

Supervised Classification Approaches to Analyze Hyperspectral Dataset

Sahar A. El_Rahman

Electrical Department, Faculty of Engineering-Shoubra, Benha University, Cairo, Egypt
sahr_ar@yahoo.com

Wateen A. Aliady

Princess Noura bint AbdAlrahman University, Riyadh, KSA
Wateen.Aliady@gmail.com

Nada I. Alrashed

Princess Noura bint AbdAlrahman University, Riyadh, KSA
nada12.ksa@gmail.com

Abstract—In this paper, Spectral Angle Mapper (SAM) and Spectral Information Divergence (SID) classification approaches were used to classify hyperspectral image of Georgia, USA, using Environment of Visualizing Images (ENVI). It is a software application used to process and analyze geospatial imagery. Spatial, spectral subset and atmospheric correction have been performed for SAM and SID algorithms. Results showed that classification accuracy using the SAM approach was 72.67%, and SID classification accuracy was 73.12%. Whereas, the accuracy of SID approach is better than SAM approach. Consequently, the two approaches (SID and SAM) have proven to be accurately converged in classification of hyperspectral image of Georgia, USA.

Index Terms—Atmospheric Correction, Hyperspectral image, Spectral Angle Mapper, Spectral Information Divergence, Supervised classification.

I. INTRODUCTION

Hyperspectral imaging is spectral imaging technique that is able to find materials, identify and distinguish spectrally unique materials. This is done by collecting and processing hundreds of contiguous narrow wavebands from the scene, which provide spectral information [1]. Collecting the information is done by using an airborne or satellite sensor at a short, medium or long distance from the scene [2].

The main advantage of hyperspectral image is the potential to provide more accurate results than any other type of remotely sensed data, because they commonly collect more than 200 spectral bands to perform a detailed information extraction in order to classify, identify, and detect objects. [2][3][4].

In contrast to traditional multispectral sensors such as AVHRR (Advanced Very High Resolution Radiometer) that measures radiation reflected from a scene in three to six spectral bands of data [4][5]. This small range of

spectral bands is a primary disadvantage to multispectral sensors [5].

The main disadvantage of hyperspectral images is the need for sensitive detectors, fast computers, and significant data storage capacity for processing hyperspectral data. This led to an increase of the cost of acquiring and processing hyperspectral data [6]. Some of the practical applications for hyperspectral image classification are: Agricultural, Traffic recognition, Locate objects in satellite images, Medical. The majority usage of hyperspectral imaging is for vegetation and minerals extraction [8]. Classification is identified as the Information extraction technique that is mostly based on analyzing the spectral reflectance properties of the study scene and performing certain algorithms designed for spectral analysis [9]. It is known by the method that group pixels with similar characteristics together in an image and, indeed, the spectral pattern present within the data for each pixel is used as the numerical basis for classification. The objective of image classification is to identify the features occurring in an image in terms of the object or type of land cover these features actually represent on the ground as shown in Fig.1.

Image classification is an important part of the remote sensing, image analyzing and pattern recognition. It forms a significant tool for digital images examination. Image classification is perhaps the most important part of digital image analysis. It is really nice to have a colorful image, having a magnitude of colors illustrating various features of the underlying terrain, but it is quite useless, unless to know what the colors mean. The analyst must choose a classifier that will accomplish the best for a certain task. Now a days, it is difficult to state which classifier is optimum for all situations as the characteristic of each data set and the circumstances for each study vary so greatly [10]. There are two main approaches used in hyperspectral classification: Supervised and Unsupervised [9].

