

# UNSUPERVISED INFORMATION EXTRACTION USING ABSORPTION LINE IN HYPERION IMAGES

Y. Rezaei<sup>a,b</sup>, M.R. Mobasher<sup>c</sup>, M.J. Valadan Zoej<sup>c</sup>

<sup>a</sup> Remote Sensing Department, KNTOOSI University of Technology, Mirdamad Cross,  
Valiasr Av. , Tehran, Iran [y.rezaei@gmail.com](mailto:y.rezaei@gmail.com)

<sup>b</sup> Civil engineering college, Hamedan , Iran

<sup>c</sup> Remote Sensing Department, KNTOOSI University of Technology,([@kntu.ac.ir">mobasher, valadanzouj](mailto:mobasher, valadanzouj))@kntu.ac.ir

Commission WG VII/3

**KEY WORDS:** Hyperspectral Imaging, Absorption Bands, Clustering, Derivative Analysis

## ABSTRACT:

Using recently developed techniques, a significant amount of Information could be extracted from spectral signatures of surface material in their continuous spectra. One of the main applications in which these kinds of information are applicable is the vegetation and agricultural crop monitoring.

Absorption lines in the spectrum of different pixels provide useful information regarding the presence of different substances and species as well as the amount of each and distribution of each substance in the scene. Reflectance characteristics of any vegetation biomass, is strongly dependent on absorption lines of their compositions. In this research, a technique in automatic calculation of absorption line parameters is presented. These parameters then will be utilized in the procedure of vegetation cover classification using hyperspectral imagery in an unsupervised manner. Moreover using different vegetation indices and hyperspectral images the scene was classified and the results were thoroughly discussed and analyzed.

## 1. INTRODUCTION

Recent advances in spatial and spectral resolution of satellite sensors caused remote sensing technology to be increasingly used for accurate measurements of vegetated canopies and forests indices. With the deployment of early broadband sensors a great amount of data were produced that could not be supplied through traditional field methods otherwise. Through these methods, for the first time vegetation mapping could be implemented on a large scale and the data updated regularly. However, this technology has shown that although data obtained from broad band sensors have been useful in many respects, they also have their limitations. Because of the limited number of channels and wide bandwidths, lots of information regarding surface cover material that usually happens in narrow region of spectrum, is lost mostly due to averaging. This deficiency however could be compensated for using newly offered hyperspectral imaging technology.

The absorption line in the spectrum of the material may provide useful information regarding the type of different materials and substances under consideration. Reflectance characteristics of forest canopies and reflectance behaviour of the leaves, strongly depends upon absorption lines of their constituents materials. Organic components absorb mid infrared and thermal part of spectrum in fundamental stretching frequencies. Absorption features of the organic components in visible and near infrared are generally due to main frequency and other overtones of fundamental stretching frequencies of junctions such as C-N, N-H and O-H (Hunt 1980).

The important biochemical components present in vegetation are cellulose; Lignin, starch and compounds having nitrogen

such as protein and chlorophyll. The chlorophyll absorption is due to electron transfer in the visible region of spectrum. The absorption line have wavelengths of around 430 nm and 660 nm (for chlorophyll-a) and 460 nm, 640 nm (for chlorophyll-b). The most abundant component in protein is nitrogen where it mostly can be found in the green leaves.

The spectral reflectance of natural surfaces is sensitive to the special chemical junction of the molecules of substances in either states of solid, liquid and gas. Diversity of the elements in a substance compound causes the variety of the shape and location of the absorption bands of that substance in the electromagnetic spectrum. Then the vast diversity of substances in the real world can produce a complicated signal that sometimes is hard to interpret.

In this research, absorption line parameters are automatically calculated and then utilized in the procedure of vegetation cover classification using hyperspectral imagery in an unsupervised manner. Moreover, with the usage of different vegetation indices a hyperspectral image was classified and the results were thoroughly discussed and compared.

## 2. DATA AND METHODOLOGY

### Region of study

The image used for this work was a Hyperion image of May21, 2005 from a region in the south of Tehran capital of I. R. Iran (Fig. 1). Hyperion is a Hyperspectral sensor on board of EO-1. It has 224 bands in the 400 to 2500nm spectral region. Its spatial resolution is 30m and its swath width is 7.5km.



