Bayesian segmentation of hyperspectral images

Adel Mohammadpour[†*], Olivier Féron[†] and Ali Mohammad-Djafari[†]

[*]*Department of Statistics, Faculty of Mathematics & Computer Science*
Amirkabir University of Technology, 424 Hafez Ave., 15914 Tehran, Iran
[†]*Laboratoire des Signaux et Systèmes, (CNRS-Supélec-UPS)*
Supélec, Plateau de Moulon, 91192 Gif-sur-Yvette, France

Abstract. In this paper we consider the problem of joint segmentation of hyperspectral images in the Bayesian framework. The proposed approach is based on a Hidden Markov Modeling (HMM) of the images with common segmentation, or equivalently with common hidden classification label variables which is modeled by a Potts Markov Random Field. We introduce an appropriate Markov Chain Monte Carlo (MCMC) algorithm to implement the method and show some simulation results.

INTRODUCTION

The most significant recent advances in remote sensing has been the development of hyperspectral sensors and software to analyze the resulting image data. Over the past decade hyperspectral image analysis has matured into one of the most powerful and fastest growing technologies in the field of remote sensing. The "hyper" in hyperspectral means "over" as in "too many" and refers to the large number of measured wavelength bands. Hyperspectral images are spectrally overdetermined, which means that they provide ample spectral information to identify and distinguish spectrally unique materials. Hyperspectral imagery provides the potential for more accurate and detailed information extraction than possible with any other type of remotely sensed data. However the huge amount of data in hyperspectral images make its information exploitation difficult and image processing tools (classification, segmentation, comprising and coding) are needed to summarize the information included in these data.

This paper will introduce a segmentation method for hyperspectral images. Several unsupervised and supervised algorithms have been developed for segmentation of multispectral images. However, these algorithms fail to deliver high accuracies for classifying hyperspectral images [1, 2, 3, 4, 5]. In this paper we consider the problem of joint segmentation of hyperspectral images in the Bayesian framework.

The proposed approach is based on a Hidden Markov Modeling (HMM) of the images with common segmentation, or equivalently with common hidden classification label variables which is modeled by a Potts Markov Random Field. We introduce an appropriate Markov Chain Monte Carlo (MCMC) algorithm to implement the method and show some simulation results. This approach has previously been considered by [6] for multispectral images.

In that work, the pixels of the same region in different images are assumed independent. This independence assumption is a valid hypothesis for multispectral images.

However in hyperspectral images the pixel values in each channel are not independent. This work is then an extension to that work by considering a Markov model for these pixels along each channel.

This paper is organized as follows: In the next two sections first we introduce our method for segmentation of hyperspectral images in the Bayesian framework. Then we propose an appropriate MCMC Gibbs sampling particularly designed for this segmentation task. Finally, in the last section we present some simulation results to show the performances of the proposed methods.

MODELING FOR BAYESIAN SEGMENTATION

Let $g_i(r)$ be the observed value of the pixel r, $r \in \boldsymbol{Z}^2$, in the spectral band i of a hyperspectral image. We model the observations by

$$g_i(r) = f_i(r) + \varepsilon_i(r), \qquad i = 1, \ldots, M \tag{1}$$

where $f_i(r)$ is the unknown perfect value of $g_i(r)$ and $\varepsilon_i(r)$ is a noise. Note that if we consider images, the pixels r belong to a finite lattice $\mathcal{S}$, and we will note S the number of pixels of this lattice. In the following we also use the notations

$$\boldsymbol{g}_i = \boldsymbol{f}_i + \boldsymbol{\varepsilon}_i \qquad or \qquad \underline{\boldsymbol{g}} = \underline{\boldsymbol{f}} + \underline{\boldsymbol{\epsilon}},$$

where $\boldsymbol{g}_i = \{g_i(r), r \in \mathcal{S}\}$ and $\underline{\boldsymbol{g}} = \{\boldsymbol{g}_i, i = 1, \ldots, M\}$ and a similar definition for $\underline{\boldsymbol{f}}$ and $\underline{\boldsymbol{\epsilon}}$. We introduce a label variable $z(r)$ for the regions and consider the region labels as common feature between all images. Thus the hidden variable $\boldsymbol{z} = \{z(r), r \in \mathcal{S}\}$ represent a common classification of the images for different bands. The main result of this paper is estimation of joint segmentation label $\boldsymbol{z}$.

