A Novel Derivative-Based Classification Method for Hyperspectral Data Processing

Yucel CIMTAY, Hakki Gokhan ILK

Electrical and Electronics Engineering, Faculty of Engineering, Ankara University, Emniyet Mahallesi, Dogol Cd. 6A, 06560 Yenimahalle/Ankara, Turkey

yucel.cimtay@gmail.com, ilk@ieee.org

Abstract. *In hyperspectral classification, a derivative of reflectance spectra is used directly or by fusion with the reflectance spectra. In this way, classification performance is improved. However, on the land cover, especially for plant species, the reflectance spectra may exhibit differences depending on a plant age and maturity level. This situation makes traditional classification methods which are based on time-dependent spectral similarity. In addition, the problem of classification of the species which have similar spectral properties is still valid. As a solution to time dependency and spectral similarity problems, in this study, a new and more generic method based on the spectral derivative is proposed. The method is tested for hyperspectral images which are captured at different time of the year and different places, in the life cycle of species. Test results show that proposed method successfully classifies the land cover time-independent and it is superior to the classical classification methods.*

Keywords

Classification, land cover, life cycle, spectral derivative, time dependency.

1. Introduction

Hyperspectral data is obtained from space and/or satellite platforms. In hyperspectral land cover classification, traditionally, reflectance data that is driven by hyperspectral data is used. However, the reflectance spectra -especially for plants- changes rapidly during their life-cycle. In addition, for the same species, the reflectance is varying from place to place. This situation makes achievement of hyperspectral classification dependent to the time and/or place at which hyperspectral data is captured. Another disadvantage of using the reflectance spectra is that it is very sensitive to noise. In this case, it is possible to see false alarms in the supervised classification results. As an alternative to reflectance data, the derivative of reflectance data is widely used in the literature directly or with reflectance data. Because of keeping the change points and amount of change, spectral derivative has the ability to catch the distinctive feature of different signals on the spectral plane. Another advantage of the spectral derivative is that it is not so much sensitive to noise. It is used in many classification studies directly [1] and [2] or with the reflectance spectra [3]. By this way, it can improve the classification results. Nevertheless, the variation of the reflectance spectra naturally changes the spectral derivative of these spectra. So, it can be commented that classification results are still time and region dependent.

Another important problem for hyperspectral classification is determining a specific threshold value for methods based on the spectral similarity. Especially, it is nearly impossible to find a threshold value that can be used for accurate classification of the species which exhibits similar spectral features. For instance, when spectra are searched in a hyperspectral image, with any spectral similarity based method, whether this species exists or not, if similar spectral species exist in the data, the method automatically matches the other similar species though an adaptive threshold value is used. In this case, for an accurate classification, one should know about the ground truth at the classification time which is not possible in every case. For a solution to spectral similarity and time-dependency problems in hyperspectral classification, a spectral derivative-based approach is proposed in this study.

2. Proposed Method

2.1. Pre-Processing of Hyperspectral Data

In this study, hyperspectral images which cover the life cycle of different species are used. The images are captured at a different time of the year. A lot of wheat, corn and cotton spectra samples are collected from these images. Because of different capture times, environmental effects and also non-ideal atmospheric corrections, a normalization process is applied to the samples. By this way, for the same plant, the spectral signatures which are captured at different time are made close to each other. Figure 1 and Fig. 2 show the original and normalized spectral signatures of the corn plant, respectively.
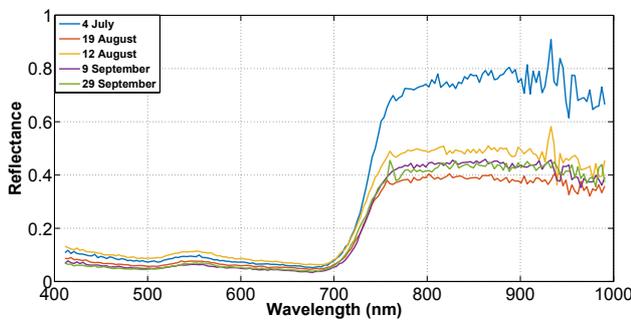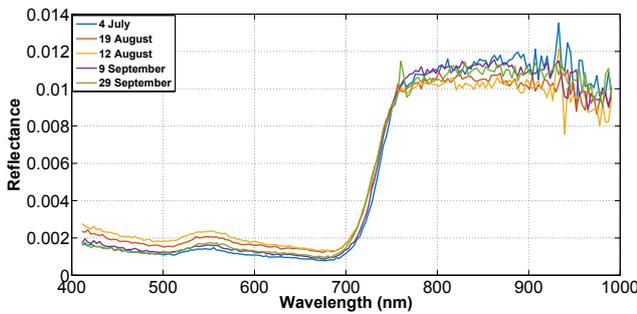

