Target Class Oriented Subspace Detection for Effective Hyperspectral Image Classification

Presented By
Md. Tanvir Ahmed
Roll: 123086

Supervised by
Dr. Md. Ali Hossain
Assistant Professor
Department of CSE

December 9, 2017

Outline

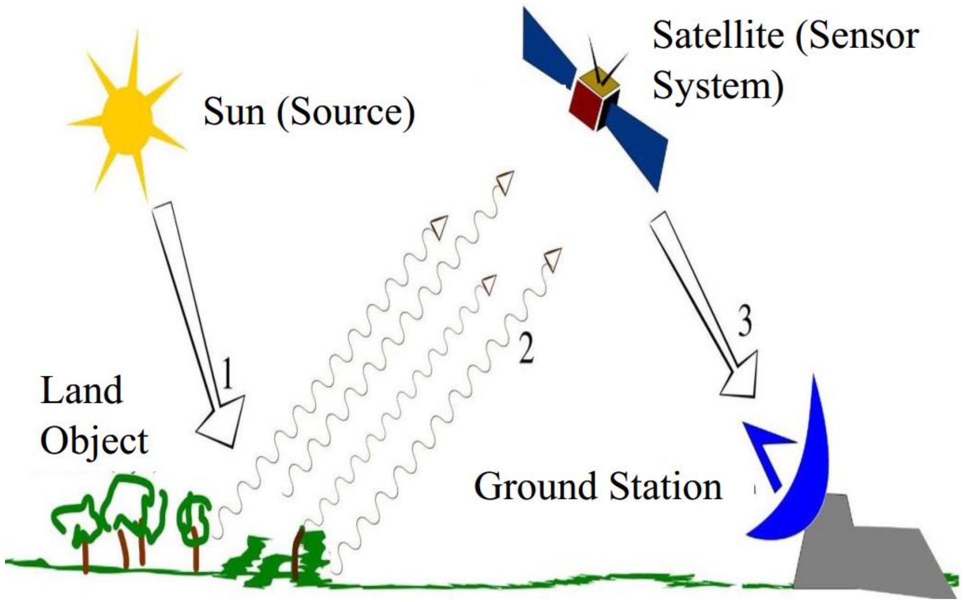
Remote Sensing

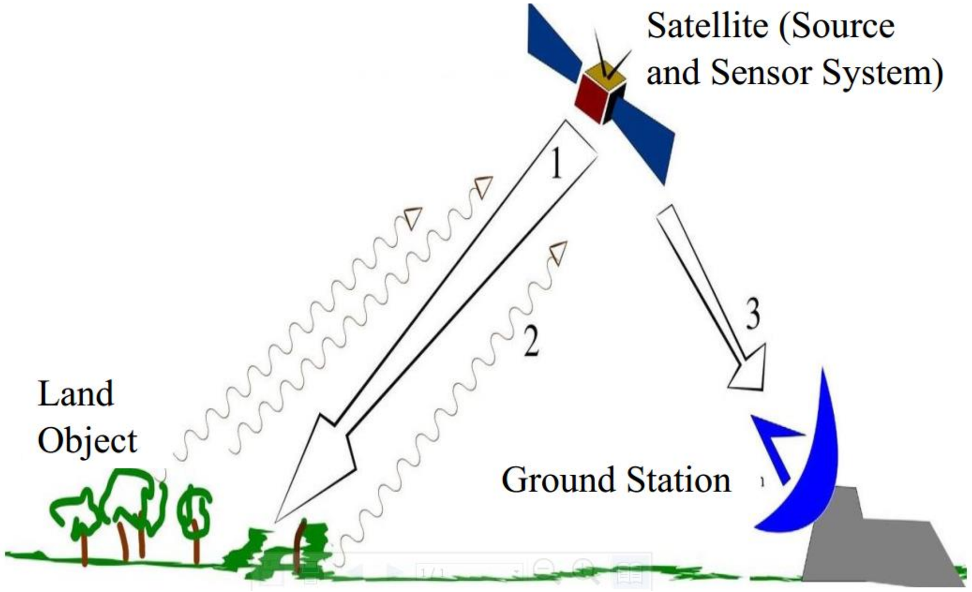
❖ What is remote sensing?

- ❖ Acquisition of information about an object without making physical contact with the object
- ❖ Used in geography, land surveying, military, intelligence and planning

❑ Types:

- ❑ Passive Remote Sensing
- ❑ Active Remote Sensing





Remote Sensing (Con't)

