

# Hyperspectral Image Classification Using Unsupervised Algorithms

Sahar A. El\_Rahman<sup>1,2</sup>

<sup>1</sup>Electronics, Computers Systems and Communication,  
Electrical Department  
Faculty of Engineering-Shoubra, Benha University  
Cairo, Egypt

<sup>2</sup> Computer Science Department, College of Computer  
and Information Sciences  
Princess Nourah Bint Abdulrahman University  
Riyadh, Saudi Arabia

**Abstract**—Hyperspectral Imaging (HSI) is a process that results in collected and processed information of the electromagnetic spectrum by a specific sensor device. It's data provide a wealth of information. This data can be used to address a variety of problems in a number of applications. Hyperspectral Imaging classification assorts all pixels in a digital image into groups. In this paper, unsupervised hyperspectral image classification algorithms used to obtain a classified hyperspectral image. Iterative Self-Organizing Data Analysis Technique Algorithm (ISODATA) algorithm and K-Means algorithm are used. Applying two algorithms on Washington DC hyperspectral image, USA, using ENVI tool. In this paper, the performance was evaluated on the base of the accuracy assessment of the process after applying Principle Component Analysis (PCA) and K-Means or ISODATA algorithm. It is found that, ISODATA algorithm is more accurate than K-Means algorithm. Since The overall accuracy of classification process using K-Means algorithm is 78.3398% and The overall accuracy of classification process using ISODATA algorithm is 81.7696%. Also the processing time increased when the number of iterations increased to get the classified image.

**Keywords**—hyperspectral imaging; unsupervised classification; K-Means algorithm; ISODATA algorithm; ENVI

## I. INTRODUCTION

Remote sensing is the art and science to obtain information about an object, area. It is viewed as the measurement and analysis of electromagnetic radiation transmitted through, reflected from, or absorbed and dissipated by the ambience, the hydrosphere and by material at or near the land surface, for the purpose of interpreting and managing the Earth's resources and surroundings. Optical remote sensing makes use of visible, near infrared and short-wave infrared sensors to make pictures of the earth's surface by observing the solar radiation reflected from targets on the background as indicated in Fig. 1. Different materials reflect and absorb differently at different wavelengths. Thus, the targets can be differentiated by their spectral reflectance signatures in the remotely sensed images [1][2][3].

Hyperspectral sensors such as the Airborne Imaging Spectro-radiometer for applications (AISA) enabled the construction of an effective, continuous reflectance spectrum for every pixel in the scene. These schemes can be applied to discriminate among earth surface features [1][2][4].

By analogy to grasp the hyperspectral imaging concept better is that the human can sight visible light in the main three (RGB) bands, i.e. red, green, and blue, whereas spectral imaging divides the spectrum into many more bands which are infrared bands, RGB bands, and ultraviolet bands (see Fig. 2) [5][6].

The “hyper” in the word "hyperspectral" refer to “too many” as in “over” and indicate the massive number of wavelength bands. Hyperspectral imaging is spectrally specified, which indicates that it provides vast spectral information to distinguish and identify spectrally singular materials. Hyperspectral imaging supplies the possibility of further precise and exhaustive information extraction than potential with any other kind of remote sensing data [7]. It owns an increased capability that enhances the chance of detecting concerning materials and supplies more information needful for recognizing and classifying these materials [8][9]. The HSI pixels form spectral vectors demonstrate the spectral characteristics of these materials in the sight [10].

There are some limitations in hyperspectral images are images distortion that are resulting from the spherical of the earth, giving inaccuracies in the properties such as directions, distances, and scale. These distortions of images can be shadows; such as the shadow covers a specific area to be studied, and the brightness of the light on a specific area [11]. The size of the pixel where it is possible to be relatively large so that a pixel can contain a lot of properties and that is difficult to classify, or be very small in terms of not contains characteristics can be classified is one of the hyperspectral imaging limits [12].