Fig. 1: Original corn spectra samples.

Fig. 2: Normalized corn spectra samples.

As s represents the spectra, $s(\lambda_i)$ the reflectance value at λ_i wavelength (value at i^{th} band) and N the number of bands, a spectrum that is collected from a hyperspectral data can be represented as a function of wavelength as in Eq. (1):

$$s = [s(\lambda_1), s(\lambda_2), s(\lambda_3), \ldots, s(\lambda_N)]. \qquad (1)$$

The normalization process applied this hyperspectral data can be shown as in Eq. (2):

$$s(\lambda_i) = \frac{s(\lambda_i)}{\sum\limits_{k=1}^{N} s(\lambda_k)}. \qquad (2)$$

2.2. Spectral Derivative Analysis

The derivative is the measure of change of a function versus the changes in independent variable [4]. In general, finite approach technique is used for calculating the derivative in hyperspectral studies. An advantage of this technique is that one can calculate the derivative for different band resolutions in order to extract the spectral properties of a signal. Forward, backward and central finite approaches are used for calculating the derivative [5]. In this study, central finite approach technique is preferred due to its low estimation error [6]. First degree derivative in central finite approach can be calculated as in Eq. (3). Here, $\lambda_{i+1} - \lambda_{i-1} > 0$.

$$\left.\frac{ds}{d\lambda}\right|_i \approx \frac{s(\lambda_{i+1}) - s(\lambda_{i-1})}{\lambda_{i+1} - \lambda_{i-1}}. \qquad (3)$$

The derivative operator is sensitive to noise in hyperspectral data processing, as it is in general. Therefore, this random noise should be smoothed or minimized before calculating derivative. For smoothing, smoothing filters such as a mean filter, moving average filter or Savitzky-Golay filter are used in the literature [5] and [7]. In this study, for giving more weight to central value, a 1×3 size Gaussian filter is used. The new value of reflectance value to which filter is applied is calculated as in Eq. (4):

$$H(\lambda_i) = [0.25s(\lambda_{i-1}) + 0.5s(\lambda_i) + 0.25s(\lambda_{i+1})]. \quad (4)$$

The filtered corn spectra are shown in Fig. 3. When compared to Fig. 2, it is obvious that the noise in the signal is attenuated, especially between 400–750 nm.
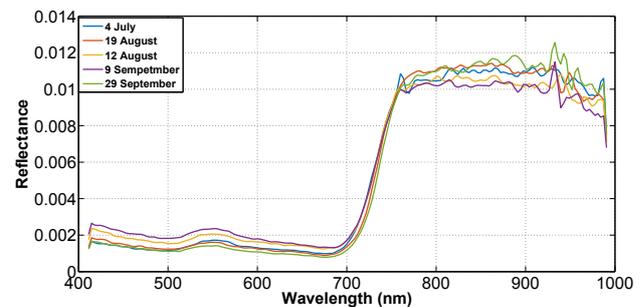

Fig. 3: Smoothed corn spectra.

