Approach Based on Fuzzy C-Means for Band Selection in Hyperspectral Images

Diego Saqui, José H. Saito, José R. Campos, Lúcio A. de C. Jorge

***Abstract*—**Hyperspectral images and remote sensing are important for many applications. A problem in the use of these images is the high volume of data to be processed, stored and transferred. Dimensionality reduction techniques can be used to reduce the volume of data. In this paper, an approach to band selection based on clustering algorithms is presented. This approach allows to reduce the volume of data. The proposed structure is based on Fuzzy C-Means (or K-Means) and NWHFC algorithms. New attributes in relation to other studies in the literature, such as kurtosis and low correlation, are also considered. A comparison of the results of the approach using the Fuzzy C-Means and K-Means with different attributes is performed. The use of both algorithms show similar good results but, particularly when used attributes variance and kurtosis in the clustering process, however applicable in hyperspectral images.

***Keywords*—**Band selection, fuzzy C-means, K-means, hyperspectral image.

I. Introduction

REMOTE SENSING (RS) is the technique that uses sensors to capture information from objects on the earth's surface without physical contact [1]. For RS be performed, sensors can be used to collect hyperspectral images (HI) that allow many applications [2]. RS and HI are applied in different areas such as agriculture, civil engineering, among others [1]. An example of application of these techniques is the identification of pests and plant diseases in agriculture [3].

HI increases sensitivity beyond the visible region because there are different bands in the spectral signature, i.e., beyond RGB (red-green-blue) bands. [3]. In an HI, each pixel contains a spectral signature composed by a large number of bands with different characteristics enabling distinction materials [2].

HI has a large number of bands (hundreds) and enough resolution to allow experts evaluates the spectrum of different materials with great precision. The result of the high spectral resolution is a large volume of data to be processed and analyzed [4]. Because of this, the practical use and research on HI has challenges related to storage, transmission and processing time. Another problem in the use of HI is that many of these bands contain redundant information due to high correlation existence between them. This redundant information can cause instability during the convergence of computing processes, such as classification and others that make use of these images [2]. Thus, based on the presented problems, a research interest in HI is to identify bands which are relevant to an analysis process, i.e., reduce the volume of data.

By reducing the number of bands, another problem that can be solved is related to the cost of manufacture of hyperspectral sensors. If the selected bands are close to each other in the spectral signature, multispectral sensors (which are cheaper than hyperspectral sensors) can be used.

Methods that reduce or identifying relevant characteristics of data set are described as dimensionality reduction and can be divided into two categories [2]: Feature extraction and feature selection. Feature selection is presented as a more viable alternative for dimensionality reduction. Feature extraction requires the original data representation to operate and during processing of feature extraction the information can be distorted. This is because of the fact that information obtained from real data could contain unknown signal sources [5]. A feature extraction example is the technique principal component analysis (PCA); In comparison with the feature extraction, feature selection preserves the original information. This is very important to reduce the data in HI because during the process is necessary to determine the specific bands. There are many techniques for feature selection such as, Search Based Methods [6], Transform Based Methods [7], Independent Component Analysis (ICA) - based band selection method [8] and information based methods [9] and recently, approaches based on clustering such as K-Means [10] and Ward's hierarchical clustering [11].

Based on clustering, in this paper the use of the Fuzzy C-Means (FCM) algorithm and new attributes are analyzed for the purpose of feature selection. One of the advantages of using clustering techniques for feature selection is the ability of these methods allows finding hidden patterns in data without prior knowledge of these data categories.

The use of FCM is motivated by the fact that: Based on a process of clustering, bands that have similar characteristics are grouped together. If the data of spectral bands are displayed in a Cartesian plane, it is difficult to establish precisely how these data are grouped characterizing uncertainty in the decision-making process. This analysis is similar to an expert, where for example, through the variance of the spectral bands, they try to determine which bands that have similar characteristics. This analysis is often characterized by uncertainty.

