Joint Bayesian Hyperspectral Unmixing for change detection

Walma Gharbi, Lotfi Chaari, Amel Benazza-Benyahia

JOINT BAYESIAN HYPERSPECTRAL UNMIXING FOR CHANGE DETECTION

W. Gharbi [1,2] , L. Chaari [3], and A. Benazza-Benyahia[1]*

[1] University of Carthage, COSIM Lab, SUP'COM, LR11TIC04, Tunis, Tunisia
[2] University of Sfax, Digital Research Center of Sfax, Tunisia
[3] University of Toulouse, IRIT-ENSEEIHT, France

ABSTRACT

Spectral unmixing allows to extract endmembers and estimate their proportions in hyperspectral data. Each observed pixel is considered to be a linear combination of several endmembers spectra. Based on a novel hierarchical Bayesian model, change detection into hyperspectral images is achieved by unmixing. A Gibbs sampler is proposed to overcome the complexity of integrating the resulting posterior distribution. The performance of the proposed Bayesian change detection method is evaluated on real data. It provides binary detection with a precision rate up to 98.90%.

Index Terms— Change Detection, Hyperspectral Imaging, Bayesian methods, MCMC methods, Linear unmixing.

1. INTRODUCTION

Among the new generation of remote sensing sensors, HyperSpectral (HS) ones have a very high spectral resolution (e.g., 5-10 nm) that allows acquiring images in hundreds and even thousands of contiguous spectral channels. Due to the rich information in HS Images (HSIs) and their detailed spectral sampling, Change Detection (CD) is dramatically improved [1]. Therefore, it is possible to detect subtle changes in contrast to multispectral images where only significant changes are detected. Thus, it is important to exploit the fine spectral variations of HSIs for CD while taking into consideration the challenges of high dimensionality, redundancy and noise.

In this context, a Multivariate Alteration Detection (MAD) technique based on canonical correlation analysis is proposed for HS-CD in [2]. This technique allows to derive a number of components which are then separated to no-change components and change components based on a defined change probability measure. In [3], an iteratively reweighted MAD method was proposed as an improved version of MAD with more reliable components related to the uncertainty of observations. A no change measure was computed based on MAD variates and a reweighting function at each iteration. Despite the effectiveness of supervised HS-CD methods, they require prior knowledge to extract most informative components related to change which can be time consuming. Besides, they all assume that pure pixels (pixels that contain only one material) are present in HS data. This pure pixel assumption is not always true given the spatial resolution of HSIs (mostly around 30m). Thus, it is more common to find more than one distinct material in the area covered by a pixel (called a mixed pixel). As a solution for this phenomenon, spectral unmixing was proposed for HS-CD. Its aim is to recover the spectral signatures of the pure materials in the scene (called endmembers) and to estimate their proportions in each pixel (called fractional abundances). Despite the fact that single image unmixing has been extensively studied in the literature [1], multitemporal spectral unmixing was slightly investigated.

These methods are able to analyze the spectral variation of HSIs for CD. The authors of [4] have proposed a Linear Mixture Model for endmember and abundance estimation of each image separately. Then, the binary CM is obtained by comparing the estimated abundance vectors and performing a similarity thresholding using the Otsu's method [5]. In [6], the authors analyzed the Difference Image (DI) between the obtained abundances of the bi-temporal images. The threshold selection was obtained through a statistical analysis of 25 unchanged land-cover sample plots. The posterior probability of classification was also analyzed as an attempt to reduce the error rate [7]. However, the selection of the training samples for post-classification comparison requires an expert intervention. These methods are mostly developed in a supervised framework and rely on the availability of training samples for classification or thresholding.

To this date, the number of proposed HS-CD by unmixing is still narrow. However, there are several approaches to tackle the problem of HS unmixing which can be divided into geometrical, sparse unmixing and statistical approaches [1]. As aforementioned, geometrical methods are based on pure pixel assumption [8]. Sparse unmixing allows to find the optimal subset of spectral signatures from a spectral library that can model each mixed pixel of the studied area [9]. However, the sparse-based technique assumes the availability of standard and publicly available spectral libraries. Statistical methods have proved their efficiency in HS unmixing [10].

