Hyperspectral Image Classification Using Sparse Representations of Morphological Attribute Profiles

Mauro DALLA MURA

`mauro.dalla-mura@gipsa-lab.grenoble-inp.fr`

Grenoble Images sPeech Signals and Automatics Lab (GIPSA-Lab),
Département Images et Signal,
Grenoble Institute of Technology

gipsa-lab

Grenoble INP ENSE3

3^{e} Colloque du Groupe SFPT-GH Porquerolles, 14-16 mai 2014

SFPT

Groupe hyperspectral

A joint work with...

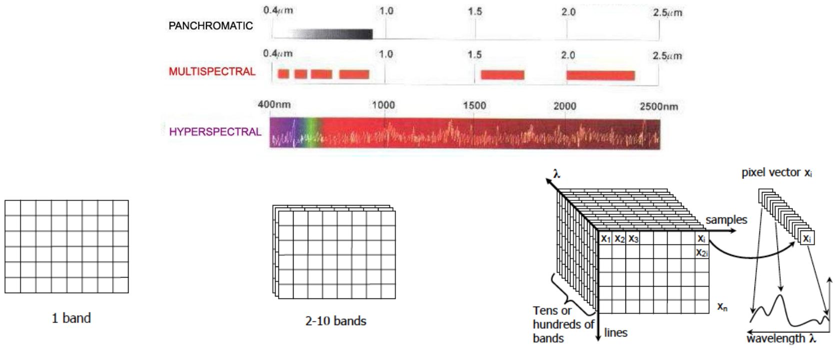
- **Benqin Song** (School of Earth and Space Sciences, Peking University, China)
- **Jun Li** (University of Extremadura, Cáceres, Spain)
- **Peijun Li** (School of Earth and Space Sciences, Peking University, China)
- **Antonio Plaza** (University of Extremadura, Cáceres, Spain)
- **José M. Bioucas Dias** (Telecommunications Institute, Lisbon, Portugal)
- **J. A. Benediktsson** (University of Iceland, Iceland)
- **J. Chanussot** (GIPSA-lab, Grenoble INP, France)

B. Song, J. Li, M. Dalla Mura, P. Li, A. Plaza, J. M Bioucas-Dias, J. A. Benediktsson, and J. Chanussot,"Remotely sensed image classification using sparse representations of morphological attribute profiles," *Geoscience and Remote Sensing, IEEE Transactions on*, accepted.

Outline

1. Introduction on Spectral-spatial Classification
2. Attribute Profiles
3. Sparse Representation Classification
4. Experimental Results
    - Simulated data
    - Real data
5. Conclusion

Optical remote sensing



Panchromatic

- One greylevel per pixel
- Large spectral range

Multispectral

- 2-10 bands
- limited spectral info

Hyperspectral

- ten-hundreds of narrow bands
- detailed spectral info

Exploiting Spatial Features

Example: image classification

True color image

Spectral features
(OA 70.42%)

Spectral+spatial features
(OA 89.89%)

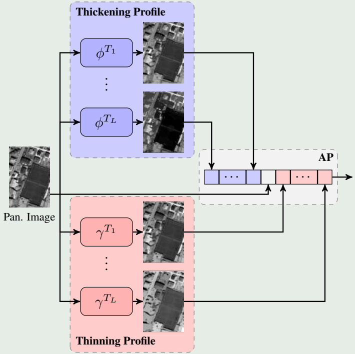
When dealing with high geometrical resolution, the use of spatial features increases the discrimination of the thematic classes leading to more accurate results.

Attribute Profiles

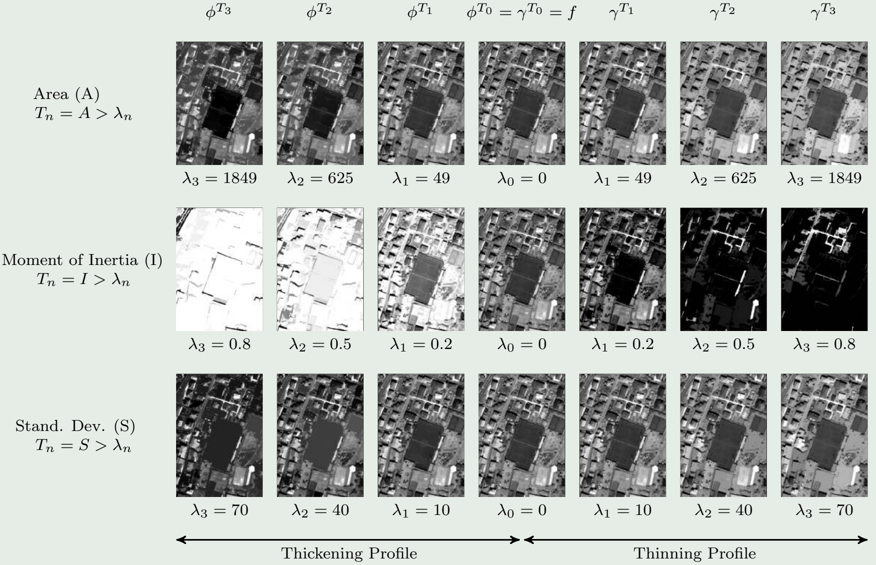
Definition

$$\mathrm{AP}(f) := \{\phi_n(f), \ldots, \phi_1(f), f, \gamma_1(f), \ldots, \gamma_n(f)\},$$

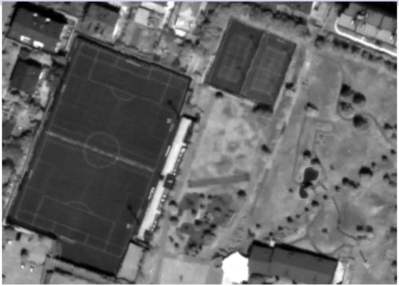
being ϕ and γ attribute thickening and attribute thinning operators, respectively.