Assuming independent noises $\boldsymbol{\varepsilon}_i$ among the different observations we have

$$p(\underline{\boldsymbol{g}}|\underline{\boldsymbol{f}}, \boldsymbol{\varepsilon}_i) = \prod_{i=1}^{M} p(\boldsymbol{g}_i|\boldsymbol{f}_i) = \prod_{i=1}^{M} p_{\boldsymbol{\varepsilon}_i}(\boldsymbol{g}_i - \boldsymbol{f}_i).$$

Assuming $\boldsymbol{\varepsilon}_i$ centered, white and Gaussian $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(0, \sigma^2_{\boldsymbol{\varepsilon}_i}\boldsymbol{I})$, and S the number of pixels of an image, we have:

$$p(\boldsymbol{g}_i|\boldsymbol{f}_i, \boldsymbol{\varepsilon}_i) = \left(\frac{1}{2\pi\sigma^2_{\boldsymbol{\varepsilon}_i}}\right)^{\frac{S}{2}} \exp\left\{-\frac{1}{2\sigma^2_{\boldsymbol{\varepsilon}_i}}||\boldsymbol{g}_i - \boldsymbol{f}_i||^2\right\},$$

where $\sigma^2_{\boldsymbol{\varepsilon}_i} \sim \mathcal{IG}(\alpha_0^{\boldsymbol{\varepsilon}_i}, \beta_0^{\boldsymbol{\varepsilon}_i})$ with unknown fixed parameters $\alpha_0^{\boldsymbol{\varepsilon}_i}$ and $\beta_0^{\boldsymbol{\varepsilon}_i}$ (inverse gamma is conjugate prior for a random variance in the Gaussian case).

To assign $p(\boldsymbol{f}_i|\boldsymbol{z}, .)$ we first define the sets of pixels which are in the same class:

$$\begin{aligned} R_k &= \{r : z(r) = k\}, \qquad |R_k| = n_k \\ \boldsymbol{f}_{ik} &= \{f_i(r) : z(r) = k\}. \end{aligned}$$

We assume that all the pixels $\boldsymbol{f}_{ik}$ of an image $\boldsymbol{f}_i$ which are in the same class k will be Gaussian with a random mean m_{ik} and a random variance $\sigma^2_{i\,k}$, i.e.

$$f_i(r)|z(r)=k, m_{ik}, \sigma^2_{i\,k} \sim \mathcal{N}(m_{ik}, \sigma^2_{i\,k}) \qquad \forall r \in \mathcal{S},$$

With these notations we have :

$$p(\boldsymbol{f}_{ik}|m_{ik}, \sigma^2_{i\,k}) = \left(\frac{1}{\sqrt{2\pi\sigma^2_{i\,k}}}\right)^{n_k} \exp\left\{-\frac{1}{2\sigma^2_{i\,k}}||\boldsymbol{f}_{ik} - m_{ik}\mathbf{1}||^2\right\},$$

and thus for $i = 1, \ldots, M$

$$p(\boldsymbol{f}_i|\boldsymbol{z}, m_{ik}, \sigma^2_{i\,k}) \;=\; \prod_{k=1}^{K}\left(\frac{1}{\sqrt{2\pi\sigma^2_{i\,k}}}\right)^{n_k} \exp\left\{-\frac{1}{2\sigma^2_{i\,k}}||\boldsymbol{f}_{ik} - m_{ik}\mathbf{1}||^2\right\},$$

where $\mathbf{1}$ is a vector with all components equal to 1, $\sigma^2_{i\,k} \sim \mathcal{IG}(\alpha_{i0}, \beta_{i0})$, with unknown fixed parameters, and m_{ik} is an autoregressive of order 1, $AR(1)$, for each class k i.e.

$$m_{ik} \;=\; \phi_k m_{i-1\,k} + \eta_{ik}, \tag{2}$$

where $\eta_{ik} \sim \mathcal{N}(0, \sigma^2_{i\,0})$, ϕ_k and $\sigma^2_{i\,0}$ are unknown fixed parameters. Therefore

$$m_{ik}|m_{i-1\,k} \sim \mathcal{N}(\phi_k m_{i-1\,k}, \sigma^2_{i\,0}).$$

The assumption of (2) is the main difference of this paper with [6], i.e. in hyperspectral images the pixel values of a class k, in each channel, are not independent.