Diego Saqui, José H. Saito and José R. Campos are with the Computing Department, Federal University of São Carlos (UFSCar), São Carlos/SP, Brazil (e-mail: diego.saqui@dc.ufscar.br, saito@dc.ufscar.br, jrcampos82@gmail.com)

José H. Saito is with the Department of Computer Science, Faculty Campo Limpo Paulista (FACCAMP), Campo Limpo Paulista/SP, Brazil (e-mail: saito@dc.ufscar.br)

Lúcio A. de C. Jorge is with Embrapa Instrumentation, São Carlos/SP, Brazil (e-mail: lacjorge@gmail.com)

For evaluation criteria, the K-Means is also considered in this paper. Discussion about the attributes, methods and algorithms used in this study are presented in the following sections.

This paper is organized into seven sections: I. Introduction; II. Attributes for Band Selection; III. HFC and NWHFC; IV. K-Means and Fuzzy C-Means; V. The Proposed Methodology; VI. Experiments and Results; and VII. Conclusion.

II. Attributes for Band Selection

As shown previously, the method which the dimensionality is reduced by selecting a set of features is known as feature selection (band selection in this study). Therefore, it is necessary and very important to understand the application area where the feature selection is applied.

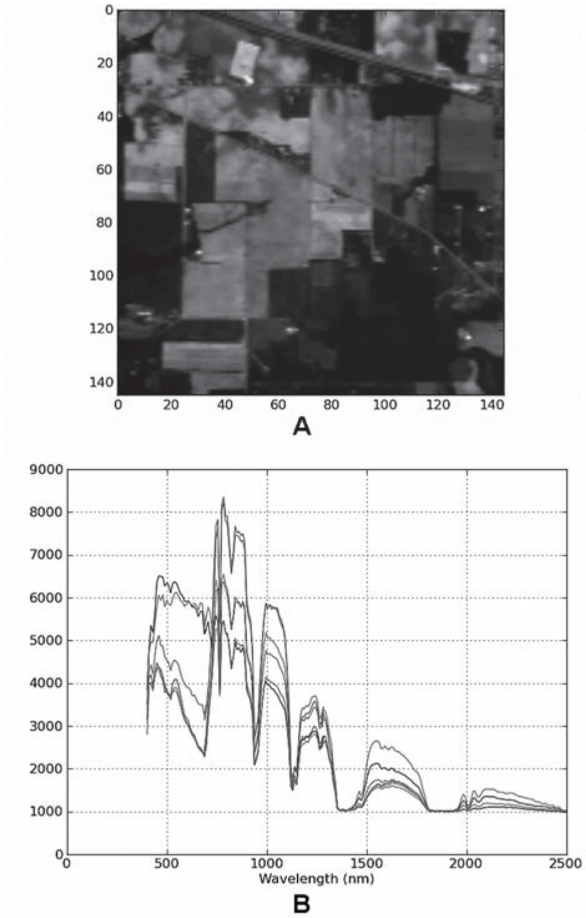

Fig. 1 A: 92AV3C.spc image in RGB format consisting by the bands 9, 29 and 19; B: Examples of spectral signatures generated from Fig. 1 A (Adapted from [13])

Figs. 1 A and B represent respectively an image in ERDAS/LAN format and spectral signature of some pixels. This image is titled 92AV3C.spc, has size of 145 x 145 pixels, 220 spectral bands and represents an area with buildings, cultures of agriculture and other natural vegetation [12].

Spectral signature is characterized by the wavelength on the abscissa and reflectance on the ordinate as shown in Fig. 1 B. Hyperspectral data can be statistically measured using variables as mean, standard deviation, variance, correlation and other measures. This can be seen in several papers that can be found in the literature, such as in [10].

