

# A comparative study of multiscale representations for spatial-spectral classification of hyperspectral imagery

Maria C. Torres-Madronero

*Facultad de Ingenierías, Instituto Tecnológico Metropolitano, Medellín, Colombia. mariatorres@itm.edu.co*

Received: May 17<sup>th</sup>, 2015. Received in revised form: August 10<sup>th</sup>, 2016. Accepted: November 08<sup>th</sup>, 2016

## Abstract

Hyperspectral remote sensors acquire data coming from hundreds of narrow bands through the electromagnetic spectrum; this allows the terrestrial and maritime surfaces to be characterized for Earth observation. Hyperspectral image processing requires algorithms that combine spatial and spectral information. One way to take full advantage of spatial-spectral data within hyperspectral imagery is to use multiscale representations. A multiscale representation generates a family of images where fine details are systematically removed. This paper compares two multiscale representation approaches in order to improve the classification of hyperspectral imagery. The first approach is based on nonlinear diffusion, which obtains a multiscale representation by successive filtering. The second is based on binary partition tree, an approach inspired in region growing. The comparison is performed using a real hyperspectral image and a support vector machine classifier. Both representation approaches allowed the classification of hyperspectral imagery to be improved. However, nonlinear diffusion results surpassed those obtained using binary partition tree.

*Keywords:* nonlinear diffusion; binary partition tree; classification; hyperspectral imagery; multiscale representation; remote sensing.

# Estudio comparativo de representaciones multiescala para clasificación de imágenes hiperespectrales

## Resumen

Los sensores remotos hiperespectrales adquieren datos a lo largo de cientos de bandas estrechas a través del espectro electromagnético permitiendo la caracterización de las superficies terrestres y marítimas para la observación de la Tierra. El procesamiento de imágenes hiperespectrales requiere de algoritmos que combinen la información espacial y espectral. Una forma de tomar ventaja de los datos espaciales y espectrales en las imágenes hiperespectrales es usar representaciones multiescala. Una representación multiescala genera una familia de imágenes donde los detalles finos son sistemáticamente removidos. Este artículo compara dos enfoques de representación multiescala para mejorar la clasificación de imágenes hiperespectrales. El primer enfoque se basa en difusión no lineal la cual obtiene la representación multiescala por medio de sucesivos filtros. El segundo se basa en un árbol de partición binaria, un enfoque inspirado en crecimiento de regiones. La comparación se realiza usando la imagen hiperespectral Indian Pines y un clasificador de máquinas de soporte vectorial. Ambos enfoques de representación permiten mejorar la clasificación de la imagen hiperespectral. Pero, los resultados de difusión no lineal superan los obtenidos usando el árbol de partición binaria.

*Palabras clave:* difusión no lineal; árbol de partición binaria; clasificación; imagen hiperespectral; representación multiescala; percepción remota.

## 1. Introduction

Remote sensors allow the atmosphere, the terrestrial surface, and the oceans to be characterized for Earth observation. The number of remote sensing applications has increased over the years, and this has encouraged the

development of diverse systems and sensors. One example of remote sensing systems is the passive sensors that collect both spatial and spectral information. Multispectral and hyperspectral sensors measure the emitted and reflected radiance of a surface along the electromagnetic spectrum [1]. Multispectral sensors have a relatively small number of non-

**How to cite:** Torres-Madronero, M.C., A comparative study of multiscale representations for spatial-spectral classification of hyperspectral imagery. DYNA 84(200), pp. 129-134, 2017.

contiguous bands (i.e. less than 25 bands). Furthermore, hyperspectral sensors capture spectral information along hundreds of narrow bands. Spatial and spectral information from hyperspectral and multispectral sensors can be seen as a cube, in which each pixel is a spectral signature that characterizes the components on the surface [1].

