

# JOINT DIMENSIONALITY REDUCTION, CLASSIFICATION AND SEGMENTATION OF HYPERSPECTRAL IMAGES

*Ali MOHAMMAD-DJAFARI, Nadia BALI and Adel MOHAMMADPOOR*

Laboratoire des Signaux et Systèmes,  
Unité mixte de recherche 8506 (CNRS-Supélec-UPS)  
Supélec, Plateau de Moulon, 3 rue Joliot Curie, 91192 Gif-sur-Yvette, France.

## ABSTRACT

Dimensionality reduction, spectral classification and segmentation are the three main problems in hyperspectral image analysis. In this paper we propose a Bayesian estimation approach which gives a solution for these three problems jointly. The data reduction problem is modeled as a blind sources separation (BSS) where the sources are the images which must be mutually independent and piecewise homogeneous. To insure these properties, we propose a hierarchical model for the sources with a common hidden classification variable which is modelled as a Potts-Markov field. The joint Bayesian estimation of this hidden variable as well as the sources and the mixing matrix of the BSS problem gives a solution for all the three problems of dimensionality reduction, spectra classification and segmentation of hyperspectral images. For the Bayesian computation, we propose to use either Gibbs Sampling (GS) or Mean Field Approximation (MFA) methods. A few simulation results illustrate the performances of the proposed method and some comparison with other classical methods of PCA and ICA used for BSS.

## 1. INTRODUCTION

Hyper-spectral images data can be represented either as a set of images  $x_\omega(\mathbf{r})$  or as a set of spectra  $x_{\mathbf{r}}(\omega)$  where  $\omega \in \Omega$  indexes the wavelength and  $\mathbf{r} \in \mathcal{R}$  a pixel position [?, ?, ?]. In both representations, the data are dependent in both spatial positions and in spectral bands. Classical methods of hyperspectral image analysis try either to classify the spectra  $x_\omega(\mathbf{r})$  in  $K$  classes  $\{s_k(\omega), k = 1, \dots, K\}$  or to classify the images  $x_\omega(\mathbf{r})$  in  $N$  classes  $\{s_j(\mathbf{r}), j = 1, \dots, N\}$ , using the classical classification methods such as distance based methods (like  $K$ -means) or probabilistic methods using the mixture of Gaussian (MoG) modeling of the data. These methods thus either neglect the spatial structure of the spectra or the spectral natures of the pixels along the wavelength bands.

When the data are considered as a set of spectra, then the

dimensionality reduction problem can be written as:

$$x_{\mathbf{r}}(\omega) = \sum_{k=1}^K A_k(\mathbf{r}) s_k(\omega) + \epsilon_{\mathbf{r}}(\omega), \quad (1)$$

where the  $s_k(\omega)$  are the  $K$  spectral source components and each column of the mixing matrix  $\mathbf{A}$  is in fact an image  $A_k(\mathbf{r})$ . The ideal case here would be to obtain an estimate for  $\mathbf{A}$  such that each column  $A_k(\mathbf{r})$  represents an image where only non-zero values for the pixels in the regions which are associated to the spectrum  $s_k(\omega)$ .

When the data are considered as a set of spectra, then the dimensionality reduction problem can be written as:

$$x_\omega(\mathbf{r}) = \sum_{j=1}^N A_j(\omega) s_j(\mathbf{r}) + \epsilon_\omega(\mathbf{r}), \quad (2)$$

where the sources  $s_j(\mathbf{r})$  are the  $N$  source images and each column of the mixing matrix  $\mathbf{A}$  in this case correspond to the spectrum  $A_j(\omega)$ . The ideal case here would be to obtain an estimate for the sources such that the pixels of each image  $s_j(\mathbf{r})$  be non-zero only for the positions in the regions which are associated to the spectrum  $A_j(\omega)$ .

As we mentioned, the classical methods in hyperspectral imaging consider these two problems independently. But we see that the two representations are really linked, because the columns  $A_k(\mathbf{r})$  of the mixing matrix in the first model is related to the sources  $s_j(\mathbf{r})$  in the second model and the columns  $A_j(\omega)$  of the mixing matrix in the second model is related to the sources  $s_k(\omega)$  in the first model. This becomes still more explicit if we choose  $K = N$ .

In this paper, we propose to use this specificity of the hyperspectral images and consider the dimensionality reduction problem as the blind sources separation (BSS) of equation ?? and use a Bayesian estimation framework with a hierarchical model for the sources with a common hidden classification variable which is modelled as a Potts-Markov field. The joint estimation of this hidden variable, the sources and the mixing matrix of the BSS problem gives a solution for all the three problems of dimensionality reduction, spectra classification and segmentation of hyperspectral images.