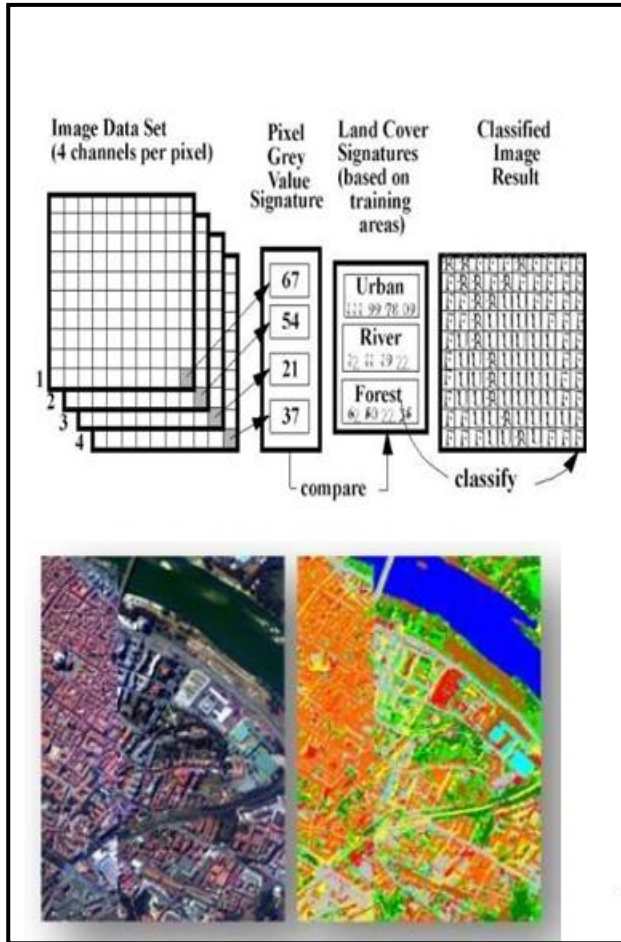


Fig. 1. Supervised classification

In supervised techniques, the training areas are used which are homogeneous representative samples of the different surface types of interest. All the spectral bands of the pixels comprising these areas have numerical information [11]. The algorithm assigns some pixels to information classes based on fieldwork, map analysis, and personal experience. Then, the algorithm classifies the rest of the pixels with unknown identities. The procedure starts by the user selecting and naming areas on the image, which correspond to the classes of interest. These classes correspond to information classes. Then, the algorithm will evaluate and assign unknown identity pixels to the class that has the highest likelihood of being a member.

Unlike the unsupervised classification, that depends on algorithms with statistically determined criteria to automatically organize pixels into unique groups with similar spectral characteristics [12].

There are a number of supervised approaches that have been developed to tackle the hyperspectral data classification problem. Each giving different classification accuracy. Two approaches are demonstrated in this work to compare their accuracy results, which are Spectral Angle Mapper (SAM), and Spatial Information Divergence (SID).

The paper is organized as follows. The study area and data set used in our work is given in Section II. The steps followed in our work in sequence are: applying Hyperion

tool, sub setting, atmospheric correction, classifying is given in Section III. The results obtained using this methodology are presented and discussed in Section VI. Section V, concludes and summarizes the observations obtained by using this approach.

II. STUDY AREA AND DATA SET

The study area is located in Georgia, US. Hyperspectral data was acquired on August 2009, using hyperspectral data from EO-1 Hyperion system. The test area covers about 1 km². This area has a lot of vegetation scene. It is downloaded from United States Geological Survey (USGS), which is a scientific agency of the United States government.

III. METHODOLOGY

First, Hyperion tool is applied on study dataset. Then, preprocessing of hyperspectral dataset include Spatial, spectral subset and atmospheric correction have been performed. Finally, SAM and SID supervised classification algorithms are applied.

A. Hyperion Tool

Hyperion tool is used to convert Geo TIFF datasets into ENVI format files that contain wavelength, and band information [12] [13]. The study dataset was in a form of 242 files with .TIF extension each representing a certain band so these bands have to be collected to form one image having all bands using this tool.

B. Preprocessing of Hyperspectral Data (Spatial and Spectral Subset)

It is often mandatory to perform the preprocessing on hyperspectral data to extract useful information from scene. This utilizes the processor by only processing needed data for the study area and improves the classification performance in hyperspectral imagery [15]. The data has been subjected to spatial and spectral subset to extract unwanted information.

▪ Spatial Subset

Performing image spatial sub setting is resizing the hyperspectral image to any size or aspect ratio by using ordered cutting that is focused on selecting the area of study in a square shape.

▪ Spectral Subset

It is based on identifying bad bands the ones that will not help in the study area. It will only cause over processing on the processor. ENVI headers may have associated information for the bad bands list. Mostly first and last bands are bad in hyperspectral images. The study area image contains 242 bands, but after the elimination of bad bands it only has 155.

C. Atmospheric correction

The atmosphere particles reduces the amount of incoming energy from the sun reaching Earth's surface and further reduce the amount of reflected energy reaching the sensor. Therefore, the energy reached to the sensor may be changed due to atmosphere interaction with incoming and reflected solar energy. So, little information would be gained from the scene.

