

CLASS-BASED AFFINITY PROPAGATION FOR HYPERSPECTRAL IMAGE DIMENSIONALITY REDUCTION AND IMPROVEMENT OF MAXIMUM LIKELIHOOD CLASSIFICATION ACCURACY

Propagação de Afinidade Baseada na Classe para redução de dimensionalidade em imagens hiperespectrais e melhoramento da acurácia na classificação por Máxima Verossimilhança

André Moiane^{1,2} – ORCID: 0000-0001-8242-6264

Alvaro Muriel Lima Machado² - ORCID: 0000-0002-7371-3461

¹ Universidade Federal do Paraná, Departamento de Geomática, Curitiba, Paraná, Brasil
E-mail: andremoiane16@gmail.com

² Instituto de Formação em Administração de Terras e Cartografia, Departamento de Investigação e Extensão, Matola, Maputo, Moçambique
E-mail: alvaroml@ufpr.br; alvaromlmachado@gmail.com

Received in May 2nd 2018

Accepted in August 19th 2018

Abstract:

This paper investigates an alternative classification method that integrates class-based affinity propagation (CAP) clustering algorithm and maximum likelihood classifier (MLC) with the purpose of overcome the MLC limitations in the classification of high dimensionality data, and thus improve its accuracy. The new classifier was named CAP-MLC, and comprises two approaches, spectral feature selection and image classification. CAP clustering algorithm was used to perform the image dimensionality reduction and feature selection while the MLC was employed for image classification. The performance of MLC in terms of classification accuracy and processing time is determined as a function of the selection rate achieved in the CAP clustering stage. The performance of CAP-MLC has been evaluated and validated using two hyperspectral scenes from the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) and the Hyperspectral Digital Imagery Collection Experiment (HYDICE). Classification results show that CAP-MLC observed an enormous improvement in accuracy, reaching 94.15% and 96.47% respectively for AVIRIS and HYDICE if compared with MLC, which had 85.42% and 81.50%. These values obtained by CAP-MLC improved the MLC classification accuracy in 8.73% and 14.97% for these images. The results also show that CAP-MLC performed well, even for classes with limited training samples, surpassing the limitations of MLC.

Keywords: Affinity Propagation; Band selection; Maximum Likelihood Classifier; Classification; Hyperspectral image

How to cite this article: MOIANE, A. and MACHADO, A. M. L. Class-Based Affinity Propagation for Hyperspectral Image Dimensionality Reduction and Improvement of Maximum Likelihood Classification Accuracy. *Bulletin of Geodetic Sciences*, 25(1): e2019004, 2019.



This content is licensed under a Creative Commons Attribution 4.0 International License.

Resumo:

Este artigo investiga um método de classificação alternativo que integra o algoritmo de clusterização propagação de afinidade baseado nas classes (PAC) e o Classificador Máxima Verossimilhança (MAXVER) com a finalidade de superar as limitações do MAXVER na classificação de dados de alta dimensionalidade e, assim, melhorar a sua acurácia. O novo classificador foi designado PAC-MAXVER, e compreende duas abordagens, seleção de características espectrais e classificação de imagem. O algoritmo de clusterização PAC foi usado para realizar a redução de dimensionalidade da imagem e seleção de características enquanto o MAXVER foi utilizado para a classificação da imagem. O desempenho do MAXVER em termos de acurácia da classificação e tempo de processamento é determinado em função da taxa de seleção realizada na fase de clusterização PAC. O desempenho de PAC-MAXVER foi avaliado e validado usando duas cenas hiperespectrais do AVIRIS (Airborne Visible Infrared Imaging Spectrometer) e HYDICE (Hyperspectral Digital Imagery Collection Experiment). Os resultados da classificação mostram que PAC-MAXVER observou uma enorme melhoria na acurácia, atingindo 94.15% e 96.47%, respectivamente, para AVIRIS e HYDICE se comparado com o MAXVER, que obteve 85.42% e 81.50%. Esses valores obtidos pelo PAC-MAXVER melhoraram a acurácia da classificação MAXVER em 8.73% e 14.97% para essas imagens. Os resultados também mostraram que o PAC-MAXVER teve um bom desempenho, mesmo para as classes com número limitado de amostras de treinamento, superando as limitações do MAXVER.

Palavras-chave: Propagação de Afinidade; Seleção de bandas; Classificador Máxima Verossimilhança; Classificação; Imagem hiperespectral

1. Introduction

High dimensionality data can offer a discriminating power much higher than traditional data at low dimensionality (Lee and Landgrebe, 1993; Jimenez and Landgrebe, 1999; Serpico and Bruzzone, 2001; Ablin and Sulochana, 2013). According to Fukunaga (1990), classes very similar spectrally can be separated satisfactorily in higher dimension spaces. This is one of the motivations for the development of sensor systems with large number of spectral bands, known as hyperspectral sensors.

Classification of hyperspectral data is a challenging research topic in remote sensing domains and pattern recognition (Bartels and Wei, 2006; Brzank and Heipke, 2007). However, one of the main difficulties that arise in the classification process of hyperspectral images through parametric classifiers like Maximum Likelihood Classifier (MLC) refers to the low number of training samples (limited in general) in comparison with the number of parameters to be estimated (Cortes and Vapnik, 1995; Camps-Valls *et al.*, 2014). A limited number of training samples results in an unreliable estimation of parameters in parametric classifiers and, consequently, in a reduced value on the accuracy of the produced thematic image (Landgrebe, 2002).

By starting the classification process with reduced dimensionality data, the thematic image accuracy tends, initially, to increase to the extent that additional information (spectral bands) is included. At a given moment, the accuracy reaches a maximum and then reduces, to the extent that the data dimensionality continues to increase. This problem is known as *Hughes phenomenon*,

and has been studied by researchers such as Hoffbeck and Landgrebe (1996) and Jimenez and Landgrebe (1999), among others.

Data dimensionality reduction by means of extraction techniques or selection of variables, introduction of semi-labelled training samples, regularised discriminant analysis techniques, are approaches that have been investigated by researchers in order to minimize the consequences of such phenomenon. Several approaches have been proposed for hyperspectral image classification. Roessner *et al.* (2001) have combined MLC with linear spectral unmixing. Segl *et al.* (2003) used thermal hyperspectral imagery for building detection improvement. A comparative study on hyperspectral data classification, including multilayer neural network, MLC and support vector machine (SVM) was conducted by Mather (2003).

In this paper, we explore the use of affinity propagation (AP) algorithm for feature selection prior to supervised classification of hyperspectral images using MLC. The proposed classification strategy here called CAP-MLC (Class-based AP-MLC) aims to overcome the limitations of traditional MLC, and improve the mapping accuracy in hyperspectral images. For comparison and evaluation purposes, two hyperspectral data sets, namely the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) and Hyperspectral Digital Imagery Collection Experiment (HYDICE) were used. This work has as main contributions the following: (i) the use of limited training samples for classification of high dimensionality data; and (ii) class-based hyperspectral image band selection and dimensionality reduction through supervised AP.

