

Journal of Electronic Imaging

JElectronicImaging.org

Stable and unsupervised fuzzy C-means method and its validation in the context of multicomponent images

Kacem Chehdi
Akar Taher
Claude Cariou

Stable and unsupervised fuzzy C-means method and its validation in the context of multicomponent images

Kacem Chehdi,* Akar Taher, and Claude Cariou

University of Rennes 1, IETR/TSI2M, Engineering School of Applied Sciences and Technology, 6 rue de Kérampont, Lannion 22300, France

Abstract. A stable and unsupervised version of the fuzzy C-means algorithm, named FCM-optimized (FCMO), is presented. The originality of the proposed algorithm stems from (1) the introduction of an adaptive incremental procedure to initialize class centers, which makes the algorithm stable and deterministic; therefore, the classification results do not vary from one run to another and (2) the use of an unsupervised evaluation criterion to estimate the optimal number of classes. The validation of FCMO with regard to stability, reliability in class number estimation, and classification efficiency is shown through experimental results on synthetic monocomponent and real multicomponent images. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JEI.24.6.061117](https://doi.org/10.1117/1.JEI.24.6.061117)]

Keywords: clustering; unsupervised; stability; fuzzy C-means; multicomponent; image.

Paper 15515SSP received Jun. 28, 2015; accepted for publication Nov. 2, 2015; published online Dec. 10, 2015.

1 Introduction

Data classification is an essential step in automated decision making which still remains unsolved. Classification methods can be categorized into three groups: supervised, semisupervised, and unsupervised.

“Supervised methods” need *a priori* knowledge (e.g., training samples) to accomplish classification. However, *a priori* knowledge may not be available in all cases of applications. Maximum likelihood,¹ support vector machines (SVM),² and expectation-maximization³ are some of the most commonly used supervised methods.

“Semisupervised methods” require at least some minimal input from the operator as *a priori* knowledge (e.g., the expected number of classes, some threshold value, or a fixed number of iterations). *K*-means⁴ is an example of a semisupervised algorithm that requires *a priori* knowledge of the number of classes. This algorithm is highly sensitive to the initial randomly selected class centers,⁵ which makes it unstable; hence, it may yield varying results for the same dataset for different runs. The principle of *K*-means is used in the Linde–Buzo–Gray (LBG) algorithm, which, as a result, is also unstable.⁶ The fuzzy C-means (FCM) is another semisupervised algorithm.⁷ It extends the *K*-means algorithm using a fuzzification operation to solve ambiguous classification problems. This algorithm is one of the most widely and successfully used methods. However, it also requires the number of classes to be known *a priori*. Furthermore, it is sensitive to the initial class centers and also to the choice of the fuzzification parameter.⁷

“Unsupervised” classification is a kind of approach that does not need training samples or *a priori* knowledge. One of the most recent unsupervised methods is affinity propagation (AP).⁸ Unfortunately, AP is found to highly overestimate the number of classes, and it is inapplicable to large image

datasets owing to its computational memory complexity. In Ref. 9, an unsupervised version of FCM that is adapted to high-dimensional multiclass pattern recognition problems is presented. This method estimates the number of classes automatically, but it is not completely stable for some complex datasets.

In this paper, we propose an optimized version of the baseline FCM algorithm, FCM-optimized (FCMO), which is stable and completely unsupervised.

The remainder of the paper is organized as follows. In Sec. 2, we present a description of the proposed method, followed by experimental results obtained on monocomponent and multicomponent images in Sec. 3. Finally, in Sec. 4, we present the conclusions.

2 Proposed FCM-Optimized Algorithm

From a general point of view, a robust classification method must check the following three properties:

- nonrequirement of any ground truth (GT) knowledge (training samples),
- ability to automatically estimate the correct number of classes,
- insensitivity to the random choice of the initial class centers (stability).

The FCMO algorithm that we propose meets these three properties.

2.1 Overview of the Proposed Algorithm

Let X be the dataset where each element is characterized by a vector of features and K is the number of classes.

To partition this dataset, the proposed iterative algorithm involves the following four steps:

- Step 1: Choice of the class to be subdivided:
At the beginning ($K = 1$), the class to be divided is the whole dataset, X ; whenever $K > 1$, the most

*Address all correspondence to: Kacem Chehdi, E-mail: kacem.chehdi@univ-rennes1.fr

expanded class is chosen as the candidate class for further splitting, if the unsupervised evaluation criterion in step 4 is true.

