

BAYESIAN CLASSIFICATION, SEGMENTATION AND DATA REDUCTION OF HYPERSPECTRAL IMAGES

Ali MOHAMMAD-DJAFARI, Adel MOHAMMADPOUR and Nadia BALI

Laboratoire des Signaux et Systèmes

CNRS-ESE-UPS

Supélec, Plateau de Moulon

91192 Gif-sur-Yvette, FRANCE.

{Adel.Mohammadpour, Nadia.Bali}@lss.supelec.fr

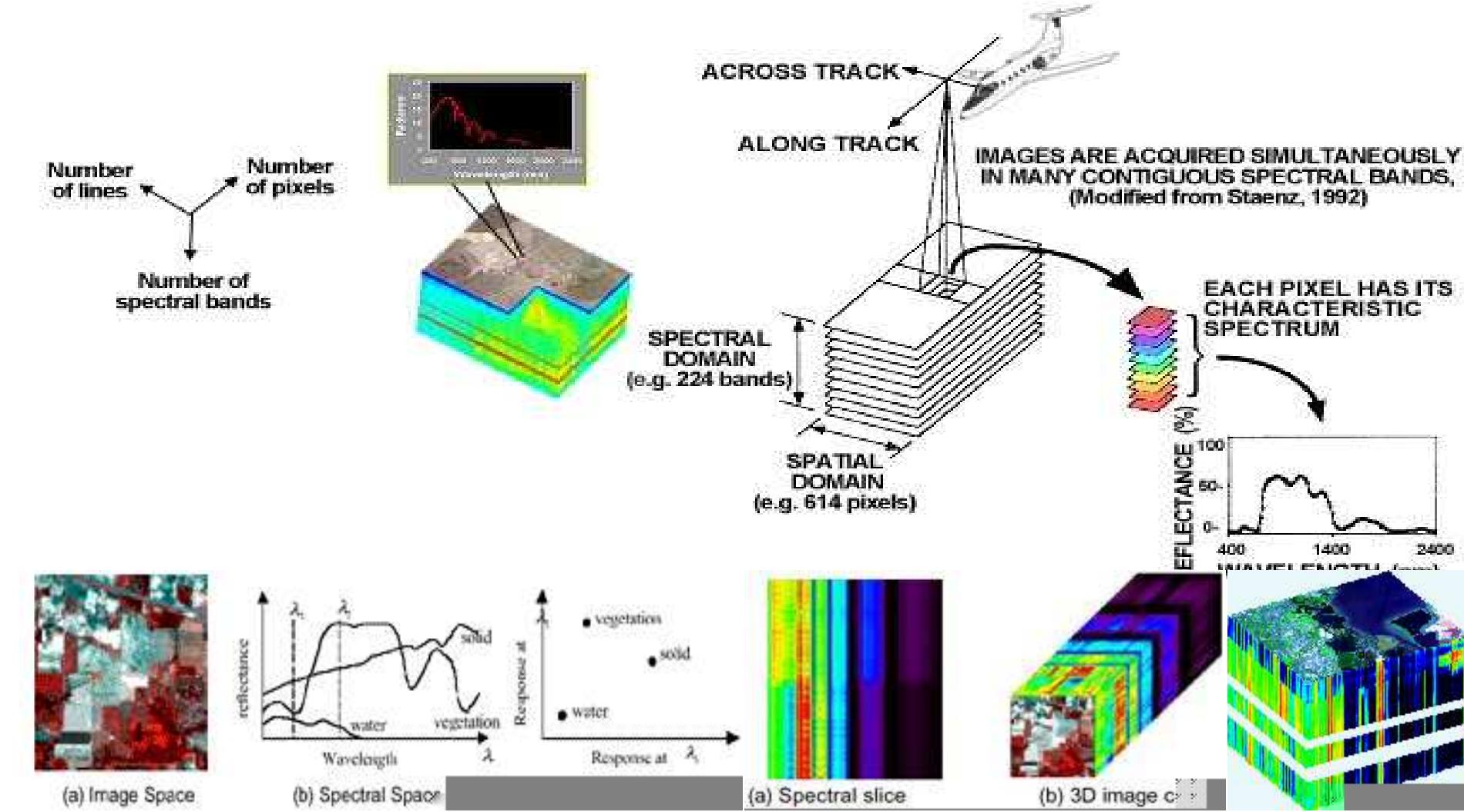
djafari@lss.supelec.fr

<http://djafari.free.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 and perspectives

Introduction to hyperspectral images

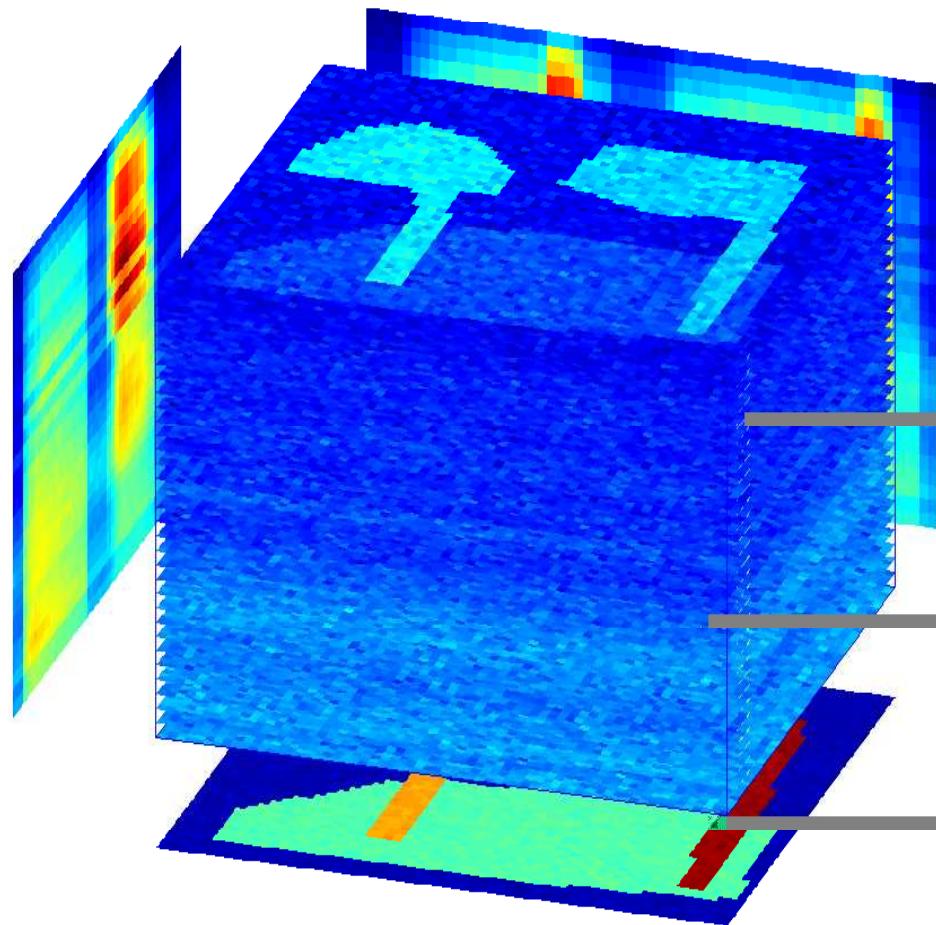
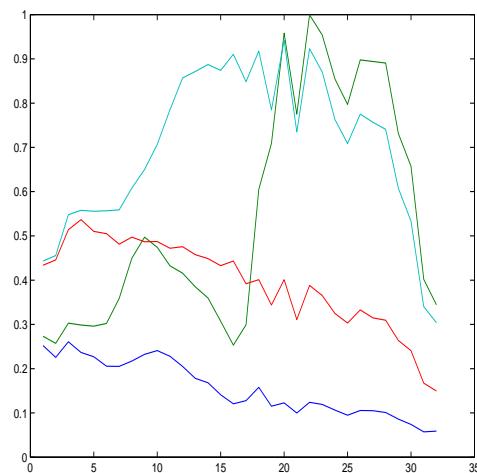
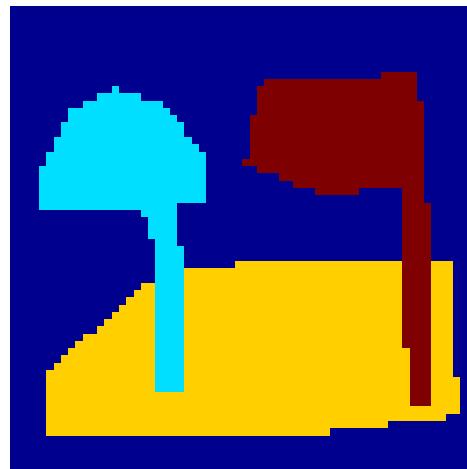


Classification, segmentation and data reduction

- Hyperspectral data : $g(\omega, \mathbf{r})$
 - A set of spectra : $g_r(\omega)$
 - A set of images : $g_\omega(\mathbf{r})$
- 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, classification and segmentation simultaneously and optimally.