2. Affinity Propagation algorithm

Affinity propagation (AP) is an algorithm that identifies centres of clusters, also called *exemplars* to form its clusters around them. This algorithm simultaneously considers all the points in the set as probable candidates to become centres of the clusters and propagate exchanges of messages between the points until the emergence of good exemplars and clusters (Frey and Dueck, 2007).

AP uses as input real-valued similarities $S(i,j)$, describing how well the j -th point is appropriated to become an exemplar for the i -th point. When the points lay along the matrix diagonal, i.e., $i = j$, the similarity matrix $S(i,j)$ is called *preference*, and indicates how probable the i -th point is to be selected as an exemplar. Preferences can be set to a global value, or for particular data points. High preference values will cause AP to find many clusters, while low values will lead to a small number of clusters. A good initial choice to determine the preference is to take the minimum or the median similarities.

The similarity is commonly expressed as a negative squared Euclidean distance according to equation (1), in which the parameters x_i and x_j are the positions of data points i and j in 2D space (Dueck, 2009).

$$S(i, j) = -\|x_i - x_j\|^2 \quad (1)$$

The number of defined centres of clusters is mainly influenced by the values of preference, but it either emerges from the message exchanging process in the factor graph. A factor graph is defined as a bi-partite graph consisting of a set of nodes representing random variables and a set

of functions. This graphical model represents global functions or probability distributions that can be factored into simpler local functions (Frey and Dueck, 2007).

The process of sending messages is presented in Figure 1. In the figure, availability and responsibility messages are exchanged. *Responsibilities* are sent from data point i to candidate exemplar point k , and show how evident point k is to be an exemplar for point i , counting with other potential exemplars for point i . *Availabilities*, are sent from candidate exemplar point k to point i , and show the chance the point k has to be selected as its exemplar, considering the support the other points give (Dueck, 2009).

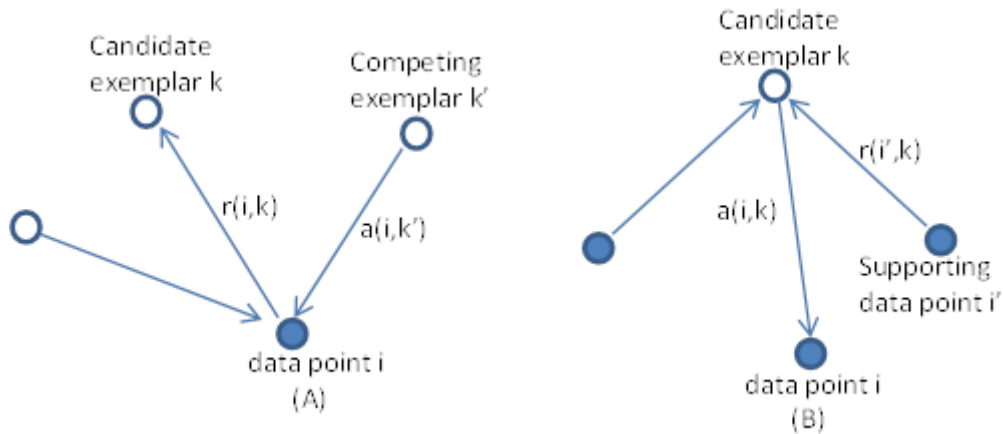


Figure 1. Propagation of two messages between data points: (A) “responsibilities” $r(i,k)$ are sent for data point i to candidate exemplar k , and (B) “availabilities” $a(i,k)$ are sent from candidate exemplar k to data point i .

The responsibility and availability values are adjusted as follows:

$$r(i, k) = S(i, k) - \max_{k': k' \neq k} [S(i, k') + a(i, k')] \quad (2)$$

$$a(i, k) = \sum \max[0, r(i', k)] \quad \text{for } k = i, i': i' \neq i \quad (3)$$

$$a(i, k) = \min \left[0, r(k, k) + \sum_{i': i' \in \{i, k\}} \max[0, r(i', k)] \right] \quad \text{for } k \neq i \quad (4)$$

In the equation (2), the letter i represents a data point and k' stands for a competing candidate exemplar. In the first iteration the availabilities are initialized to zero, and $r(i, k)$ is set to the input similarity between point i and point k as its exemplar, minus the maximum of the similarities between point i and other candidate exemplars k' .

In equations (3) and (4), the availability $a(i, k)$ is set to the self-responsibility $r(k, k)$ plus the sum of the positive responsibilities candidate exemplar k receives from other supporting points i' .

At each iteration, the assignment of items to exemplars is defined as:

$$\varphi(x_i) = \operatorname{argmax}_k \{r(i, k) + a(i, k)\} \quad (5)$$

In the equation (5), $\phi(x_i)$ is the exemplar for data point x_i . At any point, summing Responsibility $r(i,k)$ and Availability $a(i,k)$ matrices gives the clustering information needed for point i . The k with maximum $r(i, k) + a(i, k)$ represents point i 's exemplar.

The message propagation process stops as soon as it reaches a specified number of iterations or when the cluster structure stabilises with a given number of iterations, that is, the process converges if every exemplar $\phi(x_i)$ remains unchanged for some constant iterations, usually 10 (Dueck, 2009).

3. Materials and Methods

3.1 Materials

This research was performed using the following materials: MATLAB R2016a software for running affinity propagation routines; ENVI 4.6.1 software, used for removal of noisy bands and those with irrelevant information. This software was also used for image classification and for accuracy assessment; MultSpec was used for collection of training and testing samples used in the supervised clustering stage with affinity propagation. And, finally, ERDAS IMAGINE 2014 software was used to assist in the collection of training and testing samples for final image classification, and to save ENVI images into other formats readable in MATLAB.

3.1.1 AVIRIS hyperspectral image

This is a hyperspectral image 92AV3C, available at <http://www.tec.army.mil/Hypercube>. The image was provided by the AVIRIS sensor in 1992, and corresponds to the Indian Pine Test Site in North western Indiana. The image originally has 224 bands, a spatial dimension of 145 x 145 pixels and a spatial resolution of 20m per pixel (Baumgardner *et al.*, 1992). Classes range from 20 to 2468 pixels. In it, three different growing states of soya can be found, together with other three different growing states of corn. Woods, pasture and trees are the bigger classes in terms of number of samples (pixels). Smaller classes are steel towers, hay-windrowed, alfafa, drives, oats, grass and wheat. In total, the dataset has 16-labelled classes (Landgrebe, 2003). The AVIRIS hyperspectral image and the ground truth used to perform the experiments are shown in Figure 2.

