JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Feature Extraction of Hyperspectral Images

[1] B. Sucharitha, [2]Syeda Bushra Fatima
[1]Assistant Professor, [2]M.E. Student
[1]Department of Electronics and Communication Engineering,
[1]Muffakham Jah College of Engineering and Technology
Hyderabad, India

Abstract : Hundreds of closely spaced, contiguous wavelength bands conveying a plethora of spectral data make up hyperspectral images. Compression becomes necessary as processing or storing high data volume hyperspectral images puts strain on the technology. Tucker Decomposition is used to compress an image both spatially and spectrally. Peak Signal-to-Noise Ratio (PSNR) was used to calculate the information loss from compression. Spectral and spatial data are always combined in hyperspectral image classification to increase classification accuracy. This study combines Principal Component Analysis with Support Vector Machines to extract pixel information. This approach achieves remarkable classification accuracy.

IndexTerms **- Hyperspectral images, Compression, Tucker Decomposition, Support Vector Machines, Principal Component Analysis.**

I. INTRODUCTION

A Hyper spectral (HS) Image is a 3-dimensional data cube that contains data from a wide range of the objects' EM (Electro Magnetic) spectrum. Because of its capacity to preserve detailed information, HS imaging is a crucial concept in remote sensing. It has piqued the interest of academics in recent years because of its use in target recognition, classification, outlier detection, and spectral unmixing. When an object receives light, it absorbs some wavelengths and reflects others, resulting in different colours (R, G, B). This reflected light falls under separate wavelength ranges of electromagnetic radiation's visible spectrum and is responsible for the colour we perceive. Everything that is visible to the human eye is included in the electromagnetic spectrum's visible light range. It can collect wavelengths between 380 and 700 nanometres. But there are a lot of wavelengths not covered by the visible spectrum. Information regarding the wavelengths beyond the visible spectrum can be achieved by spectral imaging. Spectral imaging obtains image data at multiple bands of wavelength.

Generally, spectral imaging is of two types - multispectral and HS imaging. Multispectral imaging collects information from numerous discrete, non-contiguous wavelengths, whereas HS imaging collects enough information to rebuild a contiguous spectrum across a given spectral range.

HS imaging takes spectral data at several wavelengths while also capturing spatial data. The enhanced information provided by these wavelengths allows for more comprehensive chemical imaging. For each pixel in the picture, the recorded data is saved as a hypercube including spatial data (width and height) in two dimensions and spectral data (bands) in a third dimension.

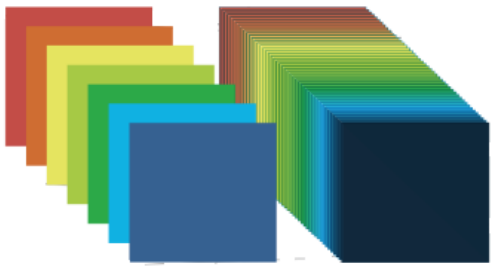

Fig. 1 A comparison of multispectral and HS imaging.

A few applications of HS imaging are explained below:

- Satellite HS sensors are employed for numerous types of surveillance, including soil tests, environmental control and monitoring, and military applications. Several wavelength channels could be used to monitor vegetation, while others can detect minerals, structures, and so on.
- In quality control of food products, HS images allow fairly accurate identification of defective products or foreign materials. Imaging sensors may be paired with computers for advanced data processing and actuators to eliminate undesired things, making digital food sorters far more efficient than man.

The acquisition, transmission, analysis, and storage of remotely sensed data presents unique challenges. The process of extracting information is perhaps the most important. In most cases, accurate analysis requires high-quality data, which increases data volume.

NASA JPL's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) captures data that has a volume frequently exceeding 500 Megabytes each flight, and is mainly used for anomaly detection, target identification, and geological mapping. NASA JPL's Atmospheric Infrared Sounder (AIRS) records thousands of bands over the infrared spectrum and produces more than 12 Gigabytes of data every day.

These numbers already exceed the transmission bandwidths that are available. Real-time distribution of data gathered remotely is made possible by effective data compression, which also lowers ground station storage needs.

The compression must be carried out in such a way that the reconstructed image has sufficient spatial and spectral quality for the application.