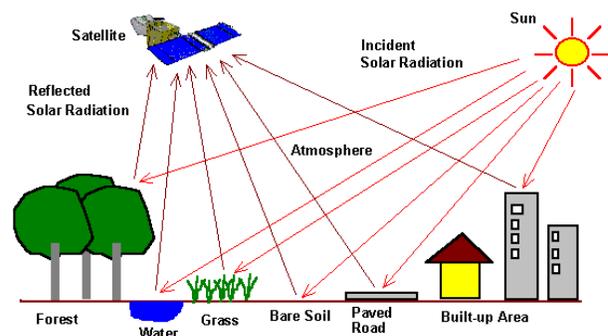


Fig. 1. Optical and Infrared Remote Sensing

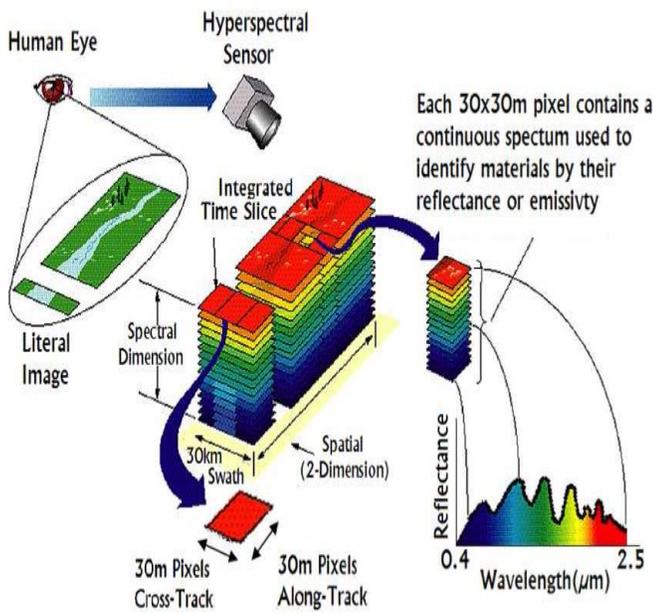


Fig. 2. HyperSpectral Imaging

## II. CLASSIFICATION

The main purpose of classification of satellite imagery is to assess landscape properties accurately and extract required information [13]. Unsupervised and supervised classification algorithms are the two prime types of classification. Unsupervised classification is shown in Fig. 3 [14]. The classification chain is unsupervised, where the classification algorithms used are K-Means algorithm and ISODATA.

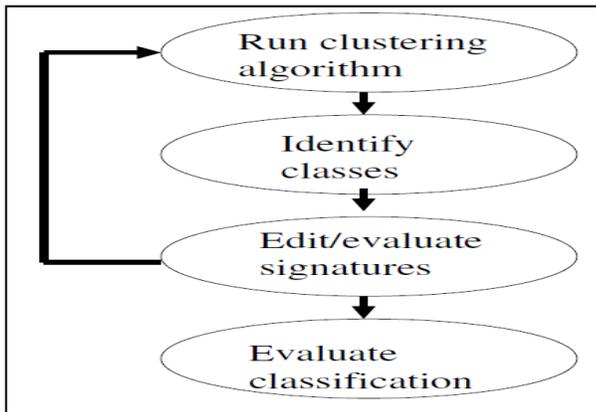


Fig. 3. Unsupervised Classification

### A. K-Means Classifier

The K-means algorithm is a straightforward process for deriving the mean of a group of K-sets. The purpose of the K-Means algorithm is to reduce the cluster variability (see Fig. 4) [15][16][17]. The process of the K-means algorithm is described in the following pseudo-code [18].

```

1   Make initial guesses for the means  $m_1, m_2,$ 
   ...,  $m_k$ 
2   Until there are no changes in any mean
3       Use the estimated means to classify the
       examples into clusters
4   For i from 1 to k
5       Replace  $m_i$  with the mean of all
       of the examples for Cluster i
6   end_for
  
```

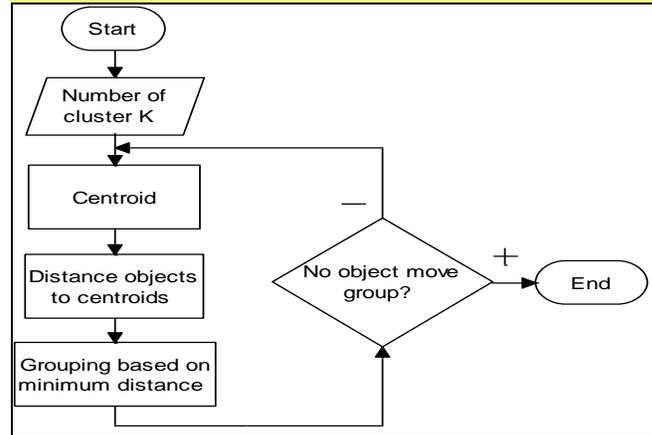


Fig. 4. K-means classifier

### B. Iterative Self-Organizing Data Analysis Technique Algorithm (ISODATA)

The ISODATA algorithm is one of the most utilized methods in the unsupervised classification (see Fig. 5). In more particular, the steps in ISODATA clustering are as follows [19][20][21]:

1. Cluster centers are randomly placed.
2. Pixels are assigned based on the shortest distance to the center.
3. The standard deviation within each cluster, and the distance between cluster centers are calculated
  - a. Clusters are split
  - b. Clusters are merged
 If one or more standard deviation > the threshold.
4. A second iteration is performed with the new cluster centers.
5. Further iterations are performed until:
  - a. The average inter-center distance falls below the user-defined threshold.
  - b. The average change in the inter-center distance between iterations is less than a threshold, or the maximum number of iterations is reached.

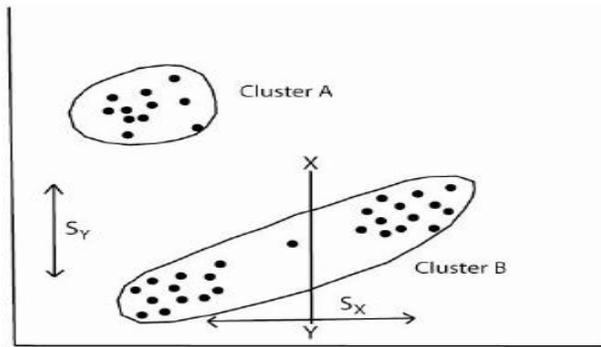


Fig. 5. ISODATA Classifier

### III. RESULTS AND DISCUSSION

The unsupervised classification was applied on a hyperspectral image using ENVI tool. The hyperspectral dataset, which has been applied to, is an image of Washington DC. The two steps that applied to the hyperspectral image are Principle Component Analysis (PCA) and K-Means or ISODATA algorithms. The result of applying K-Means algorithm and ISODATA algorithm is a classified image. Process time increased when the number of iterations increased to get the classified image. In this work, statistical information calculated from the classified image data and seen that K-Means algorithm and ISODATA algorithm are accurate since each pixel in the image is classified into a class that is not Unclassified Class. ISODATA algorithm is more accurate than K-Means algorithm since that overall accuracy of classification process using ISODATA algorithm is 81.7696% and the overall accuracy of classification process using K-Means algorithm is 78.3398%.

