

SPATIAL AND SPECTRAL METHODS FOR THE CLASSIFICATION OF URBAN HYPERSPECTRAL DATA

Mathieu Fauvel

mathieu.fauvel@gipsa-lab.inpg.fr

gipsa-lab/DIS, Grenoble National Polytechnical Institute - INPG - FRANCE
Department of Electrical and Computer Engineering, University of Iceland - ICELAND

INRIA Rhône-Alpes - Team MISTIS

June 10, 2008



Very high resolution urban remote sensing data:

- **Spatial resolution:** 0.75 to 2.5 meter by pixel
- **Spectral resolution:** 1 to more than 200 spectral bands



Panchromatic



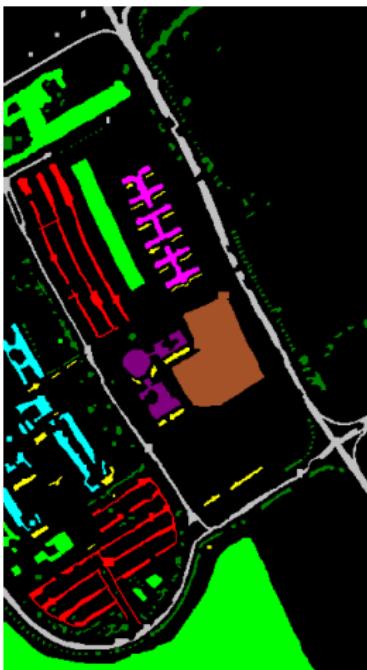
Hyperspectral

Classification: a pattern recognition approach

1. Feature extraction: one vector of attributes extracted for every pixel
2. Pattern recognition algorithms: Maximum Likelihood, Neural Network ...



Original Data



Ground-truth

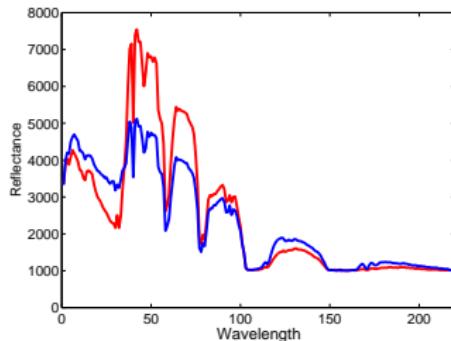


Thematic Map

Data Set: University Area, Pavia, Italy. [610 × 340 × 103], 1.5 m/pixel, 9 classes.

High resolution spectral information:

- ↗ Fine physical description
- ↗ Directly accessible
- ↘ Curse of dimensionality
- ↘ No contextual information



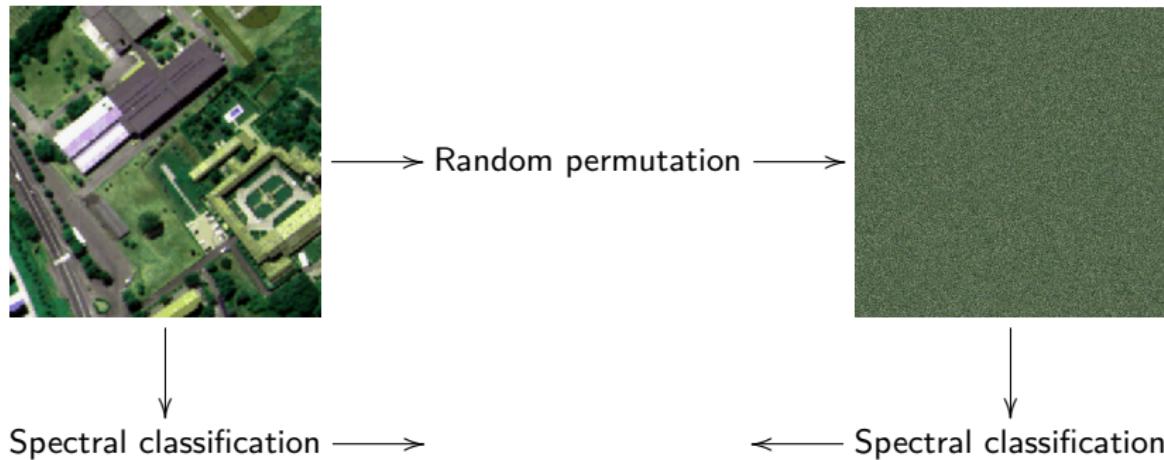
High resolution spatial information:

- ↗ Fine description of structure
- ↘ Not Directly accessible
- ↘ No spectral information



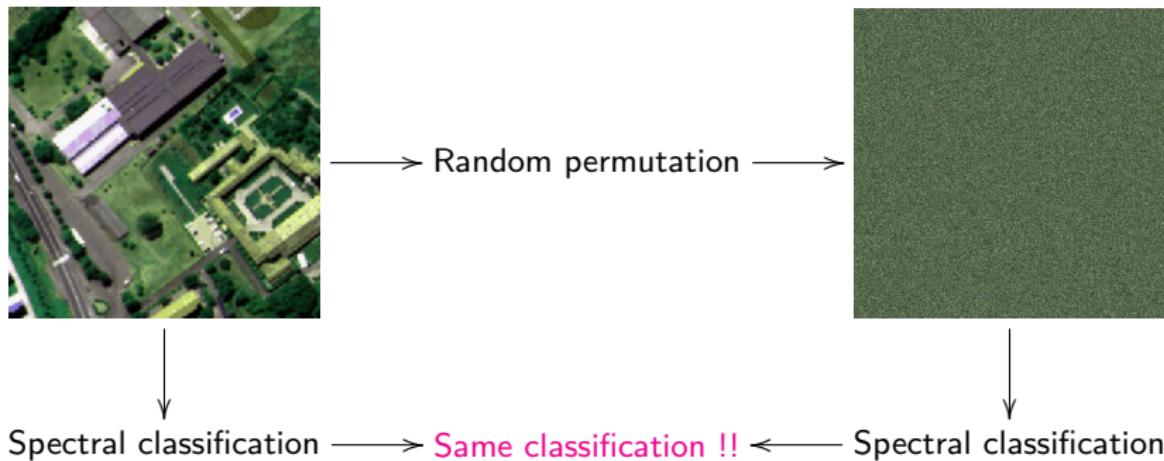
Combine the different types of information for the classification

Why?



Combine the different types of information for the classification

Why?



Need to incorporate information from the spatial domain

Combine the different types of information for the classification

How?

Prior studies: 2 steps approach

1. Feature extraction: Morphological processing
2. Classification: Neural Network, Fuzzy Logic

Contribution:

1. Feature extraction: Extraction of
 - Contextual information (self-complementary filter)
 - Spectral feature (KPCA)
2. Classification:
 - Support Vector Machines
 - Transferability of the hyperplane
3. Data Fusion:
 - at data level
 - at decision level

Combine the different types of information for the classification

How?

