Arab J Sci Eng (2014) 39:3747–3757
DOI 10.1007/s13369-014-1037-3

RESEARCH ARTICLE - COMPUTER ENGINEERING AND COMPUTER SCIENCE

# Clustering of Hyperspectral Images with an Ensemble Method Based on Fuzzy C-Means and Markov Random Fields

**Haikel Alhichri · Nassim Ammour · Naif Alajlan · Yakoub Bazi**

Received: 14 April 2012 / Accepted: 31 December 2012 / Published online: 28 March 2014


**Abstract** During the past few years, cluster ensemble methods have been introduced in the area of pattern recognition as a more accurate alternative to individual clustering algorithms as they can provide novel, robust, and stable solutions. The development of these methods is mainly motivated by the success of the combination of supervised classifiers also known as classifier ensemble. In this paper, we propose a novel ensemble method for the classification of hyperspectral images based on Fuzzy c-means (FCM) and Markov random field (MRF) theory. Firstly, in order to construct the ensemble of clustering maps, which call here "partitions", we run the FCM algorithm several times with different initializations and different subsets of features (spectral bands) selected randomly from the high-dimensional input feature vector characterizing the hyperspectral image. Secondly, since there is no-ground truth information, the class labels contained in each generated partition are symbolic, meaning that the same class of pixels can be labeled with different labels in different partitions of the ensemble. Therefore, an optimal relabeling of the ensemble with respect to a representative partition determined on the basis of the maximum entropy principle is made via a pairwise relabeling procedure. Finally, in the last step, the relabeled partitions are fused by means of an MRF method. The latter exploits two kinds of sources of contextual information: spatial and inter-partition contextual information. During the optimization process, the contributions of the different partitions are controlled through mutual information weights.

**Keywords** Hyperspectral images · Ensemble clustering · Markov random fields (MRFs) · Fuzzy c-means (FCM) · Mutual information (MI)

H. Alhichri · N. Ammour · N. Alajlan · Y. Bazi (✉)
ALISR Laboratory, Department of Computer Engineering, College of Computer and Information Sciences, King Saud University,
P.O. Box 51178, Riyadh 11543, Saudi Arabia
e-mail: ybazi@ksu.edu.sa

H. Alhichri
e-mail: hhichri@ksu.edu.sa

N. Ammour
e-mail: nammour@ksu.edu.sa

N. Alajlan
e-mail: najlan@ksu.edu.sa

**الخلاصة**

خلال السنوات القليلة الماضية، أدخلت خوارزميات تسمى فرقة التجميعات لمجال التعرف على الانماط كبديل أكثر دقة لخوارزميات التجميع الفردية لأنها يمكن أن توفر حلولا مبتكرة، وقوية، ومستقرّة. وكان الدّافع من وراء تطوير الطّرق المذكورة اعلاه هو نجاح طرق شبيهة يتمّ فيها إدماج نتائج مجموعة من خوارزميات التصنيف الغير موجهة وهو ما عرف باسم خوارزميات فرقة التصنيف الغير موجهة.

في هذه الورقة البحثية، نقترح نوع جديد من خوارزميات فرقة التجميعات ونطبّقها على الصور الفائقة الطيفية. وتعتمد هذه الحلول الجديدة على أمرين: 1) خوارزميات التجميع الضبابية أو ما يعرف بمصطلح (FCM) وذلك لانها تستخدم مفهوم المنطق الضبابي و 2) نظريات حقل ماركوف العشوائي أو ما يعرف بمصطلح (MRF).

أولا ، من أجل بناء فرقة التجميعات، نقوم بتشغيل خوارزمية التجميع الضبابية (FCM) عدة مرّات مع مدّها كل مرّة بقيم ابتدائية مختلفة وأيضا مجموعات مختلفة من الاوصاف (الصفات/الخصائص) والتي يتمّ اختيارها عشوائيا من الأبعاد العالية للصورة الفائقة الطيفية.

وثانيا ، بما أن التجميع أو التصنيف الحقيقي والاصلي للصورة غير معروف ، فإن التسميات المسندة لكل تجميع ستختلف من تجميع إلى آخر عبر فرقة التجميعات. ولكي يتم موائمة هذه التسميات وجب اختيار تجميع مرجعي يوائم عليه بقية التجميعات في الفرقة. وقد تمّ اختيار التجميع ذو أعلى إنتروبية المعلومات كتجميع مرجعي. اما طريقة الموائمة فهي تعمد على خوارزمية الموائمة المزدوجة (الازدواجية).

وفي الخطوة الأخيرة ، يتم دمج التجميعات في الفرقة ، بعد موائمتها ، باستخدام تقنية حقل ماركوف العشوائي (MRF). وهو يستغل نوعين من

مصادر المعلومات السياقية: المعلومات المكانية والمعلومات الظرفية بين الصورة. وأثناء عملية الاستمثال (التحسين) ، التي تقوم بها تقنية (MRF)، يتمّ التحكم في مساهمة كل من المصدرين من خلال أوزان تحسب بمبدأ المعلومات المتبادلة (MI).