## 2. PROPOSED DIMENSIONALITY REDUCTION MODEL AND METHOD

### 2.1. Dimensionality reduction model

We propose to consider the equation (??) written in the following vector form:

$$\mathbf{x}(\mathbf{r}) = \mathbf{A}\mathbf{s}(\mathbf{r}) + \boldsymbol{\epsilon}(\mathbf{r}) \quad (3)$$

as the dimensionality reduction where:  $\mathbf{x}(\mathbf{r}) = \{x_i(\mathbf{r}), i = 1, \dots, m\}$  is the set of  $m$  observed images in different bands  $\omega_i$ ,  $\mathbf{A}$  is the unknown mixing matrix of dimensions  $(m, n)$ ,  $\mathbf{s}(\mathbf{r}) = \{s_j(\mathbf{r}), j = 1, \dots, n\}$  is the set of  $n$  unknown components (source images) and  $\boldsymbol{\epsilon}(\mathbf{r}) = \{\epsilon_i(\mathbf{r}), i = 1, \dots, m\}$  represents the errors.

This equation can also be written as:

$$\underline{\mathbf{x}} = \mathbf{A}\underline{\mathbf{s}} + \boldsymbol{\epsilon} \quad (4)$$

where we used  $\underline{\mathbf{x}} = \{\mathbf{x}(\mathbf{r}), \mathbf{r} \in \mathcal{R}\}$ ,  $\underline{\mathbf{s}} = \{\mathbf{s}(\mathbf{r}), \mathbf{r} \in \mathcal{R}\}$  and  $\boldsymbol{\epsilon} = \{\boldsymbol{\epsilon}(\mathbf{r}), \mathbf{r} \in \mathcal{R}\}$ . In this paper, we assume that the errors  $\boldsymbol{\epsilon}(\mathbf{r})$  are centered, white, Gaussian with covariance matrix  $\boldsymbol{\Sigma}_\epsilon = \text{diag}[\sigma_{\epsilon_1}^2, \dots, \sigma_{\epsilon_m}^2]$ . This leads to

$$p(\underline{\mathbf{x}}|\underline{\mathbf{s}}, \mathbf{A}, \boldsymbol{\Sigma}_\epsilon) = \prod_{\mathbf{r}} \mathcal{N}(\mathbf{A}\mathbf{s}(\mathbf{r}), \boldsymbol{\Sigma}_\epsilon) \quad (5)$$

### 2.2. Sources model

As we mentioned in the introduction, we want to impose to all these sources  $\mathbf{s}(\mathbf{r})$  to be piecewise homogeneous and share the same segmentation, where the pixels in each region are considered to be homogeneous and associated to a particular spectrum representing the type of the material in that region. We also want that those spectra be classified in  $K$  distinct classes, thus all the pixels in regions associated with a particular spectrum share some common statistical parameters. This can be achieved through the introduction of a discrete valued hidden variable  $z(\mathbf{r})$  representing the labels associated to each type of material and thus assuming the following:

$$p(s_j(\mathbf{r})|z(\mathbf{r}) = k) = \mathcal{N}(m_{jk}, \sigma_{jk}^2), \quad k = 1, \dots, K \quad (6)$$

with the following Potts-Markov field model:

$$p(z(\mathbf{r}), \mathbf{r} \in \mathcal{R}) \propto \exp \left[ \beta \sum_{\mathbf{r} \in \mathcal{R}} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \delta(z(\mathbf{r}) - z(\mathbf{r}')) \right]. \quad (7)$$

$z(\mathbf{r})$  then will represents the common segmentation of the sources and the data. The parameter  $\beta$  controls the mean size of those regions.

We may note that, assuming *a priori* that the sources are mutually independent and that pixels in each class  $k$  are inde-

pendent from those of class  $k'$ , we have

$$\begin{aligned} p(\underline{\mathbf{s}}|\mathbf{z}) &= \sum_k \sum_{\mathbf{r} \in \mathcal{R}_k} \sum_j p(s_j(\mathbf{r})|z(\mathbf{r}) = k) \\ &= \sum_k \sum_{\mathbf{r} \in \mathcal{R}_k} \sum_j \mathcal{N}(m_{jk}, \sigma_{jk}^2) \end{aligned}$$

where  $\mathcal{R}_k = \{\mathbf{r} : z(\mathbf{r}) = k\}$  and  $\mathcal{R} = \cup_k \mathcal{R}_k$ .