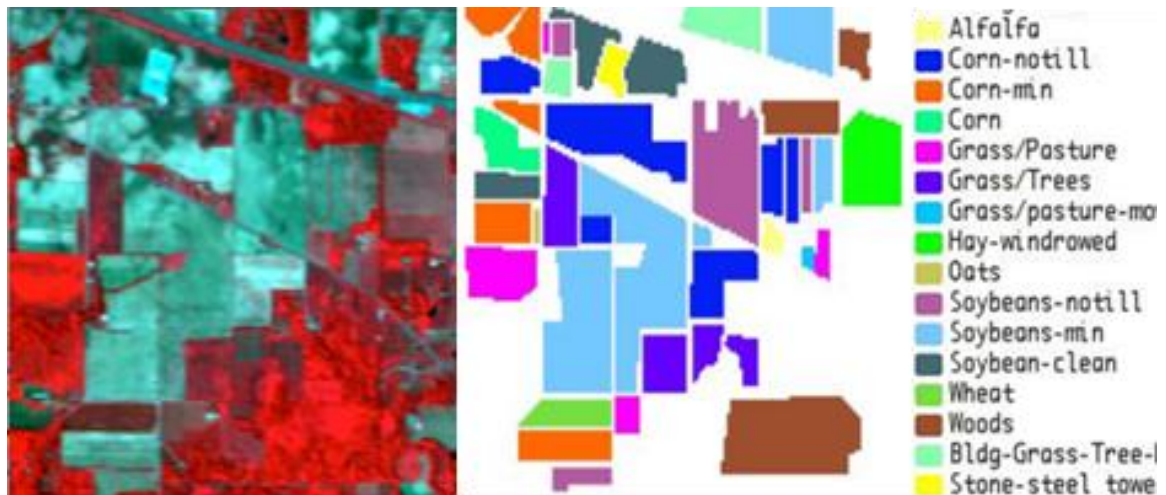


Figure 2: AVIRIS image for study site and ground truth.

The classes considered for classification, training and testing samples collected in AVIRIS image are given in the Table 1.

Table 1: Classes and ground truth for training and test.

Nr.	Classes	Training	Test	Total
1	Alfafa	22	17	39
2	Corn-notill	515	373	888
3	Corn-min	443	232	675
4	Corn	112	56	168
5	Pasture	313	107	420
6	Trees	420	149	569
7	Pasture-mowed	20	9	29
8	Hay-windrowed	355	62	417
9	Oats	16	5	21
10	Soybeans-notill	618	152	770
11	Soybeans-min	1628	239	1867
12	Soybean-clean	353	112	465
13	Wheat	147	47	194
14	Woods	558	234	792
15	Bldg-GTD	262	103	365
16	Stone-steel	80	30	110

3.1.2 HYDICE hyperspectral image

This image was collected in October 1995 by the HYDICE sensor, is available at <http://www.tec.army.mil/Hypercube>, and represents the location of an urban area in Copperas Cove, in United States. There are 307 x 307 pixels, each of which corresponds to an area of 2 x 2 m². There are 210 wavelengths in the image, ranging from 400 nm to 2500 nm, resulting in a spectral resolution of 10 Nm. There is also a ground truth with 4 classes: asphalt, grass, tree and roof. The study area image and ground truth are given in Figure 3.



Figure 3: HYDICE image for study site and ground truth.

The classes considered for classification, training and testing samples collected in HYDICE image are given in the Table 2.

Table 2: Classes and ground truth for training and test.

Nr.	Classes	Training	Test	Total
1	Asphalt	9898	7871	17769
2	Grass	16552	4704	21256
3	Tree	10228	2732	12960
4	Roof	3866	1722	5588

3.2 Methods

The steps of the methodology used in this paper are shown in the flowchart of Figure 4.

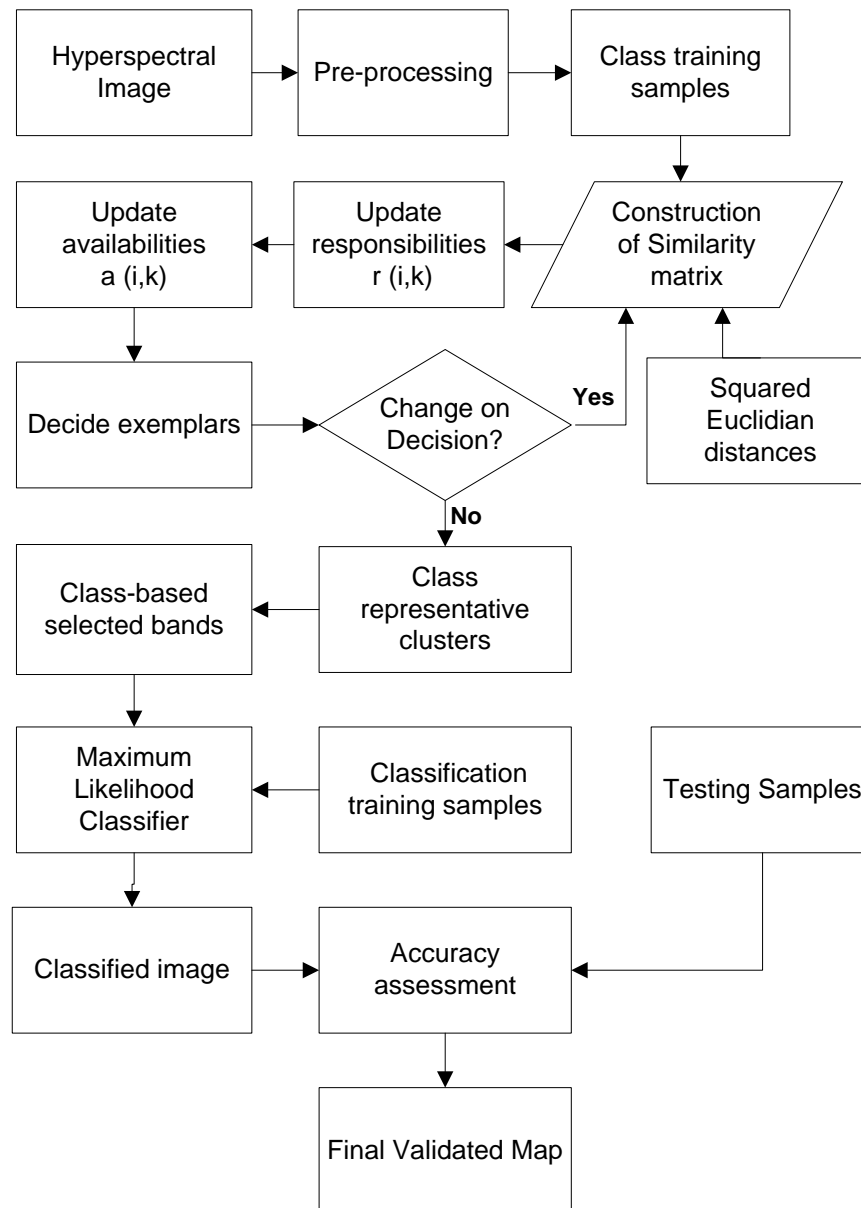


Figure 4: Flowchart for implementation of the methodology proposed.