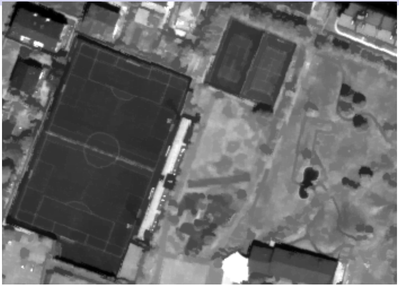
Examples of APs



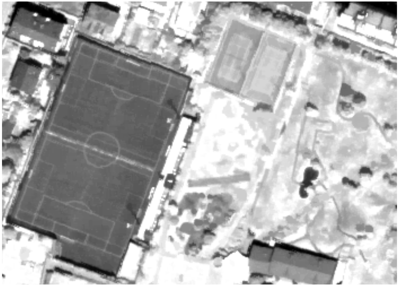
Example of APs (Particulars)



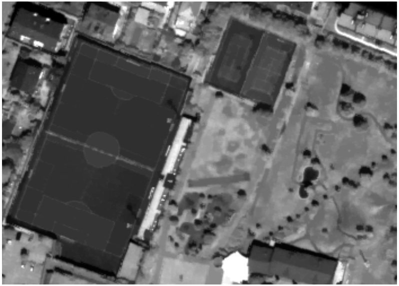
Panchromatic image (f)



$\phi^T(f)$ **Area** $\lambda = 49$

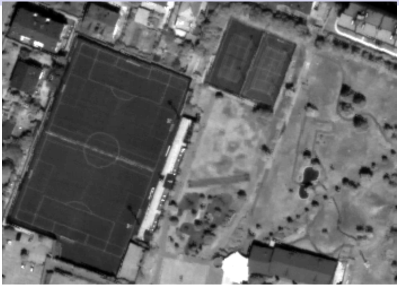


$\phi^T(f)$ **M. Inertia** $\lambda = 0.2$

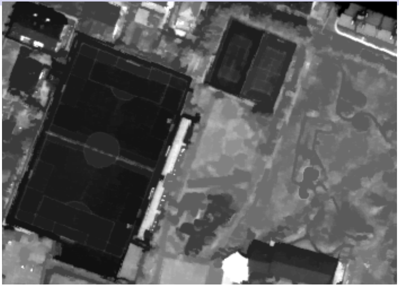


$\phi^T(f)$ **Std** $\lambda = 10$

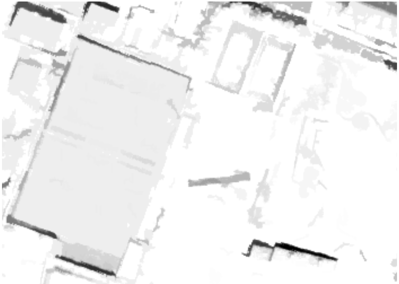
Example of APs (Particulars)



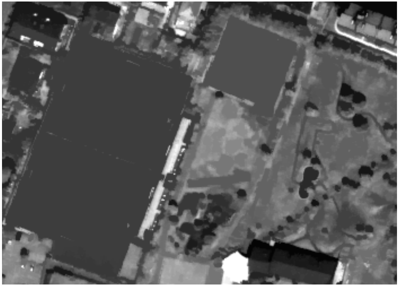
Panchromatic image (f)



$\phi^T(f)$ **Area** $\lambda = 625$

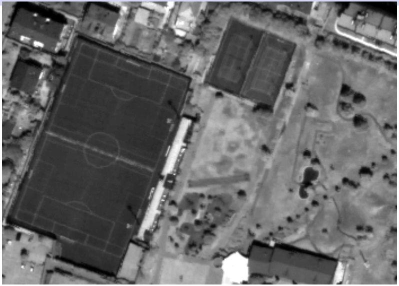


$\phi^T(f)$ **M. Inertia** $\lambda = 0.5$

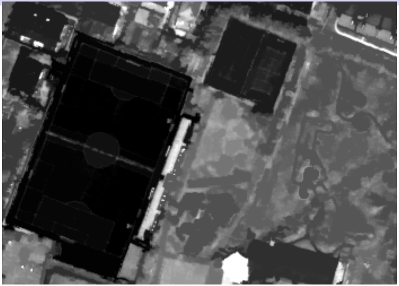


$\phi^T(f)$ **Std** $\lambda = 40$

Example of APs (Particulars)

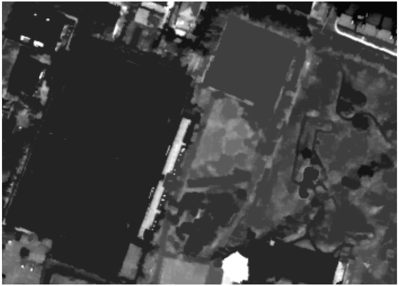


Panchromatic image (f)