For a clustering based on K-Means the mean absolute deviation (MAD), standard deviation or/and variance can be used [10]. The standard deviation, variance and MAD are data set of measures that have similar characteristics, so can be used only one of these variables. In this paper is used the variance that is shown in (1). The MAD is considered only for some tests in this study.

$$\sigma^2\_l = \frac{\sum_{i=1}^{N}(bi\_l - \overline{Bl})^2}{N} \tag{1}$$

N is the total number of pixels of an HI, $\overline{\mathrm{Bl}}$ is the average reflectance of pixels in the band l and bi_l is the reflectance value of pixel i in the band l. The variation of data in a spectral signature is relevant to band selection. If certain bands vary similarly, they can be clustered and so one element selected from this cluster for analysis.

Another measure that is used in this the Kurtosis. Kurtosis is a statistic measure of the dispersion that characterized the "tailedness" of the probability distribution. For a clustering process of this measure can be useful to group bands that have the same distribution of reflectance values independently of variance. The Kurtosis is shown in (2):

$$K_l = \left\{\frac{N(N-1)}{(N-1)(N-2)(N-3)}\sum_{i=1}^{N}\left(\frac{(bi_l - \overline{Bl})}{s}\right)^4\right\} - \frac{3(N-1)^2}{(N-2)(N-3)} \tag{2}$$

s is the standard deviation, all other variables are the same used in (1).

An example of the calculation of the variance and Kurtosis for a given image with 4 pixels and 3 bands is presented in Table I. In Table I, if only considered the variance in a clustering with K-Means (k = 2, Euclidian distance), bands 1 and 3 are in the same cluster because they have similar values. If only considered the kurtosis, the bands 2 and 3 are in the same cluster because of the same reason as before. This is because the variance and kurtosis do not exhibit the same behavior, unlike variance, standard deviation and mean absolute deviation. The variance and kurtosis are numeric and can be used in algorithms that make use of the Euclidean distance, such as K-Means and Fuzzy C-Means (FCM).

TABLE I
Variance and MAD for Spectral Bands

	Band 1	Band 2	Band 3
Pixel 1	3300	3200	5500
Pixel 2	3300	3200	5500
Pixel 3	3400	3500	5600
Pixel 4	3400	3650	5650
Variance	2500	37968,75	4218,75
Kurtosis	-6	-3,90123	-3,90123

The variance and kurtosis are used for the clustering process, but it is necessary to use a variable or method to determine which band is selected from the clusters formed. A measure that can be used to band selection is the Pearson's correlation coefficient (3) between bands of the image.

$$\rho = \frac{\sum_{i=1}^{MN}(bx_i - \overline{bx})(by_i - \overline{by})}{\sqrt{\sum_{i=1}^{MN}(bx_i - \overline{bx})^2} * \sqrt{\sum_{i=1}^{MN}(by_i - \overline{by})^2}} \quad (3)$$

ρ is the Pearson's correlation coefficient, bx and by are two specific bands to be compared by the coefficient and $\overline{bx}$ and $\overline{by}$ are the average values of the pixels in bands bx and by, respectively. The Pearson's correlation coefficient has values between -1 and 1. The closer to 0 the lower the correlation and the closer to -1 or 1 there is more correlation between two bands.

From (3), we can define some statistical metrics, such as average correlation, or the number of high/low correlation of a band compared to other bands. In this paper is used the number of bands that are not highly correlated (low correlation) from a threshold with respect to a specific band. For this is calculated the Pearson's correlation coefficient module to count the low correlation values. For example, in Table II is demonstrated some Pearson's correlation coefficients for 3 bands. In Table II, for example, if used a correlation threshold with the value of 0.6 then for Band 1 with respect to Band 2 (0.563 <0.6) and Band 3 (0616> 0.6) the low correlation is counted one time (low correlation = 1). If used another threshold, for example, with the value of 0.7 then for Band 1 with respect to Band 2 (0.563 <0.7) and Band 3 (0616 <0.7) the low correlation is counted two times (low correlation = 2). Therefore, the low correlation is determined from a specific threshold for comparison with all correlations.

TABLE II
BANDS CORRELATION

	Band 1	Band 2	Band 3
Band 1	1	0.563	0.616
Band 2	0.563	1	0.808
Band 3	0.616	0.808	1

This strategy of counting low correlation is applied in this study for the purpose of try to determine the best band in each cluster.

