BAYESIAN SEGMENTATION OF HYPERSPECTRAL IMAGES

Adel Mohammadpour, Nadia Bali and Ali Mohammad-Djafari

Laboratoire des Signaux et Systèmes
CNRS-ESE-UPS
Supélec, Plateau de Moulon
91192 Gif-sur-Yvette, FRANCE.

Adel.Mohammadpour@lss.supelec.fr
Olivier.Feron@lss.supelec.fr
djafari@lss.supelec.fr
Contents

• Introduction to hyperspectral images
• Spectral classification methods
  (Spatial distribution of spectra is neglected)
• Spatial classification methods
  (Spectral shapes of the voxels are neglected)
• Modeling for segmentation methods which account for both spatial and spectral structures of the data
• Bayesian approach and general MCMC Gibbs sampling
• Comparison of the proposed approach with classical methods
• Proposed algorithm
• Simulation results
• Conclusions
Introduction to hyperspectral images
Classification, segmentation and data reduction

- Hyperspectral data: $g(\omega, r)$
  - A set of spectra: $g_r(\omega)$
  - A set of images: $g_\omega(r)$

- Redundancy due to spectral and spatial structure

- Main objectif 1: Find the type of materials in a given position (labeling)
  - Classification
  - Segmentation

- Main objectif 2: Data reduction and compression
  - ACP, ACI and using classification and segmentation for better compression.
  - Proposing a method which does data reduction and segmentation at the same time.
Generating simulated data
Synthetic data 1:

4 classes, 32 images (64x64)
Generating simulated data

Synthetic data 2:

8 classes, 112 images (128x128)
Spectral classification methods

- Each spectral line is considered as a point in a vectorial space
- Different classification methods are used: K-means, mixture of gaussians, ...
- Spatial distribution of the spectra is neglected
112 images

Original

8 classes

112 images

Estimated

8 classes
Spatial classification methods

- Each image is considered as a point in a vectorial space
- Different classification methods are used: K-means, mixture of gaussians, ...
- Spectral structures of the image pixels are neglected
112 images

Original

8 classes

8 images

Estimated

8 classes
Modelling for accounting for both spatial and spectral structures

\[ g_i(r) = f_i(r) + \epsilon_i(r), \quad i = 1, \cdots, M \]
\[ g(r) = \{g_i(r), \ i = 1, M\} \]
\[ g(r) = f(r) + \epsilon(r) \]
\[ g_i = \{g_i(r), \ r \in R\} \]
\[ g = \{g_i(r), \ i = 1, M\} \]
\[ g = f + \epsilon \]

Segmentation:

- Hidden variables rep. regions

\[ z(r) = k, \ k = 1, \cdots, K \]
\[ R_k = \{r : z(r) = k\}, \ R = \cup_k R_k \]
- Homogeneity in regions:

\[ p(f_i(r)|z(r) = k) = \mathcal{N}(m_{i,k}, \sigma_{i,k}^2) \]
Modelling for accounting for both spatial and spectral structures

\[
\begin{align*}
g_i(r) &= f_i(r) + \epsilon_i(r), \quad i = 1, \ldots, M \\
p(f_i(r) | z(r) = k) &= \mathcal{N}(m_{ik}, \sigma_{ik}^2)
\end{align*}
\]

Prior hypothesis about \(f_i(r)\):

- Pixels values of \(f_i(r)\) in different regions of an image are independent. They may share however the same parameters \(\theta_{ik} = (m_{ik}, \sigma_{ik}^2)\)

- For pixels values in a given region of an image, two possibilities:
  - i.i.d.: \(p(f_j(r) | z_j(r) = k) = \mathcal{N}(m_{jk}, \sigma_{jk}^2)\)
    \[p(f_j(r), r \in R_{jk}) = \mathcal{N}(m_{jk} \mathbf{1}, \sigma_{jk}^2 \mathbf{I})\]
  - Markovien: \(p(f_j(r), r \in R_{jk}) = \mathcal{N}(m_{jk} \mathbf{1}, \Sigma_{jk})\)
Modelling for accounting for both spatial and spectral structures

\[
\begin{cases}
g_i(r) = f_i(r) + \epsilon_i(r), & i = 1, \ldots, M \\
p(f_i(r)|z(r) = k) = \mathcal{N}(m_{ik}, \sigma_{ik}^2)
\end{cases}
\]

Pixels along the channels represent spectra

- A Markovien model for \( f_i(r) \):

\[
p(f_i(r)|z(r) = k, f_{i-1}(r)) = \mathcal{N}(\psi_{ik} f_{i-1}(r), \sigma_{ik}^2)
\]

\[
f_{ik}(r) = \psi_{ik} f_{i-1,k}(r) + \eta_k \sim AR(1)
\]

- A Markovien model for the means:

\[
p(f_i(r)|z_i(r) = k) = \mathcal{N}(m_{ik}, \sigma_{ik}^2)
\]

\[
m_{ik} = \phi_k m_{i-1,k} + \eta_k \sim AR(1)
\]
Modeling the labels

\[ p(f_j(r)|z_j(r) = k) = \mathcal{N}(m_{jk}, \sigma_{jk}^2) \longrightarrow p(f_j(r)) = \sum_k P(z_j(r) = k) \mathcal{N}(m_{jk}, \sigma_{jk}^2) \]

- **Independent Gaussian Mixture model (IGM)**, where \( z_j = \{z_j(r), r \in \mathcal{R}\} \) are assumed to be independent and

\[ P(z_j(r) = k) = p_k, \quad \text{with} \quad \sum_k p_k = 1 \quad \text{and} \quad p(z_j) = \prod_k p_k \]

- **Contextual Gaussian Mixture model (CGM):** \( z_j \) Markovien

\[ p(z_j) \propto \exp \left[ \alpha \sum_{r \in \mathcal{R}} \sum_{s \in \mathcal{V}(r)} \delta(z_j(r) - z_j(s)) \right] \]

which is the Potts Markov random field (PMRF).

