Spectral and Spatial Methods for the Classification of Urban Remote Sensing Data

Mathieu Fauvel

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

Membres du Jury

Monsieur	Henri MAÎTRE	Président
Monsieur	Grégoire MERCIER	Rapporteur
Monsieur	Sebastiano B. SERPICO	Rapporteur
Monsieur	Albert BIJAOUI	Examinateur
Monsieur	Jordi INGLADA	Examinateur
Monsieur	Johannes R. SVEINSSON	Examinateur
Monsieur	Jocelyn CHANUSSOT	Directeur de thèse
Monsieur	Jon A. BENEDIKTSSON	Co-Directeur

Urban remote sensing data:

PLEIADES (0.75m/pix)

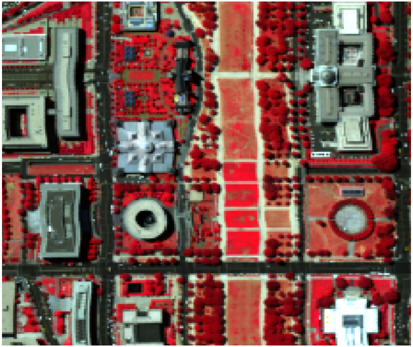
IRS-1C (5m/pix)

Spatial resolution: 0.75 to 2.5 meter by pixel

Spectral resolution: 1 to more than 200 spectral bands

Very High resolution urban remote sensing data:

Panchromatic

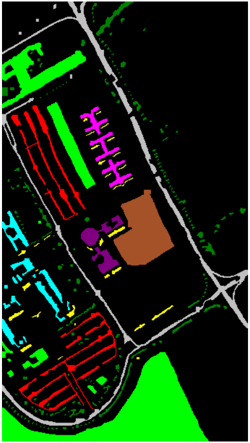


Multispectral

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



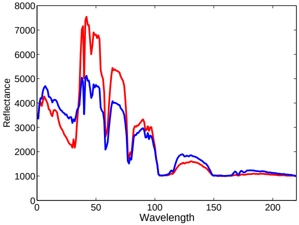
Ground-truth

Thematic Map

Experimental Data Set: *University Area*, Pavia, Italy. $[610 \times 340 \times 103]$, 1.5 m/pixel, 9 classes.

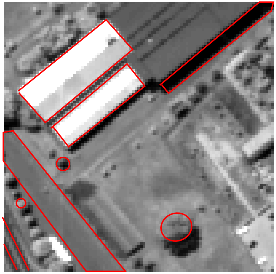
High resolution spectral information:

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



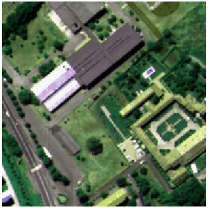
High resolution spatial information:

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

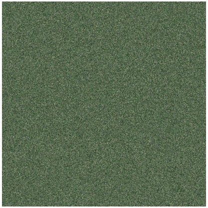


Combine the different types of information for the classification

Why?



→ Random permutation →

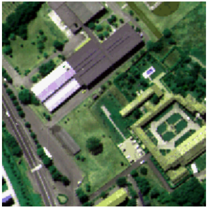


Spectral classification →

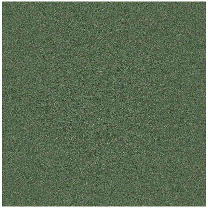
← Spectral classification

Combine the different types of information for the classification

Why?



→ Random permutation →



Spectral classification → Same classification !! ← Spectral classification

Need to incorporate information from the spatial domain

$\boxed{\text{Combine the different types of information for the classification}}$

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

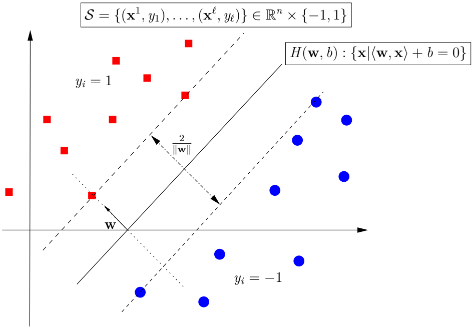
Prior studies: 2 steps approach

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

Contribution:

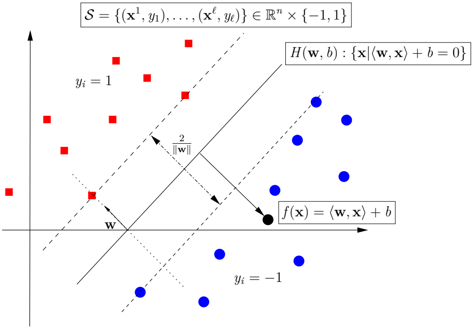
1. Feature extraction: Extraction of
   - Contextual information (self-complementary filter)
   - Spectral feature (KPCA)
2. Classification:
   - Support Vector Machines
   - Transferability of the hyperplane
3. Data Fusion:
   - at data level
   - at decision level