After filtering operation, the derivative of spectra is calculated. The graph of the derivative is shown in Fig. 4. From Fig. 2 and Fig. 3, it is obvious that spectra behave like a noise signal at the wavelengths above 750 nm. For this reason, in this study, the part (750–1000 nm) of the spectra is not used.

NSD (Normalization, smoothing, derivation) operation is applied for wheat, corn and cotton classes. It is seen that the derivative values are very close to zero and also to each other between the wavelengths 400 and 695 nm.
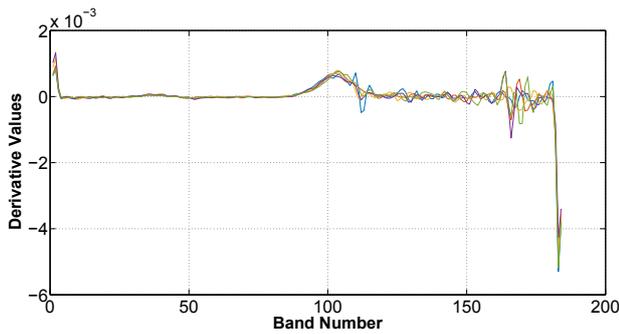
Fig. 4: Derivative of corn specta.

It is foreseen that this wavelength interval will not contribute to classification and increase the processing load. This wavelength part is also not included in this study. So, only 695–750 nm wavelength (totally 18 bands for the specific sensor used) interval which covers red-edge band and some amount of IR band is used. This decision coincides with the result that "red-edge band interval is an important argument for classification of plants" which is mentioned in [8]. For this wavelength part, NSD results are shown in Fig. 5. It is seen that wheat and corn spectra exhibit high spectral similarity for NSD results.
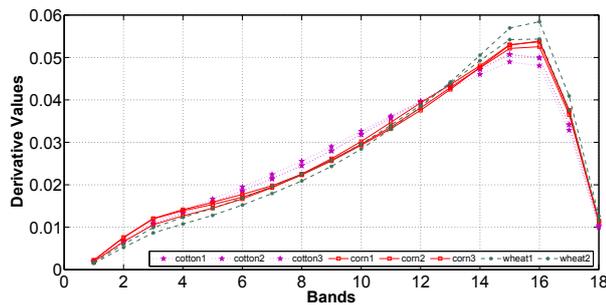

Fig. 5: Derivative values for 695–750 nm wavelength interval.

2.3. Proposed Algorithm

The proposed algorithm processes as follows:

- Many pure-pixel corn, wheat and cotton spectra are collected from hyperspectral images.

- NS (Normalization and smoothing) is applied to spectra.

- The 18 bands between wavelengths 695 and 750 nm are selected and the derivation is applied to these 18 bands.

- Derivative values are sorted in ascending ranking.

- The index number (indices from 1 to 18) also changes its position with its corresponding derivative value as a result of this ranking.

For instance, for 5 bands with derivative values $(0.11, 0.12, 0.16, 0.13, 0.09)$, index vector is naturally $(1, 2, 3, 4, 5)$. When the proposed algorithm is applied to these derivative vector, it becomes $(0.09, 0.11, 0.12, 0.13, 0.16)$ and index vector becomes $(5, 1, 2, 4, 3)$. When this ranking is investigated, it is discovered that each plant class has some characteristic features (pattern), although they exhibit spectral and derivative similarity. For the corn, wheat and cotton, these patterns are given in Tab. 1.

Tab. 1: Class patterns.

Classes	Pattern in the index vector
Cotton	The numbers 6 or 8 come just after or before 15.
Corn	The (10, 11, 18). Elements are {9 1 2}, {9 10 2} or {9 1 15} respectively.
Wheat	The (8,11,12,14). Elements are {7 9 10 11} respectively.

When classification is implementing using the proposed method based on a hyperspectral image, all stages are applied to each pixel of image data. Then, the corresponding pixel is assigned to "wheat", "corn", "cotton" or "other" classes according to the pattern found.

