate the unmixing performance of these methods, we calculate the root mean square error (RMSE) of the endmember matrix **S**. The off-line MVS-NMF was applied to the unfolded version of the image, of size 119×7600, and the parameters were set to $r = 3$, $\mu = 0.05$ and $Niter = 6000$. They were chosen to yield similar results as those obtained by the on-line version at $SNR = 50$ dB. The parameters of the on-line algorithm were set to $r = 3$, $\alpha = 0.99$, $\mu = 10^{-5}$ and $Niter = 800$. The results are shown on Fig.3.d. Both methods give very similar results. Meanwhile, the processing time (for the considered scenario) is almost double for the off-line NMF as compared to the on-line version. This difference becomes much more important as the size of the data increases.
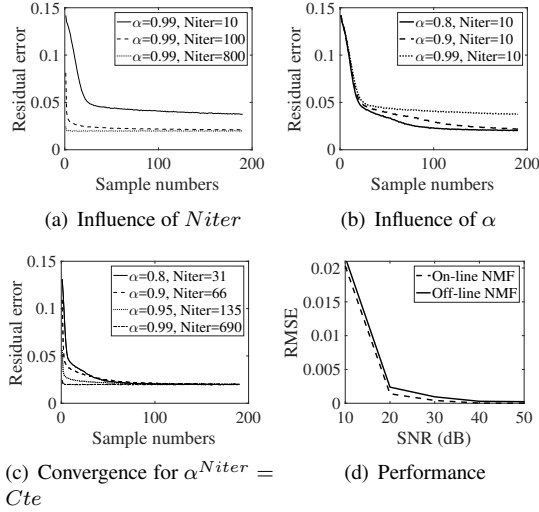

(a) Influence of $Niter$

(b) Influence of α

(c) Convergence for $\alpha^{Niter} = Cte$

(d) Performance

Fig. 3: Influence of the parameters and performances

3.2. Real data

This section presents the results obtained by the on-line MVS-NMF on real hyperspectral image provided by a Near Infra Red pushbroom imager working in the spectral range 900-2500 nm. The imaged object is a piece of wood and the hyperspectral image size is $258 \times 384 \times 1200$, where 258 represents the number of wavelengths and 384×1200 the (spatial $\times$ time) dimensions. The chosen piece of wood has a singularity of sapwood type, in the bottom right corner. This defect is hardly discernible to the naked eye. Fig.4 represents the abundances maps, the variation interval (green area) and the average of endmembers signatures estimated by the on-line MVS-NMF. The parameters were
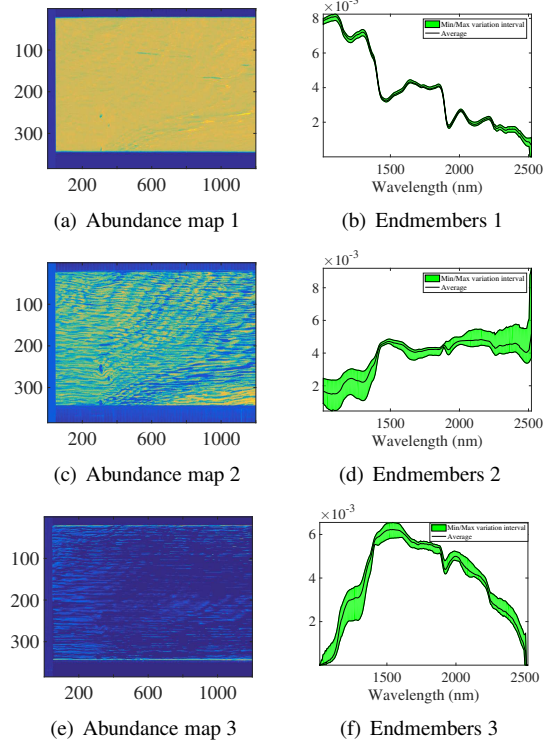

(a) Abundance map 1

(b) Endmembers 1

(c) Abundance map 2

(d) Endmembers 2

(e) Abundance map 3

(f) Endmembers 3

Fig. 4: Results on a real hyperspectral image

set to $r = 3$, $\alpha = 0.99$, $\mu = 0.001$ and $Niter = 100$. We can see that the variation interval of the estimated endmembers signatures is small; we can clearly distinguish in the bottom right corner of the Fig.4.c, an area more contrasted than the rest of the image which represents the sapwood singularity. This defect is almost impossible to detect on a visible range color image. On a Macbook Pro with 4-core processor running at 2.7 GHz and having 16GB of RAM, our algorithm took 60 seconds to process the entire image. The processing time can be considerably reduced, for example by avoiding random initializations or by modifying the α-$Niter$ pair. The off-line NMF, meanwhile, took almost half an hour to process the same image, with similar results.