## 1 Introduction

Hyperspectral remote sensing data are a set of images, where each image represents a range of the electromagnetic spectrum and is also known as a spectral band. Hyperspectral cubes are generated from airborne sensors like the NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), or from satellites like NASA's Hyperion. The higher spectral resolution of hyperspectral data means a greater quantity of data to memorize and process. Hyperspectral imaging differs from multispectral imaging in that the number of bands is much higher and the spectral bands are contiguous. The high spectral resolution characterizing these kinds of sensors allows discrimination between subtle differences in materials and ground covers. Nonetheless, the large dimensional data spaces generated by these sensors can result in degraded classification accuracies due to the Hughes effect [1].

In the literature, two main approaches to the classification problem have been proposed: based on supervised and unsupervised learning. The supervised methods require the availability of a training set for classifier learning. By contrast, unsupervised methods, known also as clustering methods, perform classification just by exploiting information present in the data, without the need for any training sample set. The supervised methods offer higher classification accuracy compared to the unsupervised ones, but in some applications, it is necessary to resort to unsupervised techniques because training information is not available. Compared to supervised classification, clustering has received less attention as it is ill-posed problem. Among the promising work available in the literature, one can find those published in [2–12]. In detail, Shi et al. [2] proposed a method of hyperspectral band reduction based on rough sets and Fuzzy c-means (FCM) clustering. First, the FCM clustering algorithm is used to classify the original bands into equivalent band groups. Then, the dimensionality on the data is reduced by selecting only the band with maximum grade of fuzzy membership from each of the groups. Gda et al. [3] discussed the limitations of $k$ means algorithm implementation and proposed a hardware solution to accelerate its implementation. Lee and Crawford [4] presented a two-stage hierarchical clustering technique for classifying hyperspectral data. In a first step, a local segmentator performs region-growing segmentation by merging spatially adjacent clusters. Then, in a second step, a global segmentator clusters the segments resulting from the previous stage using an agglomerative hierarchical clustering scheme based on a context-free similarity measure. Kersten et al. [5] compared the performances of five clustering techniques for classifying polarimetric synthetic aperture radar (SAR) images. The first two techniques are fuzzy clustering algorithms based on the standard $l_1$ and $l_2$ metrics. Two others combine a robust FCM clustering technique with a distance measure based on the Wishart distribution. Finally, the fifth technique is based on the expectation–maximization algorithm assuming that data follow a Wishart distribution. Marcal and Castro [6] proposed an agglomerative hierarchical clustering method for multispectral images, which uses both spectral and spatial information for the aggregation decision. Tran et al. [7] proposed an Markov random field (MRF) clustering method, exploiting both spectral and spatial interpixel dependency information, for polarimetric SAR images. Zhong et al. [8] proposed a two-step unsupervised artificial immune classifier for multi/hyperspectral images. Bandyopadhyay et al. [9] performed an unsupervised land cover classification by clustering pixels in the spectral domain into several fuzzy partitions. To this end, a multiobjective optimization algorithm is utilized to guide the fuzzy partitioning problem by means of a simultaneous optimization of different fuzzy cluster validity indexes. Liu et al. [10] proposed a weighted FCM clustering algorithm to carry out the fuzzy or the hard classification of multispectral images. Xia et al. [11] presented a clustering method for SAR images by embedding an MRF model in the clustering space and using graph cuts to search for data clusters optimal in the sense of the maximum a posteriori criterion. Bilgin et al. [12] performed an unsupervised classification of hyperspectral images by applying FCM clustering as well as its extended version termed as Gustafson-Kessel clustering which is based on an adaptive distance norm. Finally, Paoli et al. [13] developed an unsupervised classifier based on multiobjective particle swarm optimization. Their method optimizes criteria based on the log-likelihood function, the Bhattacharyya statistical distance between classes, and the minimum description length (MDL).

During the past few years, cluster ensemble methods have been introduced in the area of pattern recognition as a more accurate alternative to individual clustering algorithms as they can provide novel, robust, and stable solutions [14–17]. The development of these methods is mainly motivated by the success of the combination of supervised classifiers also known as classifier ensemble. The basic idea in classifier ensembles is that by considering different views of the same data, one can generate an ensemble of diverse partitions. The combination of the resulting partitions according to a consensus function can result in a better data partitioning even when the clusters are not compact and well separated. Typically, the proposed combination methods are different in two aspects: generation process and consensus function. Usually, the gen-

eration of diverse partitions can be obtained using two different strategies: data representation or clustering algorithms. In the first strategy, different partitions can be generated by employing different feature extraction mechanism, selecting a subset of features or perturbation of the data with sampling techniques. In the second strategy, clustering ensembles are obtained using different clustering algorithms or using the same algorithm with different parameters and initializations. Regarding the consensus function used for fusing the generated partitions, one can find approaches based on voting, co-association and the concept of median partition.