Atmospheric correction attempts to minimize these effects on image spectra, so, it must be applied to correct the image of the effect of atmospheric gases, and through the use of ENVI it can correct the captured image of the effects of atmosphere [15].

QUAC (QUick Atmospheric Correction) is an approach for a sophisticated atmospheric correction, whereas, from the information contained within the scene the parameters directly are determined. The QUAC method is one of the best atmospheric correction methods, because it has a user-friendly interface, extremely accurate, and significantly fast [15].

D. Supervised Classification Approaches

In the processing phase, classification is applied on corrected image, using two classification approaches SAM and SID.

▪ SAM Classification Approach

In SAM Approach the spectral similarity between two spectra is computed. This is by calculating the angle between each pixel spectrum and each target spectrum. The smaller the angle is, the more likely to belong to the reference spectra. It treats the two spectrum as vectors, not taking into account their magnitude. This technique is relatively insensitive to changes in pixel illumination because increasing or decreasing illumination doesn't change the direction of the vector, only its magnitude. Endmember spectra is extracted directly from the study area image using the library USGS library by selecting Endmembers of interest. So, it will compare the spectral signature for each pixel in study dataset to the spectral signature of selected vegetation Endmember in the library [6],[7] see Fig. 2.

▪ Spectral Information Divergence

Spectral Information Divergence (SID) is a spectral classifier that uses a divergence measure to match pixels to reference spectra. The more likely the pixels are similar, the smaller the divergence. Pixels are not classified when they with a measurement greater than the specified maximum divergence threshold. SID measures spectral variability of a single mixed pixel from a probabilistic point of view [16], [17].

Endmember spectra are extracted directly from the study area image using the USGS library by selecting Endmembers of interest. The divergence of spectral signature is calculated for each pixel in study area to the spectral signature of a selected vegetation Endmember in the library see Fig. 3.