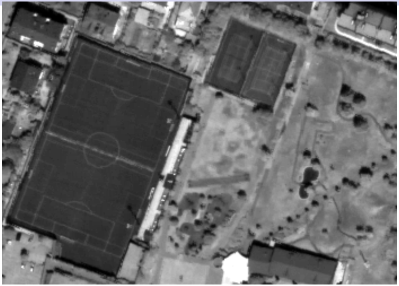
$\phi^T(f)$ **Area** $\lambda = 1849$

$\phi^T(f)$ **M. Inertia** $\lambda = 0.8$

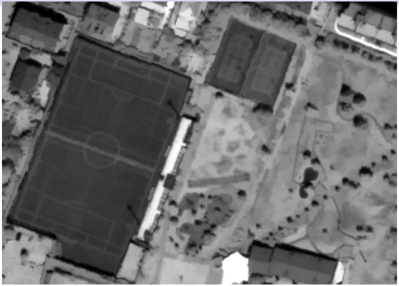


$\phi^T(f)$ **Std** $\lambda = 70$

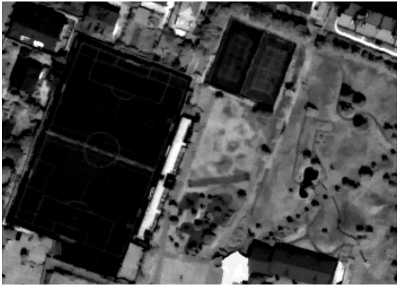
Example of APs (Particulars)



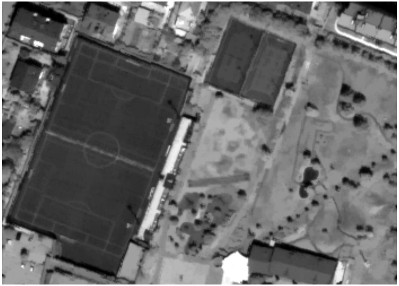
Panchromatic image (f)



$\gamma^T(f)$ **Area** $\lambda = 49$

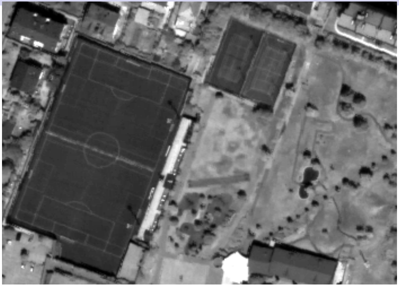


$\gamma^T(f)$ **M. Inertia** $\lambda = 0.2$

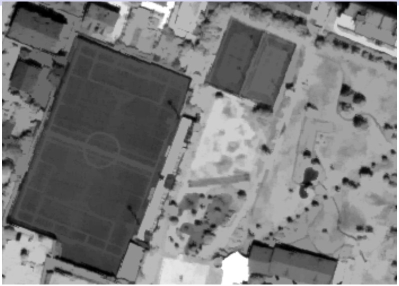


$\gamma^T(f)$ **Std** $\lambda = 10$

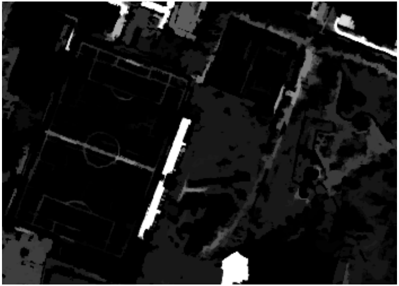
Example of APs (Particulars)



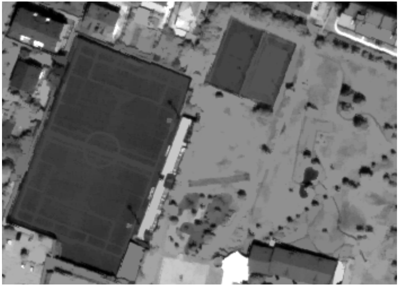
Panchromatic image (f)



$\gamma^T(f)$ **Area** $\lambda = 625$

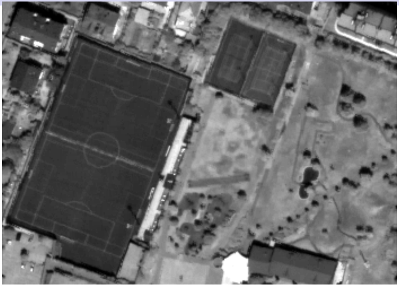


$\gamma^T(f)$ **M. Inertia** $\lambda = 0.5$

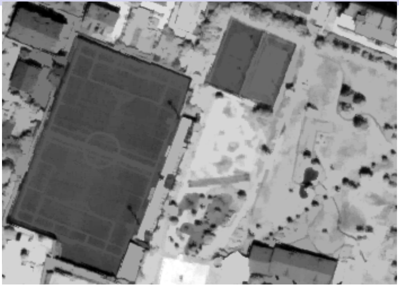


$\gamma^T(f)$ **Std** $\lambda = 40$

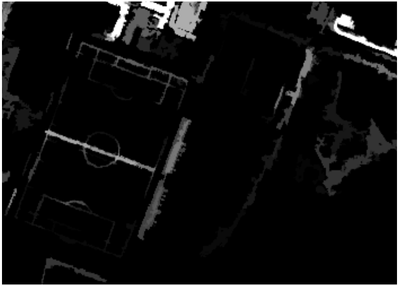
Example of APs (Particulars)



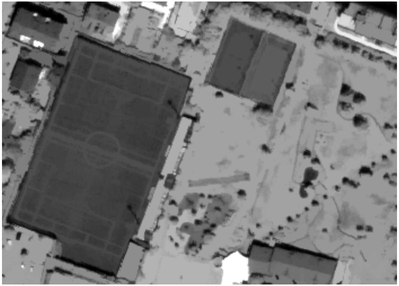
Panchromatic image (f)



$\gamma^T(f)$ **Area** $\lambda = 1849$



$\gamma^T(f)$ **M. Inertia** $\lambda = 0.8$



$\gamma^T(f)$ **Std** $\lambda = 70$

Extended Attribute Profile

- Principal Component Analysis is used for reducing the dimensionality.
- On each first n principal component (PC) extracted from the hyperspectral image, a AP is computed.
- The APs are then concatenated for obtaining an Extended Attribute Profile.

