TENSOR-BASED MACHINE LEARNING ALGORITHMS FOR MULTIWAY CLASSIFICATION OF HYPERSPECTRAL IMAGES

Guided by
Prof. N.Venkateswaran

Presented by
Prithvi Shankar.S(312212106069)
Varsha.S(312212106113)
Yogesh Kanna.R(312212106123)

OVERVIEW

The presentation is organized as follows;

INTRODUCTION

The prospect of classifying hyperspectral images is exciting as they are used in a variety of applications from mining to farming. Hyperspectral images often contain reflectance information from numerous wavelengths. Thus, efficient decomposition methods are required to compress the data and make it further suitable for other processes such as classification and pattern recognition. In this project, a tensor is subject to two different kinds of decomposition, namely PCA and MPCA. The differences are established and tabulated. Then the Support Vector Machine (SVM) is used to classify the Hyperspectral image into multiple classes and its accuracy observed.

OBJECTIVES

- To perform tensor-based multiway classification on Hyperspectral images. Thus;
  - To represent Hyperspectral in their natural form as higher order tensors.
  - To perform feature extraction using tensorial and non-tensorial methods.
  - To train the classifier using the features.
  - To automatically categorize all the pixels in the image into multiple classes using the classifier.

MOTIVATION

- Hyperspectral images contain a spectrum for each pixel of the image, i.e each pixel in an image scene is captured at many different wavelengths.
- So, it can be naturally represented as a tensor of higher dimensions.
- Thus, we intend to use tensor-based image processing techniques that operate on higher dimensional data without performing any spatial rearrangement.

HYPERSPECTRAL IMAGING

- Hyperspectral images provides a densely sampled and almost continuous spectral response over the given wavelengths.
- They are spectrally over determined, i.e, they provide ample spectral information to identify and distinguish spectrally unique materials.
- Several materials radiate the incident light in a higher proportion only in a very narrow spectral range.
- This characteristic of high spectral resolution makes differentiation of various materials on earth possible.
- The ability of hyperspectral image to capture even minor variations in scene reflectance imparts tremendous advantages as far as data classification is concerned.

TENSORS

- A tensor is basically a multidimensional or N-way array of data.
- **Order** of a tensor is defined as the number of its modes or dimensions.
- In fact tensors are merely a generalisation of scalars and vectors;
- A matrix (2-Dimensional array) is a second order tensor.
- A vector (1-Dimensional array) is a first order tensor. (eg): Electric field in space
- A scalar is a zeroth order tensor. (eg): Mass of an object

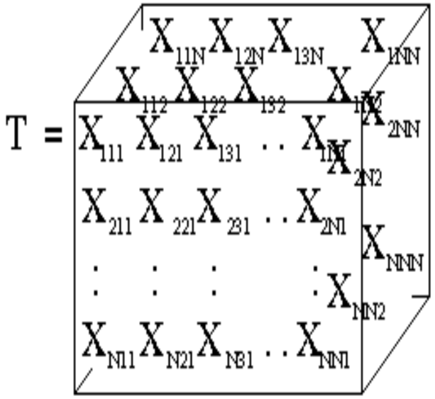


Fig. 1: A 3rd order Tensor
Courtesy: Mathis, Miles, *The trouble with tensors*. Retrieved 8 April 2016. http://milesmathis.com/tensor.html

TENSORS (contd...)

- Large sets of data which have an inherent multiway character can be naturally represented using tensors and analyzed.
- For example, Tensors are used in **Object recognition** (within computer vision application) for identifying and classifying objects in an image or video sequence.
- Fig 2 shows a sequence of images that can be represented as a third order tensor, where the 3 dimensions are the spatial row, spatial column and time.

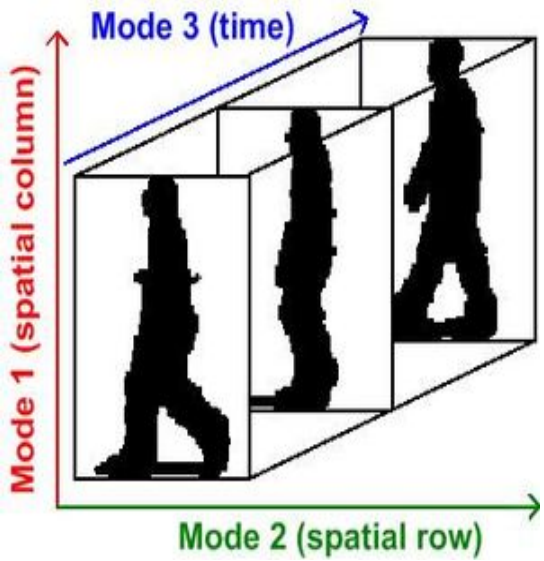


Fig. 2 : Example of tensor
Courtesy: http://www.wikiwand.com/en/Multilinear_subspace_learning