The parameter \( \alpha \) controls the mean value of the regions’ sizes.
Expressions of likelihood, prior and posterior laws

\[ g_i(r) = f_i(r) + \epsilon_i(r) \]

\[ g_i = f_i + \epsilon_i, \quad i = 1, \cdots, M \quad \rightarrow \quad g = f + \epsilon \]

- Likelihood: \( \theta_1 = \{ \sigma_{\epsilon_i}^2, i = 1, \cdots, M \} \), \( \Sigma_{\epsilon_i} = \sigma_{\epsilon_i}^2 I \)

\[ p(g|f, \theta_1) = \prod_{i=1}^{M} p(g|f, \Sigma_{\epsilon_i}) = \prod_{i=1}^{M} \mathcal{N}(f, \Sigma_{\epsilon_i}) \]

- HMM for the images: \( \theta_2 = \{(m_{ik}, \sigma_{ik}^2), j = 1, \cdots, M\} \)
  - Markovian model for \( f_i|z \):
    \[ p(f|z, \theta_2) = p(f_1|z, m_{ik}, \sigma_{ik}^2) \prod_{i=2}^{M} p(f_i|f_{i-1}, z, m_{ik}, \sigma_{ik}^2) \]
  - Markovian model for \( m_{ik} \):
    \[ p(f|z, \theta_2) = \prod_{i=1}^{M} p(f_i|z, m_{ik}, \sigma_{ik}^2) \]

but \( m_{ik} = \phi_k m_{i-1,k} + \eta_{ik} \sim AR(1) \)
• PMRF for the labels:

\[ p(z) \propto \exp \left[ \sum_{r \in \mathcal{R}} \sum_{s \in \mathcal{V}(r)} \delta(z(r) - z(s)) \right] \]

• Conjugate priors for the hyperparameters \( \theta = (\theta_1, \theta_2) \):

\[ \theta = \{ \{ \sigma_{\varepsilon_i}^2, i = 1, \ldots, M \}, \{(m_{ik}, \sigma_{ik}^2), i = 1, \ldots, M, k = 1, \ldots, K \} \} \]

\[ p(\sigma_{\varepsilon_i}) = \mathcal{IG}(\alpha_{i0}, \beta_{i0}) \]

\[ p(m_{ik}) = \mathcal{N}(\phi_km_{i-1k}, \sigma_{ik}^2) \]

\[ p(\sigma_{ik}^2) = \mathcal{IG}(\alpha_{i0}, \beta_{i0}) \]

\[ p(\Sigma_{ik}) = \mathcal{IW}(\alpha_{i0}, \Lambda_{i0}) \]

• Joint posterior law of \( \underline{f}, z \) and \( \theta \)

\[ p(\underline{f}, z, \theta | g) \propto p(g | \underline{f}, \theta_1) \ p(\underline{f} | z, \theta_2) \ p(z | \alpha) \ p(\theta) \]
**General MCMC sampling scheme**

\[
p(f, z, \theta | g) \propto p(g | f, \theta_1) p(f | z, \theta_2) p(z | \theta_2)
\]

\[
\propto (\prod_i p(g_i | f_i, \theta_1)) p(f | z, \theta_2) p(z | \alpha) p(\theta)
\]

\[
\theta = (\theta_1, \theta_2)
\]

\[
\begin{align*}
\theta_1 &= \{\sigma_{\epsilon_i}^2, i = 1, \ldots, M\} \\
\theta_2 &= \{(m_{ik}, \sigma_{ik}^2), i = 1, \ldots, M, k = 1, \ldots, K\}
\end{align*}
\]

**Gibbs sampling:**

- Generate samples \((f, z, \theta)^{(1)}, \ldots, (f, z, \theta)^{(N)}\) using

  - \(f \sim p(f | g, z, \theta) \propto p(g | f, z, \theta) p(f | z)\)

  - \(z \sim p(z | g, f, \theta) \propto p(g | f, z, \theta) p(f | z) p(z | \alpha)\)

  - \(\theta \sim p(\theta | g, f, z)\)

- Compute any statistics such as mean, median, variance, ...

## Comparison with classical methods

<table>
<thead>
<tr>
<th>Classical</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observation model: no noise</strong>&lt;br&gt;$g_i(r) = f_i(r)$</td>
<td><strong>Observation model: accounts for noise</strong>&lt;br&gt;$g_i(r) = f_i(r) + \epsilon_i(r), \ r \in \mathcal{R}$,</td>
</tr>
<tr>
<td><strong>Ind. Gaussian Mixture model:</strong>&lt;br&gt;$p(f_i(r)</td>
<td>z(r) = k) = \mathcal{N}(m_{ik}, \sigma^2_{ik})$&lt;br&gt;$z(r) \perp z(s), \ s \neq r,$&lt;br&gt;$p(f</td>
</tr>
<tr>
<td><strong>No correlation in diff. channels:</strong>&lt;br&gt;$m_{ik} \perp m_{jk}, \ i \neq j$</td>
<td><strong>Accounts for correlation in diff. channels:</strong>&lt;br&gt;$m_{ik} = \phi_k m_{i-1k} + \eta_{ik}$</td>
</tr>
</tbody>
</table>
DataOriginalKmeans1Kmeans2BFJsegment
DataOriginalKmeans1Kmeans2BFJsegment
DataOriginalKmeans1Kmeans2BFJsegment