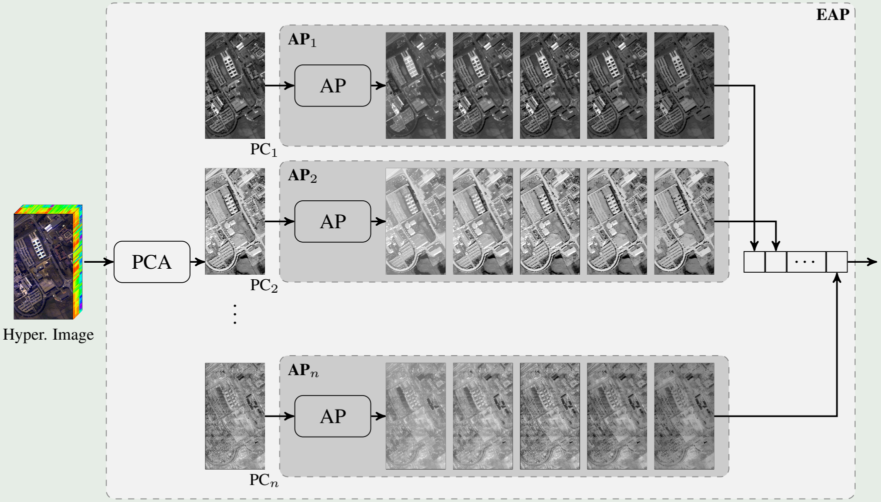
Extended Attribute Profile (EAP)[1]

$$\mathrm{EAP} = \{\mathrm{AP}(\mathrm{PC}_1), \mathrm{AP}(\mathrm{PC}_2), \ldots, \mathrm{AP}(\mathrm{PC}_n)\}.$$

1 M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *International Journal of Remote Sensing*, vol. 31, no. 22, pp. 5975–5991, Nov. 2010.

Extended Attribute Profile

Architecture



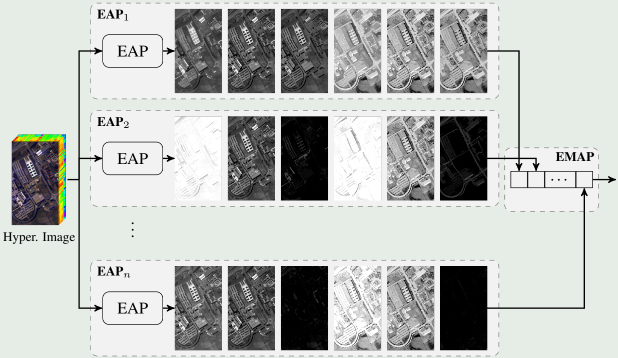
Extended Multi-Attribute Profile

Extended Multi-Attribute Profile

$$\mathrm{EMAP} = \{\mathrm{EAP}_{a_1}, \mathrm{EAP}'_{a_2}, \ldots, \mathrm{EAP}'_{a_m}\}$$

with a_i a generic attribute and $\mathrm{EAP}' = \mathrm{EAP} \setminus \{\mathrm{PC}_1, \ldots, \mathrm{PC}_\mathrm{n}\}$.

Architecture



Sparse Representation

Sparse Representation

- Principle a sample (i.e., a pixel) can be represented by a (sparse) linear combination of atoms from a training dictionary
- Dictionary $\mathbf{A} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \in \mathbb{R}^{n \times l}$ with n samples of l dimensions of c distinct classes as $\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_c]$, where $\mathbf{A}_k = \{\mathbf{x}_{k_1}, \dots, \mathbf{x}_{k_{n_k}}\}$ (*i.e.*, $\mathbf{A}_k$ holds the samples of class k in its columns, n_k is the number of samples in $\mathbf{A}_k$ and $\sum_{k=1}^{c} n_k = n$)
- Let $\mathbf{x}$ be a test sample which can be appropriately represented by a linear combination of the atoms (training samples) in the dictionary $\mathbf{A}$:

$$\mathbf{x} \approx \mathbf{x}_1\alpha_1 + \mathbf{x}_2\alpha_2 + \dots + \mathbf{x}_n\alpha_n = [\mathbf{x}_1\mathbf{x}_2 \dots \mathbf{x}_n][\alpha_1\alpha_2 \dots \alpha_n]^T = \mathbf{A}\boldsymbol{\alpha} + \epsilon$$

  where $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1^T, \dots, \boldsymbol{\alpha}_c^T]^T$ is an n-dimensional sparse vector (i.e., most elements of $\boldsymbol{\alpha}$ are zero), $\boldsymbol{\alpha}_i$ is the vector of regression coefficients associated with class i and ϵ is the representation error
- Central assumption $\mathbf{x}^i$ (belonging to class i) is well approximated by $\mathbf{A}_i\boldsymbol{\alpha}_i$, *i.e.*, $\boldsymbol{\alpha}_j = 0$, for $j \neq i$.