Using the relation (1) and the density $p(\boldsymbol{f}_i|\boldsymbol{z}, m_{ik}, \sigma^2_{i\,k})$ and $p(\varepsilon_i)$, we can calculate $p(\boldsymbol{g}_i|\boldsymbol{z}, .)$, i.e.

$$\boldsymbol{g}_i|\boldsymbol{z}, m_{ik}, \sigma^2_{i\,k}, \sigma^2_{\varepsilon_i} \sim \mathcal{N}(m_{ik}, \sigma^2_{i\,k} + \sigma^2_{\varepsilon_i}). \tag{3}$$

Finally we have to assign $P(\boldsymbol{z})$. As we introduced the hidden variable $\boldsymbol{z}$ for finding statistically homogeneous regions in images, it is natural to define a spatial dependency on these labels. The simplest model to account for this desired local spatial dependency is a Potts Markov Random Field model:

$$P(\boldsymbol{z}) = \frac{1}{T(\alpha)} \exp\left\{\alpha \sum_{r\in\mathcal{S}} \sum_{s\in\mathcal{V}(r)} \delta(z(r) - z(s))\right\}, \tag{4}$$

where $\mathcal{S}$ is the set of pixels, $\delta(0) = 1$, $\delta(t) = 0$ if $t \neq 0$, $\mathcal{V}(r)$ denotes the neighborhood of the pixel r (here we consider a neighborhood of 4 pixels), $T(\alpha)$ is the partition function or the normalization constant and α represents the degree of the spatial dependency of the variable $\boldsymbol{z}$.

ESTIMATION USING MCMC

Let $\boldsymbol{m}_i = (m_{ik})_{k=1,\ldots,K}$ and $\boldsymbol{\sigma}_i^2 = (\sigma_{i\,k}^2)_{k=1,\ldots,K}$ be the means and the variances of the pixels in different regions of the images $\boldsymbol{f}_i$. We define $\boldsymbol{\theta}_i$ as the set of all the parameters which must be estimated in the Bayesian framework:

$$\boldsymbol{\theta}_i = (\sigma_{\varepsilon_i}^2, \boldsymbol{m}_i, \boldsymbol{\sigma}_i^2), \quad i = 1, \ldots, M$$

and we note $\underline{\boldsymbol{\theta}} = (\boldsymbol{\theta}_i)_{i=1,\ldots,M}$. Now we can write the expression of the joint posterior of $p(\underline{\boldsymbol{f}}, \boldsymbol{z}, \underline{\boldsymbol{\theta}} | \underline{\boldsymbol{g}})$ by using the relations in the previous section. Then we propose the following Gibbs sampler,

$$\begin{aligned} \underline{\boldsymbol{f}} &\sim \underline{\boldsymbol{f}} | \underline{\boldsymbol{g}}, \boldsymbol{z}, \underline{\boldsymbol{\theta}} \\ \boldsymbol{z} &\sim \boldsymbol{z} | \underline{\boldsymbol{\theta}}, \underline{\boldsymbol{g}} \\ \underline{\boldsymbol{\theta}} &\sim \underline{\boldsymbol{\theta}} | \underline{\boldsymbol{f}}, \underline{\boldsymbol{g}}, \boldsymbol{z} \end{aligned}$$

to generate samples $(\underline{\boldsymbol{f}}, \boldsymbol{z}, \underline{\boldsymbol{\theta}})^{(1)}, (\underline{\boldsymbol{f}}, \boldsymbol{z}, \underline{\boldsymbol{\theta}})^{(2)}, \cdots$, and use them to compute any statistics (such as mean or median). We may note that in each of the previous Gibbs sampling steps, we again use Gibbs scheme to sample. For example $\underline{\boldsymbol{f}} | \underline{\boldsymbol{g}}, \boldsymbol{z}, \underline{\boldsymbol{\theta}}$ by alternate sampling of $\boldsymbol{f}_i | \boldsymbol{g}_i, \boldsymbol{z}, \boldsymbol{\theta}_i$. This procedure is also valid for $\underline{\boldsymbol{\theta}}$.

