Dimension Reduction on Hyperspectral Images

Meiching Fong
August 31, 2007
UCLA Department of Mathematics

Project supervisor: Prof. Todd Wittman
Work supported by NGA NURI grant HM1582-06-BAA-004

Q1: Which is the best dimension reduction method for hyperspectral images?

(Traditional linear techniques)

- **PCA** **FastICA**

(global nonlinear techniques) : preserves the global structure, and pairwise distances.

- **Diffusion maps**

(local nonlinear techniques): preserves properties of small neighborhood around the data.

- **Laplacian Eigenmaps** **LLE** **LTSA**

(extensions and variants of local nonlinear)

- **LLTSA**

Code taken from the Dimension Reduction toolbox by Laurents van der Maaten

FastICA from Laboratory of Computer and Information Science (CIS). Helsinki University of Technology

Methods we do not consider:

- Requires more input parameters that do not come with original data:

Linear Discriminant Analysis (LDA) *Generalized Discriminant Analysis (G*

Linearity Preserving Projection (LPP) *Locally Linear Coordination (LLC)*

- Takes more than ten minutes to run on just 50 x 50 images:

Stochastic Proximity Embedding (SPE) Stochastic Neighbor Embedding (SNE

Kernel PCA Coordina

Factor Analysis (CFA)

- Fails to have connected neighborhood for hyperspectral images:

Fast Maximum Variance Unfolding (FastMVU)

- Reshape the patterns of the data fed in, thus unable to compare the result:

Hessian Local Linear Embedding (HLLE)

Q2: How do we measure the performance?

- Classification Rate

Original data → **Dimension reduction**

Kmeans → **Check permutation of the Ground truth**

- Mean Improvement

Dimension reduced — Original data

+ : Improvement **—: Declination**

- Standard Deviation

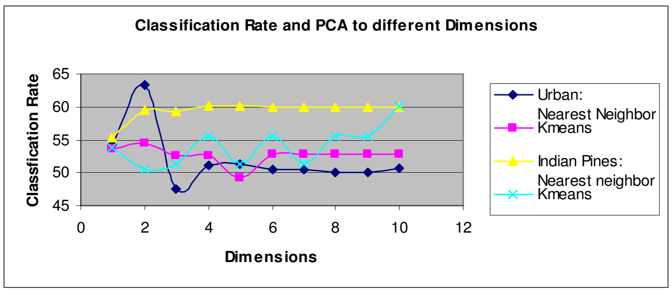
- Runtime

What is the Intrinsic Dimension?

Method	Clean Urban (162)	Noisy Urban (210)	Clean Pines (179)	Noisy Pines (220)
CorrDim Estimator	2.373	3.981	2.032	8.583
NearNb Estimator	4.974	4.981	3.627	4.319
Maximum likelihood	19.568	19.545	12.284	20.980
Eig Value	3	4	2	2
Packing Numbers	1.925	2.893	11.363	5.094
Geodesic Min Spanning Tree	9.374	13.261	4.315	133.048

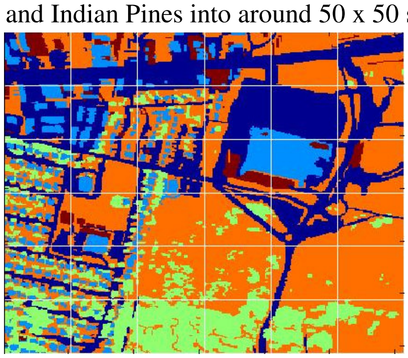
They do not agree.