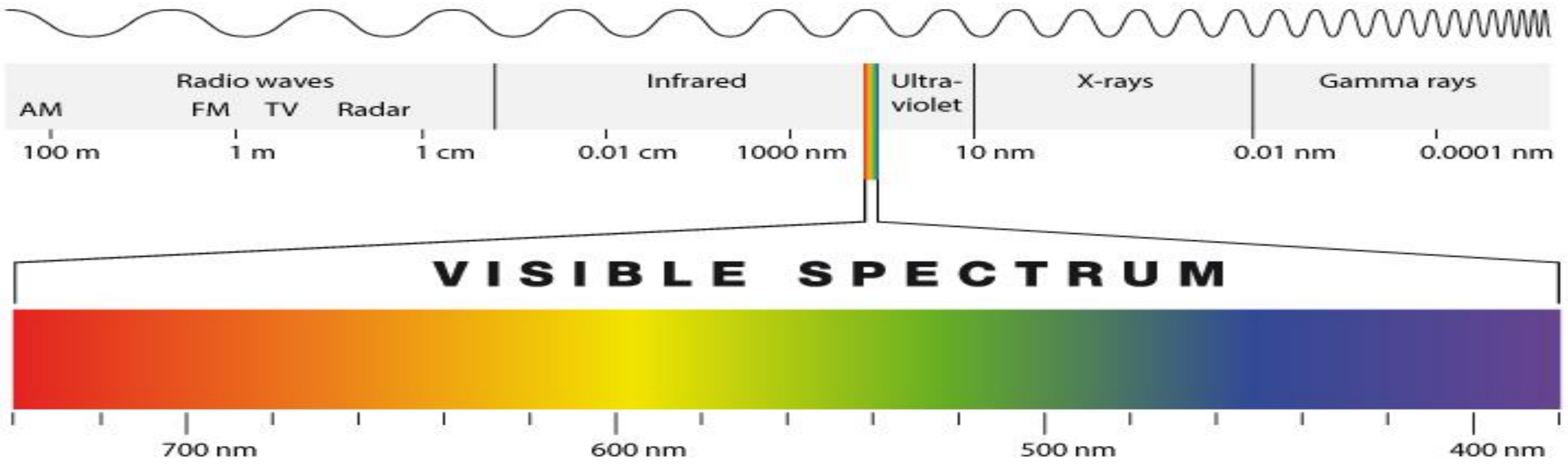


Figure: Electromagnetic Spectrum

Figure: Surveillance aircraft

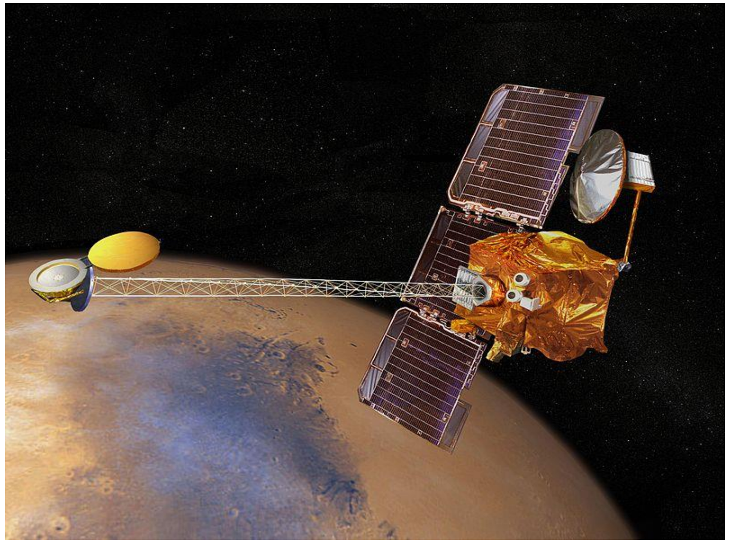


Figure: Remote sensing satellite

- ❖ Visible and infrared range of wavelengths represents only part of the story in remote sensing

- ❖ Can also image the earth in microwave range

Hyperspectral Images

Figure: 3D view of Hyperspectral data cube [5]

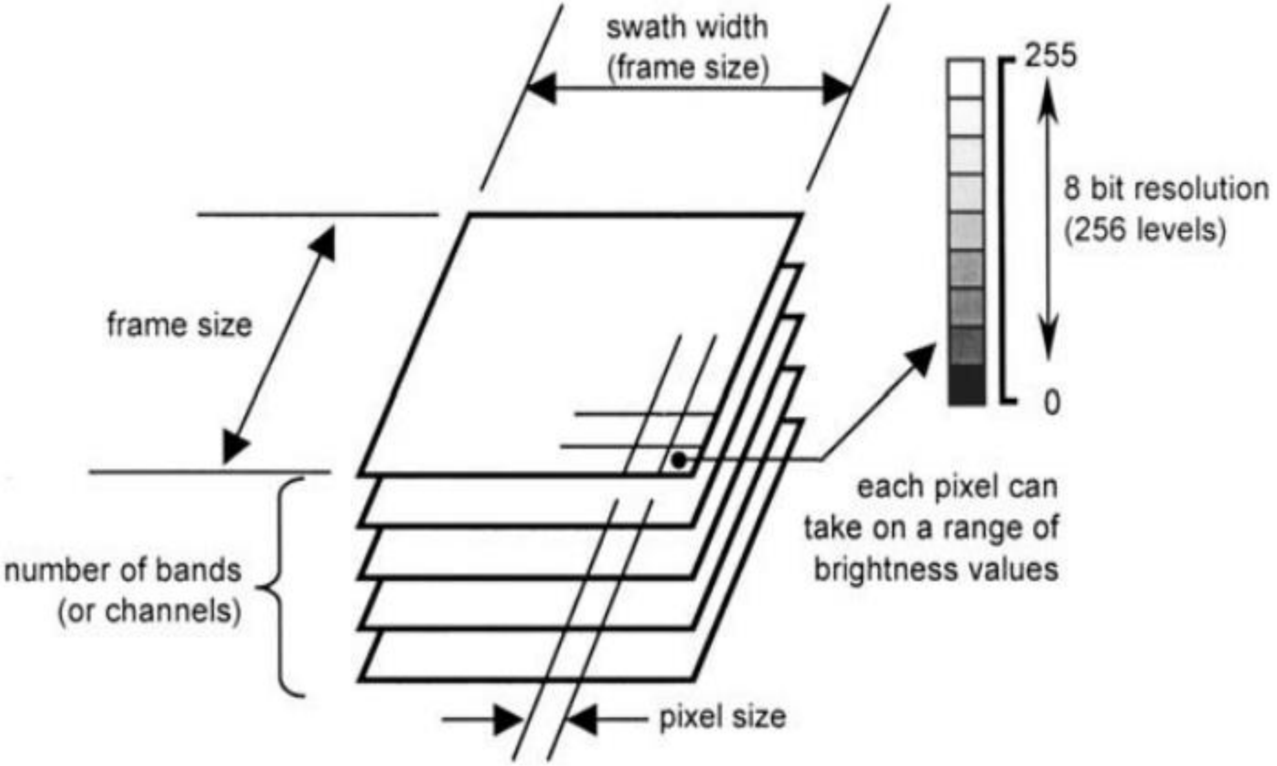


Figure: Technical characteristics of digital image data [2]

Challenges of Hyperspectral Images

- ❖ Curse of dimensionality
- ❖ High correlation among data
- ❖ All the feature are not equally important