In this work, we focus on the unsupervised classification of hyperspectral remote sensing images. In particular, we propose a novel method based on an ensemble of image partitions, generated using the FCM clustering algorithm, and then fused by an MRF technique. Firstly, in order to construct the ensemble of partitions, we run the FCM algorithm several times with different initializations and different subsets of features selected randomly from the high-dimensional input feature vector characterizing the hyperspectral image. Secondly, since there is no-ground truth information, the class labels contained in each generated partition of the ensemble are symbolic. Therefore, an optimal relabeling of the ensemble with respect to a representative partition determined on the basis of the maximum entropy principle is made via a pairwise relabeling procedure. For such purpose, the relabeling issue is formulated as a weighted bipartite matching problem, which is a well-known combinatorial optimization problem that is solvable using the Hungarian method [18]. Finally, in the last step, the relabeled partitions are fused by means of an MRF fusion strategy. The latter exploits two kinds of sources of contextual information: spatial and inter-partition contextual information. The first source defines the spatial correlation between pixels and their neighbors within the same partition whereas the second one explores the relationship between the different partitions of the ensemble. During the optimization process, the contributions of the different partitions are controlled through mutual information (MI) weights.

The rest of the paper is organized as follows: Sect. 2 is devoted to the description of the proposed fusion approach. The experimental part is reported in Sect. 3. Finally, conclusions are drawn in Sect. 4.

## 2 Proposed Ensemble Method

The proposed ensemble method is composed of the following main steps: (1) Generating an ensemble of image partitions using the FCM clustering algorithm; (2) Choosing a base partition from the cluster ensemble and align all other partitions to it; (3) Fusing the aligned partitions using an MRF approach.

### 2.1 Ensemble Generation with FCM

In this work, the generation of the ensemble is made by running the FCM algorithm $P$ times with random initializations. To increase the diversity of the ensemble of partitions, $N_b$ bands are chosen randomly at each time in the range $[N_b^{\min}, N_b^{\max}]$ and fed as input to FCM. In fuzzy clustering, each point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging to just one cluster completely. Thus, points on the edge of a cluster may be in the cluster to a lesser degree than points in the center of cluster. An overview and comparison of different fuzzy clustering algorithms can be found in [19]. For each point $x$, we have a coefficient giving the degree of being in the $i$th cluster, $G_i(x)$, which is related to the inverse of the distance to the cluster center:

$$G_i(x) = \frac{1}{\Sigma_{j=1}^{c} \left( \frac{D(x,\text{center}_i)}{D(x,\text{center}_j)} \right)^{1/2(m-1)}}, \quad (1)$$

where the sum of those coefficients over all $c$ clusters for any given $x$ is 1:

$$\forall x \sum_{i=1}^{c} G_i(x) = 1 \quad (2)$$

These fuzzy weights will be effectively used to decide the class membership of each pixel in the MRF estimation step.

### 2.2 Alignment of the Partitions

Let $(A_i, i = 1:, \ldots P)$ be the different partitions obtained by applying the FCM. All partitions are characterized by the same number of classes (or clusters) $c$. Since no prior knowledge is available, the class labels contained in each partition $A_i$ are symbolic, meaning that the same class of pixels can be labeled with different labels in different partitions of the ensemble. To illustrate the problem, consider Fig. 1, where we show two partitions in a simplified form. The classes colored red are clearly the same class even though their shapes

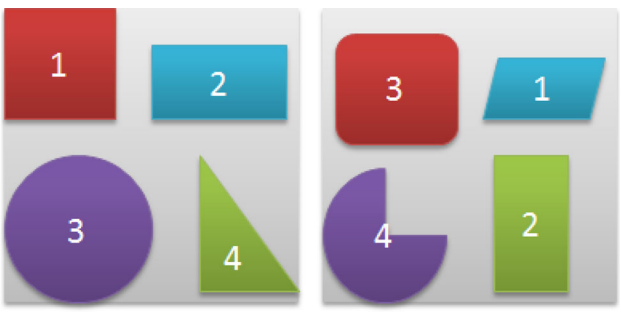


**Fig. 1** Illustration of the class label alignment problem: The labels (*1*, *2*, *3*, *4*) in partition (**b**) must be mapped to the labels (*2*, *4*, *1*, *3*) of the base partition (**a**)

are slightly changed (due to classification errors). Yet, the clustering algorithms labeled them differently in both partitions (label 1 versus label 3). Similarly all other classes have different labels. The aim of the alignment step is to find the mapping of class labels between partitions. For example, in the above case, the labels (1, 2, 3, 4) in partition (b) are mapped to (2, 4, 1, 3) to align it with partition (a).

