Unsupervised Feature Extraction for Hyperspectral Images Using Combined Low Rank Representation and Locally Linear Embedding

Mengdi Wang[1] Jing Yu[2] Lijuan Niu[3] Weidong Sun[1]

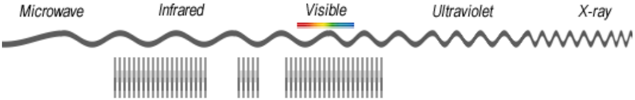
[1]Tsinghua University, China
wangmd12@mails.tsinghua.edu.cn; wdsun@tsinghua.edu.cn
[2]Beijing University of Technology, China
jing.yu@bjut.edu.cn
[3]Chinese Academy of Medical Sciences, China

ICASSP 2017

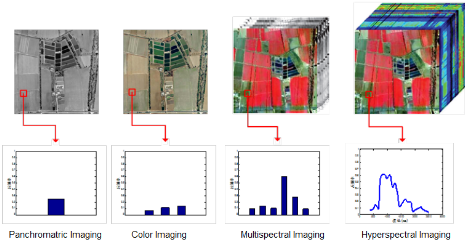
Outline

What is hyperspectral images?

- Hyperspectral images (HSIs)
  - captured by the remote sensing platforms
  - contain hundreds of bands across the spectral dimension

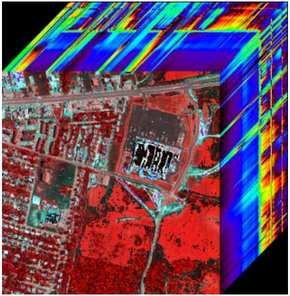


  - can provide not only spatial but also spectral information of the land-covers in a scene

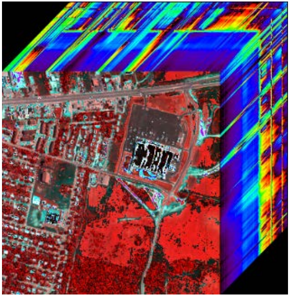


Applications and Problems of HSIs

- Applications of HSIs
  - agriculture
  - environment
  - monitoring
  - food safety
  - medicine
  - mineralogy
  - *etc.*
- Problems
  - hundreds of bands
  - curse of dimensionality



Applications and Problems of HSIs



- Applications of HSIs
  - agriculture
  - environment
  - monitoring
  - food safety
  - medicine
  - mineralogy
  - *etc.*
- Problems
  - hundreds of bands
  - curse of dimensionality

Feature extraction and dimension reduction for HSIs

Related Works

- Widely used methods
  - PCA, ICA, MNF
  - LLE, ISOMAP, Laplacian Eigenmap
- HSI-specified methods based on the endmember mixing nature
  - VCA (vertex component analysis)
  - MVC-NMF (minimum volume constrained nonnegative matrix factorization)
- Recently works
  - OTVCA (orthogonal total variation component analysis), *TGRS'16*
  - IR (Intrinsic Representation), *TGRS'16*

Structure of the spectral space in HSIs

- The spectral space in HSIs can be divided into several subspaces according to the land-covers$\{\mathcal{S}_c\}_{c=1}^{C}$, and $\mathcal{S}_{c_1} \bigcap \mathcal{S}_{c_2}\,(c_1 \neq c_2) = \emptyset$
- The spectral space $\mathcal{S}$ can be represented by $\mathcal{S} = \bigcup_{c=1}^{C} \mathcal{S}_c$
- The spectral vectors in each class share high similarity, thus $\mathcal{S}_c$ should be low-rank.
- The spectral space in HSIs is a union of multiple low-rank subspaces.

Structure of the spectral space in HSIs

- The spectral space in HSIs can be divided into several subspaces according to the land-covers$\{\mathcal{S}_c\}_{c=1}^{C}$, and $\mathcal{S}_{c_1} \bigcap \mathcal{S}_{c_2}\, (c_1 \neq c_2) = \emptyset$
- The spectral space $\mathcal{S}$ can be represented by $\mathcal{S} = \bigcup_{c=1}^{C} \mathcal{S}_c$
- The spectral vectors in each class share high similarity, thus $\mathcal{S}_c$ should be low-rank.
- The spectral space in HSIs is a union of multiple low-rank subspaces.

An informative data representation when used for FE should:

1. preserve the subspace-inherent structures
2. minimize the inter-subspace components

Framework of LRR

- Assume $\mathbf{X} \in \bigcup_{c=1}^{C} \mathcal{S}_c$ and $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_C]$, $\mathbf{X}_c \in \mathcal{S}_c$
- If there is a structured dictionary $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_C]$, $\mathbf{A}_c \in \mathcal{S}_c$
- Then if $\mathbf{X}$ is modelled as,

$$\min_{\mathbf{Z}} \ \mathrm{rank}(\mathbf{Z}) \quad \text{s.t. } \mathbf{X} = \mathbf{A}\mathbf{Z}$$

- Rank constraint on $\mathbf{Z}$ will lead to

$$\mathbf{Z}^* = \begin{bmatrix} \mathbf{Z}_1^* & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2^* & \mathbf{0} & \mathbf{0} \\ & & \ddots & \\ \mathbf{0} & \mathbf{0} & & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Z}_C^* \end{bmatrix},$$

- Dictionary selection: $\mathbf{A} = \mathbf{X}$

Unsupervised FE using LRR

FE model:

$$\min_{\mathbf{Z},\mathbf{E}} \operatorname{rank}(\mathbf{Z}) + \lambda \|\mathbf{E}\|_{2,0} \quad \text{s.t. } \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}$$

- **E** is constituted by the vectors that has the inter-subspace components
- Number of such vectors should be small
- Thus the column-sparse constraint $\ell_{2,0}$ norm is used.