The proposed hyperspectral image classification approach can be divided into four main stages: Image pre-processing, affinity propagation-based image band selection and dimensionality reduction, image classification and, accuracy assessment and validation.

3.2.1 Image pre-processing

High dimensionality images have spectral regions with noise bands caused by atmospheric effects. This noise may cause inconveniences at the time of data analysis and processing (Refianti

et al., 2016). For this reason, perform a preliminary phase for exclusion and removal of noisy bands is recommended (Ramakrishnan and Bharti, 2015). In this paper, the preprocessing stage was performed for both AVIRIS and HYDICE images.

AVIRIS image is composed of 224 spectral bands, but includes regions of the spectrum with noise bands due to the interference of water vapor from the atmosphere (Baumgardner *et al.*, 1992). ENVI 4.6.1 software was used to remove 20 bands (104 - 108, 150 - 163, 220) from the original image, due to absorption of water and low signal to noise rate. Additional 10 bands were excluded from this image due to irrelevant information, and the final preprocessed image remained with only 190 bands.

For HYDICE image, the same procedure was used, and bands 1-4, 76, 87, 101-111, 136-153 and 198-210 were removed due to the dense water vapour and atmospheric effects, and after pre-processing only 162 bands remained.

3.2.2 Class-based Affinity Propagation image band selection and dimensionality reduction

The affinity propagation image band selection and dimensionality reduction was firstly supervised (clustering based on class training samples) and then unsupervised (band-based clustering), and comprised the following steps:

3.2.2.1 Construction of Similarity Matrix

The similarity matrix is used to group points or nodes into clusters. After data pre-processing stage, Multispec software was used to collect training samples to construct the similarity matrix. Two approaches were considered in construction of the similarity matrix: class-based (supervised) and band-based (unsupervised).

In the supervised AP phase, various training samples with pixel vectors belonging to the same class were selected. Then, all pixel vectors were organised in a table (row and column) according to the class they belong, and the pixels digital values extracted using a MATLAB code developed throughout this research.

After extraction of pixel values for different training samples, class similarity matrices were constructed. To do that, the pixel digital values were used, and the negative squared Euclidean distance given by equation (1) was used to compute the distances between pairwise points (pixel values) for one class each time. These class similarities were then used as input data to AP clustering algorithm.

This procedure was repeated for all sixteen classes in AVIRIS image and for the four classes existing in HYDICE image. At the end of this stage, singular class-based similarity matrices were constructed, one for each class selected during the collection of training samples.

In the unsupervised phase, each image band obtained from supervised AP clustering was considered as a data point as can be seen a Figure 5, and similarities were constructed by computing squared Euclidian distances between these points.

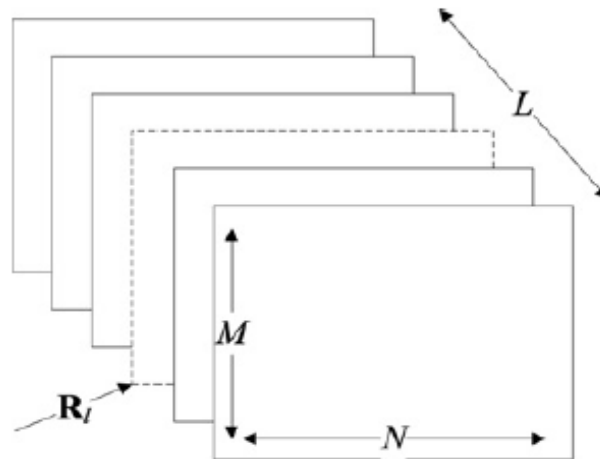


Figure 5. Representation of a hyperspectral image cube.

From the figure, can be seen that a hyperspectral imagery is a three dimensional array with the width and length corresponding to spatial dimensions and the spectral bands to the third dimension. These dimensions are denoted by M , N and L in sequence, and R is an image cube with each band $R_l \in \mathbb{R}^{M \times N}$ being an image matrix. Each band image can be considered as a data point with $M \times N$ dimensions.

3.2.2.2 Evidence Calculation

The evidence calculation comprises the responsibility and availability messages propagation process in the factor graph. This procedure initializes with availability $a(i,k) = 0$ for the first iteration, and the responsibility $r(i,k)$ is initialized as $S(i,k)$. From the second iteration, $a(i,k)$ will no longer be equal to zero, and then responsibilities and availabilities matrices are updated, using the equations (2), (3) and (4).

All the evidence computation in this class-based and supervised AP clustering was performed considering the preference parameter as the minimum of the similarities laying on the matrix main diagonal ($p = \min S(i,i)$) to obtain a reduced number of clusters, and dumping factor $\lambda = 0.9$ to avoid oscillations caused by the message passing mechanism.

The unsupervised phase was the last and performed separately using all class representative bands obtained from the supervised phase. For these phase the evidence calculation was considered with preference parameter set to the median of similarities and dumping factor $\lambda = 0.9$. The preference set to median allowed for moderate clustering solutions.

3.2.2.3 Cluster Exemplars Assignment

In order to assign exemplars or cluster centres from the similarity matrix, first the responsibility and availability matrices are created and updated. After successful creation and

update of these two matrices, the clustering results will be determined and the exemplars assigned using equation (5).

From the matrix obtained in equation (5), the exemplars can be easily assigned by selecting diagonal values greater than zero. This process of message propagation and responsibility update until cluster identification will be repeated until it reaches a specified number of iterations or when the cluster structure stabilises with a given number of iterations (Dueck, 2009).

The final step of AP-based band selection is the decorrelation, performed by choosing only the representative bands of each cluster. In these conditions, the cluster centres are generally considered preferable bands as they are highly correlated to the remaining bands in the same cluster (the similarity *within* clusters is high and the similarity *between* clusters is low). The removal of correlated bands is a very important step to reduce the dimensionality of the hyperspectral image. The selection of the centres of the clusters as representative of the bands was performed considering for one class each time, and the resulting bands as representative of that class.

3.2.2.4 Image Classification Stage

After image band selection and dimensionality reduction, ENVI 4.6.1 software was used to perform the image classification. For this purpose, MLC was adopted.

MLC is the traditional method most commonly used when it is necessary to obtain informational classes from remote sensing images. Before performing the classification, the following assumptions were considered:

- The spectral distribution of the classes is considered as being Gaussian or normal, i.e., objects belonging to the same class will present spectral responses next to the average values for that class (Maselli *et al.*, 1992; Richards, 1999).
- The method considers the weighted average distances, using statistical parameters for the distribution of pixels within a given class (Crósta, 1993).
- To achieve a good result with this classifier, it is necessary to choose a fairly high number of pixels for each sample of training class, and that they have a statistical distribution closer to the normal distribution (Crósta, 1993; Landgrebe, 2003).