Prior studies: 2 steps approach

1. Feature extraction: Morphological processing
2. Classification: Neural Network, Fuzzy Logic

Contribution:

1. Feature extraction: Extraction of
 - Contextual information (self-complementary filter)
 - Spectral feature (KPCA)
2. Classification:
 - Support Vector Machines
 - Transferability of the hyperplane
3. Data Fusion:
 - at data level
 - at decision level

OUTLINE:

Support Vector Machines

Optimal Separating Hyperplane

Hyperparameters Selection

Using spatial information with SVM

Multiple Classifiers - Fusion of SVM

Problem formulation

Fuzzy Fusion: application to SVM

Multisource Markov Fusion

Conclusions and Perspectives

Support Vector Machines

Optimal Separating Hyperplane

Hyperparameters Selection

Using spatial information with SVM

Multiple Classifiers - Fusion of SVM

Problem formulation

Fuzzy Fusion: application to SVM

Multisource Markov Fusion

Conclusions and Perspectives

Support Vector Machines

Optimal Separating Hyperplane

Hyperparameters Selection

Using spatial information with SVM

Multiple Classifiers - Fusion of SVM

Problem formulation

Fuzzy Fusion: application to SVM

Multisource Markov Fusion

Conclusions and Perspectives

Supervised classification

Learn from samples:

- Generated from an unknown $\mathcal{P}(\mathbf{x}, y)$

$$\mathcal{S} = \{(\mathbf{x}^1, y_1), \dots, (\mathbf{x}^\ell, y_\ell)\} \in \mathbb{R}^n \times \{-1; 1\}$$

- Find $f(\mathbf{x})$ minimizing:

$$R(f) = \int_{\mathcal{S}} l(f(\mathbf{x}), y) d\mathcal{P}(\mathbf{x}, y)$$

- $\mathcal{P}(\mathbf{x}, y)$ unknown!!

$$R_{emp}(f) = \frac{1}{\ell} \sum_{i=1}^{\ell} l(f(\mathbf{x}^i), y_i)$$

Convergence? $\ell \rightarrow +\infty$

- $\forall f, R_{emp}(f) \rightarrow R(f)$
- $f_1 \min R_{emp}$ and $f_2 \min R$

$$R_{emp}(f_1) \rightarrow R(f_1) \text{ but } R(f_1) \not\rightarrow R(f_2)$$

Supervised classification

Learn from samples:

- Generated from an unknown $\mathcal{P}(\mathbf{x}, y)$

$$\mathcal{S} = \{(\mathbf{x}^1, y_1), \dots, (\mathbf{x}^\ell, y_\ell)\} \in \mathbb{R}^n \times \{-1; 1\}$$

- Find $f(\mathbf{x})$ minimizing:

$$R(f) = \int_{\mathcal{S}} l(f(\mathbf{x}), y) d\mathcal{P}(\mathbf{x}, y)$$

- $\mathcal{P}(\mathbf{x}, y)$ unknown!!

$$R_{emp}(f) = \frac{1}{\ell} \sum_{i=1}^{\ell} l(f(\mathbf{x}^i), y_i)$$

Convergence? $\ell \rightarrow +\infty$

- $\forall f$, $R_{emp}(f) \rightarrow R(f)$
- $f_1 \min R_{emp}$ and $f_2 \min R$

$$R_{emp}(f_1) \rightarrow R(f_1) \text{ but } R(f_1) \not\rightarrow R(f_2)$$

Restriction on f

Bayesian approach: Distribution *a priori* $P(f)$ over the class \mathcal{F} of function f

Statistical Learning Theory: Control the capacity of \mathcal{F}

VC Dimension h : Number of points that can be learn by a class \mathcal{F} of functions f .

VC Theory: V. Vapnik and A. Chervonenkis¹

$$R(f) \leq R_{emp}(f) + \sqrt{\frac{h(\log(2\ell/h+1)) - \log(\eta/4)}{\ell}}$$

¹V. Vapnik and A. Chervonenkis, *The necessary and sufficient conditions for consistency in the empirical risk minimization method*, Pattern recognition and Image Analysis 1(3):283-305, 1991.

Bayesian approach: Distribution *a priori* $P(f)$ over the class \mathcal{F} of function f

Statistical Learning Theory: Control the capacity of \mathcal{F}

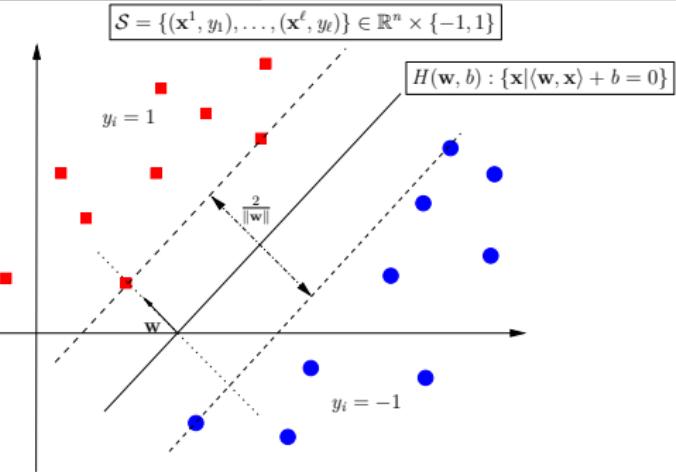
VC Dimension h : Number of points that can be learn by a class \mathcal{F} of functions f .

VC Theory: V. Vapnik and A. Chervonenkis¹

$$R(f) \leq R_{emp}(f) + \sqrt{\frac{h(\log(2\ell/h+1)) - \log(\eta/4)}{\ell}}$$

$$\mathcal{F} = \left\{ f \mid f(\mathbf{x}) = \langle \mathbf{x}, \mathbf{w} \rangle + b \right\} \implies h \leq \|\mathbf{w}\|^2 \mathcal{R}^2 + 1$$

¹V. Vapnik and A. Chervonenkis, *The necessary and sufficient conditions for consistency in the empirical risk minimization method*, Pattern recognition and Image Analysis 1(3):283-305, 1991.



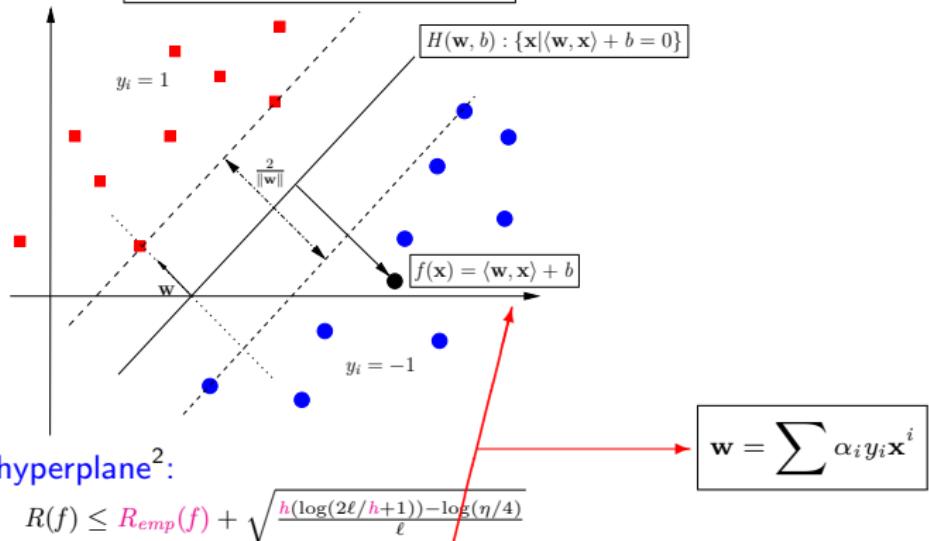
Optimal separating hyperplane²:

$$R(f) \leq R_{emp}(f) + \sqrt{\frac{h(\log(2\ell/h+1)) - \log(\eta/4)}{\ell}}$$

- $\min R_{emp}(f)$: Minimize training errors over \mathcal{S}
- $\min h$: Maximize the margin $\iff \min \|\mathbf{w}\|^2$

² V. Vapnik, *Statistical Learning Theory*, Wiley, New York 1998.

$$\mathcal{S} = \{(\mathbf{x}^1, y_1), \dots, (\mathbf{x}^\ell, y_\ell)\} \in \mathbb{R}^n \times \{-1, 1\}$$



- $\min R_{emp}(f)$: Minimize training errors over \mathcal{S}
- $\min h$: Maximize the margin $\iff \min \|\mathbf{w}\|^2$

Decision: $g(\mathbf{x}) = \text{sgn}(f(\mathbf{x})) = \text{sgn} \left(\sum_{i=1}^{\ell} \alpha_i y_i \langle \mathbf{x}^i, \mathbf{x} \rangle + b \right)$

² V. Vapnik, *Statistical Learning Theory*, Wiley, New York 1998.

