

>>> Classification of Hyperspectral Images

>>> GRSS Summer School

Name: Mathieu Fauvel (UMR Dynafor)

Date: [2017-04-26 Wed 13:30]–[2017-04-26 Wed 16:30]

1. Motivations

2. Introductory examples

Influence of the number of samples

Influence of the number of features

Comparison of state of the art classifier

Comparison spectral feature extraction

3. Spatial-spectral Classification

Introduction

Spatial filter

Data fusion

Spatial post-regularization

Composite kernel

4. References

## 1. Motivations

## 2. Introductory examples

Influence of the number of samples

Influence of the number of features

Comparison of state of the art classifier

Comparison spectral feature extraction

## 3. Spatial-spectral Classification

Introduction

Spatial filter

Data fusion

Spatial post-regularization

Composite kernel

## 4. References

>>> Classification of hyperspectral image is a challenging problem

\* Remember *Introduction*: High dimensional data

- ★ High number of features
- ★ Large volume of pixels
- ★ Low number of labelled pixels

\* Practical issues:

- ★ Intra-class spectral variability
- ★ Non-linear pre-processing (atmospheric/geometric corrections)
- ★ Noise in the data



- ★ Reference data ?
  - ★ Supervised
  - ★ Unsupervised
  - ★ Semi-supervised
- ★  $p(\mathbf{x}, y) \sim ?:$ 
  - ★ Parametric models
  - ★ Non parameterics models
- ★ Number of classes ?
  - ★ Multiclass
  - ★ One-class / detection
  - ★ Unknown



>>> Data sets used



1. Motivations

2. Introductory examples

Influence of the number of samples

Influence of the number of features

Comparison of state of the art classifier

Comparison spectral feature extraction

3. Spatial-spectral Classification

Introduction

Spatial filter

Data fusion

Spatial post-regularization

Composite kernel

4. References

## 1. Motivations

## 2. Introductory examples

Influence of the number of samples

Influence of the number of features

Comparison of state of the art classifier

Comparison spectral feature extraction

## 3. Spatial-spectral Classification

Introduction

Spatial filter

Data fusion

Spatial post-regularization

Composite kernel

## 4. References

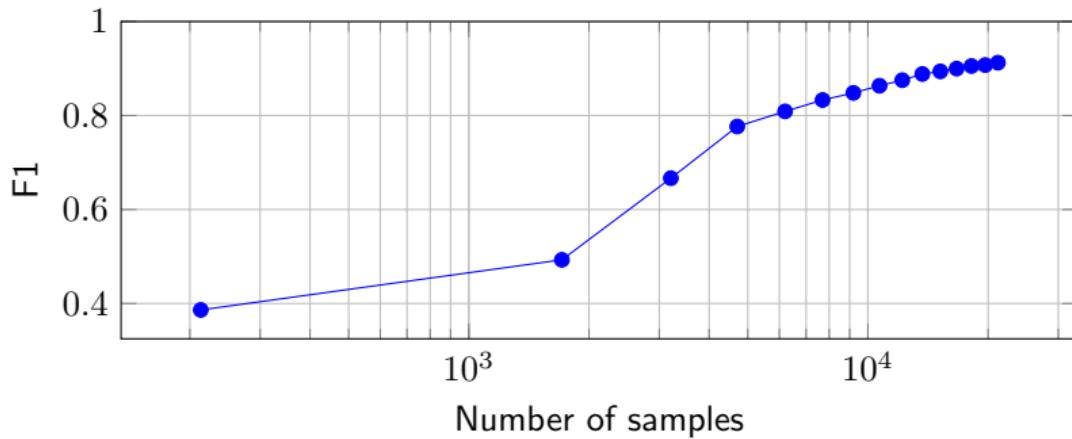
```
>>> Experimental set-up
```

```
import scipy as sp
import rasterTools as rt
from sklearn.model_selection import train_test_split
from sklearn.discriminant_analysis import QuadraticDiscriminantAnalysis
from sklearn.metrics import f1_score

# Load data set
X,y=rt.get_samples_from_roi('..../Data/university.tif','..../Data/university_gt.tif')

# Split the data
X_, X_test, y_, y_test = train_test_split(X, y, train_size=0.50,random_state=0,stratify=y)

# Try differents size of the training set
SPLIT = sp.linspace(0.01,0.99,15)
F1,NS = sp.zeros_like(SPLIT),sp.zeros_like(SPLIT)
for i,split in enumerate(SPLIT):
    # Split the data
    X_train, _, y_train, _ = train_test_split(X_, y_, train_size=split,random_state=0,stratify=y_)
    # Learn the classifier
    clf = QuadraticDiscriminantAnalysis()
    clf.fit(X_train,y_train)
    # Predict the classes
    yp = clf.predict(X_test)
    #Compute the F1
    F1[i],NS[i] = f1_score(y_test,yp,average='weighted'),y_train.size
```



- ★ Dimension of the data:  $d = 103$
- ★ Number of parameters to estimate: 49139

## 1. Motivations

## 2. Introductory examples

Influence of the number of samples

Influence of the number of features

Comparison of state of the art classifier

Comparison spectral feature extraction

## 3. Spatial-spectral Classification

Introduction

Spatial filter

Data fusion

Spatial post-regularization

Composite kernel

## 4. References

```
>>> Experimental set-up
```

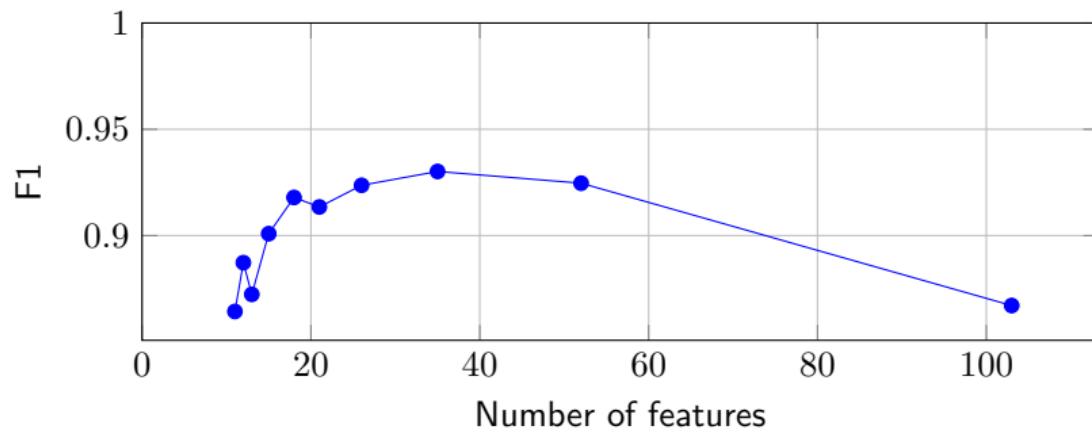
```
import scipy as sp
import rasterTools as rt
from sklearn.model_selection import train_test_split
from sklearn.discriminant_analysis import QuadraticDiscriminantAnalysis
from sklearn.metrics import f1_score

# Load data set
X,y=rt.get_samples_from_roi('..../Data/university.tif','..../Data/university_gt.tif')

# Split the data
X_, X_test, y_train, y_test = train_test_split(X, y, train_size=0.25,random_state=0,stratify=y)

# Try differents size of the training set
SKIP = sorted(range(1,11),reverse=True)
FS,NF = sp.zeros_like(SKIP,dtype='float'),sp.zeros_like(SKIP)
for i,skip in enumerate(SKIP):
    # Skip some variables
    X_train = X_[:,::skip]
    # Learn the classifier
    clf = QuadraticDiscriminantAnalysis()
    clf.fit(X_train,y_train)
    # Predict the classes
    yp = clf.predict(X_test[:,::skip])
    #Compute the FS
    FS[i], NF[i] = f1_score(y_test,yp,average='weighted'), X_train.shape[1]
```