The increasing availability of high and medium resolution remote sensors has opened new possibilities for Earth exploration, and, at the same time, image processing researchers face new challenges. One such challenge is the analysis and processing of remote sensing imagery that integrates both spatial and spectral information [2-3]. It is expected that spatial-spectral approaches improve the classification results and the estimation of sub-pixel information. Image classification identifies the classes in a dataset and assigns label to each pixel [1]. The integration of spatial information in the analysis of remote sensing data can be performed in several ways. For instance, Jimenez et al. [4] used the information within the neighborhood of a pixel to refine the classification results. The most frequent class in the neighborhood determines the final label of a pixel. Another way to include spatial information is by using segmentation techniques [5-7]. Segmentation divides the image into spectrally uniform regions, but it requires either a similarity threshold or a priori knowledge about the number of regions.

This paper explores multiscale representation to improve hyperspectral imagery classification. A multiscale representation builds a family of images, systematically removes fine details [8]. Unlike segmentation and kernel techniques, multiscale representation provides a way to explore the objects and regions within hyperspectral images in a natural and automatic way [9-11]. Different approaches are used to build a multiscale representation. For instance, nonlinear diffusion [10,11] obtains a multiscale representation by successively filtering. On other hand, binary partition tree [12-14] builds a multiscale representation by region growing. This paper aims to determine which of these approaches provides the best representation of hyperspectral imagery for classification purposes. Both representation approaches generate a family of images; thus, a scale selection criteria needs to be used. This study shows that both multiscale representations improve the classification of hyperspectral imagery, and one selection criterion is recommended in order to determine the suitable scale for the classification analysis.

The next section describes both the nonlinear diffusion and binary partition tree representations. Four scale selection criteria are subsequently presented. Section 4 shows the experiments that use a real hyperspectral image. The final section contains the conclusions and future work.

## 2. Multiscale Representation

Two multiscale representation approaches are explored for the classification of hyperspectral imagery. First, nonlinear diffusion is presented, and then binary partition tree is presented. Both approaches seek to deal with the scale notion, however, the first address the problem from a filtering perspective and the second from a region growing perspective.

### 2.1. Nonlinear diffusion

A multiscale representation of a hyperspectral image  $\mathbf{Y} = [y_1, \dots, y_m]^T$  with  $n$  spectral bands and  $m$  pixels is obtained by systematically solving the partial differential equation – PDE, given by:

$$\frac{\partial \mathbf{y}_i}{\partial t} = \nabla[g(\theta)\nabla\mathbf{y}_\sigma] \quad i = 1, \dots, m \quad (1)$$

where  $\mathbf{y}_\sigma$  is the result of a Gaussian filter with zero mean and standard deviation  $\sigma$ , which are applied to the image  $\mathbf{Y}$ ;  $\theta$  is an edge measure given by:

$$\theta = \sqrt{\frac{1}{n} \sum_{j=1}^n |\nabla\mathbf{y}_{\sigma j}|^2} \quad (2)$$

and  $g$  is the diffusion coefficient proposed by Weickert et al. [15]:

$$g(\theta) = \begin{cases} 1 & \theta = 0 \\ 1 - \exp\left(-\frac{3.31488}{(\theta/\alpha)^8}\right) & \theta > 0 \end{cases} \quad (3)$$

The  $\alpha$  parameter is a threshold to control the diffusion [9]. Duarte et al. [9] showed that semi-implicit schemes can solve the nonlinear diffusion faster than by using explicit discretization. The semi-implicit discretization of the nonlinear diffusion PDE is expressed in matrix form as:

$$g(\mathbf{I} - \mu\mathbf{G}_t)\mathbf{Y}_{t+1} = \mathbf{Y}_t \quad (4)$$

where  $\mu$  is the scale step, i.e.  $\mu \equiv \Delta t$ ,  $\mathbf{Y}_t$  is the hyperspectral image and  $\mathbf{G} = [g_{ij}]$  is the diffusion coefficients matrix. An algorithm, to solve the hyperspectral image's nonlinear diffusion using multigrid methods (AMG) was proposed by Duarte et al., [10] and it is used in this paper.