III. HFC AND NWHFC

The K-Means, Fuzzy C-Means (FCM) and some other clustering algorithms require a parameter to be initialized. This initial parameter is the number of clusters to be generated. For this purpose, in HI can be applied methods that allow estimating the virtual dimensionality (VD).

VD is the minimum number of spectrally distinct signal sources that characterize the hyperspectral data from the perspective view of target detection and classification. VD is often used as an estimated value of the minimum number of useful bands in HI. VD can be used as the number of clusters to be located by a clustering process. In the area of HI, there are two algorithms for this: HFC (Harsanyi-Farrand-Chang) and NWHFC (noise-whitened Harsanyi-Farrand-Chang) [2].

HFC implies the calculation of the eigenvalues of both the data-correlation $\{\bar{\lambda}l\}$ and covariance $\{\lambda l\}$ matrices, for the lth band. A signal source is present if their difference, $\bar{\lambda}l$ - λl, is positive. A binary-composite-hypothesis test is formulated for each spectral band such that the null (H0) and alternative (H1) hypothesis represent two scenarios, H0: $\bar{\lambda}l - \lambda l = 0$ and H0: $\bar{\lambda}l - \lambda l > 0$, corresponding to the absence and presence of a signal in the lth band, respectively. The NWHFC is an improved version of HFC, which considers information about white noise [14].

IV. K-MEANS AND FUZZY C-MEANS

There are different types of clustering algorithms, such as hierarquical or partitional. Partitional clustering is the process of categorizing the elements in clusters so that similar elements are allocated to a same cluster [15].

A classic partitional algorithm is the K-Means. K-Means is an unsupervised algorithm for solving the problem of clustering. K-Means allows the partition of a data set in K clusters so that these elements are just in one cluster (non-overlapping). The number of clusters K is defined a prior.

The elements of the data set are associated to the nearest centroid of a determined prototype (cluster). Again K new centroids are recalculated for new centers of the prototypes and a new clustering has to be done between the data set elements and the nearest new centroid. A loop is run step by step until there is no change and the centroids are fixed (convergence). The K-Means uses the Euclidean distance (or other metrics) to establish the position of the prototypes in relation to the elements [10]. The algorithm of the K-Means is:

1. Randomly choose a number of k prototypes (centers) for clusters
2. Assign each element to the nearest center cluster (according to some distance, e.g. Euclidean)
3. Move each center to the mean of the objects of the corresponding cluster
4. Repeat steps 2 and 3 until some convergence criterion is obtained:

- Maximum number of iterations
- Minimum threshold of changes in centroid

K-Means clustering technique can be used for clustering the bands in a process of band selection [10]. An algorithm similar but that considering data overlaps is the Fuzzy C-Means (FCM).

FCM is a clustering algorithm that allows elements of a data set to be in more than one cluster. This method was developed by Dunn in 1973 [16] and improved by Bezdek in 1981 [17] is commonly used in pattern recognition. As well as the K-Means, the algorithm "FCM" aims to minimize the function as shown in (4):

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^m ||x_i - c_j||^2 \quad (4)$$

where m is a real number greater than 1, N is the total number of elements, C is the total number of clusters, u_{ij} are values that representing the element i of the data set, c_j is the center of the cluster j and $\| * \|$ is the equation that expresses the similarity between the measured element and a cluster. The fuzzy partitioning is performed by an optimization of an iterative function J_m, updating the degree of relevance u_{ij} and determined cluster center by (5) and (6) respectively

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{||x_i - c_j||}{||x_i - c_k||}\right)^{\frac{2}{m-1}}} \tag{5}$$

$$c_j = \frac{\sum_{k=1}^{N} u_{ij}^{m} * x_i}{\sum_{k=1}^{N} u_{ij}^{m}} \tag{6}$$

The algorithm of the FCM is:

1. Select the initial centers: $c_1, c_2, ..., c_c$
2. Calculate u_{ij} (5)
3. Update centers c_j (6)
4. In case of no convergence, return to step 2.