3. Hyperspectral Test Data

Hyperspectral images used in this study are captured with a push broom camera which has a spectral range 411–992 nm with spectral resolution 3.2 nm (totally 182 bands). The images have 60×60 cm^2 spatial ground resolution. One of the images covers almost 10 km^2 region in *Harran* plane, *Sanliurfa*. There are 5–10 plant species but mostly corn, wheat and cotton species are on the land cover, throughout a year. The other hyperspectral image is captured from *Gerede*, *Bolu*.

Pre-processing steps (radiometric, geometric and atmospheric corrections.) are applied to the raw images. By this way, reflectance images are handled. The proposed method is compared with SAM (Spectral angle mapper) method which is used traditionally and runs with spectral similarity principle. Totally 18 bands are used both proposed and SAM methods. SAM method is applied after NS operation and band selection. The proposed method is applied to the image data used in SAM (18 bands) after taking its derivative.

4. Experimental Results

4.1. Time Dependency Problem of Hyperspectral Classification

The origin of the problem is that plant spectra change rapidly even in short period. Therefore, traditional methods cannot make an accurate detection when a plant spectrum which is extracted from an image is searched in another image even if this new image is captured after short time. This is the problem of time dependency in hyperspectral classification. In the first conducted test in this study, a cotton spectrum is extracted from a pure cotton pixel of hyperspectral image that is captured on 8th August. This spectrum is searched in 19th August image with SAM method for threshold values $5°$ and $3°$. Results are shown in Fig. 6. Figure 6(a) shows the RGB bands of hyperspectral image, Fig. 6(b) the ground truth for this image and Fig. 6(c) the SAM result for threshold value $5°$. As it is seen, almost all cotton areas are detected. However, some extra corn regions are detected as cotton (regions marked with the white circle). This means we have too many false alarms. Recall value is high; however, precision is very low. Figure 6(d) shows the SAM result for threshold value $3°$. This time, some cotton areas cannot be detected (regions marked with the red circle). This means we have low recall result. Figure 6(e) is the result of the proposed method. As it is seen, we can detect all cotton areas and there is almost no false alarm. It is also possible to fill in the detected areas with basic morphological image processing techniques without creating any false alarm.
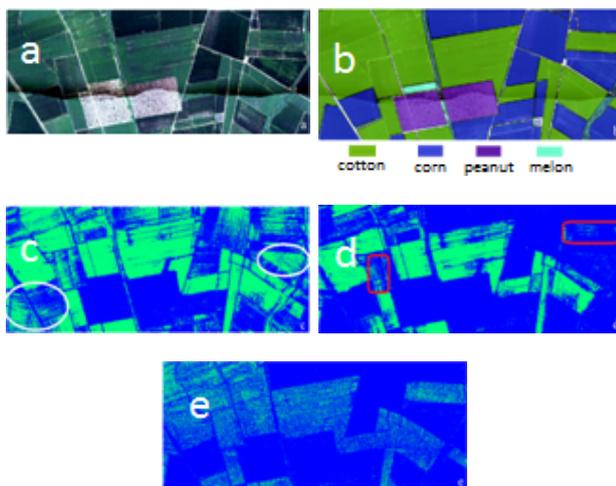

Fig. 6: '19 August' image, classification result.: (a) RGB bands of image, (b) Ground truth, (c) SAM result (Threshold value: $5°$), (d) SAM result (Threshold value: $3°$), (e) result of the proposed method.