Algorithm:

$$\begin{aligned} \text{minimize} \quad & \frac{\langle \mathbf{w}, \mathbf{w} \rangle_{\mathbb{R}^n}}{2} + C \sum_{i=1}^{\ell} \xi_i \\ \text{subject to} \quad & y_i (\langle \mathbf{w}, \mathbf{x}^i \rangle_{\mathbb{R}^n} + b) \geq 1 - \xi_i, \quad \forall i \in 1, \dots, \ell \\ & \xi_i \geq 0, \quad \forall i \in 1, \dots, \ell \end{aligned}$$

Lagrangian: $\alpha_i > 0$ et $\beta_i > 0$ Lagrange coefficients³.

$$L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\langle \mathbf{w}, \mathbf{w} \rangle_{\mathbb{R}^n}}{2} + \sum_{i=1}^{\ell} \alpha_i (1 - \xi_i - y_i (\langle \mathbf{w}, \mathbf{x}^i \rangle + b)) - \sum_{i=1}^{\ell} \beta_i \xi_i + C \sum_{i=1}^{\ell} \xi_i$$

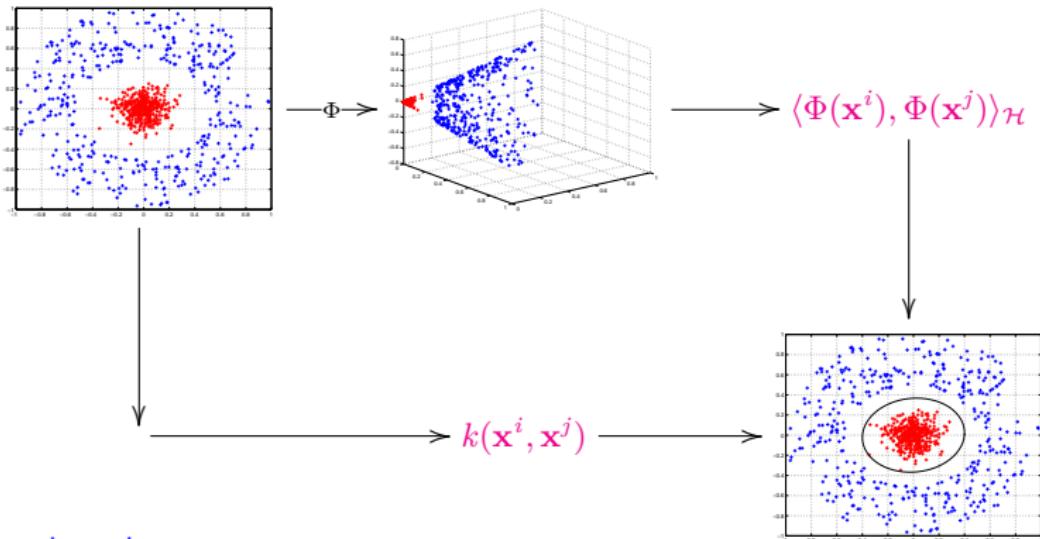
Dual formulation :

$$\begin{aligned} \max_{\boldsymbol{\alpha}} g(\boldsymbol{\alpha}) &= \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} \alpha_i \alpha_j y_i y_j \langle \mathbf{x}^i, \mathbf{x}^j \rangle_{\mathbb{R}^n} \\ \text{subject to} \quad & 0 \leq \alpha_i \leq C \\ & \sum_{i=1}^{\ell} \alpha_i y_i = 0 \end{aligned}$$

³S. Boyd and L. Vandenberghe. *Convex Optimization*, Cambridge University press, 2006.

Kernel methods: Use kernel function k (positive semi-definite)

$$k(\mathbf{x}^i, \mathbf{x}^j) = \langle \Phi(\mathbf{x}^i), \Phi(\mathbf{x}^j) \rangle_{\mathcal{H}}$$



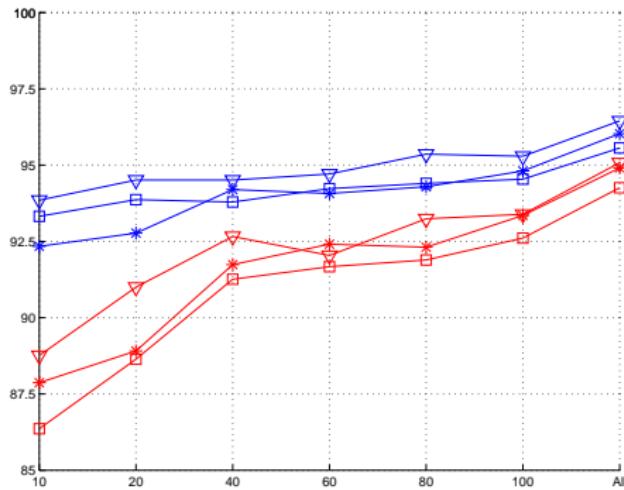
Some kernels:

$$\text{Polynomial kernel: } k(\mathbf{x}^i, \mathbf{x}^j) = (\langle \mathbf{x}^i, \mathbf{x}^j \rangle + q)^p$$

$$\text{Gaussian kernel: } k(\mathbf{x}^i, \mathbf{x}^j) = \exp(-\gamma \|\mathbf{x}^i - \mathbf{x}^j\|^2)$$

$$\text{Spectral kernel: } k(\mathbf{x}^i, \mathbf{x}^j) = \exp\left(-\gamma \arccos\left(\frac{\langle \mathbf{x}^i, \mathbf{x}^j \rangle}{\|\mathbf{x}^i\| \|\mathbf{x}^j\|}\right)^2\right)$$

Training set: 10 to 100 samples by class. Testing set: 103.504 samples.



∇ Gaussian kernel, * polynomial kernel and \square spectral kernel⁴.
 Average accuracy | Overall accuracy

| | Gaussian Kernel | | | |
|--------|-----------------|-----------|-------------|------------|
| | SVM (90) | SVM (900) | SVM (5.536) | GML(5.536) |
| OA (%) | 93,8 | 95,3 | 96,5 | 93,8 |
| AA (%) | 88,8 | 93,4 | 95,1 | 90,8 |

⁴ M. Fauvel, J. Chanussot and J. A. Benediktsson *Classification of hyperspectral remote sensing images with support vectors machines*, in IEEE-ICASSP-06, May 2006, Toulouse.