Generating simulated data

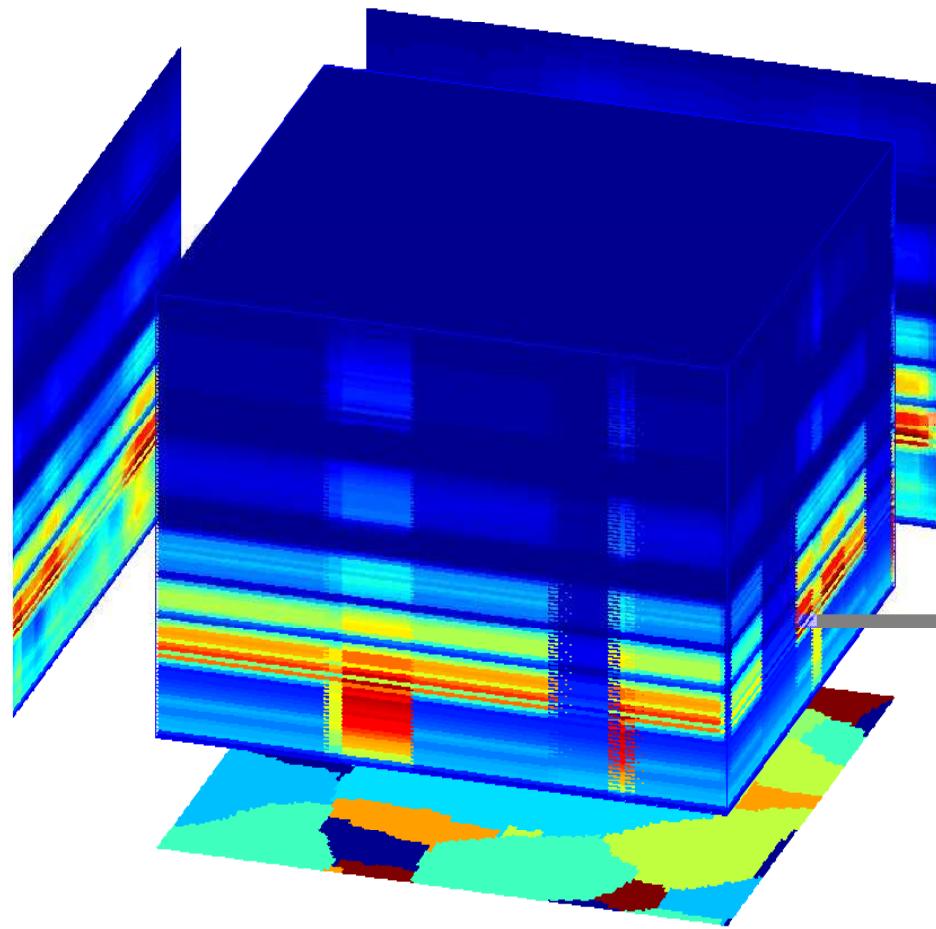
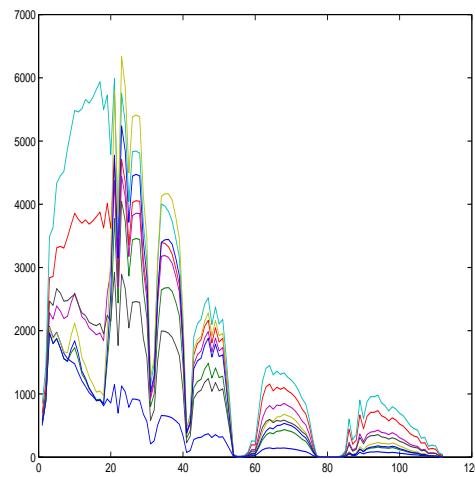
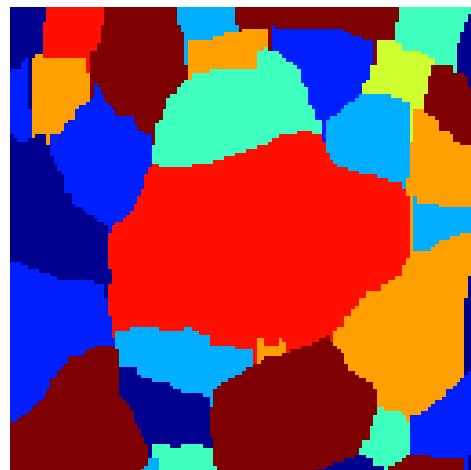
Synthetic data 1:



4 classes, 32 images (64x64)

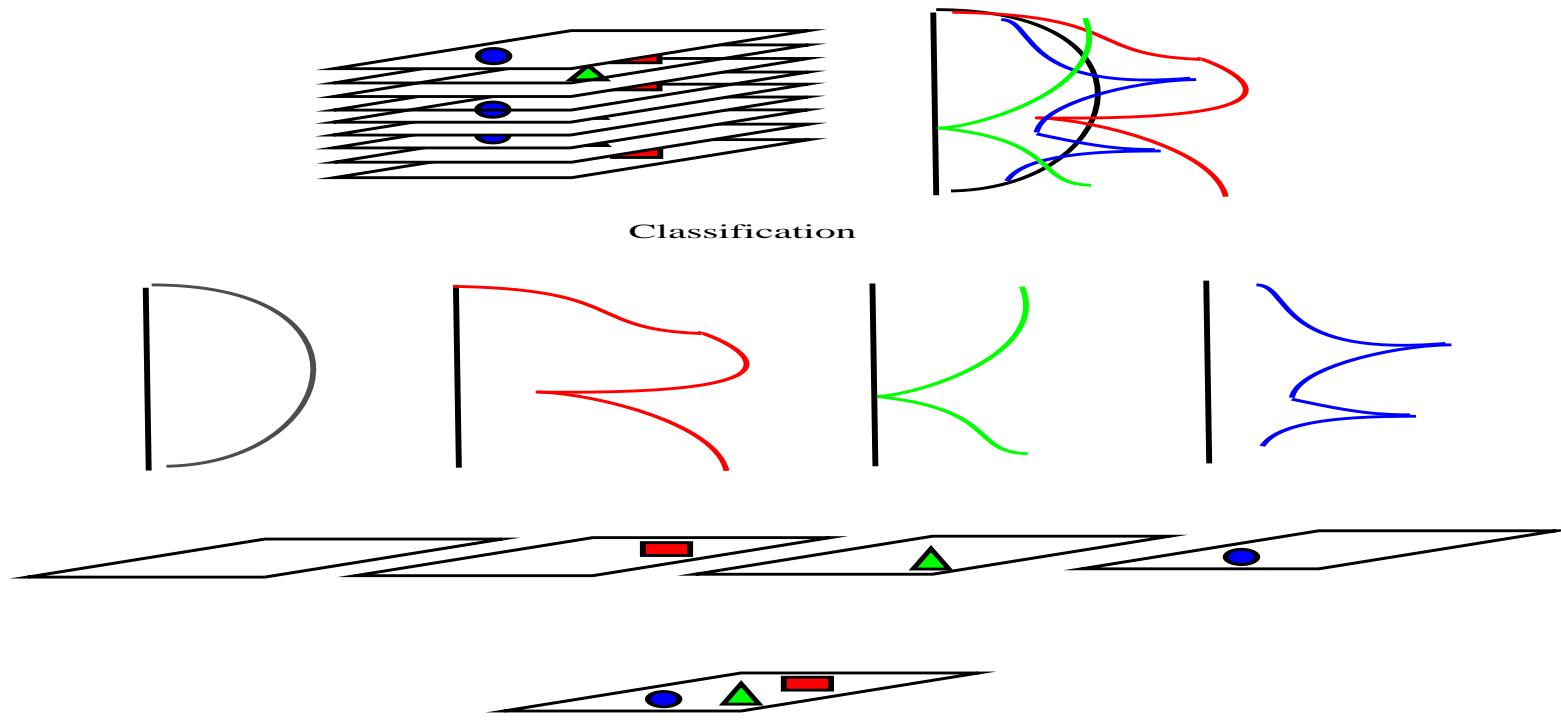
Generating simulated data

Synthetic data 2:

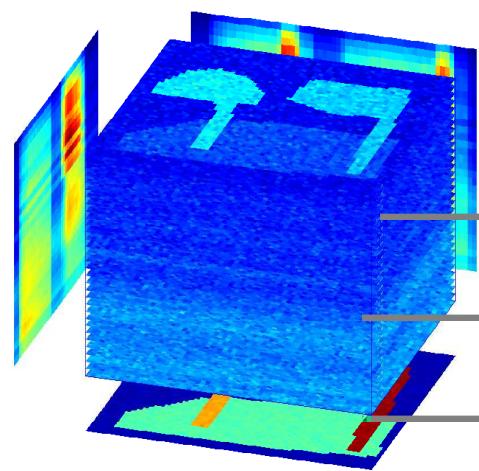


8 classes, 112 images (128x128)