HYPERSPECTRAL IMAGE AS A TENSOR

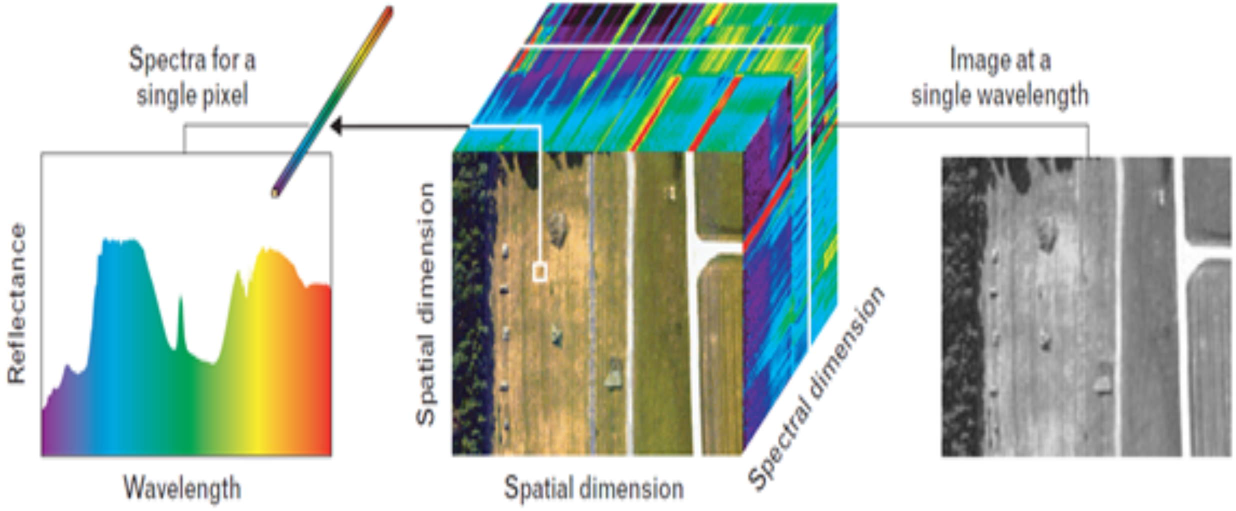


Fig. 3: Hyperspectral image
Courtesy: Prof. Tamás János, Fórián Tünde (2008), *Geoinformatics*
http://www.tankonyvtar.hu/en/tartalom/tamop425/0032_terinformatika/ch04s04.html

- The Hyperspectral image has three dimensions (2 spatial and one spectral), thus can be naturally represented as a tensor of 3rd order.

FEATURE EXTRACTION

- Feature extraction a type of dimensionality reduction that efficiently represents interesting parts of an image as a compact feature vector. This approach is useful when image sizes are large and a reduced feature representation is required to quickly complete tasks such as image matching and retrieval.

- Feature detection, feature extraction, and matching are often combined to solve common computer vision problems such as object detection and recognition, content-based image retrieval, face detection and recognition, and texture classification.

- It starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and training of the classifier.

FEATURE EXTRACTION

- We have performed **non-tensorial** and **tensor – based** feature extraction on the input Hyperspectral image.
- The following were used to form the feature vector in the non-tensorial method;
  - Spectral Magnitude.
  - Spectral First Derivative
  - Principal Component Analysis
- The tensor – based feature extraction was performed using Multilinear Principal Component Analysis (MPCA)

SPECTRAL SIGNATURE

- Different surface types reflect radiation differently in various wavelength channels.
- The difference in the reflectance or emittance characteristics of the surface with respect to wavelengths is called as **spectral signature**.
- The spectral signatures of 3 classes (corn-no till, Grass/Trees, woods) are shown in Fig. 4.

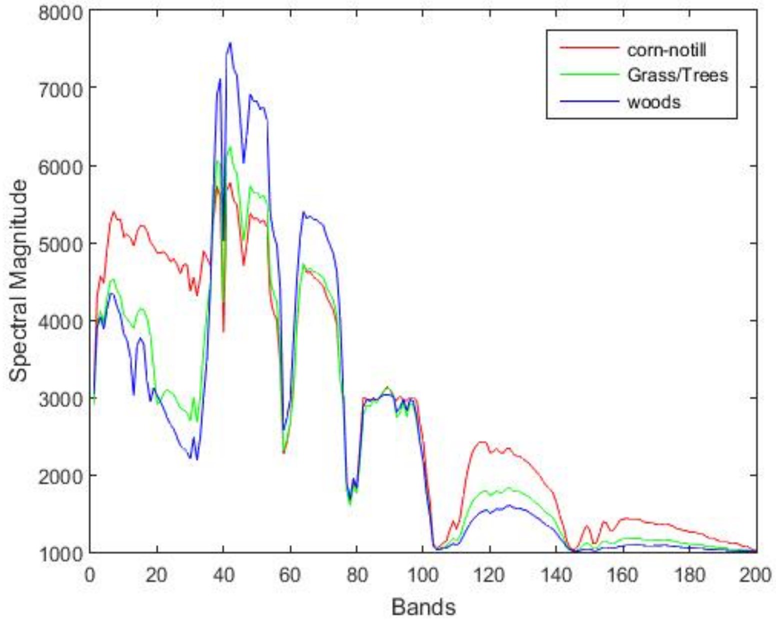


Fig. 4: Spectral Signature

SPECTRAL FIRST DERIVATIVE

•Spectral derivatives are used for discriminating different classes.