*This work is supported by the Tunisian program "Projets de Recherche Fédérés" of the Ministry of Higher Education and Scientific Research under the project "Supervision Sensitive de lieux Sensibles multi-capteurs : SuperSense"

Indeed, the inverse problem of spectral unmixing is formulated as a statistical inference problem. Despite the performance of statistical methods under a Bayesian framework for HS unmixing [10], their adoption for CD is less investigated. To this end, we propose a fully unsupervised automatic solution for HS-CD by spectral unmixing based on a Hierarchical Bayesian Model. By adopting a Bayesian perspective, the inference engine is the posterior density of the unknown endmemebers and abundances to be estimated. The proposed method analyzes the spectral variations of HSIs to perform CD without the need of reference samples. Besides, most HS-CD methods have performed endmember extraction and abundance estimation separately. Conversely, the proposed method allows to jointly estimate these parameters by adopting the equivalent priors that enforce the constraints of the abundance fractions when extracting the endmembers. This paper is organized as follows. In the Section 2, we present the proposed method for HS-CD by unmixing including the adopted Hierarchical Bayesian Model. The experimental results are given in Section 3 demonstrating its validity and performance. Conclusions are drawn in Section 4.

2. JOINT BAYESIAN UNMIXING FOR HS-CD

2.1. Problem formulation

Let $\boldsymbol{X}^1$ and $\boldsymbol{X}^2$ denote two HSIs in $\mathbb{R}^{P\times B}$, acquired over the same geographical area at different dates t_1 and t_2, respectively where B is the number of the spectral channels and P is the number of pixels. Let $\boldsymbol{x}_p^1$ and $\boldsymbol{x}_p^2$ denote the vectorized version of the p^{th} pixel for all spectral channels of $\boldsymbol{X}^1$ and $\boldsymbol{X}^2$, respectively. The DI vector $\boldsymbol{y}_p$ reads:

$$\boldsymbol{y}_p = \boldsymbol{x}_p^2 - \boldsymbol{x}_p^1. \tag{1}$$

We propose to perform CD by spectral unmixing within each observed pixel $\boldsymbol{y}_p$. The observation model writes:

$$\boldsymbol{y}_p = \sum_{r=1}^{R} a_{p,r}\boldsymbol{m}_r + \boldsymbol{n}_p, \tag{2}$$

where R is the number of endmembers , $\boldsymbol{m}_r = [m_{r,1}, \ldots, m_{r,B}]^{\mathrm{T}}$ indicates the r^{th} material spectrum, $a_{p,r}$ is its fraction in the p^{th} pixel and the abundance vector of a pixel p is $\boldsymbol{a}_p = [a_{p,1}, ..., a_{p,R}]$. The $\boldsymbol{n}_p = [n_{p,1}, \ldots, n_{p,B}]^{\mathrm{T}}$ is an additive i.i.d. Gaussian noise vector with a diagonal covariance matrix $\mathbf{S}_n = \sigma^2\boldsymbol{I}_B$. Considering all pixels, standard matrix notation yields

$$\boldsymbol{Y} = \boldsymbol{M}\boldsymbol{A} + \boldsymbol{N}, \tag{3}$$

where $\boldsymbol{Y} = [\boldsymbol{y}_1, ..., \boldsymbol{y}_P]$, $\boldsymbol{M} = [\boldsymbol{m}_1, ..., \boldsymbol{m}_R]$, $\boldsymbol{A} = [\boldsymbol{a}_1, ..., \boldsymbol{a}_P]$ and $\boldsymbol{N} = [\boldsymbol{n}_1, ..., \boldsymbol{n}_P]$. Based on the spectral signature mixture of the DI vector $\boldsymbol{Y}$, we perform a fully automatic and unsupervised CD. To this end, we propose a hierarchical Bayesian model to perform CD by unmixing and joint estimation of the endmembers and the abundance maps.