Figure. (1)- A Hyperion subimage colour composite (447.17, 548.92, 650.67 nm) of a region in the south of Tehran

#### Preprocessing and Atmospheric Correction

To correct this image for atmospheric effects the FLAASH program built in ENVI software along with the necessary synoptic parameters was used. This correction was applied to the converted DN to physical quantities such as radiance.

The Hyperion image in hand was in Level 1 on which only the radiometric corrections were done. Only 196 bands out 224 were calibrated. These were bands 8-57 in the visible and near infrared (VNIR) and 77-224 for shortwave IR (SWIR). One of the problems in Hyperion images is that the sensors VNIR and SWIR have one pixel misplacing after detector 128 in the CCD array. This misplacing was corrected at this stage resulted in disappearance of the striping effects. Then DN to radiance ( $L$ ) conversion was done using the following equations:

$$L_{\text{VNIR}} = \text{Digital Number} / 400 \quad (1)$$

$$L_{\text{SWIR}} = \text{Digital Number} / 800$$

In the next stage using FLAASH software and visibility value (from nearest weather station) the correction for the atmospheric effects were carried out. Then using MNF function (Minimum Noise Fraction) those bands with high noises were detected and omitted and so it remained only 157 bands.

After this, to correct for the remaining extra undesired points in the reflectance spectrum and smoothing, the EFFORT function was applied to the image (Mobasher, Rezaei et al. 2007). After correcting for the atmospheric effects and finishing preprocessing, the image was ready for algorithm designing.

#### Methodology in Algorithm Designation

There are different pixel-based image processing techniques in extraction of information regarding surface composition from hyperspectral images through their absorption lines. The absorption parameters can be calculated from the continuum removed spectrum. These are absorption band position, depth of the absorption lines and band asymmetry factors. These parameters are commonly applied in supervised methods. Because the supervised methods need training data and auxiliary information, they are not easily applicable.

In order to derive information on surface composition on a pixel by pixel basis for the entire image, several techniques have been proposed to process hyper spectral imagery among which some use absorption band depth, such as the relative absorption-band-depth (RBD) approach, the spectral feature fitting (SFF) technique and the Tricorder and TETRACORDER algorithms developed at the USGS spectral laboratory (Van der Meer 2004). These techniques work on continuum removed reflectance spectra (Fig. 2), thus acknowledging that the absorption in a spectrum has two components: a continuum and individual features. The continuum or background is the overall albedo of the reflectance curve. Removing this may effectively scales the spectra to 100% when the spectral curve approaches the continuum. Mathematically this can be done as follows (Van der Meer 2004)

$$L_c(\lambda) = \frac{L(\lambda)}{C_1(\lambda)} \quad \text{and} \quad O_c(\lambda) = \frac{O(\lambda)}{C_0(\lambda)} \quad (2)$$

Where  $L(w)$  is the library spectrum as a function of wavelength,  $O$  the observed spectrum,  $C_1$  the continuum for the library spectrum,  $C_0$  the continuum for the observed spectrum,  $L_c$  the continuum-removed library spectrum, and  $O_c$  is the continuum-removed observed spectrum.

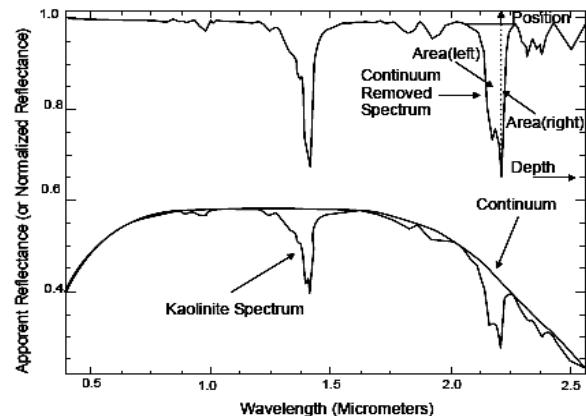


Figure. (2). Definition of the continuum and continuum removal and subsequent definition of absorption feature characteristics (Van der Meer 2004)

Since supervised extraction of preliminary information and distinction of different feathers in the hyperspectral images are highly difficult and time consuming, using unsupervised methods in image processing are a necessity. High spectral resolution of these images, make extraction of many features signal possible where detection of these features through preliminary information were difficult if not impossible.