4.2. Spectral Similarity Problem of Hyperspectral Classification

The origin of the problem is that different plant species may have very high spectral similarity. Therefore, spectral similarity-based traditional methods may easily detect the false plant. Tuning threshold value also cannot solve the problem. This is the problem of the spectral similarity in hyperspectral classification. For justification of this problem, two different cases are prepared. In the first case, a cotton spectrum which is taken from 25th June image and corn spectra which is taken from 12th August image is searched in 2nd April image data with SAM method for threshold value $1.5°$. Results are shown in Fig. 7. Figure 7(a) shows the RGB bands of hyperspectral image and Fig. 7(b) the ground truth for wheat plant. Figure 7(c) and Fig. 7(d) show the SAM results for cotton and corn spectra, respectively. As it is seen from the figures, SAM algorithm detects wheat areas as cotton and/or corn. This proves that spectral similarity is a problem in hyperspectral classification. Figure 7(e) is the SAM result of spectra and image derivatives. False alarms still exist but less than in Fig. 7(c) and Fig. 7(d). Figure 7(f), Fig. 7(g) and Fig. 7(h) show the detection results of wheat, cotton and corn, respectively, according to the proposed method. It finds the wheat regions successfully and also finds no false alarm for cotton and corn.
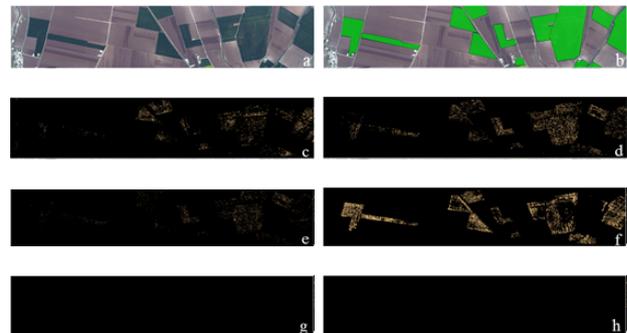

Fig. 7: '2 April' image classification results: (a) RGB bands of image, (b) wheat ground truth data, (c) SAM result with cotton spectra, (d) SAM result with corn spectra, (e) SAM result with derived corn and image, (f) wheat result of proposed method, (g) cotton result of proposed method, (h) corn result of proposed method.

In the second case, a wheat spectrum which is taken from 2nd April image (from *Harran*) is searched in 4th July image which is captured from *Gerede, Bolu*. So, we test spectral similarity for the case of hyperspectral images which are captured from different regions. Results are shown in Fig. 8. Figure 8(a) shows the RGB bands of the image, Fig. 8(b) the ground truth for corn and Fig. 8(c) the SAM result for wheat spectra with threshold value $1°$. Corn region is detected as wheat which means false alarm. Figure 8(d) shows the

SAM result for the derivative case. False alarms still exist but less. Figure 8(e), Fig. 8(f) and Fig. 8(g) show the detection results of corn, cotton and wheat, respectively, according to the proposed method. It finds the corn regions successfully and also finds no false alarm for cotton and very little false alarm for wheat which one can easily be aware of that there is no planting area at these false alarm points.
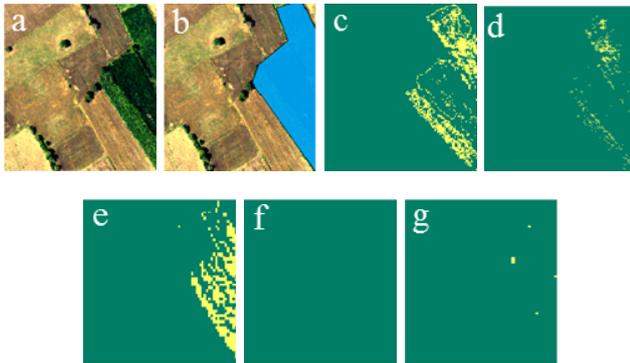

Fig. 8: '4 July' Image classification result. (a) RGB bands of image, (b) corn ground truth data, (c) SAM result with wheat spectra, (d) SAM result with derived wheat spectra and derived image, (e) corn result of proposed method, (f) cotton result of proposed method, (g) wheat result of proposed method.