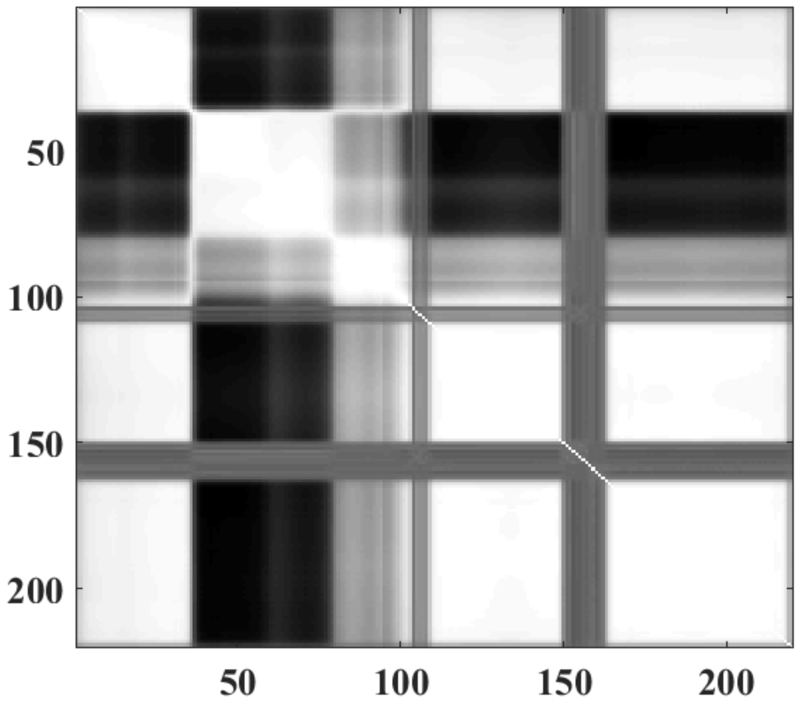


Figure: Correlation among bands

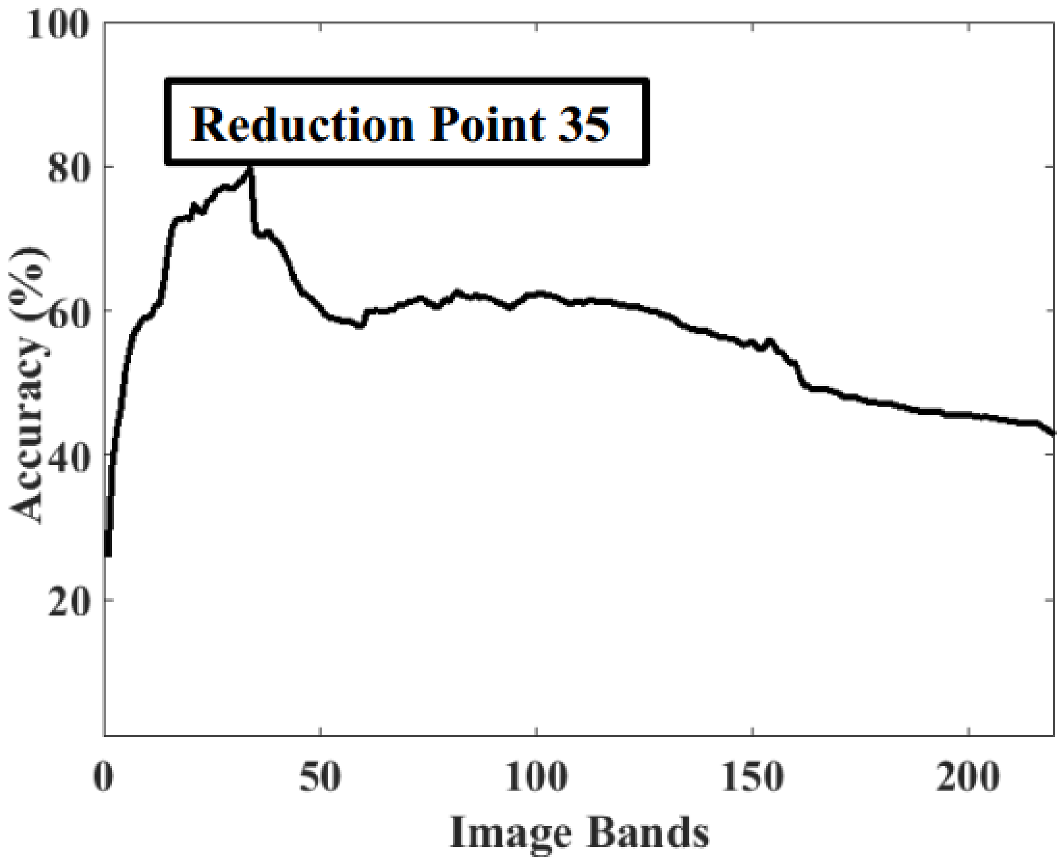


Figure: Curse of dimensionality

Motivations and Objectives

- Extract only relevant features and find an effective subspace
- Address curse of dimensionality
- Improve classification accuracy

Proposed Method

Target Class Oriented Subspace Detection (TCOSD)

Collaboration of feature mining through feature extraction and feature selection

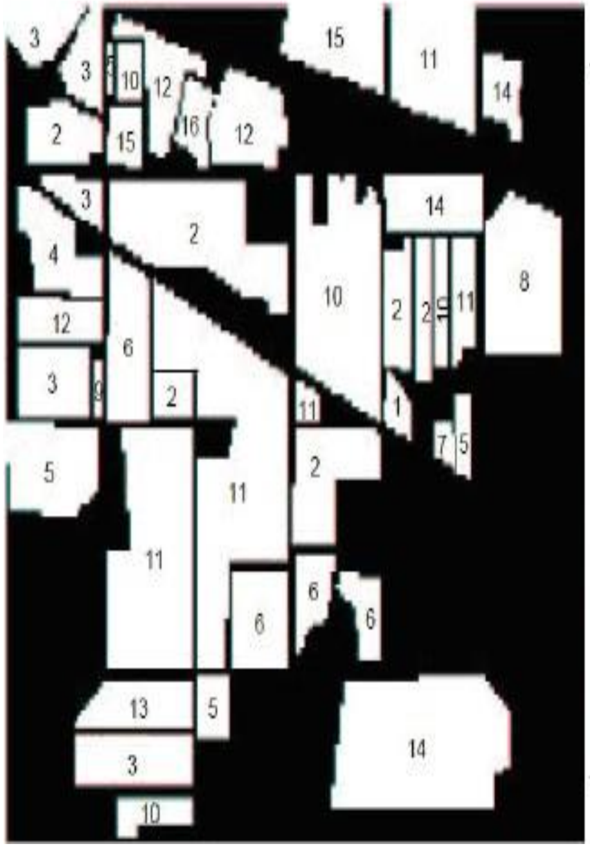
- **Feature Extraction:**
  Principal Component Analysis (PCA)
- **Feature Selection:**
  Normalized Mutual Information (NMI)
- **Classification:**
  Kernel Support Vector Machine (KSVM)

Dataset Description

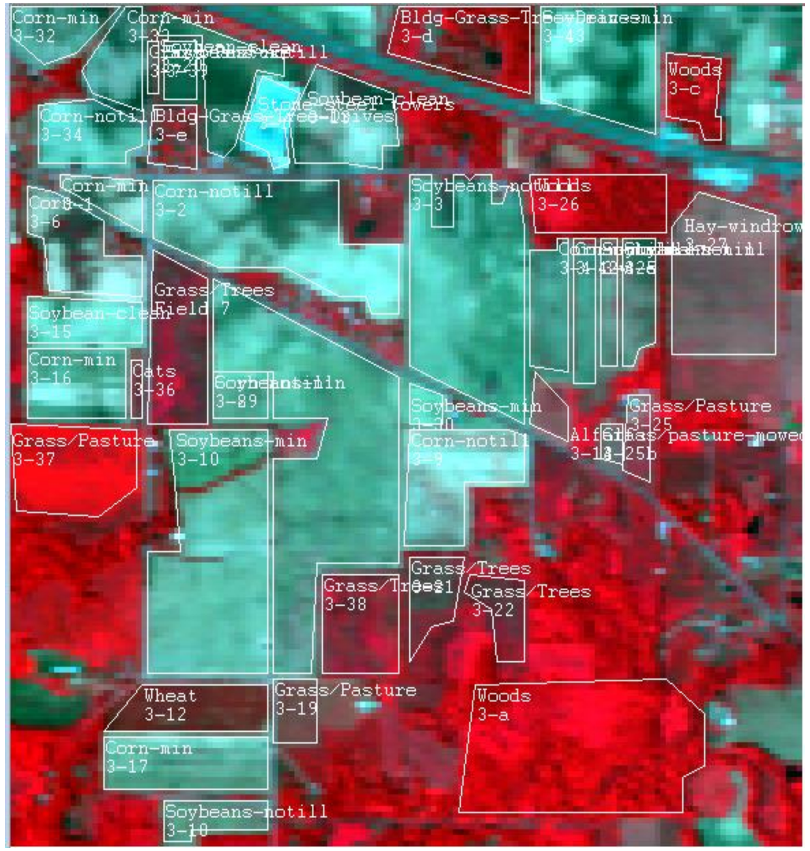
- **Dataset:** AVIRIS 92AV3C
- **Spectral Resolution:**
  0.4μm-2.5 μm
- **Spatial Resolution:** 30m
- **Band:** 220
- **Classes:**16