The use of HS images should theoretically improve our ability to classify the numerous features contained in it. However, data categorization is not as effective for HS data. The primary problem is the high dimensionality of HS images. The number of training samples necessary for image classification must increase in tandem with the complexity of the feature space.

In this paper, we propose a compression technique based on Tucker Decomposition (TD). It represents the HS Image dataset as a third-order tensor. Image classification for feature extraction is carried out using Support Vector Machines. The rest of this paper is structured as follows. Classification and compression algorithms are described in Section II. The flow of the algorithm is described in Section III. The results of the experimental analysis are shown in Section IV. Section V concludes with some suggestions for the future.

II. Literature Survey

2.1 Compression Algorithms

HS image compression is a broad topic that may be divided into a number of sections. We classed the compression algorithms depending on their methodology. Its specifics are detailed in the following sections.

2.1.1 Transform Algorithms

The most prevalent two-dimensional (2-D) image compression technology is transform-based, which has been extended to three-dimensional (3-D) for HS image compression. It applies the transformation function to all three dimensions of the image to transform the pixel values into the frequency domain. These algorithms can be used in conjunction with almost any other compression method.

Within the scope of this paper, certain transform-based compression methods are shown below. Karami et al. [1] presented 3D-DWT-TD on the HSI, which employs DWT to convert spatial domain pixels into frequency domain utilizing wavelet function on all three dimensions of the HS image. Each submatrix contains edge-based information, horizontal information, vertical information, and approximation information. TD is then applied to each of these four matrices individually, followed by the formation of mode matrices. The core tensor is then coded using an entropy coder, and the original picture is recreated using the reverse method. Dragotti et al. [2] used the 3-D Set Partitioning in Hierarchical Trees (SPIHT) technique to compress multispectral images at low bit rates. They presented and contrasted two different compression techniques. Both employ the wavelet transform (WT) in the spatial domain, but they differ in how they account for spectral dependencies: the first uses the Karhunen Loeve Transform (KLT), and the second uses tree-structured vector quantization (VQ). They initially used KLT on the entire image's spectral domain, then applied a spatial wavelet transform on the transform bands. After the transformation, the SPIHT algorithm is run on a 3-D hierarchical structure. The second SPIHT-based approach for multispectral images is developed, in which vector quantization is employed to encode spectral vectors of coefficients. More specifically, after performing WT on all bands of the picture, the similar coefficients from all bands are stacked to produce vectors, which are then vector-quantized to a certain precision. The two algorithms provide equivalent rate-distortion performance. However, given its lower design complexity, the KLT-based variant appears to be preferred.

Transformation-based techniques have also been applied with machine learning techniques such as Principal Component Analysis (PCA) in combination with DCT. The PCA-DCT [3] technique employs PCA for component identification and image data dimensionality reduction. The eigenvector is computed to choose the principal components (PCs) for determining similarities and differences in an image, which is then converted by an inverse transformation matrix to reconstruct the image. The reconstructed image is then processed using the DCT method to determine the coefficient of the generated hyperspectral image using companding quantization methods. The DCT primarily operates in the spectral domain to reduce storage space.

2.1.2 Tensor Decomposition

One of the newest image compression algorithms that work better than earlier methods is tensor decomposition. A tensor can be defined as a multidimensional array. A tensor of order one is a vector, one of order two is a matrix, and those of order three or more are referred to as higher-order tensors.

In CNN-NTD [4], Li combined tensor decomposition with deep learning techniques to propose a convolution neural network (CNN)-based transform for transforming a large-scale spectral tensor into a small-scale tensor. Then, to further decrease the dimensionality of the small-scale tensor acquired in the previous phase, Non-negative Tucker Decomposition (NTD) is used. To reduce spatial and spectral correlation, the resulting tensor was converted into a frequency domain using 3-D DCT.