### 2.2. Binary Partition Tree

Binary Partition Tree -BPT is an image representation based on regions. BPT is mainly used for image segmentation [12], and was introduced for hyperspectral image processing in [13, 14]. Each of the tree's levels is a partition of the image into a region. A region adjacency graph,  $G = (V, E)$ , is used to represent each level. BPT is built by successively merging region operations in the region adjacency graph. The first level corresponds to an initial partition obtained by a region growing algorithm [12]. BPT requires that a merging order, a merging criterion, and a region model are all selected [12]. In this paper, a first order statistic is used as the region model, i.e. each region is represented by the spectra mean of its constituent pixels:

$$V_{ik} = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{ik} \quad j = 1, \dots, n \quad i \in R_i \quad (5)$$

where  $V$  are the graph vertices,  $R_i$  is a region,  $m_i$  is the

number of pixels in  $R_i$ ,  $y_i$  is a pixel in  $R_i$ , and  $n$  is the number of bands. The merging criteria are established by using a similarity measure: the cosine distance. The merging order is given by the most similar regions, i.e. the regions with the minimum distance. Furthermore, building a BPT level requires a three-step process: computing the cosine distance between regions, merging the two most similar regions, and updating the region model [12, 13].

### 3. Scale selection

Multiscale representations build a family of images,  $\{Y_1, Y_2, \dots, Y_T\}$ , where  $Y_i$  is a scale. For nonlinear diffusion,  $Y_i$  a smoothed image obtained by solving eq. (1). For binary partition tree,  $Y_i$  is the image build from the RAG in level  $i$ . Hyperspectral image classification requires one single scale. For this reason, scale selection approaches are explored in this paper. These criteria are usually used for filtering of gray scale images. This work presents the generalizations of scale selection methods for vector-value images.

#### 3.1. Minimal Entropy Change

Sporring and Weickert [16] used the scale that achieves minimal entropy change. The minimal entropy change is an indicator of stable scales. The entropy for a gray-scale image is given by:

$$S = - \sum_{i=1}^K p_i \log p_i \quad (6)$$

where  $p_i$  is the histogram count for the intensities values ranging from 1 to  $K$ . The optimal scale is found by minimizing the entropy change  $\Delta S$ . Extending this criterion to a vector-value image is undertaken by using the average entropy across bands.

#### 3.2. Decorrelation Criterion

The decorrelation criterion proposed by Mrazek and Navarra [17] used the scale for which the minimum correlation between the smoothed image in the time  $t$ ,  $Y_t$ , and the differential image between  $Y_t$ , and the original image  $Y$ ,  $Y_t - Y$ , is found. It is expected that this correlation will decrease as more iterations are performed. When the smoothing operations remove significant details from the image, the correlation begins to increase. The scale for which the correlation is minimum guarantees that important details are not removed. A generalization of the decorrelation criterion for vector-value images is obtained by averaging the correlation in each band:

$$\rho(Y_t - Y, Y_t) = \frac{1}{n} \sum_{i=1}^n \rho(Y_t(i) - Y(i), Y_t(i)) \quad (7)$$

where  $Y_t(i)$  is the band  $i$  in the scale  $t$  and  $\rho$  is the coefficient of correlation.

#### 3.3. Diffusion Balance Criterion

Jiabin and Guizhong [18] proposed using the scale for which the multiscale representation achieves a balance:

$$\rho|\Delta E_{Y_t}| = |\Delta E_{D_t}| \quad (8)$$

where  $E_{Y_t}$  is the energy of the smoothed image at  $t$ , and  $E_{D_t}$  is the energy of the differential image  $D_t = Y_t - Y$  at the scale  $t$ . The extension of the diffusion balance criterion to vector-value images is performed by using the Frobenious norm as a measure of energy.