Convex model,

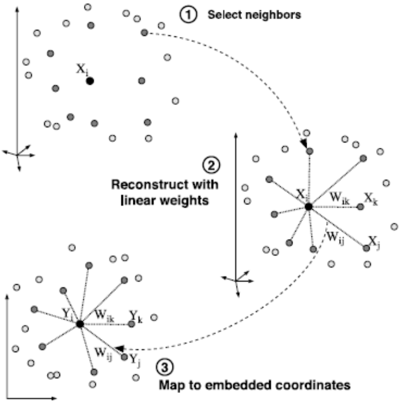
$$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_1 \quad \text{s.t. } \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}$$

Solved using the inexact augmented Lagrange multiplier(IALM) method.

Spatial constraint using LLE

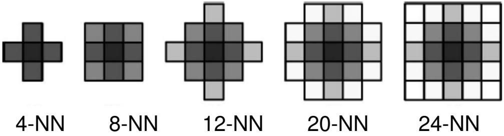
Introduce the spatial similarity in the FE procedure based on locally linear embedding (LLE)

1. Select the neighbors
2. construct the topology structure within the neighborhood in the original feature space
3. preserve this topology structure in the extracted feature space



Procedure of LLE

1. Select the neighbors



2. Construct the topology structure using the quadratic fit,

$$\{W_{ij}\} = \arg\min_{W_{ij}} \|\mathbf{X}_i - \sum_j W_{ij}\mathbf{X}_j^{(i)}\|_F^2$$

3. Preserve this topology in the extracted feature space

$$L = \sum_i \|\mathbf{Y}_i - \sum_j W_{ij}\mathbf{Y}_j^{(i)}\|_F^2 = \mathrm{Tr}\left(\mathbf{Y}\left(\mathbf{I}-\mathbf{W}\right)^T\left(\mathbf{I}-\mathbf{W}\right)\mathbf{Y}^T\right)$$

$[\mathbf{W}]_{ij}$ being W_{ij} if $\mathbf{X}_j$ is a neighbour of $\mathbf{X}_i$ and 0 if not

Combine LRR and LLE for unsupervised FE

- LRR framework

$$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_1 \quad \text{s.t. } \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}$$

- $\mathbf{Z}_i$ is actually the transform of $\mathbf{X}_i$ in the self-representation domain, therefore $\mathbf{Z}_i$ should preserve the same neighborhood topology structure as $\mathbf{X}_i$,

$$\mathrm{Tr}\left(\mathbf{Z}\left(\mathbf{I}-\mathbf{W}\right)^{\mathrm{T}}\left(\mathbf{I}-\mathbf{W}\right)\mathbf{Z}^{\mathrm{T}}\right)$$

- The combined LRR and LLE for unsupervised FE is,

$$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_1 + \frac{\beta}{2}\mathrm{Tr}\left(\mathbf{Z}\left(\mathbf{I}-\mathbf{W}\right)^{\mathrm{T}}\left(\mathbf{I}-\mathbf{W}\right)\mathbf{Z}^{\mathrm{T}}\right) \quad \text{s.t. } \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}$$

Combine LRR and LLE for unsupervised FE

- The combined LRR and LLE for unsupervised FE is,

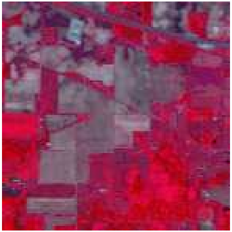
$$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_1 + \frac{\beta}{2}\mathrm{Tr}\left(\mathbf{Z}\left(\mathbf{I}-\mathbf{W}\right)^{\mathrm{T}}\left(\mathbf{I}-\mathbf{W}\right)\mathbf{Z}^{\mathrm{T}}\right)$$
$$\text{s.t. } \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}$$

- The structural extracted features are $\hat{\mathbf{X}} = \mathbf{X}\mathbf{Z}^*$
- The dimension remains unchanged, so the PCA is adopted to reduce the dimension.