Support Vector Machines

Optimal Separating Hyperplane

Hyperparameters Selection

Using spatial information with SVM

Multiple Classifiers - Fusion of SVM

Problem formulation

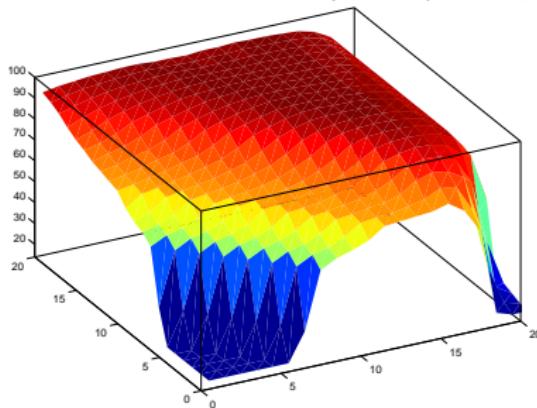
Fuzzy Fusion: application to SVM

Multisource Markov Fusion

Conclusions and Perspectives

Hyperparameters setting: Gaussian kernel $\Rightarrow \mathbf{p} = [\gamma, C]$

- Cross-validation: $-1 \leq \mathbf{x} \leq 1 \Rightarrow C \gg 1 (\approx 200)$ and $\gamma \in \{0.5, 1, 2, 4\}$



- Error bound Σ^5 : $R(f) \leq \Sigma(\mathbf{p})$

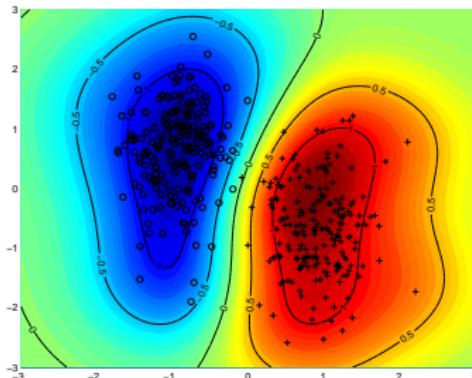
$$\text{Radius-Margin Bound} - \Sigma(\mathbf{p}) := \|\mathbf{w}\|^2 \mathcal{R}^2$$

$$\mathbf{grad}(\Sigma(\mathbf{p})) = \left[\begin{array}{c} \frac{\partial \Sigma}{\partial C} \\ \frac{\partial \Sigma}{\partial \gamma} \end{array} \right] = \left[\begin{array}{c} \frac{\partial \|\mathbf{w}\|^2}{\partial C} \mathcal{R}^2 + \frac{\partial \mathcal{R}^2}{\partial C} \|\mathbf{w}\|^2 \\ \frac{\partial \|\mathbf{w}\|^2}{\partial \gamma} \mathcal{R}^2 + \frac{\partial \mathcal{R}^2}{\partial \gamma} \|\mathbf{w}\|^2 \end{array} \right]$$

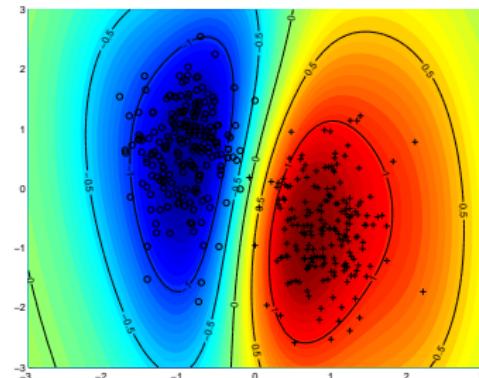
⁵ O. Chapelle, V. Vapnik, O. Bousquet and S. Mukherjee, *Choosing Multiple Parameters for Support Vector Machines*, Machine Learning 1(46):131-159, 2002.

Multiple hyperparameters setting: Gaussian kernel with adapted bandwidth

$$k(\mathbf{x}^i, \mathbf{x}^j) = \exp \left(- \sum_{k=1}^n \gamma_k (x_k^i - x_k^j)^2 \right)$$



$$\dim(\gamma) = 1$$



$$\dim(\gamma) = 2$$

University data set:

| Model selection | CV | Gradient | |
|-----------------|-------|--------------------|----------------------|
| | | $\dim(\gamma) = 1$ | $\dim(\gamma) = 103$ |
| OA(%) | 95.58 | 94.19 | 93.87 |
| N° Optim. | 30 | 15 | 24 |

Support Vector Machines

Optimal Separating Hyperplane

Hyperparameters Selection

Using spatial information with SVM

Multiple Classifiers - Fusion of SVM

Problem formulation

Fuzzy Fusion: application to SVM

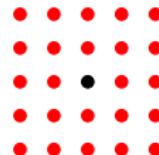
Multisource Markov Fusion

Conclusions and Perspectives

Markov Random Field [Lafarge-05]: fixed neighborhood (cliques)



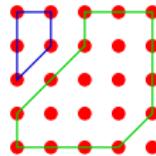
Contextual features [Camps-Valls-06; Bruzzone-06]: fixed neighborhood ($p \times p$ square)



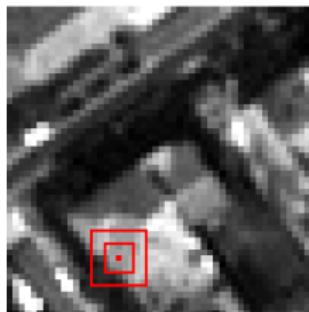
Texture features [Mercier-06] : two 1-D wavelets (on x and y)



Morphological processing [Benediktsson-03]: structures



Not well suited to urban area data: discontinuity (no stationarity)



Morphological Neighborhood: adapt the neighborhood to the structures

- Previous works: Granulometry with geodesic filters (Morphological profile and its derivative).

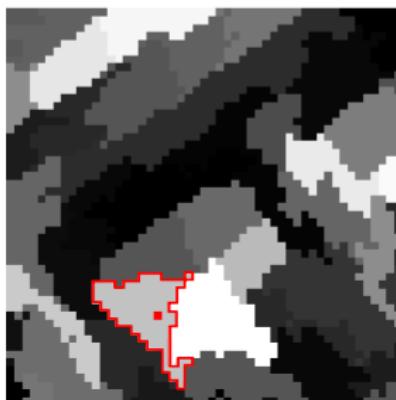
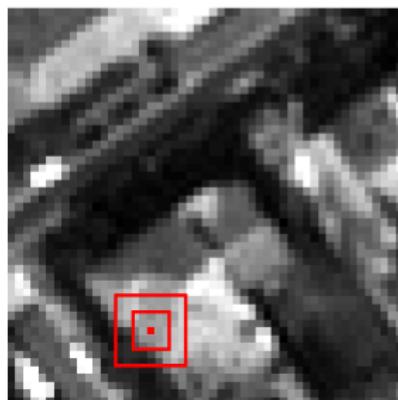
Acts only on extrema structures

- Proposed works: Self-complementary filters.