But first, we have to select a base partition $A_r$ from the ensemble, then all other are partitions are aligned with the base partitions so that similar classes are roughly in the same regions in all partitions. This base partition is selected on the basis of the maximum entropy principle [14,15], i.e,

$$A_r; \text{Such that } r = \max_{i=1:P} H(A_i) \tag{3}$$

Where

$$H(A_i) = -\sum_{j=1}^{c} \frac{n_j}{n} \log\left(\frac{n_j}{n}\right), \tag{4}$$

where $n_j$ represent the number of pixels in cluster $j$ and $n$ is the total number of pixels. Then, the remaining partitions are successively aligned to this base partition via a pairwise relabeling procedure. In this context, the relabeling issue is formulated as a weighted bipartite matching problem, which is a well-known combinatorial optimization problem that is solvable in $O(N^3)$, using the Hungarian method [18].

### 2.3 Markovian Fusion

It is well known from the literature of image processing that the analysis of image pixels under spatial independence assumption may lead to inconsistencies mainly due to the presence of noise. Hence, making a decision for a pixel by taking into account its neighborhood often represents an effective way for increasing the accuracy of the result [20]. To this end, MRFs have proved a powerful and successful mathematical framework as shown by various works dealing with different remote sensing problems [21–25].

Under the Markovian assumption, the generation of an optimal partition $Y^*$ of all the pixels of the original image $X$ of size $N \times M$, given the aligned partitions $A_i$ ($i = 1, 2, \ldots, P$), one can resort to the maximum a posteriori probability (MAP) decision criterion:

$$P(Y^*|A_1, A_2, \ldots, A_P) = \max_{Y} \{P(Y|A_1, A_2, \ldots, A_P)\} \tag{5}$$

The adoption of the MRF approach simplifies the complexity of this maximization problem by passing from a global model to a model of the local image properties. The combination of the MAP method with the MRF modeling makes the classification task equivalent to the minimization of a total energy function $U_T$ expressed in the following relationship:

$$P(Y|A_1, A_2, \ldots, A_P) = \frac{1}{Z} \exp\left[-U_T(Y, A_1, A_2, \ldots, A_P)\right], \tag{6}$$

where $Z$ is a normalizing constant.

The total energy function $U_T$ can be rewritten in terms of local energy functions $U_{mn}$ using the concept of neighborhood [20]:

$$U_T(Y, A_1, A_2, \ldots, A_P) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} U_{mn}(y_{mn}, Y^S(m,n), A_1^S(m,n), A_2^S(m,n), \ldots, A_P^S(m,n)) \tag{7}$$

where $Y^s(m,n)$ and $A_i^s(m,n)$ stand for the set of labels of the pixels of the partition $Y$ and the partitions $A_i$ ($i = 1, 2, \ldots, P$), respectively, in a predefined neighborhood system $S$ associated with pixel $(m,n)$.

The minimization of (7) can be carried out by means of different algorithms such as simulated annealing (SA), the maximizer of posterior marginals (MPM), and the iterated conditional modes (ICM) algorithms [20]. In this work, the ICM algorithm is adopted since it represents a simple and computationally moderate solution to optimize the MRF-MAP estimates, to converge to a local, but usually good minimum of the energy function.

Similar to what is done previously in the context of change detection [21], we decompose the local energy function $U_{mn}$ into two kinds of sources of contextual information, which contribute to the optimization process. These are: (1) the spatial contextual information source, which defines the spatial correlation in partition $Y$ between the label of pixel $(m,n)$ and the labels of its neighbors; and (2) the inter-partition information sources, which express the relationship between the partition $Y$ and each of the partitions $A_i$ ($i = 1, 2, \ldots, P$). Accordingly, the local energy function $U_{mn}$ to be minimized for the pixel $(m,n)$ can be written as follows:

$$U_{mn} = \beta_{SP} \cdot U_{SP}(y_{mn}, Y^S(m,n)) + \sum_{i=1}^{P} \beta_i \cdot U_{II}(y_{mn}, A_i^S(m,n)), \tag{8}$$

where $U_{SP}$ and $U_{II}$ refer to the spatial and inter-partition energy functions, respectively, while $\beta_{SP}$ and $\beta_i$ ($i = 1, 2, \ldots, P$) represent the spatial and inter-partition parameters, respectively.

The neighborhood system $S = Y^S \cup A_1^S \cup \cdots A_P^S$ adopted to define the two kinds of energy functions required to compute the local energy function in (7) is based on a second-order neighborhood. On the basis of this neighborhood sys-

tem, the spatial energy function can be expressed as:

$$U_{SP}(y_{mn}, Y^S(m,n)) = -\sum_{y_{pq}\in Y^S(m,n)} I(y_{mn}, y_{pq}), \quad (9)$$

where $I(.,.)$ is the indicator function which allows us to count the number of occurrences of $y_{mn}$ in $Y^S$ (the spatial part of $S$) and is defined as:

$$I(y_{mn}, y_{pq}) = \begin{cases} 1, & \text{if } y_{mn} = y_{pq} \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

The inter-partition energy function characterizing the spatial correlation between the partition $Y$ and the partitions $A_i$ ($i = 1, 2, \ldots, P$) is given by:

$$U_{II}(y_{mn}, A_i^S(m,n)) = -\sum_{A_i(p,q)\in A_i^S(m,n)} \times G_i(p,q) I(y_{mn}, A_i(p,q)), \quad (11)$$

where $G_i(p,q)$ is the class membership grade, defined in Eq. (1) for the pixel at location $(p,q)$ as classified in partition $A_i$.

