HYPERSPECTRAL IMAGING

CONTENTS

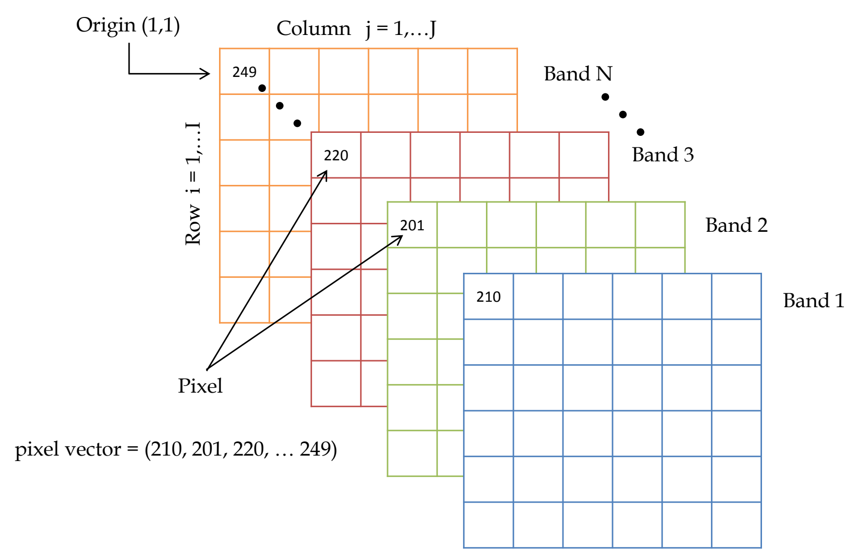
Hyperspectral Imagery

- Simultaneous acquisition of hundreds of spectral wavelengths for each image pixel.
- Bands are continuous, regularly spaced.
- Spectral resolution ~10nm.

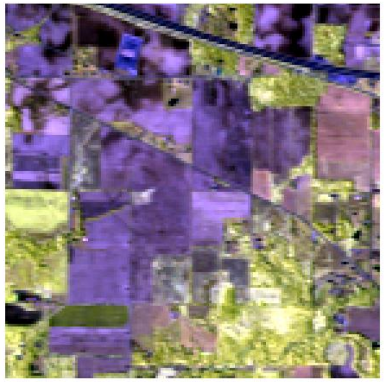
Why Hyperspectral Imagery

- Possibility of more accurate discrimination among materials of interest.
- Higher spectral resolution, thus giving the opportunity to push further the information extraction capability.
- Inter class and intra class delineation.
- Lots of application in diverse fields.