Due to the fact that the spectrum of hyperspectral data is a nearly continuum, in order to locate the absorption bands one could take a derivative of the spectrum with respect to wavelength. The resulting derivatives are sensitive to shape rather than the magnitudes of spectra and do not show any sensitivity to the variation of illumination intensity due to the changes in sun angle, cloud cover, or topography. In general the spectral shape is more important than the spectral magnitude because the former has a direct relationship with the composition of the target were the latter represent the amount of substance present in the pixel. The derivative of a spectrum can model the variety of the main spectrum geometry precisely.

Derivatives were estimated using a finite approximation (Tsai and Philpot 1998). An advantage of the finite approximation is that the derivatives can be computed according to different finite band resolutions (band separations) to extract spectral features of interest at different spectral scales (Zhang, Rivard et al. 2004). The spectrum extracted from the hyperspectral data is a function of the wavelength ( $\lambda$ ) that can be represented in discrete form as follow (Zhang, Rivard et al. 2004).

$$S = [s(\lambda_1), s(\lambda_2), s(\lambda_3), \dots, s(\lambda_n)] \quad (3)$$

where  $S$  is the spectrum of the reflectance value at different wavelength (e.g., the reflectance value at the  $i$ th band). The first derivative can be estimated by

$$\frac{ds}{d\lambda} \Big|_i \approx \frac{S(\lambda_i) - S(\lambda_j)}{\Delta\lambda} \quad (4)$$

Where  $\frac{ds}{d\lambda} \Big|_i$  is the first derivative at wavelength  $\lambda_i$ ,  $\Delta\lambda$  is the separation between adjacent bands,  
 $\Delta\lambda = \lambda_j - \lambda_i$ ,  $\lambda_j > \lambda_i$

Similarly, the second derivative can be approximated as

$$\frac{d^2s}{d\lambda^2} \Big|_j \approx \frac{S(\lambda_i) - 2S(\lambda_j) + S(\lambda_k)}{(\Delta\lambda)^2} \quad (5)$$

Where  $\frac{d^2s}{d\lambda^2} \Big|_i$  is the second-derivative value at wavelength  $\lambda_i$ ,

$$\Delta\lambda = \lambda_j - \lambda_i = \lambda_k - \lambda_j, \lambda_k > \lambda_j > \lambda_i.$$

The choice of band separation  $\Delta\lambda$  is closely related to the magnitude and resolution of the derivative spectrum and can have an important impact not only on the wavelength location of inflection points but also on that of zero cross-overs (Zhang, Rivard et al. 2004). As the band separation  $\Delta\lambda$  increases, the magnitude of the spectral derivative will be depressed because the derivative is normalized by a power of  $\Delta\lambda$ . A large band separation can result in the loss of spectral detail and the attenuation of key spectral features.

Derivative process enhances the high frequency noise; therefore, in order to remove this effect, the low-pass filters were used. The mean filter smoothing algorithm was used in this study because it is straightforward and requires the least computation time. The equation of this operator can show below:

$$\hat{S}(\lambda_j) = \frac{\sum_{i=1}^n S(\lambda_i)}{n} \quad (6)$$

Where  $n$  (number of sampling points) is the size of the filter, and  $j$  is the index of the midpoint.  $\hat{S}(\lambda_j)$  is the new value of the midpoint in the window.

This fact should be taken into consideration that the larger the filter window is chosen, the smoother the results will be and consequently the loss of useful detailed spectral information is more likely to occur (Tsai and Philpot 1998). In this research, the window size of smoothing filter was chosen to be 3.

Having calculated the first and second derivatives of the spectrum, the absorption band parameters such as absorption band location, absorption band width and depth can be extracted.

For calculating these parameters, we used the IDL programming. In figure (3) a sample spectrum of a full vegetated pixel is illustrated.