- Step 2: Choice of the initial subclasses centers:
The selected candidate class is subdivided into two subclasses; its center of gravity is chosen as the first subclass center, whereas the second subclass center is an element randomly selected within the same class.
- Step 3: Classification with class center fine-tuning:
The dataset is partitioned using standard FCM. To make the approach independent of the initial class centers, a removal-insertion fine-tuning process is used.
- Step 4: Evaluation of the obtained intermediate partitioning using an unsupervised criterion:
If this criterion is satisfied, partitioning into $K + 1$ classes is validated; then go back to step 1. If the criterion is not satisfied, go to step 1 and change the class to be subdivided by choosing the next most expanded class and so on. The algorithm stops if no subdivision satisfies the criterion. In this case, the current number of classes is considered as optimal.

In the following, we present the details of the four aforementioned steps.

2.2 Algorithm Details

We denote $P_K = \{C_1, \dots, C_K\}$ as a partition of X into K classes. We attempt to subdivide one of these classes to obtain $K + 1$ classes.

To choose a candidate class to subdivide, we calculate a dispersion measure for each class C_i , for $i = 1, \dots, K$ as follows:

$$\text{Dispersion}(C_i) = \frac{1}{|C_i|} \sum_{j=1}^{|C_i|} d[g(C_i), C_i^j], \quad (1)$$

where $g(C_i)$ and $|C_i|$, respectively, are the center of gravity and the number of elements of class C_i ; $d(.,.)$ is the Euclidean distance; and C_i^j is the j 'th element of class C_i .

The dispersion values of the classes are then arranged in decreasing order so that the most expanded class is selected as the first candidate to be subdivided. Recall that if the subdivision of the selected class is not validated in step 4, the next most expanded class is selected in turn.

After selecting the class to subdivide into two subclasses, followed by the initialization of the class centers (step 2), the whole dataset is then partitioned into $K + 1$ classes by using FCM with the selected centers. To ensure that the proposed method is independent of the initial class centers, a removal-insertion process to fine-tune the class centers is used, based on the assumption that the dispersion values of the classes are identical.¹⁰ According to this assumption, the empty class and the class with the lowest dispersion (called the "losing" class) are removed directly. The elements of the class adjacent to the losing class (called "neighbor" of losing class) and the elements of the losing class are merged. The center of gravity of this new class (losing and neighbor of losing) is then recalculated. Then a new class center is introduced near the class with the highest local dispersion (called the

"winning" class). This new class center is inserted by randomly selecting an individual from the winning class. The FCM algorithm is then applied on the whole dataset. This process is repeated until the decrease in the dispersion achieves a lower limit.

The final step (step 4) validates or rejects the partitioning obtained with $K + 1$ classes. This step allows estimating the final partition of the dataset X and the associate number of classes.

To validate the partition result P_{K+1} of the dataset X after splitting of the most expanded class, the following condition must be met:

$$D(P_{K+1}) - D(P_K) > \eta D(P_K), \quad (2)$$

where η is a small value which guarantees the stopping of the subdivision algorithm (we have chosen $\eta = 10^{-3}$ in our experiment), and $D(P_K)$ is the criterion defined as¹¹

$$D(P_K) = \frac{1 + \bar{D}(P_K) - \underline{D}(P_K)}{2}, \quad (3)$$

where $\underline{D}(P_K)$ and $\bar{D}(P_K)$ are the global within-class and between-class disparities of partition P_K , respectively.

$\underline{D}(P_K)$ is calculated from the within-class disparity $\underline{D}(C_i)$ of each class in partition P_K

$$\underline{D}(P_K) = \frac{1}{NC} \sum_{i=1}^{NC} \frac{|C_i|}{N} \underline{D}(C_i), \quad (4)$$

where NC is the number of classes of partition P_K and N is the number of elements in the dataset X .

$\bar{D}(P_K)$ is calculated by including the disparity between each class and the other classes of P_K

$$\bar{D}(P_K) = \frac{1}{NC} \sum_{i=1}^{NC} \frac{|C_i|}{N} \bar{D}(C_i). \quad (5)$$

If a partition P_{K+1} is validated by satisfying Eq. (2), the algorithm is repeated in order to subdivide the most expanded class from the ones created thus far. If not, the algorithm is repeated, attempting to subdivide the following class in the sorting order. If none of the classes allow for valid subdivision, the current number of class K is considered as optimal.