It can be shown that $f_i(r) | g_i(r), z(r), \boldsymbol{\theta}_i$ has a Gaussian distribution and it can be sampled very easily. On the other hand,

$$p(\boldsymbol{\theta}_i | \boldsymbol{f}_i, \boldsymbol{g}_i, \boldsymbol{z}) \propto p(\sigma_{\varepsilon_i}^2 | \boldsymbol{f}_i, \boldsymbol{g}_i)\; p(\boldsymbol{m}_i, \boldsymbol{\sigma}_i^2 | \boldsymbol{f}_i, \boldsymbol{z}).$$

For the last term $p(\boldsymbol{m}_i, \boldsymbol{\sigma}_i^2 | \boldsymbol{f}_i, \boldsymbol{z})$ we have to use a Gibbs algorithm and then sample following the conditional distributions $p(\boldsymbol{m}_i | \boldsymbol{\sigma}_i^2, \boldsymbol{f}_i, \boldsymbol{z})$ and $p(\boldsymbol{\sigma}_i^2 | \boldsymbol{m}_i, \boldsymbol{f}_i, \boldsymbol{z})$. It can be shown that $\sigma_{\boldsymbol{\varepsilon}_i}^2 | \boldsymbol{f}_i, \boldsymbol{g}_i$ and $\sigma_{i\,k}^2 | \boldsymbol{f}_i, \boldsymbol{z}$ have inverse gamma distributions and $m_{ik} | \boldsymbol{f}_i, \boldsymbol{z}, \sigma_{i\,k}^2, m_{i-1\,k} \sim \mathcal{N}(\mu_{ik}, v_{i\,k}^2)$, with

$$\begin{aligned} \mu_{ik} &= v_{i\,k}^2 \left(\frac{\phi_k m_{i-1\,k}}{\sigma_{i\,0}^2} + \frac{1}{\sigma_{i\,k}^2} \sum_{r \in R_k} f_i(r) \right), \\ v_{i\,k}^2 &= \left(\frac{n_k}{\sigma_{i\,k}^2} + \frac{1}{\sigma_{i\,0}^2} \right)^{-1}. \end{aligned} \tag{5}$$

Note that if m_{ik}s are independent as it is the case of multispectral images in [6] then we had

$$\mu_{ik} = v_{i\,k}^2 \left(\frac{m_{i0}}{\sigma_{i\,0}^2} + \frac{1}{\sigma_{i\,k}^2} \sum_{r \in R_k} f_i(r) \right),$$

where m_{i0} is a fixed number.

Finally, we can write the posterior probability of $\boldsymbol{z}$ by

$$P(\boldsymbol{z} | \underline{\boldsymbol{g}}, \underline{\boldsymbol{\theta}}) \propto p(\underline{\boldsymbol{g}} | \boldsymbol{z}, \underline{\boldsymbol{\theta}})\; P(\boldsymbol{z}) = \prod_{i=1}^{M} p(\boldsymbol{g}_i | \boldsymbol{z}, \boldsymbol{\theta}_i)\; P(\boldsymbol{z}), \tag{6}$$

where can be calculated by using (3) and (4). As we choose a Potts Markov Random Field model for the labels $\boldsymbol{z}$, we may note that an exact sampling of the *a posteriori* distribution $P(\boldsymbol{z}|\underline{\boldsymbol{g}},\underline{\boldsymbol{\theta}})$ is impossible. But we can still use a Gibbs sampling to generate parallel samples of $\boldsymbol{z}$.

For simplicity sake, we estimate the parameters ϕ_k, $k=1,\ldots,K$ with a classical method and we consider it as constant in this section. If the series $\{m_{ik}\}_{i=1,\ldots,M}$ has an AR model, (k is fixed), then we can estimate ϕ_k efficiently, because the number of images M is large.

Here we give the summary of the proposed algorithm for estimating $\boldsymbol{z}$ which has the following steps:

1. Find an initial joint segmentation of the hyperspectral image by using any simple segmentation or classification method such as k-means method,
2. Calculate $\{m_{ik}\}_{i=1,\ldots,M}$ for $k=1,\ldots,K$,
3. Fit an AR model for each series $\{m_{ik}\}_{i=1,\ldots,M}$, $k=1,\ldots,K$ with a classical method,
4. Use the proposed Gibbs algorithm to generate samples of $\{m_{ik}\}$ using (5) and $\boldsymbol{z}$ using (6).