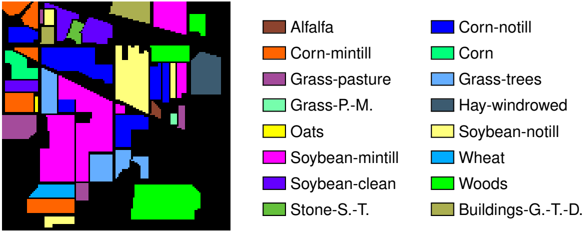
Experiments set-up

- Evaluation way
  - The following classification task is used as evaluation way
  - Support vector machine (SVM) with the radial basis function (RBF).
- Datasets
  - AVIRIS data: *Indian Pines*, $145 \times 145 \times 200$
  - ROSIS data: *Pavia University*, $610 \times 340 \times 103$
- Compared methods
  - PCA, ICA
  - MVC-NMF (TGRS'07)
  - IR (TGRS'16)
  - LLE
  - LRR

Indian Pines

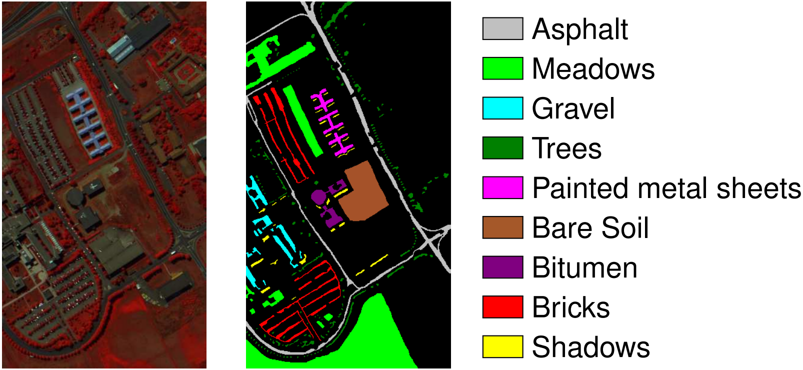


(a) False color



(b) Ground truth

Pavia University



(a) False color (b) ground truth

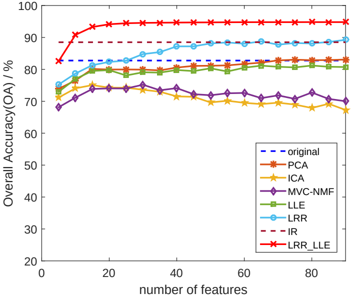
Classification results

	Indian Pines			Pavia University		
	Reduce dimension: 20 *Training set: 10%*			*Reduce dimension: 15* *Training set: 1%*		
	OA	AA	kappa	OA	AA	kappa
original	82.76	80.76	0.8034	88.31	90.45	0.8479
PCA	79.95	79.87	0.7712	74.47	82.35	0.6777
ICA	74.27	70.71	0.7057	83.27	87.38	0.7840
MVC-NMF	74.04	71.12	0.7023	82.96	85.78	0.7775
LLE	79.76	77.47	0.7694	87.77	90.04	0.8411
LRR	82.34	78.47	0.7984	91.22	92.34	0.8852
IR	88.5	88.1	0.869	93.1	94.3	0.909
LRR_LLE	**94.13**	**93.30**	**0.9330**	**95.03**	**95.43**	**0.9345**

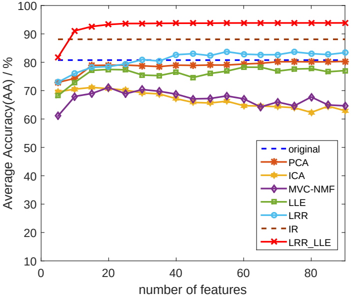
Classification results w.r.t. feature dimension

Indian Pines

Number of training sets is fixed.



(a) OA

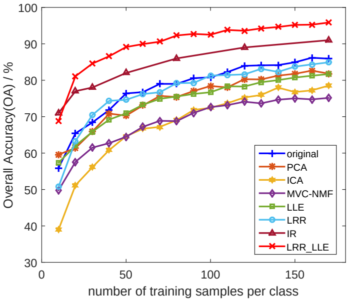


(b) AA

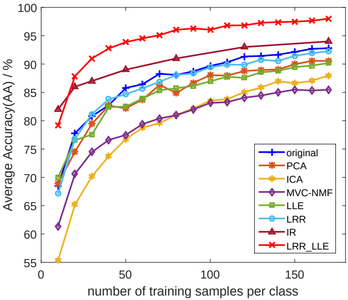
Classification results w.r.t. number of training samples

Indian Pines

Reduced dimension is fixed.



(a) OA



(b) AA

- We proposed a novel unsupervised feature extraction method using combined LRR and LLE:
  - LRR is capable to structurally represent the union spectral space of multiple low-rank subspaces, therefore can help preserve the subspace-inherit components;
  - LLE is a nonlinear dimension reduction method, help to preserve the locally geometric manifold in the spatial domain;
  - The combination model can simultaneously employ the spectral correlation and the locally spatial correlation information during the FE procedure.
- Experiments with a following classification task using SVM show that the proposed method LRR_LLE outperforms the state-of-art methods when used for unsupervised FE in HSIs

Thank you

wangmd12@mails.tsinghua.edu.cn
jing.yu@bjut.edu.cn
wdsun@tsinghua.edu.cn