V. The Proposed Methodology

Based on the presented algorithms, the proposed methodology is shown in Fig. 2 and described below:

(1) An HI is used as an input parameter; this image contains the number of bands and the positions of each of its pixels;
(2) The variance, the kurtosis and the number of bands with low correlation (values lower than 0.7) are calculated as shown previously in (1)-(3);
(3) The NWHFC algorithm is applied. This algorithm is executed 4 times with of false alarms probability values between 10-4 to 10-1. The average obtained is utilized as the VD. Details about this algorithm can be found in [14];
(4) VD is used as the number of clusters of Fuzzy C-Means (FCM), which is applied, and so the clusters to be considered are defined. This algorithm is applied to the bands represented by the variance and kurtosis variables;
(5) For each band, the cluster (highest degree of pertinence) most representative is indicated;
(6) For each indicated cluster, the band with lower correlation is selected;
(7) The selected bands are used to generate a new image.

The new image generated is then segmented to be applied to an evaluation process. The process of segmentation used in this study is based on Gaussian maximum likelihood classifier (GMLC) algorithm.

The GMLC operates on the variance and covariance of the spectral signature patterns to classify an unknown pixel. This algorithm assumes that the data distribution takes the form of a multivariate Gaussian. Probability density functions are used to classify an unknown pixel, calculating the probability of the pixel value belonging to each class. After evaluating the probability of each class, the pixel is assigned to the most likely class.

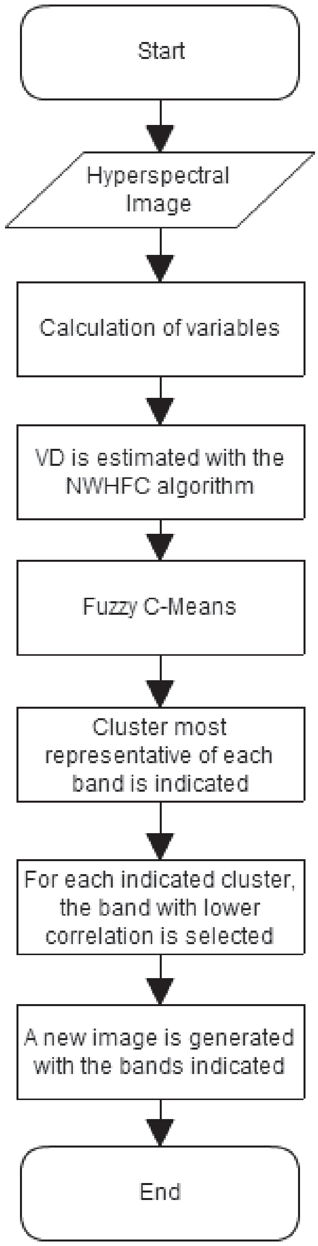

Fig. 2 Flow chart of proposed methodology

VI. Experiments and Results

The image '92AV3C.spc' with 220 bands is used for testing. This image is described with 17 classes that represent types of areas, as vegetation or soil conditions [12]. This image with 17 segmented classes in regions is represented in Fig. 3.

In the proposed methodology are calculated variance and kurtosis of each band hyperspectral image (HI). The NWHFC is applied to the data set and estimating the VD that which resulted in that same number of classes that the image '92AV3C.spc' (17 classes). The accuracy of the number of classes does not mean that clustering and segmentation should occur completely as shown in Fig. 3. In the segmented image, merged regions can be created by different classes generated by different bands. However, this value is an approximation of the variations that occur in the spectral signature indicating the possibility of finding different clusters around these parts representing the classes of Fig. 3.

Fig. 3 Image '92AV3C.spc' with 17 segmented regions (Adapted from [12])

The variance and the kurtosis of the bands image '92AV3C.spc' are used in the K-Means and FCM. In order to reduce the chance of local optimum in the algorithms, they were executed 100 times with random initialization of cluster centers, always with the number of clusters set to the value of 17. For each algorithm is selected the result with the lowest value of the objective function ((4) for FCM).