The general procedures for MLC are the following:

1. The number of land cover types (classes) within the study area was determined for both AVIRIS and HYDICE hyperspectral images with support of Figures 2 and 3.
2. The training samples (pixels) for each of the desired classes were collected from the AVIRIS and HYDICE hyperspectral images based on the information of the study area (ground truth), as can be seen in Tables 1 and 2.
3. The collected and trained samples were then used to estimate the mean vector and covariance matrix of each class.
4. And, finally, each pixel in the image was classified in one of the desired land cover types or labelled as unknown.

In this stage of the proposed approach the final classification maps have already been performed. To enable classification performance comparison, three traditional classification methods namely PCA, MLC and AP were used.

3.2.2.5 Classification Accuracy Assessment and Validation

The last step in our approach consists to assess the image classifier accuracy, and validate the classification. To do that, confusion or error matrix and kappa coefficient were used. The confusion matrix is given by the overall accuracy (OA), producer's accuracy (PA) and user's accuracy (UA), while the kappa statistics is represented by the kappa coefficient (KC). The OA represents the ratio between the number of samples correctly recognized by the classification algorithm and the total number of test samples. According to Scepan (1999), the minimum acceptable OA is 85%.

The producer's accuracy (PA) informs the image analyst about the number of correctly classified pixels in a specific category, and measures the omission errors. The user's accuracy (UA) is computed using the number of correctly classified pixels and the total number of pixels assigned to a particular category (Story and Congalton, 1986).

The Kappa coefficient (KC) is a measure of the relationship between the possibility of agreement and disagreement expected. This is the second measure of classification accuracy that incorporates the elements off the diagonal as well as those of the diagonal of the confusion matrix, giving a more robust accuracy assessment than the overall accuracy. The value of KC is in the interval $[-1, 1]$. The more the KC value close to one, the better the classification. Although negative KC values are possible, Cohen (1960) notes that are unlikely to happen in practice, and when it happens it is an indicator of a serious problem, because negative values of KC represent a disagreement. According to (Cohen, 1960; Landis and Koch, 1977), the KC coefficient can be interpreted in accordance with Table 3.

Table 3: Interpretation of kappa coefficient values.

Agreement levels values for KC		
0-0.2	None	0-4%
0.21-0.39	Minimum	4-15%
0.40-0.59	Weak	15-35%
0.60-0.79	Moderate	35-63%
0.80-0.90	Strong	64-81%
>0.90	Almost perfect	82-100%

4. Results and Discussion

The experiments were performed considering two hyperspectral images, AVIRIS and HYDICE. In order to assess the accuracy of the proposed, here called CAP-MLC, three commonly used classifiers namely PCA, MLC and the original AP were used. The final experimental results are presented in Figures 6 - 8 and Tables 4 - 7.

AVIRIS Image Experiments

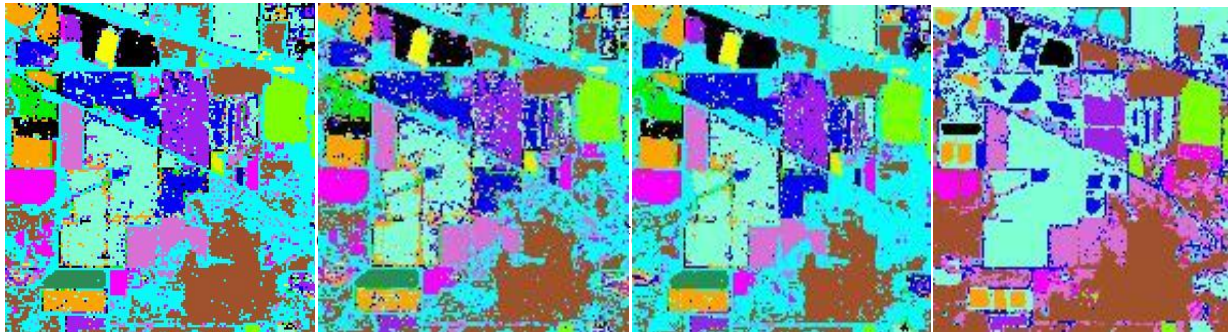


Figure 6. Classified hyperspectral images for CAP-MLC, AP-MLC, PCA-MLC and MLC.

Table 4: Accuracies and kappa statistics for the experiments.

Accuracy (%)	CAP-MLC				AP-MLC				PCA-MLC				MLC			
	PA	OE	UA	CE	PA	OE	UA	CE	PA	OE	UA	CE	PA	OE	UA	CE
Alfafa	92.9	7.1	100	0	100	0	100	0	88.2	11.8	100	0	Not classified			
Cornot	43.3	56.7	81.3	18.8	48	52	63.2	36.8	46.2	53.9	92.3	7.7	57.1	42.9	99.5	0.5
Cormi	100	0	100	0	91.4	8.6	100	0	89.5	10.5	100	0	73.7	26.3	100	0
Corn	100	0	96.6	3.5	100	0	73.7	26.3	100	0	80	20	Not classified			
Grasp	100	0	100	0	100	0	100	0	100	0	100	0	97.2	2.8	100	0
Grasst	100	0	92.6	7.4	93.1	6.9	93.1	6.9	100	0	100	0	98.7	1.3	96.1	3.9
Grasspm	100	0	100	0	100	0	100	0	100	0	100	0	Not classified			
Haywind	100	0	97.9	2.1	100	0	100	0	100	0	100	0	100	0	100	0
Oats	100	0	100	0	100	0	83.3	16.7	100	0	100	0	Not classified			
Soynot	81.3	18.8	100	0	61.1	38.9	91.7	8.3	80	20	100	0	94.7	5.3	98	2
Soymin	100	0	54.8	45.2	100	0	66.7	33.3	100	0	70.6	29.4	100	0	50.3	49.7
Soyclean	100	0	100	0	94.3	5.7	89.2	10.8	100	0	97	3	90.2	9.8	100	0
Wheat	100	0	100	0	100	0	100	0	100	0	100	0	Not classified			
Woods	100	0	92.3	7.7	100	0	91.8	8.2	100	0	83.8	16.2	100	0	95.5	4.5
Bgtd	87.1	12.9	100	0	84.4	15.6	96.4	3.6	81.8	18.2	81.8	18.2	88.4	11.7	100	0
Stonst	100	0	100	0	100	0	100	0	100	0	100	0	Not classified			
OA	94.15				92.20				93.24				85.42			
KC	0.94				0.92				0.93				0.83			

PA- Producer's Accuracy, OE= Omission Errors, UA= User's Accuracy, CE= Commission Errors, OA=Overall Accuracy, KC=kappa coefficient.

Table 5: Summary of overall accuracies and kappa coefficients for the experiments.