The proposed FCMO algorithm can be applied to any dataset X specified in the standard array format (N elements $\times M$ quantitative features).

In the case of image partitioning, two schemes were considered:

1. For a monocomponent image, where each pixel is characterized by a set of M features, the described algorithm is applied directly.
2. For a multicomponent image with B bands, the FCMO algorithm could be applied directly on the set of N pixels, each characterized by $B \times M$ features. In this paper, we have chosen to apply FCMO to each component independently in parallel, thereby generating B partitions. In order to obtain a global partition from these B partitions, a genetic fusion algorithm is used.^{11,12} This strategy is motivated by the algorithm's

ability to solve problems with multiple solutions, and therefore, synthesize relevant information owing to its regularizing capability.

3 Experimental Results

We assessed the proposed unsupervised classification approach FCMO with respect to three criteria: the stability of the classification results as a function of the initial choice of cluster centers, the accuracy of the estimated number of classes, and the correct classification rate.

In order to partition an image, each pixel is characterized by a set of 15 spatial features (for each spectral band) calculated using a window of spatial size 9×9 pixels. These features include the statistical moments of orders 1 to 4, nine co-occurrence features¹³ (contrast, correlation, inverse difference moment, sum average, sum entropy, entropy, first information measure of correlation, second information measure of correlation, and contour information) and two features from the curvilinear integral with two angles.¹¹

3.1 Assessment on Synthetic Monocomponent Images

We first checked the stability of the FCMO algorithm by executing it 100 times on 50 different synthetic monochrome images. Each image in the database is composed of five textured regions extracted from the Brodatz album.¹⁴ Figure 1(a) shows an example of one synthetic image of this database. To check the stability, the number of classes is fixed to $K = 5$ in this experiment for all the compared methods. Table 1 shows the stability performance of FCMO for all the values of the fuzzification parameter m as compared to that of FCM and LBG. These results show that FCM yields identical classification results in only 77% of the runs when $m = 2$ and 94% of the runs when $m = 4$. Moreover, LBG is stable in 73% of the runs.

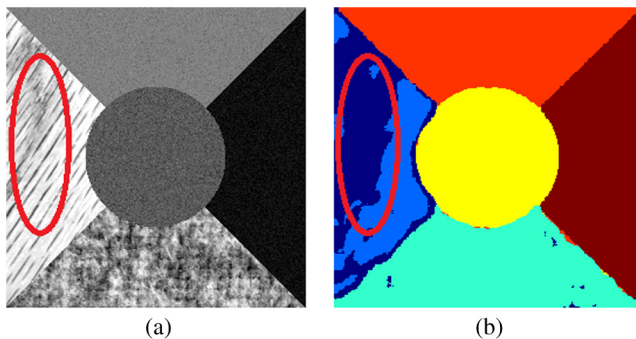


Fig. 1 Example of estimation of the number of classes: (a) original image and (b) classification result (six classes).

Table 1 Stability performance comparison.

FCMO $m = 2, 4, 6, 8$	100%
FCM $m = 2 (m = 4)$	77% (94%)
LBG	73%

With regard to the estimation of the number of classes, we tested the FCMO algorithm on the same image database without specifying any prior knowledge. The average correct class number (ACCN) estimation was 90% over the same 50 images. This rate is coherent because in some images of the database, high fluctuations could occur within a class, which consequently would be detected as distinct classes. For example, in Fig. 1(a), the region on the left of the image displays a wood texture with defect (surrounded area). It is clear that this area rigorously should not be classified in a single class. Indeed, the FCMO algorithm, as expected, divides this class into two subclasses, which is coherent [Fig. 1(b)]. This poses the problem of the reliability of GT data, which does not account for class heterogeneity. Consequently, if the GT was accurate, the true ACCN would be superior to 90%.

The average correct classification rates (ACCR) obtained by FCMO and other methods given in Table 2 confirm the superiority of FCMO. We emphasize that the best overall classification results are given when $m = 4$, which confirms the suggestion proposed by Wu.⁷

Additionally, we compare the performance of FCMO with supervised and semisupervised classification methods: SVM, ISODATA, as well as a parallel cooperation system using SVM with ISODATA algorithms.¹⁵ For SVM, 10% of the GT pixels are used for training, the kernel function used is the Gaussian RBF, and the optimal parameters are chosen by fivefold cross validation. The optimal parameters fixed for the ISODATA algorithm are 4, 10, 2, and 5%, respectively, for the minimum number of classes, the maximum number of classes, the minimum number of pixels in a class, and the change threshold. Table 3 shows two examples of this performance comparison on the images of Fig. 2. Note that \hat{K} denotes the estimated number of classes. For SVM and SVM with ISODATA, the number of classes K is fixed to 5.