2.2 Classification Algorithms

The large dimensionality of HS image data leads to the "dimensionality curse," which lowers classifier capability and worsens classification performance, especially when the number of labeled samples is constrained. Dimensionality reduction is commonly used as a pre-processing step to eliminate highly correlated and redundant observations from the original high-dimensional HS Image spectrum and preserve crucial data in a low-dimensional subspace. Dimensionality reduction of HS image data may be divided into two basic categories: feature selection and feature extraction. The former aims to extract from the original bands a small subset of the most characteristic bands, whilst the latter determines an optimal transformation matrix that maps the original high-dimensional spectral features onto a low-dimensional subspace. Feature extraction can use whole bands to create new

discriminative features. The major focus of this study is on using feature extraction to minimize the feature dimensions of HS images. Feature extraction may be categorized into unsupervised and supervised procedures based on whether or not label information is employed.

2.2.1 Unsupervised Procedures

Non-labeled hyperspectral data that is directly collected from the measurement is commonly reduced using unsupervised dimensionality reduction algorithms. Unsupervised approaches compress HS Images by using a set of criteria to find a representation in a low-dimensional space.

By taking into consideration the variation in distinct homogenous regions, the super pixel-wise PCA [6] approach enhances the dimensionality reduction impact of classic PCA. SuperPCA combines spatial context into unsupervised dimensionality reduction using superpixel segmentation and retrieves possible low-dimensional features even in the presence of noise.

2.2.2 Supervised Procedures

To infer class separability, supervised dimensionality reduction techniques use labelled hyperspectral data. Li et al. [7] established a supervised spectral-spatial segmentation system, integrating a multinomial logistic regression algorithm with a multilevel logistic Markov–Gibbs prior to help with segmentation of the spectral and spatial contextual data of the HS image. This approach characterizes noise and mixed pixels, and also yields accurate segmentation.

III. Proposed Method

In this section, we discuss the proposed algorithm. This algorithm consists of two phases: compression and classification. Compression of HS images is carried out by implementing Tucker decomposition (TD), whereas the Classification phase involves two stages i.e., PCA and Support Vector Machines (SVM).

3.1 Tucker Decomposition

TD breaks the tensor down into a tiny core tensor and factor matrices. A tensor in its decomposed Tucker form is just a core tensor with about the same rank as that of the source tensor plus a set of basis matrices, one for each mode of the core tensor.

The proposed algorithm makes use of the classical Tucker approach i.e., the Tucker decomposition via Higher Order Orthogonal Iteration (HOOI).

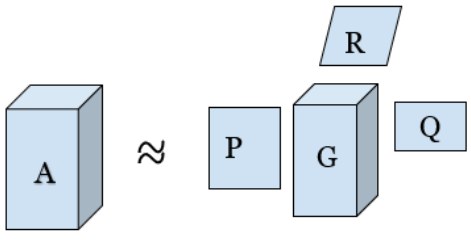

Fig. 2

Fig. 2 depicts a Tucker decomposition of a three-way array, where $A \in X^{I \times J \times K}$, we have

$$A \approx G \times_1 P \times_2 Q \times_3 R$$

$$= \sum_{l=1}^{L} \sum_{m=1}^{M} \sum_{n=1}^{N} g_{lmn}\, p_l \cdot q_m \cdot r_n \quad (1)$$

Here, $P \in X^{I \times L}$, $Q \in X^{J \times M}$, $R \in X^{K \times N}$ are the orthogonal factor matrices and the tensor $G \in X^{L \times M \times N}$ is called the core tensor. L, M, N are the number of components in the factor matrices P, Q, R respectively.

Algorithm 2: Tucker Decomposition via Higher Order Orthogonal Iterations

Input:

Tensor $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \ldots \times I_N}$

Truncation $(R_1, R_2, \ldots, R_N)$

Initial Guess $\{U_0^{(n)} : n = 1, 2, \ldots, N\}$

Output:

Core Tensor $\mathcal{G}$

Factor Matrices $\{U_k^{(n)} : n = 1, 2, \ldots, N\}$

1: k ← 0
2: **while** not convergent **do**
3: **for all** $n \in \{1, 2, \ldots, N\}$ **do**
4: $\mathcal{B} \leftarrow \mathcal{T} \times_1 (U_{k+1}^{(1)})^T \ldots \times_{n-1} (U_{k+1}^{(n-1)})^T \times_{n+1} (U_k^{(n+1)})^T \ldots \times_N (U_k^{(N)})^T$
5: $B_{(n)} \leftarrow \mathcal{B}$ in matrix format
6: $U, \Sigma, V^T \leftarrow$ truncated rank - R_n SVD of $B_{(n)}$
7: $U_{k+1}^{(n)} \leftarrow U$
8: k ← k + 1
9: **end for**
10: **end while**
11: $\mathcal{G} \leftarrow \Sigma V^T$ in tensor format