In Fig. 4, the K-Means and FCM result in many near each other characterizing clusters with poor definition. Therefore, it is difficult to analyze which of the two algorithms had better results.

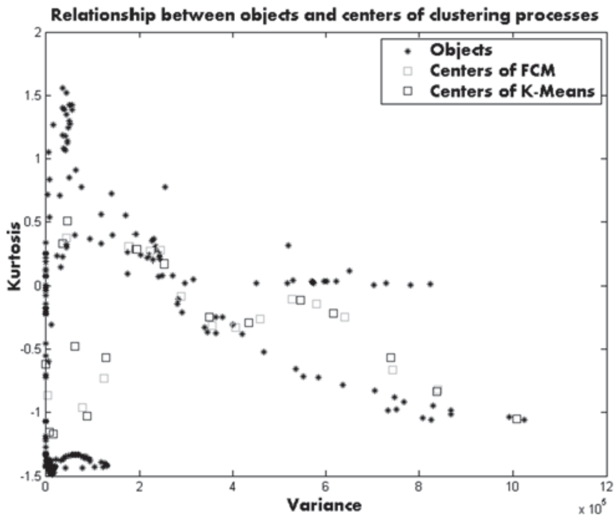

Fig. 4 Relationship between objects and clustering algorithms

One way to evaluate the result of the segmentation with selected bands is to compare the results with images generated from knowledge of a specialist. For this comparison, an expert can indicate the areas of interest in the image, and thus be verified that the developed method can perform the same operation. The image with the areas indicated by experts is demonstrated in Fig. 5.

Fig. 5 Image with the regions indicated by experts (Adapted from [13])

For data obtained from Fig. 5, a training process is applied and a supervised classifier can be used to image segmentation using all bands. In this paper is used the GMLC and the result of the segmentation using all bands is presented in Fig. 6.

Fig. 6 Segmented image by GMLC with all bands (220 bands)

The image shown in Fig. 6 is used as a criterion for comparison with the images generated from the bands selected by the K-Means and FCM.

The metric used to compare the images generated with the reduced number of bands and the segmented image with all the bands is the structural similarity index (SSIM). SSIM is a method for establishing the similarity between two images.

For K-Means are selected the bands 10, 14, 25, 28, 43, 46, 52, 57, 66, 81, 103, 117, 123, 130, 137, 196, 209. With these bands, the GMLC is applied and the result of the segmentation is shown in Fig. 7. Using the SSIM between this image and the original image the result was 0.858.

For FCM are selected the bands 15, 16, 22, 25, 28, 36, 38, 39, 61, 63, 66, 77, 81, 90, 91, 95, 104. With these bands, the GMLC is applied and the result of the segmentation is shown in Fig. 8. Using the SSIM between this image and the original image the result was 0.840.

Fig. 7 Segmented image by GMLC with 17 bands selected by K-Means with Variance and Kurtosis

Fig. 8 Segmented image by GMLC with 17 bands selected by FCM with Variance and Kurtosis

Fig. 9 Segmented image by GMLC with 17 bands selected by K-Means with Variance and (MAD).

For comparison is also generated an image (Fig. 9) based on previous research. The method used is similar to the structure proposed in this paper. For this, is used the K-Means, with the attributes variance and mean absolute deviation (MAD). The band selection for each cluster is based on the band nearest of the center of the clusters. The selected bands are 15, 16, 18, 22, 25, 28, 35, 36, 38, 39, 61, 63, 66, 67, 74, 90, 104. Using the SSIM between the image generated in this process and the original image the result was 0.837.

Based on the tests, the approach proposed using any of the two algorithms (K-Means or FCM) showed the ability to select significant bands to a segmentation process. If the method is used the variance and kurtosis the K-Means and FCM presented better performance than when using the variance and the mean absolute deviation with the K-Means. In the comparisons of the images generated by K-Means and FCM with variance and kurtosis, the first algorithm (K-Means) showed a slight improvement.