SIMULATION

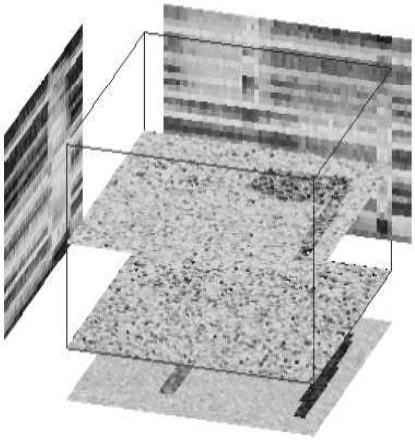
Synthetic data

To measure the performance of the proposed method, first we generate artificially a set of data, starting by a known segmentation $\boldsymbol{z}$ and generate the images $\boldsymbol{f}_i|\boldsymbol{z}=k$ as homogeneous Gaussian with known mean m_{ik} and variance $\sigma^2_{i\,k}$ with $m_{ik}=\phi_k m_{i-1k}+\eta_{ik}$, where $|\phi_k|<1$ is a fixed value. Then we generate the $\boldsymbol{g}_i$ by adding a Gaussian noise. Finally we use these data as input for different segmentation methods and compare their relative performance.

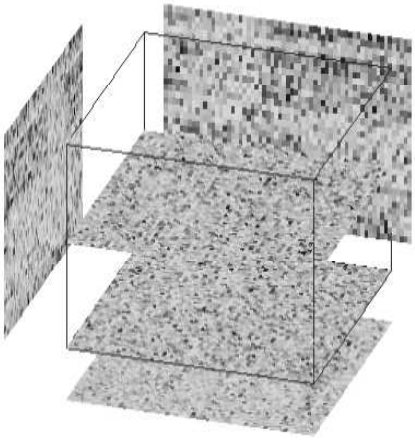
In this example, we know the original $\boldsymbol{z}$. Therefore, comparison can be done by counting the number of misclassified pixels.

Figure 1 shows the results of this experiment. The first row of Figure 1 shows $\boldsymbol{z}$, $\underline{\boldsymbol{f}}$, and $\underline{\boldsymbol{g}}$. The second row shows the results of segmentation of images $\underline{\boldsymbol{g}}$ by two methods: the first assume independence of m_{ik} ($i=1,\cdots,M$) and second which is the proposed method use an AR modeling of m_{ik}. The number of miss classified pixels for $\hat{\boldsymbol{z}}_1$ is more than 1000 and for $\hat{\boldsymbol{z}}_2$ is less than 200.