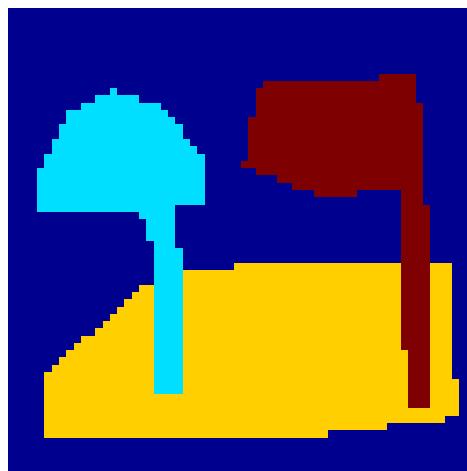
Spectral classification methods



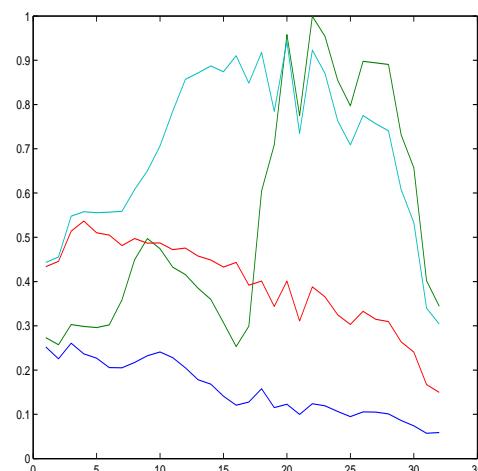
- Each spectral line is considered as a point in a vectorial space
- Different classification methods are used: K-means, mixture of gaussians, ...
- Spatial distribution of the spectra is neglected