Method	OA	KC	Bands	Classes
CAP-MLC	94.15	0.94	14	16
AP-MLC	92.20	0.92	09	16
PCA-MLC	93.24	0.93	09	16
MLC	85.42	0.83	190	10

In the experiments performed with the AVIRIS hyperspectral image, a total of sixteen (16) classes were considered, namely: alfalfa, cornot, cormi, corn, grassp, grasst, grasspm, haywind, oats, soynot, soymin, soyclean, wheat, woods, bgtd and stonst. The classification method proposed for this study was CAP-MLC, which integrates class-based supervised AP and MLC, and the methods AP-MLC, PCA-MLC and MLC were used for comparative purposes.

According to the results presented in Table 5, the method CAP-MLC presented better separation between the classes, with an overall accuracy of 94.15% and kappa coefficient of 0.94. The classification using the PCA-MLC method resulted in an overall accuracy of 93.24% and kappa coefficient of 0.93, being the second best method to separate the classes, as illustrated in Table.

The AP-MLC method had an overall accuracy of 92.20%, slightly below PCA-MLC, and a kappa coefficient of 0.92. The worst separation between classes has been verified by the traditional MLC method, with an overall accuracy of only 85.42%, which according to Scepan (1999) this is the minimum acceptable value. The value of the kappa coefficient for MLC was 0.83.

According to Table 5, the accuracy values obtained by employing these methods explain that 94.15%, 93.24%, 92.20% and 85.42% of the pixels in the image were correctly classified, and respectively 5.85%, 6.75%, 7.8% and 14.58% of the pixels were erroneously classified.

The MLC method was applied with the original image, considering all 190 bands, and classified only 10 classes, contrarily to the other methods presented here which classified the 16 classes, as shown in Figure 6 and in Table 4. This limitation of the MLC in classifying some of the classes such as alfalfa, corn, grasspm, oats, wheat and stonst, is attributed to the fact that these classes have a small number of training samples in a high dimensionality feature space, as is evidenced by Landgrebe (2003) and Camps-Valls *et al.* (2014).

From the Table 4, it can be observed that the proposed CAP-MLC method resulted in a producer's accuracy of 100% for a total of 12 classes, meaning that only 4 classes, namely alfalfa, cornot, soynot and bgtd had errors in classification. During the classification process, the class alfalfa was omitted in 7.14%, cornot was omitted in 56.67% and allocated by commission in 18.75%, soynot was omitted in 18.75% and bgtd omitted in 12.9%. Within the classes with errors in the producer's accuracy, cornot class was the one that had lower accuracy in all experiments. The values obtained were 43%, 48%, 46% and 57% respectively for CAP-MLC, AP-MLC, PCA-MLC and MLC.

The best accuracy in the classification obtained by the proposed CAP-MLC method, has proved that this method presents discriminative features that make it not sensitive to the limited number of training samples when classifying high dimensionality data, as is supported by the literature (Peng *et al.*, 2016).

As regards the producer's accuracy obtained by the PCA-MLC method, this was 100% for the same classes that of the CAP-MLC, differing only in a class, cormi. The PCA-MLC method classified correctly 11 classes. The remaining 5 classes with classification errors, namely alfalfa, cornot, cormi, soynot and bgtd, were omitted in respectively 11.76%, 53.85%, 10.53%, 20% and 18.18%. The commission errors in this method were 7.69% and 18.18% respectively for cornot and bgtd.

From the 10 classes classified by MLC method, only 3 (haywind, soymin and woods) were correctly classified, being the remaining 7 with classification errors. During the Classification, the classes were erroneously omitted in the following way: cornot (42.90%), cormi (26.29%), grassp (2.80%), grasst (1.34%), soynot (5.26%), soyclean (9.82%) and bgtd (11.65%). In the same process, commission errors were observed, and classes were wrongly attributed to cornot in 0.47%, grasst in 3.92% and 2.04% to soynot.

The values of the kappa coefficients for the experiments with CAP-MLC, AP-MLC, PCA-MLC and MLC were respectively 0.94, 0.93, 0.93 and 0.83. According to Cohen (1960) and Landis and Koch (1977), these values suggest that the methods CAP-MLC, AP-MLC and PCA-MLC that reach Kappa coefficient above 0.9, showed an almost perfect agreement between classes and good classification accuracy. The MLC method, with kappa coefficient equal to 0.83 showed an agreement between classes from moderate to strong, and a moderate accuracy.

As these KC values evaluate the homogeneity of the samples among themselves and, once they have been taken from the image, between the samples and the rest of the image, there is a difficulty of using old images and lacking of ground truth. This causes the homogeneity of the samples on the image, thus limits the resulting accuracy, and in some cases the generalization of the methods.

HYDICE Image Experiments

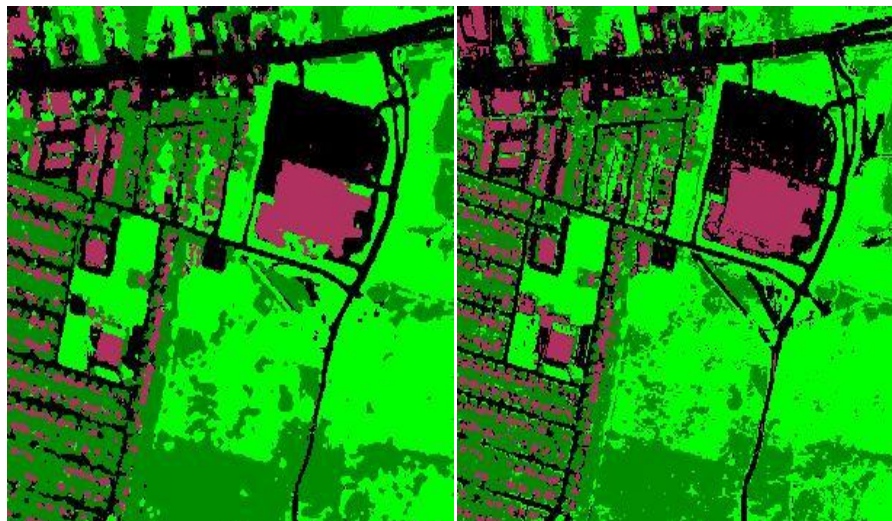


Figure 7: Classified hyperspectral images for CAP-MLC and PCA-MLC.