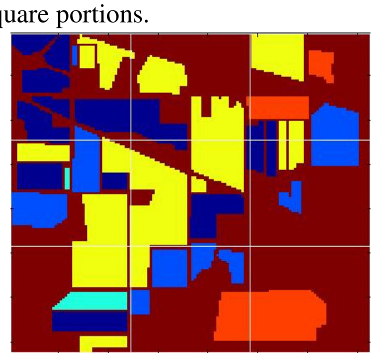
Classification rate after PCA reducing to different dimensions on the noisy datasets

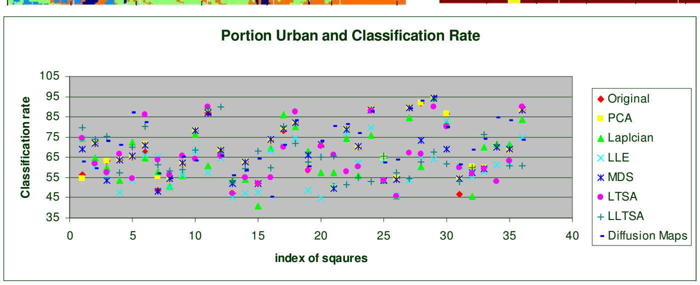


The graph generally flats out at 4 to 6 dims for our data.

Since most of the methods cannot handle large data, we divided Urban and Indian Pines into around 50 x 50 square portions.

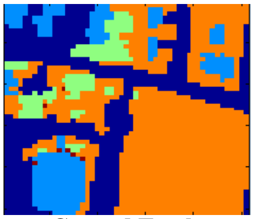
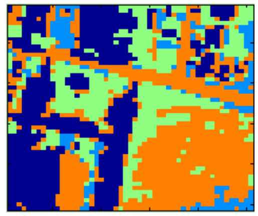
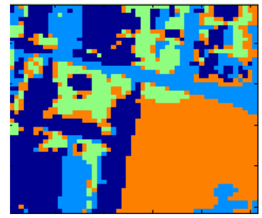






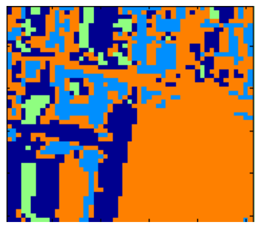
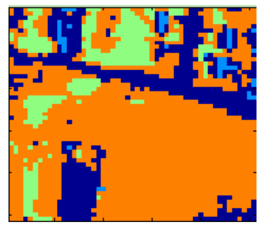
The result is not always consistent.

Classification rate of a portion of clean Urban dataset with some example of dimension reduction.

Ground Truth **No dimension reduction** **PCA 54.47 %**

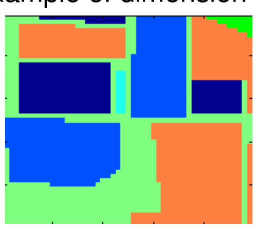
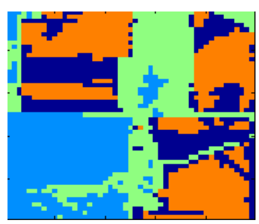
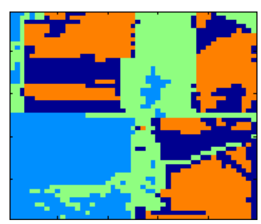
42.56 %

Diffusion map 58.13 % **LTSA 52.51%**

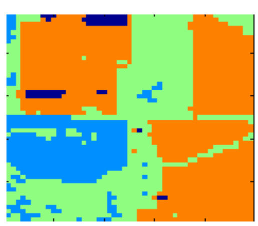
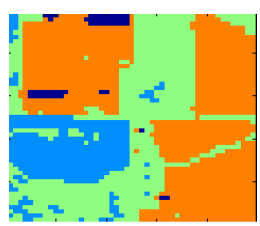
6.32 sec **17.56 sec**

Classification rate of a portion of clean Pines dataset with some example of dimension reduction.