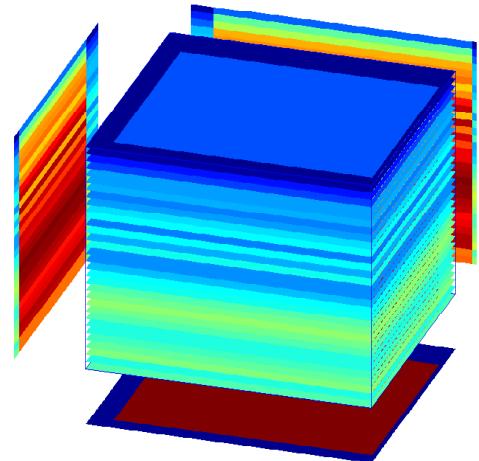
32 images



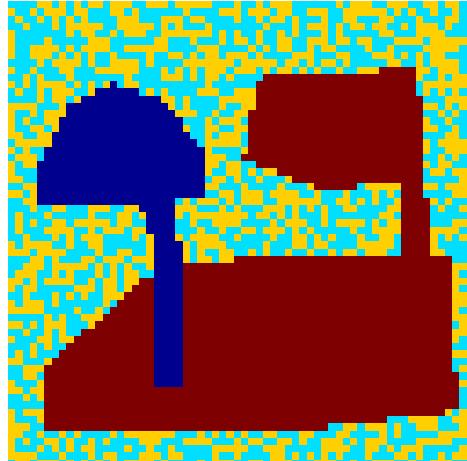
Original



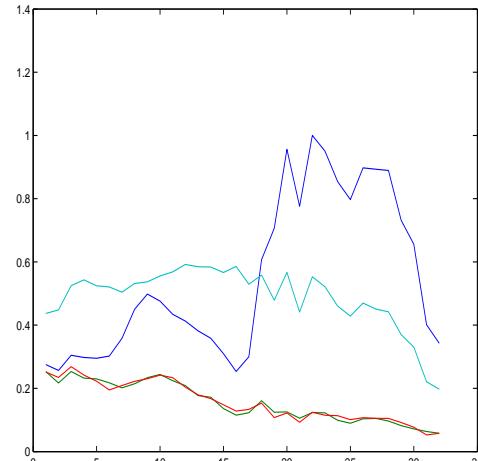
4 classes



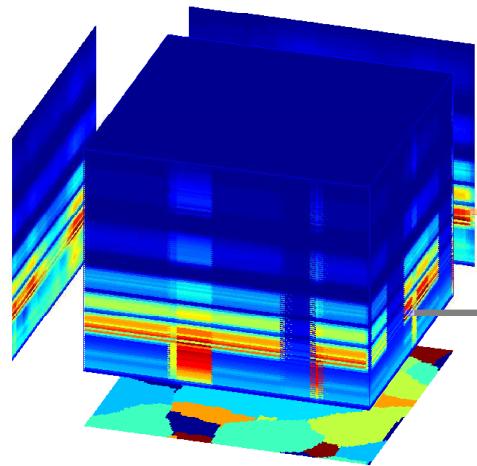
32 images



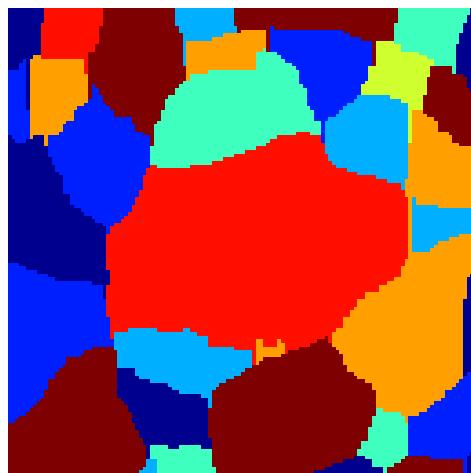
Estimated



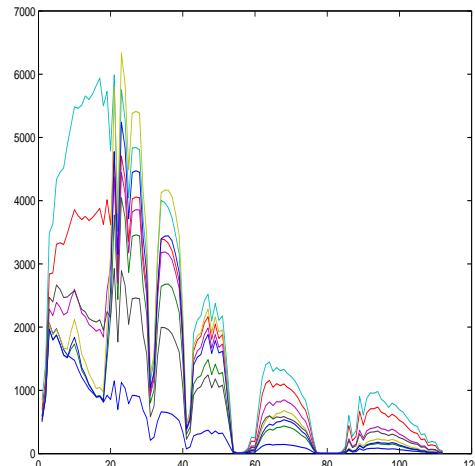
4 classes



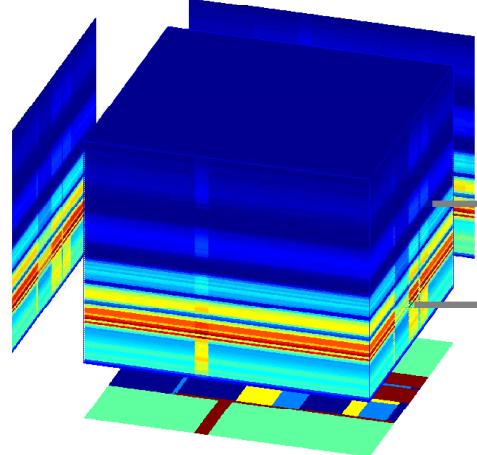
112 images



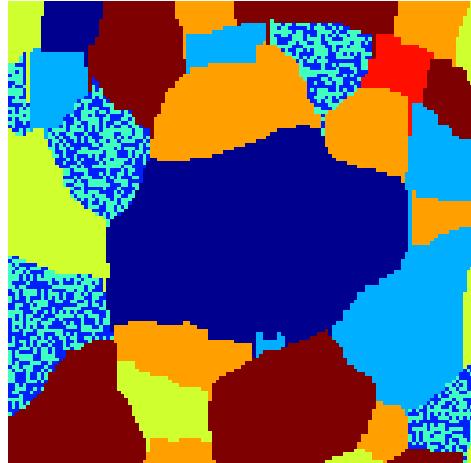
Original



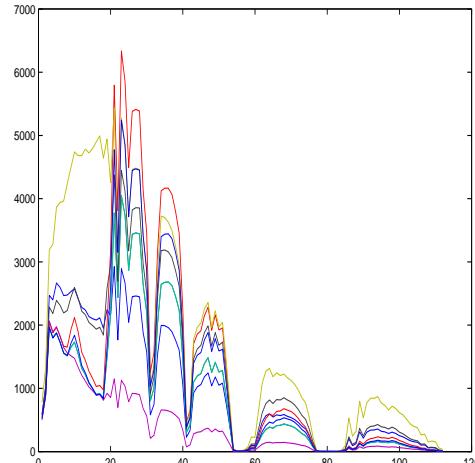
8 classes



112 images

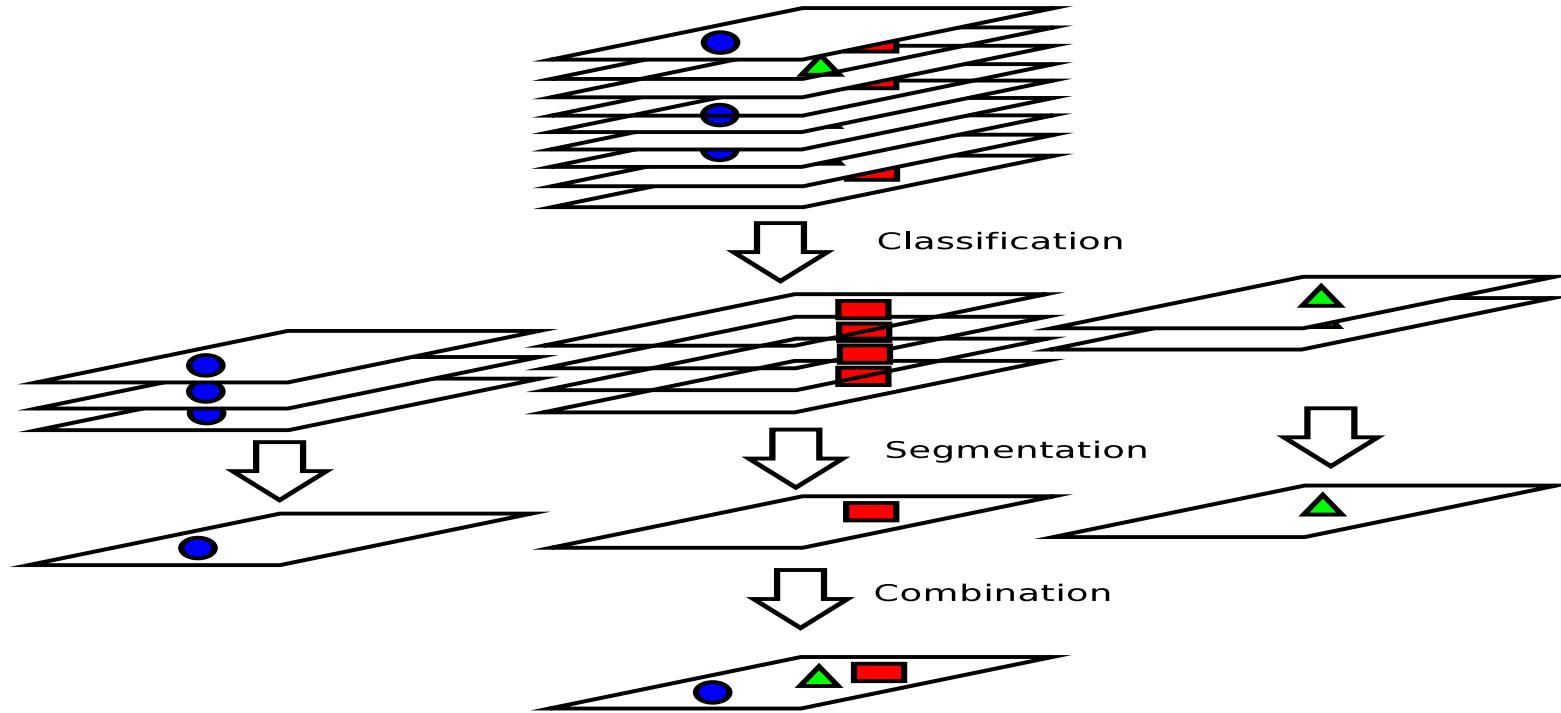


Estimated

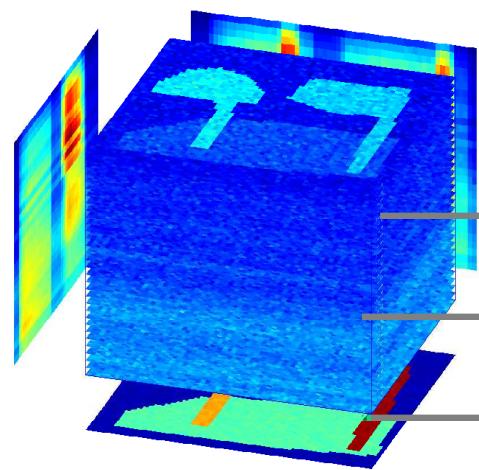


8 classes

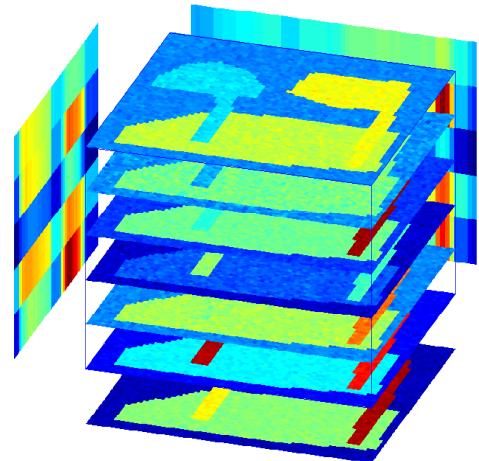
Spatial classification methods



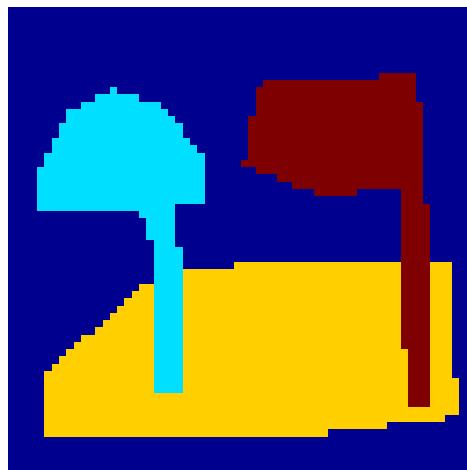
- Each image is considered as a point in a vectorial space
- Different classification methods are used: K-means, mixture of gaussians, ...
- Spectral structures of the image pixels are neglected