>>> Results



## 1. Motivations

## 2. Introductory examples

Influence of the number of samples

Influence of the number of features

Comparison of state of the art classifier

Comparison spectral feature extraction

## 3. Spatial-spectral Classification

Introduction

Spatial filter

Data fusion

Spatial post-regularization

Composite kernel

## 4. References

>>> Experimental set-up

```
if __name__ == '__main__':
    # Load data set
    X,y=rt.get_samples_from_roi('../Data/university.tif','../Data/university_gt.tif')
    sc = StandardScaler()
    X = sc.fit_transform(X)

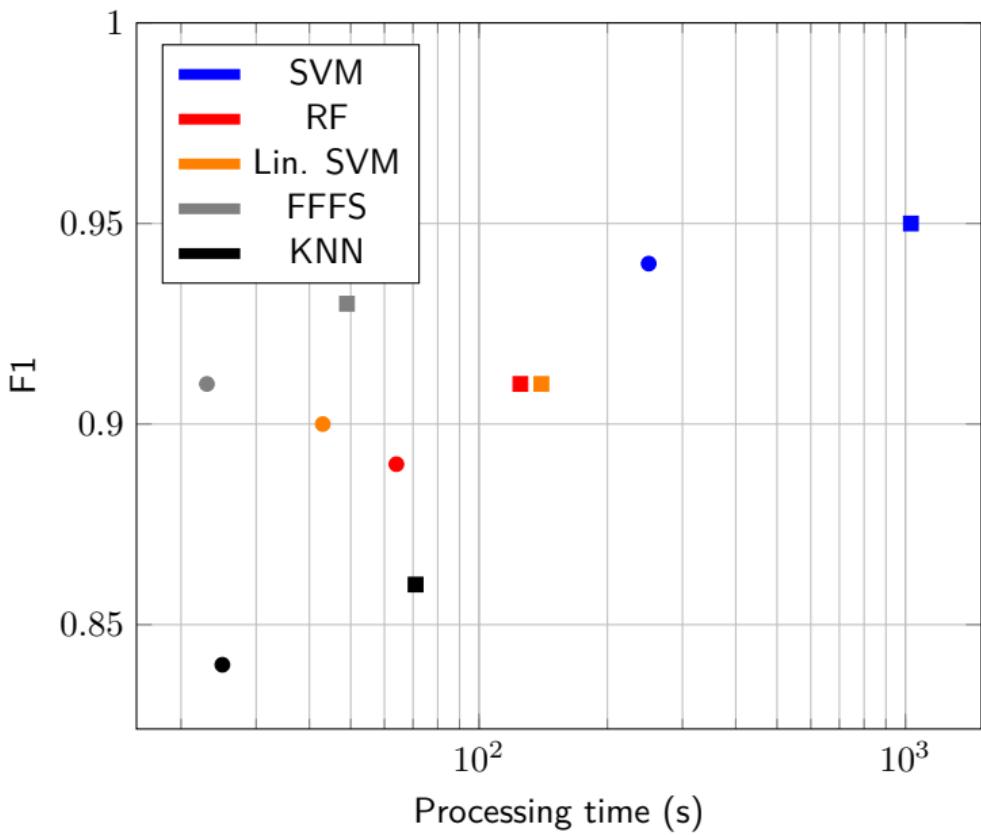
    # Split the data
    X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.1,random_state=0,stratify=y)

    # Parameters
    param_grid_svm = dict(gamma=2.0**sp.arange(-4,4), C=10.0**sp.arange(0,3)) # SVM
    param_grid_linear_svm = dict(C=10.0**sp.arange(-2,3)) # LinearSVM
    param_grid_rf = dict(n_estimators=sp.arange(10,150,10)) # RF
    param_grid_fffs = dict(maxvar=20,threshold=0.001) # FFFS
    param_grid_knn = dict(n_neighbors = sp.arange(1,50,5))
    F1,CT=[],[]

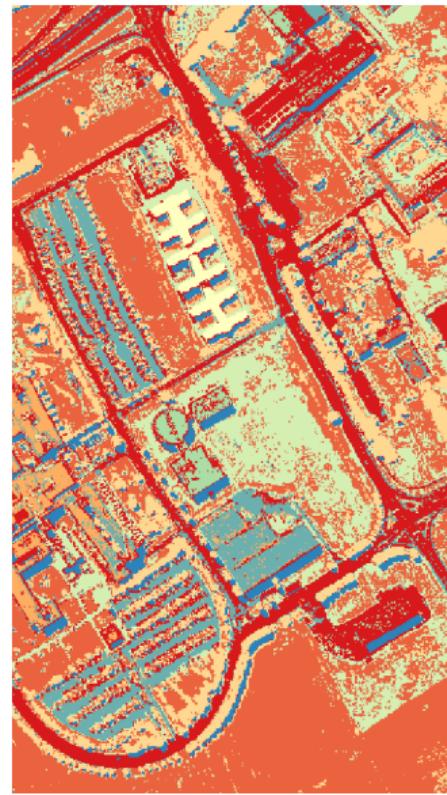
    # Start the classification: SVM
    ts=time.time()
    F1.append(compute_SVM(X_train,y_train,X_test,y_test,param_grid_svm))
    CT.append(time.time()-ts)

    # Start the classification: RF
    ts=time.time()
    F1.append(compute_RF(X_train,y_train,X_test,y_test,param_grid_rf))
    CT.append(time.time()-ts)
```

Class	•	■
1	663	1326
2	1865	3730
3	210	420
4	306	613
5	134	269
6	503	1006
7	133	266
8	368	736
9	95	189



>>> Classification Map



## 1. Motivations

## 2. Introductory examples

Influence of the number of samples

Influence of the number of features

Comparison of state of the art classifier

Comparison spectral feature extraction

## 3. Spatial-spectral Classification

Introduction

Spatial filter

Data fusion

Spatial post-regularization

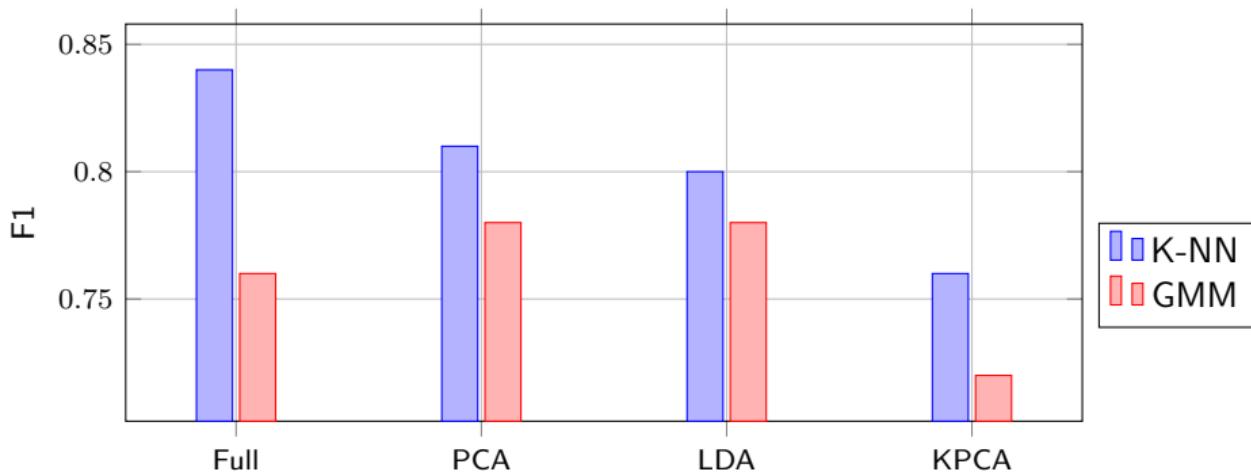
Composite kernel

## 4. References

```
>>> PCA, LDA and KPCA
```

```
DATA = ['../Data/university.tif', '../Data/pca_university.tif', '../Data/lda_university.tif',
        '../Data/kpca_university.tif']
GT = '../Data/university_gt.tif'

F1_knn,F1_gmm = [], []
for data in DATA:
    print data
    # Load data set
    X,y=rt.get_samples_from_roi(data,GT)
    sc = StandardScaler()
    X = sc.fit_transform(X)
```