Outline:



Optimal separating hyperplane [Vapnik-98]:

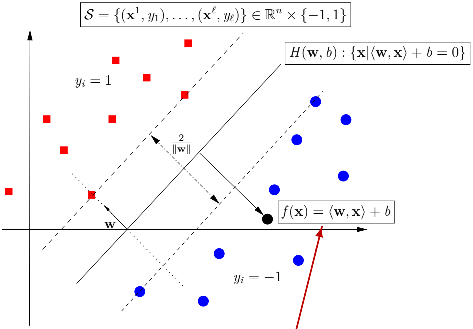
- Minimize training errors over $\mathcal{S}$
- Maximize the margin $\iff$ minimize $\|\mathbf{w}\|^2$



Optimal separating hyperplane [Vapnik-98]:

- Minimize training errors over $\mathcal{S}$
- Maximize the margin $\iff$ minimize $\|\mathbf{w}\|^2$

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



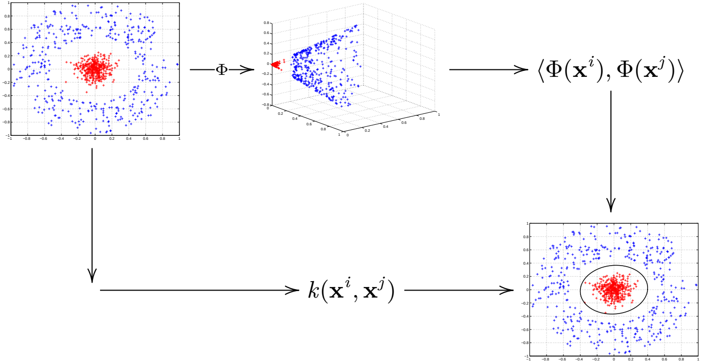
Optimal separating hyperplane [Vapnik-98]:

- Minimize training errors over $\mathcal{S}$
- Maximize the margin $\iff$ minimize $\|\mathbf{w}\|^2$

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

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

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

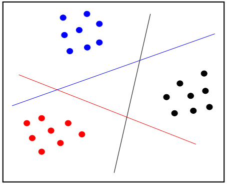


Some kernels:

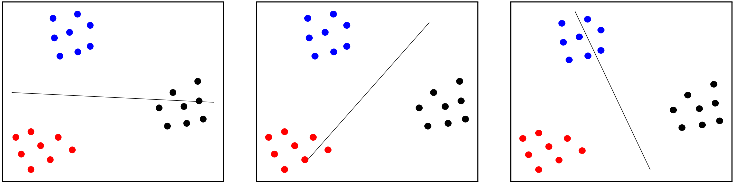
- Polynomial kernel: $k(\mathbf{x}^i, \mathbf{x}^j) = \left(\langle \mathbf{x}^i, \mathbf{x}^j \rangle + q\right)^p$
- Gaussian kernel: $k(\mathbf{x}^i, \mathbf{x}^j) = \exp\left(-\frac{\|\mathbf{x}^i - \mathbf{x}^j\|^2}{\gamma^2}\right)$

Multiclass problem: m classes

1. One versus All: m binary classifiers

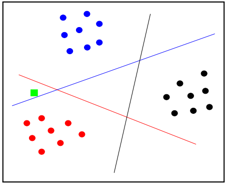


2. One versus One: m(m-1)/2 classifiers

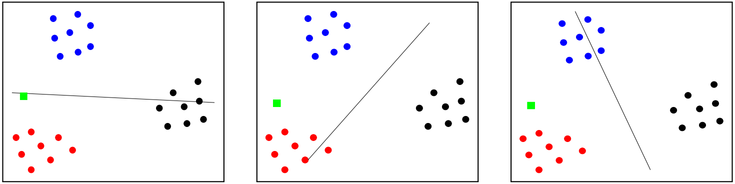


Multiclass problem: m classes

1. One versus All: m binary classifiers



2. One versus One: m(m-1)/2 classifiers



Classification in the spectral space: ▸ spatial & spectral

Gaussian ML (78%)

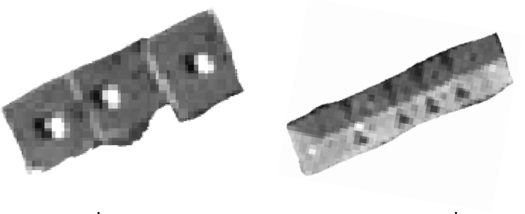
Neural Network (67%)