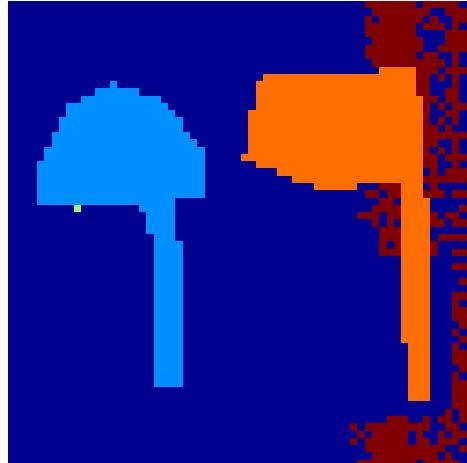
32 images



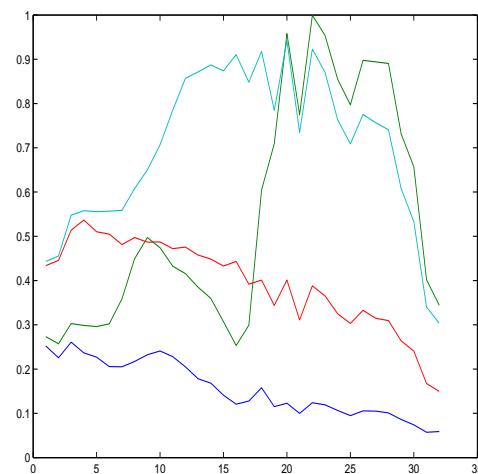
8 images



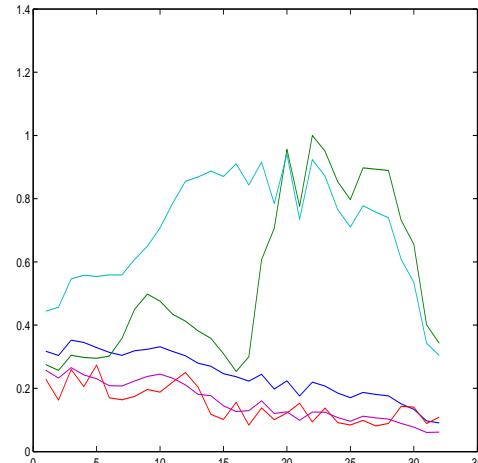
Original



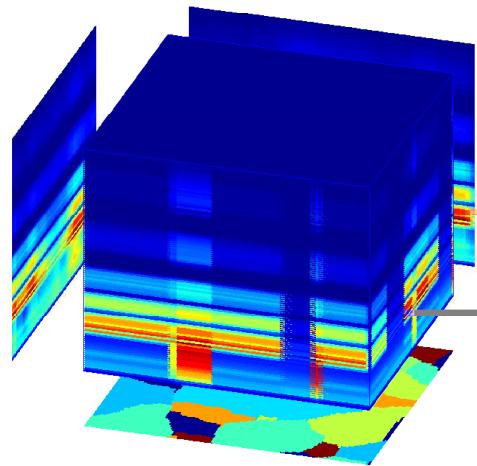
Estimated



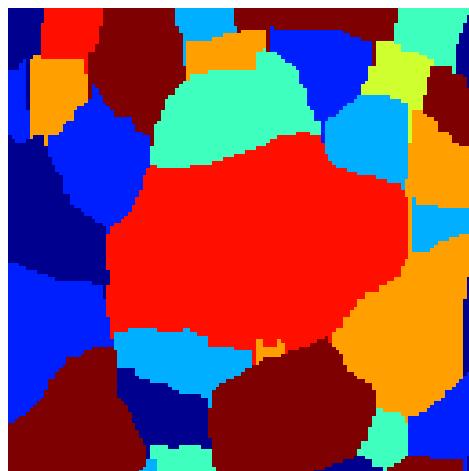
4 classes



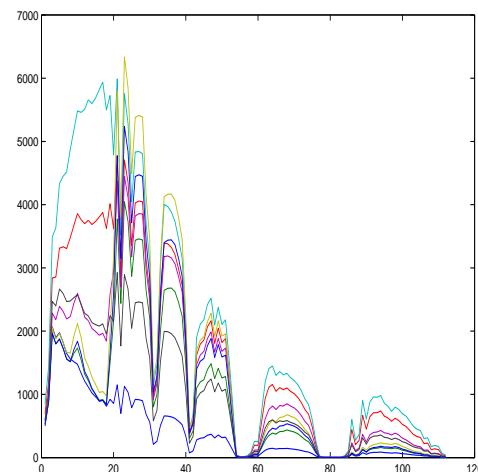
4 classes



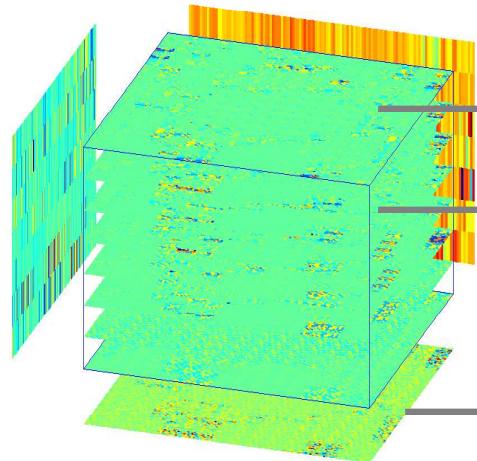
112 images



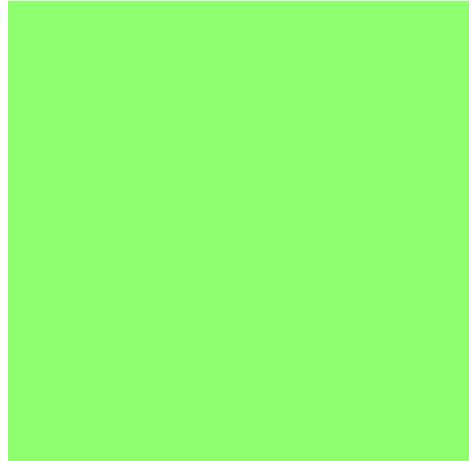
Original



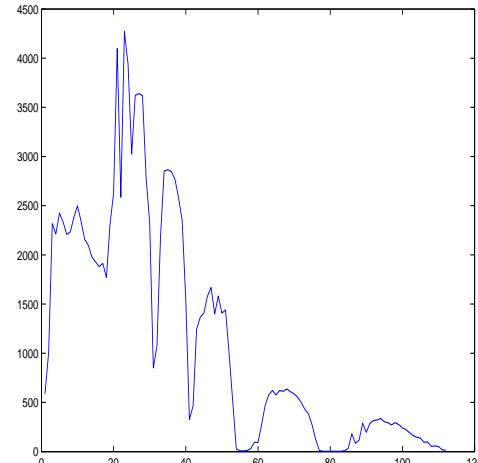
8 classes



8 images

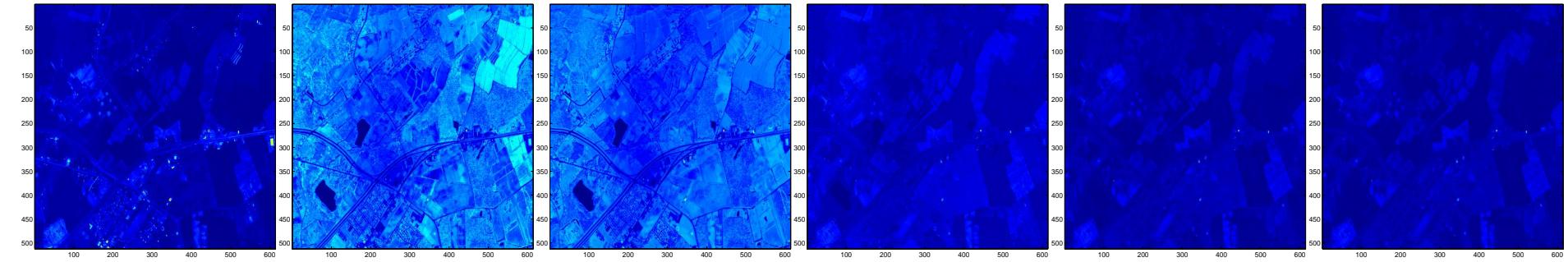


Estimated



8 classes

Modelling for accounting for both spatial and spectral structures



$$g_i(\mathbf{r}) = f_i(\mathbf{r}) + \epsilon_i(\mathbf{r}), \quad i = 1, \dots, M$$

$$\underline{g}(\mathbf{r}) = \{g_i(\mathbf{r}), i = 1, M\}$$

$$\underline{g}(\mathbf{r}) = \underline{f}(\mathbf{r}) + \underline{\epsilon}(\mathbf{r})$$

$$\mathbf{g}_i = \{g_i(\mathbf{r}), \mathbf{r} \in \mathcal{R}\}$$

$$\underline{\mathbf{g}} = \{\mathbf{g}_i(\mathbf{r}), i = 1, M\}$$

$$\underline{\mathbf{g}} = \underline{\mathbf{f}} + \underline{\boldsymbol{\epsilon}}$$

Segmentation:

- Hidden variables rep. regions

$$z(\mathbf{r}) = k, \quad k = 1, \dots, K$$

$$\mathcal{R}_k = \{\mathbf{r} : z(\mathbf{r}) = k\}, \quad \mathcal{R} = \cup_k \mathcal{R}_k$$

- Homogeneity in regions:

$$p(f_i(\mathbf{r}) | z(\mathbf{r}) = k) = \mathcal{N}(m_{ik}, \sigma_{ik}^2)$$

Modelling for accounting for both spatial and spectral structures

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

Prior hypothesis about $f_i(\mathbf{r})$:

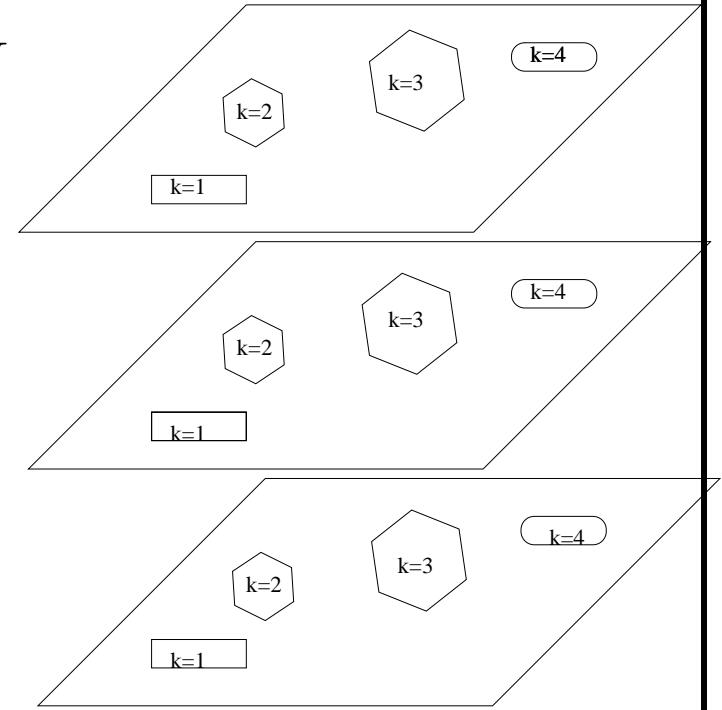
- Pixels values of $f_i(\mathbf{r})$ in different regions of an image are independent.