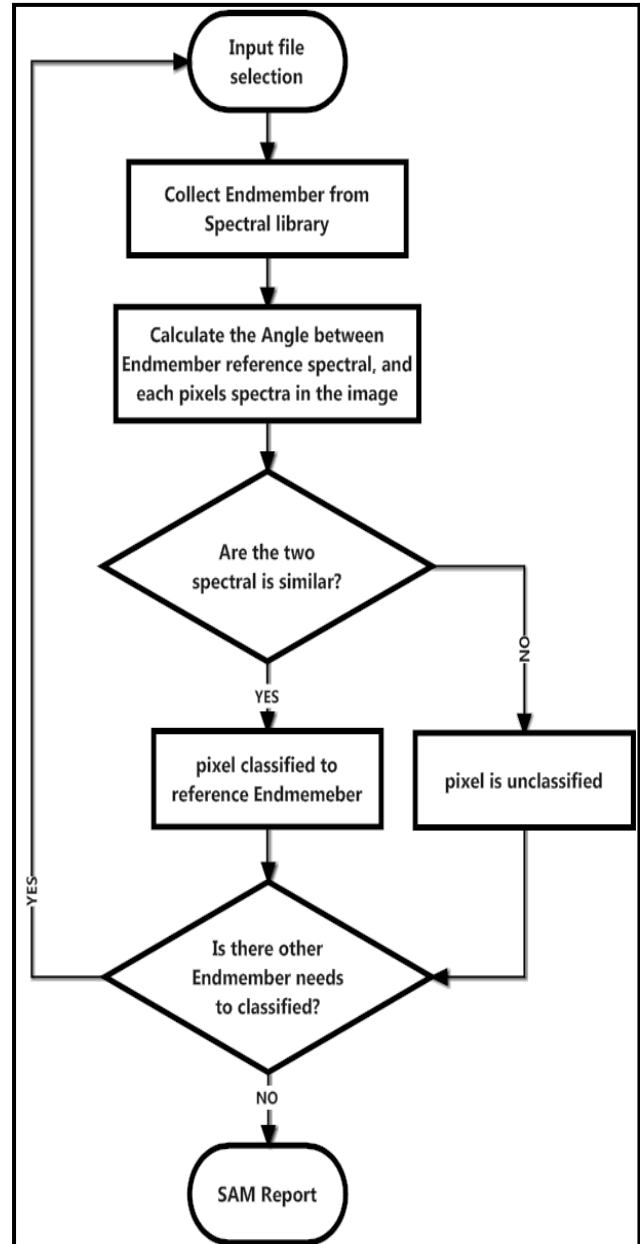


Fig. 2. SAM classification approach

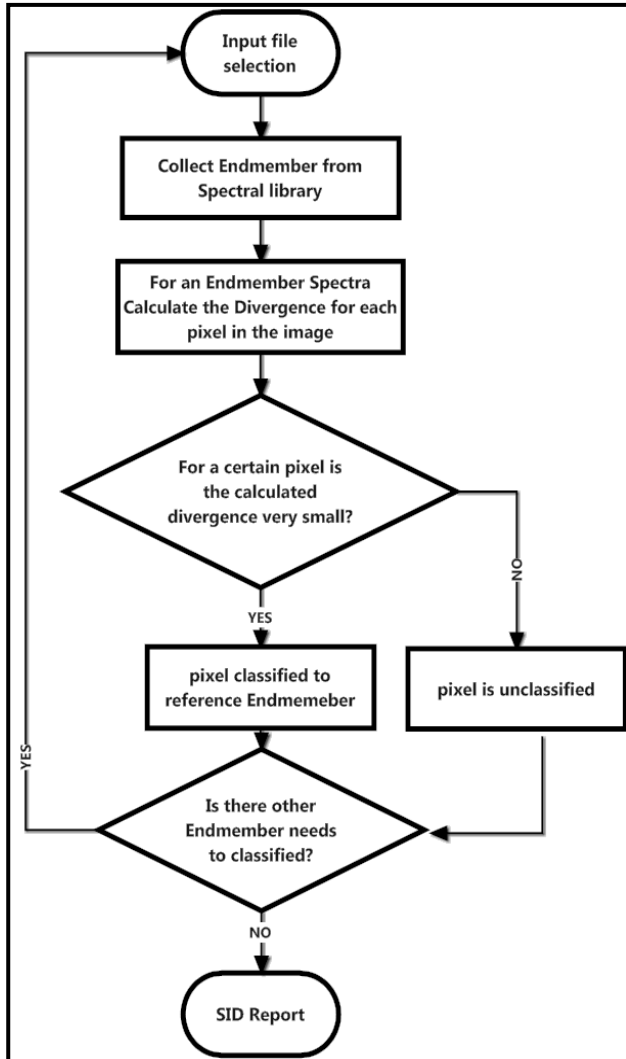


Fig. 3. SID classification approach

IV. RESULTS AND DISCUSSION

The classification result of study dataset using SAM approach is shown in Fig.4. Table 1 shows SAM classification of the entire study area image where Bay Laurel is the mostly found vegetation in this area of Georgia, whereas, 4.172% of the image is classified as Bay Laurel. Also, there were no identification for Chamise (Flower), Chamise (Green), and Coast Redwood (Green). The classification result for the study dataset after applying SID is represented in Fig.5. Table 2 shows SIM classification. I shows that Jasper Ridge Serpentine is the mostly found vegetation in this area of Georgia, whereas, 41.534% of the image is classified as Jasper Ridge Serpentine. Also, there were no identification for California Valley Oak, and Coast Redwood (Green). In order to differentiate statistically between the two classifications results, accuracy assessments has to be performed on SAM and SID to differentiate between their classifications results.

The correctness of classified images is determined by the accuracy assessments. The correlation between a standard that is assumed to be correct and an image classification of unknown quality is considered as the

measure of accuracy. So, at the beginning, the verification samples must be stated, which are used in ENVI as a standard for the accuracy assessments of the classifications performed. Then, generate random sampling for them. Finally, calculate the accuracy assessment by the confusion matrix.

The verification samples were chosen by using certain pixels that has spectral signature with a close match to the spectral signature of materials used in this work classification, which are found in the USGS spectral library, as shown in Fig.5, the spectral signature for Leather Oak in the USGS Spectral library, which is presented with the color green. This signature is used as a reference spectral to find pixels in study area image that has a spectral signature close to it, where the closest match was for pixels having the spectral signature colored in green, as shown in Fig.6.