Fig. 3

The algorithm for TD via HOOI is shown in Fig. 3. It works by breaking down the HS image tensor into a tiny core tensor and factor matrices. Compression is performed using truncated SVD. It is an iterative process which repeats until the relative error between original tensor and compressed tensor is ideal.

3.2 Principal Component Analysis

Principal Component Analysis is a technique for reducing dimensionality. PCA is required to cut down on the number of variables since fewer variables are easier to deal with and can be processed more quickly. To put it another way, we strive to retain as much information as feasible with fewer variables.

PCA reduces dimensionality to improve the data interpretability by creating new uncorrelated variables that maximize the variance. These newly formed variables are commonly known as Principal Components (PCs).

3.3 Support Vector Machines

Among supervised learning techniques are support vector machines. The well-known machine learning technique SVM for classification issues looks for a hyperplane that optimises the distance between various classes while changing margin, which is the distance between the decision border and the closest training samples (support vectors).

IV. Simulation Results

4.1 Datasets

The project employed hyperspectral data: Indian Pines Corrected and Pavia University Centre. Indian Pines Corrected comprises of 145×145 reflectance values and 200 bands. The dataset Pavia comprises of 610×340 pixels and 103 bands.

Table 1 Specifications of Dataset

Dataset	Image size	Spectral bands	No. of Classes
Indian Pines Corrected	145 x 145 pixels	200	16
Pavia University	610 x 340 pixels	103	9

4.2 Image Compression

The hyperspectral image is compressed in both spatial and spectral aspects using TD into a core tensor and three orthogonal basis matrices. Table 2 shows the relative error in reconstructed and original HS images. The relative error reported in Table 2 shows that the reconstructed HS image is very much similar to the original HS image. Results are compared with compression results of [9].

In the paper [9], compression is carried out using Higher Order Singular Value Decomposition (HOSVD). The HOSVD creates factor matrices and core tensors from the singular vectors of the n-mode matricizations of a tensor. Since HOSVD is not iterative, it is clear that computing the SVD of a tensor's n-mode matricizations accounts for the majority of the computational cost.

Table 2 Relative Error

Dataset	Relative Error (HOSVD) [9]	Relative Error (HOOI)
Indian Pines Corrected	0.03	0.019
Pavia University	0.07	0.047

The size of the original tensor, compressed core tensor, and the factor matrices is given in Table 3. The time taken to compress and reconstruct the image is also tabulated in the table below. The computation time is an average of ten iterations.

Table 3 Compression Results

Dataset	Original Size	Compressed Core Size	Factor Matrices	Computation Time
Indian Pines Corrected	145 x 145 x 200	100 x 100 x 90	145 x 100 145 x 100 200 x 75	8.48 sec
Pavia University	610 x 340 x 103	500 x 200 x 75	610 x 500 340 x 200 103 x 75	66.46 sec

The performance of the compression algorithm is evaluated by measuring Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM). The Peak Signal-to-Noise ratio is used to compare the quality of the original HS image to the reconstructed HS image. The better the quality of the reconstructed image, the higher the PSNR. The PSNR plot of each band of the datasets is depicted in Fig.4(a-b).