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.: $p(f_j(\mathbf{r}) | z_j(\mathbf{r}) = k) = \mathcal{N}(m_{jk}, \sigma_{jk}^2)$
 $p(f_j(\mathbf{r}), \mathbf{r} \in R_{jk}) = \mathcal{N}(m_{jk}\mathbf{1}, \sigma_{jk}^2 \mathbf{I})$
- Markovien: $p(f_j(\mathbf{r}), \mathbf{r} \in R_{jk}) = \mathcal{N}(m_{jk}\mathbf{1}, \Sigma_{jk})$



Modelling for accounting for both spatial and spectral structures

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

Pixels along the channels represent spectra

- A Markovien model for $f_i(\mathbf{r})$:

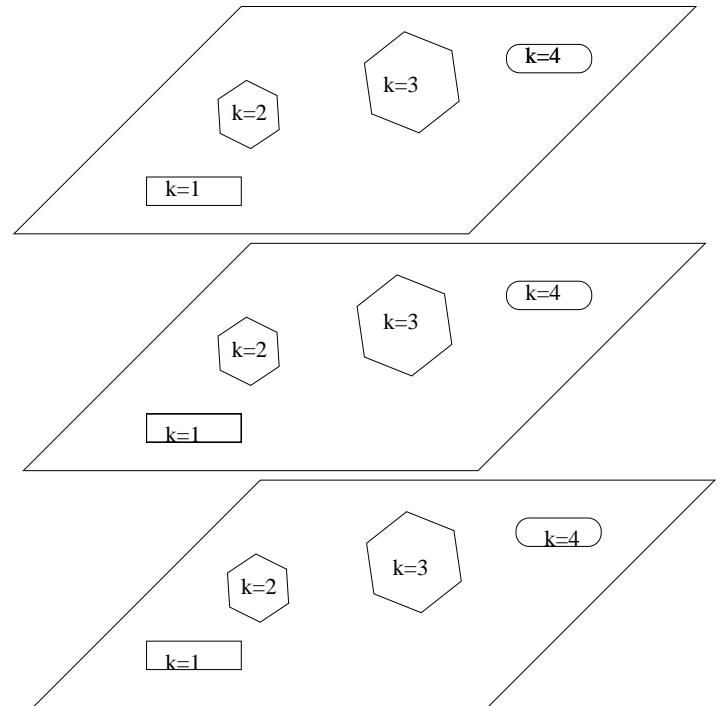
$$p(f_i(\mathbf{r})|z(\mathbf{r}) = k, f_{i-1}(\mathbf{r})) = \mathcal{N}(\psi_{ik}f_{i-1}(\mathbf{r}), \sigma_{ik}^2)$$

$$f_{ik}(\mathbf{r}) = \psi_{ik}f_{i-1,k}(\mathbf{r}) + \eta_{ik} \sim AR(1)$$

- A Markovien model for the means:

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

$$m_{ik} = \phi_k m_{i-1,k} + \eta_k \sim AR(1)$$



Modeling the labels

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

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

$$P(z_j(\mathbf{r}) = k) = p_k, \quad \text{with} \quad \sum_k p_k = 1 \quad \text{and} \quad p(\mathbf{z}_j) = \prod_k p_k$$

- Contextual Gaussian Mixture model (CGM): \mathbf{z}_j Markovien

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

which is the Potts Markov random field (PMRF).

The parameter α controls the mean value of the regions' sizes.

Expressions of likelihood, prior and posterior laws

$$g_i(\mathbf{r}) = f_i(\mathbf{r}) + \epsilon_i(\mathbf{r})$$

$$\mathbf{g}_i = \mathbf{f}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, M \quad \longrightarrow \quad \underline{\mathbf{g}} = \underline{\mathbf{f}} + \underline{\boldsymbol{\epsilon}}$$

- Likelihood: $\boldsymbol{\theta}_1 = \{\sigma_{\epsilon_i}^2, i = 1, \dots, M\}, \quad \longrightarrow \quad \boldsymbol{\Sigma}_{\epsilon_i} = \sigma_{\epsilon_i}^2 \mathbf{I}$

$$p(\underline{\mathbf{g}}|\underline{\mathbf{f}}, \boldsymbol{\theta}_1) = \prod_{i=1}^M p(\underline{\mathbf{g}}_i|\underline{\mathbf{f}}_i, \boldsymbol{\Sigma}_{\epsilon_i}) = \prod_{i=1}^M \mathcal{N}(\underline{\mathbf{g}}_i, \boldsymbol{\Sigma}_{\epsilon_i})$$

- HMM for the images: $\boldsymbol{\theta}_2 = \{(m_{ik}, \sigma_{ik}^2), j = 1, \dots, M\}$

– Markovian model for $\mathbf{f}_i|\mathbf{z}$:

$$p(\underline{\mathbf{f}}|\mathbf{z}, \boldsymbol{\theta}_2) = p(\mathbf{f}_1|\mathbf{z}, m_{ik}, \sigma_{ik}^2) \prod_{i=2}^M p(\mathbf{f}_i|\mathbf{f}_{i-1}, \mathbf{z}, m_{ik}, \sigma_{ik}^2)$$

– Markovian model for m_{ik} :

$$p(\underline{\mathbf{f}}|\mathbf{z}, \boldsymbol{\theta}_2) = \prod_{i=1}^M p(\mathbf{f}_i|\mathbf{z}, m_{ik}, \sigma_{ik}^2)$$

but $m_{ik} = \phi_k m_{i-1,k} + \eta_{ik} \sim AR(1)$

- PMRF for the labels:

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

- Conjugate priors for the hyperparameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$:
 $\boldsymbol{\theta} = \{\{\sigma_{\epsilon_i}^2, i = 1, \dots, M\}, \{(m_{ik}, \sigma_{ik}^2), i = 1, \dots, M, k = 1, \dots, K\}\}$

$$p(\sigma_{\epsilon_i}) = \text{IG}(\alpha_{i0}, \beta_{i0})$$

$$p(m_{ik}) = \mathcal{N}(\phi_k m_{i-1k}, \sigma_{ik0}^2)$$

$$p(\sigma_{ik}^2) = \text{IG}(\alpha_{i0}, \beta_{i0})$$

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

- Joint posterior law of $\underline{\mathbf{f}}$, \mathbf{z} and $\underline{\boldsymbol{\theta}}$

$$p(\underline{\mathbf{f}}, \mathbf{z}, \underline{\boldsymbol{\theta}} | \underline{\mathbf{g}}) \propto p(\underline{\mathbf{g}} | \underline{\mathbf{f}}, \boldsymbol{\theta}_1) p(\underline{\mathbf{f}} | \mathbf{z}, \boldsymbol{\theta}_2) p(\mathbf{z} | \alpha) p(\underline{\boldsymbol{\theta}})$$

General MCMC sampling scheme

$$\begin{aligned} p(\underline{\mathbf{f}}, \mathbf{z}, \underline{\boldsymbol{\theta}} | \underline{\mathbf{g}}) &\propto p(\underline{\mathbf{g}} | \underline{\mathbf{f}}, \boldsymbol{\theta}_1) p(\underline{\mathbf{f}} | \mathbf{z}, \boldsymbol{\theta}_2) p(\mathbf{z} | \boldsymbol{\theta}_2) \\ &\propto (\prod_i p(\mathbf{g}_i | \mathbf{f}_i, \boldsymbol{\theta}_1)) p(\underline{\mathbf{f}} | \mathbf{z}, \boldsymbol{\theta}_2) p(\mathbf{z} | \alpha) p(\underline{\boldsymbol{\theta}}) \end{aligned}$$

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \left\{ \begin{array}{l} \boldsymbol{\theta}_1 = \{\sigma_{\epsilon_i}^2, i = 1, \dots, M\} \\ \boldsymbol{\theta}_2 = \{(m_{ik}, \sigma_{ik}^2), i = 1, \dots, M, k = 1, \dots, K\} \end{array} \right.$$

Gibbs sampling:

- Generate samples $(\underline{\mathbf{f}}, \underline{\mathbf{z}}, \underline{\boldsymbol{\theta}})^{(1)}, \dots, (\underline{\mathbf{f}}, \underline{\mathbf{z}}, \underline{\boldsymbol{\theta}})^{(N)}$ using
 - $\underline{\mathbf{f}} \sim p(\underline{\mathbf{f}} | \underline{\mathbf{g}}, \mathbf{z}, \underline{\boldsymbol{\theta}}) \propto p(\underline{\mathbf{g}} | \underline{\mathbf{f}}, \mathbf{z}, \underline{\boldsymbol{\theta}}) p(\underline{\mathbf{f}} | \mathbf{z})$
 - $\mathbf{z} \sim p(\mathbf{z} | \underline{\mathbf{g}}, \underline{\mathbf{f}}, \underline{\boldsymbol{\theta}}) \propto p(\underline{\mathbf{g}} | \underline{\mathbf{f}}, \mathbf{z}, \underline{\boldsymbol{\theta}}) p(\underline{\mathbf{f}} | \mathbf{z}) p(\mathbf{z} | \alpha)$
 - $\underline{\boldsymbol{\theta}} \sim p(\underline{\boldsymbol{\theta}} | \underline{\mathbf{g}}, \underline{\mathbf{f}}, \mathbf{z})$
- Compute any statistics such as mean, median, variance, ...

