Journal of Basic and Applied Engineering Research
p-ISSN: 2350-0077; e-ISSN: 2350-0255; Volume 3, Issue 10; July-September, 2016, pp. 831-837

http://www.krishisanskriti.org/Publication.html

Hyperspectral Remote Sensing Data Processing and Classification: A Tutorial

Sayan Mukhopadhaya

Indian Institute of Remote Sensing, ISRO
E-mail: sayanmukhopadhaya@gmail.com

Abstract—*This work provides guidelines on the practical tool to help in the processing of hyperspectral imagery data. Topics such as image preprocessing, classification and also ultimate image classification have been discussed with some guidelines. Due to large applicability and numerous methods available, this paper has focused mainly on the framework based on processing and classification in a step wise mode to segregate between different classes in a satellite image using hyperspectral imaging and error correction method along with data dimensionality reduction method. Hence, the reader is directed by each step and also uses the approaches as per the requirement of the problems faced. This work has been divided into several sections like the introduction of the work, the materials used and the several processing and pre-processing required to handle the hyperspectral data while used for classification of the images at certain spectral wavelengths.*

1. INTRODUCTION

The usage of hyperspectral images in remote sensing started from the early 1970's (Schowengerdt, 2006; van der Meer et al., 2012). Mukhopadhaya (2016) has shown that multispectral data can be used to detect the changes of a land use/land cover over many years by analyzing the different spectral reflectance values of the areas of the images. Hyperspectral imaging has developed as a new method to obtain spectrum in each area of larger spatial positions, so that one of the spectral wavelength could be used to form a detectable image (Clark, 1999).

This work is mainly aimed at focusing on the challenging task of executing a multivariate analysis on a hyperspectral remotely sensed imagery. Recognizing the main aim of discussing all the possible options will be much longer, so the main issues faced by a researcher are discussed here while processing hyperspectral imagery with their possible solutions. This will support to form a framework covering the two main aspects of hyperspectral data: a) processing and classification as one of the main focal point; and b) the removal of all the errors in the pre-processing of the data.

In this esteem, it has been hypothesized that the sensors used in the hyperspectral images can help to prevail over the customary problems handled while using the images of the multi-spectral sensors, like the problem faced in estimating the quantity and also while estimating the quality (Sahoo, 2000). Hyperspectral images can also be used to do many GIS related works, like the works stated by Richards and Jia (2006), Mukhopadhaya (2016a), etc. In view of all this, this work aims at presenting how classification models are successful using Hyperspectral-NIR to differentiate between land use/land cover using the ENVI tool. This will let us in handling the hyperspectral images properly as per our need and also to implement an economically viable process of handling larger areas that are not properly covered using multispectral remotely sensed data.

2. ADVANTAGES OF HYPERSPECTRAL IMAGERY

Frequent sampling done by destructive method is required for estimation of attributes of soil and vegetation by direct field practices. These practices are challenging, requirement of intense labour and costlier with respect to both time and money. Hence, these prevailing methods cannot be extended over large areas (Sahoo, 2000). Nevertheless, hyperspectral image remote sensing practice can be used to perform a reckonable evaluation of soil and vegetation. One of the major drawback of broadband remotely sensed products is that of the usage of average of the information collected spectrally, hence results in the loss of important information that are present in a specific band (Blackburn, 1998, Thenkabail et al., 2000).

On a general note, a huge range of information can be collected from the reflectance characteristics associated to the nature of the material of soil and other classes (Stoner & Baumgardner, 1981; Baumgardner et al., 1985). The data obtained from reflectance have been used for predicting numerous properties of the class types. For characterization of the classes hyperspectral data can be used, like Kumar et al. (2006) studied 87 samples of soil.

3. MATERIALS USED

The material used was created by the United States Geological Survey (USGS) and contains E0-1 Hyperion data of Jharia District with 242 bands dated May 4th, 2009, with pixel size of 30m spatial resolution. The ground elevation of the area is of 220m with central latitude of 23º36.75'36.7275''N and central

longitude of 86º23'5.75''E. The materials used in this work was provided by the Indian Institute of Remote Sensing (IIRS), Indian Space Research Organization.