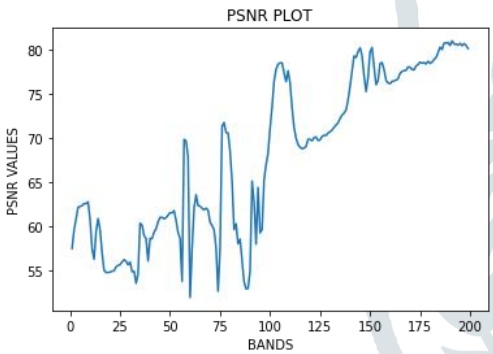

Fig. 4(a) PSNR plot of Indian Pines Corrected

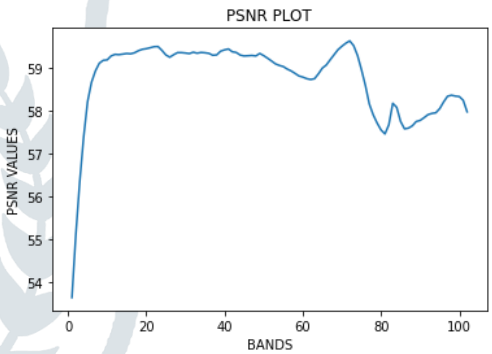

Fig. 4(b) PSNR plot of Pavia University

The human visual perception system is extremely capable of extracting structural information from a scene and hence distinguishing between information acquired from a reference and a sample scene. The Structural Similarity Index mimics this tendency while distinguishing between a sample and a reference image.

The Structural Similarity Index assigns a number between 0 and 1 to the similarity between two images. The value 1 indicates that the two images are highly similar or the same, whilst a value of 0 shows that the two images are extremely dissimilar. These numbers are frequently adjusted to be within the range [0, 1], where the extremes have the same meaning. SSIM is calculated for each band of the dataset and plotted as illustrated in Fig. 5(a-b).

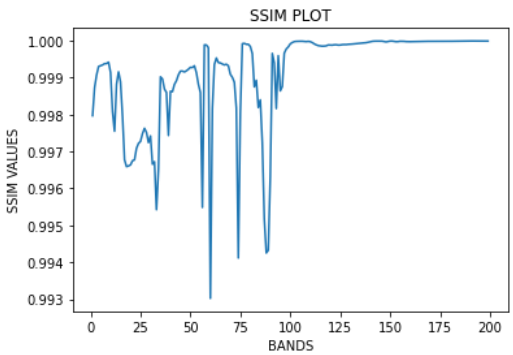

Fig. 5(a) SSIM plot of Indian Pines Corrected

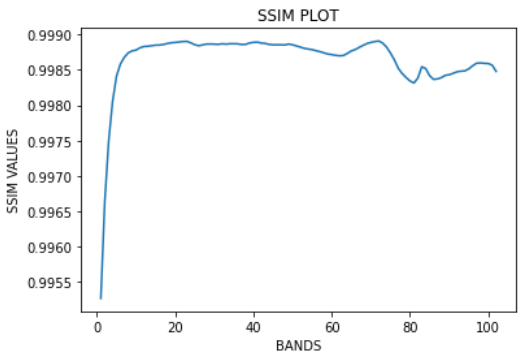

Fig. 5(b) SSIM plot of Pavia University

4.3 Image Classification

Dimensionality Reduction (DR) has proven critical in improving pixel categorization accuracy. DR reduces the number of dimensions of the data, allowing classifiers to create complete models at a minimal computational cost. Among the various DR techniques, we make use of PCA to reduce data dimensionality.

PCA increases the interpretability at the same time minimizes information loss. This is accomplished by creating Principal Components (PCs). The Explained Variance Ratio is an array containing the variance of the data explained by each PC. We may then choose the ideal number of dimensions to minimize. These PCs are then used for building a classification model via SVM.

SVM maximizes the difference between the data and the hyperplane. To project the data into higher dimensions, several kernel functions such as linear, polynomial, radial basis function (RBF), and sigmoid are utilized. Class-wise classification accuracy is analyzed. This can be seen in the classification report which also specifies the performance metrics namely Precision and Recall. The classification report for Indian Pines Corrected and Pavia University is depicted in Fig. 6 and Fig. 7 respectively.