Data Set:
Captured Date:
June 12, 1992 Indian Pine
Test Site 3
Version 1.0

220 Band AVIRIS
Hyperspectral Image
Each band has 145 x 145
=21025 pixel



1. Alfalfa
2. Corn-notill
3. Corn-min
4. Corn
5. Grass/Pasture
6. Grass/Trees
7. Grass/pasture-mowed
8. Hay-windrowed
9. Oats
10. Soybeans-notill
11. Soybeans-min
12. Soybean-clean
13. Wheat
14. Woods
15. Bldg-Grass-Tree-Drives
16. Stone-steel towers



Feature Extraction

- Feature means unique measurable characteristics of an image/observed object (e.g., each band)
- Feature extraction means creating new features which have the essence of original features.
- Feature extraction helps to obtain relevant information from large dataset.

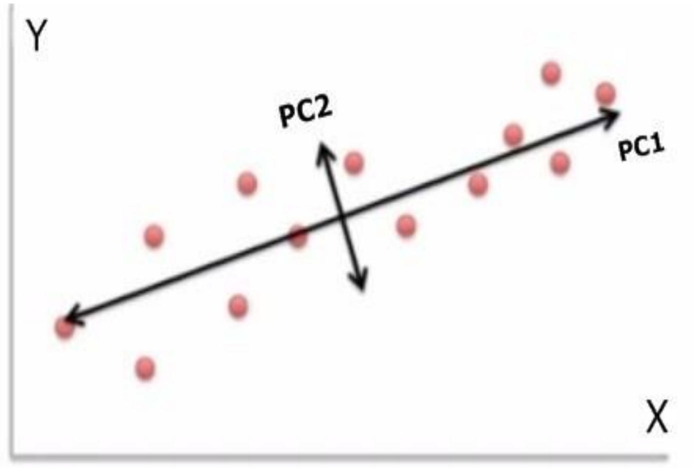
Principal Component Analysis(PCA)

- Perform orthogonal transformation
- Data are ordered based on high variance
- Dimensionality reduced by taking only first few components
- Classification accuracy Increased from the original set

Feature extraction using Principal Component Analysis(PCA)

- Principal Component Analysis (PCA) is a process of reduction input dimensionality. PC2 can be removed

- Transformed data is uncorrelated



Feature extraction using Principal Component Analysis(PCA)(Con’t)

All the steps we need to perform Principal Component Analysis[1]:

- **Step 1:** Input the image
- **Step 2:** Subtract the mean
- **Step 3:** Calculate the covariance matrix
- **Step 4:** Calculate the eigenvectors and eigenvalues of the covariance matrix
- **Step 5:** Choose components and form a feature vector

  FeatureVector = (eig1, eig2, eig3 ,…, eign)
- **Step 6:** Derive the new data set

  FinalData = RowFeatureVector x RowDataAdjust

Original Band

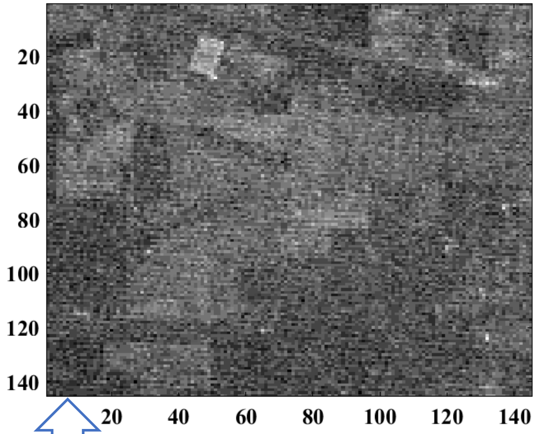
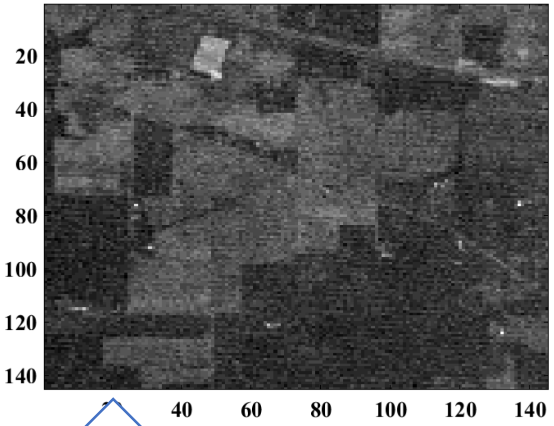
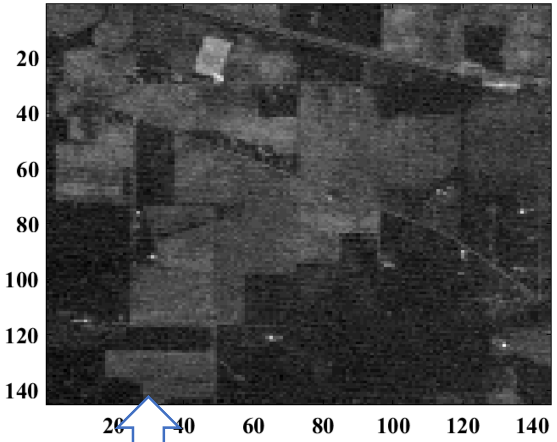