Hyperspectral Imagery

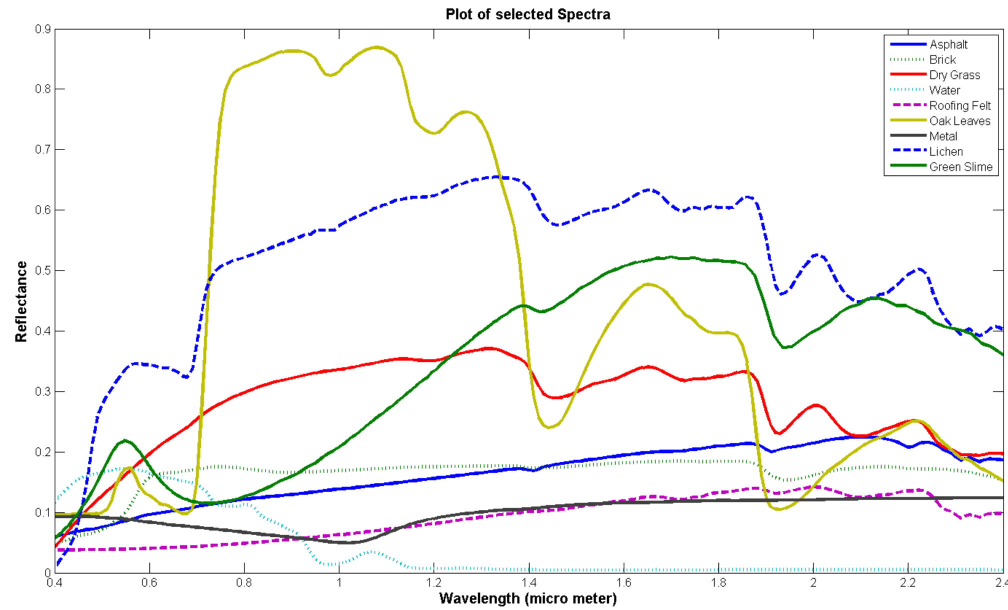


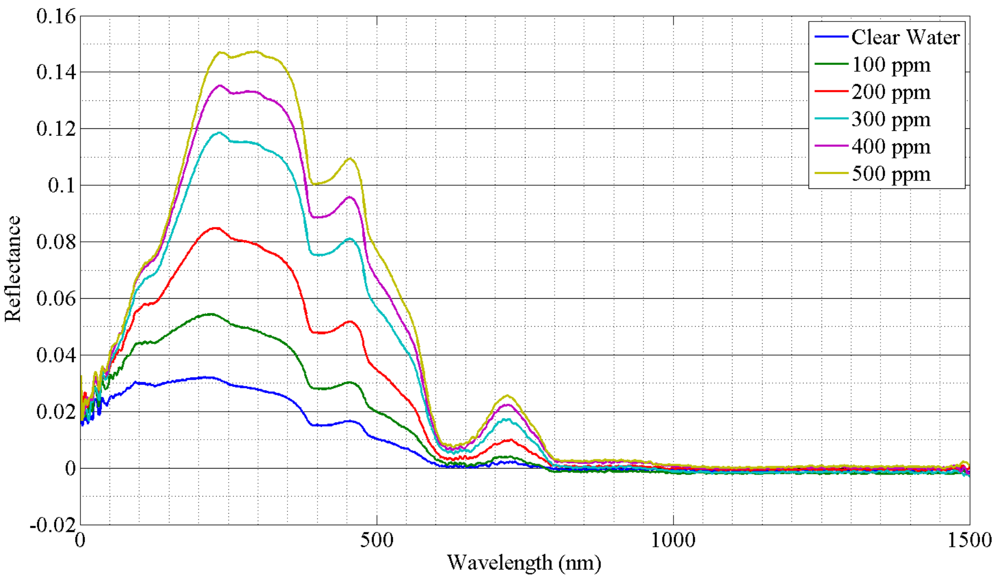
191 Band HYDICE image of Washington DC Mall



220 Band AVIRIS image Indian Pine Test Site

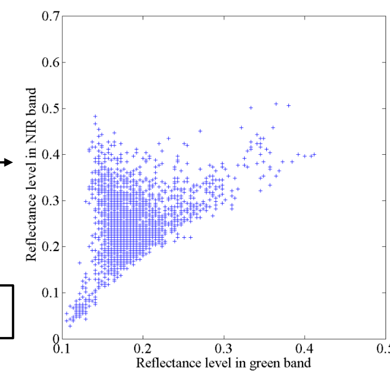
https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html