•The first spectral derivative is given by;

$$x_j' = x_{j+1} - x_j, j = 1, 2, \ldots, N-1$$

Where, x_j' shows the jth derivative

•The spectral first derivatives of 3 classes (corn-no till, Grass/Trees, woods) are shown in Fig. 5.

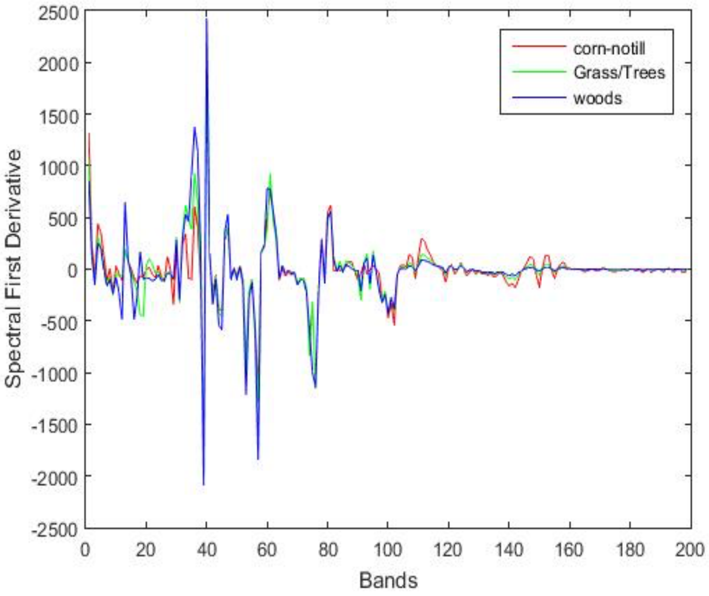


Fig. 5: Spectral First Derivative

PRINCIPAL COMPONENT ANALYSIS

- Principal Component Analysis (PCA) is commonly used for dimensionality reduction by performing orthogonal transformation on a set of correlated variables to transform into a set of uncorrelated linear variables.

- This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components.

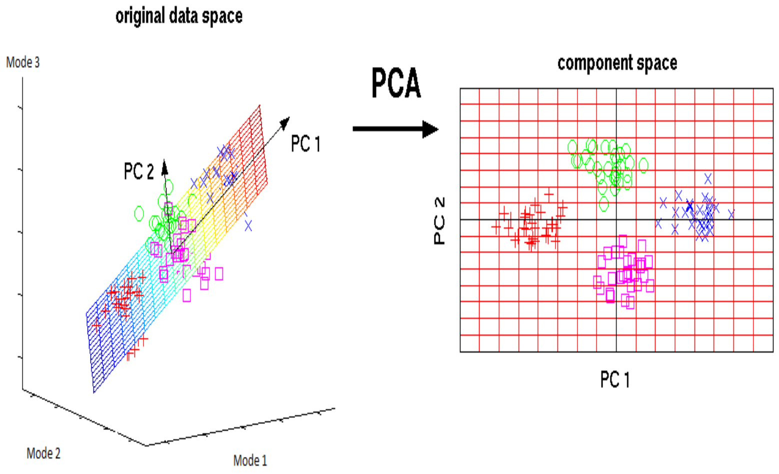


Fig. 6: PCA Transformation.
Scholz, M.(2006), *Principal Component Analysis(PCA)*. Retrieved 2 February, 2016 http://www.nlpca.org/pca_principal_component_analysis.html.

MULTILINEAR PRINCIPAL COMPONENT ANALYSIS

- MPCA is a multilinear extension of PCA.
- The major difference is that PCA needs to reshape a multidimensional object into a vector, while MPCA operates directly on multidimensional objects through mode-wise processing.
- For example, for 100x100 images, PCA operates on vectors of 10000x1 while MPCA operates on vectors of 100x1 in two modes. Thus, MPCA is more efficient and better conditioned in practice.

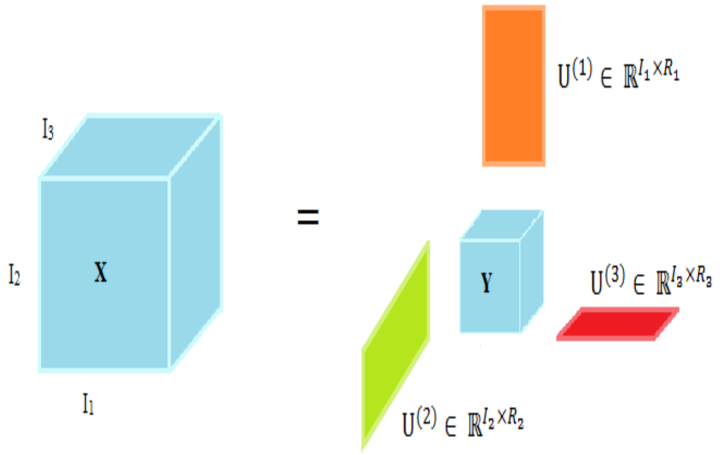


Fig. 7: Tensor Decomposition

MPCA ALGORITHM

INPUT: A tensor of N-dimensions

$$\{\chi_m \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}, m = 1, \ldots, M\}$$