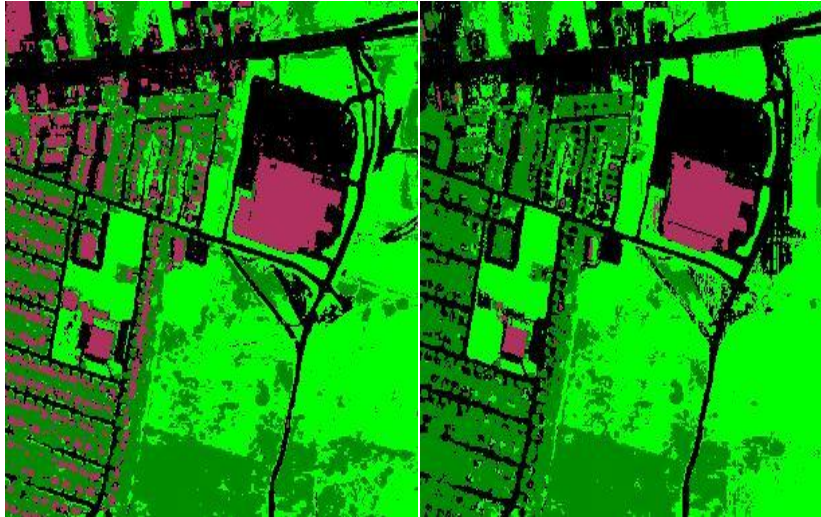


Figure 8: Classified hyperspectral images for AP-MLC and MLC.

Table 6: Accuracies and kappa statistics for CAP-MLC, AP-MLC, PCA-MLC and MLC.

Accuracy (%)	CAP-MLC				AP-MLC				PCA-MLC				MLC			
	PA	OE	UA	CE	PA	OE	UA	CE	PA	OE	UA	CE	PA	OE	UA	CE
Asphalt	97.5	2.5	99.7	0.4	95.6	4.4	99.3	0.7	97	3	99.9	0.1	100	0	99.9	0.1
Tree	99.4	0.6	85.1	14.9	98.6	1.4	86.7	13.3	98.9	1.1	89	11	0	0	0	100
Grass	92.6	7.4	98.6	1.5	92.1	7.9	96.5	3.5	93.9	6.1	96.9	3.1	0.5	99.5	100	0
Roof	97.6	2.4	98.1	1.9	97.1	2.9	89.9	10.1	99.1	0.9	95.7	4.3	0	100	0	0
OA	96.47				95.26				96.69				81.50			
KC	0.95				0.93				0.95				0.45			

Table 7: Summary of overall accuracies and kappa coefficients for CAP-MLC, AP-MLC, PCA-MLC and MLC.

Method	OA(%)	KC	Bands	Classes
CAP-MLC	96.47	0.95	09	04
AP-MLC	95.26	0.93	09	04
PCA-MLC	96.69	0.95	09	04
MLC	81.50	0.45	162	04

The HYDICE hyperspectral image was used to perform the experiments. To do that, a total of four (04) classes, namely: asphalt, grass, tree and roof were considered. The classification method proposed in this paper is CAP-MLC, and three additional classification methods, namely AP-MLC, PCA-MLC and MLC were used for comparison purposes.

According to the results presented in Table 7, the PCA-MLC method presented the best separation between classes with an overall accuracy of 96.69% and kappa coefficient of 0.95. The classification obtained by using the CAP-MLC method has resulted in an overall accuracy of 96.47% and kappa coefficient of 0.95 as well, being jointly with the PCA-MLC method, the best to separate

the classes. The traditional AP-MLC method was the third method which classified better the image, resulting in an overall accuracy of 95.26% and kappa coefficient of 0.93.

In analogy to the results obtained using the AVIRIS image, the traditional MLC was the method with the lowest accuracy to separate the classes, having an overall accuracy of only 81.50% and kappa coefficient of 0.45. According to Scepan (1999), this accuracy value is below 85% which is the minimum acceptable, revealing the poor performance of this classifier for urban environments.

From Table 7 it can be observed that the accuracy values obtained by employing the four methods explain that 96.47%, 95.26%, 96.69% and 81.50% of the pixels in the image were correctly classified, and respectively 3.53%, 4.74%, 3.31%, and 18.5% of the pixels were miss classified.

The MLC method was applied with the original image, considering 162 bands remaining after the pre-processing stage. The method classified only the major features of the 4 classes, contrarily with CAP-MLC, AP-MLC and PCA-MLC that presented detailed classification for all the features in all classes as shown in Figures 7 and 8.

According to Landgrebe (2003) and Camps-Valls *et al.* (2014), this limitation of MLC in classifying some of the classes can be attributed to the fact that the non-classified features present a small number of training samples in a feature space of high dimensionality.

According to the results in Table 6, the method CAP-MLC resulted in a producer's accuracy of 97.54% for asphalt, 99.38% for tree, 92.60% for grass and 97.56% for roof. It can also be observed that the errors committed to classify the classes asphalt, grass and roof were below 2%, and that only the class tree has experienced a commission error of approximately 15%.

During the classification process, the class asphalt was omitted in 2.46%, tree was omitted in 0.62% and allocated by Commission in 14.94%, grass omitted in 7.40% and 2.44% for roof. Within the classes with errors in producer's accuracy, the classes tree and roof had the lowest accuracy when using the MLC method, being 0% for both. In other methods such as CAP-MLC, AP-MLC and PCA-MLC, the same classes obtained accuracies above 90%. The best classification accuracies were obtained by PCA-MLC, CAP-MLC and AP-MLC methods, confirming the theoretical evidence that these methods as the discriminative are not limited to the size of training samples when performing classification in high-dimensionality data (Peng *et al.*, 2016).

The producer accuracies obtained by the PCA-MLC method were 97.04%, 98.94%, 93.92% and 99.07% respectively for the classes asphalt, tree, grass and roof. The four classified classes were omitted respectively in 2.96%, 1.06%, 6.08% and 0.93%. In this method, the commission errors were 0.13%, 11.03%, 3.14% and 4.26% respectively for the classes asphalt, tree, grass and roof. The MLC method, classified correctly only the class asphalt, while the remaining 3 classes (tree, grass and roof) were classified with errors. During the Classification, the classes grass and roof were omitted in 99.54% and 100% respectively. This process also observed commission errors, where they were wrongly allocated classes for asphalt at 0.10% and tree at 100%.

The kappa coefficient values for the experiments with CAP-MLC, AP-MLC, PCA-MLC and MLC were respectively 0.95, 0.93, 0.95 and 0.45. According to Cohen (1960) and Landis and Koch (1977), these values suggest that the first three methods (CAP-MLC, AP-MLC and PCA-MLC) with Kappa coefficient above 0.9 present an almost perfect agreement between classes and good accuracy. The last method (MLC) had kappa coefficient equal to 0.45, the reason why it presents a weak agreement between classes.

5. Conclusion

This paper proposes a method named CAP-MLC which integrates Class-based AP and MLC to improve the MLC accuracy even in cases where limited number of training samples is used to classify hyperspectral images. After discussion of results, the following conclusions were made:

Compared with PCA-MLC, AP-MLC and MLC, using AVIRIS and HYDICE hyperspectral data, the proposed CAP-MLC method presented the best classification with an overall accuracy of 94.15% and kappa coefficient of 0.94. The MLC method obtained the lowest separation between classes, with overall accuracies of only 85.42% and 81.50% and kappa coefficients of 0.83 and 0.45, respectively, for the AVIRIS and HYDICE images. Considering the second image, the PCA-MLC method presented a slight improvement in the classification accuracy (96.69%) if compared with the proposed CAP-MLC method, with 96.47% of overall accuracy, and both with a kappa coefficient of 0.95.