Comparison with classical methods

Classical

Observation model: no noise

$$g_i(\mathbf{r}) = f_i(\mathbf{r})$$

Ind. Gaussian Mixture model:

$$p(f_i(\mathbf{r})|z(\mathbf{r}) = k) = \mathcal{N}(m_{ik}, \sigma_{ik}^2)$$

$$z(\mathbf{r}) \perp z(\mathbf{s}), \quad \mathbf{s} \neq \mathbf{r},$$

$$p(\underline{\mathbf{f}}|\mathbf{z}) = \prod_i p(\mathbf{f}_i|\mathbf{z})$$

No correlation in diff. channels:

$$m_{ik} \perp m_{jk}, \quad i \neq j$$

Proposed

Observation model: accounts for noise

$$g_i(\mathbf{r}) = f_i(\mathbf{r}) + \epsilon_i(\mathbf{r}), \quad \mathbf{r} \in \mathcal{R},$$

Hidden Markov Model:

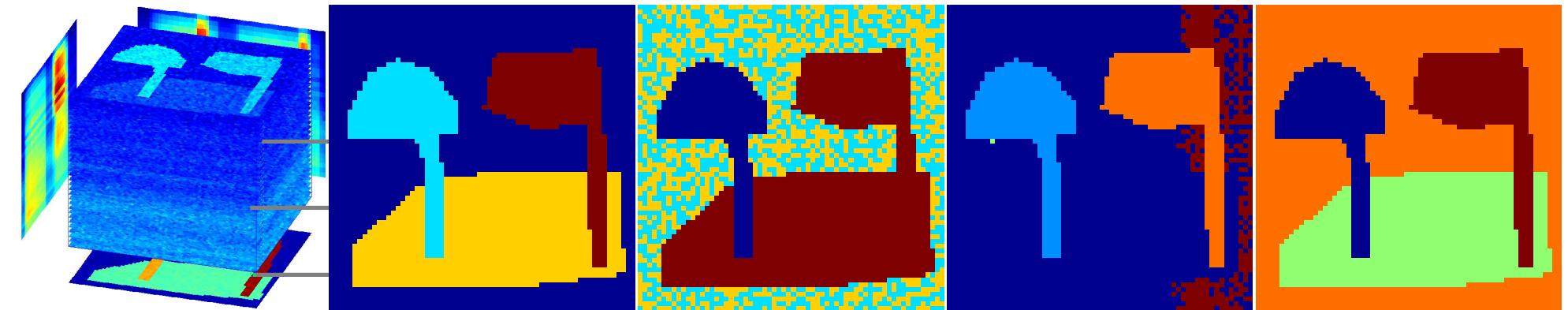
$$p(f_i(\mathbf{r})|z(\mathbf{r}) = k) = \mathcal{N}(m_{ik}, \sigma_{ik}^2)$$

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

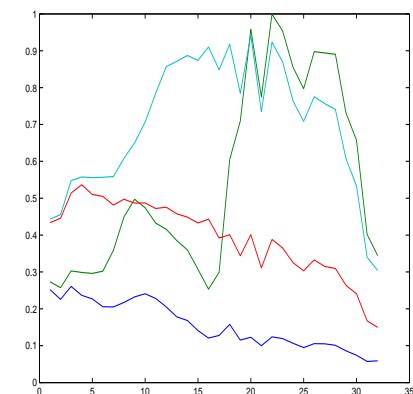
$$p(\underline{\mathbf{f}}|\mathbf{z}) = \prod_i p(\mathbf{f}_i|\mathbf{z})$$

Accounts for correlation in diff. channels:

$$m_{ik} = \phi_k m_{i-1k} + \eta_{ik}$$

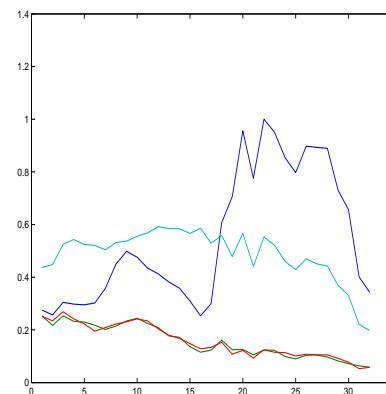


Data

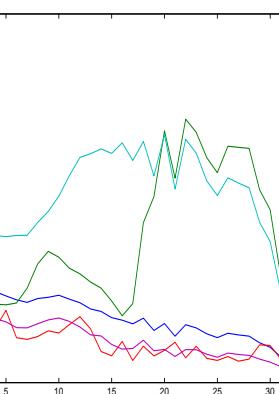


Original

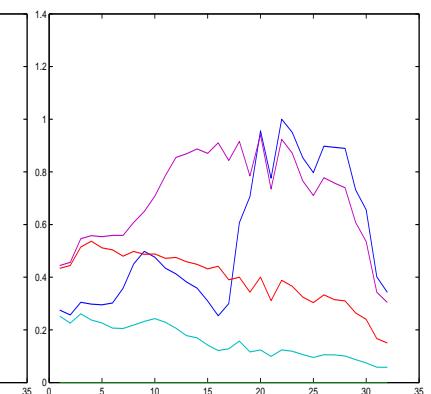
Kmeans1

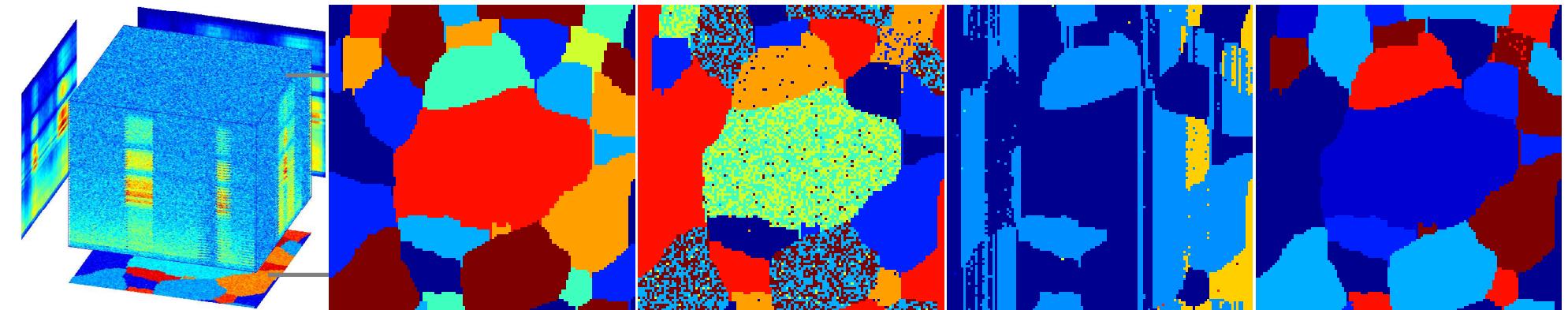


Kmeans2

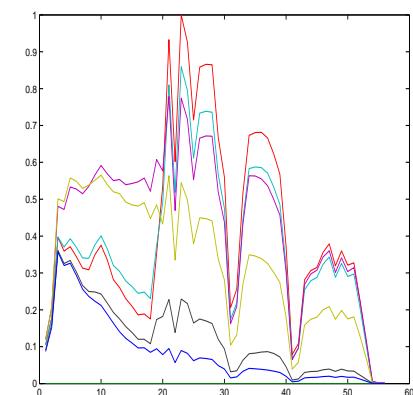


BFJsegment



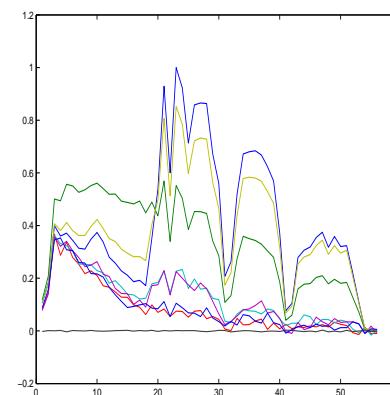


Data

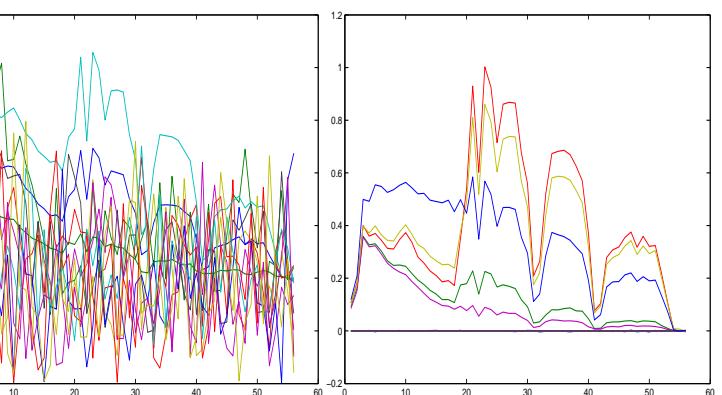


Original

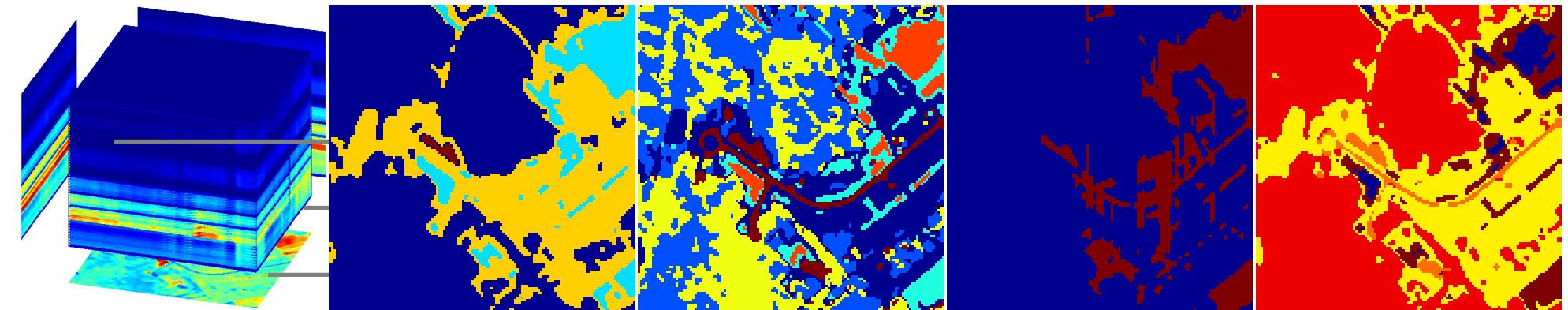
Kmeans1



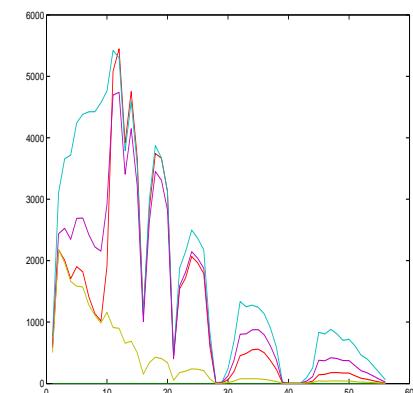
Kmeans2



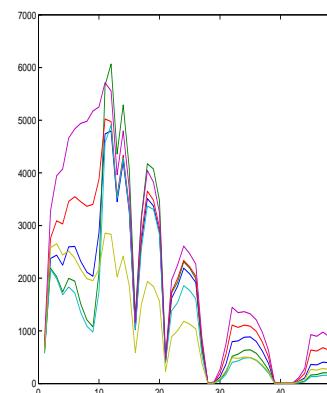
BFJsegment



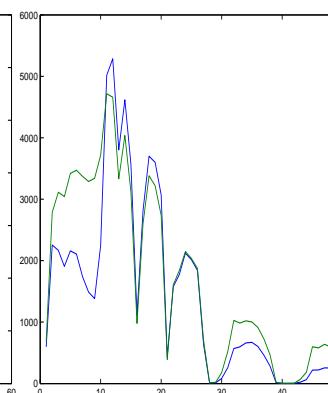
Data



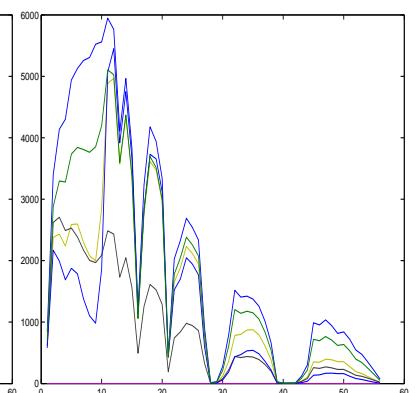
Original



Kmeans1



Kmeans2

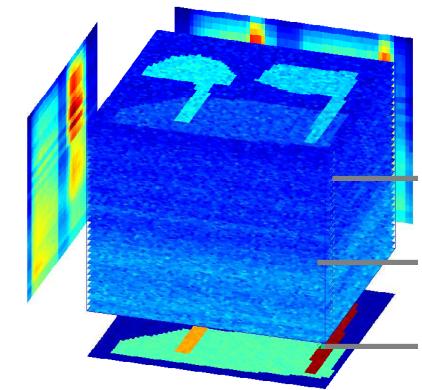
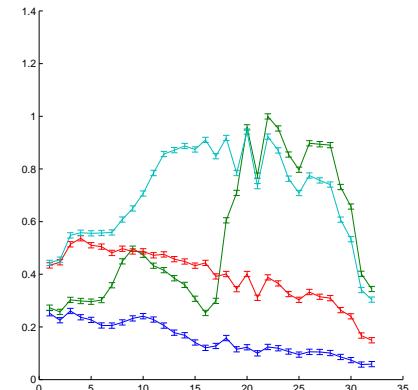
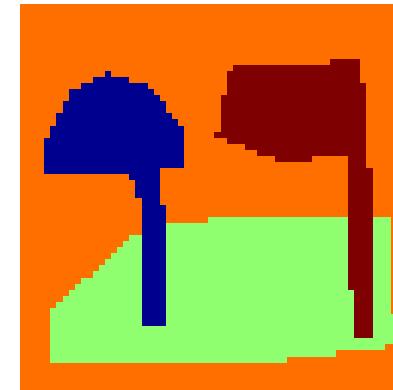
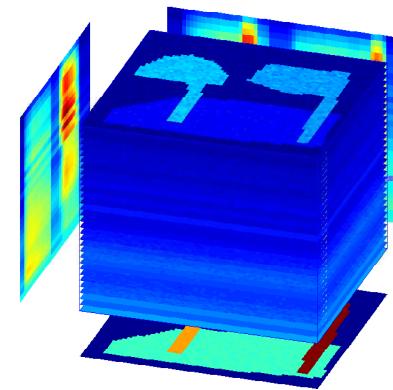
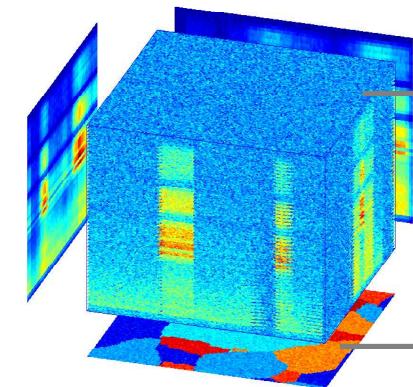
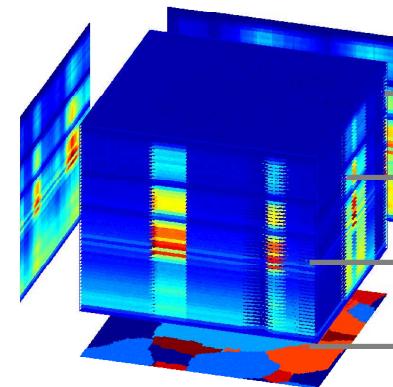
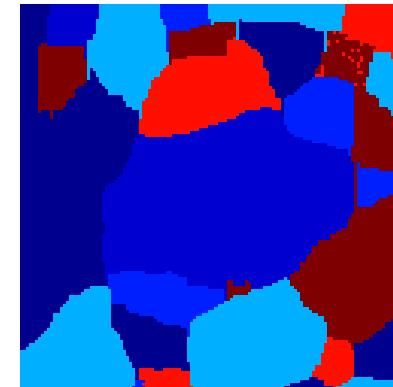
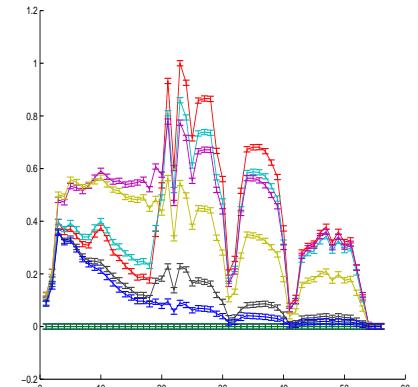


BFJsegment

Conclusions

- Spectra classification methods do not account for their spatial structures.
- Images classification methods do not account for their spectral structures.
- We proposed a classification and segmentation method which accounts both for spectral and spatial dependancies of data in hyperspectral images.
- This unsupervised segmentation results can be used efficiently for data compression
- At the end of this unsupervised segmentation, we have a discrete valued image $z(\mathbf{r})$ representing the segmentation as well as the classification of the spectra in K classes with mean values m_{ik} and their dispersions σ_{ik} .

Outputs of our segmentation algorithm

 \Rightarrow Data $g_i(r)$  \Rightarrow images $f_i(r)$  (m_{ik}, σ_{ik}) 

Perspectives

- We are working for extensions of this method for doing both data reduction and segmentation jointly and optimally. The main idea is in the following modeling:

$$g_i(\mathbf{r}) = \sum_{j=1}^N A_{ij} f_j(\mathbf{r}) + \epsilon_i(\mathbf{r}), \quad i = 1, \dots, M, \quad \text{with } N < M$$

where the sources $f_j(\mathbf{r})$ are the most independent and share the same segmentation $z(\mathbf{r})$.

Prospective results: outputs of our data reduction and segmentation algorithm