The software used was of ENVI with Hyperion tool embedded onto it.

4. STEPS USED

The steps used to process the data are as follows:

Creating ROI and sub setting the image spatially

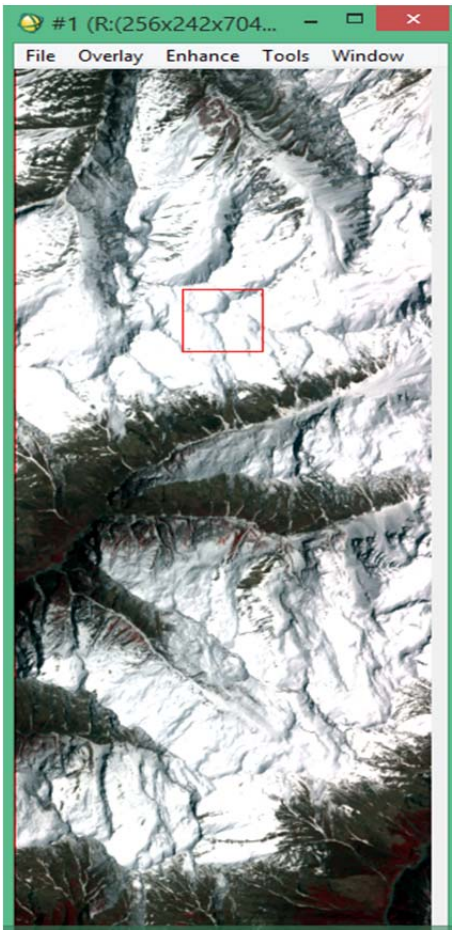


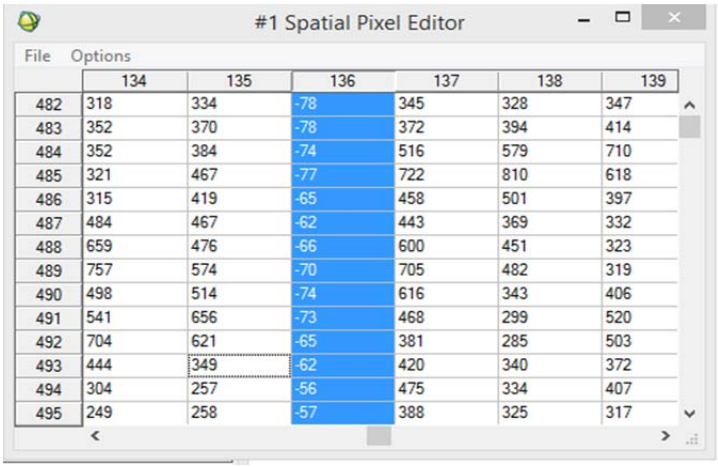
Fig. 1: Input Image

The raw input has been shown in Fig.1 which consists of 242 bands along with some bad bands and columns. The next step is to create a region of interest (ROI), which is shown in red box.

Removal of drop lines, bad bands and Spectral subset:

This step shows the region of interest (ROI) both spatially and spectrally. The spectral subset is used for removing bad bands and also the removal of drop lines. The bad bands are those bands where there is no data shown, and the image is shown in either total black or total white. These band numbers are to be noted for further handling of the imagery used. Now, the spectral subset is formed, and the bands having no information are excluded. For removal of drop lines, the column having zero pixel values are to be selected individual as shown in. The selected lines are replaced by the average values of the corresponding two lines near to the line missing as shown in figures 2, 3 and 4. The same method is to be constant for all the other bands having drop lines or bad columns.