#### A. ENVI (Environment of Visualizing Images)

ENVI is an image processing system. It was designed to process remotely sensed data. It provides comprehensive data visualization and analysis for images. It has the ability of treating a broad set of scientific data formats [22][23].

#### B. Cases Studies

- Comparison between the results of applying different RGB bands on the same hyperspectral image.
- Studying the effects of changing the number of iterations in the process of classification on the accuracy of classification.
- Comparison between the result of applying K-Means algorithm in the first time with the result of applying the ISODATA algorithm at the second time.
- Comparison between the result of applying the PCA algorithm then applying K-Means algorithm in the first time with the result of applying only K-Means algorithm in the second time.

1) *Case Study 1:* It is using different RGB bands (see Table I) to the same hyperspectral image (Washington DC), and applying Principle Component Analysis (PCA) and K-means. The results of applying PCA on the image using Test values in Table I are shown in Fig. 6. The number of classes

is 6 and the number of iterations is 3, where Fig. 7 shows the results of applying K-Means algorithm on the output images of PCA.

TABLE I. CASE STUDIES USING DIFFERENT RGB BANDS

Test Values	R	G	B
1	180	95	35
2	176	88	23
3	29	59	129

2) *Case Study 2:* It is seen that when the number of iterations to classify a hyperspectral image is increased, the overall accuracy will increase while the overall accuracy will decrease when the number of iterations is decreased. This is approved through the above experiments. Table II is a summary of the results of the experiments.

3) *Case Study 3:* It is about using different algorithms on the same hyperspectral image (Washington DC), these algorithms are K-means and Iterative Self-Organizing Data Analysis Technique Algorithm (ISODATA). In both case as shown in Fig. 8, it is applied as a first step PCA then either K-Means algorithm or ISODATA as a second step. The selected bands are PC Band 172 for R, PC Band 86 for G, and PC Band 24 for B. Number of classes is 6 classes and the maximum iterations is 3 are chosen as K-Means parameters. Fig. 8-a shows the result of applying K-Means algorithm. Then, the number of classes range is from 4 to 6 and the maximum iterations is 3 are chosen as ISODATA parameters. Fig. 8-b shows the result of applying ISODATA.

4) *Case Study 4:* It is about studying the effect of implementing PCA on classification result, as shown in Fig. 9-a, comparing to implementing just K-Means algorithm without applying PCA as shown in Fig. 9-b. This study is implemented on the hyperspectral image (Washington DC).

#### C. Class Statistics

1) *Calculating Class statistics based on applying the K-means algorithm:* Calculating statistics based on applying K-means algorithm on Washington DC image results, Fig. 10-a shows the Means for all classes and Fig. 10-b shows the standard deviation for all classes, those represent the relation between the band number and the value. Fig. 10-c shows the basic statistics include minimum, maximum and mean for each band for Tree class. Fig. 10-d shows the standard deviation for Tree class. Table III shows the class distribution summary and Table IV indicates confusion matrices using the ground truth image. A total class error is indicated in Fig. 11.

2) *Calculating Class Statistics based on applying ISODATA algorithm:* Calculate statistics based on applying the ISODATA algorithm on Washington DC hyperspectral image results, Fig. 12-a shows the Means for all classes and Fig. 12-b shows the Stdev for all classes, those represent the relation between the band number and the value. Fig. 12-c shows the basic statistics include minimum, maximum and mean for each band for Tree class. Fig. 12-d shows the standard deviation for Tree class. Table V shows the class distribution summary and Table VI indicates confusion

matrices using the ground truth image. A total class error is indicated in Fig. 13.

TABLE II. SUMMARY OF THE EXPERIMENTS OF CASE STUDY 2

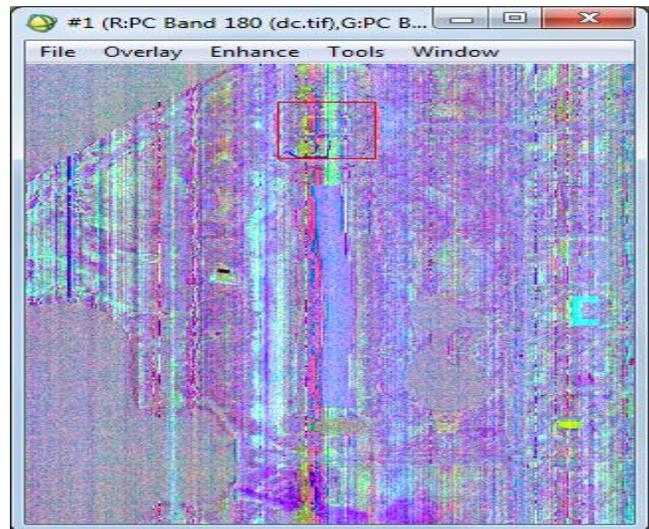
Experiment Description	Overall Accuracy	Class Percentage						
		Unclassified	Roof	Grass	Land	Trail	Road	Water
Comparison between the result of a classified image with only one iteration (Ground Truth) and a classified image with three iterations.	62.7079%	0.0	100	100	100	100	83.29	21.93
Comparison between the result of a classified image with three iterations (Ground Truth) and a classified image with ten iterations.	87.6110%	0.0	100	100	100	100	90.34	8.87
Comparison between the result of a classified image with only one iteration (Ground Truth) and a classified image with ten iterations.	57.8171%	0.0	100	100	100	100	89.59	27.06
Comparison between the result of a classified image with three iterations (Ground Truth) and a classified image with fifteen iterations.	87.6110%	0.0	100	100	100	100	90.34	8.87