Acts on structures whatever their gray-level

Self-complementary area filter⁶ : Remove all the structures that are smaller than a given area threshold

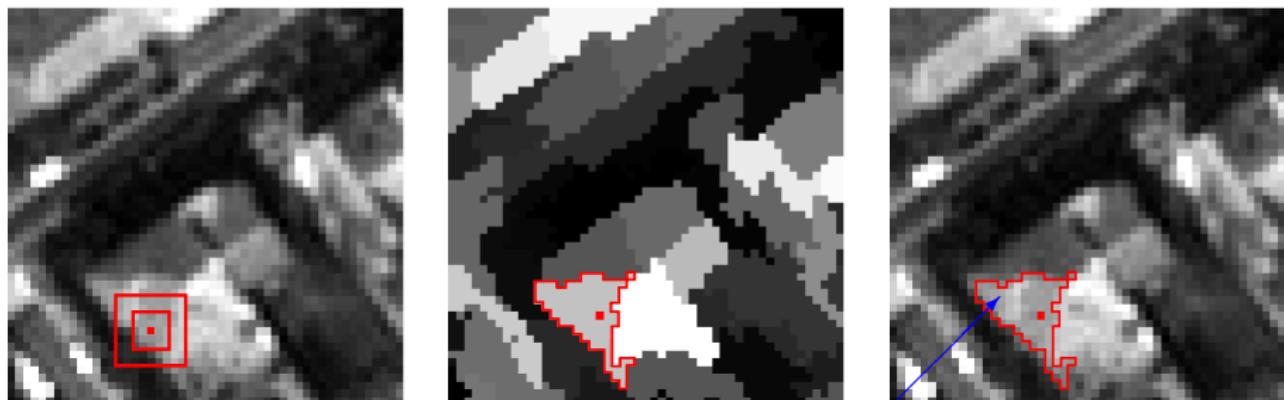
1. Labelling all the flat zones that satisfy the area criterion λ
2. Growing the labelled flat zones until an image partition is reached



⁶P. Soille, *Beyond self-duality in morphological image analysis*, Image and Vision Computing 2:249-257, 2005

Self-complementary area filter⁶ : Remove all the structures that are smaller than a given area threshold

1. Labelling all the flat zones that satisfy the area criterion λ
2. Growing the labelled flat zones until an image partition is reached



Extracting the inter-pixel dependency :

$$\Upsilon_x = \text{med}(\Omega_x)$$

⁶P. Soille, *Beyond self-duality in morphological image analysis*, Image and Vision Computing 2:249-257, 2005

How to use conjointly the spatial and the spectral information?

- Kernel approach: mixture of kernels
- SVM classifier

Combination of kernels: Spatio-spectral kernel (SS kernel)

$$K = \mu k^{spect} + (1 - \mu)k^{spat}$$

- k^{spect} acts on the spectral information
- k^{spat} acts on the spatial information
- μ control the amount of each type of information

Extension to hyperspectral data?

Ordering relation for pixels is needed: does not exist.

- Marginal ordering \Rightarrow by band filtering
- Total Pre-ordering $\Rightarrow h : \mathbb{R}^n \rightarrow \mathbb{R}^1$

Our approach: Dimension reduction with PCA⁷.

1. Projection on the first PC
2. Area filtering and generalization of the neighborhood mask for each band
3. Spatial feature extraction
4. Classification with SS SVM

$$\begin{aligned} h : \mathbb{R}^n &\rightarrow \mathbb{R}^1 \\ \mathbf{x} &\mapsto x = \langle \mathbf{x}, \mathbf{v}_p^1 \rangle_{\mathbb{R}^n} \end{aligned}$$

⁷ M. Fauvel, J. Chanussot and J. A. Benediktsson *Adaptive pixel neighborhood definition for the classification of hyperspectral images with support vector machines and composite kernels*, in IEEE-ICIP-08, October 2008, San Diego.

Classification using spectral and spatial information:



Spectral



EMP



SS SVM

Classes: asphalt, meadow, gravel, tree, metal sheet, bare soil, bitumen, brick and shadow.

Classification accuracies:

| % | SVM | EMP | SS SVM |
|-------------------|-----------|-----------|-----------|
| OVERALL ACCURACY | 80 | 80 | 86 |
| AVERAGE ACCURACY | 88 | 85 | 92 |
| KAPPA COEFFICIENT | 75 | 74 | 82 |
| ASPHALT | 80 | 93 | 84 |
| MEADOW | 68 | 73 | 78 |
| GRAVEL | 74 | 52 | 84 |
| BARE SOIL | 95 | 62 | 95 |

Classification accuracies:

| % | SVM | EMP | SS SVM |
|-------------------|-----------|-----------|-----------|
| OVERALL ACCURACY | 80 | 80 | 86 |
| AVERAGE ACCURACY | 88 | 85 | 92 |
| KAPPA COEFFICIENT | 75 | 74 | 82 |
| ASPHALT | 80 | 93 | 84 |
| MEADOW | 68 | 73 | 78 |
| GRAVEL | 74 | 52 | 84 |
| BARE SOIL | 95 | 62 | 95 |



Data Fusion

Spatial information is helpful

Benefits

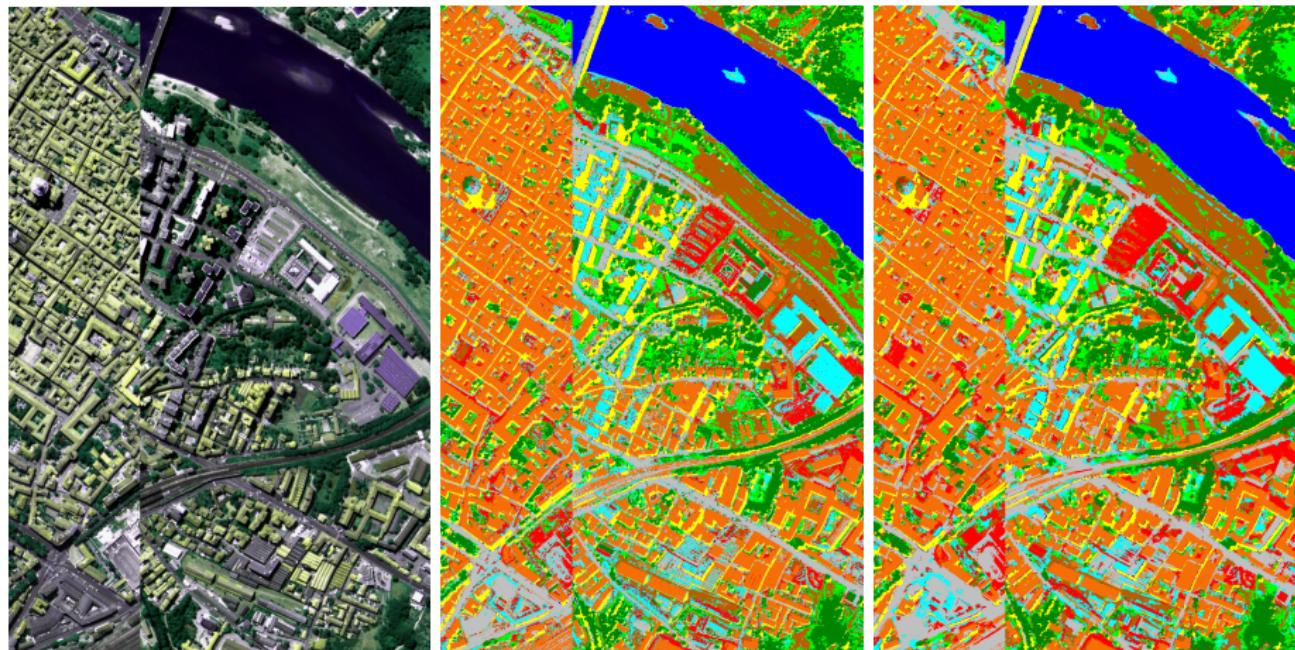
- Adaptive neighborhood definition
- Kernel formulation
- Performance for peri-urban area