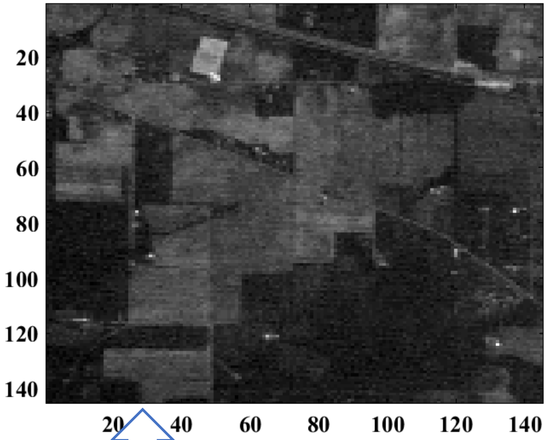
Image Band 1



Band 2

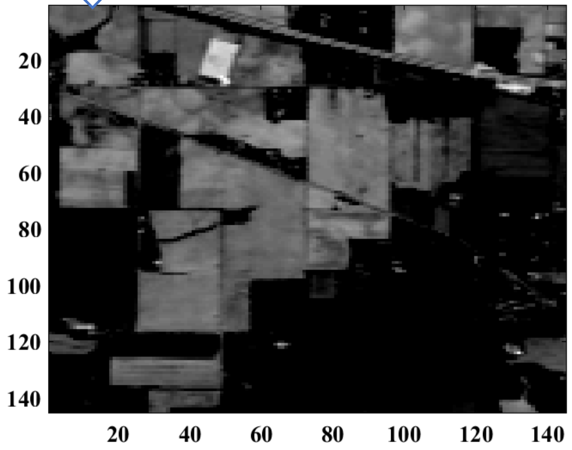


Band 3

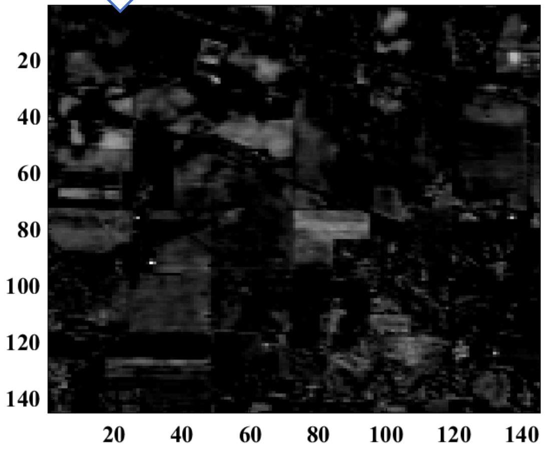


Band 4

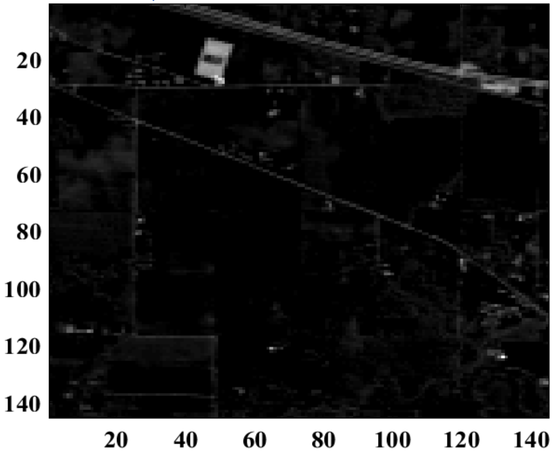
PCA Band



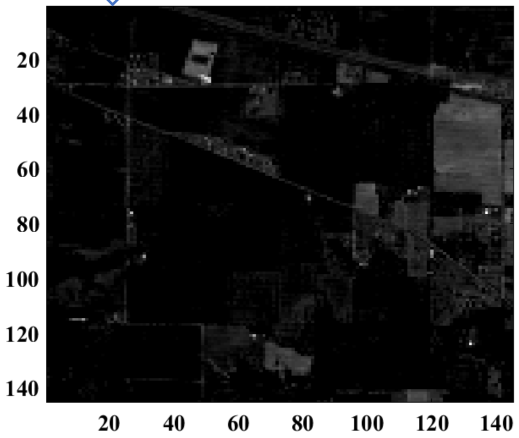
PC 1



PC 2



PC 3



PC 4

Cumulative variance of Transformed data

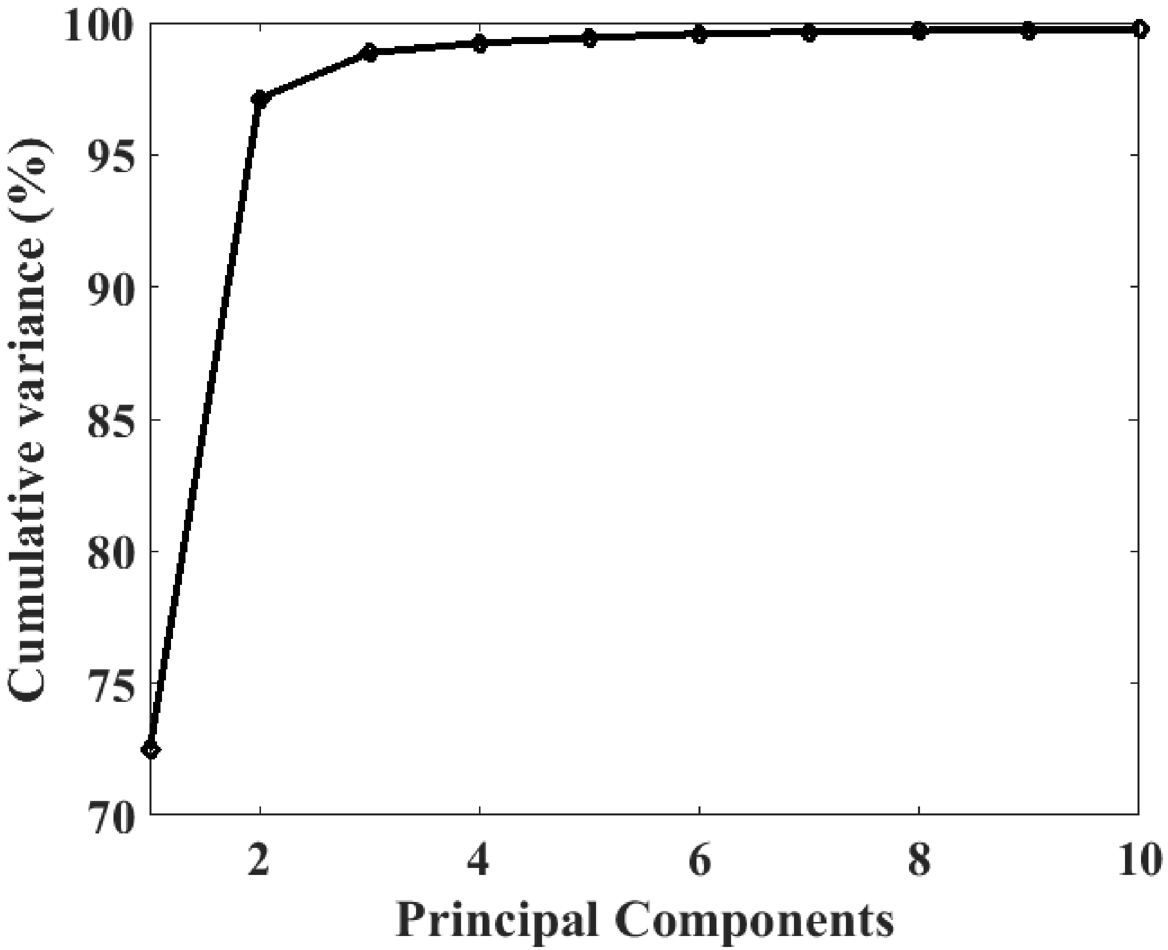


Figure: Cumulative variance

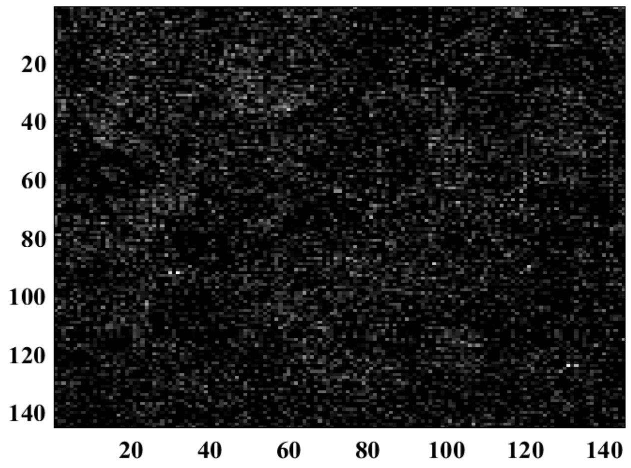
After principal component analysis first **10** components have **99.65%** variance of the total