Sparse Representation

Sparse Representation

- Principle a sample (i.e., a pixel) can be represented by a (sparse) linear combination of atoms from a training dictionary
- Dictionary $\mathbf{A} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \in \mathbb{R}^{n \times l}$ with n samples of l dimensions of c distinct classes as $\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{A}_c]$, where $\mathbf{A}_k = \{\mathbf{x}_{k_1}, \ldots, \mathbf{x}_{k_{n_k}}\}$ (*i.e.*, $\mathbf{A}_k$ holds the samples of class k in its columns, n_k is the number of samples in $\mathbf{A}_k$ and $\sum_{k=1}^{c} n_k = n$)
- Let $\mathbf{x}$ be a test sample which can be appropriately represented by a linear combination of the atoms (training samples) in the dictionary $\mathbf{A}$:

$$\mathbf{x} \approx \mathbf{x}_1\alpha_1 + \mathbf{x}_2\alpha_2 + \cdots + \mathbf{x}_n\alpha_n = [\mathbf{x}_1\mathbf{x}_2 \ldots \mathbf{x}_n][\alpha_1\alpha_2 \ldots \alpha_n]^T = \mathbf{A}\boldsymbol{\alpha} + \epsilon$$

  where $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1^T, \ldots, \boldsymbol{\alpha}_c^T]^T$ is an n-dimensional sparse vector (i.e., most elements of $\boldsymbol{\alpha}$ are zero), $\boldsymbol{\alpha}_i$ is the vector of regression coefficients associated with class i and ϵ is the representation error
- Central assumption $\mathbf{x}^i$ (belonging to class i) is well approximated by $\mathbf{A}_i\boldsymbol{\alpha}_i$, *i.e.*, $\boldsymbol{\alpha}_j = 0$, for $j \neq i$.

Sparse Representation

Sparse Representation

- Principle a sample (i.e., a pixel) can be represented by a (sparse) linear combination of atoms from a training dictionary
- Dictionary $\mathbf{A} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \in \mathbb{R}^{n \times l}$ with n samples of l dimensions of c distinct classes as $\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_c]$, where $\mathbf{A}_k = \{\mathbf{x}_{k_1}, \dots, \mathbf{x}_{k_{n_k}}\}$ (*i.e.*, $\mathbf{A}_k$ holds the samples of class k in its columns, n_k is the number of samples in $\mathbf{A}_k$ and $\sum_{k=1}^{c} n_k = n$)
- Let $\mathbf{x}$ be a test sample which can be appropriately represented by a linear combination of the atoms (training samples) in the dictionary $\mathbf{A}$:

$$\mathbf{x} \approx \mathbf{x}_1\alpha_1 + \mathbf{x}_2\alpha_2 + \cdots + \mathbf{x}_n\alpha_n = [\mathbf{x}_1\mathbf{x}_2 \dots \mathbf{x}_n][\alpha_1\alpha_2 \dots \alpha_n]^T = \mathbf{A}\boldsymbol{\alpha} + \epsilon$$

  where $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1^T, \dots, \boldsymbol{\alpha}_c^T]^T$ is an n-dimensional sparse vector (i.e., most elements of $\boldsymbol{\alpha}$ are zero), $\boldsymbol{\alpha}_i$ is the vector of regression coefficients associated with class i and ϵ is the representation error
- Central assumption $\mathbf{x}^i$ (belonging to class i) is well approximated by $\mathbf{A}_i\boldsymbol{\alpha}_i$, *i.e.*, $\boldsymbol{\alpha}_j = 0$, for $j \neq i$.

Sparse Representation

Sparse Representation

- Principle a sample (i.e., a pixel) can be represented by a (sparse) linear combination of atoms from a training dictionary
- Dictionary $\mathbf{A} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \in \mathbb{R}^{n \times l}$ with n samples of l dimensions of c distinct classes as $\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{A}_c]$, where $\mathbf{A}_k = \{\mathbf{x}_{k_1}, \ldots, \mathbf{x}_{k_{n_k}}\}$ (*i.e.*, $\mathbf{A}_k$ holds the samples of class k in its columns, n_k is the number of samples in $\mathbf{A}_k$ and $\sum_{k=1}^{c} n_k = n$)
- Let $\mathbf{x}$ be a test sample which can be appropriately represented by a linear combination of the atoms (training samples) in the dictionary $\mathbf{A}$:

$$\mathbf{x} \approx \mathbf{x}_1\alpha_1 + \mathbf{x}_2\alpha_2 + \cdots + \mathbf{x}_n\alpha_n = [\mathbf{x}_1\mathbf{x}_2 \ldots \mathbf{x}_n][\alpha_1\alpha_2 \ldots \alpha_n]^T = \mathbf{A}\boldsymbol{\alpha} + \epsilon$$

  where $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1^T, \ldots, \boldsymbol{\alpha}_c^T]^T$ is an n-dimensional sparse vector (i.e., most elements of $\boldsymbol{\alpha}$ are zero), $\boldsymbol{\alpha}_i$ is the vector of regression coefficients associated with class i and ϵ is the representation error
- Central assumption $\mathbf{x}^i$ (belonging to class i) is well approximated by $\mathbf{A}_i\boldsymbol{\alpha}_i$, *i.e.*, $\boldsymbol{\alpha}_j = 0$, for $j \neq i$.

Sparse Representation

Finding $\boldsymbol{\alpha}$

- The sparse vector $\boldsymbol{\alpha}$ can be estimated by solving the following optimization problem (NP hard):

$$\hat{\boldsymbol{\alpha}} = \arg\min \quad \|\boldsymbol{\alpha}\|_0 \quad \text{subject to} \quad \|\mathbf{x}_i - \mathbf{A}\boldsymbol{\alpha}\|_2 \leq \delta,$$

  where $\|\boldsymbol{\alpha}\|_0$ denotes the ℓ_0-(pseudo) norm which counts the nonzero components in the coefficient vector and δ is an error tolerance.