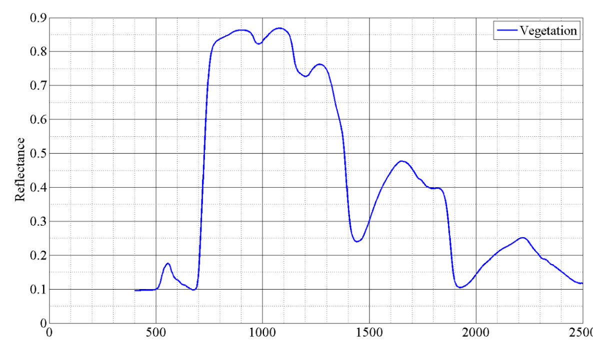




Image/Spatial space

Feature/Spectral space

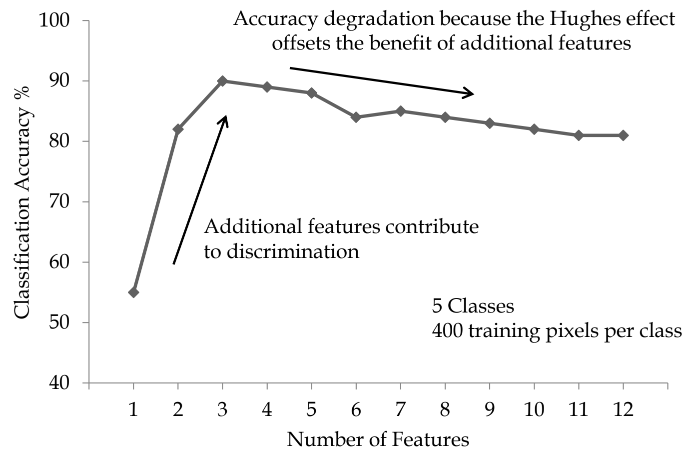




Single pixel vegetation spectrum

Issues

- Huge quantity of data to store.
- Computationally intensive.
- Exhibit redundant information.
- *Curse of Dimensionality*: With a limited training set, beyond a certain limit, the classification accuracy actually decreases as the number of features increases.



Fu *et al.* (1969); Richards (2013)

Small Training Sample Size

- GML
  - Effectiveness: reliable estimation of mean vector and covariance matrix for each class
  - Sufficient training sample for each class
- *N* dimensional spectral space
  - Mean vector: *N* elements
  - Covariance matrix: $N(N+1)/2$ elements

Clustering High Dimensional Data

- Problem 1: Visualization and tabularization of data becomes extremely difficult.

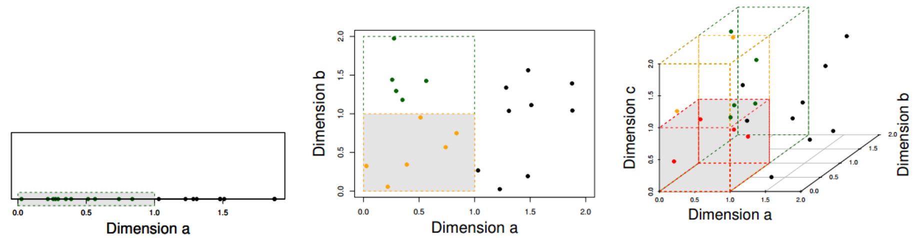
Kriegel *et al.* (2009)

Clustering High Dimensional Data

- Problem 1: Visualization and tabularization of data becomes extremely difficult.
- Problem 2: Discrimination between nearest and farthest point in high dimension becomes poor (Beyer *et al.* 1999).

Kriegel *et al.* (2009)

Curse of Dimensionality



(a) 11 Objects in One Unit Bin (b) 6 Objects in One Unit Bin (c) 4 Objects in One Unit Bin

Parsons, Haque and Liu (2004)

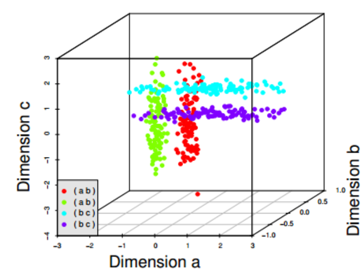
Clustering High Dimensional Data

- Problem 1: Visualization and tabularization of data becomes extremely difficult.
- Problem 2: Discrimination between nearest and farthest point in high dimension becomes poor (Beyer *et al.* 1999).
- Problem 3: Subset of some bands may be relevant to one cluster and subset of some other bands may be relevant to other cluster, and so on.