TABLE III. CLASS DISTRIBUTION SUMMARY (K-MEANS ALGORITHM)

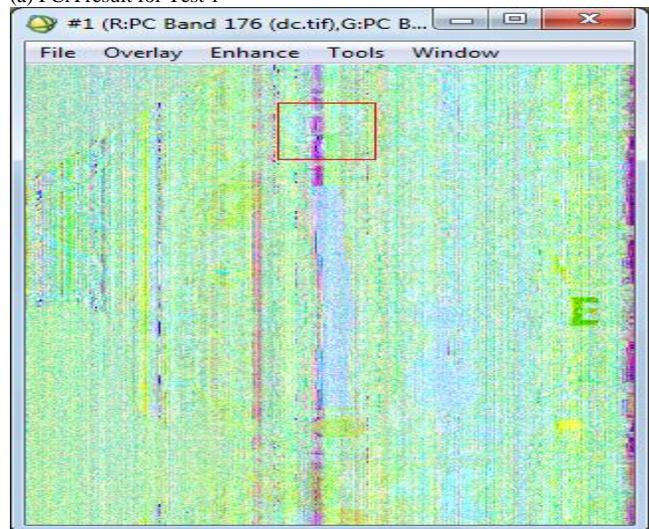
Unclassified	0 points (0.000%)
Roof	51,689 points (13.154%)
Grass	98,094 points (24.963%)
Tree	56,529 points (14.385%)
Trail	58,722 points (14.944%)
Road	69,678 points (17.732%)
Water	58,248 points (14.823%)

TABLE IV. CONFUSION MATRIX USING K-MEANS ALGORITHM

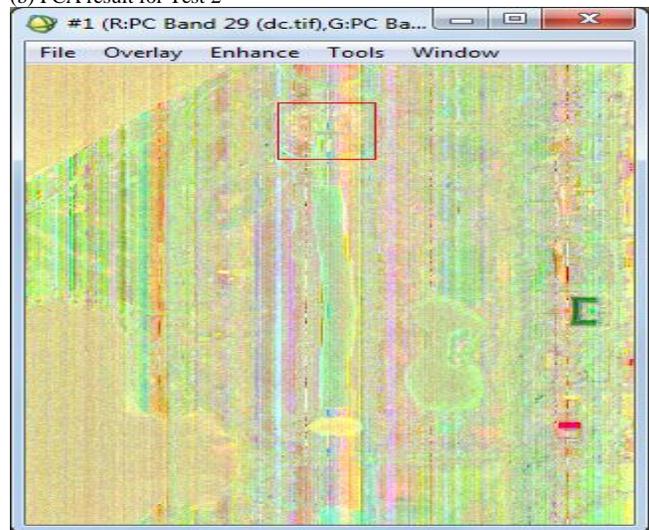
Overall Accuracy = (307844/392960) 78.3398%								
Kappa Coefficient = 0.7373								
Ground Truth (Percent)								
Class	Unclassified	Roof	Grass	Tree	Trail	Road	Water	Total
Unclassified	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Roof	0.00	100.00	1.59	0.00	14.92	0.00	0.00	13.15
Grass	0.00	0.00	93.34	17.02	0.15	0.00	0.00	24.96
Tree	0.00	0.00	5.07	46.06	32.74	0.00	0.00	14.39
Trail	0.00	0.00	0.00	33.14	51.08	6.26	0.00	14.94
Road	0.00	0.00	0.00	3.77	1.11	84.79	0.00	17.73
Water	0.00	0.00	0.00	0.00	0.00	8.95	100.00	14.82
Total	0.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00