Considering that the KC values evaluate the homogeneity of the samples among themselves and, that they have been taken from the image, between the samples and the rest of the image, the use of old images, lacking of ground truth causes the homogeneity of the samples on the image, limiting the classification accuracy, and in some cases the generalization of the methods.

The CAP-MLC method proposed in this research has proven to be very efficient for classification of land covers on hyperspectral images. The accuracies obtained from the proposed method were at a level above the recommended for use in urban and rural applications.

The proposed CAP-MLC method improved the MLC classification accuracy in 8.73% and 14.97% respectively for AVIRIS and HYDICE image. The accuracies obtained by using this method were higher than those of other methods, and the classification was detailed even for classes with small training samples, enabling all the cover types in the images to be correctly classified regardless of the training sample size.

Despite the fact of reducing the image dimensionality, the proposed approach has shown the potential to remove redundant information between bands and keep only relevant spectral information. Thus, the hyperspectral image classification performed, resulted in a high accuracy for both AVIRIS and HYDICE image, and outperformed the other three methods evaluated, indicating that the proposed method is promising for classification of hyperspectral data sets.

The main contribution of the proposed CAP-MLC method is that it improved the Maximum Likelihood classified hyperspectral image accuracy even in cases of limited training samples and produces the clusters basing on the classes collected.

Acknowledgements

The authors would like to thank the Purdue University for the provision of AVIRIS and HYDICE hyperspectral data sets. They would also like to express their gratefulness to Professor Jorge Silva Centeno from Federal University of Paraná (UFPR) for the valuable support in this research.

References

- Ablin, R; Sulochana, C.H. A Survey of Hyperspectral Image Classification in Remote Sensing. *International Journal of Advanced Research in Computer and Communication Engineering*, v. 2, n. 8, 2013.
- Bartels, M; Wei, H. Rule-based Improvement of Maximum Likelihood Classified LIDAR Data fused with coregisterd band. *Computational Vision Group*, School of Systems Engineering the University of Reading, 2006.
- Baumgardner, M; Biehl, L; Landgrebe, D. 220 band AVIRIS hyperspectral image data set: June 12, 1992 indian pine test site 3. *Purdue University Research Repository* vol. 10, p. R7RX991C, 2015.
- Brzank, A; Heipke, C. Classification of LIDAR into Water and Land in Coastal Areas. Institute of Photogrammetry and Geoinformation, 2007.
- Camps-Valls, G; Tuia, D; Bruzzone, L; Benediktsson, J.A. Advances in hyperspectral image classification. *IEEE Signal Processing. Mag*, vol. 31, p. 45-54, 2014.
- Cohen, J. A. Coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, vol. 20, p. 37-46, 1969.
- Cortes, V; Vapnik, V. Support vector networks. *Machine Learning*, vol. 20, p. 273-297, 1995.
- Crósta, A. P. *Processamento Digital de Imagens de Sensoriamento Remoto*. Campinas, P, Unicamp, ed. rev, 1993.
- Dueck, D. *Affinity Propagation: clustering data by passing messages*. University of Toronto, September 24, 2009.
- Frey, B. J; Dueck, D. Clustering by passing messages between data points. *Science*, vol. 315, n. 5814, p. 972-976, 2007.
- Fukunaga, K. *Introduction to Statistical Pattern Recognition*. Second Edition. Academic Press, 1990.
- Hoffbeck, J. P; Landgrebe, D.A. Covariance Matrix Estimation and Classification with Limited Training Data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, n. 7, 1996.
- Jimenez, L. O; Landgrebe, D.A. Hyperspectral Data Analysis and Supervised Feature Reduction Via Projection Pursuit. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, n. 6, p. 2653-2667, 1999.
- Landgrebe, D. A. Hyperspectral image data analysis. *IEEE Signal Process. Mag*, vol. 19, n. 1, p. 17-28, 2002.
- Landgrebe, D. A. *Signal Theory Methods in Multispectral Remote Sensing*. John Wiley & Sons: New York, NY, USA, p. 503, 2003.
- Landis, J. R; Koch, G.G. The measurement of observer agreement for category data. *Biometrics*, vol. 33, n. 1, p. 159-174, 1977.
- Lee, C; Landgrebe, D. A. Analyzing high-dimensional multispectral data. *IEEE Transactions on Geoscience and Remote Sensing*, v. 31, p. 792 - 800, 1993.

- Maselli, F; Conese, C; Petkov, L; Resti, R. Inclusion of prior probabilities derived from a nonparametric process into the maximum likelihood classifier. *Photogrammetric Engineering and Remote Sensing*, vol. 58, p. 201-207, 1992.
- Mather, P. M. Support Vector classifiers for Land Cover Classification. *Map India 2003 Image Processing & Interpretation*, 2003.
- Peng, Y; Pan, Z; Zheng, Z; Li, X. Hyperspectral Image Classification by Fusion of Multiple Classifiers. *International Journal of Database and Theory and Application*, vol. 9, n. 2, 2016.
- Ramakrishnan, D; Bharti, R. Hyperspectral remote sensing and geological applications. *Current Science*, vol. 108, n. 5, 2015.
- Refianti, R; Mutiara, A.B; Syamsudduha, A.A. Performance Evaluation of Affinity Propagation Approaches on Data Clustering. *International Journal of Advanced Computer Science and Applications*, vol. 7, n. 3, 2016.
- Richards, J. A. *Remote Sensing Digital Image Analysis*. Springer-Verlag: Berlin, Germany, p. 240, 1999.
- Roessner, S; Segl, K; Heiden, U; Kaufmann, H. Automated differentiation of urban surfaces based on airborne hyperspectral imagery. *IEEE Transactions on Geosciences and Remote Sensing*, vol. 39, n. 7, p. 1525-1532, 2001.
- Scepan, J. Thematic validation of high-resolution global land-cover data sets. *Photogrammetric Engineering and Remote Sensing*, vol. 65, p. 1051-1060, 1999.
- Segl, K; Roessner, S; Heiden, U; Kaufmann, H. Fusion of spectral and shape features for identification of urban surface cover types using reflective and thermal hyperspectral data. *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 58, n. 1-2, p. 99-112, 2003.
- Serpico, S. B; Bruzzone, L. A New Search Algorithm for Feature Selection in Hyperspectral Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, n. 7, July 2001.
- Story, M; Congalton, R. Accuracy assessment: a user's perspective. *Photogrammetric Engineering and Remote Sensing*, vol. 52, p. 397-399, 1986.