Also, Red Willow sandstone verification samples were generated following the same method, where the spectral signature for Red Willow in the USGS Spectral library, is shown in Fig.5, and the closest match to it was for pixels having the spectral signature, as shown in Fig.6. Using region of interest (ROI) Tool, the verification samples on the original hyperspectral image are drawn manually.

After that, a random sample is generated, which is used to find pixels in the image that has a spectral signature with a close match to the spectral signature of ROI pixels, because it is helpful and can be valuable in supporting classification accuracy assessments. The stratified random sampling is used, also called proportional or quota random sampling. It involves dividing the population (all of the ROIs) into homogeneous subgroups (the individual ROIs) then taking a simple random sample in each subgroup. The used sampling technique proportionate, which means the sampling produces sample sizes that are directly related to the size of the classes (that is, the larger the class, the more samples will be drawn from it).

Finally, the Confusion Matrix is used to show the accuracy of a classification by comparing a classification result with ground truth information. A confusion matrix is calculated using ground truth ROIs previously determined. Table 3 shows the Confusion Matrix for SAM, and Table 4 shows the confusion matrix of SID.

The overall accuracy for SAM is 72.67%, and 73.12 for the SID. So, the SID has given a better classification for the study area image.

V. CONCLUSIONS

In this paper, the potential use of SAM, and SID classifiers combined with the EO-1 Hyperion imagery analysis for deriving total vegetation is achieved. They are applied in a test site representative in study area in Georgia, USA, as that is one of the famous vegetation areas. SID and SAM approaches use the same set of training and validation points selected over the acquired EO-1 Hyperion imagery, which allowed a direct comparison of their performance. The overall accuracy was reported as 72.67% for the SAM classification

approach, and 73.12% for the SID classification approach. SID has a better result on study area image, because SAM approach is insensitive since it depends on the spectrum direction and not the length of the spectra unlike the SID that measures the discrepancy between each pixel spectrum and a reference spectrum.

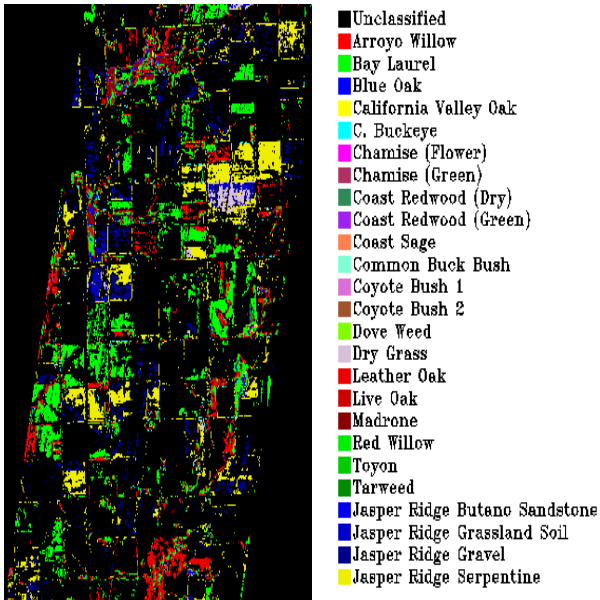


Fig. 4. Classified image using SAM classification approach.

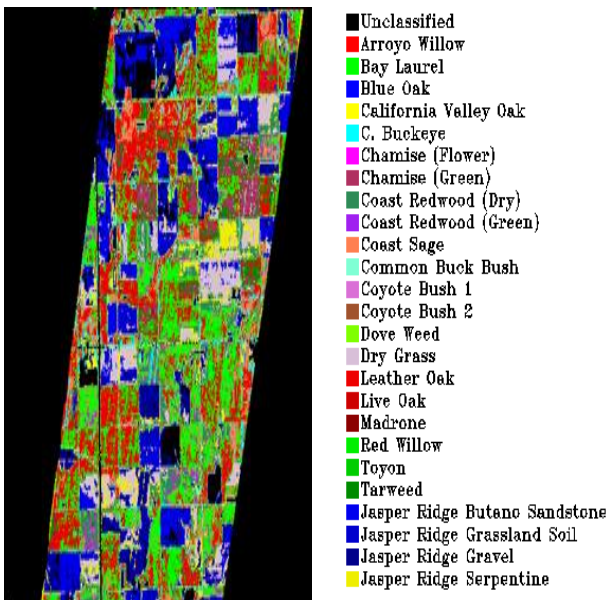


Fig. 5. Classified image using SID classification approach.