1. Motivations

2. Introductory examples

Influence of the number of samples

Influence of the number of features

Comparison of state of the art classifier

Comparison spectral feature extraction

3. Spatial-spectral Classification

Introduction

Spatial filter

Data fusion

Spatial post-regularization

Composite kernel

4. References

## 1. Motivations

## 2. Introductory examples

Influence of the number of samples

Influence of the number of features

Comparison of state of the art classifier

Comparison spectral feature extraction

## 3. Spatial-spectral Classification

Introduction

Spatial filter

Data fusion

Spatial post-regularization

Composite kernel

## 4. References

## Spatial inter-pixel dependency

- ★ Spatial feature extraction

- ★ Texture
- ★ Mathematical morphology
- ★ Convolution

- ★ Image Segmentation

- ★ kmeans
- ★ MeanShift

- ★ Markov Random Field

## Spatial inter-pixel dependency

- ★ Spatial feature extraction

- ★ Texture
- ★ Mathematical morphology
- ★ Convolution

- ★ Image Segmentation

- ★ kmeans
- ★ MeanShift

- ★ Markov Random Field

## Joint use of spectral and spatial information

- ★ Data fusion

- ★ Input: Feature stacking,
- ★ Output: fusion of classifier outputs.

- ★ Post classification regularization

- ★ Majority vote,
- ★ Region growing from markers.

- ★ Spatial-spectral classifiers:

- ★ Composite kernel
- ★ MRF

>>> Question to solve

1. What kind of information is needed ?
2. How to extract it from the data ?
3. How to combine it with the spectral information ?

## 1. Motivations

## 2. Introductory examples

Influence of the number of samples

Influence of the number of features

Comparison of state of the art classifier

Comparison spectral feature extraction

## 3. Spatial-spectral Classification

Introduction

Spatial filter

Data fusion

Spatial post-regularization

Composite kernel

## 4. References

- ★ Template filters
  - ★ Mean, variance, median, entropy, range, ...
- ★ Gabor features
- ★ Wavelet features [ML02]
- ★ Co-occurrence [HSD73]

```
otbcli_HaralickTextureExtraction -in ../Data/pca_university.tif -channel 1 \
-out ../Data/university_haralick.tif -parameters.min 789 parameters.max 64897
```

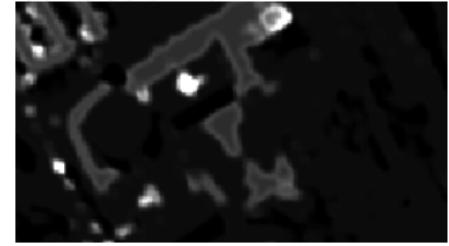
Color Image

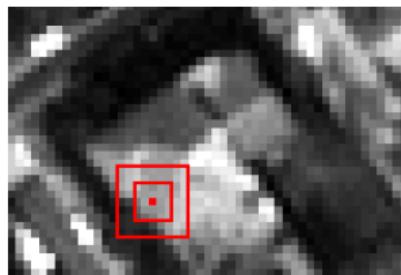


Energy for PC1



Correlation for PC1

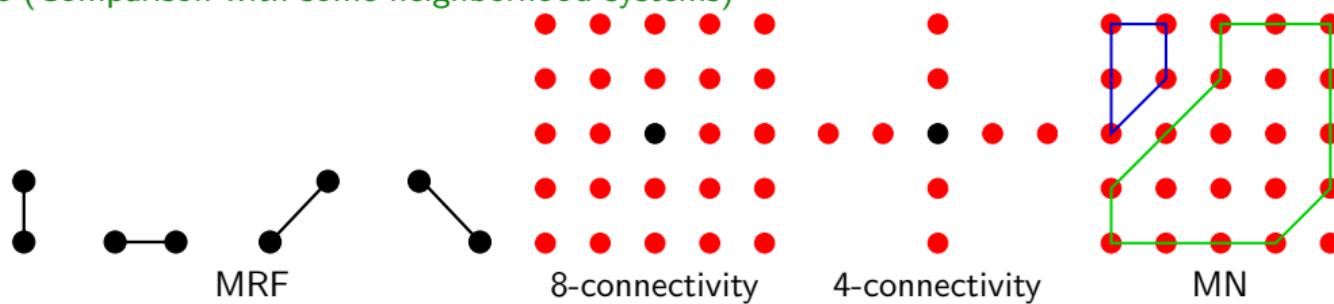




### Definition (Morphological neighborhood)

The Morphological Neighborhood of a pixel  $x$  is the set of pixels that belongs to the same spatial structure than  $x$ .

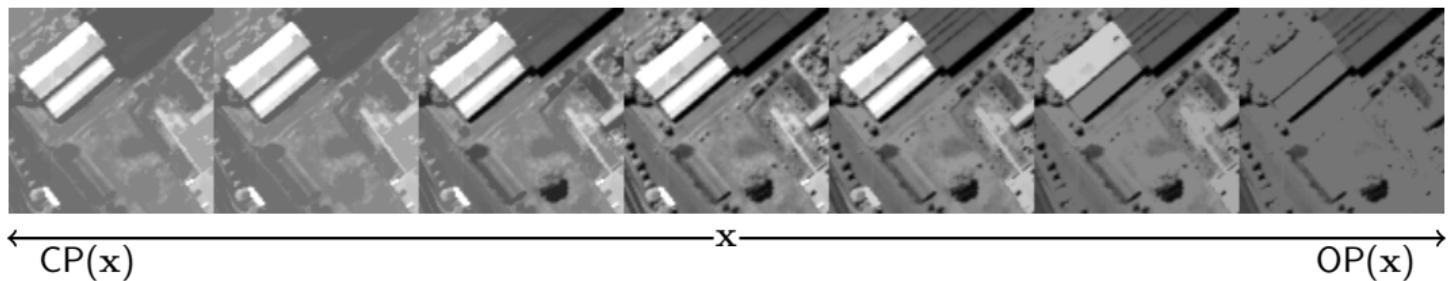
### Example (Comparison with some neighborhood systems)



### Definition (Morphological profile)

The Morphological Profile of size  $n$  is a  $(2n + 1)$ -dimensional vector such as:

$$\text{MP}(\mathbf{x}) = [\text{CP}_n(\mathbf{x}), f(\mathbf{x}), \text{OP}_n(\mathbf{x})].$$