OUTPUT: Low dimensional representation of the input tensor i.e output

$$\{\Upsilon_m \in \mathbb{R}^{P_1 \times P_2 \times \cdots \times P_N}, m = 1, \ldots, M\}$$

Step 1: Center the input samples;

$$\bar{\chi} = \frac{1}{M} \sum_{m=1}^{M} \chi_m$$

Step 2: Calculate the Eigen decomposition of mode-n total scatter matrix in full projection using;

$$\Phi^{(n)*} = \sum_{m=1}^{M} \tilde{\mathrm{X}}_{m(n)} \cdot \tilde{\mathrm{X}}_{m(n)}^{T}$$

Step 3: Set the projection matrix $\tilde{\mathrm{U}}^{(n)}$ to consist of Eigen vectors corresponding to the most significant P_n Eigen values, where n=1,2…..N

Step 4: Calculate $\{\tilde{\Upsilon}_m = \tilde{\chi}_m \times_1 \tilde{\mathrm{U}}^{(1)^{\mathrm{T}}} \times_2 \tilde{\mathrm{U}}^{(2)^{\mathrm{T}}} \ldots \times_N \tilde{\mathrm{U}}^{(n)^{T}}, m = 1, \ldots, M\}$

$$\Psi_{y_0} = \sum_{m=1}^{M} ||\tilde{\Upsilon}_m||_F^2$$

Step 5: For local optimization, iterative process is carried as follows

For k=1,2…,K

For n=1,2,…,N

Set the projection matrices $\widetilde{\mathrm{U}}^{(n)}$ to consist of Eigen vectors corresponding to the largest eigen values where, n=1,2…..N

Calculate $\{\widetilde{\mathrm{Y}}_{\mathrm{m}}, \mathrm{m} = 1, \ldots, \mathrm{M}\}$ and Ψ_{Υ_k}

If,

$$\Psi_{\Upsilon_k} - \Psi_{\Upsilon_{k-1}} < \eta$$

break

else

Repeat step 5.

Step 6: The feature tensor after projection is obtained as

$$\left\{\Upsilon_m = \chi_m \times_1 \widetilde{\mathrm{U}}^{(1)^{\mathrm{T}}} \times_2 \widetilde{\mathrm{U}}^{(2)^{\mathrm{T}}} \ldots \times_N \widetilde{\mathrm{U}}^{(N)^{\mathrm{T}}}, m = 1, \ldots, M\right\}$$

The feature tensor of dimensions less than that of the original tensor is obtained.

MULTIWAY CLASSIFICATION

- **Multiway** or **Multiclass classification** is the problem of classifying instances into one of the more than two classes.
- Some classification algorithms are binary in nature. However, they can be turned into multiway classifiers by the following strategies;

1. One Vs Rest:

It involves training a single classifier per class, with the samples of that class as positive samples and all other samples as negatives.

2. One Vs One:

In this, one trains K (K − 1) / 2 binary classifiers for a K-way multiclass problem. Each receives the samples of a pair of classes from the original training set, and must learn to distinguish these two classes.

SUPPORT VECTOR MACHINE

- **Support Vector Machines (SVM)** are supervised learning models with associated learning algorithms that analyze data used for classification.

- Given a set of training examples, each marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a binary classifier.

- However SVM can be modified to carry out Multiway classification.

SUPPORT VECTOR MACHINE

- The SVM algorithm constructs an hyperplane that separates the 2 different classes in the case of binary classification.
- For a linear SVM classification, the two classes(output) are labeled as Y=+1 and Y=-1 such that;

$$Y_i(w_i\, x + b) \geq 1, \text{ for any ith vector point}$$

- It must be true for all the data points, where x represent vector points, w and b represents weight of the vector point and a constant.

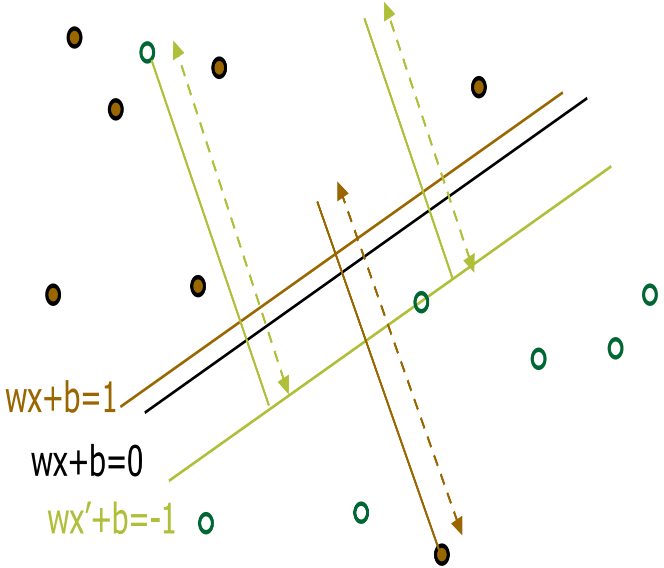


Fig. 8: Representation of Hyper planes.