- This can be approximated by a convex problem (and solved using linear programming) by replacing ℓ_0-norm with the ℓ_1-norm:

$$\hat{\boldsymbol{\alpha}} = \arg\min \quad \|\boldsymbol{\alpha}\|_1 \quad \text{subject to} \quad \|\mathbf{x}_i - \mathbf{A}\boldsymbol{\alpha}\|_2 \leq \delta,$$

- This is equivalent to the unconstrained optimization problem:

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\|\mathbf{x}_i - \mathbf{A}\boldsymbol{\alpha}\|_2^2 + \tau\|\boldsymbol{\alpha}\|_1,$$

  where the parameter τ is a Lagrange multiplier which balances the tradeoff between the reconstruction error and the sparse solution

Sparse Representation

Finding $\boldsymbol{\alpha}$

- The sparse vector $\boldsymbol{\alpha}$ can be estimated by solving the following optimization problem (NP hard):

$$\hat{\boldsymbol{\alpha}} = \arg\min \quad \|\boldsymbol{\alpha}\|_0 \;\; \text{subject to} \;\; \|\mathbf{x}_i - \mathbf{A}\boldsymbol{\alpha}\|_2 \leq \delta,$$

  where $\|\boldsymbol{\alpha}\|_0$ denotes the ℓ_0-(pseudo) norm which counts the nonzero components in the coefficient vector and δ is an error tolerance.

- This can be approximated by a convex problem (and solved using linear programming) by replacing ℓ_0-norm with the ℓ_1-norm:

$$\hat{\boldsymbol{\alpha}} = \arg\min \quad \|\boldsymbol{\alpha}\|_1 \;\; \text{subject to} \;\; \|\mathbf{x}_i - \mathbf{A}\boldsymbol{\alpha}\|_2 \leq \delta,$$

- This is equivalent to the unconstrained optimization problem:

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\|\mathbf{x}_i - \mathbf{A}\boldsymbol{\alpha}\|_2^2 + \tau\|\boldsymbol{\alpha}\|_1,$$

  where the parameter τ is a Lagrange multiplier which balances the tradeoff between the reconstruction error and the sparse solution

Sparse Representation

Finding $\boldsymbol{\alpha}$

- The sparse vector $\boldsymbol{\alpha}$ can be estimated by solving the following optimization problem (NP hard):

$$\hat{\boldsymbol{\alpha}} = \arg\min \quad \|\boldsymbol{\alpha}\|_0 \quad \text{subject to} \quad \|\mathbf{x}_i - \mathbf{A}\boldsymbol{\alpha}\|_2 \leq \delta,$$

  where $\|\boldsymbol{\alpha}\|_0$ denotes the ℓ_0-(pseudo) norm which counts the nonzero components in the coefficient vector and δ is an error tolerance.

- This can be approximated by a convex problem (and solved using linear programming) by replacing ℓ_0-norm with the ℓ_1-norm:

$$\hat{\boldsymbol{\alpha}} = \arg\min \quad \|\boldsymbol{\alpha}\|_1 \quad \text{subject to} \quad \|\mathbf{x}_i - \mathbf{A}\boldsymbol{\alpha}\|_2 \leq \delta,$$

- This is equivalent to the unconstrained optimization problem:

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\|\mathbf{x}_i - \mathbf{A}\boldsymbol{\alpha}\|_2^2 + \tau\|\boldsymbol{\alpha}\|_1,$$

  where the parameter τ is a Lagrange multiplier which balances the tradeoff between the reconstruction error and the sparse solution

Classification

- Once $\boldsymbol{\alpha}$ is estimated, each sample is assigned to the class that shows the least residual in the reconstruction (i.e., the atoms in the dictionary belonging to that class are contributing most to the representation of the sample)
- $$\widehat{\text{class}}(\mathbf{x}_i) = \arg \min_{j \in \{1,\ldots,c\}} \|\mathbf{x}_i - \mathbf{A}_j \boldsymbol{\alpha}_j\|_2.$$

Synthetic image

- The simulated image is generated with a size of $N = 128 \times 128$ pixels made up of linear mixtures between 3 components as follows:

$$\mathbf{x}_i = \sum_{k=1}^{c} \mathbf{m}_k \mathbf{s}_i^k + \mathbf{n}_i,$$

  where $\mathbf{s} = \{\mathbf{s}_1, \ldots, \mathbf{s}_N\}$ is the fractional abundances matrix which is generated according to a uniform distribution over the simplex.
- $\mathbf{m} = \{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3\}$ is the mixing matrix where the spectral signatures used were randomly obtained from the United States Geological Survey (USGS) digital library[a].
- Zero-mean Gaussian noise with variance $\sigma^2\mathbf{I}$, *i.e.*, $\mathbf{n}_i \sim (0, \sigma^2\mathbf{I})$ is added to our simulated image ($\sigma = 0.3182$ approx. SNR 5dB)
- Spectra highly mixed (all pixels have abundance fractions less than 0.5)

[a] http://speclab.cr.usgs.gov

Synthetic image

- The simulated image is generated with a size of $N = 128 \times 128$ pixels made up of linear mixtures between 3 components as follows:

$$\mathbf{x}_i = \sum_{k=1}^{c} \mathbf{m}_k \mathbf{s}_i^k + \mathbf{n}_i,$$

where $\mathbf{s} = \{\mathbf{s}_1, \ldots, \mathbf{s}_N\}$ is the fractional abundances matrix which is generated according to a uniform distribution over the simplex.

- $\mathbf{m} = \{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3\}$ is the mixing matrix where the spectral signatures used were randomly obtained from the United States Geological Survey (USGS) digital library[a].
- Zero-mean Gaussian noise with variance $\sigma^2\mathbf{I}$, *i.e.*, $\mathbf{n}_i \sim (0, \sigma^2\mathbf{I})$ is added to our simulated image ($\sigma = 0.3182$ approx. SNR 5dB)
- Spectra highly mixed (all pixels have abundance fractions less than 0.5)

[a] http://speclab.cr.usgs.gov

Synthetic image

- The simulated image is generated with a size of $N = 128 \times 128$ pixels made up of linear mixtures between 3 components as follows:

$$\mathbf{x}_i = \sum_{k=1}^{c} \mathbf{m}_k \mathbf{s}_i^k + \mathbf{n}_i,$$

where $\mathbf{s} = \{\mathbf{s}_1, \ldots, \mathbf{s}_N\}$ is the fractional abundances matrix which is generated according to a uniform distribution over the simplex.