#### 3.4. Entropy Change

The entropy can be seen as a measure of information. As more smoothing iterations are performed, more information is removed from the image. Thus, the entropy change of  $D_t = Y_t - Y$  is significant in the first iterations. The entropy change criterion proposed in [11] selects the scale where the entropy change achieved a break point. The perceptual difference between  $Y_t$  and  $Y$  is used to generalize the criterion for any image:

$$\bar{D}_t(i, j, k) = \frac{|Y_t(i, j, k) - Y(i, j, k)|}{Y(i, j, k)} \quad (9)$$

## 4. Experiments

### 4.1. Dataset

The comparison of multiscale representation approaches with scale selection criteria for classification purposes was performed using the AVIRIS image, which was collected over the West Lafayette (USA) agriculture region. This hyperspectral image, named Indian Pines, has 145x145 pixels and 220 spectral bands from 400 to 2500 nm. There is a classification map with 16 classes of Indian Pines available online (<https://engineering.purdue.edu/~biehl/MultiSpec/>). Fig. 1 (a) shows band 42 of the Indian Pines. Only 197 spectral bands are used. Water absorption bands, 107 to 114 (1353 nm to 1422 nm) and 153 to 167 (1811 nm to 1939 nm) were removed.

### 4.2. Multiscale Representations and Classification

For India Pines, 20 smoothing iterations were performed using nonlinear diffusion in order to obtain a multiscale representation. The diffusion coefficient  $\alpha$  was set equal to 0.01, and the scale step  $\mu$  was set equal to 5. For the binary partition tree, an initial partition was performed with a region growing algorithm. Each level of the binary partition tree is taken as a scale. Fig. 1 (b,c) shows the 42nd band for the smoothed image with nonlinear diffusion in the third iteration and the image in the 120th level in BPT.

First, the original image and the family of images from the nonlinear diffusion and binary partition tree representations were processed using a super vector machine classifier with a second order polynomial kernel. The classification was performed using

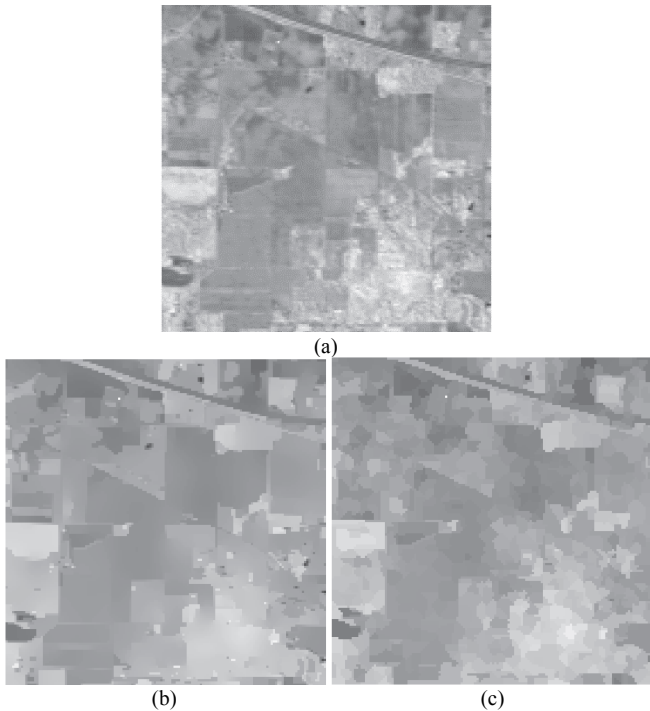


Figure 1. Spectral band 42 for (a) Original Image, (b) fourth iteration of nonlinear diffusion, and (c) 120th level of BPT. Source: The authors.

Table 1. Classification Accuracy for Indian pines

Original Image			
Accuracy (%)		80.5	
Nonlinear Diffusion		Binary Partition Tree	
Iteration	Accuracy (%)	Iteration	Accuracy (%)
1	96.6	20	94.3
2	97.8	40	94.3
3	<b>98.0</b>	60	94.4
4	<b>98.0</b>	80	94.4
5	97.8	100	94.6
6	97.7	<b>120</b>	<b>95.1</b>
7	97.4	140	94.9
8	97.2	160	94.8
9	97.2	180	94.8
10	97.4	200	94.8
11	97.4	220	94.7
12	97.1	240	94.8

Source: The authors.