SUPPORT VECTOR MACHINE

KERNEL TRICK:

If data is linear, a separating hyper plane may be used to divide the data. However it is often the case that the data is far from linear and the datasets are inseparable. To allow for this kernels are used to non-linearly map the input data to a high-dimensional space. The new mapping is then linearly separable.

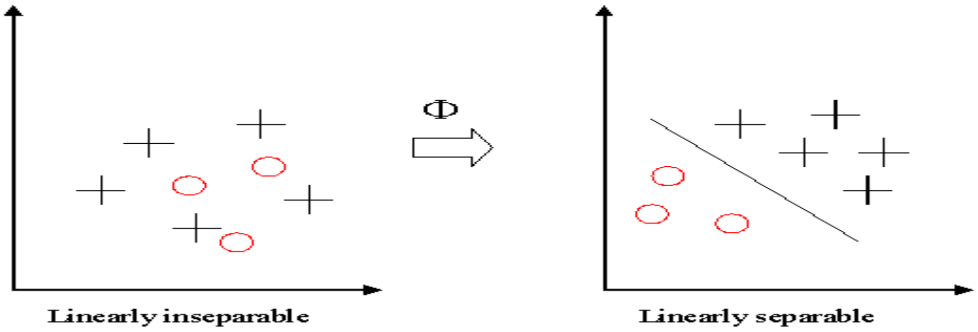


Fig. 9: Why use kernels?

PCA Vs MPCA

- First phase of the experiment involved the comparison of the performance of the dimensionality reduction methods, PCA and MPCA.
- For this, A sample Hyperspectral image of a landscape was obtained from Stanford Center for Image Systems Engineering (SCIEN), taken across 148 different wavelengths, ranging from 400nm to 950nm.

Fig. 10: RGB rendition of the Hyperspectral image.

PCA Vs MPCA - RESULTS

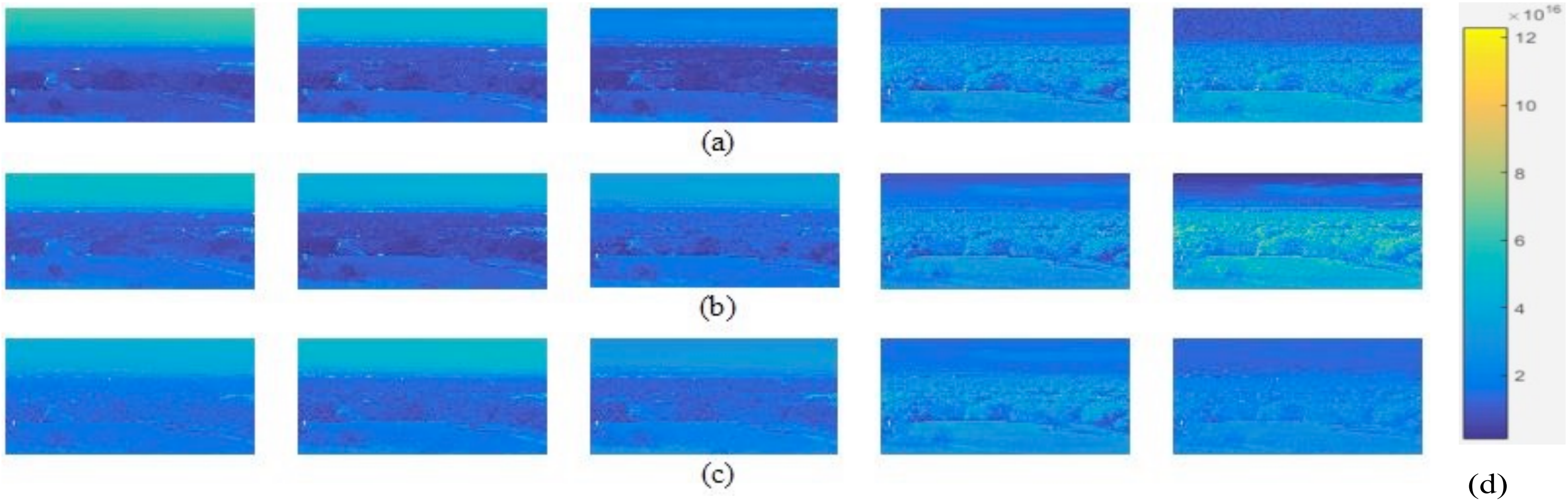


Fig. 11: Row (a) contains colour scaled versions of the original hyperspectral image at 5 wavelengths, namely; 414.7243nm, 538.6450nm, 684.4339nm, 848.4465nm, 950.4988nm. Row (b) contains the corresponding colour scaled versions of the reconstructed hyperspectral image obtained after applying PCA. Row (c) contains the colour scaled versions of the reconstructed hyperspectral image obtained after applying MPCA. (d) shows the colorbar used.