4. CONCLUSION AND FUTURE WORK

We proposed an on-line NMF algorithm specifically devised for the processing of pushbroom imager data of pieces of wood. We showed that by adding a minimum volume constraint on the endmembers, the stability of the solution is guaranteed. We also showed that the results generated by our method are similar to off-line NMF, for a much smaller computation time. The tests on real data confirmed that our method is well-suited for wood defect detection on industrial production lines. Future work will study the influence of the regularization parameter μ on the estimation performance, and develop an efficient method to estimate it.

5. REFERENCES

[1] S.-I. Amari. Natural gradient works efficiently in learning. *Neural computation*, 10(2):251–276, 1998.

[2] S. S. Bucak and B. Gunsel. Incremental subspace learning via Non-negative Matrix Factorization. *Pattern recognition*, 42(5):788–797, 2009.

[3] B. Cao, D. Shen, J.-T. Sun, X. Wang, Q. Yang, and Z. Chen. Detect and track latent factors with On-line Non-negative Matrix Factorization. In *International Joint Conferences on Artificial Intelligence*, volume 7, pages 2689–2694, 2007.

[4] N. Guan, D. Tao, Z. Luo, and B. Yuan. On-line Non-negative Matrix Factorization with robust stochastic approximation. *IEEE Transactions on Neural Networks and Learning Systems*, 23(7):1087–1099, 2012.

[5] P. O. Hoyer. Non-negative Matrix Factorization with sparseness constraints. *Journal of machine learning research*, 5(Nov):1457–1469, 2004.

[6] D. D. Lee and H. S. Seung. Algorithms for Non-negative Matrix Factorization. In *Advances in neural information processing systems*, pages 556–562, 2001.

[7] A. Lefèvre, F. Bach, and C. Févotte. On-line algorithms for Non-negative Matrix Factorization with the Itakura-Saito divergence. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 313–316, 2011.

[8] L. Miao and H. Qi. Endmember extraction from highly mixed data using minimum volume constrained Non-negative Matrix Factorization. *IEEE Transactions on Geoscience and Remote Sensing*, 45(3):765–777, 2007.

[9] S. Miron, M. Dossot, C. Carteret, S. Margueron, and D. Brie. Joint processing of the parallel and crossed polarized raman spectra and uniqueness in blind Non-negative source separation. *Chemometrics and Intelligent Laboratory Systems*, 105(1):7–18, 2011.

[10] S. Moussaoui, D. Brie, and J. Idier. Non-negative source separation: range of admissible solutions and conditions for the uniqueness of the solution. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, volume 5, pages v–289. IEEE, 2005.

[11] B. Peng. The determinant: A means to calculate volume. *Recall*, 21:a22, 2007.

[12] K. B. Petersen, M. S. Pedersen, et al. The matrix cookbook. *Technical University of Denmark*, 7:15, 2008.

[13] R. Schachtner, G. Pöppel, and E. W. Lang. Towards unique solutions of Non-negative Matrix Factorization problems by a determinant criterion. *Digital Signal Processing*, 21(4):528–534, 2011.

[14] F. Wang, P. Li, and A. C. König. Efficient document clustering via On-line Non-negative Matrix Factorizations. In *Society for Industrial and Applied Mathematics*, volume 11, pages 908–919, 2011.

[15] Y. Wu, B. Shen, and H. Ling. Visual tracking via On-line Non-negative Matrix Factorization. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(3):374–383, 2014.

[16] Y. Yu and W. Sun. Minimum distance constrained Non-negative Matrix Factorization for the endmember extraction of hyperspectral images. In *MIPPR 2007: Remote Sensing and GIS Data Processing and Applications; and Innovative Multispectral Technology and Applications*, volume 6790, page 679015. International Society for Optics and Photonics, 2007.

[17] G. Zhou, Z. Yang, S. Xie, and J.-M. Yang. On-line blind source separation using incremental Non-negative Matrix Factorization with volume constraint. *IEEE transactions on neural networks*, 22(4):550–560, 2011.