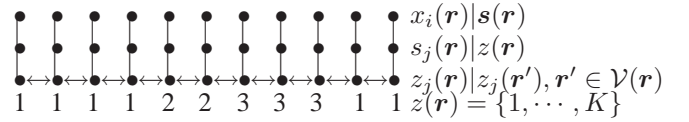
To insure that each image  $s_j(\mathbf{r})$  is only non-zero in those regions associated with the  $k$ th spectrum, we impose  $K = n$  and  $m_{jk} = 0, \forall j \neq k$  and  $\sigma_{jk}^2 = 0, \forall j \neq k$ . We may then write

$$\begin{aligned} p(\underline{\mathbf{s}}|\mathbf{z}) &= \sum_{\mathbf{r} \in \mathcal{R}} p(\mathbf{s}(\mathbf{r})|z(\mathbf{r}) = k) \\ &= \sum_{\mathbf{r} \in \mathcal{R}} \mathcal{N}(\mathbf{m}_k(\mathbf{r}), \boldsymbol{\Sigma}_k(\mathbf{r})) \quad (8) \end{aligned}$$

where  $\mathbf{m}_k(\mathbf{r})$  is a vector of size  $n$  with all elements equal to zero except the  $k$ -th element  $k = z(\mathbf{r})$  and  $\boldsymbol{\Sigma}_k(\mathbf{r})$  is a diagonal matrix of size  $n \times n$  with all elements equal to zero except the  $k$ -th main diagonal element where  $k = z(\mathbf{r})$ .

### 2.3. Data and sources hierarchical model

Combining the observed data model (??) and the sources model (??) of the previous section, we obtain the following hierarchical model:



**Fig. 1.** Proposed hierarchical model for hyperspectral images: the sources  $s_j(\mathbf{r})$  are hidden variables for the data  $x_i(\mathbf{r})$  and the common classification and segmentation variables  $z(\mathbf{r})$  is a hidden variable for the sources.

### 2.4. Bayesian estimation framework

Using the prior data model (??), the prior source model (??) and the prior Potts-Markov model (??) and also assigning appropriate prior probability laws  $p(\mathbf{A})$  and  $p(\boldsymbol{\theta})$  to the hyperparameters  $\boldsymbol{\theta} = \{\boldsymbol{\theta}_\epsilon, \boldsymbol{\theta}_s\}$  where  $\boldsymbol{\theta}_\epsilon = \mathbf{R}_\epsilon$  and  $\boldsymbol{\theta}_s = \{(m_{jk}, \sigma_{jk}^2)\}$ , we obtain an expression for the posterior law

$$p(\underline{\mathbf{s}}, \mathbf{z}, \mathbf{A}, \boldsymbol{\theta}|\underline{\mathbf{x}}) \propto p(\underline{\mathbf{x}}|\underline{\mathbf{s}}, \mathbf{A}, \boldsymbol{\theta}_\epsilon) p(\underline{\mathbf{s}}|\mathbf{z}, \boldsymbol{\theta}_s) p(\mathbf{z}) p(\mathbf{A}) p(\boldsymbol{\theta}) \quad (9)$$

In this paper, we used conjugate priors for all of them, i.e., Gaussian for the elements of  $\mathbf{A}$ , Gaussian for the means  $m_{jk}$  and inverse Gamma for the variances  $\sigma_{jk}^2$  as well as for the noise variances  $\sigma_{\epsilon_i}$ .

When given the expression of the posterior law, we can then use it to define an estimator such as Joint Maximum A Posteriori (JMAP) or the Posterior Means (PM) for all the unknowns. The first needs optimization algorithms and the second integration methods. Both are computationally demanding. Alternate optimization is generally used for the first while the MCMC techniques are used for the second.

In this work, we propose to separate the unknowns in two sets  $(\underline{s}, z)$  and  $(\mathbf{A}, \underline{\theta})$  and then use the following iterative algorithm:

- Estimate  $(\underline{s}, z)$  using  $p(\underline{s}, z | \hat{\mathbf{A}}, \hat{\underline{\theta}}, \underline{x})$  by  $\hat{\underline{s}} \sim p(\underline{s} | \hat{z}, \hat{\mathbf{A}}, \hat{\underline{\theta}}, \underline{x})$  and  $\hat{z} \sim p(z | \hat{\mathbf{A}}, \hat{\underline{\theta}}, \underline{x})$
- Estimate  $(\mathbf{A}, \underline{\theta})$  using  $p(\mathbf{A}, \underline{\theta} | \hat{\underline{s}}, \hat{z}, \underline{x})$  by  $\hat{\mathbf{A}} \sim p(\mathbf{A} | \hat{\underline{s}}, \hat{z}, \underline{\theta}, \underline{x})$  and  $\hat{\underline{\theta}} \sim p(\underline{\theta} | \hat{\underline{s}}, \hat{z}, \hat{\mathbf{A}}, \underline{x})$

In this algorithm,  $\sim$  represents either *argmax* or *generate sample using* or still *compute the Mean Field Approximation (MFA)*.