Table 2 Average correct classification rates (ACCR) performance comparison.

FCMO $m = 2 (m = 4)$	99.92% (99.97%)
FCM $m = 2 (m = 4)$	87% (94%)
LBG	74%

Table 3 Examples of performance comparison of fuzzy C-means optimized (FCMO) with other methods.

Methods	ACCR (\hat{K} : estimated number of classes)	
	Image 1	Image 2
Support vector machines (SVM)	93.68%	85.97%
ISODATA	74.34% ($\hat{K} = 7$)	74.54% ($\hat{K} = 9$)
SVM with ISODATA	94.27%	87.62%
FCMO	99.47% ($\hat{K} = 5$)	89.80% ($\hat{K} = 5$)

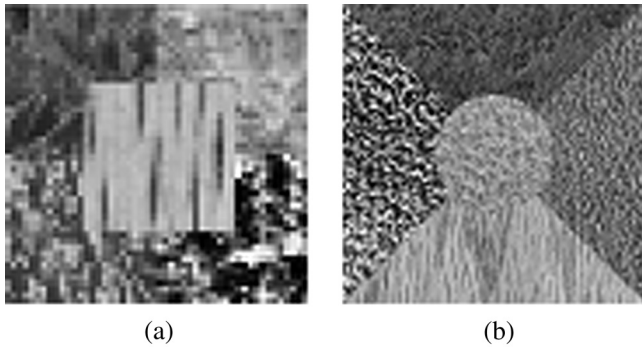


Fig. 2 Examples of images used for performance comparison: (a) original image 1 and (b) original image 2.

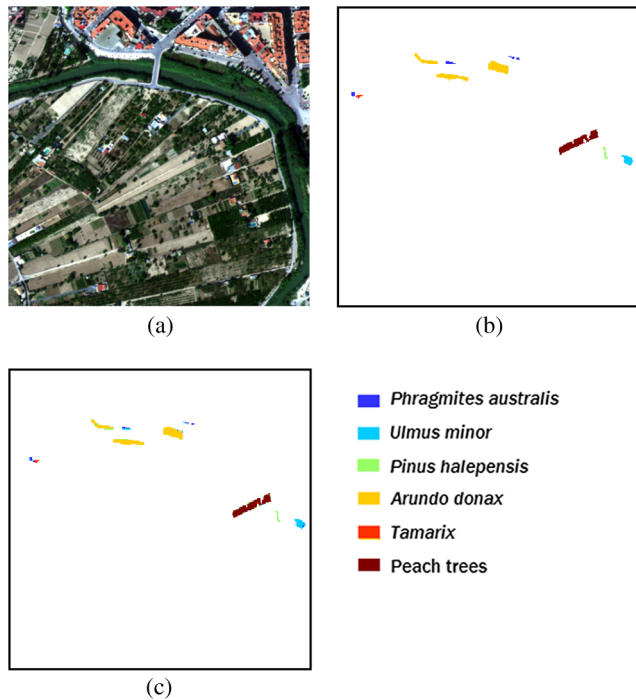


Fig. 3 Identification results of invasive and noninvasive vegetation on the hyperspectral image using fuzzy C-means optimized (FCMO): (a) original image (three components over 62 for visualization), (b) ground truth (GT) areas, and (c) FCMO classification result.

These results show that FCMO yields the best results despite its unsupervised nature.

3.2 Assessment on Real Image

FCMO is also evaluated on two real applications. In the first application, a hyperspectral image is used for invasive vegetation identification, whereas in the second application, a multispectral image is used for the detection of pine trees.

3.2.1 Assessment on hyperspectral image

To assess our unsupervised method, GT data were provided with the image. The hyperspectral image has 62 spectral bands and its size is 1000×1000 pixels. The GT related to this image includes six different invasive and noninvasive vegetation classes, namely, *Phragmites australis*, *Arundo donax*, *Tamarix*, *Ulmus minor*, *Pinus halepensis*, and peach trees.¹⁶ The GT comprises 9059 pixels.