2.2. Hierarchical Bayesian model

Adopting a Bayesian approach, one needs to define the used likelihood and priors.

2.2.1. Likelihood

According to the linear mixture model defined in Eq. (2) and the assumption of an additive Gaussian noise of variance σ^2, the likelihood writes as

$$f_p(\boldsymbol{y}_p|\boldsymbol{M}, \boldsymbol{a}_p, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{B}{2}} \exp \quad -\frac{\|\boldsymbol{y}_p - \boldsymbol{M}a_p\|^2}{2\sigma^2}\Bigg), \tag{4}$$

where $\|\cdot\|$ denotes the Euclidean norm. Since the noise vectors $\boldsymbol{n}_p (p = 1, .., P)$ are assumed to be independent, the likelihood of all the observations $\boldsymbol{Y}$ is defined as follows

$$f(\boldsymbol{Y}|\boldsymbol{M}, \boldsymbol{A}, \sigma^2) = \prod_{i=1}^{P} f_p(\boldsymbol{y}_p|\boldsymbol{M}, \boldsymbol{a}_p, \sigma^2). \tag{5}$$

2.2.2. Prior for endmembers

The Principal Component Analysis (PCA) technique is performed to reduce the dimension and it is applied on $\boldsymbol{Y}$ to obtain a subspace $\mathcal{V}_{\mathcal{K}}$ spanned by K orthogonal axes $(\boldsymbol{v}_1, ..., \boldsymbol{v}_K)$. We define $\mathbf{D} = \text{diag}(\lambda_1, \ldots, \lambda_K)$ and $\mathbf{V} = [\boldsymbol{v}_1, ..., \boldsymbol{v}_K]^{\mathrm{T}}$ respectively as the diagonal matrix of the K highest eigenvalues and the corresponding eigenvector matrix of the $(B \times B)$ empirical covariance matrix of the observation given by

$$\Upsilon = \tfrac{1}{P}\sum_{p=1}^{P}(\boldsymbol{y}_p - \bar{\mathbf{y}})(\boldsymbol{y}_p - \bar{\mathbf{y}})^{\mathrm{T}}.$$

Let $\boldsymbol{t}_r$ denote the projected endmembers spectra $\boldsymbol{m}_r$ onto $\mathcal{V}_{\mathcal{K}}$. The PCA projection $\boldsymbol{t}_r \in \mathbb{R}^K$ of the endmember $\boldsymbol{m}_r \in \mathbb{R}^B$ is defined as follows:

$$\boldsymbol{t}_r = \boldsymbol{P}(\boldsymbol{m}_r - \bar{y}), \tag{6}$$

with $\boldsymbol{P} = \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{V}$ and $\bar{\boldsymbol{y}}$ is the empirical mean of of $\mathbf{Y}$. The prior of the projected spectra $\boldsymbol{t}_r$ is a conjugate multivariate Gaussian Distribution (MGD) on the set $\mathcal{T}_r \subset \mathcal{V}_{\mathcal{K}}$. It is defined according the nonnegativity constraint of the corresponding reconstructed $(B \times 1)$ spectrum $\boldsymbol{m}_r$. Under the assumption of the independence between all the $\boldsymbol{t}_r$ vectors, the prior distribution for all the projected endmembers matrix $\boldsymbol{T}$ is

$$f(\boldsymbol{T}|\boldsymbol{E}, \boldsymbol{s}^2) = \prod_{r=1}^{R} \mathcal{N}_{\mathcal{T}_r}(\boldsymbol{t}_r|\boldsymbol{e}_r, \boldsymbol{s}_r^2\boldsymbol{I}_K). \tag{7}$$