Limitations of PCA

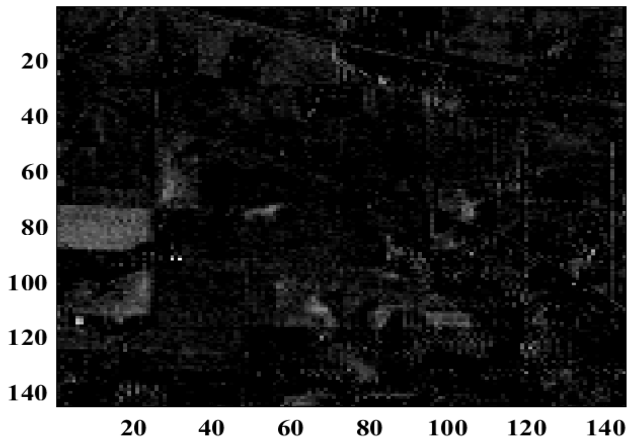
- PCA doesn't consider the class structure of input data and depends solely on the global variance . Assumes high variance means high information

- Here though PC 8 has high variance but it contains less information of image contents.

- On the other hand PC 9 has low variance but it contains more information of image contents.



PC 8



PC 9

Feature Selection

Normalized Mutual Information

$$\mathrm{I}(\mathbf{X},\mathbf{Y})=\sum_{y=1}^{L_1}\sum_{x=1}^{L_2}\mathrm{P}(x,y)\log\frac{\mathrm{P}(x,y)}{\mathrm{P}(x)\,\mathrm{P}(y)} \quad \text{........(1)}$$

$$\hat{\mathrm{I}}(\mathbf{X},\mathbf{Y})=\frac{\mathrm{I}(\mathbf{X},\mathbf{Y})}{\sqrt{\mathrm{I}(\mathbf{X},\mathbf{X})}\;\sqrt{\mathrm{I}(\mathbf{Y},\mathbf{Y})}} \quad \text{........(2)}$$

$$\hat{\mathrm{I}}(\mathbf{X},\mathbf{C})=\frac{\mathrm{I}(\mathbf{X},\mathbf{C})}{\sqrt{\mathrm{I}(\mathbf{X},\mathbf{X})}\;\sqrt{\mathrm{I}(\mathbf{C},\mathbf{C})}} \quad \text{........(3)}$$

$$\hat{R}(\mathbf{X_i},k)=\hat{\mathrm{I}}(\mathbf{X_i},\mathbf{C})-\frac{1}{k}\sum_{\mathbf{X}=S_k}\hat{\mathrm{I}}(\mathbf{X_i},\mathbf{X}),\ \mathbf{X_i}\notin S_k \quad \text{........(4)}$$

$$\mathbf{Y}=\mathbf{E}^T\mathbf{X} \quad \text{........(5)}$$

$$\hat{\mathrm{I}}(\mathbf{Y},\mathbf{C})=\frac{\mathrm{I}(\mathbf{Y},\mathbf{C})}{\sqrt{\mathrm{I}(\mathbf{Y},\mathbf{Y})}\;\sqrt{\mathrm{I}(\mathbf{C},\mathbf{C})}} \quad \text{........(6)}$$

$$\hat{R}(\mathbf{Y_i},k)=\hat{\mathrm{I}}(\mathbf{Y_i},\mathbf{C})-\frac{1}{k}\sum_{\mathbf{Y}=S_k}\hat{\mathrm{I}}(\mathbf{Y_i},\mathbf{Y}),\ \mathbf{Y_i}\notin S_k \quad \text{........(7)}$$