To implement this algorithm, we need the following expressions:

- $p(\underline{s} | z, \mathbf{A}, \underline{\theta}, \underline{x}) \propto p(\underline{x} | \underline{s}, \mathbf{A}, \Sigma_\epsilon) p(\underline{s} | z, \underline{\theta})$ .

It is then easy to see that  $p(\underline{s} | z, \mathbf{A}, \underline{\theta}, \underline{x})$  is separable in  $\mathbf{r}$ :

$$\begin{aligned} p(\underline{s} | z, \underline{\theta}, \underline{x}) &= \prod_{\mathbf{r}} p(s(\mathbf{r}) | z(\mathbf{r}), \underline{\theta}, \mathbf{x}(\mathbf{r})) \\ &= \prod_{\mathbf{r}} \mathcal{N}(\bar{s}(\mathbf{r}), \mathbf{B}(\mathbf{r})) \end{aligned}$$

with

$$\begin{cases} \mathbf{B}(\mathbf{r}) = [\mathbf{A}^t \Sigma_\epsilon^{-1} \mathbf{A} + \Sigma_{z(\mathbf{r})}^{-1}]^{-1} \\ \bar{s}(\mathbf{r}) = \mathbf{B}(\mathbf{r}) [\mathbf{A}^t \Sigma_\epsilon^{-1} \mathbf{x}(\mathbf{r}) + \Sigma_{z(\mathbf{r})}^{-1} \mathbf{m}_{z(\mathbf{r})}] \end{cases} \quad (10)$$

In this relation  $\mathbf{m}_{z(\mathbf{r})}$  is a vector of size  $n$  with all elements equal to zero except the  $k$ -th element where  $k = z(\mathbf{r})$  and  $\Sigma_{z(\mathbf{r})}$  is a diagonal matrix of size  $n \times n$  with all elements equal to zero except the  $k$ -th diagonal where  $k = z(\mathbf{r})$ .

- $p(z | \mathbf{A}, \underline{\theta}, \underline{x}) \propto p(\underline{x} | z, \mathbf{A}, \underline{\theta}) p(z)$ , where

$$\begin{aligned} p(\underline{x} | z, \mathbf{A}, \underline{\theta}) &= \prod_{\mathbf{r}} p(\mathbf{x}(\mathbf{r}) | z(\mathbf{r}), \mathbf{A}, \underline{\theta}) \\ &= \prod_{\mathbf{r}} \mathcal{N}(\mathbf{A} \mathbf{m}_{z(\mathbf{r})}, \mathbf{A} \Sigma_{z(\mathbf{r})} \mathbf{A}^t + \Sigma_\epsilon). \end{aligned}$$

It is then easy to see that, even if  $p(\underline{x} | z, \mathbf{A}, \underline{\theta})$  is separable in  $\mathbf{r}$ ,  $p(z | \mathbf{A}, \underline{\theta}, \underline{x})$  is not and it has the same markovian structure that  $p(z)$ .

- $p(\mathbf{A} | z, \underline{\theta}, \underline{x}) \propto p(\underline{x} | z, \mathbf{A}, \underline{\theta}) p(\mathbf{A})$

It is easy to see that, with a Gaussian or uniform prior for  $p(\mathbf{A})$  we obtain a Gaussian expression for this posterior law. Indeed, with an uniform prior, the posterior mean is equivalent to the posterior mode and equivalent to the Maximum

Likelihood (ML) estimate  $\hat{\mathbf{A}} = \arg \max_{\mathbf{A}} \{p(\underline{x} | z, \mathbf{A}, \underline{\theta})\}$  whose expression is:

$$\hat{\mathbf{A}} = \left[ \sum_{\mathbf{r}} \mathbf{x}(\mathbf{r}) \bar{s}'(\mathbf{r}) \right] \left[ \sum_{\mathbf{r}} \bar{s}(\mathbf{r}) \bar{s}'(\mathbf{r}) + \mathbf{B}(\mathbf{r}) \right]^{-1} \quad (11)$$

where  $\bar{s}(\mathbf{r})$  and  $\mathbf{B}(\mathbf{r})$  are given by (??).