PCA Vs MPCA – RESULTS (Contd…)

Amount of variation kept in each mode	95%	97%	99%
SNR(dB)	20.2031	22.2045	26.5725
Size of Feature tensor	13 x 10 x 148	27 x 23 x 148	79 x 72 x 148
Relative error (%)	15.1750	11.9983	7.3370

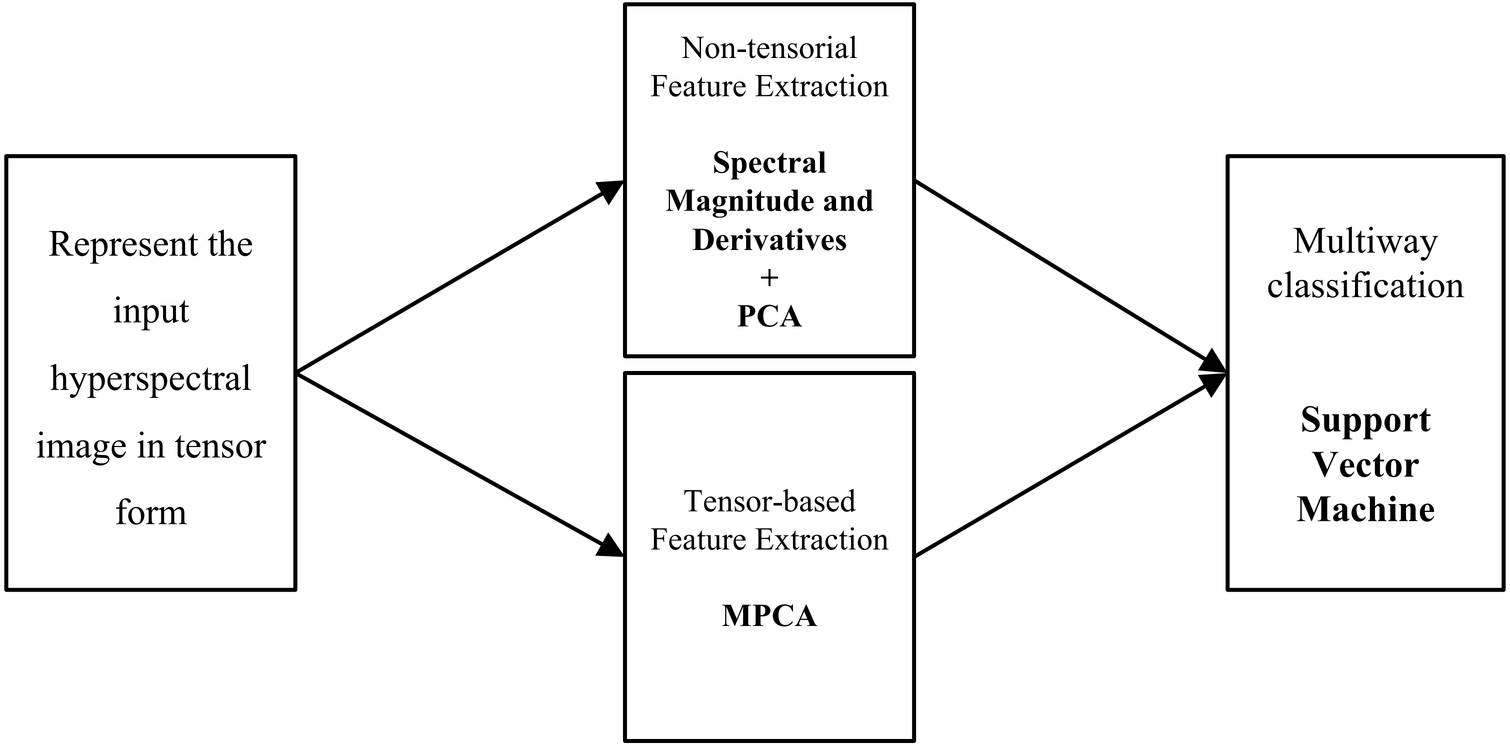
Table 1: Values of SNR (in dB) and Relative error (%) for different amounts of variation kept in each mode while applying MPCA.

PCA Vs MPCA – RESULTS (Contd…)

Algorithm	Computational time	
	SNR=22Db	SNR=27dB
MPCA	13.936887s	15.390155s
PCA	14.257096s	15.657825s

Table 2: Computational time for MPCA and PCA with similar SNR performed using Intel i7-4500U processor @ 1.80 GHz 2.40 GHz.