Ground Truth **No dimension reduction** **PCA 62.29 %**

62.23 %

Diffusion map 64.96 % **LTSA 69.89 %**

5.47 sec **13.4**

Euclidean Distance :

	Clean Urban 4 Dimensions			
Method	Classify-Rate	Mean Improvement	Std.	Runtime
No DR	61.481	---------	12.007	---------
PCA	61.109	-0.372	12.949	0.090
Laplacian	57.775	-3.706	12.060	5.635
LLE	66.143	4.662	14.555	19.074
LTSA	67.426	5.945	13.725	17.344
LLTSA	64.699	3.219	11.529	17.743
Diff-Maps	66.203	4.723	12.519	6.750
	Clean Indian Pines 4 Dimensions			
Method	Classify-Rate	Mean Improvement	Std.	Runtime
No DR	72.793	---------	9.939	--------
PCA	72.826	0.033	7.664	0.187
Laplacian	72.290	-0.503	13.685	4.899
LLE	81.139	8.346	15.999	14.024
LTSA	70.124	-2.669	10.091	13.356
LLTSA	78.901	6.108	13.529	13.294
Diff-Maps	83.612	10.818	13.602	5.430
FastICA	73.063	0.270	15.389	4.179

Cosine and Correlation Distance on Clean Indian Pines:

	Clean Pines 4 Dimensions Cosine				Clean Pines 4 Dimensions Correlation			
Method	Classify-Rate	Mean Improvement	Std.	Runtime	Classify -Rate	Mean Improvement	Std.	Runtime
No DR	77.925	--------	12.327	--------	80.386	--------	13.782	--------
PCA	74.615	-3.310	9.265	0.116	79.617	-0.769	12.474	0.136
Laplacian	76.661	-1.263	12.562	4.838	72.888	-7.498	12.190	5.240
LLE	79.209	1.285	12.696	16.363	75.823	-4.563	16.707	18.995
LTSA	82.221	4.296	13.869	15.124	82.914	2.528	12.054	17.074
LLTSA	77.972	0.048	11.692	14.784	LLTSA and Diff-Maps produced errors in original code when running with correlation distance.			
Diff-Maps	68.736	-9.189	18.085	2.836				
FastICA	74.438	-3.4872	13.233	28.414	74.465	-5.921	13.430	24.231

More trials show Diff-Maps gives the highest classification rate under Euclidean distance and when the data is as original without variations. However, Diffusion Maps has a higher standard deviation which implies it is not a very stable method.

Wavelet Transform on Clean Pines 4 dimensions:

	Do wavelet transforms before Dimension Reduction			
Method	Classify-Rate	Mean Improvement	Std.	Runtime
No DR	72.847	--------	16.196	--------
PCA	70.257	-2.590	7.722	0.131
Laplacian	70.329	-2.519	13.047	4.895
LLE	69.984	-2.864	10.160	15.788
LTSA	**78.54**	**5.693**	12.075	14.773
LLTSA	70.560	-2.287	16.581	16.360
Diff-Maps	76.127	3.279	19.206	3.687
FastICA	76.160	3.313	17.995	3.968

Spatial Coherent on 3 x 3 Neighborhood.

(Mohan, Sapiro and Bosch April 2007)

1	2	3
4	**X**	5
6	7	8

→

X	1	2	3	4	5	6	7	8

	First apply PCA then apply 3 x3 spatial coherent			
Method	Classify-Rate	Mean Improvement	Std.	Runtime
No DR	68.302	--------	8.919	--------
PCA	69.574	1.272	14.956	0.046
Laplacian	78.005	9.704	14.695	4.466
LLE	**80.682**	**12.38**	12.028	19.376
LTSA	78.414	10.112	15.431	16.450
LLTSA	75.094	6.792	15.547	15.606
Diff-Maps	68.614	0.312	13.704	11.171

Result:

- PCA is the fastest with a relatively low standard deviation.
- Diffusion maps gives the best result under Euclidean distance but it is unstable under data variation.
- LTSA also gives pretty good performance with a slightly lower standard deviation but it has a higher runtime.

Future Work:

- A more stable clustering technique other than Kmeans.
- To improve classification and ground truth comparison.