- $p(\mathbf{R}_\epsilon | z, \mathbf{A}, \underline{\theta}, \underline{x}) \propto p(\underline{x} | z, \mathbf{A}, \underline{\theta}) p(\mathbf{R}_\epsilon)$

It is also easy to show that, with an uniform prior on the logarithmic scale or an inverse gamma prior for the noise variances, the posterior is also an inverse gamma.

- $p(\underline{\theta} | z, \mathbf{A}, \underline{x}) \propto p(\underline{x} | z, \mathbf{A}, \underline{\theta}) p(\underline{\theta})$ :

Again here, using the conjugate priors for the means  $m_{j_k}$  and inverse gamma for the variances  $\sigma_{j_k}^2$  we can obtain easily the expressions of the posterior laws for them.

Details of the expressions of  $p(\mathbf{A} | z, \underline{\theta}, \underline{x})$ ,  $p(\mathbf{R}_\epsilon | z, \mathbf{A}, \underline{\theta}, \underline{x})$  and  $p(\underline{\theta} | z, \mathbf{A}, \underline{x})$  as well as their modes and means can be found in [?].

### 3. COMPUTATIONAL CONSIDERATIONS AND MEAN FIELD APPROXIMATION

As we can see, the expression of the conditional posterior of the sources is separable in  $\mathbf{r}$  but this is not the case for the conditional posterior of the hidden variable  $z(\mathbf{r})$ . So, even if it is possible to generate samples from this posterior using a Gibbs sampling scheme, the cost of the computation is very high for real applications. The Mean Field Approximation (MFA) then becomes a natural tool for obtaining approximate solutions with lower computational cost.

The mean field approximation is a general method for approximating the expectation of a Markov random variable. The idea consists in, when considering a pixel, to neglect the fluctuation of its neighbor pixels by fixing them to their mean values. Another interpretation of the MFA is to approximate a non separable

$$\begin{aligned} p(z) &\propto \exp \left[ \beta \sum_{\mathbf{r}} \sum_{\mathbf{r}'} \delta(z(\mathbf{r}) - z(\mathbf{r}')) \right] \\ &\propto \prod_{\mathbf{r}} p(z(\mathbf{r}) | z(\mathbf{r}'), \mathbf{r}' \in \mathcal{V}(\mathbf{r})) \end{aligned}$$

with the following separable one:

$$q(z) \propto \prod_{\mathbf{r}} q(z(\mathbf{r}) | \bar{z}(\mathbf{r}'), \mathbf{r}' \in \mathcal{V}(\mathbf{r}))$$

where  $\bar{z}(\mathbf{r}')$  is the expected value of  $z(\mathbf{r}')$  computed using  $q(zb)$ . This approximate separable expression is obtained in such a way to minimize  $KL(p, q)$  for a given class of separable distributions  $q \in \mathcal{Q}$ .

Using now this approximation in the expression of the conditional posterior law  $p(z|\mathbf{A}, \underline{\theta}, \underline{x})$  gives the separable MFA

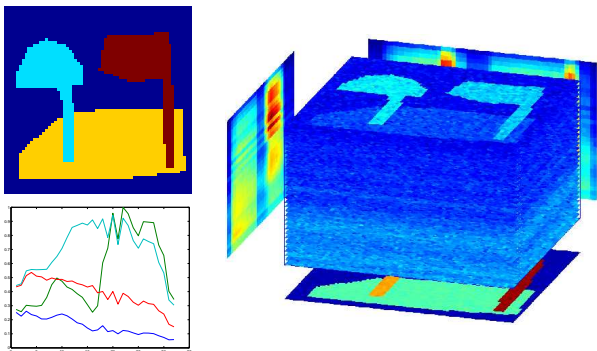
$$q(z|\mathbf{A}, \underline{\theta}, \underline{x}) \propto \prod_r p(\mathbf{x}(r)|z(r), \mathbf{A}, \theta) q(z(r)|\bar{z}(r'), r' \in \mathcal{V}(r)) \\ \propto \prod_r q(z(r)|\bar{z}(r'), r' \in \mathcal{V}(r), \mathbf{A}, \theta, \mathbf{x}(r))$$

where  $\bar{z}(r)$  can be computed by

$$\bar{z}(r) = \frac{\sum_{z(r)} z(r) q(z(r)|\bar{z}(r'), r' \in \mathcal{V}(r), \mathbf{A}, \theta, \mathbf{x}(r))}{\sum_{z(r)} q(z(r)|\bar{z}(r'), r' \in \mathcal{V}(r), \mathbf{A}, \theta, \mathbf{x}(r))}$$