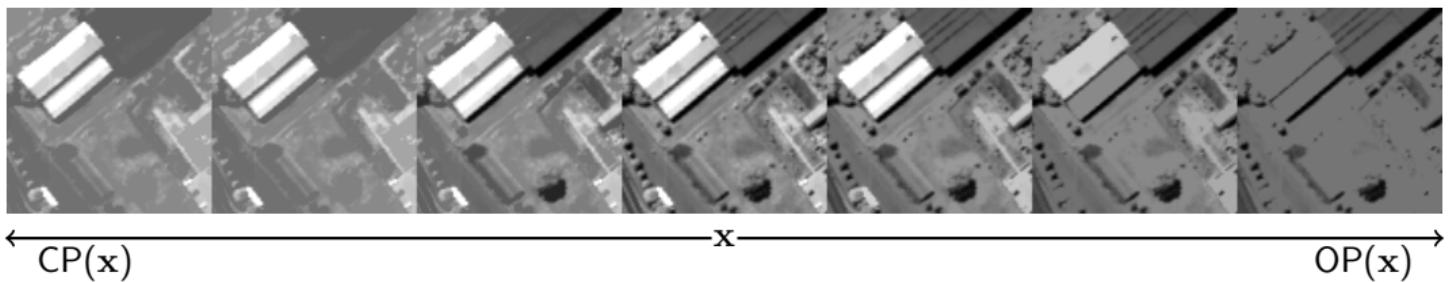


For a given pixel  $x$ , information include in the  $\text{MP}(x)$  are:

### Definition (Morphological profile)

The Morphological Profile of size  $n$  is a  $(2n + 1)$ -dimensional vector such as:

$$\text{MP}(\mathbf{x}) = [\text{CP}_n(\mathbf{x}), f(\mathbf{x}), \text{OP}_n(\mathbf{x})].$$



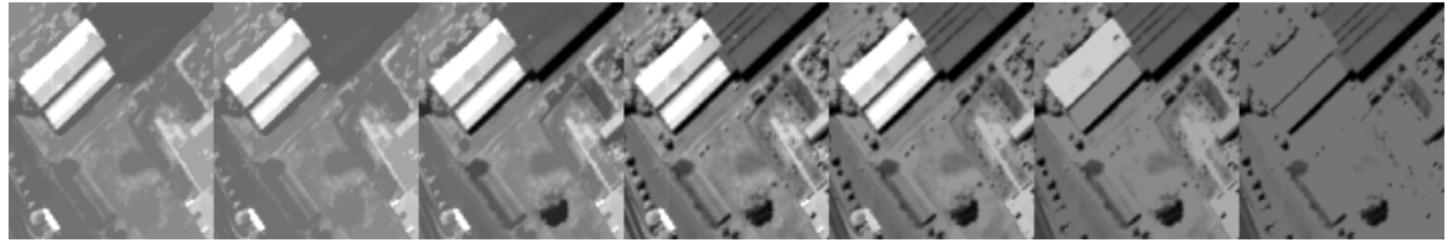
For a given pixel  $x$ , information included in the  $\text{MP}(x)$  are:

- ★ Contrast: Is the structure to which the pixel belongs darker or lighter than his surrounding neighbors?
- ★ Size: Is the structure to which the pixel belongs small or big compared to  $G$ ?

### Definition (Derivative of the morphological profile)

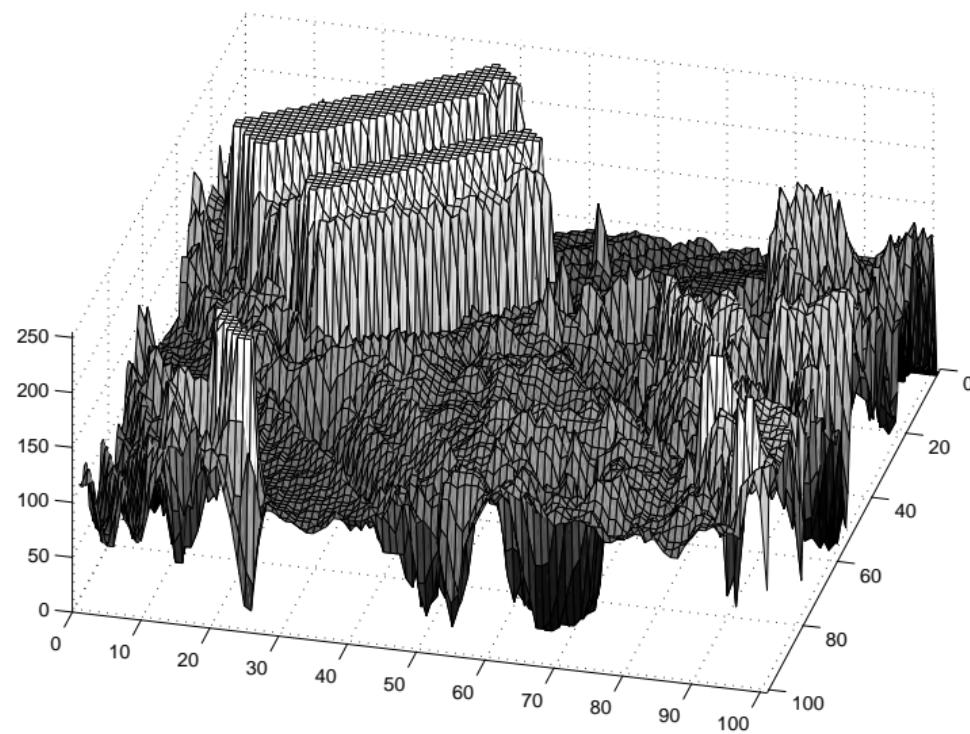
The Derivative of the Morphological Profile of size  $n$  is a  $(2n)$ -dimensional vector such as:

$$\text{DMP}(\mathbf{x}) = \left[ |\phi_n(\mathbf{x}) - \phi_{n-1}(\mathbf{x})|, \dots, |\gamma_{n-1}(\mathbf{x}) - \gamma_n(\mathbf{x})| \right].$$