Fig. 2: Spatial/Spectral ROI



	134	135	136	137	138	139
482	318	334	-78	345	328	347
483	352	370	-78	372	394	414
484	352	384	-74	516	579	710
485	321	467	-77	722	810	618
486	315	419	-65	458	501	397
487	484	467	-62	443	369	332
488	659	476	-66	600	451	323
489	757	574	-70	705	482	319
490	498	514	-74	616	343	406
491	541	656	-73	468	299	520
492	704	621	-65	381	285	503
493	444	349	-62	420	340	372
494	304	257	-56	475	334	407
495	249	258	-57	388	325	317

Fig. 3: Spatial Pixel Editor Interface

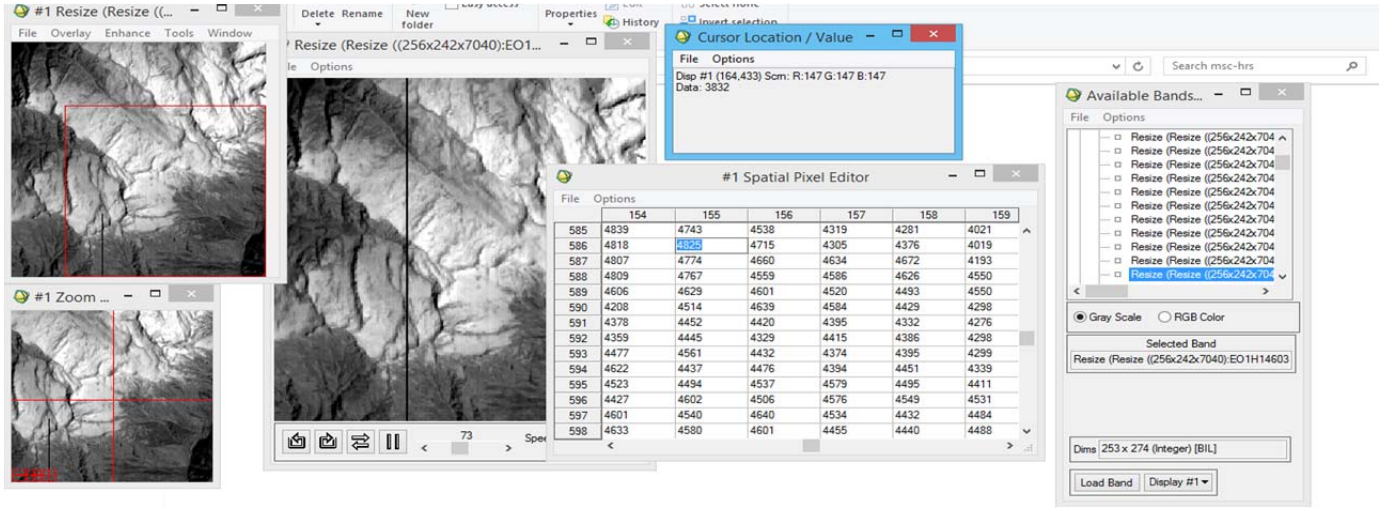


Fig. 4. Removal of Bad bands using subset of image.

Atmospheric Corrections:

The need of atmospheric correction is required as the solar radiation contains the atmospheric information, hence the information received by the sensor contains both earth surface information as well as the atmospheric information. So, to study the earth surface information, the atmospheric information need to be removed. In ENVI, the FLAASH (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes) module has been used for retrieving the spectral information from the hyperspectral images.

Reduction of Data dimensionality using PCA:

For, this work we have also taken multivariate regression into account as linear regression can only estimate the coefficients of the equation used (linear).

Hyperspectral sensor collects a huge amount of data in several number of bands, but these large amount of data subsequently increases the level of computation to reach a noticeable result and this also results a large degree of redundancy of data. Hence, to remove the redundant bands thus decreasing in computational load, it is required to reduce the data dimensionality. There are lot of processes for reduction of dimensionality, but here PCA (Principal Component Analysis) is used.

PCA has been used to yield uncorrelated bands, for separating noise components and also to lessen the dimensionality of the data used. Principal Component Transformation has been used to yield the uncorrelated bands in the output by locating a newly set orthogonal axes having the center of origin at the mean of the data. These outputs are rotated so that the variance is maximum. PC bands are formed by combination in a linear manner of the original spectral bands. These PC bands are also uncorrelated.

PCA was performed on atmospheric corrected image (Fig. 5) and fig. 6 is the obtained output based on the Eigen values for each band. The band with high Eigen values consists of more information.