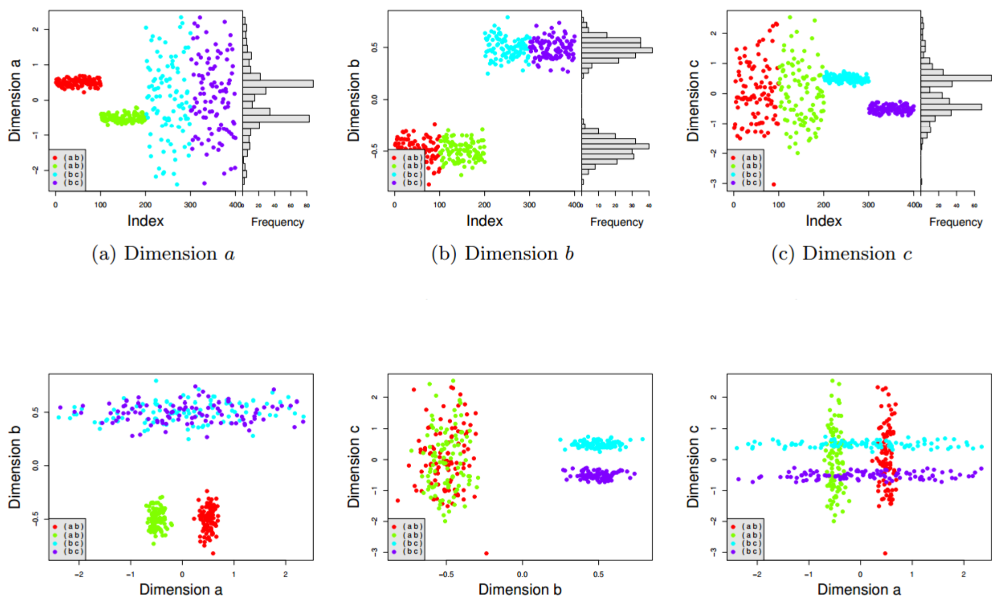
Kriegel *et al.* (2009)

Example

Parsons, Haque and Liu (2004)



400 data points
4 clusters of *100* data points each
Normally distributed, $\mu = 0.5$ and -0.5 in dim *a* and 0.5 in dim *b*, and $\sigma = 0.2$, dim c $\mu = 0$ and $\sigma = 1$
Other two clusters in dim *b* and c (same manner)



(a) Dimension a (b) Dimension b (c) Dimension c

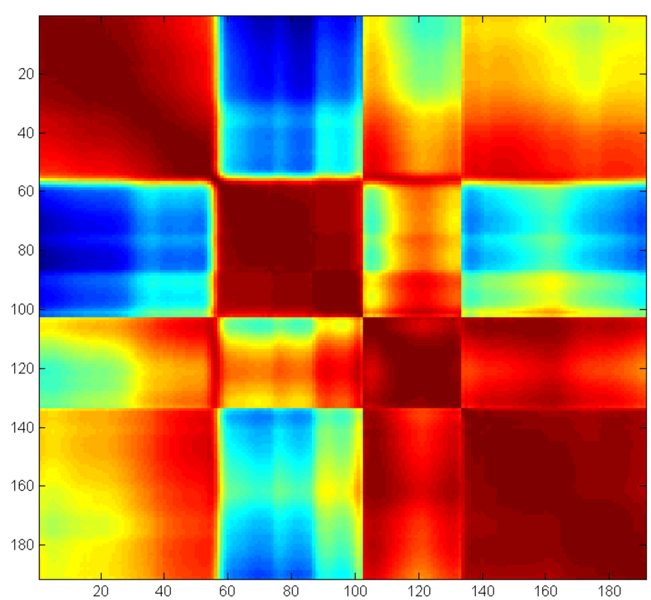
(a) Dims a & b (b) Dims b & c (c) Dims a & c

Parsons, Haque and Liu (2004)