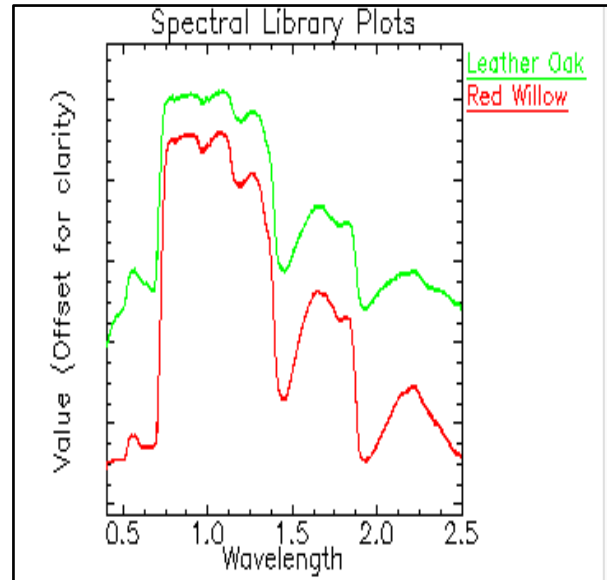


Fig. 6. Spectral signature for leather oak, and Red Willow in the USGS Spectral library

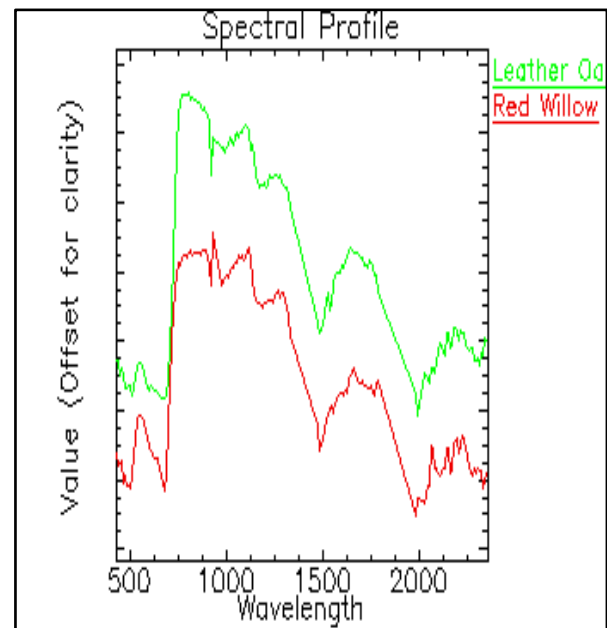


Fig. 7. Spectral signature for pixels in study area image that has a close spectral signature to leather oak, and Red Willow signature in USGS Spectral library.

ACKNOWLEDGMENT

We would like to thank our families for encouraging and supporting us. Also thank the people who have been instrumental in the successful completion of this work. The editing and comments of the reviewers is gratefully appreciated.

Table 1. Classification Of The Entire Study Area Image By SAM

Class	Points	Percent	Area
Unclassified	169,176	85.7%	152,061,436.4 m ²
Arroyo Willow	1,256	0.6%	1,128,937.69 m ²
Bay Laurel	8,230	4.172%	
Blue Oak	482	0.244%	433,238.8 m ²
California Valley Oak	75	0.038%	67,412.6 m ²
C. Buckeye	357	0.181%	320,884.3 m ²
Chamise (Flower)	0	0.00%	0.0000 m ²
Chamise (Green)	0	0.00%	0.0000 m ²
Coast Redwood (Dry)	3	0.002%	2,696.5072 m ²
Coast Redwood (Green)	0	0.000%	0.0000 m ²
Coast Sage	2	0.001%	1,797.67 m ²
Common Buck Bush	3	0.002%	2,696.5072 m ²
Coyote Bush 1	61	0.031%	54,828.98 m ²
Coyote Bush 2	184	0.09%	165,385.7 m ²
Dove Weed	681	0.345%	612,107.14 m ²
Dry Grass	840	0.426%	755,022.02 m ²
Leather Oak	4,305	2.182%	3,869,487.8 m ²
Live Oak	26	0.013%	23,369.72 m ²
Madrone	394	0.200%	354,141.2 m ²
Red Willow	610	0.309%	548,289.8 m ²
Toyon	5	0.003%	4,494.1 m ²
Tarweed	0	0.000%	0 m ²
Jasper Ridge Butano andstone	17	0.009%	15,280.2 m ²
Jasper Ridge Grassland Soil	1,353	0.686%	1,216,124.7 m ²
Jasper Ridge Gravel	2,827	1.433%	2,541,008.6 m ²
Jasper Ridge Serpentine	6,393	3.241%	5,746,256.9 m ²