	precision	recall	f1-score	support
Alfalfa	1.00	0.89	0.94	9
Corn-notill	0.83	0.80	0.81	286
Corn-mintill	0.90	0.81	0.85	166
Corn	0.79	0.79	0.79	47
Grass-pasture	0.92	0.96	0.94	97
Grass-trees	0.96	0.98	0.97	146
Grass-pasture-mowed	1.00	0.80	0.89	5
Hay-windrowed	0.99	1.00	0.99	96
Oats	0.50	0.50	0.50	4
Soybean-notill	0.88	0.79	0.83	194
Soybean-mintill	0.84	0.91	0.87	491
Soybean-clean	0.84	0.89	0.87	119
Wheat	0.95	1.00	0.98	41
Woods	0.96	0.98	0.97	253
Buildings Grass Trees Drives	0.92	0.73	0.81	77
Stone Steel Towers	1.00	1.00	1.00	19
accuracy			0.89	2050
macro avg	0.89	0.86	0.88	2050
weighted avg	0.89	0.89	0.88	2050

Fig. 6

	precision	recall	f1-score	support
Asphalt	0.96	0.96	0.96	1989
Meadows	0.97	0.98	0.98	5595
Gravel	0.89	0.83	0.86	630
Trees	0.99	0.96	0.98	919
Painted metal sheets	1.00	1.00	1.00	403
Bare soil	0.94	0.92	0.93	1509
Bitumen	0.92	0.91	0.91	399
Self-blocking bricks	0.89	0.92	0.90	1105
Shadows	1.00	1.00	1.00	284
accuracy			0.96	12833
macro avg	0.95	0.94	0.95	12833
weighted avg	0.96	0.96	0.96	12833

Fig. 7

The confusion matrix is a N x N matrix that is used to evaluate the performance of a classification model, where N is the number of target classes. The matrix compares the actual values to the model's predictions. This provides us with a comprehensive picture of how well our classification model is working and the kind of errors it is producing. Fig. 8 and Fig. 9 illustrate the classification performance for Indian Pines Corrected and Pavia University via the confusion matrix.

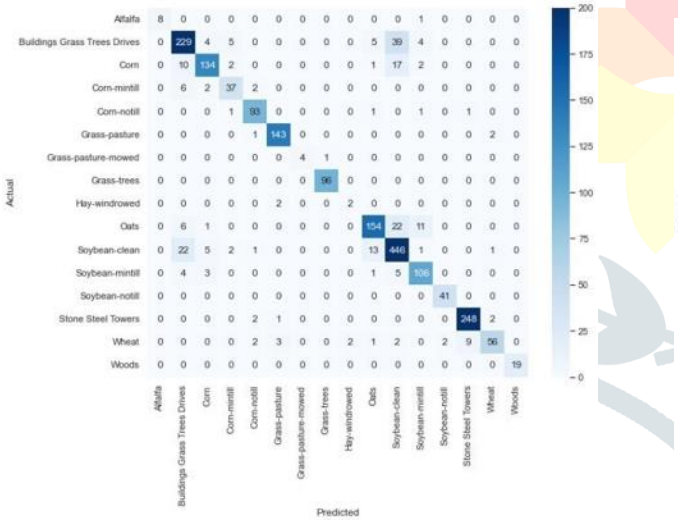

Fig. 8

Fig. 9

Classification maps are demonstrated for clear analysis of the original image and reconstructed image from tensor decomposition, in Fig. 10 and Fig. 11 for Indian Pines Corrected and Pavia University respectively.

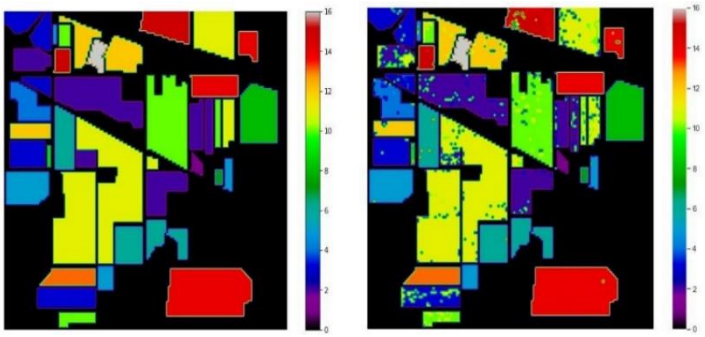

Fig. 11 Ground Truth on the left and Classification after processing on the right.