4.3. Comparison with Other Classification Methods

In the third conducted test, the first conducted test for SAM method is repeated for also some other classification methods by using "ENVI Target Detection Wizard" [9]. These methods are MF (Match Filter), ML (Maximum Likelihood) and CEM (Constrained Energy Minimization). Results are shown in Fig. 9. Figure 9(a) shows the RGB bands of hyperspectral image, Fig. 9(b) the ground truth for this image and Fig. 9(c) the SAM result. Figure 9(d) shows the MF result, Fig. 9(e) the ML result, Fig. 9(f) the CEM result and Fig. 9(g) the result of the proposed method. Due to the usage of pure pixels in this study, some automatic image processing techniques are used in order to fill in the holes (the pixels which are vegetation but not pure pixel) in the results of each method. ML and CEM methods produce the worst results. SAM and MF methods partially produce accurate results but the best result in terms of covering almost all cotton areas (high accuracy of recall) and producing very little false positives (high accuracy of precision) are generated by the proposed method.

Table 2 shows the accuracy of the methods in terms of recall and precision values. In the classification of land cover, both recall and precision are important.
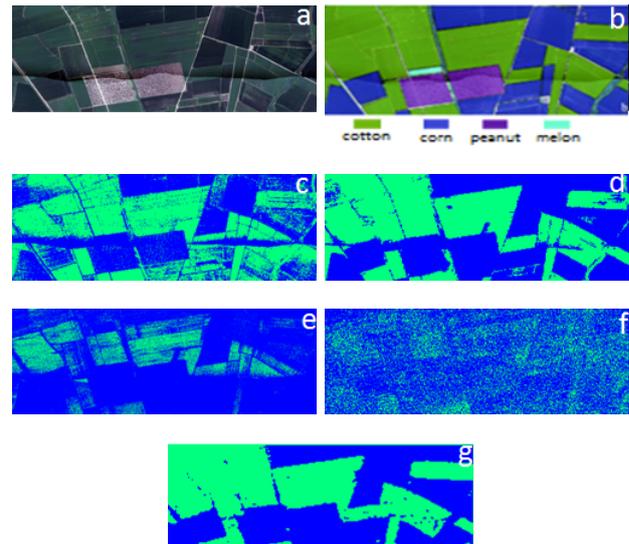

Fig. 9: '19 August' image, classification results. (a) RGB bands of image, (b) ground truth, (c) SAM result, (d) MF result, (e) ML result, (f) CEM result, (g) result of proposed method.

Precision and recall are calculated according to Eq. (5):

$$\begin{aligned} Precision &= \frac{tp}{tp + fp}, \\ Recall &= \frac{tp}{tp + fn}. \end{aligned} \tag{5}$$

The value of tp (true positive) increases when method assigns a pixel as target and the pixel is actually a target point in ground truth data. The value of fp (false positive) increases when method assigns a pixel as a target and the pixel is actually not a target point in ground truth data. The value of fn (false negative) increases when the method does not assign a pixel as a target but the pixel is actually a target point in ground truth data.

As it is seen from the table, the maximum recall and precision score are achieved by the proposed method. Both SAM and MF exhibit good recall scores however they are not successful enough in terms of precision. That means they detect some false positives. ML is successful for precision; however, it cannot detect all cotton areas.

Tab. 2: Accuracy of different classification methods.

Method	Recall	Precision
SAM	92.27	86.73
MF	94.77	95.47
ML	82.16	96.65
CEM	64.13	51.38
Proposed	**98.05**	**98.81**

5. Discussion and Results

In this study, a new derivative-based method is proposed for hyperspectral classification. This method determines patterns which best define each class of plants and distinguish this class from others. These patterns are used for classification of hyperspectral data. By this way, time dependency and spectral similarity problems of hyperspectral classification are solved. In the image capturing region, corn, cotton and wheat plants are seen on the land cover, in most of the year. Therefore, these plants are studied in this study. Results indicate that the proposed method is superior to traditional methods in terms of target detection and reducing false alarms. Only 18 bands are used in the method. Therefore, it is available to be run as near real time. In the future, different classes will be studied and in addition to VNIR $(400 - 1000$ nm) images used in this study, SWIR $(1000 - 2500$ nm) images will be also used in order to increase the separability of different species.