(a) PCA result for Test 1



(b) PCA result for Test 2



(c) PCA result for Test 3

Fig. 6. Case Study 1: Using Different RGB Bands after Applying PCA



(a) K-Means result for Test 1

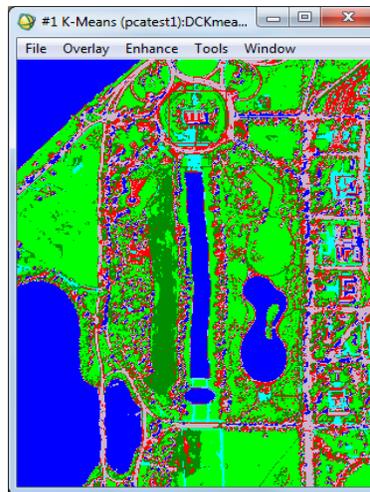


(b) K-Means result for Test 2

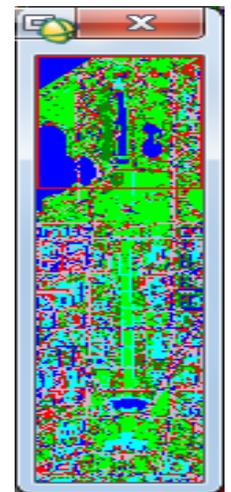
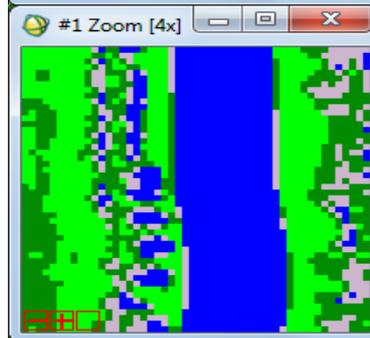
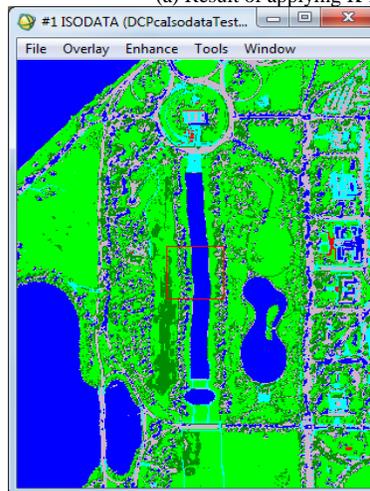


(c) K-Means Result for Test 3

Fig. 7. Case Study 1: Using Different RGB Bands after Applying PCA and K-Means



(a) Result of applying K-means algorithm.



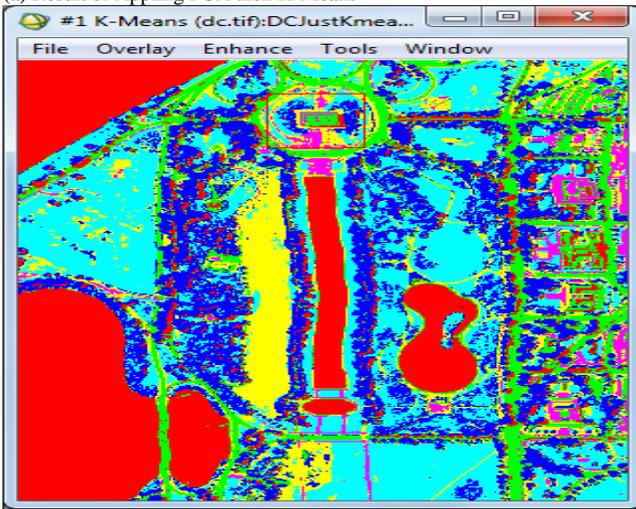
Water	Road
Grass	Roof
Tree	Trail

Water	Road
Grass	Roof
Tree	Trail

Fig. 8. The Results of Case Study 3



(a) Result of Applying PCA then K-Means



(b) Result of Applying K-Means without PCA

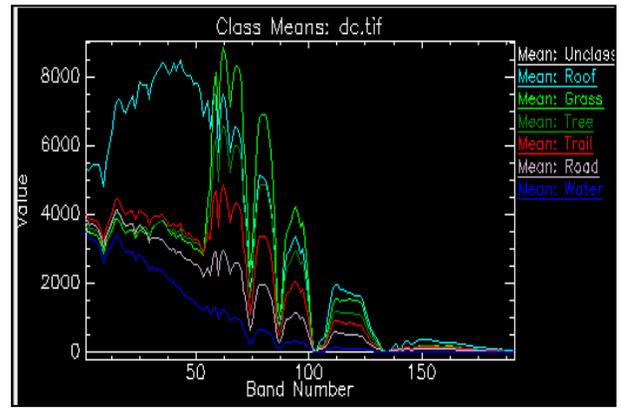
Fig. 9. The Results of Case Study 4

TABLE V. CLASS DISTRIBUTION SUMMARY (ISODATA ALGORITHM)

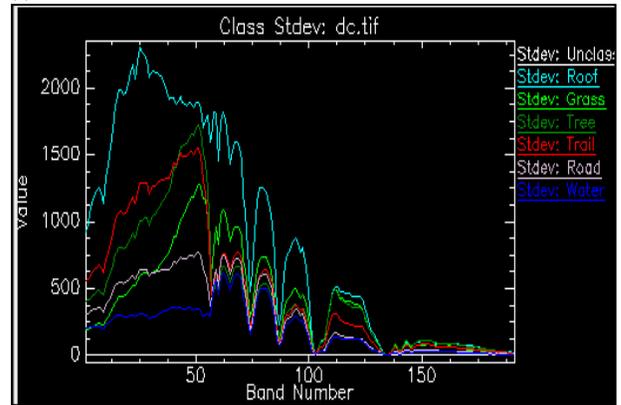
Unclassified	0 points (0.000%)
Trail	6.697 points (1.704%)
Grass	119.476 points (30.404%)
Tree	62.947 points (16.019%)
Road	76.234 points (19.400%)
Water	70.155 points (17.853%)
Roof	57.451 points (14.620%)