Table 3. Confusion Matrix For Sam Classification Approach

Reference data		
Classified data	leather oak	Red Willow
leather oak	17.11	0.75
Red Willow	0.01	2.14
<i>Overall Kappa Statistic:</i> 0.1354		

Table 2. Classification Of The Entire Study Area Image By SID

Class	Points	Percent	Area
Unclassified	7,883	3.9%	7,085,522.2 m ²
Arroyo Willow	1,169	0.593%	1,050,738.98 m ²
Bay Laurel	20,418	10.350%	8,352,428.30 m ²
Blue Oak	198	0.100%	177,969.4781 m ²
California Valley Oak	0	0.000%	0.0000 m ²
C. Buckeye	4,315	2.187%	3,878,476.2 m ²
Chamise (Flower)	1,070	0.542%	961,754.2502 m ²
Chamise (Green)	2,559	1.297%	2,300,120.6 m ²
Coast Redwood (Dry)	9,024	4.574%	8,111,093.7 m ²
Coast Redwood (Green)	0	0.000%	0.0000 m ²
Coast Sage	9,229	4.678%	8,295,355.1 m ²
Common Buck Bush	363	0.184%	326,277.3 m ²
Coyote Bush 1	200	0.101%	179,767.1496 m ²
Coyote Bush 2	481	0.244%	432,339.9947 m ²
Dove Weed	1,011	0.512%	908,722.9411 m ²
Dry Grass	9,095	4.610%	8,174,911.1 m ²
Leather Oak	22,022	11.163%	19,794,160.8 m ²
Live Oak	9	0.005%	8,089.5217 m ²
Madrone	2	0.001%	1,797.6715 m ²
Red Willow	953	0.483%	856,590.46 m ²
Toyon	20	0.010%	17,976.71 m ²
Tarweed	2,414	1.224%	2,169,789.495m ²
Jasper Ridge Butano Sandstone	10,597	5.372%	9,524,962.4 m ²
Jasper Ridge Grassland Soil	4,021	2.038%	3,614,218.5 m ²
Jasper Ridge Gravel	8,288	4.201%	7,449,550.6 m ²
<i>Jasper Ridge Serpentine</i>	81,939	41.534%	73,649,702.3 m ²

Table 4. Confusion Matrix For Sid Classification Approach

Reference data		
Classified data	leather oak	Red Willow
leather oak	100	89.28
Red Willow	0.0	10.31
<i>Overall Kappa Statistic:</i> 0.5575		