- $\mathbf{m} = \{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3\}$ is the mixing matrix where the spectral signatures used were randomly obtained from the United States Geological Survey (USGS) digital library[a].
- Zero-mean Gaussian noise with variance $\sigma^2\mathbf{I}$, *i.e.*, $\mathbf{n}_i \sim (0, \sigma^2\mathbf{I})$ is added to our simulated image ($\sigma = 0.3182$ approx. SNR 5dB)
- Spectra highly mixed (all pixels have abundance fractions less than 0.5)

[a] http://speclab.cr.usgs.gov

Synthetic image

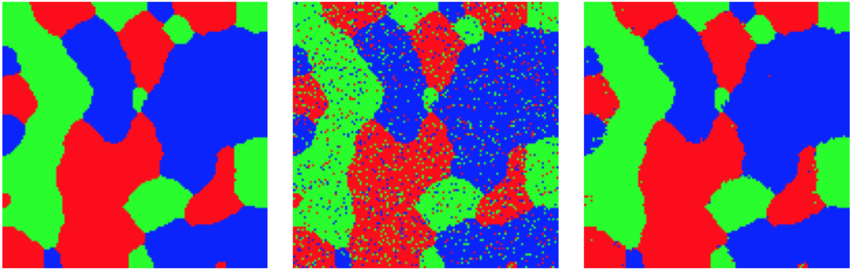
- The simulated image is generated with a size of $N = 128 \times 128$ pixels made up of linear mixtures between 3 components as follows:

$$\mathbf{x}_i = \sum_{k=1}^{c} \mathbf{m}_k \mathbf{s}_i^k + \mathbf{n}_i,$$

  where $\mathbf{s} = \{\mathbf{s}_1, \ldots, \mathbf{s}_N\}$ is the fractional abundances matrix which is generated according to a uniform distribution over the simplex.
- $\mathbf{m} = \{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3\}$ is the mixing matrix where the spectral signatures used were randomly obtained from the United States Geological Survey (USGS) digital library[a].
- Zero-mean Gaussian noise with variance $\sigma^2\mathbf{I}$, *i.e.*, $\mathbf{n}_i \sim (0, \sigma^2\mathbf{I})$ is added to our simulated image ($\sigma = 0.3182$ approx. SNR 5dB)
- Spectra highly mixed (all pixels have abundance fractions less than 0.5)

[a] http://speclab.cr.usgs.gov

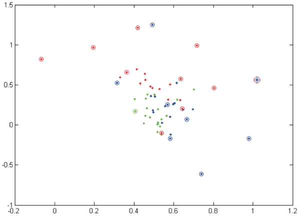
Synthetic image



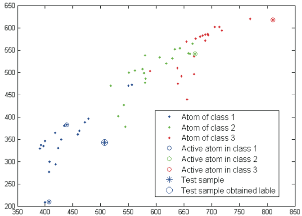
(a) (b) (c)

- (a) Image of class labels for a simulated data set made up with highly mixed pixels and noise
- (b) Sparse classification based on the original spectral information (OA=89.34%)
- (c) Sparse classification based on EMAPs (OA=99.11%)

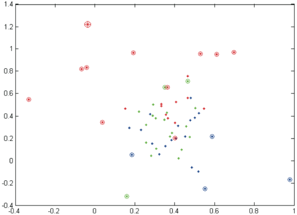
Synthetic image



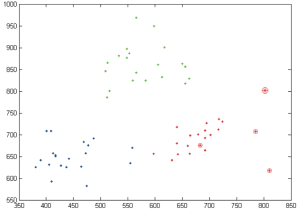
(a) Sparse representation of a sample from *blue* class (spectral space)



(b) Sparse representation of a sample from *blue* class (EMAP space)



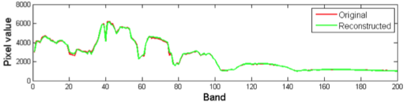
(c) Sparse representation of a sample from *red* class (spectral space)



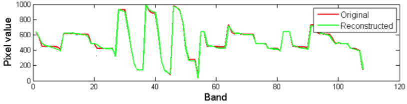
(d) Sparse representation of a sample from *red* class (EMAP space)

Sparse representation of two different test samples from a simulated hyperspectral scene in spectral and EMAP space

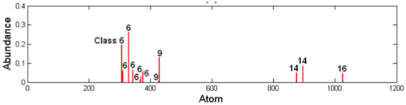
Details on the SRC



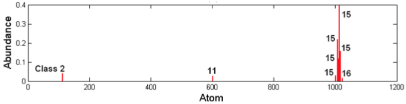
(a) Original vs reconstructed signature of a sample in class 15 (spectral space)



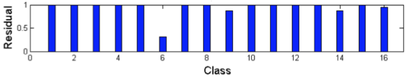
(b) Original vs reconstructed signature of a sample in class 15 (EMAP space)



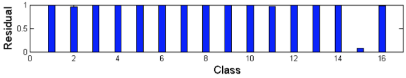
(c) Sparse representation (spectral space)



(d) Sparse representation (EMAP space)



(e) Residuals for each class (spectral space)



(f) Residuals for each class (EMAP space)