Clustering High Dimensional Data

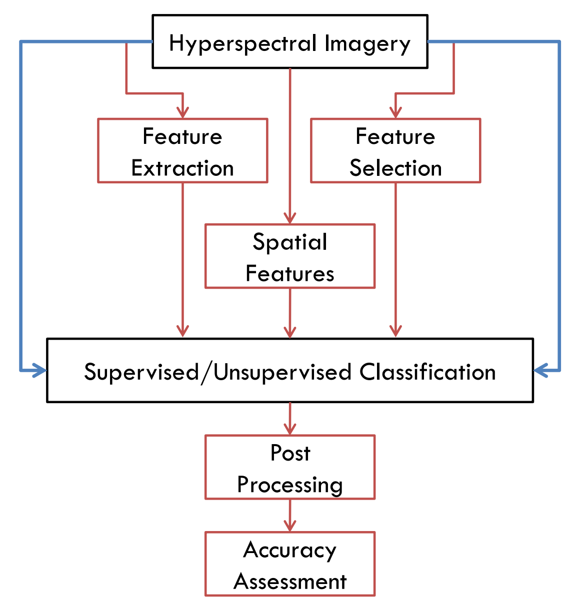
- Problem 1: Visualization and tabularization of data becomes extremely difficult.
- Problem 2: Discrimination between nearest and farthest point in high dimension becomes poor (Beyer *et al*. 1999).
- Problem 3: Appropriate subset of bands to describe the similarity of pixel belonging to the same group and possibly different subsets of bands for different groups of pixel.
- Problem 4: Correlation among bands.

Kriegel *et al.* (2009)



Color coded visualization of correlation matrix. Red color indicates high correlation and blue color indicates low correlation.

General Framework



Feature Reduction

- Original feature space: not necessarily most effective.
- Not designed for any particular problem.
- Redundant information: contiguous bands highly correlated.
- Classification cost increases with the number of features.
- Small training sample size issue.

Feature Selection

- Subset of spectral bands
- Advantages
  - Data Transmission
  - Interpretability of the results
  - Extrapolation of results
- Limitations
  - Information loss: features are completely discarded
  - Computationally intensive

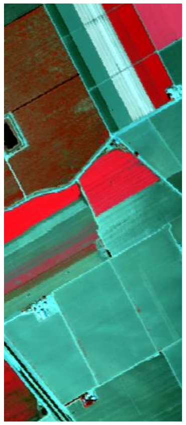
Feature Extraction

- Redistribution of information: higher dimensional to lower dimensional space without loss of significant amount of information
- Advantages
  - Can use all the available features
- Limitations
  - Transformed features losses physical meaning
  - Large number of features are noisy/irrelevant

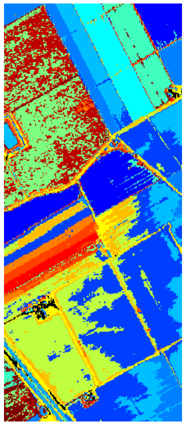
Classification

Examples

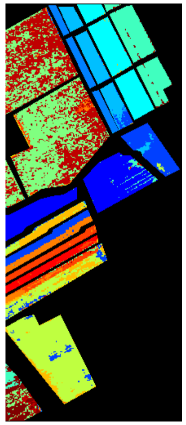
Clustering



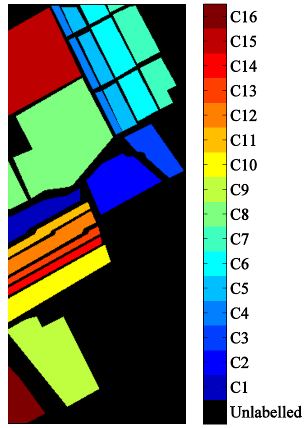
Salinas



k-means



Masked

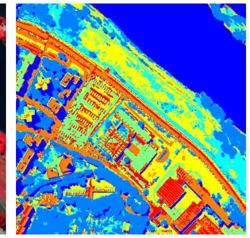


Reference Image

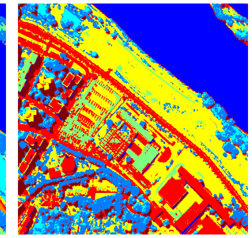
	Purity	NMI
k-means	0.6963	0.7453

Clustering

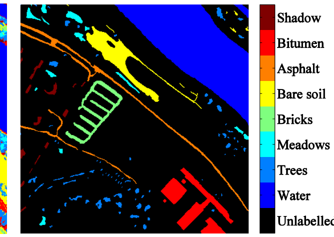
Pavia City



k-means



k-means + PCA



Reference Image

	Purity	NMI
k-means	0.8268	0.7598
k-means + PCA	0.8484	0.7677

Clustering Evaluation

Purity[1]