2.2.3. Prior for abundances

Due to physical considerations [11], the fraction vectors have to satisfy both the nonnegativity and full-additivity constraints. In our case, the abundance vector $\boldsymbol{a}_p$ is defined $\boldsymbol{a}_p = [\boldsymbol{c}_p, a_{p,r}]^{\mathrm{T}}$ where $\boldsymbol{c}_p = [a_{p,1}, \ldots, a_{p,R-1}]^{\mathrm{T}}$ and $a_{p,R-1} = -\sum_{r=1}^{R-1} a_{p,r}$. The prior chosen for $\boldsymbol{c}_p$ is a uniform Bernoulli distribution which reinforces the ability to separate changed pixels from unchanged ones that are captured by the Bernoulli part of the prior

$$f(\boldsymbol{c}_p|\omega) = (1-\omega)\delta(\boldsymbol{c}_p) + \omega \times \mathcal{U}_{[-\prime,\prime]}(\boldsymbol{c}_p), \tag{8}$$

where ω is a weight belonging to $[0,1]$ that reflects the rate of non-zero coefficients. Under the assumption of statistical independence between the vectors, the full prior distribution for partial abundance matrix $\boldsymbol{C} = [\boldsymbol{c}_1, ..., \boldsymbol{c}_P]^T$ can be written:

$$f(\boldsymbol{C}) \sim \prod_{p=1}^{P} f_p(\boldsymbol{c}_p|\omega). \tag{9}$$

2.2.4. Prior for noise variance σ^2

Since σ^2 is a real positive scalar, we use an inverse gamma (IG) prior given by

$$f(\sigma^2|\alpha,\beta) = \mathcal{IG}(\sigma^2|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\sigma_n^{-2(\alpha+1)}\exp\left(-\frac{\beta}{\sigma^2}\right), \tag{10}$$

where $\Gamma(.)$ is the standard gamma function and the positive reals α and β refer to the shape and scale hyperparameters. This is a convenient choice for σ^2 since the inverse gamma distribution is the conjugate distribution of the normal distribution [12]. A non-informative prior distribution is chosen to reflect uncertainty of σ^2 in HSIs.

2.2.5. Hyperprior for ω

A non informative prior is used for ω which is the uniform distribution on $[0,1]$.

2.3. Bayesian inference scheme

We propose to estimate the unknown parameter vector $\boldsymbol{\theta} = \{\boldsymbol{C}, \boldsymbol{T}, \sigma^2\}$ as well as the hyperparameter ω. The joint posterior distribution of the vector $\{\boldsymbol{\theta}, \omega\}$ can be expressed as

$$f(\boldsymbol{\theta},\omega|\boldsymbol{Y},\alpha,\beta) \propto f(\boldsymbol{Y}|\boldsymbol{\theta})f(\boldsymbol{\theta}|\omega)f(\omega). \tag{11}$$

This hierarchical structure of the proposed model allows one to integrate out the hyperparameter ω from the joint distribution. Deriving a simple closed-form expression of the Bayesian estimators such as minimum mean square error or the maximum a posteriori from the posterior distribution is difficult. Therefore, we resort to the markov chain monte carlo methods to generate samples asymptotically distributed according to the posterior distribution. The Bayesian estimators can be approximated using these samples. Gibbs sampling is an iterative procedure that generates samples according to the conditional posteriors $f(\boldsymbol{C}|\boldsymbol{Y},\boldsymbol{T},\sigma^2)$, $f(\sigma^2|\boldsymbol{Y},\boldsymbol{C},\boldsymbol{T})$, $f(\boldsymbol{T}|\boldsymbol{Y},\boldsymbol{C},\sigma^2)$ and $f(\omega|\boldsymbol{Y},\boldsymbol{C})$. The maximum a posteriori was used to select the most probable value sampled for $\boldsymbol{c}_p$: $\widehat{\boldsymbol{c}}_{pMAP} = \arg\max_{\boldsymbol{c}_p} p(\boldsymbol{c}_p|\boldsymbol{Y})$. The estimated value of $c_{p,r}$ is computed as follows

If γ_i=0 **then**

$c_{p,r}$=0

Else

$c_{p,r} \sim \mathcal{N}_{\mathcal{S}}(\mu_p, \mathbf{S}_p)$

The estimators $\widehat{\sigma}^2$, $\widehat{\boldsymbol{T}}$ and $\widehat{\omega}$ are obtained by minimizing the average squared error by employing the related laws conditioned by Y.