TABLE VI. CONFUSION MATRIX USING ISODATA ALGORITHM

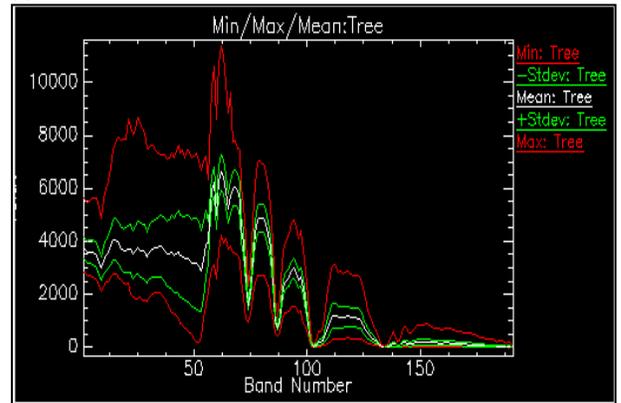
Overall Accuracy = (321322/392960) 81.7696%								
Kappa Coefficient = 0.7726								
Ground Truth (Percent)								
Class	Unclassified	Trail	Grass	Tree	Road	Water	Roof	Total
Unclassified	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Trail	0.00	36.54	0.10	0.00	0.00	0.00	2.32	1.70
Grass	0.00	0.00	99.81	21.07	0.00	0.00	10.95	30.40
Tree	0.00	0.00	0.09	60.90	14.40	0.00	8.88	16.02
Road	0.00	0.00	0.00	18.03	78.73	4.96	0.00	19.40
Water	0.00	0.00	0.00	0.00	1.53	95.04	0.00	17.85
Roof	0.00	63.46	0.00	0.00	5.34	0.00	77.85	14.62
Total	0.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00



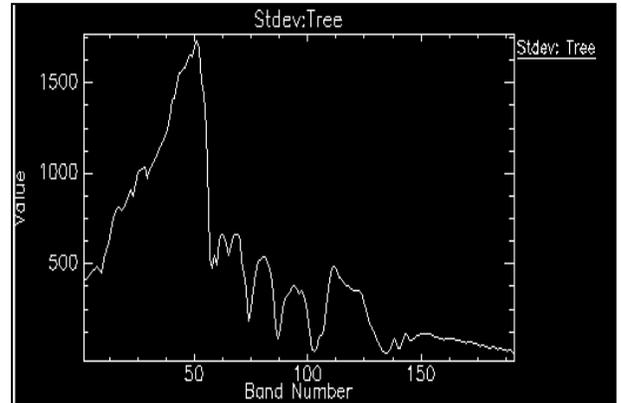
(a) Means for all Classes



(b) Standard Deviation for all Classes



(c) Minimum, Maximum and Mean for Tree class



(d) Standard Deviation for Tree Class

Fig. 10. Class statistics based on applying K-means algorithm

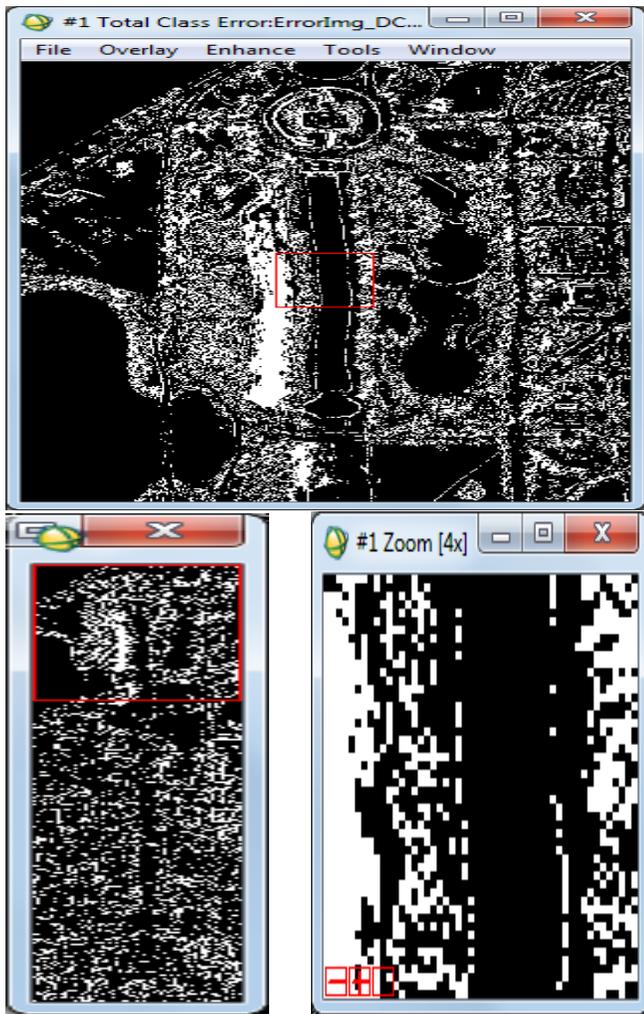
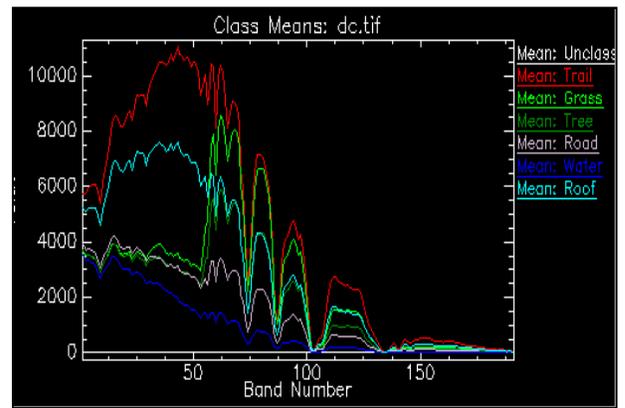