The weight $\beta_i$ associated with the inter-partition function aims at controlling, during the fusion process, the effect of unreliable decisions of the clustering step. The unavailability of training samples makes the determination of these weights particularly difficult. A possible solution to this problem consists of utilizing the MI between a pair of partitions to compute the weight for each partition. For two partitions $A_a$ and $A_b$ from the ensemble $P$, the normalized MI is given by:

$$\Phi^{\mathrm{NMI}}(A_a, A_b) = \sum_{i=1}^{c}\sum_{j=1}^{c} \frac{n_{ij}^{ab}}{n} \log\left(\frac{\frac{n_{ij}^{ab}}{n}}{\frac{n_i^a}{n}\cdot\frac{n_j^b}{n}}\right), \quad (12)$$

where $n_{ij}^{ab}$ denotes the number of shared patterns between the clusters $i$ and $j$ belonging to the partitions $A_a$ and $A_b$, respectively. Then, for every partition $A_i$, the weight $\beta_i$ can be computed as follows:

$$\beta_i = \frac{1}{P}\sum_{j=1,\, j\neq i}^{P} \Phi^{\mathrm{NMI}}(A_i, A_j) \quad (13)$$

Accordingly, with this weighting mechanism, a clustering algorithm is penalized if it exhibits a value that is statistically incompatible with those of the ensemble.

In the following, we present the algorithm of the propose Markovian fusion method:

**Algorithm: MRF-Fusion**

*Inputs:*

**X** – hyperspectral image of size $N\times M\times d$

$P$ – size of the ensemble

$c$ – Number of clusters

$\beta_{SP}$ – Spatial regularization parameter

*Iter* – Number of MRF iterations

$N_b$ – Number of selected bands (such that $N_b<d$)

*Outputs:*

$\mathbf{Y}^*$ – final classification map

1. *for* $i$=1:$P$
   1.1. Select randomly $N_b$ spectral channels from X.
   1.2. Generate a partition (a clustering map) by applying an FCM to these $N_b$ spectral channels.
2. end
3. Identify the base partition using the maximum entropy principle as shown in (4) and (5).
4. Align the remaining partitions to the base partition using the Hungarian method [18].
5. Compute the weights $\beta_i$ associated to these aligned partitions according to (13) and (14).
6. Initialize **Y** by minimizing for each pixel $(m, n)$ the local energy function $U_{mn}$ defined in (9) without the spatial energy term (i.e., by setting $\beta_{SP} = 0$).
7. *for* $k = 1$: *Iter*
8. Update **Y** by minimizing for each pixel $(m, n)$ the local energy function $U_{mn}$ defined in (9) including the spatial energy term $\beta_{SP}$.
9. end

## 3 Experimental Results

### 3.1 Synthetic Dataset

To illustrate the effectiveness of the new solution clearly, synthetic data are used in the first part of the experiments.