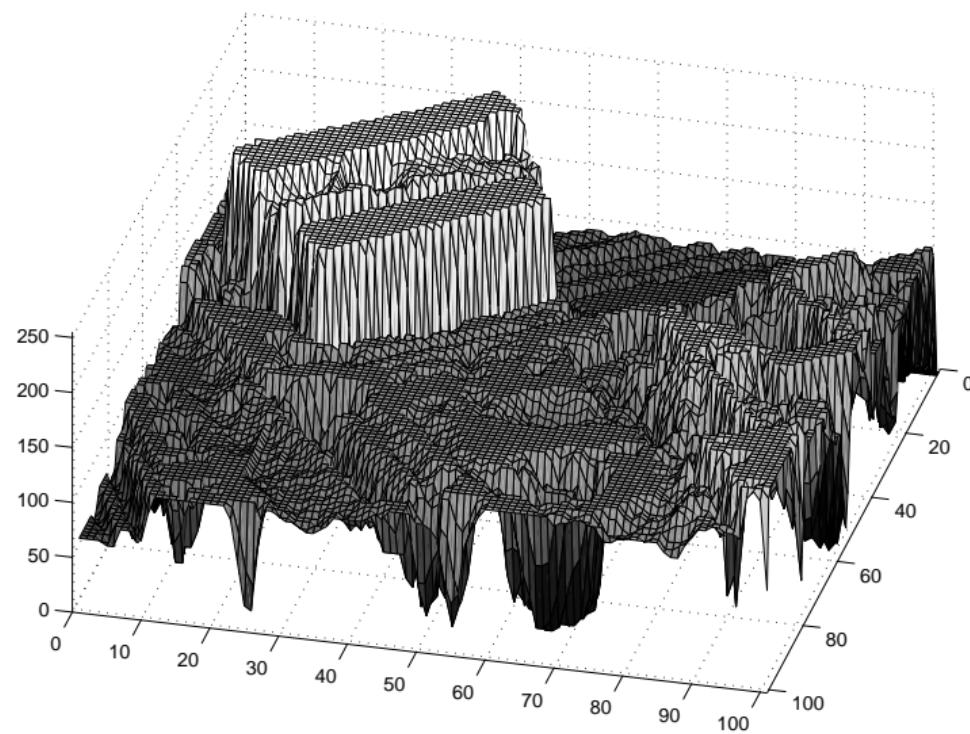


>>> Limits of the (D)MP

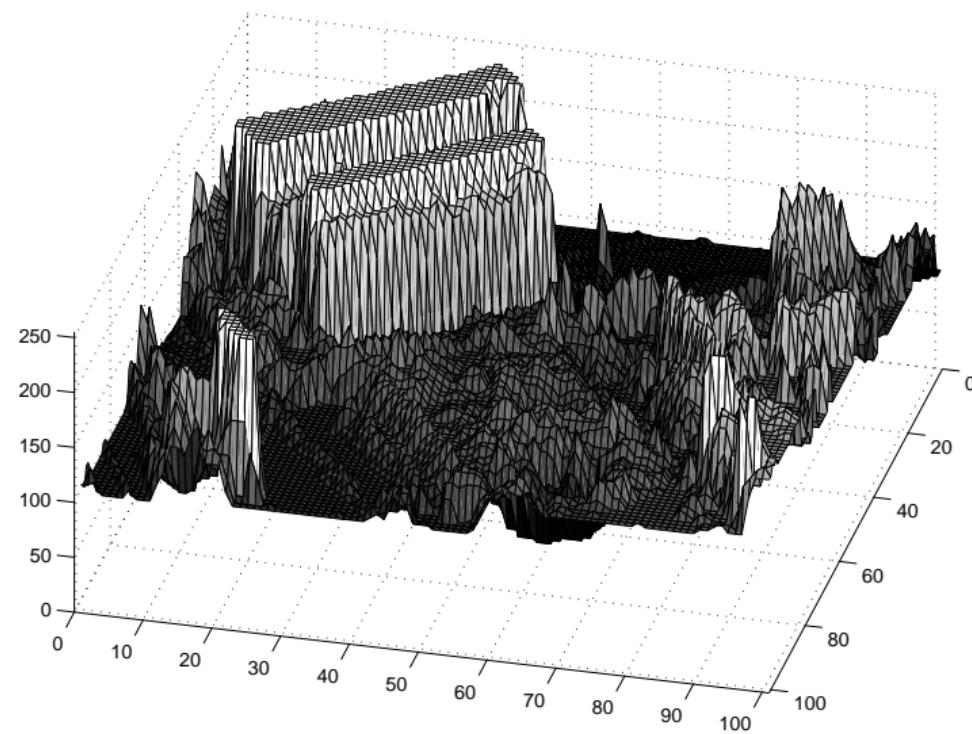
- ★ Geodesics filters only act on extrema structures