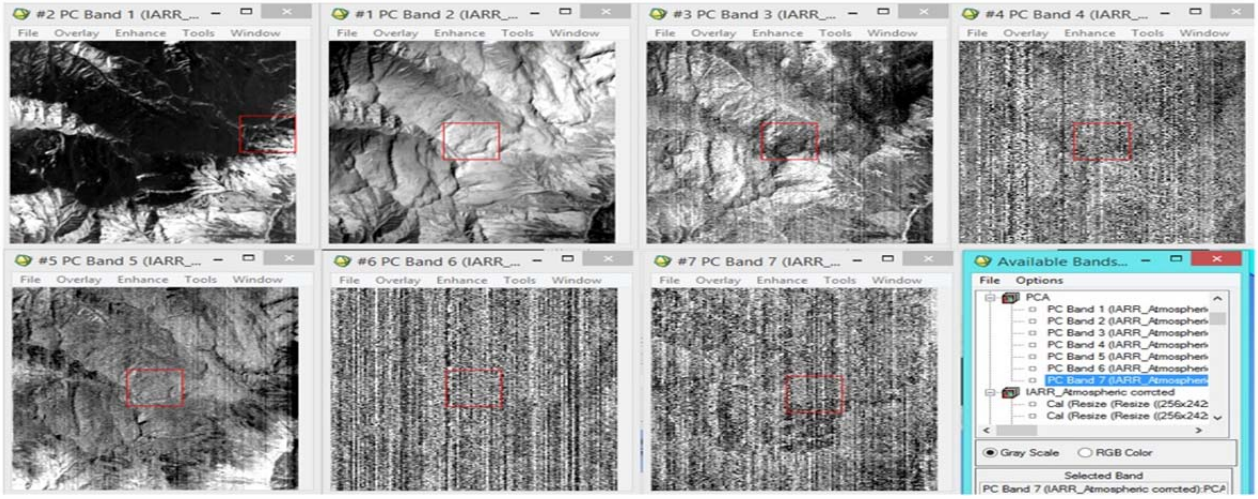
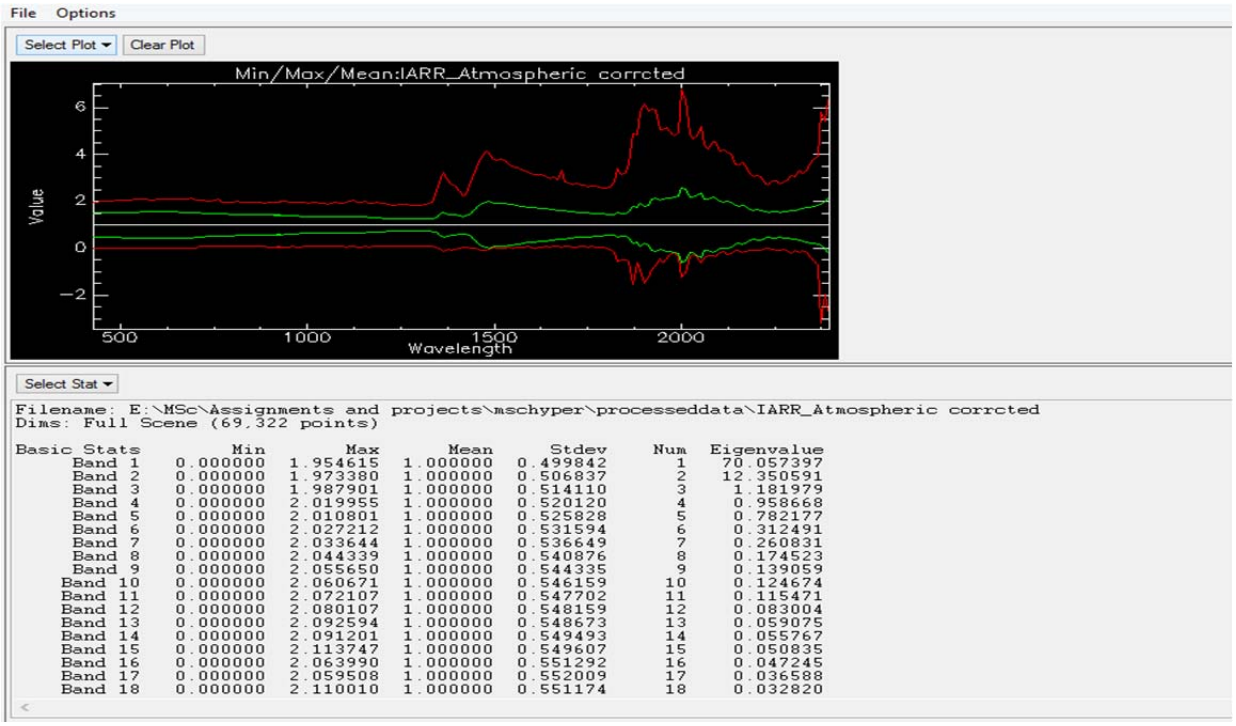