$$purity(\Omega,\mathbb{C}) = \frac{1}{N}\sum_k \max_j |\omega_k \cap c_j|$$

Normalized Mutual Information[1,2] (NMI)

$$NMI(\Omega,\mathbb{C}) = \frac{MI(\Omega,\mathbb{C})}{\max(H(\Omega),H(\mathbb{C}))}$$

$$MI(\Omega,\mathbb{C}) = \sum_k\sum_j p(\omega_k \cap c_j)\log_2\frac{p(\omega_k \cap c_j)}{p(\omega_k)p(c_j)}$$

$$H(\Omega) = -\sum_k p(\omega_k)\log_2 p(\omega_k)$$

$\Omega = \{\omega_1,\omega_2,\ldots,\omega_k\}$ is the set of clusters

$\mathbb{C} = \{c_1,c_2,\ldots,c_j\}$ is the set of classes obtained from reference image

N is total number of pixel samples

[1] http://nlp.stanford.edu/IR-book/html/htmledition/evaluation-of-clustering-1.html
[2] Cai, D., X. He, and J. Han. 2005. "Document Clustering Using Locality Preserving Indexing." *IEEE Transactions on Knowledge and Data Engineering*, 17 (12): 1624-1637.

$$\Omega = \begin{bmatrix} 1 & 1 & 1 & 2 \\ 2 & 2 & 2 & 1 \end{bmatrix}; \mathbb{C} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \end{bmatrix}$$

$$purity = 0.7500$$
$$NMI = 0.1888$$

Purity

$$purity(\Omega,\mathbb{C}) = \frac{1}{N}\sum_k \max_j |\omega_k \cap c_j|$$

Normalized Mutual Information (NMI)

$$NMI(\Omega,\mathbb{C}) = \frac{MI(\Omega,\mathbb{C})}{\max(H(\Omega),H(\mathbb{C}))}$$

$$MI(\Omega,\mathbb{C}) = \sum_k\sum_j p(\omega_k \cap c_j)\log_2\frac{p(\omega_k \cap c_j)}{p(\omega_k)p(c_j)}$$

$$H(\Omega) = -\sum_k p(\omega_k)\log_2 p(\omega_k)$$

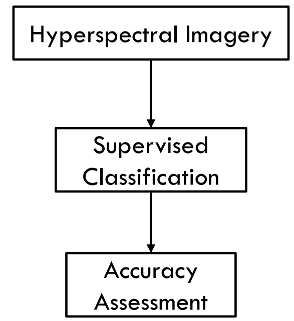
$\Omega = \{\omega_1,\omega_2,\ldots,\omega_k\}$ is the set of clusters

$\mathbb{C} = \{c_1,c_2,\ldots,c_j\}$ is the set of classes obtained from reference image