#### 4. SIMULATION RESULTS

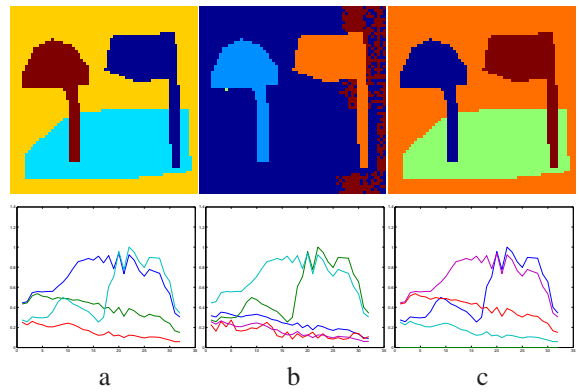
The main objectives of these simulations are: first to show that the proposed algorithm gives the desired results, and second to compare its relative performances with respect to some classical methods. For this purpose, first we generated some simulated data according to the data generating model, i.e.; starting by generating  $z(r)$ , then the sources  $s(r)$ , then using some given spectral signatures obtained from real materials construct the mixing matrix  $\mathbf{A}$  and finally generate data  $\mathbf{x}(r)$ . Fig. 2 shows an example of such data generated with the following parameters:  $m = 32, n = 4, K = 4$  and SNR=20 dB and Fig. 3 shows a comparison of the results obtained by two classical spectral and image classification methods using the classical  $K$ -means with the results obtained by the proposed method. Some other simulated results as well as the results obtained on real data will be given in near future.



**Fig. 2.** Data generating process:  $z(r)$  with  $K = 4$ , 4 spectral signatures used to construct the mixing matrix  $\mathbf{A}$  and  $m = 32$  images of size (64x64) used as data.

#### 5. CONCLUSION

Classical methods of dimensionality reduction in hyperspectral imaging use classification methods either to classify the spectra or to classify the images in  $K$  classes where  $K$  is, in general, much less than the number of spectra or the number



**Fig. 3.** Dimensionality reduction by different methods: a) Spectral classification using  $K$ -means, b) Image classification using  $K$ -means, c) Proposed method. Upper row shows estimated  $z(r)$  and lower row the estimated spectra. These results have to be compared to the original  $z(r)$  and spectra in previous figure.

of observed images. However, these methods neglect either the spatial organization of the spectra or the spectral property of the pixels along the spectral bands. In this paper, we considered the dimensionality reduction problem in hyperspectral images as a blind source separation and presented a Bayesian estimation approach with a particular hierarchical prior model for the observations and sources which accounts for both spectral and spatial structure of the data, and thus, gives the possibility to jointly do dimensionality reduction, classification of spectra and segmentation of the images.

#### 6. REFERENCES

- [1] K. Sasaki, S. Kawata, and S. Minami, "Component analysis of spatial and spectral patterns in multispectral images. I. basics," *Journal of the Optical Society of America. A*, vol. 4, no. 11, pp. 2101–2106, 1987.
- [2] L. Parra, C. Spence, A. Ziehe, K.-R. Mueller, and P. Sajda, "Unmixing hyperspectral data," in *Advances in Neural Information Processing Systems 13, (NIPS'2000)*. 2000, pp. 848–854, MIT Press.
- [3] Nadia Bali and Ali Mohammad-Djafari, "Mean Field Approximation for BSS of images with compound hierarchical Gauss-Markov-Potts model," in *MaxEnt05, San Jos CA, US*. Aug. 2005, American Institute of Physics (AIP).
- [4] Nadia Bali and Ali Mohammad-Djafari, "Approxiamtion en champ moyen pour la sparation de sources appliqu aux images hyperspectrales," in *Actes 20<sup>e</sup> coll. GRETSI*, Louvains Belgium, Sep. 2005.
- [5] Hichem Snoussi, *Approche bayssienne en sparation de sources. Applications en imagerie.*, Phd thesis, Université de Paris-Sud, Orsay, France, Sep. 2003.
- [6] J. Zhang, "The mean field theory in EM procedures for blind Markov random field image restoration," *IEEE Trans. Image Processing*, vol. 2, no. 1, pp. 27–40, Jan. 1993.