METHODOLOGY



DATASET DESCRIPTION

- The Indian pines dataset was gathered by Airborne Visible/infrared Imaging Spectrometer (AVIRIS) developed by NASA.
- It is a scene of 145 x 145 pixels with 220 bands acquired over Indiana's Indian pine in June 1992.
- The scene comprises of several forests and agriculture fields with 16 classes as shown in Fig. 12.

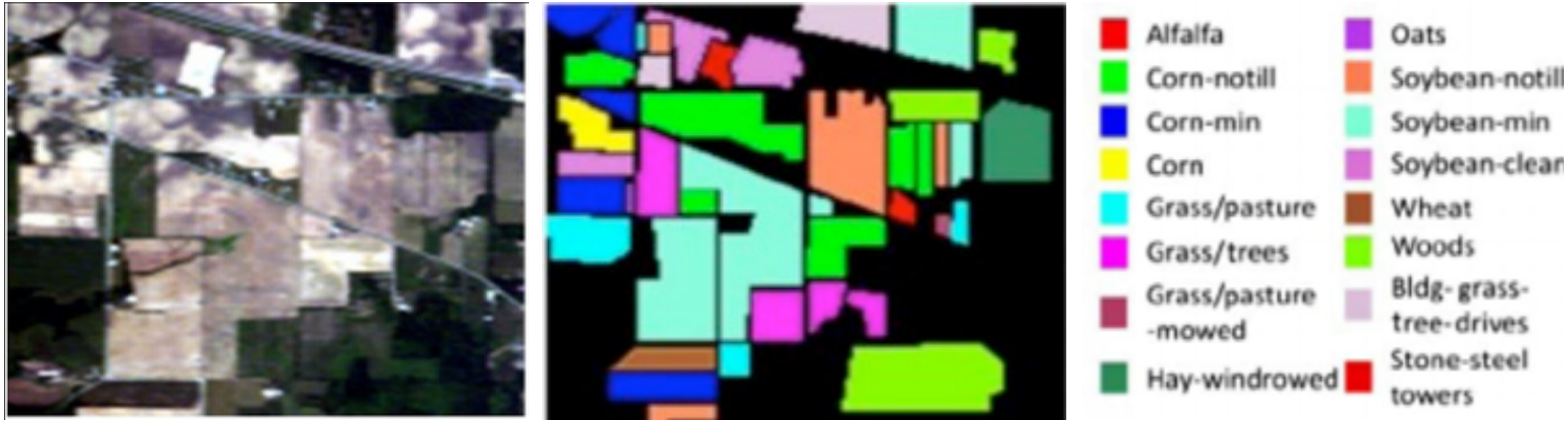


Fig. 12: A sample band of the Hyperspectral image, ground truth and legend.

The number of pixels(samples)/wavelength band for each class is as shown in Table 3;

Sl. No.	Class	Samples (Pixels)
1	Alfalfa	46
2	Corn-notill	1428
3	Corn – mintill	830
4	Corn	237
5	Grass – Pasture	483
6	Grass – trees	730
7	Grass – pasture – mowed	28
8	Hay – windrowed	478
9	Oats	20
10	Soybean – notill	972
11	Soybean – mintill	2455
12	Soybean – clean	593
13	Wheat	205
14	Woods	1265
15	Buildings – Grass – Trees – Drives	386
16	Stone – Steel – Towers	93

Table 3: Number of samples (pixels)/class

From Table 3 it can be seen that the classes 1, 4,7,9,13,15 and 16 have a small number of samples which is not sufficient for providing both training and test samples. Thus we don't consider these classes for classification. So we consider only 9 classes in the HSI for multiclass classification. The number of samples taken per class for training and testing are tabulated in Table 4.

Class	Training Samples	Test Samples
Corn – no till	720	708
Corn – min till	430	400
Grass/Pasture	250	233
Grass/Trees	370	360
Hay – windrowed	235	243
Soybean – no till	485	487
Soybean – min till	1245	1210
Soybean – clean till	300	293
Woods	640	625

Table 4: Number of Training and Test Samples for each class

SIMULATION RESULTS

Fig 13: Corn-notill Vs Hay-windrowed

Fig 14: Soybean-mintill Vs Woods

One Vs One Classification considering two classes at a time using the spectral magnitude and spectral first derivative features is shown in Fig. 13 and Fig. 14.

Fig. 15 shows all the 9 classes after all the one vs one classifications are performed using spectral magnitude and spectral first derivative features..

	Corn - notill
	Corn – mintill
	Grass/Pasture
	Grass/Tree
	Hay-windrowed
	Soybean-notill
	Soybean-mintill
	Soybean-cleantill
	Woods

Fig. 15: Complete classification of the input HSI using non-tensorial Feature extraction.

Fig. 16 shows all the 9 classes after all the one vs one classifications are performed using SVM on the feature vector obtained after applying PCA.

	Corn - notill
	Corn – mintill
	Grass/Pasture
	Grass/Tree
	Hay-windrowed
	Soybean-notill
	Soybean-mintill
	Soybean-cleantill
	Woods

Fig. 16: Classification of the input HSI after applying PCA on the feature vector.

Fig. 17 shows the complete classification into 9 classes by using SVM after MPCA based feature extraction.