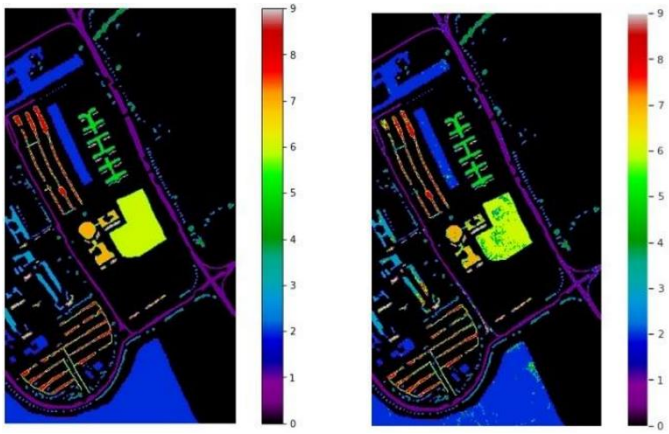

Fig. 12 Ground Truth on the left and Classification after processing on the right.

The classification accuracy achieved is tabulated in Table 4 below.

Table 4

Dataset	Classification Accuracy
Indian Pines Corrected	87.6%
Pavia University	95.6%

V. CONCLUSION

This study employs Tucker Decomposition via HOOI to compress data and achieves minimal relative error. SVM also does deep feature analysis and achieves excellent classification accuracy. As a future scope, non-iterative decomposition algorithms can be investigated so as to reduce computation time.

Due to the high dimensionality of the data and the scarcity of training examples, HS images present significant hurdles for supervised classification algorithms. These problems could limit the efficacy of classifiers together with the considerable intraclass variability (and interclass similarity) that are frequently present in HSI data. Several deep learning-based architectures have recently been developed, showing excellent potential in HS image classification, to address these issues. To extract the increasingly intricate properties layer by layer, the deep learning models for HS image classification can be utilized as a future scope.

REFERENCES

[1] Karami, M. Yazdi, and G. Mercier, “Compression of hyperspectral images using discrete wavelet transform and tucker decomposition,” in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 444-450, April 2012, doi: 10.1109/JSTARS.2012.2189200.

[2] P. Luigi Dragotti, G. Poggi and A. R. P. Ragozini, "Compression of multispectral images by three-dimensional SPIHT algorithm," *in IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 1, pp. 416-428, Jan. 2000, doi: 10.1109/36.823937.

[3] Yadav R.J and Nagmode M.S,” Compression of Hyperspectral Image Using PCA–DCT Technology” in *Saini, H., Singh, R., Reddy, K. (eds) Innovations in Electronics and Communication Engineering. Lecture Notes in Networks and Systems*, vol 7. Springer, Singapore., 2018, https://doi.org/10.1007/978-981-10-3812-9_28

[4] Li, Jin and Z. Liu, “Multispectral Transforms Using Convolution Neural Networks for Remote Sensing Multispectral Image Compression” in *Remote. Sensing*, 11(7), 759 (2019)

[5] Y. Zhang, P. Yuan, L. Jiang and H. T. Ewe, "Novel Data-Driven Spatial-Spectral Correlated Scheme for Dimensionality Reduction of Hyperspectral Images," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 3877-3890, 2022, doi: 10.1109/JSTARS.2022.3173999.

[6] J. Jiang, J. Ma, C. Chen, Z. Wang, Z. Cai and L. Wang, "SuperPCA: A Super pixelwise PCA Approach for Unsupervised Feature Extraction of Hyperspectral Imagery," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4581-4593, Aug. 2018, doi: 10.1109/TGRS.2018.2828029.

[7] J. Li, J. M. Bioucas-Dias and A. Plaza, "Spectral–Spatial Hyperspectral Image Segmentation Using Subspace Multinomial Logistic Regression and Markov Random Fields," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 3, pp. 809-823, March 2012, doi: 10.1109/TGRS.2011.2162649.

[8] Dua, Yaman & Kumar, Vinod & Singh, Ravi,” Comprehensive review of hyperspectral image compression algorithms” in *Optical Engineering*, vol 59, 090902 – 090902, 2020.

[9] Renu R K, Sowmya V, K P Soman, “Spatio-Spectral Compression and Analysis of Hyperspectral Images using Tensor Decomposition”, in Twenty Fourth National Conference on Communications (NCC), 2018.