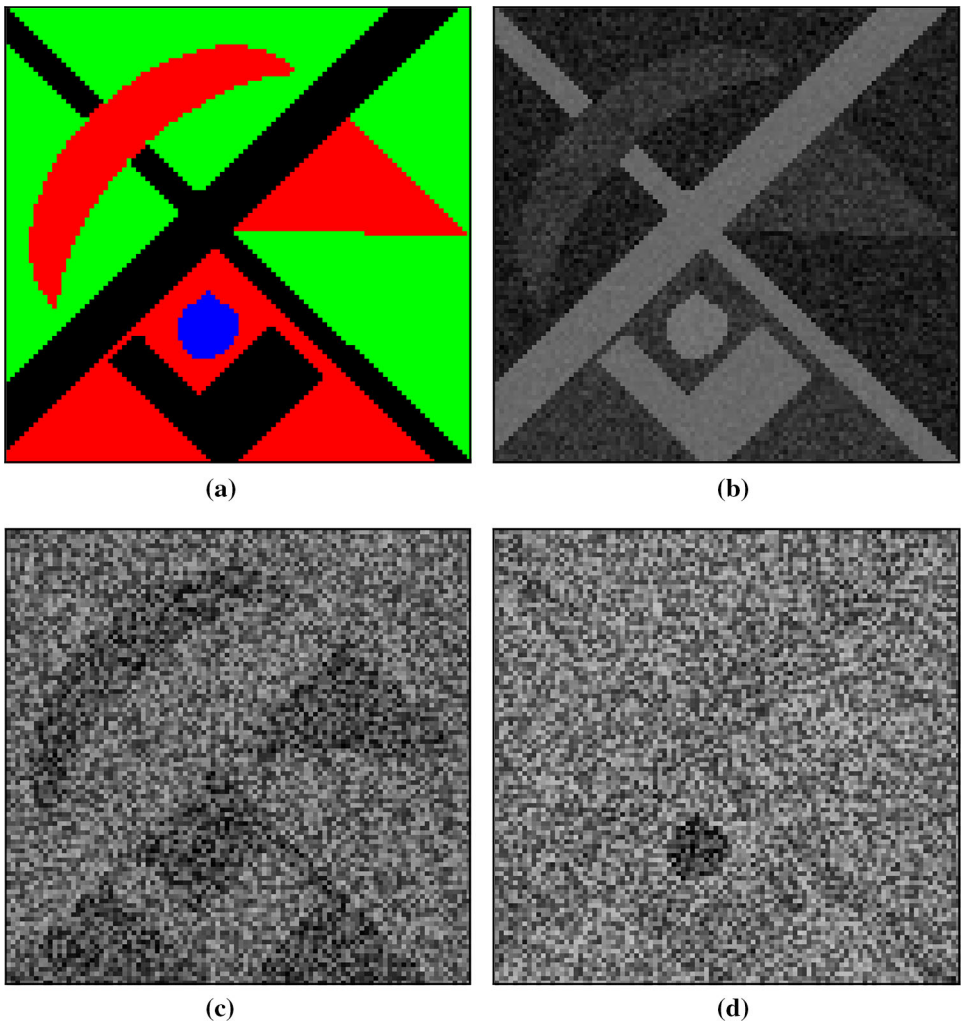


**Fig. 2** Synthetic hyperspectral image: **a** ground truth (each color is a class); **b** noiseless band; **c** noisy band; and **d** very noisy band

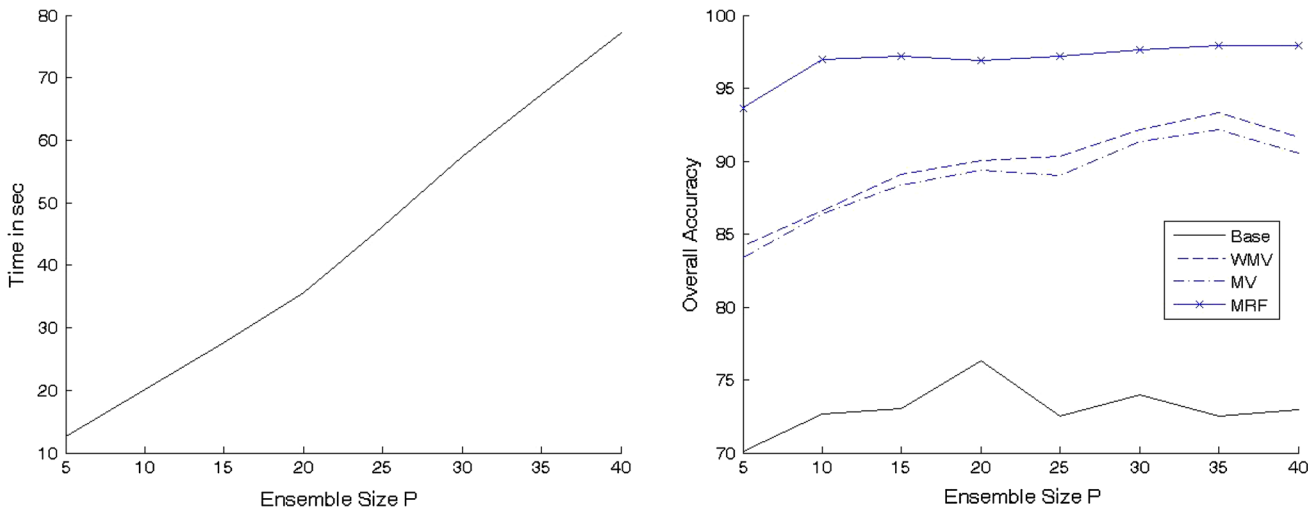


**Fig. 3** Computation time and overall accuracy versus the size of the ensemble $P$

To this end, 35 different synthetic hyperspectral images were generated randomly based on the following parameters:

- Each image is of size $100 \times 100$ pixels and composed of 100 bands.
- The number of classes in each band is fixed to 4.

**Table 1** Classification results in terms of overall accuracy (OA), average accuracy (AA), class-by-class accuracy and computation time for the synthetic hyperspectral dataset

| Method | Time (s) | (OA, $\sigma_{OA}$) | AA | C1 | C2 | C3 | C4 |
|---|---|---|---|---|---|---|---|
| Base partition | 1.51 | (76.31, 12.00) | 66.059 | 82.0567 | 82.8194 | 9.8632 | 29.4964 |
| MV | 32.35 | (90.02, 6.18) | 73.2189 | 98.7764 | 99.9217 | 80.3647 | 13.8129 |
| WMV | 32.35 | (89.35, 5.87) | 72.8093 | 97.847 | 99.7534 | 79.4742 | 14.1624 |
| MRF | 40.85 | (96.92, 3.22) | 74.7344 | 99.7813 | 99.5668 | 96.4034 | 3.186 |