For each component of this image, FCMO is first applied using the feature set described above. This step, therefore, provides 62 independent classification results, which are then fused by a genetic algorithm.^{11,12} Figure 3 and Table 4 show the result of FCMO and the corresponding confusion matrix, respectively.

The obtained result shows that the proposed method provides a very good classification result with an ACCR of 98.71%.

To confirm the relevance of the FCMO method, we compared it with SVM, ISODATA, FCM, and also a cooperative approach¹⁵ that uses SVM with ISODATA. The results of SVM and this cooperative approach with a postrelaxation are also presented.

Table 5 summarizes the results of this comparative study. The results validate the superiority of the proposed FCMO method in unsupervised classification of hyperspectral data.

3.2.2 Assessment on multispectral image

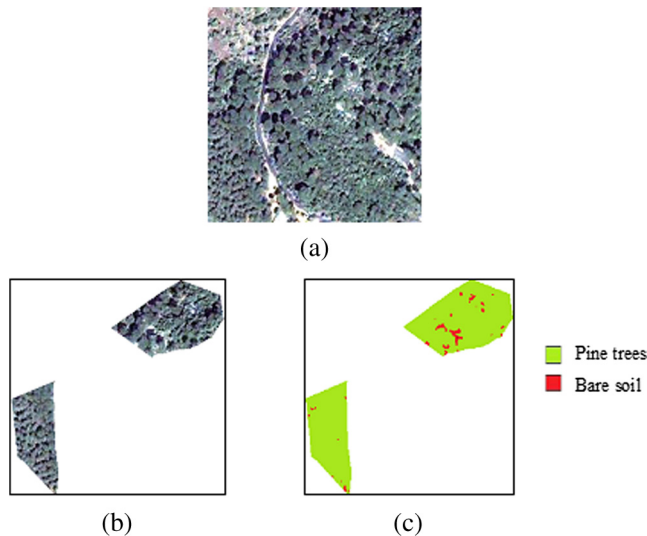
The performance of FCMO is investigated in another context: the detection of the presence of Pine trees in a region of Lebanon from a multispectral image [Fig. 4(a)] acquired by the Earth observation satellite Ikonos (2005). In order to assess FCMO, a GT is provided where 11,736 pixels are

Table 4 Confusion matrix for the proposed approach in the classification of invasive and noninvasive vegetation species [correct classification rate in %, (.) : number of pixels].

Classes predicted automatically by our approach	Ground truth classes (number of pixels)					
	<i>Phragmites australis</i> (544)	<i>Arundo donax</i> (4200)	<i>Tamarix</i> (162)	<i>Ulmus minor</i> (764)	<i>Pinus halepensis</i> (274)	<i>Peach trees</i> (3115)
<i>Phragmites australis</i>	98.90% (538)	1.15% (48)	2.47% (4)	0.52% (4)	1.10% (3)	0.93% (29)
<i>Arundo donax</i>	0	98.85% (4152)	0	0	0	0
<i>Tamarix</i>	0	0	97.53% (158)	0	0	0
<i>Ulmus minor</i>	0.73% (4)	0	0	99.48% (760)	0	0.48% (15)
<i>Pinus halepensis</i>	0.37% (2)	0	0	0	98.90% (271)	0
<i>Peach trees</i>	0	0	0	0	0	98.59% (3071)

Table 5 Compared ACCR for different unsupervised and supervised classification methods applied to invasive and noninvasive vegetation species.

Methods	SVM	SVM (with postrelaxation)	ISODATA	SVM with ISODATA	SVM with ISODATA (with postrelaxation)	FCM	FCMO
ACCR	93.20%	96.03%	23.13%	91.11%	93.67%	46.94%	98.71%

**Fig. 4** Example of pine trees detection result by FCMO: (a) original multicomponent image, (b) zones of GT identified as pine trees, and (c) FCMO detection result.

labeled in the original multispectral image as pine trees [Fig. 4(b)]. For this application, the FCMO algorithm is applied in the same manner as on the hyperspectral image described in Sec. 3.2.1. Figure 4(c) shows the detection result, where 11,328 of 11,736 pixels are detected by FCMO as pine trees. Therefore, the correct detection rate obtained is 96.52%. This application perfectly illustrates the problem of nonprecision of the GT provided. Indeed, all the pixels of the two areas in Fig. 4(b) are reported as corresponding to pine trees, whereas we see that some pixels actually correspond to bare soil. Therefore, we can consider that the good detection rates obtained by FCMO is higher than 96.52% if the GT was established with greater accuracy, which is difficult to achieve in practice. This fully justifies the development of completely unsupervised methods such as FCMO.