- ★ Geodesics filters only act on extrema structures

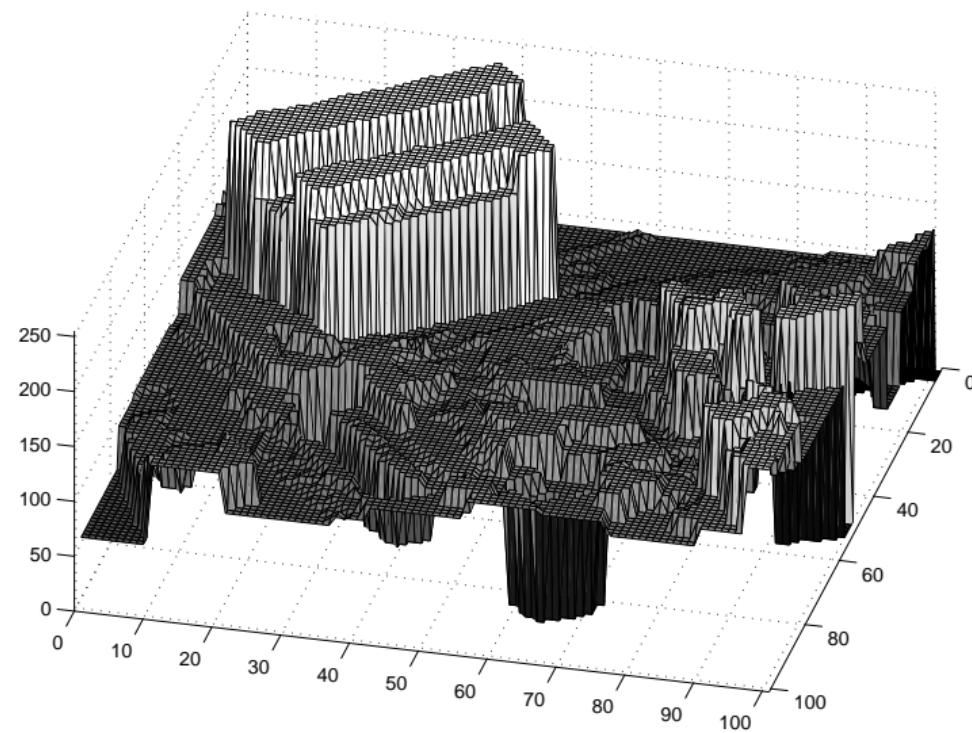


- ★ Geodesics filters only act on extrema structures

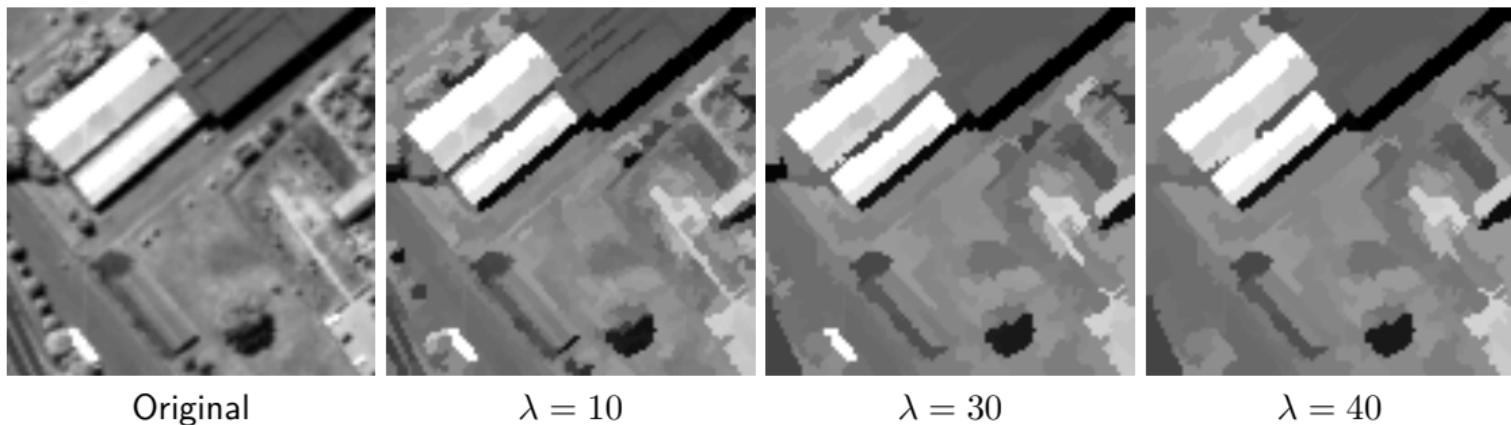


>>> Limits of the (D)MP

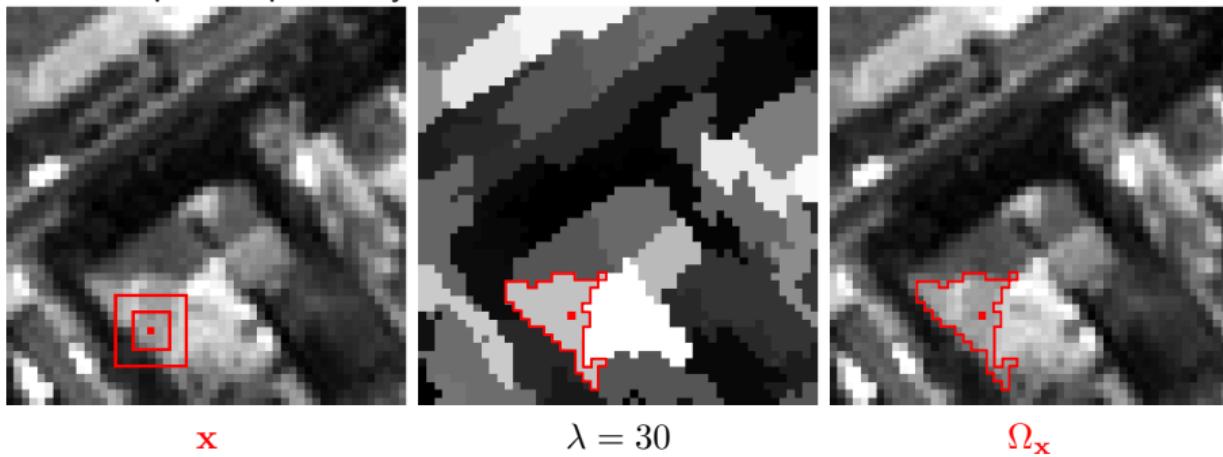
- ★ Geodesics filters only act on extrema structures
- ★ Self-complementary area filter



- ★ Self-complementarity:  $\Psi = \mathbf{C}\Psi \Rightarrow$  each structure is processed equally.
- ★ Area filter: Removes small structures (area = number of pixels).
- ★ Algorithm:
  1. Label all the flat zones that satisfy the area criterion  $\lambda$ ,
  2. Grow the labelled flat zones until a partition of the image is reached.



- ★ Extract the inter-pixel dependency  $\Upsilon$ :



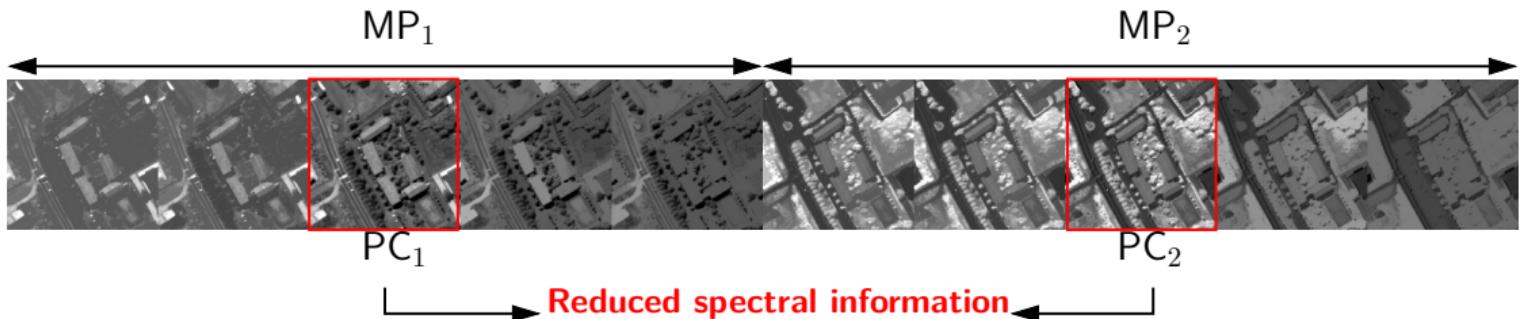
- ★  $\Upsilon_x = \text{median}(\Omega_x)$ :

- ★ Structure: What are the pixels related to  $x$ ?
- ★ Contrast: Local gray-level distribution

### Definition (Extended Morphological Profile)

The EMP of size  $n \times p$  is a  $(2n + 1)p$ -dimensional vector made of the MP build with the  $p$  first principal components:

$$\text{EMP}(\mathbf{x}) = [\text{MP}_1(\mathbf{x}), \dots, \text{MP}_p(\mathbf{x})].$$



- \* Fusion of morphological and spectral features
- \* PCA, FDA, Kernel-PCA ...
- \* *Same methods for self-complementary area filter*

>>> Application case: Extended Morphological Profile

```
# Start the computation
for i in xrange(no):
    # Structuring elements
    se = disk(radius+i*2)

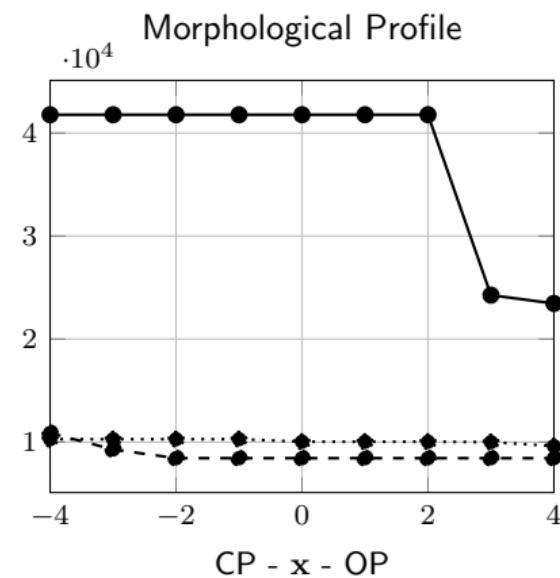
    # Compute opening per reconstruction
    temp = erosion(im,se)
    out[:, :, no+1+i] = reconstruction(temp,im,method='dilation')

    # Compute closing per reconstruction
    temp = dilation(im,se)
    out[:, :, no-1-i] = reconstruction(temp,im,method='erosion')

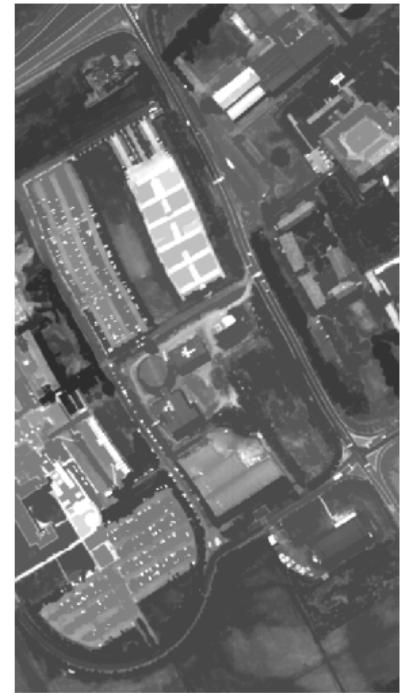
return out

if __name__ == '__main__':
    # Load image
    im,GeoT,Proj = rt.open_data('../Data/pca_university.tif')

    # Apply the Morphological profile on each PC
    EMP = []
    for i in xrange(3):
        EMP.append(morphological_profile(im[:, :, i]))
    EMP = sp.concatenate(EMP, axis=2)
    rt.write_data("../Data/emp_pca_university.tif",EMP,GeoT,Proj)
```



Where is the *closing* and the *opening* ?