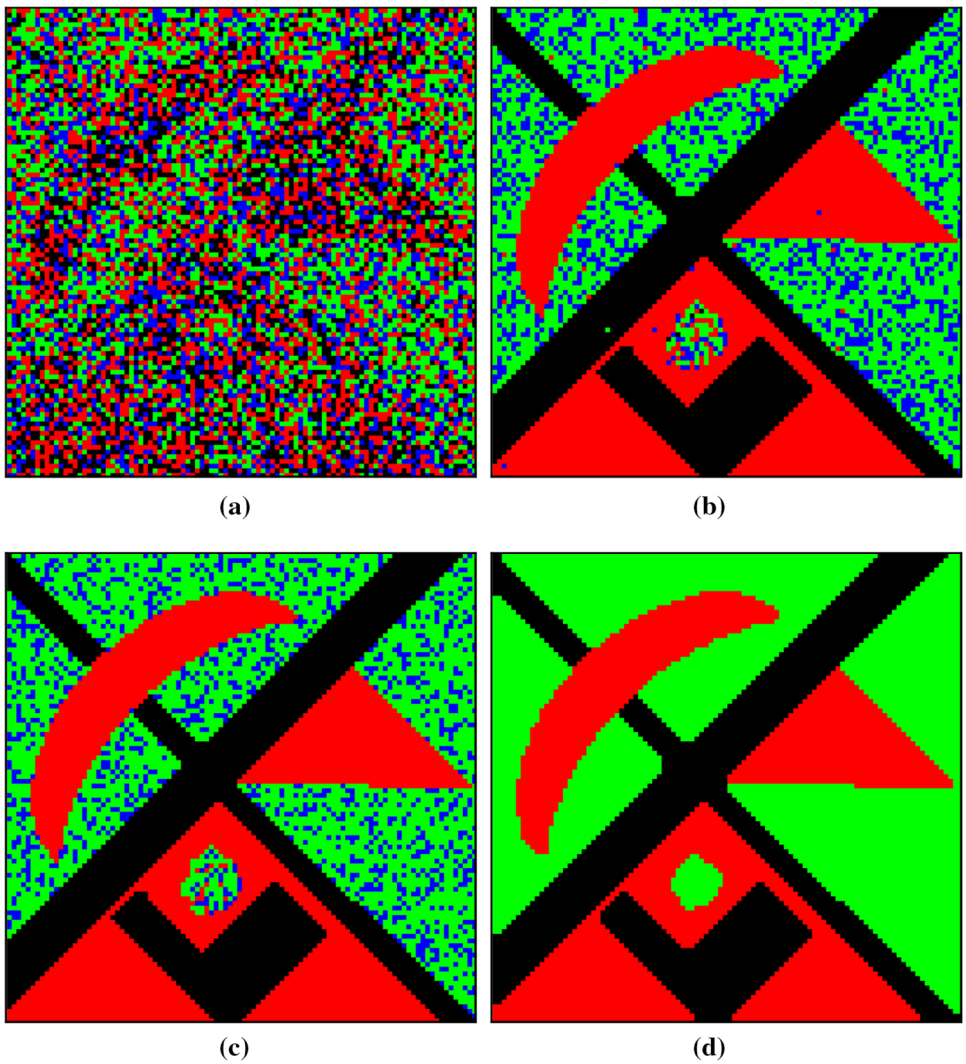


**Fig. 4** Qualitative results obtained for template 1 of the synthetic hyperspectral dataset: **a** base partition; **b** MV method; **c** WMV method; and **d** MRF method

- The classes occupy well-defined spatial shapes as specified in Fig. 2, which is also the ground truth used to compute the accuracy values.
- Each class has a Gaussian spectral distributions with random mean values in the range [0, 100] and random variance in the range [0, 100].
- In each image, the last 20 bands (out of 100) are corrupted by an additive Gaussian noise yielding a signal-to-noise ratio $SNR_{dB}$ randomly chosen such that ($SNR_{dB} \in [0, 5]$).

In the following results, accuracy is defined as the percentage of pixels that are classified correctly, i.e. the class assigned to them by our method matches the true class in the ground truth classification. Pixels for which we do not know the true class in the ground truth image are not considered. Also, computational time is reported based on a SONY VAIO computer (Intel Core i5 M540@2.53GHz) with 4.00 GB RAM.

Figure 3 shows the computation time and the overall accuracy (OA) with respect to the size of the ensemble $P$ for the