4 Conclusion

In this paper, a deterministic and unsupervised version of the FCM algorithm, named FCMO, is presented. FCMO is optimized in two aspects: the first one is stability, which is achieved through the use of an adaptive incremental approach that ensures independency in the initialization of the class centers; the second one is class number estimation using an objective evaluation criterion, which makes the algorithm fully unsupervised.

Experimental results, obtained on a synthetic image database and on two real applications in the context of hyperspectral and multispectral image partitioning, are presented. These results prove the effectiveness of the proposed

algorithm with regard to both stability and class number estimation, as well as classification efficiency in comparison with other unsupervised and supervised classification approaches.

Acknowledgments

The authors would like to thank Dr. C. Reyes Hurtado and Dr. M. Awad from Infraestructura y Ecología S.A. Aretech Group, Chile, and Lebanese National Remote Sensing Center, respectively, for providing the hyperspectral/multispectral images and the GT datasets.

References

1. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. Ser. B* **39**(1), 1–38 (1977).
2. C. Cortes and V. Vapnik, "Support-vector networks," *J. Mach. Learn.* **20**(3), 273–297 (1995).
3. T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.* **13**(6), 47–60 (1996).
4. A. K. Jain, "Data clustering: 50 years beyond k-means," *Pattern Recognit. Lett.* **31**(8), 651–666 (2010).
5. S. Bubeck, M. Meilă, and U. Von Luxburg, "How the initialization affects the stability of the k-means algorithm," *ESAIM Probab. Stat.* **16**, 436–452 (2012).
6. Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.* **28**(1), 84–95 (1980).
7. K.-L. Wu, "Analysis of parameter selections for fuzzy c-means," *Pattern Recognit.* **45**(1), 407–415 (2012).
8. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science* **315**(5814), 972–976 (2007).
9. C.-C. Hung, S. Kulkarni, and B.-C. Kuo, "A new weighted fuzzy C-means clustering algorithm for remotely sensed image classification," *IEEE J. Sel. Top. Signal Process.* **5**(3), 543–553 (2011).
10. F. Shen and O. Hasegawa, "An adaptive incremental LBG for vector quantization," *Neural Netw.* **19**(5), 694–704 (2006).
11. C. Rosenberger and K. Chehdi, "Genetic fusion: application to multi-components image segmentation," in *IEEE Proc. ICASSP'00*, Vol. 6, pp. 2223–2226 (2000).
12. M. Awad, K. Chehdi, and A. Nasri, "Multicomponent image segmentation using a genetic algorithm and artificial neural network," *IEEE Geosci. Remote Sens. Lett.* **4**, 571–575 (2007).
13. C. Rosenberger and K. Chehdi, "Toward a complete adaptive analysis of an image," *J. Electron. Imaging* **12**(2), 292–298 (2003).
14. P. Brodatz, *Textures: A Photographic Album for Artists and Designers*, Dover Publications, Inc., New York (1999).
15. Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitioning clustering techniques," *IEEE Trans. Geosci. Remote Sens.* **47**(8), 2973–2987 (2009).
16. K. Chehdi, M. Soltani, and C. Cariou, "Pixel classification of large size hyperspectral images by affinity," *J. Appl. Remote Sens.* **8**(1), 083567 (2014).

Kacem Chehdi received his PhD and the "Habilitation à diriger des Recherches" degrees in signal processing from University of Rennes 1, France, in 1986 and 1992, respectively. From 1986 to 1992, he was an assistant professor at the University of Rennes 1. Since 1993, he has been a professor of signal and image processing. His research activities pertain to adaptive processing at every level in the pattern recognition chain by vision.

Akar Taher received his BS degree in computer engineering from Cyprus International University in 2004 and his MS degree in computer science and IT from Koya University in 2006 (Kurdistan-Iraq). He obtained his PhD in signal and image processing from

University of Rennes 1, France, in 2014. His current research interests include unsupervised data classification.

Claude Cariou received his PhD in electronics from the University of Brest, France, in 1991. Since 1992, he has been with the Engineering School of Applied Sciences and Technology, where he is

currently with the Institute of Electronics and Telecommunications of Rennes, France. His research interests include image analysis, pattern recognition, unsupervised classification, texture modeling and segmentation, image registration and feature extraction/selection, mostly dedicated to multispectral and hyperspectral imagery.