Drawbacks

- Parameters λ (area threshold) and μ (weight in SS kernel)
- Statistical spatial information only
- Extension to multi- or hyperspectral data

Some results

Pavia Center [1096 × 610 × 102]



Original

SVM (98.06%)

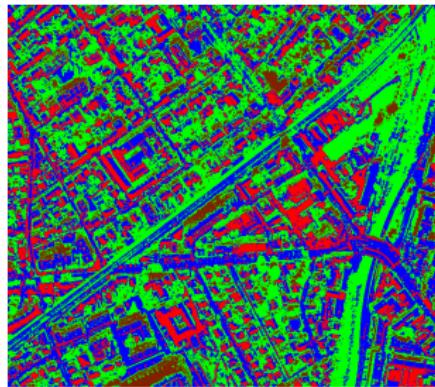
SS SVM(98.43%)

Some results

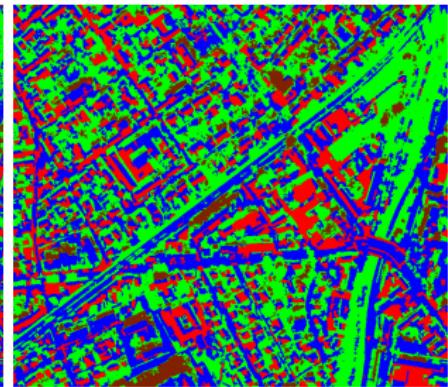
Toulouse [886 × 780 × 1]



Original



SVM (76.99%)



SS SVM (82.25%)

Support Vector Machines
Optimal Separating Hyperplane
Hyperparameters Selection
Using spatial information with SVM

Multiple Classifiers - Fusion of SVM
Problem formulation
Fuzzy Fusion: application to SVM
Multisource Markov Fusion

Conclusions and Perspectives

Support Vector Machines

Optimal Separating Hyperplane
Hyperparameters Selection
Using spatial information with SVM

Multiple Classifiers - Fusion of SVM

Problem formulation
Fuzzy Fusion: application to SVM
Multisource Markov Fusion

Conclusions and Perspectives

Different classifiers:

- **Statistic Theory**: Gaussian Maximum Likelihood, Fisher Discriminant Analysis...
- **Machine learning Theory**: Neural Network, Support Vector Machines...
- **Fuzzy Set Theory**: Fuzzy NN, Fuzzy model...

Different sources:

- Panchromatic and multispectral data
- Multi-valued data
- Multi-temporal data
- Extracted data: texture or geometrical features

Different classifiers:

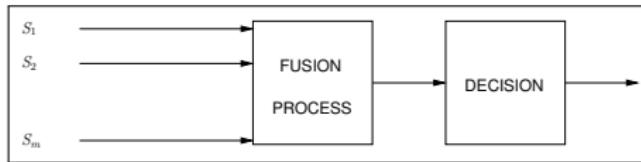
- **Statistic Theory**: Gaussian Maximum Likelihood, Fisher Discriminant Analysis...
- **Machine learning Theory**: Neural Network, Support Vector Machines...
- **Fuzzy Set Theory**: Fuzzy NN, Fuzzy model...

Different sources:

- Panchromatic and multispectral data
- Multi-valued data
- Multi-temporal data
- Extracted data: texture or geometrical features

Complementary information

Decision fusion scheme (for classification):

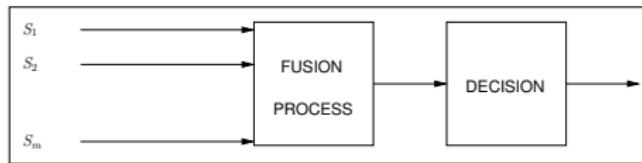


Problems:⁸

- Sources of different natures
- Sources of different reliabilities
- Conflicting situations

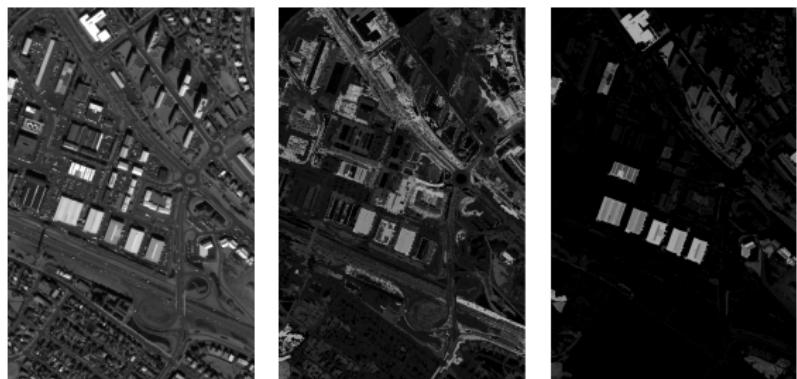
⁸ M. Fauvel, J. Chanussot and J. A. Benediktsson *Decision fusion for the classification of urban remote sensing images*, in IEEE-Trans. Geosci. Remote Sens., Vol. 44, N° 10, October 2006.

Decision fusion scheme (for classification):



Problems:⁸

- Sources of different natures
- Sources of different reliabilities
- Conflicting situations



⁸ M. Fauvel, J. Chanussot and J. A. Benediktsson *Decision fusion for the classification of urban remote sensing images*, in IEEE-Trans. Geosci. Remote Sens., Vol. 44, N° 10, October 2006.

Support Vector Machines

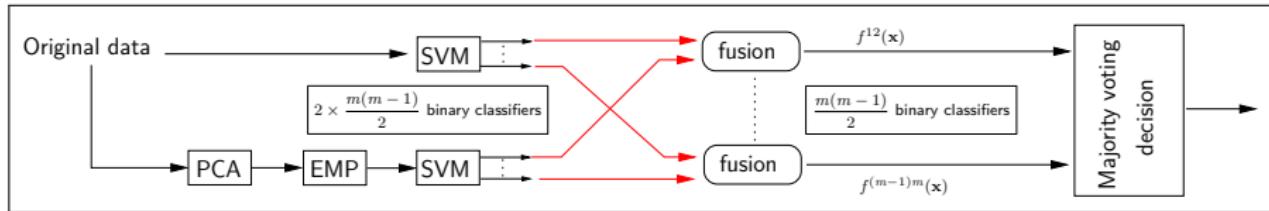
Optimal Separating Hyperplane
Hyperparameters Selection
Using spatial information with SVM

Multiple Classifiers - Fusion of SVM

Problem formulation
Fuzzy Fusion: application to SVM
Multisource Markov Fusion

Conclusions and Perspectives

SVM fusion scheme:

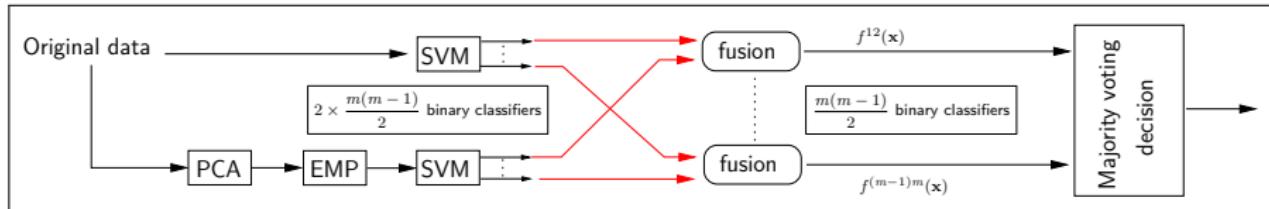


SVM output: distance to the hyperplane $f(\mathbf{x})$

- sign numbers
- not bounded

⁹ M. Fauvel, J. Chanussot and J. A. Benediktsson *A Joint Spatial and Spectral SVMs Classification of Panchromatic Images*, in IEEE-IGARSS, July 2007, Barcelona.