	Corn - notill
	Corn – mintill
	Grass/Pasture
	Grass/Tree
	Hay-windrowed
	Soybean-notill
	Soybean-mintill
	Soybean-cleantill
	Woods

Fig. 17: Classification of the input HSI using tensor - based (MPCA) feature extraction.

Table 5 shows the accuracy (%) of the classification when the different methods of feature extraction are used.

Feature vector	**Spectral magnitude + Spectral First Derivative**	**PCA**	**MPCA**
Accuracy (%)	99.63	86.94	94.44

Table 5: Accuracy in different cases.

- It can be seen that Spectral Magnitude and Spectral First Derivative features when used directly yields the highest accuracy.
- But the size of the feature vector is very high in this case.
- The size is considerably reduced in the case of PCA and MPCA based feature extraction.
- Out of the two the tensor-based (MPCA) method is found to perform better than the non-tensorial (PCA) method of feature extraction.

CONCLUSION

- The input Hyperspectral image of size 145 x 145 pixels with 220 bands acquired over Indiana's Indian Pines gathered using AVIRIS (Airborne Visible/infrared Imaging Spectrometer) was represented as a tensor. Then tensor decomposition was done using PCA and MPCA. After which, the feature vectors were extracted and were classified using Support Vector Machines (SVM). It was observed that MPCA is marginally faster. The size of the feature tensor and the relative error depended upon the number of pixels preserved.

- Various combinations of tensor decomposition techniques and feature extraction techniques were used to measure the classifier accuracy. It is found that the tensorial method of feature extraction combined with linear SVM is more efficient than the PCA based feature extraction and polynomial-kernel based non-linear SVM. Also, the effect of an additional principal component was found.

FUTURE WORK

- This project can be extended by introducing an unsupervised learning algorithm which does not require existing classified datasets (Ground truth) to train the classifier.
- The new algorithm should classify the desired Hyperspectral image by utilizing the underlying information in the image.

REFERENCES

1. B. W. Bader and T. G. Kolda (2015). Algorithm 862: MATLAB tensor classes for fast algorithm prototyping, ACM Transactions on Mathematical Software 32(4):635-653
2. Brett W. Bader, Tamara G. Kolda and others (2015). MATLAB Tensor Toolbox Version 2.6, Available online
3. Baumgardner, M., Biehl, L., Landgrebe, D. (2015). 220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3. Purdue University Research Repository. doi: 10.4231/R7RX991C
4. S. Bourennane , C. Fossati and A. Cailly (2010) , ‘Improvement of classification for hyperspectral images based on tensor modeling’, IEEE Geosci. Remote Sens. Lett. , vol. 7 , no. 4 , pp.801 -805
5. B. Demir and S. Erturk (2008), ‘Spectral magnitude and spectral derivative feature fusion for improved classification of hyperspectral images’, Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS\'08)
6. M. Grigoryan and M. M. Grigoryan (2012) , Image Processing: Tensor Transform and Discrete Tomography with MATLAB :CRC Press
7. H. P. Lu, K. N. Plataniotis and A. N. Ventsanopoulos (2008), ‘MPCA: Multilinear principal component analysis of tensor objects’ , IEEE Trans. Neural Netw. , vol. 19 , no. 1 , pp.18 -39
8. Haiping Lu, K. N. Plataniotis, and A. N. Venetsanopoulos (2013) **‘Multilinear Subspace Learning: Dimensionality Reduction of Multidimensional Data’** , Chapman & Hall/CRC Press Machine Learning and Pattern Recognition Series, Taylor and Francis, ISBN: 978-1-4398572-4-3.

REFERENCES

9. B. Luo and J. Chanussot (2009) , ‘Hyperspectral image classification based on spectral and geometrical features’, Proc. IEEE Int. Workshop Mach. Learn. Signal Process. , pp.1 -6.
10. D. Manolakis and G. Shaw (2002) , ‘Detection algorithms for hyperspectral imaging applications’ , IEEE Signal Proc. Mag. , vol. 19, no. 1 , pp.29 -43.
11. M. Mudrova and A. Prochazka (2005), ‘Principal Component Analysis in Image Processing’ in Proceedings of International Conference Technical Computing, Prague, Czech Republic, Nov. 2005.
12. A.-H. Phan and A. Cichocki (2010), "Tensor decompositions for feature extraction and classification of high dimensional data sets", Nonlinear Theory Appl., IEICE, vol. 1, no. 1, pp. 37-68
13. Scholz, M.(2006), Principal Component Analysis(PCA). Retrieved 2 February, 2016 http://www.nlpca.org/pca_principal_component_analysis.html.
14. G. Shaw and D, Manolakis (2002), ‘Signal Procesing for Hyperspectral Image Exploitation.’ IEEE Signal Processing Magazine, Vol.19, N.1.
15. T. Skauli and J. Farrell (2013) ‘A collection of hyperspectral images for imaging systems research’. In Proceedings of the SPIE Electronic Imaging

THANK YOU