10% of pixels as training samples. Table 1 shows the classification accuracy that was obtained for 12 images, and Fig. 2 presents the classification accuracy for all images in the multiscale representation that was obtained by (a) nonlinear diffusion and by using a (b) binary partition tree.

Table 2 shows the classification accuracy per class obtained from the fourth iteration of nonlinear diffusion and the 120th level of binary partition tree.

### 4.3. Comparison Scale Selection Criteria

The scale selection criteria were compared according the classification accuracy obtained from the smoothed image

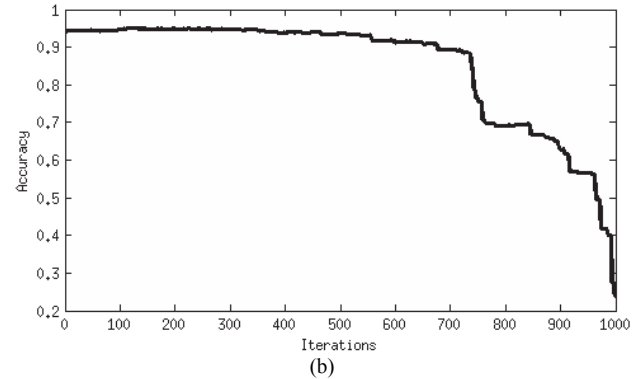
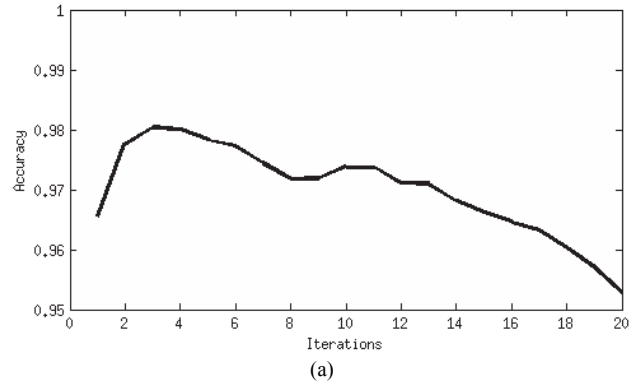


Figure 2. Classification accuracy for (a) nonlinear diffusion, and (b) binary partition tree. Source: The authors.

Table 2. Classification Accuracy per class for Indian pines

	Original Image	Nonlinear Diffusion	Binary Partition Tree
1. Alfalfa	48.15	<b>98.15</b>	74.07
2. Buildings, grass, tress, drives	67.37	<b>95.53</b>	95.00
3. Corn	53.42	<b>94.02</b>	88.46
4. Corn min	68.94	<b>97.12</b>	95.56
5. Corn notill	77.89	<b>98.47</b>	95.19
6. Grass pasture	86.12	95.57	<b>98.19</b>
7. Grass pasture mowed	57.69	<b>100</b>	0
8. Grass trees	97.46	<b>97.99</b>	95.85
9. Hay windrowed	98.57	99.39	<b>99.59</b>
10. Oats	70.00	<b>100</b>	<b>100</b>
11. Soybean clean	76.71	<b>97.23</b>	92.67
12. Soybean mintill	78.24	<b>98.14</b>	96.23
13. Soybean notill	66.01	<b>91.63</b>	89.88
14. Stone steel towers	85.26	91.58	<b>96.84</b>
15. Wheat	<b>100</b>	99.53	<b>100</b>
16. Woods	96.29	<b>99.00</b>	98.61

Source: The authors.

Table 3. Selected Scale and their Classification Accuracy for Nonlinear Diffusion Representation.