SVM fusion scheme:



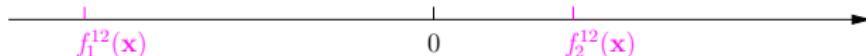
SVM output: distance to the hyperplane $f(\mathbf{x})$

- sign numbers
- not bounded

Fusion rule:⁹

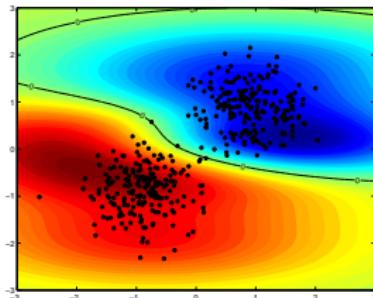
The greater the distance to the hyperplane, the more reliable the source

$$f^{ij}(\mathbf{x}) = \text{maxabs}\left(f_1^{ij}(\mathbf{x}), f_2^{ij}(\mathbf{x})\right)$$

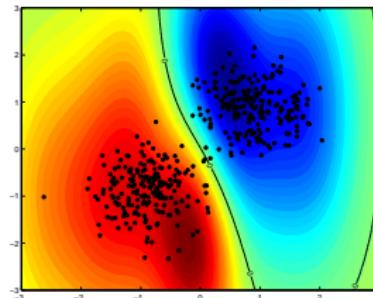


⁹ M. Fauvel, J. Chanussot and J. A. Benediktsson *A Joint Spatial and Spectral SVMs Classification of Panchromatic Images*, in IEEE-IGARSS, July 2007, Barcelona.

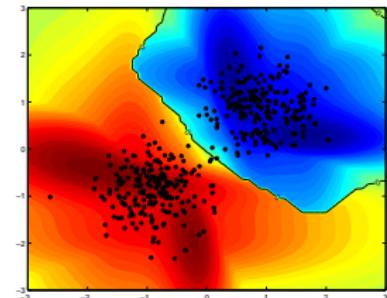
Fusion - Toy example:



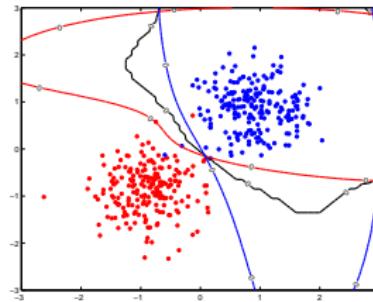
$f_1^{ij}(\mathbf{x})$



$f_2^{ij}(\mathbf{x})$



$f^{ij}(\mathbf{x})$



Fusion - real example:

| | SVM | EMP | DECISION FUSION | MAJORITY VOTING |
|-------------------|-----------|-----------|-----------------|-----------------|
| OVERALL ACCURACY | 81 | 85 | 90 | 86 |
| AVERAGE ACCURACY | 88 | 91 | 94 | 88 |
| KAPPA COEFFICIENT | 76 | 81 | 87 | 82 |
| ASPHALT | 84 | 95 | 93 | 94 |
| MEADOW | 70 | 80 | 84 | 85 |
| GRAVEL | 70 | 88 | 82 | 65 |
| BARE SOIL | 92 | 64 | 91 | 62 |



SVM



EMP



Decision fusion

Support Vector Machines

Optimal Separating Hyperplane
Hyperparameters Selection
Using spatial information with SVM

Multiple Classifiers - Fusion of SVM

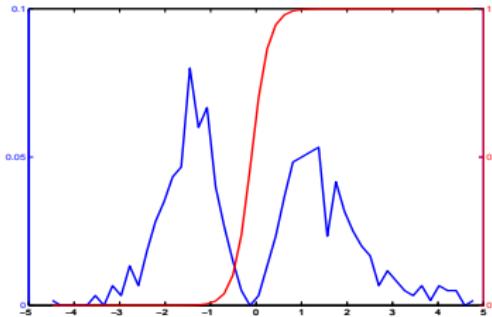
Problem formulation
Fuzzy Fusion: application to SVM
Multisource Markov Fusion

Conclusions and Perspectives

Still spatial aberration : regularization

- $f(\mathbf{x}) \longrightarrow p(y = 1|\mathbf{x}) = \frac{1}{1 + \exp(Af(\mathbf{x}) + B)}$
- Potts model: $\mathcal{U}(y|\mathbf{x}) = \sum_s \log(p(y_s = 1|\mathbf{x}_s)) + \beta \sum_{s,t} \phi(y_s, y_t)$
- Gibbs sampler with temperature: simulated annealing

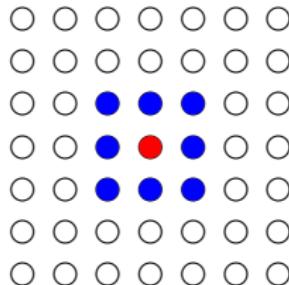
$$P_T(Y = y) = \frac{1}{Z(T)} \exp\left(-\frac{\mathcal{U}(y|\mathbf{x})}{T}\right)$$



Still spatial aberration : regularization

- $f(\mathbf{x}) \longrightarrow p(y = 1|\mathbf{x}) = \frac{1}{1 + \exp(Af(\mathbf{x}) + B)}$
- Potts model: $\mathcal{U}(y|\mathbf{x}) = \sum_s \log(p(y_s = 1|\mathbf{x}_s)) + \beta \sum_{s,t} \phi(y_s, y_t)$
- Gibbs sampler with temperature: simulated annealing

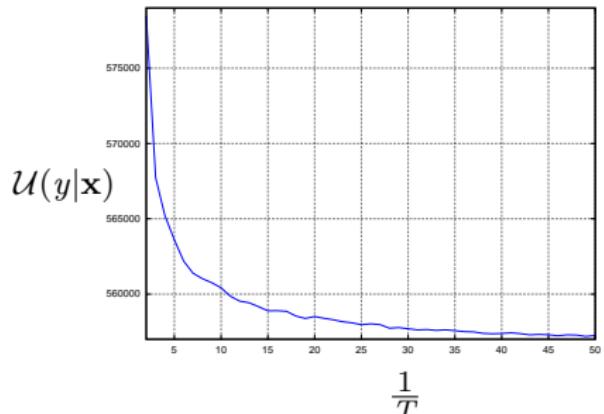
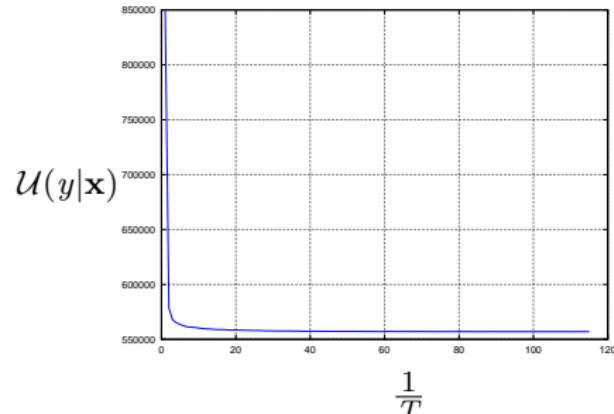
$$P_T(Y = y) = \frac{1}{Z(T)} \exp\left(-\frac{\mathcal{U}(y|\mathbf{x})}{T}\right)$$