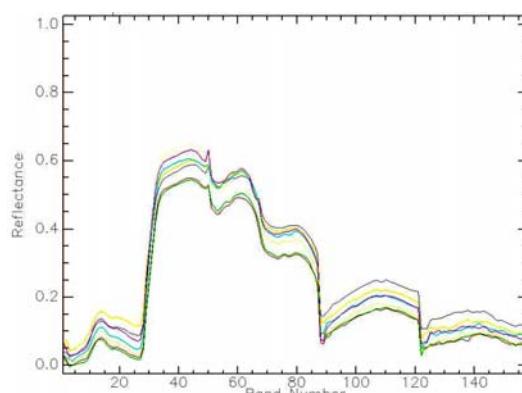


Figure. (3) : Spectrum of a vegetated pixel

In order to analyze the extracted derivatives, visible portion of the spectrum (0.4-0.7μm) was used. For this the spectrum of few pixels in some selected wavelength range is shown in fig.4 to 6.

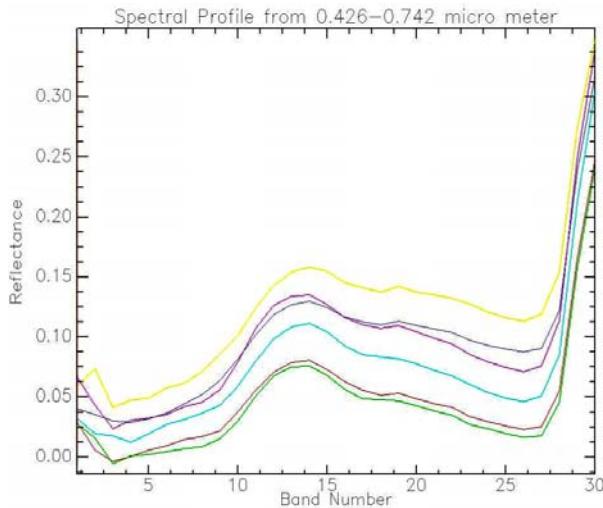


Figure. (4): Spectral profile of vegetated pixels in 0.426-0.742 micrometer range.

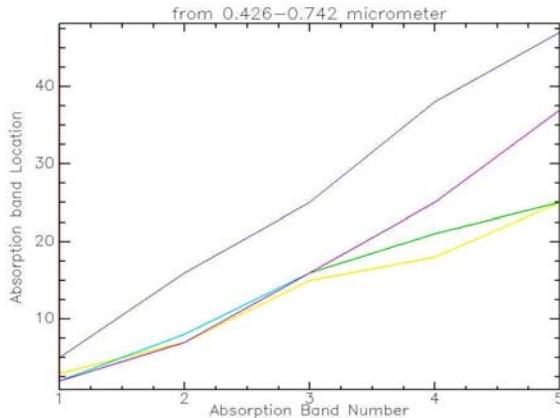


Figure. (5): Absorption bands location in 0.426-0.742 micrometer range.

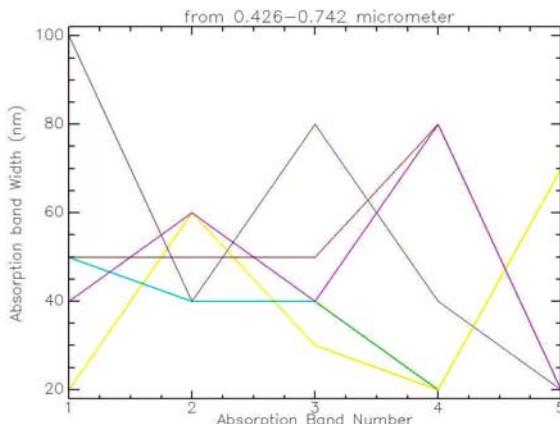


Figure.(6): Absorption bands width in 0.426-0.742 micrometer range.

The next step of the proposed method is calculating the absorption band parameters. For this, the spectrum of each pixels were extracted. Using these spectrums, the absorption depth and width in each absorption line were determined. Through derivative analysis the spectral angle distance with respect to reference spectrum was also calculated.

Having calculated the essential parameters, pixels were classified using c-means clustering. After which the red edge

index was applied on the image and again c-means clustering was performed.