2.3.1. Sampling according to $f(\sigma^2|\boldsymbol{Y},\boldsymbol{C},\boldsymbol{T})$

The conditional distribution of $\sigma^2|\boldsymbol{Y},\boldsymbol{C},\boldsymbol{T}$ is the following

$$\sigma^2|\boldsymbol{Y},\boldsymbol{C},\boldsymbol{T} \sim \mathcal{IG} \quad \alpha + \frac{P\times B}{2}, \beta + \sum_{p=1}^{P}\frac{\|\boldsymbol{y}_p - \boldsymbol{M}\boldsymbol{a}_p\|^2}{2}\Bigg). \tag{12}$$

2.3.2. Sampling according to $f(\omega|\boldsymbol{Y},\boldsymbol{C})$

Straightforward calculations indicate that the posterior of ω is a uniform distribution according to which it is easy to sample.

$$\omega|\boldsymbol{Y},\boldsymbol{C} \sim \mathcal{U}(0,1). \tag{13}$$

2.3.3. Sampling according to $f(\boldsymbol{C}|\boldsymbol{Y},\boldsymbol{T},\sigma^2)$

It can be easily shown that

$$f(\boldsymbol{c}_p|\boldsymbol{Y},\omega,\boldsymbol{T},\sigma^2) = \omega_{1,p}\delta(\boldsymbol{c}_p) + \omega_{2,p}\mathcal{N}_{\mathcal{S}}(\mu_p,\mathbf{S}_p), \tag{14}$$

where $\mathcal{N}_{\mathcal{S}}(\mu_p,\mathbf{S}_p) = \exp\left(-\frac{(\boldsymbol{c}_p-\boldsymbol{v}_p)^{\mathrm{T}}\mathbf{S}_p^{-1}(\boldsymbol{c}_p-\boldsymbol{v}_p)}{2}\right)1_{\mathcal{S}}(\boldsymbol{c}_p)$

and $\begin{cases}\mu_p = \boldsymbol{H}_p^{\mathrm{T}}\mathbf{S}_n^{-1}(\boldsymbol{y}_p - \boldsymbol{m}_R)\\ \mathbf{S}_p = \boldsymbol{H}_p^{\mathrm{T}}\mathbf{S}_n^{-1}\boldsymbol{H}_p\end{cases}$ with $\begin{cases}\mathbf{S}_n^{-1} = \frac{1}{\sigma}\boldsymbol{I}_B\\ \boldsymbol{H}_p = \boldsymbol{M}_{-R} - \boldsymbol{m}_R\mathbf{1}_{R-1}^{\mathrm{T}}\end{cases}$

$\boldsymbol{M}_{-R}$ denotes the matrix $\boldsymbol{M}$ whose R^{th} column has been removed. As a consequence, if the $\boldsymbol{c}_p$ coefficients are different to zero, they are distributed according to an MGD. The weights $(\omega_{l,p})_{1\le l\le 2}$ in Eq. (14) are computed using:

$\omega_{l,p} = \frac{\mu_{l,p}}{\sum_{l=1}^{2}\mu_{l,p}}$, where $\begin{cases}\mu_{1,p} = 1-\omega\\ \mu_{2,p} = \frac{\omega}{2}\exp\left(\frac{\mu_p^2}{2\sigma^2}\right)\sqrt{2\pi\sigma^2}\end{cases}$