REFERENCES

- [1] Lamyaa G. Taha , Atia A. Shahin, "Assessment of Cartographic potential of airborne hyperspectral data for large scale mapping", Recent Advances in Image, Audio and Signal Processing, wseas2013, PP 143-153. EGYPT, 2013. ISBN: 978-960-474-350-6. <http://www.wseas.us/e-library/conferences/2013/Budapest/IPASRE/IPASRE-19.pdf>.
- [2] U. Heiden, W. Heldens, S. Roessner, K. Segl, T. Esch and A. Mueller, "Landscape and Urban Planning in Urban structure type characterization using hyperspectral remote sensing and height information", Elsevier, pp 361-375, 2012.
- [3] Stefan A. Robila, Andrew Gershman, " Spectral Matching Accuracy in Processing Hyperspectral Data", IEEE, 0-7803-9029-6/05, 2005. PP 163-166.
- [4] Schurmer, "Air Force Research Laboratories Technology," J.H., U.K., 2003.
- [5] S. J. Purkis and V. V. Klemas, "Remote Sensing and Global Environmental Change," Remote Sensing and Global Environmental Change, vol. 3, no. 10, 2011.
- [6] Chein-I Chang, " An Information-Theoretic Approach to Spectral Variability, Similarity, and Discrimination for Hyperspectral Image Analysis", IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 46, NO. 5, AUGUST 2000, 1927-1932.
- [7] S. Rashmi, A. Swapna, Venkat and S. Ravikiran, "Spectral Angle Mapper Algorithm for Remote Sensing Image Classification," *Spectral Angle Mapper Algorithm for Remote Sensing Image Classification*, vol. 1, no. 4, pp. 201-205, 2014.
- [8] Belkacem Baassou, Mingyi He, Shaohui Mei, Yifan Zhang, "Unsupervised Hyperspectral Image Classification Algorithm By Integrating Spatial-Spectral Information", ICALIP 2012, 978-1-4673-0174-9/12, ©2012 IEEE, PP 610-615.
- [9] S. Pignatti, M. R. Cavalli, V. Cuomo, L. Fusilli, M. Poscolieri and San, "Remote Sensing of Environment," Evaluating Hyperion capability for land cover mapping in a fragmented ecosystem: Pollino National Park, Italy., vol. 3, no. 12, pp. 622-634, 2009.
- [10] Khamael Abbas, Mustafa Rydh, "Satellite Image Classification and Segmentation by Using JSEG Segmentation Algorithm", I.J. Image, Graphics and Signal Processing, Copyright © 2012 MECS, 10, pp 48-53. <http://www.mecs-press.org/>.
- [11] J. Senthilnath, Nitin Karnwal, D. Sai Teja, "Crop Type Classification Based on Clonal Selection Algorithm for High Resolution Satellite Image", Image, Graphics and Signal Processing, Copyright © 2014 MECS, 9, pp 11-19. <http://www.mecs-press.org/>.
- [12] Samuel Rosario Torres, "Implementation of the SVDSS in the ENVI/IDL Environment", vol. 15, no. 12, 2002.
- [13] D. White, "Hyperion Tools 2.0 Installation and User Guide, 2013.
- [14] Jinguo Yuan, Zheng Niu, "Classification Using EO-1 Hyperion Hyperspectral and ETM+ Data", Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007), IEEE, 0-7695-2874-0/07, 2007.
- [15] P. Shipper, "Introduction to Hyperspectral Image Analysis," An International Electronic Journal, 2003. <http://spacejournal.ohio.edu/pdf/shippert.pdf>.
- [16] Chein-I Chang, " Spectral Information Divergence for Hyperspectral Image Analysis", IEEE, 0-7803-5207-6/99, 1999, pp 509-511.
- [17] E. Zhang, X. Zhang, Y. Shuyuan and W. Shuang, "Improving Hyperspectral Image Classification Using Spectral Information Divergence," *Improving Hyperspectral Image Classification Using Spectral Information Divergence*, vol. 11, no. 1, pp. 249-253, 2013.

Authors' profiles

Sahar Abd El_Rahman was born Cairo, Egypt, B.Sc. Electronics & communication, Electrical Engineering Department. Benha University, Shoubra Faculty of Engineering, Cairo-Egypt. M.Sc. in an AI Technique Applied to Machine Aided Translation, Electronic Engineering, Electrical Engineering Department, Benha University, Shoubra Faculty of Engineering, Cairo-Egypt, May2003. PHD. in Reconstruction of High-Resolution Image from a Set of Low-Resolution Images, Electronic Engineering, Electrical Engineering Department, Benha University, Shoubra Faculty of Engineering, Cairo-Egypt in Jan2008.

She is assistant professor from 2008 till now at Electrical Engineering Department, Faculty of Engineering, Shoubra., Benha University, Cairo, Egypt. She was a lecture in the same location from 2003 and instructor in the same location in 1998.

Her research interests include computer vision, digital image processing, Signal processing, Robotics and Networks.

Wateen A. Aliady received B.Sc. degree in Computer Science from College of Computer Science and Information, Princess Noura Bint Abdul Rahman University in 2014, Saudi Arabia.

Nada I. Alrashed received B.Sc. degree in Computer Science from College of Computer Science and Information at Princess Noura bint AbdulRahman University in 2014, Saudi Arabia, and now I'm working as teacher in Computer Sciences department in Princess Noura bint AbdulRahman University.

How to cite this paper: Sahar A. El_Rahman, Wateen A. Aliady, Nada I. Alrashed, "Supervised Classification Approaches to Analyze Hyperspectral Dataset", IJIGSP, vol.7, no.5, pp.42-48, 2015. DOI: 10.5815/ijigsp.2015.05.05