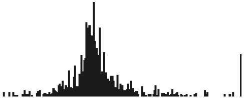
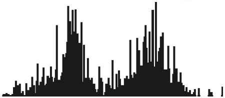
SVM (80%)

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

Spatial information:

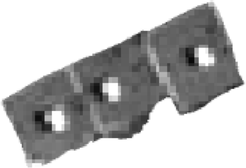
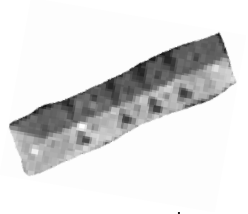
Statistical Information	Geometrical Information
Inter-pixels dependency	Shape
Texture	Area
Gray level distribution	Orientation

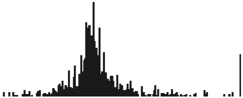
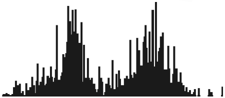


Histogram		
Mean	107	136
Variance	35	48
Shape	Rectangular	Rectangular
Size	8049	9741

Spatial information:

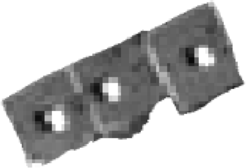
Statistical Information	**Geometrical Information**
Inter-pixels dependency	Shape
Texture	Area
Gray level distribution	Orientation

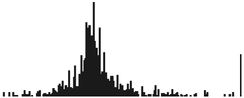
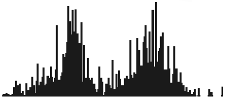



Histogram		
Mean	107	136
Variance	35	48
Shape	Rectangular	Rectangular
Size	8049	9741

Spatial information:

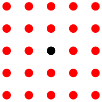
Statistical Information	**Geometrical Information**
Inter-pixels dependency	Shape
Texture	Area
Gray level distribution	Orientation



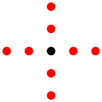
Histogram		
Mean	107	136
Variance	35	48
Shape	Rectangular	Rectangular
Size	8049	9741

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

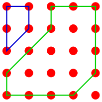
Contextual features [Camps-Valls-06; Bruzzone-06]: fixed neighborhood ($p \times p$ square)



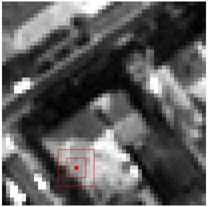
Texture features [Mercier-06] : two 1-D wavelets (on x and y)



Morphological processing [Benediktsson-03]: structures



Not well suited to urban area data: discontinuity



Morphological Neighborhood: adapt the neighborhood to the structures

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

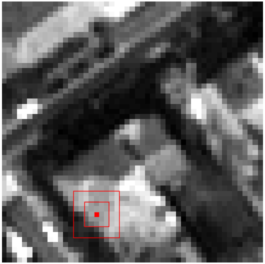
Acts only on extrema structures

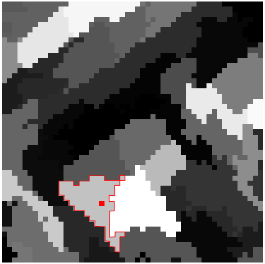
- Proposed works: Self-complementary filters.

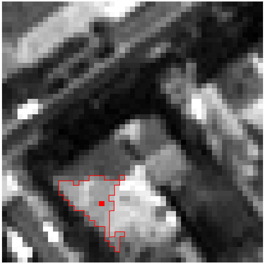
Acts on structures whatever their gray-level

Self-complementary area filter [Soille-05]: 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.