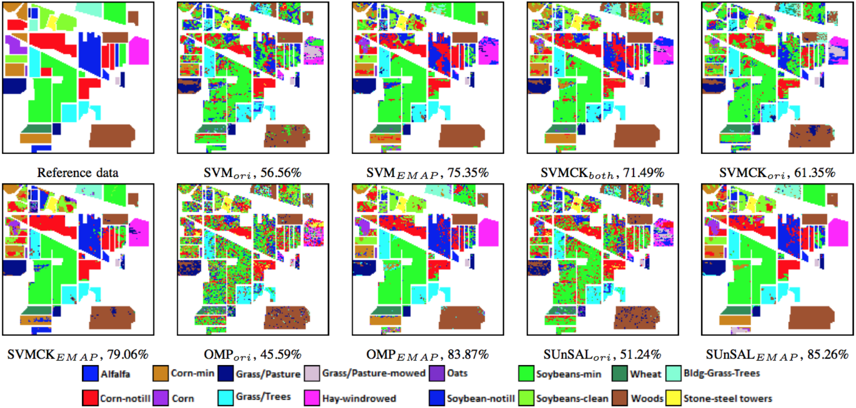
Sparse representation of a sample from a highly mixed class (building-grass-tree-drives, class number 15) of the AVIRIS Indian Pines data in spectral and EMAP space.

Numerical results

Class	Train	Test	SVM_{ori}	SVM_{EMAP}	$SVMCK_{both}$	$SVMCK_{ori}$	$SVMCK_{EMAP}$	OMP_{ori}	OMP_{EMAP}	$k\mathrm{NN}_{ori}$	$k\mathrm{NN}_{EMAP}$	$SUnSAL_{ori}$	$SUnSAL_{EMAP}$	$SUnSAL_{EMAP}(\tau = 0)$
Alfalfa	3	43	54.88	94.88	88.84	79.09	94.88	34.42	96.51	74.41	88.14	51.86	**97.67**	83.95
orn-notill	14	1414	42.63	68.06	60.19	50.92	69.34	39.05	72.23	40.62	60.58	40.83	**75.57**	15.57
Corn-min	8	822	29.53	60.22	64.57	41.24	**74.04**	22.71	66.34	28.44	54.45	22.21	68.22	26.13
Corn	3	234	17.65	35.73	31.92	35.38	44.27	8.63	**51.92**	24.52	26.03	12.52	47.01	27.35
Grass/Pasture	5	478	57.51	74.10	69.27	65.69	**76.86**	55.67	74.29	59.12	67.51	58.28	71.74	41.90
Grass/Trees	7	723	81.02	91.59	91.27	83.91	93.68	72.67	88.73	72.13	92.13	83.90	**95.09**	33.98
Grass/Pasture-mowed	3	25	86.40	95.20	93.20	96.40	97.60	61.60	**98.00**	88.00	95.20	77.60	96.80	87.60
Hay-windrowed	5	473	62.37	94.52	82.85	73.81	95.58	46.70	99.98	67.51	90.95	65.60	**100**	71.21
Oats	3	17	70.59	85.88	84.71	81.18	89.41	44.12	**97.06**	58.82	89.41	59.41	91.76	58.24
Soybeans-notill	10	962	39.73	75.60	66.65	51.19	79.59	25.34	83.37	43.67	66.67	29.18	**85.25**	26.72
Soybeans-min	25	2430	73.31	87.37	80.88	72.40	88.06	54.28	88.61	60.91	79.85	64.12	**92.26**	15.29
Soybeans-clean	6	587	21.72	57.68	41.93	29.40	59.93	18.94	**70.49**	27.10	43.20	18.93	68.18	42.54
Wheat	3	202	87.28	96.44	93.66	90.54	97.97	72.87	98.61	79.80	97.67	81.19	**99.50**	76.78
Woods	13	1252	84.03	97.32	91.73	85.93	96.45	68.95	95.67	75.80	89.03	82.72	**98.43**	6.88
Bldg-Grass-Trees	4	382	17.38	65.16	57.57	34.82	74.48	20.26	75.68	18.40	61.99	17.07	**77.88**	35.68
Stone-steel towers	3	90	70.67	86.89	89.44	89.22	**98.78**	65.11	95.11	80.33	93.78	85.44	95.44	81.11
OA			56.73	79.07	73.08	62.95	81.96	45.68	82.70	52.94	72.32	52.58	**84.90**	26.71
AA			56.04	79.17	74.29	66.32	83.18	44.46	84.54	56.23	74.79	53.18	**85.05**	45.81
κ			49.88	76.08	69.29	57.52	79.47	37.82	80.26	46.33	68.43	45.40	**82.76**	19.24
Time(s)			0.49	13.39	16.52	6.07	19.28	29.01	35.94	0.35	13.28	4.93	18.03	16.39

Overall accuracy (OA), average accuracy (AA), kappa statistic (κ) and class individual accuracies ([%]) obtained by different classifiers on the AVIRIS Indian Pines data (here, we use a total of 115 samples for training, which represents about 1% of the available labeled data for the scene).

Classification maps



Classification results obtained by different classifiers for the AVIRIS Indian Pines scene (using a total of 115 samples for training, which represents about 1% of the available labeled data for the scene).

Conclusion

- We have proposed a new classification strategy that integrates sparse representations and extended multi-attribute profiles (EMAPs) for spatial-spectral classification of remote sensing data.
- The proposed approach can appropriately exploit the inherent sparsity present in EMAPs in order to provide state-of-the-art classification results.
- This is mainly due to the fact that the samples in EMAP space can be approximately represented by a few number of atoms in the training dictionary after solving the optimization problem, while the same samples could not be represented in the original spectral space with the same level of sparsity.
- A comparison with state-of-the-art classifiers shows very promising results for the proposed approach, particularly when a very limited number of training samples is available.