2.3.4. Sampling according to $f(\boldsymbol{T}|\boldsymbol{Y},\boldsymbol{C},\sigma^2)$

The conditional posterior distribution of $\boldsymbol{t}_r (r = 1, ..., R)$ writes

$$f(\boldsymbol{t}_r|\boldsymbol{T}_{-r},\boldsymbol{c}_p,\sigma^2,\boldsymbol{Y}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{t}_r-\boldsymbol{\tau}_r)^{\mathrm{T}}\boldsymbol{\Lambda}_r^{-1}(\boldsymbol{t}_r-\boldsymbol{\tau}_r)\right]1_{\mathcal{T}_r}(\boldsymbol{t}_r), \tag{15}$$

where $\begin{cases}\boldsymbol{\Lambda}_r = \left[\sum_{p=1}^{P}{\boldsymbol{a}_{p,r}}^2\boldsymbol{U}^{\mathrm{T}}\mathbf{S}_n^{-1}\boldsymbol{U} + \frac{1}{\boldsymbol{s}_r^2}\boldsymbol{I}_k\right]^{-1}\\ \boldsymbol{\tau}_r = \boldsymbol{\Lambda}_r\left[\sum_{p=1}^{P}\boldsymbol{a}_{p,r}\boldsymbol{U}^{\mathrm{T}}\mathbf{S}_n^{-1}\epsilon_{p,r} + \frac{1}{\boldsymbol{s}_r^2}\boldsymbol{e}_r\right]\end{cases}$

and $\epsilon = \boldsymbol{y}_p - \boldsymbol{a}_{p,r}\bar{\mathbf{y}} + \mathbf{S}_{j\neq r}$, $\boldsymbol{m}_j = \boldsymbol{U}\boldsymbol{t}_j + \bar{\mathbf{y}}$. Therefore, the posterior distribution of $\boldsymbol{t}_r$ is the a truncated MGD that writes

$$f(\boldsymbol{t}_r|\boldsymbol{T}_{-r},\boldsymbol{c}_p,\sigma^2,\boldsymbol{Y}) \sim \mathcal{N}_{\mathcal{T}_r}(\boldsymbol{\tau}_r,\boldsymbol{\Lambda}_r). \tag{16}$$

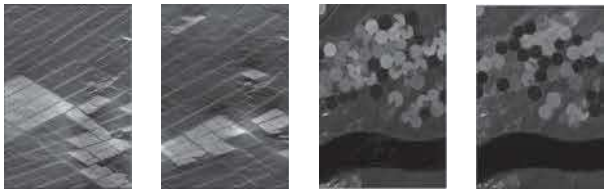
The conditional distributions used in the GS (namely Algorithm 1) are detailed in the following.

3. EXPERIMENTS

The first dataset covers a farmland near the city of Yancheng, Jiangsu province, China, with a size of 450×140 pixels. The second dataset is an irrigated agricultural field of Hermiston city in Umatilla County, Oregon, USA of a size 307×241.

Algorithm 1: Gibbs Sampler (GS).

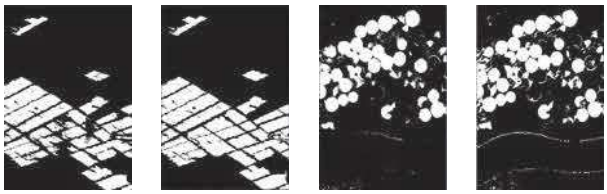
- Initialize $\boldsymbol{\sigma}^2$,$\boldsymbol{T}$,$\boldsymbol{C}$.
for $s = 1 \dots S$ **do**
  Sample σ^2 according to Eq.(12);
  Sample ω according to Eq.(13);
  for *for* $p = 1, \dots, R$ **do**
    Sample $\boldsymbol{c}_p$ according to Eq. ((14));
  end
end
for $r = 1, \dots, R$ **do**
  Sample $\boldsymbol{t}_r$ according to Eq.(15);
end



(a) (b) (c) (d)

Fig. 1. China dataset on: (a) May 3, 2006, (b) April 23, 2007, USA imagery on (c) May 1, 2004, (d) May 3, 2007

The bi-temporal HSIs are displayed in Fig. 1. Both datasets have 154 bands after noise elimination. For each dataset, two HSIs are used to compute the DI at a pixel level. However, the spectral unmixing exploits the spectral information of the HSIs and attempts to identify different spectra (called endmembers) within the spectrum of each pixel. It is worthnoting that even if different spectra are modeled, CD is only performed at a pixel level. The GS has been run with 200 iterations, including 100 iterations as a burn-in period to reach convergence. The estimated binary CM is obtained by taking