Fig. 5: The image output at different steps in PCA during reduction of data



Basic Stats	Min	Max	Mean	Stdev	Num	Eigenvalue
Band 1	0.000000	1.954615	1.000000	0.499842	1	70.057397
Band 2	0.000000	1.973380	1.000000	0.506837	2	12.350591
Band 3	0.000000	1.987901	1.000000	0.514110	3	1.181979
Band 4	0.000000	2.019955	1.000000	0.520120	4	0.958668
Band 5	0.000000	2.010801	1.000000	0.525828	5	0.782177
Band 6	0.000000	2.027212	1.000000	0.531594	6	0.312491
Band 7	0.000000	2.033644	1.000000	0.536649	7	0.260831
Band 8	0.000000	2.044339	1.000000	0.540876	8	0.174523
Band 9	0.000000	2.055650	1.000000	0.544335	9	0.139059
Band 10	0.000000	2.060671	1.000000	0.546159	10	0.124674
Band 11	0.000000	2.072107	1.000000	0.547702	11	0.115471
Band 12	0.000000	2.080107	1.000000	0.548159	12	0.083004
Band 13	0.000000	2.092594	1.000000	0.548673	13	0.059075
Band 14	0.000000	2.091201	1.000000	0.549493	14	0.055767
Band 15	0.000000	2.113747	1.000000	0.549607	15	0.050835
Band 16	0.000000	2.063990	1.000000	0.551292	16	0.047245
Band 17	0.000000	2.059508	1.000000	0.552009	17	0.036588
Band 18	0.000000	2.110010	1.000000	0.551174	18	0.032820

Fig. 6. The output statistics file

Data dimensionality reduction and Noise removal using MNF:

MNF (Minimum Noise Fraction) rotation transforms are useful in determining the inherent dimensions of the input image, so that to separate noise in the data and also in the reduction of the computational necessities for the further processing. MNF consists of different PCA rotations.

In this step the input data was the atmospherically corrected data and MNF method was applied to reduce the noise and the dimensionality of the dataset to get the uncorrelated bands from the data. Finally, the MNF output has 8 bands with first 2 bands having more information with very less noise (Fig. 7). Along with generated output statistics file. This method used the shift difference to calculate the noise and uses PCA twice for creating the noise free components.