Normalized Mutual Information

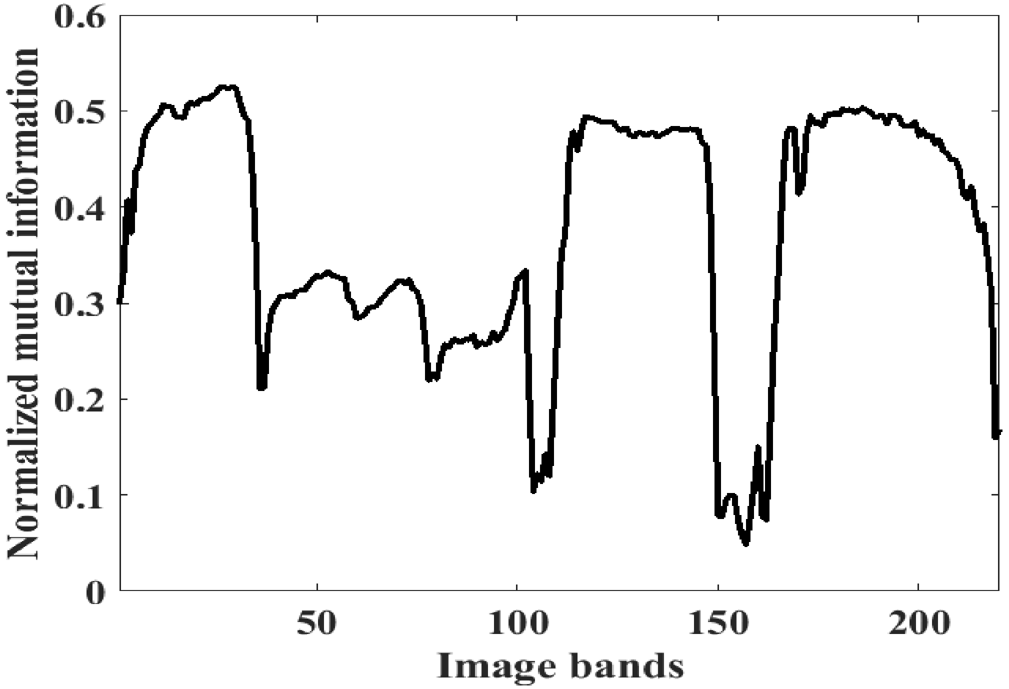


Figure: NMI between Class labels and Original Image bands

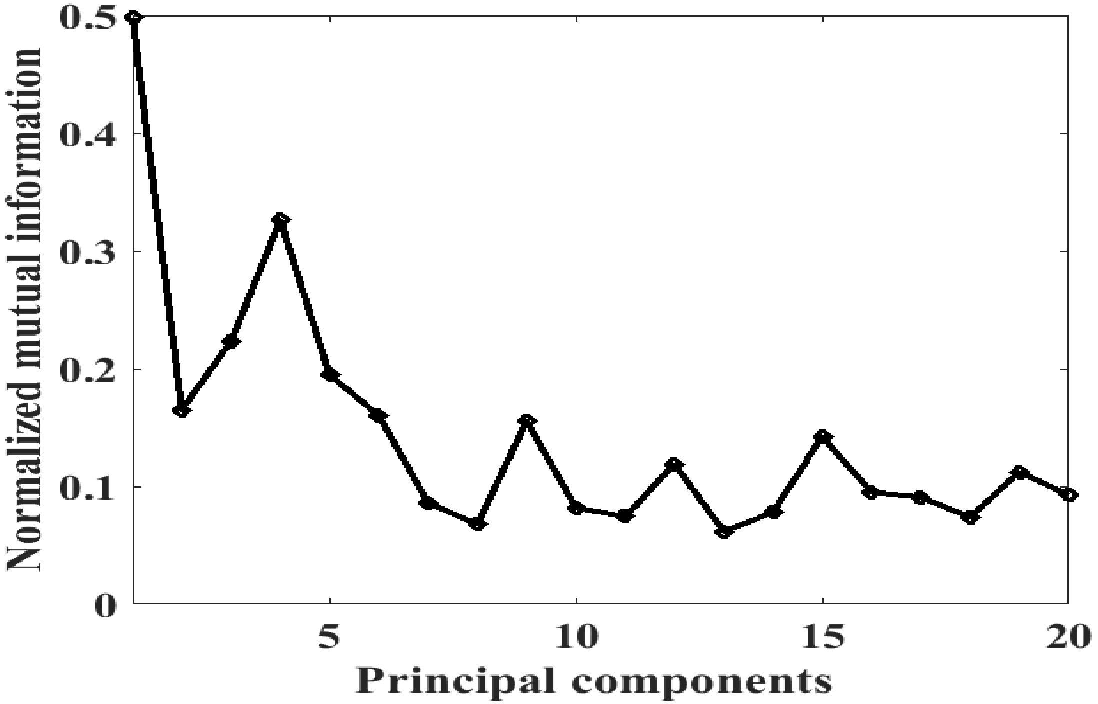


Figure: NMI between Class labels and Principal Components

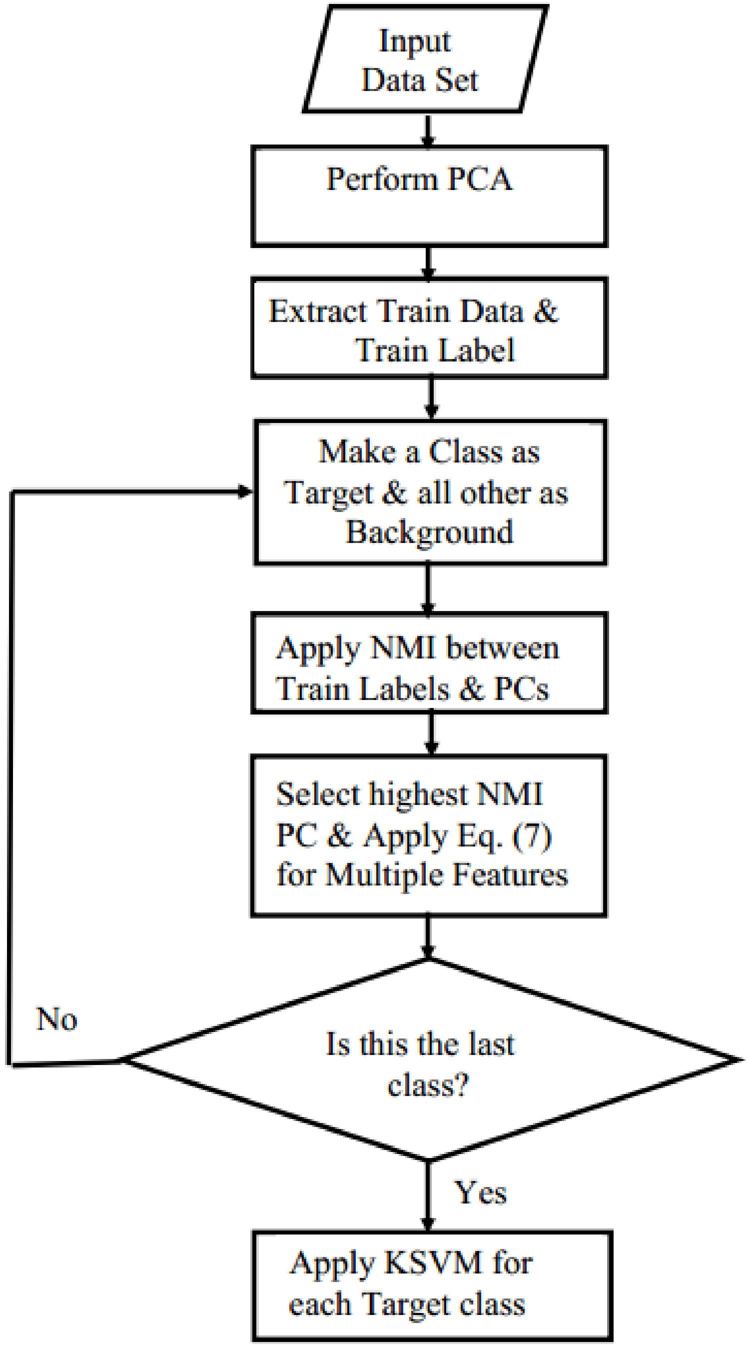
Method	Order of selected features
Org+NMI	Band: 27,1,167,3,139,14,2,8
PCA	PC: 1,2,3,4,5,6,7,8
PCA+NMI	PC: 1,4,3,5,9,6,2,15

Table: Selected Component of Standard Approaches Studied

Experimental Procedure

The algorithm of proposed TCOSD method is summarized as follows:

Step 1: Perform PCA and extract features of the input data set.
Step 2: Extract the training data and training label from the PCA features.
Step 3: Consider one class as target at a time, i.e, make a class as target and all other class as background.
Step 4: Apply NMI between the training labels and the principal components.
Step 5: Select the best principal component based on the high value of NMI. List the selected subset.
Step 6: Apply the Eq. (7) for selecting multiple features based on NMI for the target class.
Step 7: Apply the selected features to the KSVM classifier.
Step 8: Repeat step 3 to 6 by making another class as target and the remaining as background.



Experimental Results

Target class	Order of selected features
Hay-windrowed	PC: 4,1,17,5,12,3,16,2
Soybean-notil	PC: 1,17,16,11,20,14,13,3
Woods	PC: 1,16,17,3,15,11,5,20
Wheat	PC: 6,19,17,3,12,16,11,5
Grass/trees	PC: 4,9,5,6,16,17,2,1
Soybean-min	PC: 1,17,16,3,5,12,11,9
Corn-min	PC: 1,17,16,5,11,3,20,19
Stone-steel	PC: 3,4,1,17,16,10,5,11
Alfalfa	PC: 4,17,15,20,3,16,6,12
Grass/Pasture	PC: 9,3,17,6,16,11,1,15
Corn-notill	PC: 1,17,2,16,11,18,20,19
Soybean-clean	PC: 1,15,17,3,16,20,12,11
Corn	PC: 1,19,17,16,18,8,11,3
Bldg-Grass	PC: 15,16,17,3,18,11,19,4

Table: Selected features for each target class with proposed method (TCOSD)

Classification

Kernel support vector machine (KSVM)[3] classifier is used to asses the performance of the proposed TCOSD method.