	DE	DBC	MEC	EC
Selected Scale	9	12	<b>10</b>	8
Accuracy (%)	97.2	97.1	<b>97.4</b>	97.2

Source: The authors.

selected as the optimal scale. Table 3 shows the selected scale for the decorrelation criterion (DE), the minimal entropy change (MEC), the diffusion balance criterion (DBC), and the entropy change (EC).

For the binary partition tree representation, only the decorrelation criterion proposed by [17] obtained a unique scale. This criterion selected the 100th iteration in the BPT where a classification accuracy of 94.6% is achieved. The other methods did not provide a unique solution.

#### 4.4. Discussion

Comparing the classification results obtained using the super vector machine from the original image and the images from the multiscale representations (Table 1), it can be seen that the accuracy increases with the latter. The best accuracy can be found in all scales from the nonlinear diffusion representation (Fig. 2(a)), and in the first 700 iterations in the binary partition tree representation (Fig. 2(b)). Scale three and four is where the best accuracy was achieved for nonlinear diffusion representation (98%). For the binary partition tree representation, the best accuracy was achieved in the 120th iteration (95.1%). Table 2 shows the accuracy improvement using the nonlinear diffusion representation. Although the binary partition tree has a better overall performance than the original image, one class is not detected by this method: BPT loses grass pasture mowed.

When scale selection criteria were compared, it was discovered that the criteria found in the literature work as expected for the nonlinear diffusion representation but not for the binary partition tree. The selected scales were different for all criteria for the nonlinear diffusion representation, as can be seen in Table 3. For the binary partition tree, only the decorrelation criterion obtained a unique solution. It is important to note that the criteria compared in this work not used labeled data, i.e. they do not require training data. All the criteria only used the information in the original image and the image family generated by the multiscale representation.

#### 5. Conclusions

A comparison between two multiscale representation approaches was performed. Nonlinear diffusion and binary partition tree were employed to generate a family of images. The first approach was based on nonlinear filtering, in which fine details are smoothed by successively solving a partial differential equation. The second approach was inspired by region growing, and fine details are removed by a successive merging operation. A real hyperspectral image was used to compare both approaches for classification. Both nonlinear diffusion and binary partition tree representation achieved better classification performance than when the original image was used. However, nonlinear diffusion representation obtained the best overall accuracy classification as well as most of the best per class accuracy for Indian Pines.

Since the multiscale representation obtained a family of images, scale selection criteria are necessary to select an image for processing. A decorrelation criterion was enabled in order to select a unique scale for both representations. However, these selected scales did not contain the scale that

obtained the best accuracy. Note that the decorrelation criterion and the other criteria do not use training data, they only used the information in the original image and the image containing the multiscale representation.

#### Acknowledgements

This project was developed as part of the P14125 research project in the Instituto Tecnológico Metropolitano, which was undertaken by the research group on Automatic, Electronic and Computational Science.