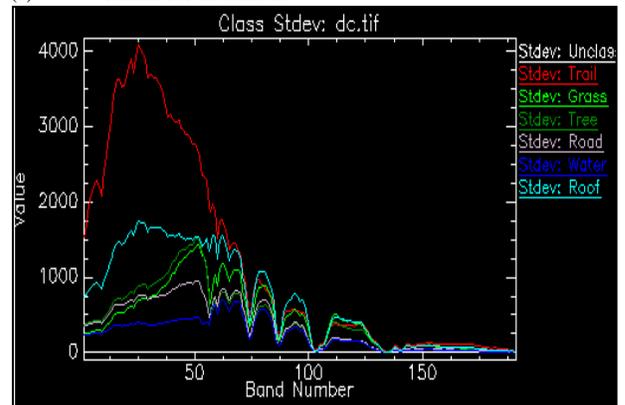
Fig. 11. Total Class Error (K-means)

#### IV. CONCLUSION

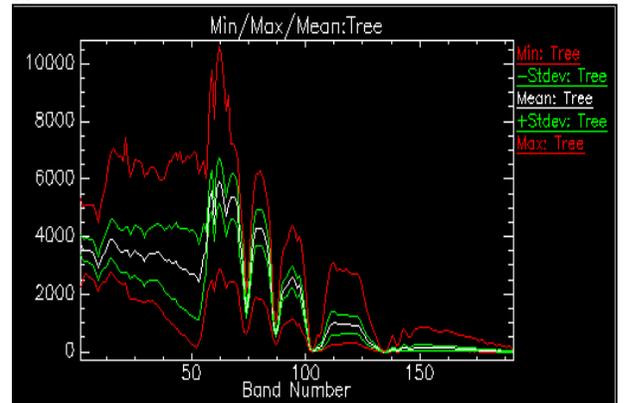
Hyperspectral images have broad spectral information to identify and distinguish materials spectrally unique. Classification of hyperspectral image means assigning objects with the same level of a class with homogeneous characteristics. In this work, unsupervised classification algorithms (K-Means algorithm and ISODATA algorithm) are used after applying Principle Component Analysis (PCA) using ENVI tool. PCA is used before the classification process as a technique in data analysis to reduce hyperspectral image dimensions. They are applied in a test site representative in the study area in Washington DC, USA. The overall accuracy was reported as 78.3398% for the K-Means classification approach, and 81.7696% for the ISODATA classification approach. It is found that K-Means algorithm and ISODATA algorithm give accurate results, but ISODATA algorithm has a better result on the study area image.



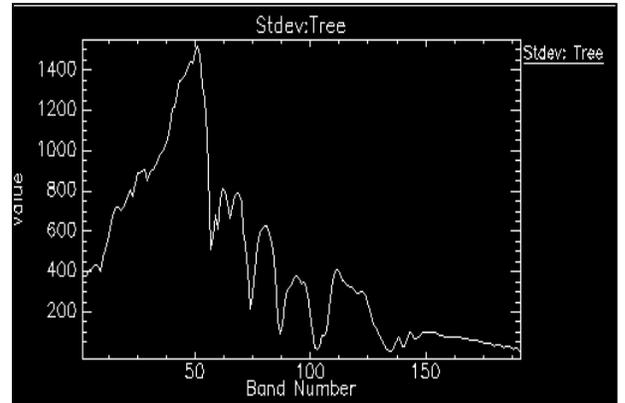
(a) Means for all Classes



(b) Standard Deviation for all Classes



(c) Minimum, Maximum and Mean for Tree class



(d) Standard Deviation for Tree Class

Fig. 12. Class statistics based on applying ISODATA algorithm

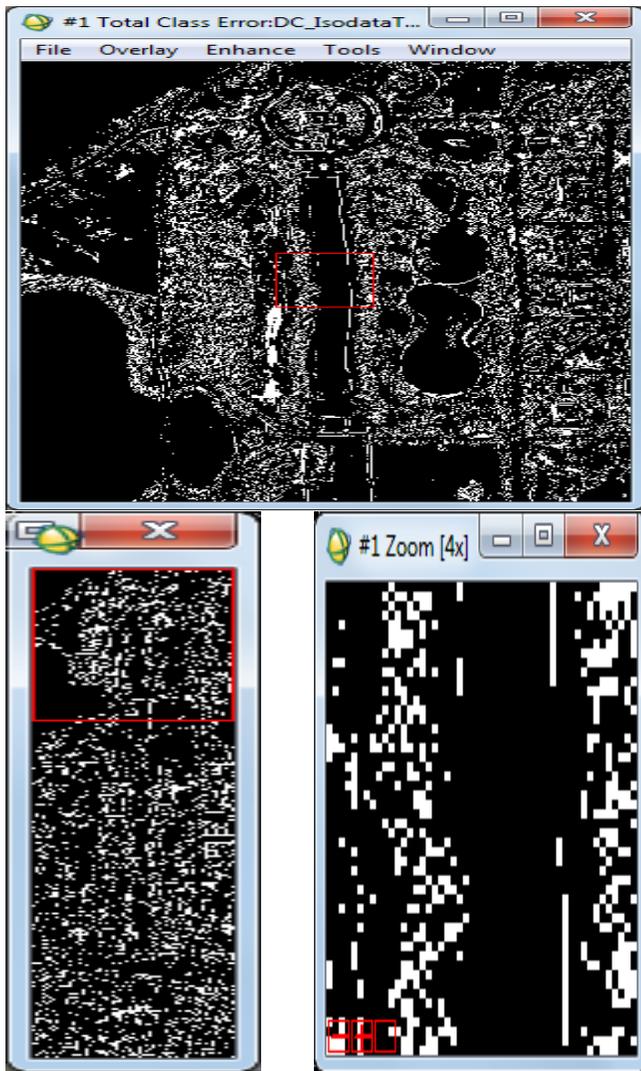


Fig. 13. Total Class Error (ISODATA)

## V. FUTURE WORK

Further studies should be conducted in studying the underground water, and soil in different regions in the Arab region for starting new wheat pivots. Also, the regions are not suitable for growing wheat in order to save money and effort if it gets infected.