Steps of classification:

Step 1: Get train and test data
Step 2: Scale data set into 0 to 1
Step 3: Calculate best C and g (kernel width)for RBF kernel of SVM

- 10 fold cross validation was used

Step 4: Assess the test sample accuracy

- Radial basis function (RBF) kernel was used

Class name	Train	Test
Hay-windrowed	187	77
Soybean-notil	128	105
Woods	367	341
Wheat	54	78
Grass/trees	249	115
Soybean-min	253	115
Corn-min	253	115
Stone-steel	36	30
Alfalfa	24	24
Grass/Pasture	156	92
Corn-notill	172	60
Soybean-clean	96	78
Corn	30	15
Bldg-Grass	40	12
Total	1900	1179

Conclusion

- 14 classes are selected
- Selecting first 8 features for proposed TCOSD gives 96.57% accuracy

Data set	C	g
Org+NMI	10	2.85
PCA	19	2.40
PCA+NMI	16	0.7
TCOSD (Proposed)	5	0.75

Data set	Classification Result (First 8 features only)
Org+NMI	72.26%
PCA	89.90%
PCA+NMI	93.12%
TCOSD (Proposed)	96.57%

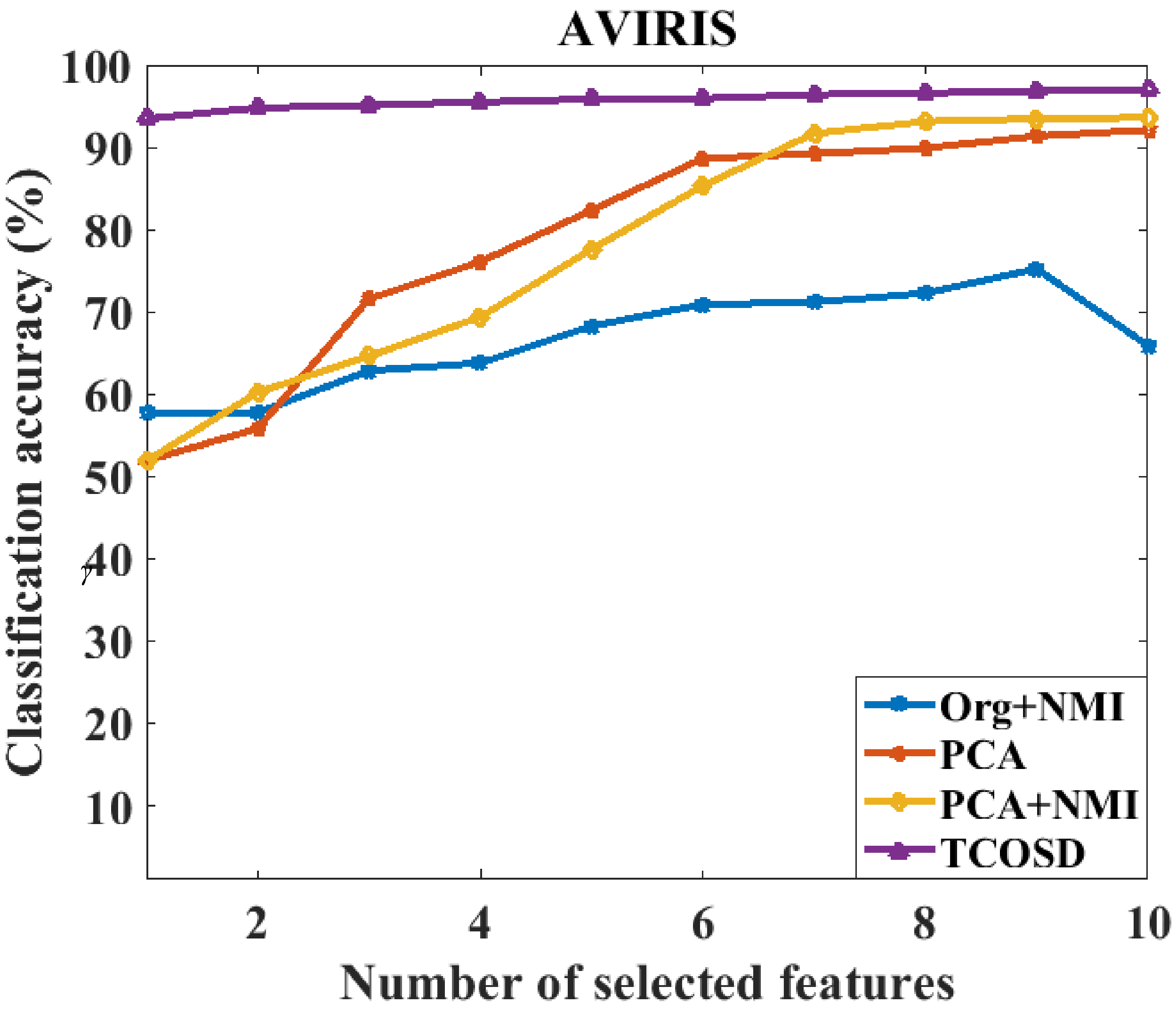


Figure: Comparison of classification result of Proposed TCOSD and standard approaches studied

Future Work

- ❖ This method needs some further improvement to handle the complex class relationships where only a few feature may not capable to complete the task.

- ❖ Selection will be performed by introducing adaptive thresholding T.

Acknowledgement

Special thanks to
Professor Dr. David A. Landgrebe, Purdue University
For the dataset and MUltiSpec

References

[1]L. I. Smith, “A tutorial on Principal Component Analysis”, February 26,2002.

[2]J. A. Richards and X. Jia, Remote Sensing Digital Image Analysis 4th edition, Verlag: Springer 2006.

[3]C. Hsu, C. Chang, and C. Lin,Department of Computer Science,National Taiwan University, Taipei 106, Taiwan,"A Practical Guide to Support Vector Classication", Initial version: 2003, Last updated: May 19, 2016.

[4]M. A. Hossain, X. Jia and M. Pickering, “Subspace detection using a mutual-information measure for hyperspectral image classification,” IEEE Geoscience and Remote Sensing Letters, vol. 11, no. 2, pp. 424-428, Feb. 2014.

[5][Online]https://en.wikipedia.org/wiki/Hyperspectral_imaging

References

[6]M. A. Hossain, X. Jia and M. Pickering, " Improved feature selection based on a mutual information measure for hyperspectral image classification," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Munich, Germany, pp. 3058-3061, Jul. 2012.

[7]M. A. Hossain, M. Pickering and X. Jia, " Unsupervised feature extraction based on a mutual information measure for hyperspectral image classification," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*,Vancouver, Canada, pp. 1720-1723, Jul. 2011.

Thank You