Acknowledgment

References

[1] BAO, J., M. CHI and J. A. BENEDIKTSSON. Spectral Derivative Features for Classification of Hyperspectral Remote Sensing Images: Experimental Evaluation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2013, vol. 6, iss. 2, pp. 594–601. ISSN 2151-1535. DOI: 10.1109/JSTARS.2013.2237758.

[2] DEHNAVI, S., Y. MAGHSOUDI, M. J. V. ZOEJ and F. BANIADAM. Beneficiary of High Order Derivative Spectrum In Target Detection. In: *IEEE Geoscience and Remote Sensing Symposium*. Quebec City: IEEE, 2014, pp. 4608–4611. ISBN 978-1-4799-5775-0. DOI: 10.1109/IGARSS.2014.6947519.

[3] KALLURI, H. R., S. PRASAD and L. M. BRUCE. Fusion of spectral reflectance and derivative information for robust hyperspectral land cover classification. In: *First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*. Grenoble: IEEE, 2009, pp. 1–4. ISBN 978-1-4244-4686-5. DOI: 10.1109/WHISPERS.2009.5288982.

[4] KALLURI, H. R., S. PRASAD and L. M. BRUCE. Decision-Level Fusion of Spectral Reflectance and Derivative Information for Robust Hyperspectral Land Cover Classification. *IEEE Transactions on Geoscience and Remote Sensing*. 2010, vol. 48, iss. 11, pp. 4047–4058. ISSN 1558-0644. DOI: 10.1109/TGRS.2010.2072787.

[5] KIM, R. S. *Spectral Matching using Bitmap Indices of Spectral Derivatives for the Analysis of Hyperspectral Imagery*. Ohio, 2011. Master Thesis. The Ohio State University. Supervisor: Alper Yilmaz.

[6] Finite difference approximation of derivatives. *Rosenstiel School of Marina & Atmospheric Science I* [online]. 2017. Available at: https://www.rsmas.miami.edu/users/miskandarani/Courses/MSC321/lectfiniteDifference.pdf.

[7] TSAI, F. and W. PHILPOT. Derivative Analysis of Hyperspectral Data. *Remote Sensing of Environment*. 1998, vol. 66, iss. 1, pp. 41–51. ISSN 0034-4257. DOI: 10.1016/S0034-4257(98)00032-7.

[8] HORLER, D. N. H., M. DOCKRAY and J. BARBER. The red edge of plant leaf reflectance. *International Journal of Remote Sensing*. 1983, vol. 4, no. 2, pp. 273–288. ISSN 1366-5901. DOI: 10.1080/01431168308948546.

[9] ENVI Classic Tutorial: Target Detection. In: *Harris Geospatial Solutions* [online]. 2014. Available at: http://www.harrisgeospatial.com/portals/0/pdfs/envi/Target_Detection.pdf.

About Authors

Yucel CIMTAY was born in Elazig, Turkey in 1984. He received his B.Sc. degree from Bilkent University, Ankara, Turkey in 2008 and the M.Sc. degree from Eskisehir Osmangazi University, Eskişehir, Turkey in 2012. He is a doctorate student in Ankara University in the department of Electrical and Electronics engineering. His research interests include image and video processing, computer vision and machine learning.

Hakki Gokhan ILK was born in Ankara, Turkey in 1971. He received the B.Sc. degree from Ankara University, Ankara, Turkey, in 1993 and the M.Sc. degree instrument design and application, and the Ph.D. degree from the University of Manchester, Manchester, U.K., in 1994 and 1997, respectively. He is a Professor with Electrical and Electronics Engineering Department, Ankara University. His current research interests are digital signal processing, namely speech, image and video processing.