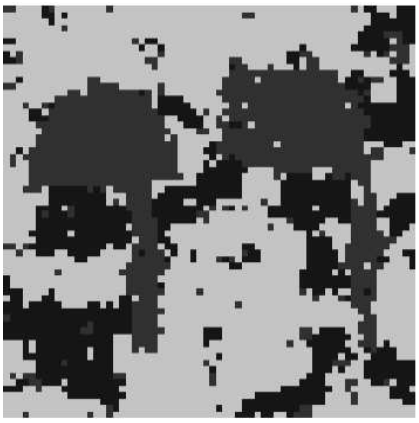
$\boldsymbol{z}$



$\underline{\boldsymbol{f}}$ and their marginals



$\underline{\boldsymbol{g}}$ and their marginals



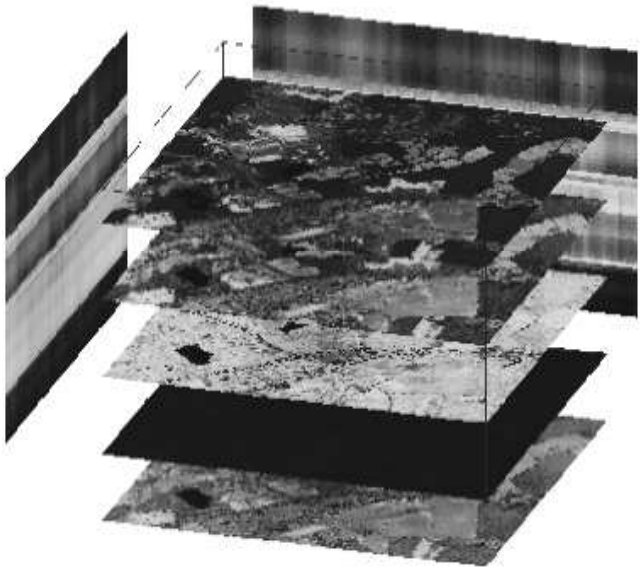
$\hat{\boldsymbol{z}}_1$ with independence and

$\hat{\boldsymbol{z}}_2$, with AR assumption on $\{m_{ik}\}$

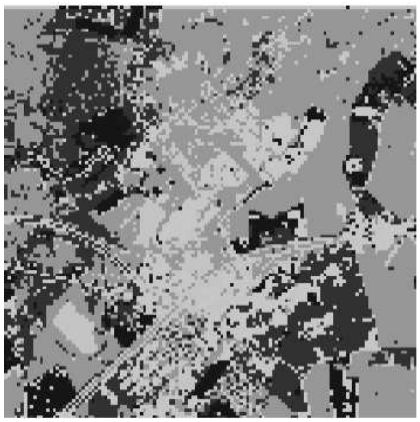
FIGURE 1. Bayesian segmentation of a simulated hyperspectral image with independence and AR assumption on $\{m_{ik}\}$ with 50 iterations.

Real data

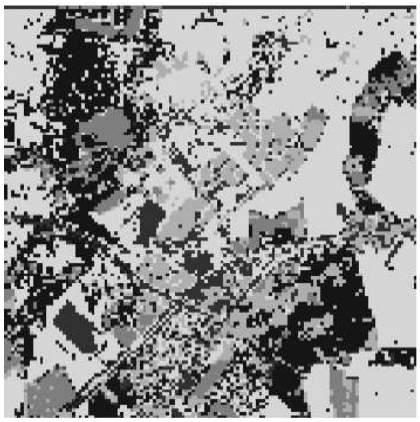
In the next step we applied the proposed method to a part of real data (Aviris). In Figure 2 we illustrate a real example hyperspectral images $\underline{\boldsymbol{g}}$ which are 56, (128×124) images. This figure shows the reconstruction and joint segmentation results i.e. $\hat{\boldsymbol{z}}_1$ with *independence* and $\hat{\boldsymbol{z}}_2$, with *AR* assumption on $\{m_{ik}\}$ after 200 iterations. Unfortunately, in this case we cannot give a quantitative comparing of the results.



$\underline{\boldsymbol{g}}$ and their marginals



$\hat{\boldsymbol{z}}_1$ with *independence* and



$\hat{\boldsymbol{z}}_2$, with *AR* assumption on $\{m_{ik}\}$

FIGURE 2. Bayesian segmentation of a part of real hyperspectral image with dimension $128 \times 128 \times 56$ with independence and AR assumption on $\{m_{ik}\}$ with 200 iterations.

REFERENCES

1. J.R. Jensen, *Introductory digital image processing: A remote sensing perspective. 2nd ed.*, Upper Saddle River, NJ: Prentice Hall, 1995.
2. X. Jia and J.A. Richards, *Remote sensing digital image analysis: An introduction. 3rd ed.*, Berlin, NY: Springer Verlag, 1999.
3. D. Landgrebe, "Hyperspectral image data analysis," *IEEE Transactions on Signal Processing*, vol. 19, pp. 17–28, 2002.
4. G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," *IEEE Transactions on Signal Processing*, vol. 19, pp. 12–16, 2002.
5. C.A. Shah, P. Watanachaturaporn, P.K. Varshney, and M.K. Arora, "Some recent results on hyperspectral image classification," in *Advances in Techniques for Analysis of Remotely Sensed Data, IEEE Workshop on, 27-28 Oct. 2003*, 2003, vol. 19, pp. 346–353.
6. Olivier Féron and Ali Mohammad-Djafari, "Image fusion and joint segmentation using an mcmc algorithm," *Submitted to: JEI*, April 2004.