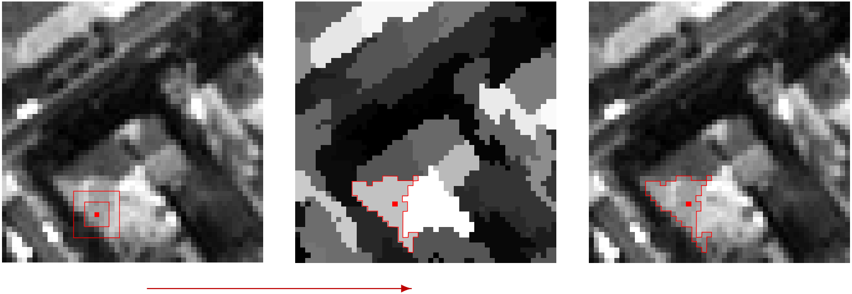






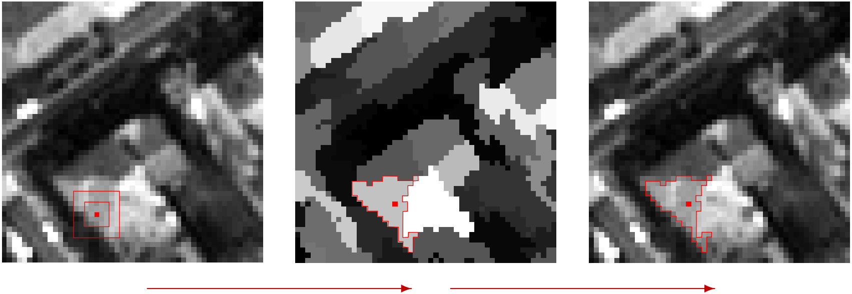
Self-complementary area filter [Soille-05]: Remove all the structures that are smaller than a given area threshold

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



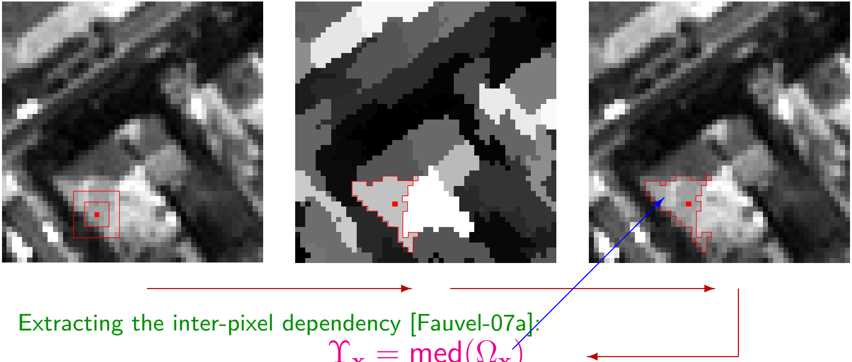
Self-complementary area filter [Soille-05]: Remove all the structures that are smaller than a given area threshold

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



Self-complementary area filter [Soille-05]: 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 [Fauvel-07a]:

$$\Upsilon_{\mathbf{x}} = \mathrm{med}(\Omega_{\mathbf{x}})$$

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)

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

- 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 exit.

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

Our approach: Dimension reduction with PCA.

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

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

Experiments:

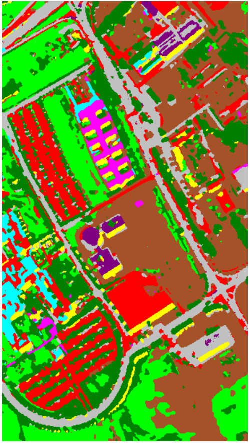
- EMP: Morphological approach (EMP): 3 PCs and 4 opening/closing.
- MRF: Markov Random Field (Potts model):
- SS SVM: Spatio-Spectral SVM

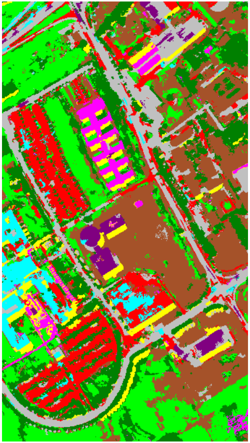
Parameters setting: Cross-validation

Multiclass strategy: One versus all.

Classification using spectral and spatial information:

spectral only





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

Classification accuracies:

%	SVM	EMP	MRF	SS SVM
Overall Accuracy	80	80	83	**86**
Average Accuracy	88	85	89	**92**
Kappa coefficient	75	74	79	**82**
Asphalt	80	**93**	92	84
Meadow	68	73	70	**78**
Gravel	74	52	62	**84**
Bare soil	95	62	**98**	95

Classification accuracies:

%	SVM	EMP	MRF	SS SVM
Overall Accuracy	80	80	83	**86**
Average Accuracy	88	85	89	**92**
Kappa coefficient	75	74	79	**82**
Asphalt	80	**93**	92	84
Meadow	68	73	70	**78**
Gravel	74	52	62	**84**
Bare soil	95	62	**98**	95

$\Downarrow$

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 [Soille-07]

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:

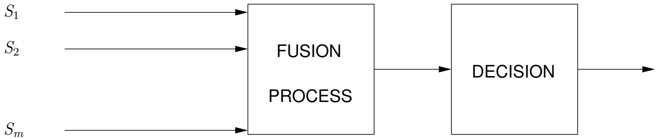
- Statistic Theory: Gaussian Maximum Likelihood, Fisher Discriminant Analysis...
- Machine learning Theory: Neural Network, Support Vector Machines...
- Fuzzy Set Theory: Fuzzy NN, Fuzzy model...