## 1. Motivations

## 2. Introductory examples

Influence of the number of samples

Influence of the number of features

Comparison of state of the art classifier

Comparison spectral feature extraction

## 3. Spatial-spectral Classification

Introduction

Spatial filter

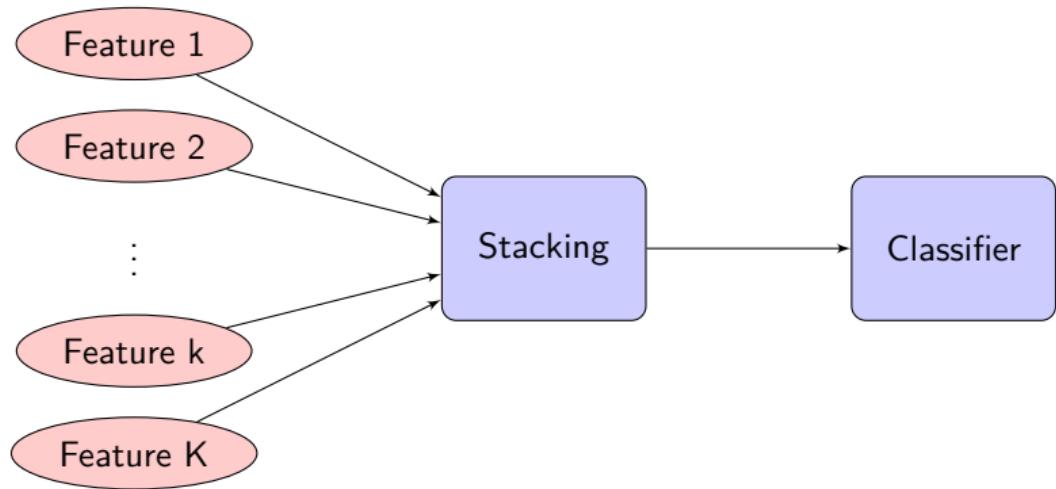
Data fusion

Spatial post-regularization

Composite kernel

## 4. References

>>> Feature fusion 1/2



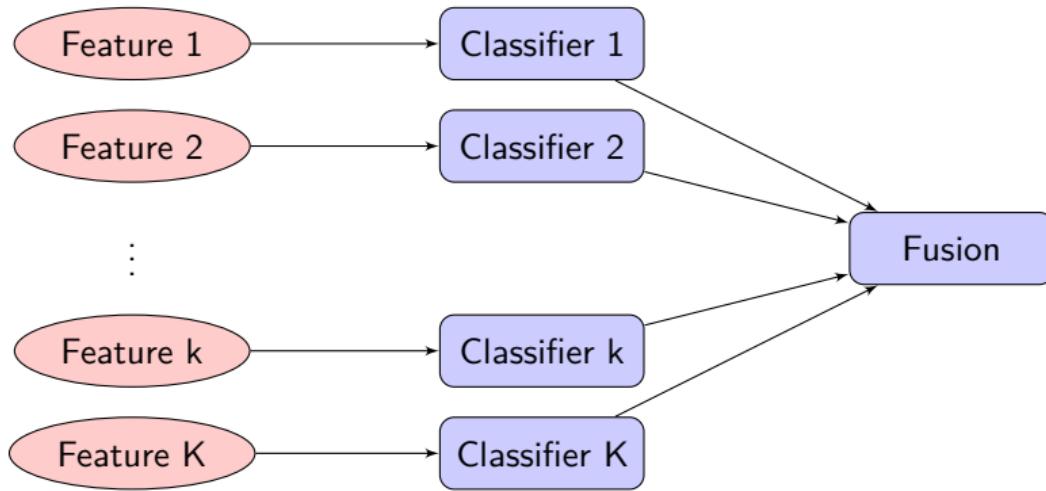
- ★ Extract several spatial descriptors
  - ★ EMP,
  - ★ Texture,
  - ★ Histogram of oriented gradients (HOG),
  - ★ ...
- ★ *Optional:* apply feature extraction
  - ★ Spectral features,
  - ★ Spatial features,
  - ★ Both
- ★ Stack all the features into a "big vector"

## ★ In [Fau+08]:

- ★ Extrat EMP
- ★ Apply PCA/LDA to the spectral and spatial features
- ★ Stack the first PCs of spectral/spatial feautres
- ★ Classification with SVM

Method	# Features	OA
Spectral	103	79.5
EMP	27	79.1
S+EMP	130	83.5
S-DBFE+EMP-DBFE	27+10	88.0

>>> Classifier fusion 1/2



## ★ Fusion of classifier outputs:

- ★ At the decision level

$$C_1 : \{y_1\};$$

$$C_2 : \{y_2\};$$

$$\vdots$$

$$C_K : \{y_K\}$$

- ★ At the membership level

$$C_1 : \{m_{11}, \dots, m_{1C}\};$$

$$C_2 : \{m_{21}, \dots, m_{2C}\};$$

$$\vdots$$

$$C_K : \{m_{K1}, \dots, m_{KC}\}$$

- ★ Decision level: Majority vote
- ★ Membership level: Probabilistics methods, fuzzy logic, Dempster-Shafer ...

- ★ In [FCB06]: Fusion of SVM
- ★ Use the distance to the hyperplane
- ★ *Absolute maximum* fusion rule
- ★ Two classifiers with different intput: Spectral and EMP

Feature	OA
Spectral	81.0
EMP	85.2
Output fusion	89.6

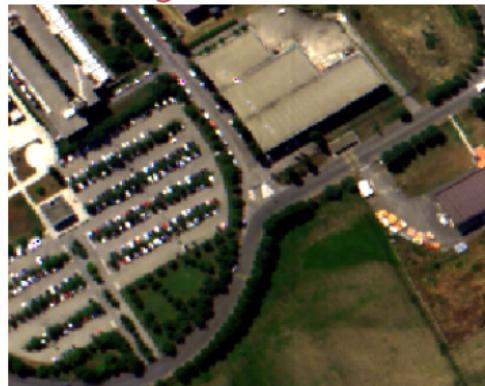
- ★ Simple to implement:

```
# Concatenate the spectral and spatial features and do scaling  
IM_EMP = sp.concatenate((im,EMP),axis=1)  
sc = StandardScaler()  
IM_EMP = sc.fit_transform(IM_EMP)
```

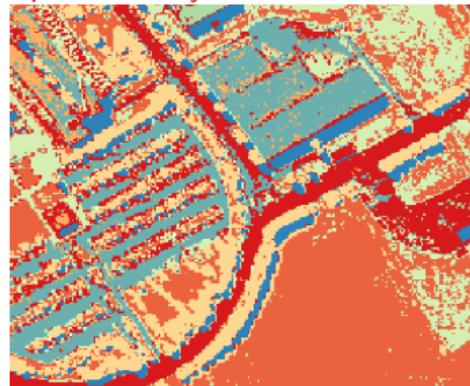
- ★ Good classification accuracy:  $F1 = 0.99$

- ★ sptial

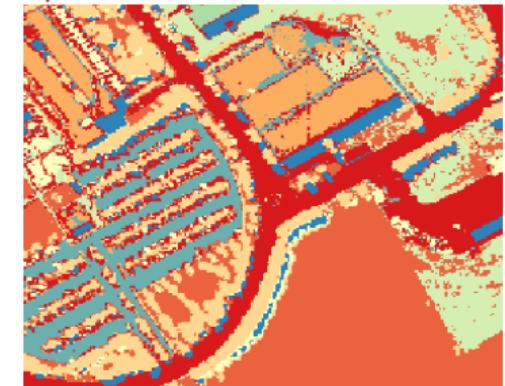
Color Image



Spectral only



Spectral + EMP



## 1. Motivations

## 2. Introductory examples

Influence of the number of samples

Influence of the number of features

Comparison of state of the art classifier

Comparison spectral feature extraction

## 3. Spatial-spectral Classification

Introduction

Spatial filter

Data fusion

Spatial post-regularization

Composite kernel

## 4. References

★ Main ideas

- ★ Segmentation of the image: partition the image into non-overlapping homogeneous zones
- ★ Spatial regularization of the thematic map