VII. Conclusion

This paper presents an approach to band selection of HI based on clustering algorithm Fuzzy C-Means (FCM). The band selection structure based on clustering algorithms partitional, had previously been proposed with the K-Means algorithm. In addition to the FCM, in this study is considered the variance and kurtosis in the clustering process. An additional step based on the correlation analysis to determine the best band for each cluster is also included.

For validation, the K-Means and FCM are both applied for band selection with different attributes. The results for both algorithms are similar and favorable in reducing the number of bands especially when using the attributes kurtosis and variance. In future work, we will explore other clustering algorithms in band selection, such as DBSCAN.

Acknowledgment

Resources that were used in this study were provided by the Laboratory of Images and modeling of Embrapa instrumentation and CNPq with the project 403426/2013-8.

References

[1] E. Novo, Sensoriamento Remoto: Princípios e Aplicações, 4 ed., Edgard Blücher, 2011.

[2] C. I. Chang, Techniques for Spectral Detection and Classification, New York: Kluwer Academic/Plenum Publishers, 2003.

[3] L. A. C. Jorge and R. Y. Inamasu, "Uso de veículos aéreos não tripulados (VANT) em Agricultura de Precisão," in Agricultura de Precisão. Resultados de um Novo Olhar., Embrapa, 2014.

[4] F. Baret, G. Guyot and D. J. Major, "TSAVI: a vegetation index which minimizes soil brightness effects on LAI and APAR estimation," International Geoscience and Remote Sensing Symposium (IGARSS'89); Canadian Symposium on Remote Sensing. IEEE, 1989., pp. 1355-1358, 1989.

[5] C. I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," IEEE Transactions on Geosciences and Remote Sensing, vol. 44, no. 6, pp. 1575-1585, 2006.

[6] G. Petrie, P. Heasler, and T. Warner, "Optimal band selection strategies for hyperspectral data sets," Geoscience and Remote Sensing Symposium Proceedings, 1998. IGARSS '98. 1998 IEEE International, vol. 3, pp. 1582-1584, July, 1998.

[7] P. Groves and P. Bajcsy, "Methodology for hyperspectral band and classification model selection," IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, pp. 120-128, 2003.

[8] H. Du, H. Qi, X. Wang, R. Ramanath, and W. E. Snyder, "Band selection using independent component analysis for hyperspectral image processing," in Applied Imagery Pattern Recognition Workshop, IEEE Computer Society, Los Alamitos, CA, USA, pp. 93- 98, 2003.

[9] A. M. U. Uso, F. Pla, J. M. Sotoca, and P. G. Sevilla, "Clustering based hyperspectral band selection using information measures," IEEE

Transactions on Geosciences and Remote Sensing, vol. 45, pp.158-4171, Dec. 2007.
[10] M. Sohaib, I. U. Haq and Q. Mushtaq, "Dimensional Reduction of Hyperspectral Image Data Using Band Clustering and Selection Based on Statistical Characteristics of Band Images," International Journal of Computer and Communication Engineering, vol. 2, Março 2013.
[11] A. M. U. Uso, F. Pla, J. M. Sotoca and P. G. Sevilla, "Clustering based hyperspectral band selection using information measures," IEEE Transactions on Geosciences and Remote Sensing, vol 45, 2007.
[12] D. Landgrebe, "Multispectral Data Analysis: A Signal Theory Perspective," School of Electr. Comput. Eng., Purdue Univ., West Lafayette, 1998.
[13] T. Boggs, Specytral Python (SPy). URL:http://www.spectralpython.net, 2014. Accessed April 2016.
[14] C. I. Chang, Hyperspectral data Processing: Algorithm Design and Analysis, Maryland, USA: Wiley, 2013.
[15] A. K. Jain and R. C. Dubes, Algorithms for Clustering Data, Prentice Hall, 1988.
[16] J. C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters," Journal of Cybernetics, vol. 3, pp. 32-57, 1973.
[17] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms," Plenum Press, 1981.