## ACKNOWLEDGMENT

The author would like to thank all the participants involved in this work.

## REFERENCES

- [1] Sachin T. Jadhav and Mandar M. Kolap, "A Review on Methodologies of Remote sensing, GPS And GIS", IOSR Journal of Electronics & Communication Engineering (IOSR-JECE), ISSN : 2278-2834, ISBN : 2278-8735, PP. 1-6.
- [2] Foudan Salem, Tarek El-Ghazawi, and Menas Kafatos, "Remote Sensing And Image Analysis For Oil Spill Mitigation In The Red

- Sea", 2nd Biennial Coastal GeoTools Conference Charleston, SC, January 2001.
- [3] Principles Of Remote Sensing, ITC Educational Textbook series, ©ITC 2001. [Online]. Available: <http://www.gdmc.nl/oosterom/PoRSHyperlinked.pdf>
- [4] Foudan Salem and Menas Kafatos, "Hyperspectral Image Analysis For Oil Spill Mitigation", 22nd Asian Conference on Remote Sensing, Singapore, November 2001.
- [5] "Hyperspectral imaging", Fleurus (Municipality, Province of Hainaut, Belgium): CRW Flags Inc., 2010.
- [6] "HyperSpectral Image: How it happens". <http://www2.brgm.fr/mineo/RapportFinal/Image16.gif>
- [7] S. Peg, "Introduction to Hyperspectral Image Analysis", Earth Science Applications Specialist Research Systems, Inc.
- [8] P. Shipper, "Introduction to Hyperspectral Image Analysis," An International Electronic Journal, 2003. [Online]. Available: <http://spacejournal.ohio.edu/pdf/shippert.pdf>. [Accessed 10 october 2013].
- [9] SpecTIR, "Hyperspectral Image," SpecTIR, 2012. [Online]. Available: <http://www.spectir.com/technology/hyperspectral-imaging/>. [Accessed 10 Nov. 2013].
- [10] G. Shaw and H. Burke, "Spectral imaging for remote sensing," Lincoln Laboratory J, p. 3-28, 2003.
- [11] S. Dr Ropert, "Introduction To Remote Sensing," New Mexico State University.
- [12] J. B. Campbell, "Introduction to Remote Sensing" (3rd Edition), London: Taylor and Francis, 2002.
- [13] [Online]. Available: <http://disi.unitn.it/seminars/281>. [Accessed 22 September 2013].
- [14] Tan, "The K-Means algorithm," December 2006. [Online]. Available: <http://www.cs.uvm.edu/~xwu/kdd/Slides/Kmeans-ICDM06.pdf>. [Accessed 2013].
- [15] A. Villa, J. Chanussot, J.A. Benediktsson, C. Jutten c, R.Dambrevelle, "Unsupervised methods for the classification of hyperspectral images with low spatial resolution", Journal of Pattern Recognition, Volume 46 Issue 6, pp. 1556-1568, June 2013.
- [16] S.Praveena S.P.Singh I.V.Murali Krishna, "An Approach for the Segmentation of Satellite Images using K-means", KFCM, Moving KFCM and Naive Bayes Classifier. International Journal of Computer Applications, Volume 65, No.20, 2013.
- [17] M.Priyadharshini,R.Karthi,S.N.Sangeethaa, Premalatha, k.S.Tamilselvan, "Implementation of Fuzzy Logic for the High-Resolution Remote Sensing Images with Improved Accuracy", IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE) e-ISSN: 2278-1676,p-ISSN: 2320-3331, Volume 5, Issue 3, 2013, PP 13-17.
- [18] "K-Means Clustering," Home.deib.polimi.it, [Online]. Available: [http://home.deib.polimi.it/matteucc/Clustering/tutorial\\_html/kmeans.htm](http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/kmeans.htm)
- [19] "ERDAS Field Guide", 2005 Leica. Geosystems Geospatial Imaging, LLC, 2005.
- [20] Mather, Paul, "Computer Processing of Remotely-Sensed Images", 2004 John Wiley & Sons, Ltd
- [21] Memarsadeghi, N., Netanyahu, N.S., LeMoigne, J., "A Fast Implementation of the ISODATA Clustering Algorithm", International Journal of Computational Geometry and Applications, IJCGA, Vol. 17, No. 1, PP. 71-103, 2007.
- [22] "Multispectral Processing using ENVI's Hyperspectral Tools," 10 December 1997. [Online]. Available: <http://www.ltid.inpe.br/tutorial/tut14.htm>. [Accessed 6 November 2013].
- [23] ENVI, "ENVI User's Guide", Research Systems, Inc., 2004.