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:

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

Requirement:

- At least one accurate classifier
- Variable reliability of classifier:
  - for each pixel
  - for each class

General framework [Fauvel-06]:

- Modeling the classifier's output with fuzzy set
- Reliability is estimated by fuzziness

Requirement:

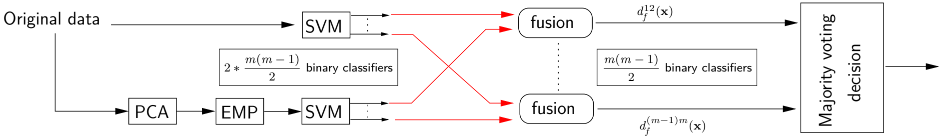
- At least one accurate classifier
- Variable reliability of classifier:
  - for each pixel
  - for each class

General framework [Fauvel-06]:

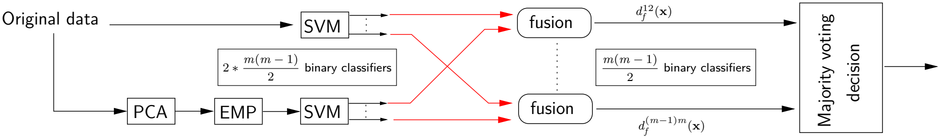
- Modeling the classifier's output with fuzzy set
- Reliability is estimated by fuzziness

Application to SVM classifiers: needs specific derivations

SVM fusion scheme:



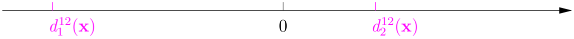
SVM fusion scheme:



Fusion rule [Fauvel-07b]:

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

$$d_f^{ij}(\mathbf{x}) = \text{maxabs}\left(d_1^{ij}(\mathbf{x}), d_2^{ij}(\mathbf{x})\right)$$



Same classifier with different inputs:

- A classifier based on the spectral information: SVM
- A classifier based on the spatial information: EMP + SVM

Experiments:

- spectral features: 103
- spatial features: 33
- parameters fit with CV
- comparison with simple majority vote rule

Classification results:

	SVM	EMP	Desision 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

Classification using decision fusion:

SVM

EMP

Decision fusion

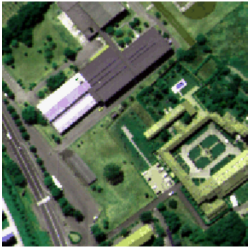
Classification accuracies:

	SVM	EMP	SS SVM	Decision Fusion
Overall Accuracy	80	85	86	**90**
Average Accuracy	88	90	92	**94**
Kappa coefficient	75	80	83	**87**

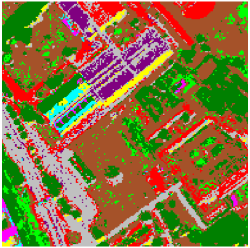
Processing complexity:

	SVM	EMP	SS SVM	Decision Fusion
Pre-processing	$+$	$--$	$-$	$--$
Training	$+$	$+$	$-$	$--$
Classification	$+$	$+$	$+$	$-$

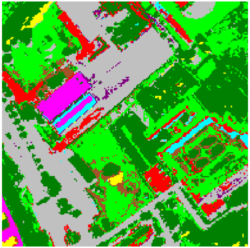
Visual inspection:



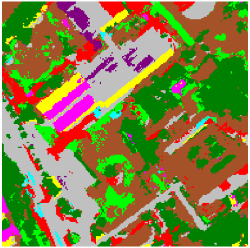
Original



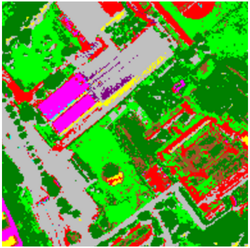
SVM



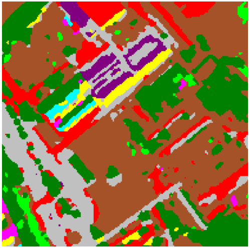
EMP



SS SVM



Decision Fusion



MRF

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

Classification of remote sensing data of urban area

1. Include spatial information: Self-complementary filter
2. Support Vector Machines: kernel approach to include spatial information
3. Decision fusion: adaptive approach based on reliability estimation

Spatial information:

- Extract other spatial information: geometric ...

Classification:

- Kernel for *spectral* data

Data fusion

- Estimation of global accuracy: bound of performances

Spectral and Spatial Methods for the Classification of Urban Remote Sensing Data

Mathieu Fauvel

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