★ Issues:

- ★ Segmentation of hyperspectral images is tricky !
- ★ Spatial regularization

★ From [Fau+13]:

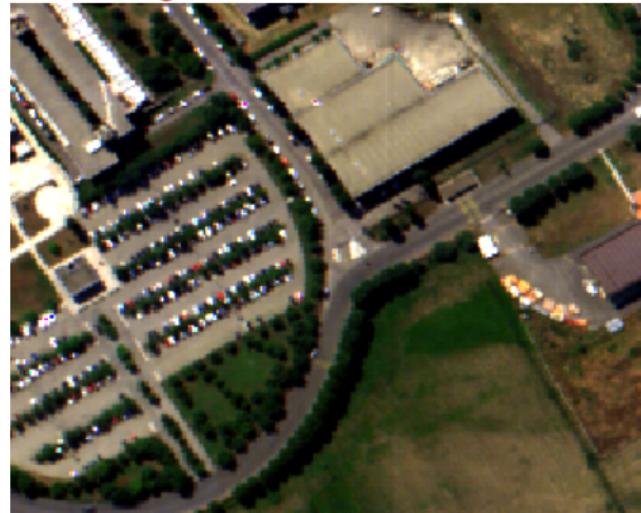
- ★ Segmentation:
  - ★ Image processing: Watershed, region growing, mean-shift, ...
  - ★ Statistical: GMM, K-means ...
- ★ Regularization:
  - ★ Majority voting,
  - ★ Region growing

>>> Segmentation 2/4

# Using Mean Shift

```
otbcli_Segmentation -in ../Data/pca_university.tif -mode raster -mode.raster.out ../Data/mean_shift_unive  
-filter.meanshift.minsize 50
```

Color Image



Segmented



>>> Segmentation 3/4

```
import rasterTools as rt
import scipy as sp
from scipy.stats import mode

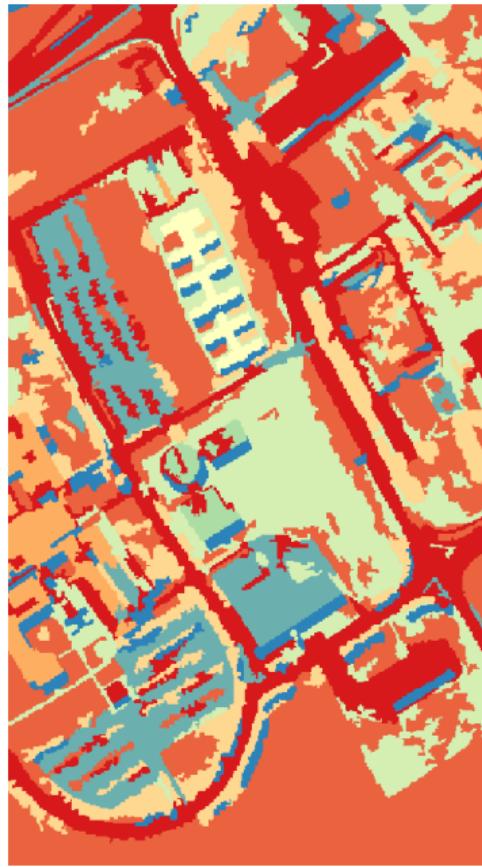
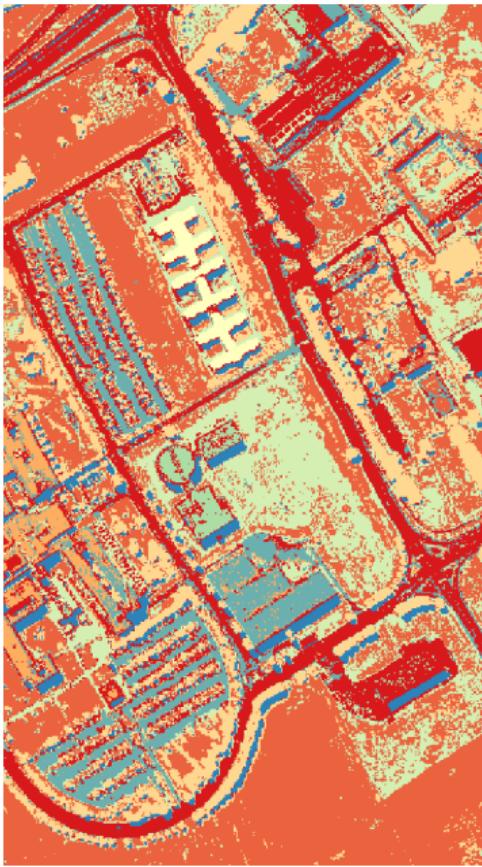
# Load Thematic Map
im,GeoT,Proj = rt.open_data('../Data/tm_university_svm.tif')
out = sp.empty_like(im)

# Load segmented image
segmented,GeoT,Proj = rt.open_data('../Data/mean_shift_university.tif')

# Do the majority vote
for l in sp.unique(segmented):
    t = sp.where(segmented==l)
    y = im[t]
    out[t] = mode(y, axis=None)[0][0]

# Write the new image
rt.write_data("../Data/tm_university_fusion_mv.tif",out,GeoT,Proj)
```

>>> Segmentation 4/4



>>> Markov Random Field

## 1. Motivations

## 2. Introductory examples

Influence of the number of samples

Influence of the number of features

Comparison of state of the art classifier

Comparison spectral feature extraction

## 3. Spatial-spectral Classification

Introduction

Spatial filter

Data fusion

Spatial post-regularization

Composite kernel

## 4. References

1. Motivations

2. Introductory examples

Influence of the number of samples

Influence of the number of features

Comparison of state of the art classifier

Comparison spectral feature extraction

3. Spatial-spectral Classification

Introduction

Spatial filter

Data fusion

Spatial post-regularization

Composite kernel

4. References

- Fauvel, Mathieu, Jocelyn Chanussot, and Jon Atli Benediktsson. "A combined support vector machines classification based on decision fusion". In: *Proc. IEEE Intl. Geoscience and Remote Sensing Symposium*. 2006, pp. 2494–2497.
- Fauvel, Mathieu et al. "Advances in spectral-spatial classification of hyperspectral images". In: *Proceedings of the IEEE* 101.3 (2013), pp. 652–675.
- Fauvel, Mathieu et al. "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles". In: *Geoscience and Remote Sensing, IEEE Transactions on* 46.11 (2008), pp. 3804–3814.
- Haralick, R. M., K. Shanmugam, and I. Dinstein. "Textural Features for Image Classification". In: *IEEE Transactions on Systems, Man, and Cybernetics SMC-3.6* (Nov. 1973), pp. 610–621. ISSN: 0018-9472. DOI: [10.1109/TSMC.1973.4309314](https://doi.org/10.1109/TSMC.1973.4309314).
- Mercier, G. and M. Lennon. "On the characterization of hyperspectral texture". In: *IEEE International Geoscience and Remote Sensing Symposium*. Vol. 5. 2002, 2584–2586 vol.5. DOI: [10.1109/IGARSS.2002.1026708](https://doi.org/10.1109/IGARSS.2002.1026708).

Creative Commons Attribution-ShareAlike 4.0 Unported License