Still spatial aberration : regularization

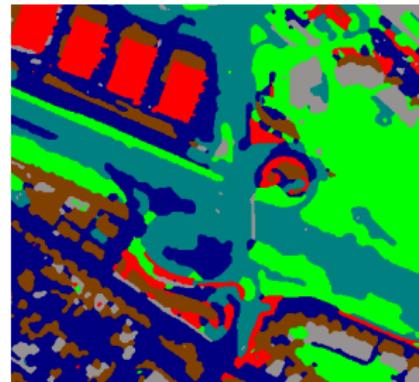
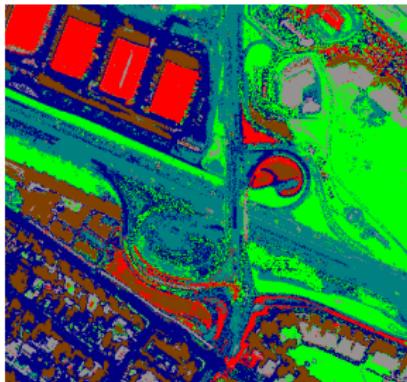
- $f(\mathbf{x}) \longrightarrow p(y = 1|\mathbf{x}) = \frac{1}{1 + \exp(Af(\mathbf{x}) + B)}$
- Potts model: $\mathcal{U}(y|\mathbf{x}) = \sum_s \log(p(y_s = 1|\mathbf{x}_s)) + \beta \sum_{s,t} \phi(y_s, y_t)$
- Gibbs sampler with temperature: **simulated annealing**

$$P_T(Y = y) = \frac{1}{Z(T)} \exp\left(-\frac{\mathcal{U}(y|\mathbf{x})}{T}\right)$$

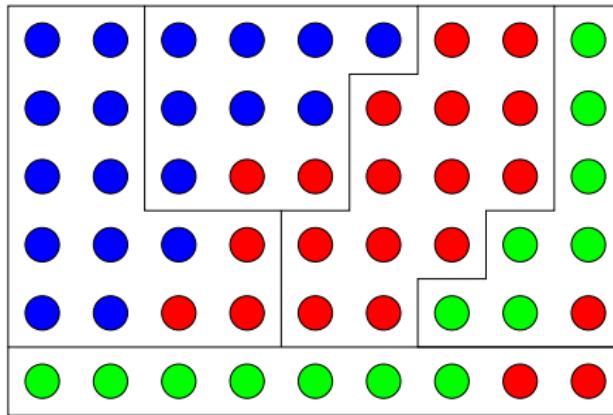




| | | |
|----------|------|------|
| OA | 80 % | 83 % |
| AA | 88 % | 89 % |
| κ | 75 % | 79 % |



| | | |
|----------|------|------|
| OA | 77 % | 82 % |
| AA | 78 % | 84 % |
| κ | 72 % | 78 % |



- Multi-scale model: $\mathcal{U}(\text{within connected set}) + \mathcal{U}(\text{between connected set})$
- Optimization algorithm and parameters setting?

Support Vector Machines
Optimal Separating Hyperplane
Hyperparameters Selection
Using spatial information with SVM

Multiple Classifiers - Fusion of SVM
Problem formulation
Fuzzy Fusion: application to SVM
Multisource Markov Fusion

Conclusions and Perspectives

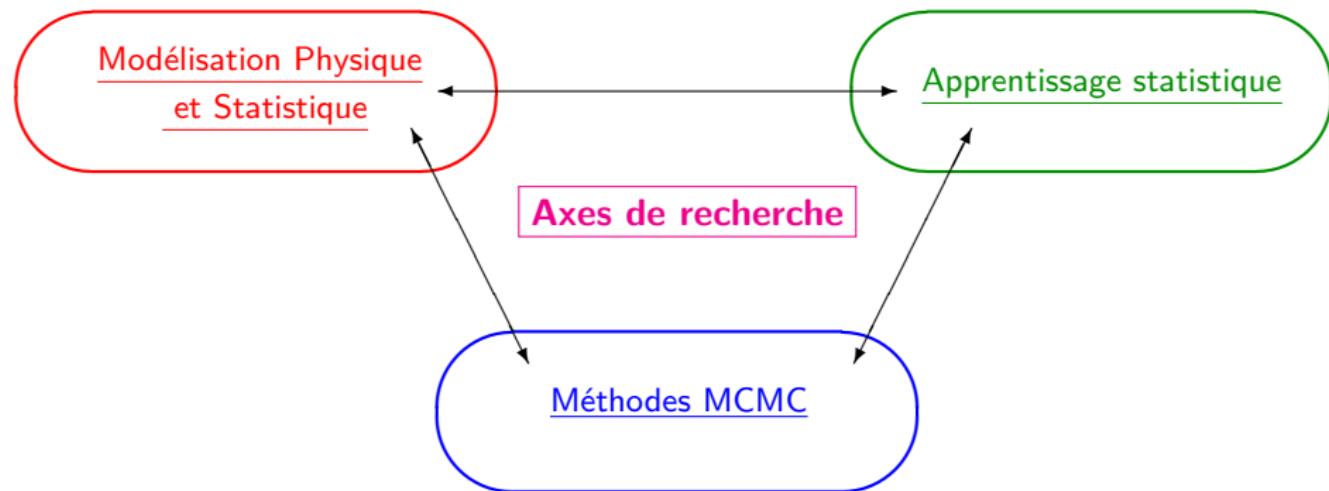
Classification of hyperspectral urban data

- High dimensional space: **Support Vectors Machines**
- Spatial and Spectral information:
 1. Neighbors definition
 2. Mixture of kernels
- Fusion of **complementary** classifiers
 1. Estimation of the reliability
 2. Appropriate fusion scheme

Ongoing works

- Spatial information:
 - Geometric and **textural** feature
 - **Segmentation** and regularization
- Definition of new kernel
 - Kernel for hyperspectral data
 - Use of **semi-supervised SVM**
- **Errors bound** estimation (Radius Margin Bound)
- Markov regularization on **primitive graph: multi-scale approach**

Traitements des signaux multicomposantes et multicapteurs à très grande dimension



Applications: Images et signaux vectoriels, séparation de sources, classification, apprentissage semi-supervisé, extraction de caractéristiques, réseau et coopération multicapteurs.

SPATIAL AND SPECTRAL METHODS FOR THE CLASSIFICATION OF URBAN HYPERSPECTRAL DATA

Mathieu Fauvel

`mathieu.fauvel@gipsa-lab.inpg.fr`

`gipsa-lab/DIS`, Grenoble National Polytechnical Institute - INPG - FRANCE
Department of Electrical and Computer Engineering, University of Iceland - ICELAND

INRIA Rhône-Alpes - Team MISTIS

June 10, 2008