In the next step of this research, the absorption bands parameters were calculated. For doing so, the spectrum of each pixels were extracted. Using these spectrums the absorption depth and width for every absorption lines were determined through derivative analysis. Also spectral angle distance with respect to reference spectrum was calculated for each pixel. In the next step using this calculated parameters, the pixels were grouped in different classes using c-means clustering. Then, red index is applied and compared with c-means clustering. The Red edge index can calculate by:

$$mNDVI_{705} = \frac{\rho_{750} - \rho_{705}}{\rho_{750} + \rho_{705} - 2\rho_{445}} \quad (7)$$

Where:  $\rho_{750}$  =reflectance at 750 nm

$\rho_{705}$  =reflectance at 705 nm

$\rho_{445}$  =reflectance at 445 nm

Also to compare the results, the image was classified using c-means clustering algorithm in the spectral bands of 711.72 to 752.250 nm.

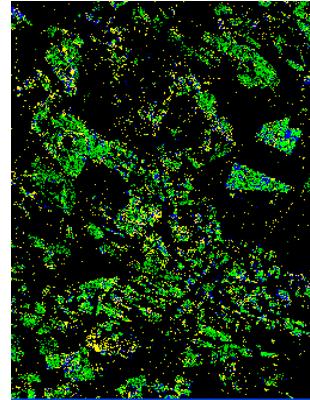


Figure. (7): The results of clustering using absorption bands

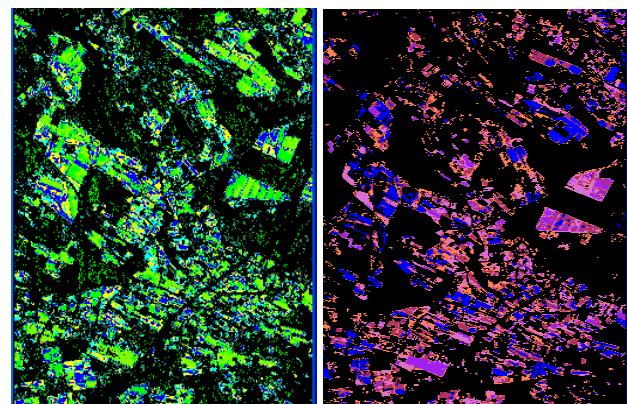


Figure. (7): The results of clustering using Red Edge Index (left), and the results of clustering using the spectral bands of 711.72 to 752.250 nm (right)

Comparing the clustered images (by Red edge and absorption band parameters), it is seen that the absorption band parameters can provide useful information about features in the clustered image.

### 3. DISCUSSION AND CONCLUSION

Since absorption band parameters provide useful information on different features, utilizing them in the procedure of feature identification and detection is of significant importance. The absorption lines of each feature in the corresponding pixel spectrum can be extracted using these parameters and after that classification can be carried out.

In order to assess the results, a comprehensive comparison between the output of the proposed method and those of red edge index and clustered image by using red edge bands, was done.

The final results indicated that, without prior information about vegetation of a specific area, classification could be performed well in an unsupervised manner. In order to improve the results, other information including spectral angle could be used along with absorption band parameters.

### REFERENCES

- "Preprocessing EO-1 Hyperion Hyperspectral Data to Support the Application of Agricultural Indexes." *GeoRS* 41 ( 6): 1246-1259. (2003)
- Hunt, G. R. (1980). Electromagnetic radiation: The communication link in remote sensing, In *Remote Sensing in Geology*. N.Y., John Wiley & Sons,.
- Mobasher, M. R., Y. Rezaei, and Valadan Zanjani, M. J.. (2007). "A Method in extraction vegetation quality parameters using Hyperion images,with application to precision farming." *World Applied Science Journal* 2(5): 476-483.
- Tsai, F. and W. D. Philpot (1998). "Derivative analysis of hyperspectral data." *Remote Sens. Environ.* 66: 41–51.
- Van der Meer, F. (2004). "Analysis of spectral absorption features in hyperspectral imagery." *International Journal of Applied Earth Observation and Geoinformation* 5: 55–68.
- Zhang, J., B. Rivard, et al. (2004). "Derivative Spectral Unmixing of Hyperspectral Data Applied to Mixtures of Lichen and Rock." *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING* 42(9).