#### References

- [1] Richards, J.A. and Xia, J., Remote sensing digital image analysis: An introduction. 5th edition. Springer, 2013.
- [2] Plaza, A., Benediktsson, J., Boardman, J., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., Marconcini, M., Tilton, J. and Trianni, G., Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment*, 113, pp. s110–s122, 2009. DOI: 10.1016/j.rse.2007.07.028
- [3] Blaschke, T., Object based image analysis for remote sensing. *Journal of Photogrammetry and Remote Sensing*, 65, pp. 2-16, 2010. DOI: 10.1016/j.isprsjprs.2009.06.004
- [4] Jimenez, L., Rivera-Medina, J., Rodriguez-Diaz, E., Arzuaga-Cruz, E. and Ramirez-Velez, M., Integration of spatial and spectral information by means of unsupervised extraction and classification for homogeneous objects applied to multispectral and hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4), pp. 844-851, 2005. DOI: 10.1109/TGRS.2004.843193
- [5] Tilton, J.C., Image segmentation by region growing and spectral clustering with a natural convergence criterion. *IEEE Geoscience & Remote Sensing Symposium*, 1998, pp. 1766-1768.
- [6] Tarabalka, Y., Benediktsson, J.A., Chanussot, J. and Tilton, J.C., Multiple spectral-spatial classification approach for hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11), pp. 4122-4132, 2010. DOI: 10.1109/TGRS.2010.2062526
- [7] Li, J., Bioucas-Dias, J.M. and Plaza, A., Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 50(3), pp. 809-823, 2012. DOI: 10.1109/TGRS.2011.2162649
- [8] Lindeberg, T., Scale-space theory in computer vision. Kluwer Academic Publisher, 1994.
- [9] Duarte, J.M., Castillo, P. and Velez-Reyes, M., Comparative study of semiimplicit schemes for nonlinear diffusion in hyperspectral imagery. *IEEE Transactions on Image Processing*, 16(5), pp. 1303-1314, 2007. DOI: 10.1109/TIP.2007.894266
- [10] Duarte, J.M., Sapiro, G., Velez, M. and Castilo, P., Multiscale representation and segmentation of hyperspectral imagery using geometric partial differential equations and algebraic multigrid methods. *IEEE Transactions on Geosciences and Remote Sensing*, 46(8), pp. 1303-1314, 2008. DOI: 10.1109/TGRS.2008.916478
- [11] Torres-Madronero, M.C. and Velez-Reyes, M., Integrating spatial information in unsupervised unmixing of hyperspectral imagery using multiscale representation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6), pp. 1985-1993, 2014. DOI: 10.1109/JSTARS.2014.2319261
- [12] Salembier, P. and Garrido, L., Binary partition tree as an efficient representation for image processing, segmentation and information retrieval. *IEEE Transactions on Image Processing*, 9(4), pp. 561-576, 2000. DOI: 10.1109/83.841934
- [13] Valero, S., Salembier, P. and Chanussot, J., Hyperspectral image representation and processing with binary partition trees. *IEEE Transactions on image processing*, 22(4), pp. 1430-1443, 2013. DOI: 10.1109/TIP.2012.2231687
- [14] Drumetz, L., Veganzones, M.A., Marrero, R., Tochon, G., Dalla, M., Plaza, A. and Chanussot, J., Binary partition tree-based local spectral

- unmixing. International IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing 2 WHISPERS. 2014.
- [15] Weickert, J., Romenu, B. and Viergever, M., Efficient and reliable schemes for nonlinear diffusion filtering. IEEE Transactions on Image Processing, 7(3), pp. 398-410, 1998. DOI: 10.1109/83.661190
- [16] Sporring, J. and Weickert, J., Information measures in scale spaces. IEEE Transactions on Information Theory, 45, pp. 1051-1058, 1999. DOI: 10.1109/18.761342
- [17] Mrazek, P. and Navarra, M., Selection of optimal stopping time for nonlinear diffusion filtering. International Journal of Computer Vision, 52, pp. 189-203, 2003.
- [18] Jiabin, Y. and Guizhong, L., A novel anisotropic diffusion time estimation method. Congress on Image and Signal Processing, 2008. pp. 241-244.

**M.C. Torres-Madronero**, was awarded her BSc. degree in Electronic Engineering from the Universidad Nacional de Colombia, Manizales, Colombia, in 2006. She gained her MSc. degree in Electrical Engineering and PhD. in Computing and Information Science and Engineering from the University of Puerto Rico, Mayaguez (UPRM), in 2008 and 2013, respectively. She is a faculty member of the Instituto Tecnológico Metropolitano, Medellín, Colombia. Her research interests include hyperspectral data, image processing, unmixing analysis, pattern recognition, and machine learning.  
ORCID: <http://orcid.org/0000-0002-9795-2459>



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN  
FACULTAD DE MINAS

Área Curricular de Ingeniería  
Eléctrica e Ingeniería de Control

Oferta de Posgrados

Maestría en Ingeniería - Ingeniería Eléctrica

Mayor información:

E-mail: [ingelcontro\\_med@unal.edu.co](mailto:ingelcontro_med@unal.edu.co)  
Teléfono: (57-4) 425 52 64