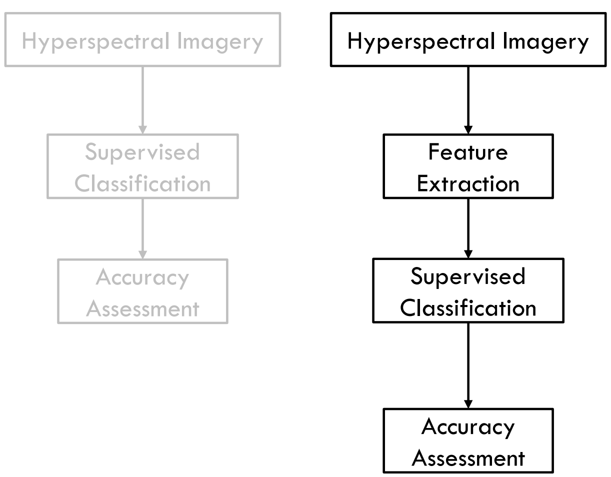
N is total number of pixel samples

Application of clustering in Supervised Classification

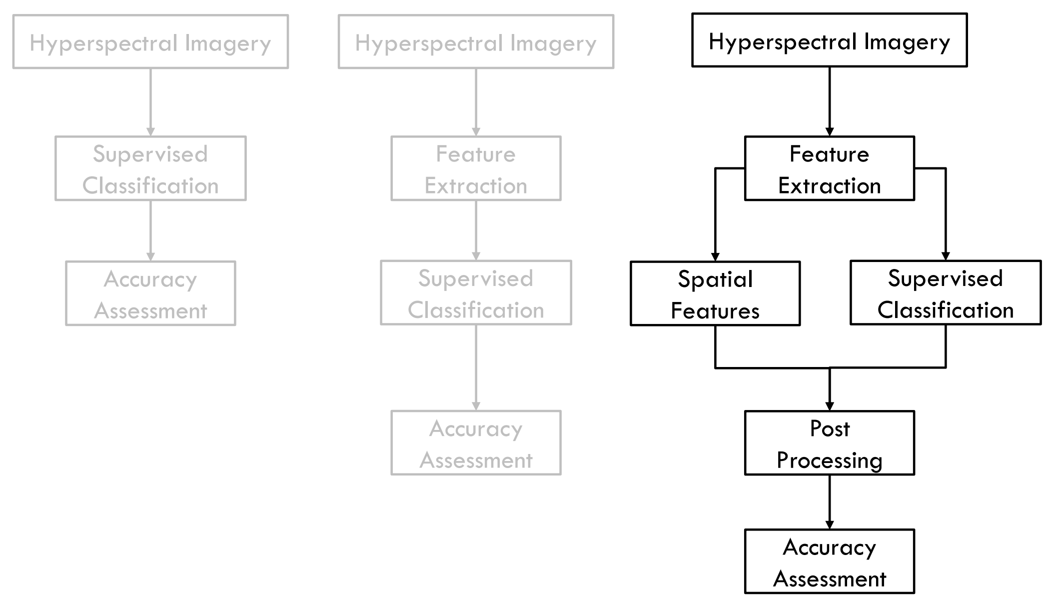
Classification Strategy I

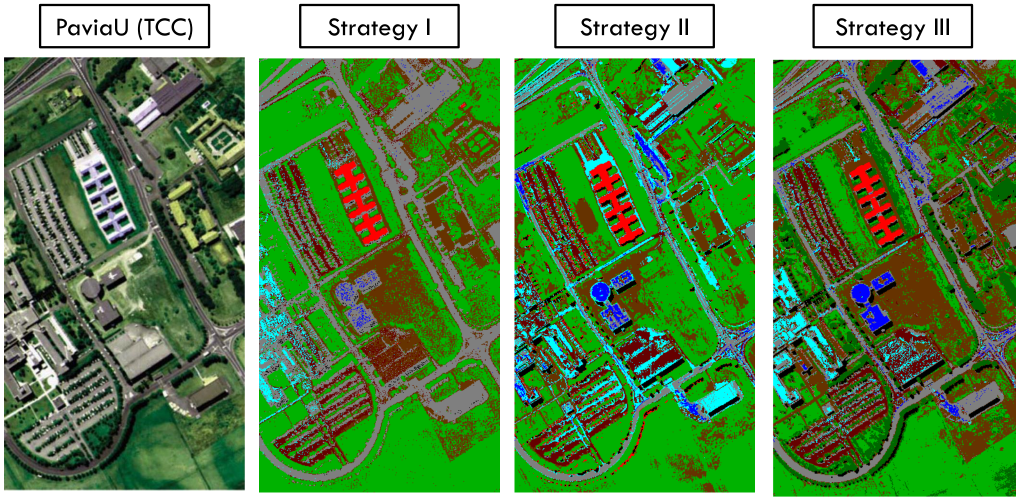


Classification Strategy II



Classification Strategy III





	Strategy I	Strategy II	Strategy III
Overall Kappa (%)	57.98	73.67	82.80

Thank You