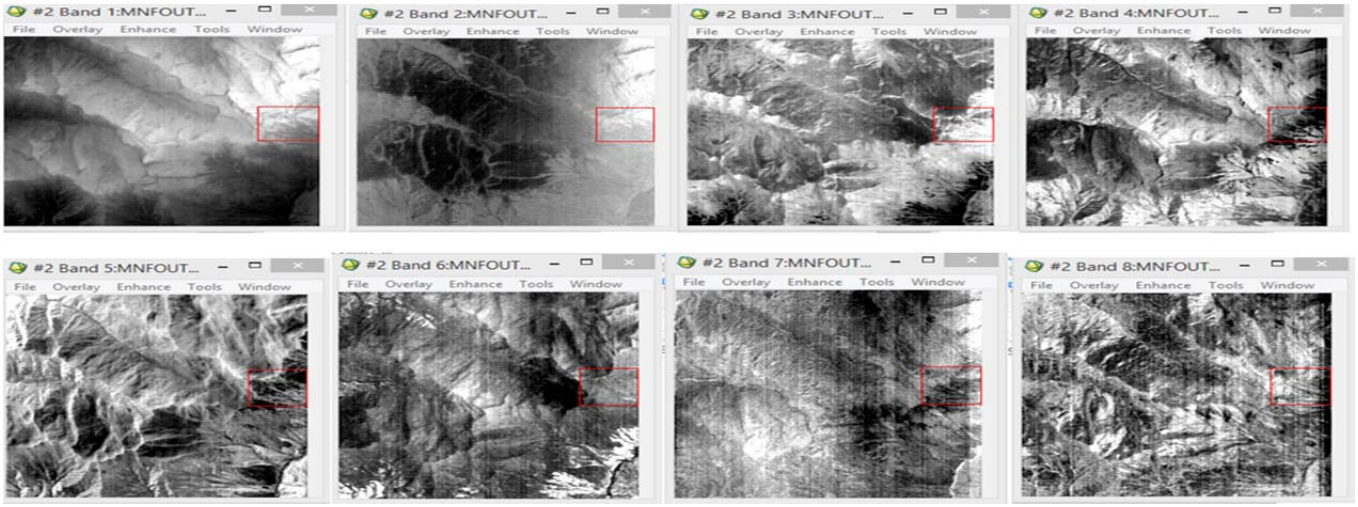


Fig. 7: The output having 8 bands, with the first 2 bands having less noise.

End member collection

Firstly using FAST Pixel purity index (PPI), the purest pixels were collected and the ROI is used for n-d visualizer to generate classes of end members. N-D visualizer of clouds of pixels obtained from PPI which are the reflectance values of each pixel. For the different combination of input bands (8 as in MNF) the classes of end members were collected.

PPI has been used to detect the most spectrally pure pixels in the input image. These corresponds to endmember mixing. PPI is calculated by iteratively projecting n-D scatter plots on a unit vector taken randomly. FAST PPI method was used to place the data obtained from image and then performing the following computations in the memory. This method is much quicker than the PPI which is mainly disk based, but for FAST PPI there is a requirement of adequate memory. ENVI indicates the adequate memory required for the process of FAST PPI.

The brighter pixels indicate the purest pixels. Later a threshold PPI was given 150 and 2000 clouds of pixels were collected and given as input to n-D visualizer where new classes were defined.

The n-D visualizer is a window plot showing n-D scatter plot of the image data, which is selected for the visual interpretation. N-D visualization is the projection of n-dimensional data on a 2-dimensional plane.

Fig. 9 shows the sample visualizer view of the collected points for 8-bands. The classes obtained were exported as region of interest. The classes obtained were snow cover, barren land and less dense vegetation.