(a) (b) (c) (d)

Fig. 2. China dataset: (a) The binary CM, (b) ground truth map, USA dataset: (c) binary CM and (d) USA groundtruth map

into consideration all the spectral bands and it is illustrated in Fig. 2. This binary CM indicates for each pixel if it is changed or not. Our method provides a good estimation of change that minimizes the presence of noise or disconnected pixels. It is clear that all the changed pixels are correctly classified and the errors are mostly related to misclassified edge pixels. The accuracy of the binary CM is evaluated in terms of overall accuracy, precision and Kappa. The overall accuracy that represents the total percentage of correctly classified pixels, Kappa coefficient which refers to the percentage of agreement amended by the agreement expected by chance and precision which reflects the number of correctly detected pixels as changed. Table 1 presents the quantitative results that are consistent with the visual results.

Table 1. CD accuracy measures for China and USA datasets

Accuracy measures	China	USA
Overall Accuracy (%)	95.81	94.89
Precision (%)	98.90	92.65
Kappa (%)	95.03	94.48

4. CONCLUSION

A joint Bayesian unmixing model for HSIs was adopted to perform CD. The proposed method was validated using two different HS datasets and obtained successful rates of accuracy for both experiments. As future work, we aim at extending our model to consider super-resolution for CD that improve the spatial resolution of HSIs. Other spatial resolution of HSIs will be considered for additional experiments.

5. ACKNOWLEDGEMENTS

The authors would expend their sincere appreciation to Dr. Abdelrahim Halimi for providing them Matlab code that was updated for the problem of HS-CD.

6. REFERENCES

[1] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, "A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 140–158, 2019.

[2] M. Frank and M. Canty, "Unsupervised change detection for hyperspectral images," in *Process. 12th JPL Airborne Earth Sci. Workshop*, 2003, pp. 63–72.

[3] A. Nielsen, "The regularized iteratively reweighted mad method for change detection in multi-and hyperspectral data," *IEEE Trans. on Image Process..*, vol. 16, no. 2, pp. 463–478, 2007.

[4] Q. Du, L. Wasson, and R. King, "Unsupervised linear unmixing for change detection in multitemporal airborne hyperspectral imagery," in *Int. Workshop Anal. Multi-Temp RS Images*. IEEE, 2005, pp. 136–140.

[5] N. Otsu, "A threshold selection method from gray-level histograms," *Trans. on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[6] D. Lu, M. Batistella, and E. Moran, "Multitemporal spectral mixture analysis for amazonian land-cover change detection," *Canadian Journal of Remote Sensing*, vol. 30, no. 1, pp. 87–100, 2004.

[7] J. Chen, X. Chen, X. Cui, and J. Chen, "Change vector analysis in posterior probability space: A new method for land cover change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 317–321, 2010.

[8] A. Ertürk and A. Plaza, "Informative change detection by unmixing for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1252–1256, 2015.

[9] A. Ertürk, M. D. Iordache, and A. Plaza, "Sparse unmixing-based change detection for multitemporal hyperspectral images," *J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 9, no. 2, pp. 708–719, 2015.

[10] N. Dobigeon, S. Moussaoui, M. Coulon, J. Y. Tourneret, and A. O. Hero, "Joint bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. on Signal Process.*, vol. 57, no. 11, pp. 4355–4368, 2009.

[11] N. Dobigeon, J. Y. Tourneret, and C. I. Chang, "Semi-supervised linear spectral unmixing using a hierarchical bayesian model for hyperspectral imagery," *IEEE Trans. on Signal Process.*, vol. 56, no. 7, pp. 2684–2695, 2008.

[12] G. C. Tiao and W. Y. Tan, "Bayesian analysis of random-effect models in the analysis of variance. I. posterior distribution of variance-components," *Biometrika*, vol. 52, no. 1/2, pp. 37–53, 1965.