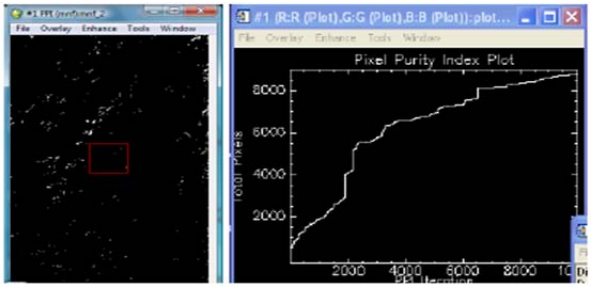


Fig. 8: PPI image and the plot

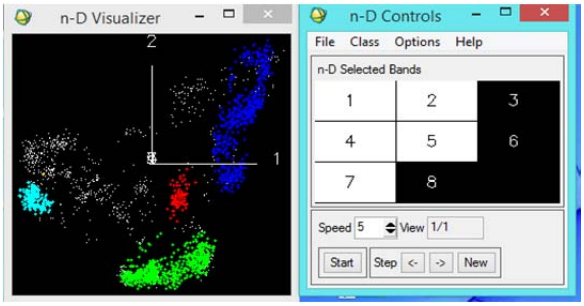


Fig. 9: The visualization in n-D visualizer

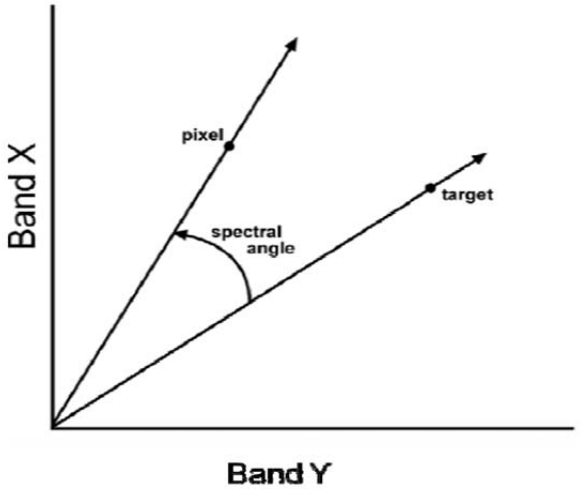


Fig. 10. Spectral Angle

Classification:

Using the End member collection option the classification was then performed by importing the ROI (classes) for the atmospheric corrected data. The classification used was Spectral angle mapper. Spectral angle mapper is extensively used as a spectral measure of similarity. The algorithm is based on the properties of the data taken and the training data selected. It evaluates the similarity based on spectral properties between the reference spectrum and the spectrum of the data. The measure of segregation is the angle between the two spectra. The following equation shows the formulation of spectral angle mapper.

$$\cos\phi = \frac{\sum_{i=1}^{n} e_i\, r_i}{\sqrt{\sum_{i=1}^{n} e_i^2}\ \sqrt{\sum_{i=1}^{n} r_i^2}}$$

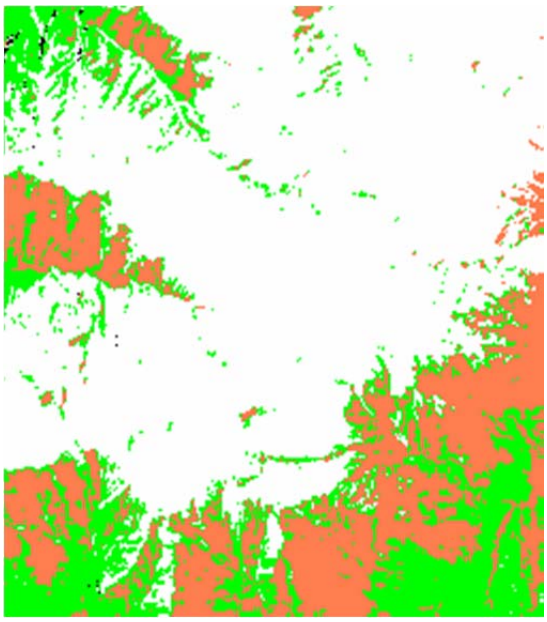
where, ϕ = spectral angle (fig. 10)

e = given spectra of image

r = spectra of end member (reference)

n = number of classes.

The angle of discrimination was given in radians of 0.8. The output of the classification is shown in fig. 11.



= snow, = barren land, = vegetation

Fig. 11: The classified image.

The output image shows the classified outputs containing 3 classes (snow covered, barren land and less dense vegetation).

5. CONCLUSION

This work shows some ideas on dealing with many issues related to the development of process of classification using hyperspectral imagery in a practical point of view. Besides, key points and some basic hints of how to deal with hyperspectral image classification are also discussed. As stated before, every single process of image classification requires a custom-made solution. So, the steps taken in each level should be taken as an optimum key for this case only.

Going by the obtained results, the image was properly classified among the different classes. Lastly, all should look forward for the application of hyperspectral and digital image processing that suits best as per the purpose.

Here, we have discussed about how ENVI with hyperion tool can be used to do the image analysis of the hyperspectral data. However, there are many other commercial tools available. Hence, to fulfill our requirement we should always answer the following scientific questions:

-*Do hyperspectral imagery needed for the work?* There are two main spheres of applications of hyperspectral data: the application where the structural part of a model data is explored, and the other application is regarding where the total object is taken as a single homogeneous object but it has been distributed on a surface with other objects (like, discussed in this paper, the image classification of LULC). For the case one, the distribution of any component can be accomplished by using imaging techniques. For the other case, the methods used for imaging really hangs on the ultimate purpose of the work to be achieved. For instance, in the case of this work, the final aim was to classify the image as per the classes using the NIR imaging techniques. On principle, a LULC image can be well thought of as a homogeneous surface. This fact is very essential for the development of the model for classifying the image, such that a single NIR spectrum can be used to characterize the whole ROI. This shows that development of a model does not need the knowledge of the imaging techniques but, requires knowledge of the final product.

-*What kind of spectral radiation is required?* The answer of the section is very important and depends mainly on the final purpose and also the complexity of the image used. The sole advice can be thought of is that to study the features of the image properly to reach the target. Nonetheless, there are a lot of cases where visible imaging could be enough.

-*What spatial resolution is required?* It is evident that to illustrate the distribution of all the aspects of an image, a spatial resolution in meters is required.

-*What spectral resolution is required?* Along with the extent of modality, the final application of the work helps in deciding the spectral resolution. However, in this case, it has been also a question of what is the extent of spectral radiation to be chosen, such that the discrimination between the classes can be done easily.

-*What multivariate model should be used?* This aspect is very important as the final output depends on the nature of the multivariate chosen for the work. There are many methods out of them some are highlighted in this work like PCA and MNF.

-*What kind of software is required?* As explained earlier, this is fully dependent on the ability to handle big data analysis. However, the main aim to master hyperspectral image analysis, it is recommended the usage of software that will allow to set down our own modules.

6. ACKNOWLEDGEMENT

The work has been possible due to the help of Mr.Vinay Kumar, scientist/engineer-SD from Indian Institute of Remote Sensing, ISRO.

REFERENCES

[1] Baumgardner, M. F., Silva, L. F., Biehl, L. L., & Stoner, E. R. (1985). Reflectance Properties of Soils. *Advances in Agronomy*, *38*(C), 1–44. http://doi.org/10.1016/S0065-2113(08)60672-0

[2] Blackburn, G. A. (1998). Quantifying Chlorophylls and Caroteniods at Leaf and Canopy Scales. *Remote Sensing of Environment*, *66*(3), 273–285. http://doi.org/10.1016/S0034-4257(98)00059-5

[3] Clark, R. N. (1999). *Spectroscopy of rocks and minerals, and principles of spectroscopy*. *Remote sensing for the earth sciences: Manual of remote sensing* (Vol. 3). http://doi.org/10.1111/j.1945-5100.2004.tb00079.x

[4] Kumar, M., & Philip, L. (2006). Adsorption and desorption characteristics of hydrophobic pesticide endosulfan in four Indian soils. *Chemosphere*, *62*(7), 1064–1077. http://doi.org/10.1016/j.chemosphere.2005.05.009

[5] Mukhopadhaya, S. (2016a). GIS-based Site Suitability Analysis : Case Study for Professional College in Dehradun. *Journal of Civil Engineering and Environmental Technology*, *3*(1), 60–64.

[6] Mukhopadhaya, S. (2016b). Land use and land cover change modelling using CA-Markov Case study : Deforestation Analysis of Doon Valley. *Journal of Agroecology and Natural Resource Management*, *3*(1), 1–5.

[7] Richards, J. A., & Jia, X. (2006). *Remote Sensing Digital Image Analysis*. *New York*. Berlin/Heidelberg: Springer-Verlag. http://doi.org/10.1007/3-540-29711-1

[8] Sahoo, R. N. (2000). *Hyperspectral Remote Sensing* (Vol. 2). New Delhi: I.A.R.I. Retrieved from www.iasri.res.in/ebook/GIS_TA/M2_4_HYSRS.pd

[9] Stoner, E. R., & Baumgardner, M. F. (1981). Characteristic variations in reflectance from surface soils. *Soil Science Society of America Journal*, *45*, 1161–1165.

[10] Thenkabail, P. S., Smith, R. B., & De Pauw, E. (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment*, *71*(2), 158–182. http://doi.org/10.1016/S0034-4257(99)00067-X

[11] Van der Meer, F. D., van der Werff, H. M. A., van Ruitenbeek, F. J. A., Hecker, C. A., Bakker, W. H., Noomen, M. F., Woldai, T. (2012). Multi- and hyperspectral geologic remote sensing: A review. *International Journal of Applied Earth Observation and Geoinformation*. http://doi.org/10.1016/j.jag.2011.08.002