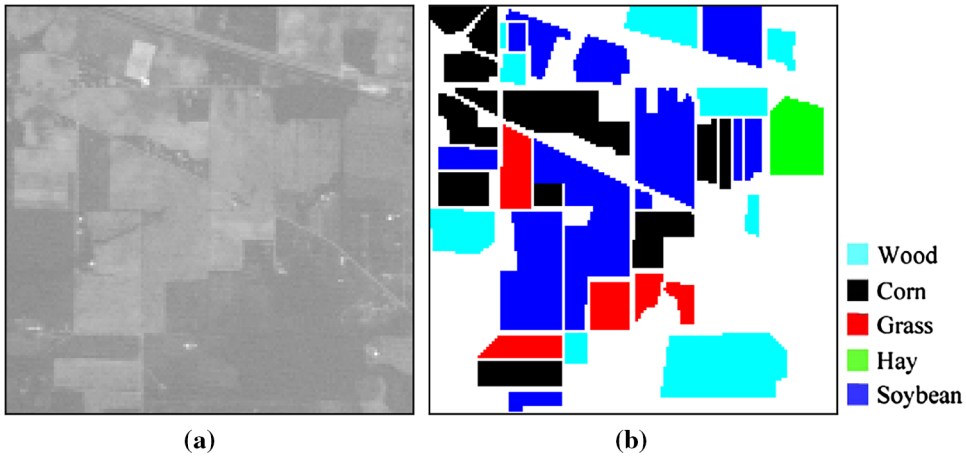


**Fig. 5** The Indiana hyperspectral dataset: **a** band 5; **b** ground truth

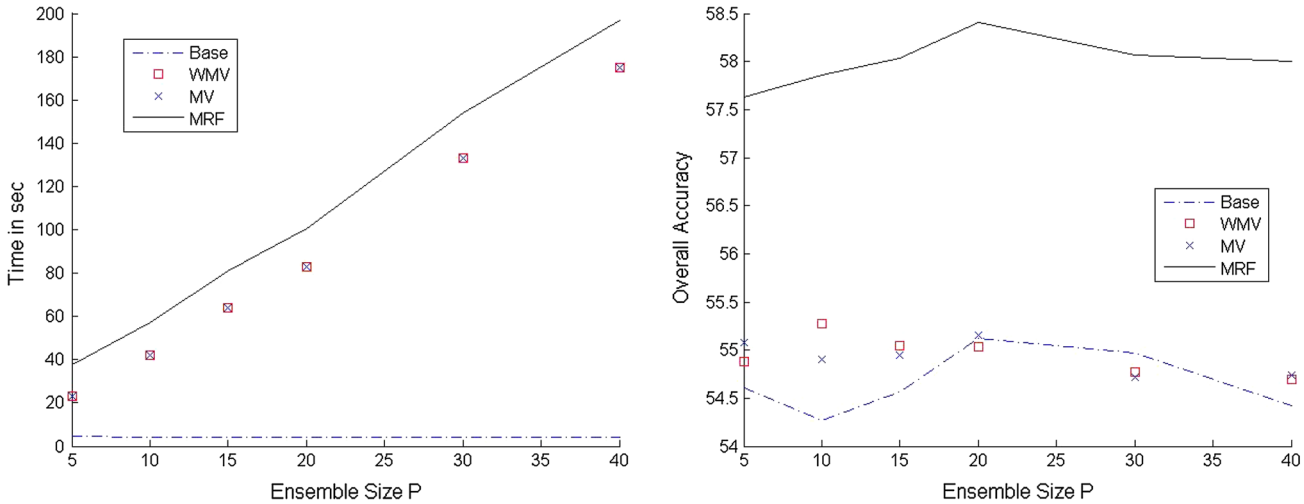


**Fig. 6** Computation time and overall accuracy versus the size of the ensemble $P$

**Table 2** Classification results in terms of overall accuracy (OA), average accuracy (AA), class-by-class accuracy and computation time for the Indiana hyperspectral dataset

| Method | Time (s) | (OA, $\sigma_{OA}$) | AA | Soybean | Wood | Corn | Grass | Hay |
|---|---|---|---|---|---|---|---|---|
| Base partition | 3.93 | (54.31, 4.50) | 54.6189 | 63.8929 | 45.4939 | 67.8058 | 43.1142 | 52.7880 |
| MV | 82.85 | (55.23, 2.56) | 58.9098 | 60.1632 | 53.0604 | 48.1156 | 80.6544 | 52.5552 |
| WMV | 82.85 | (55.23, 2.48) | 59.0948 | 59.6868 | 52.7452 | 47.0907 | 82.9798 | 52.9714 |
| MRF | 100.61 | (58.64, 2.27) | 63.4869 | 66.5250 | 54.6740 | 53.0672 | 88.5948 | 54.5735 |

first template of the synthetic dataset. It is interesting to notice that we have executed the algorithm 35 times for every value of $P$ and computed the mean and standard deviations for the accuracy results. The spatial regularization parameter $\beta_{SP}$ is set to 1.5 as suggested in [21,24] and the number of bands $N_b$ is selected each time randomly within the range [5,20] to reduce the Hughes effect.

As can be seen, the overall accuracy is stable in the range between $P=20$ and $P=40$. Thus, $P=20$ is a good compromise between accuracy and computational time. The proposed method yielded better and stable results compared to the base partition identified on the basis of the maximum entropy principle in addition to two state of the art methods based on majority voting (MV) and weighted majority (WMV) voting rules. Table 1 reports the classification results in terms of class-by-class, average (AA), and overall (OA) accuracies averaged over the 35 templates. In details, the proposed MRF fusion method yielded on average (OA

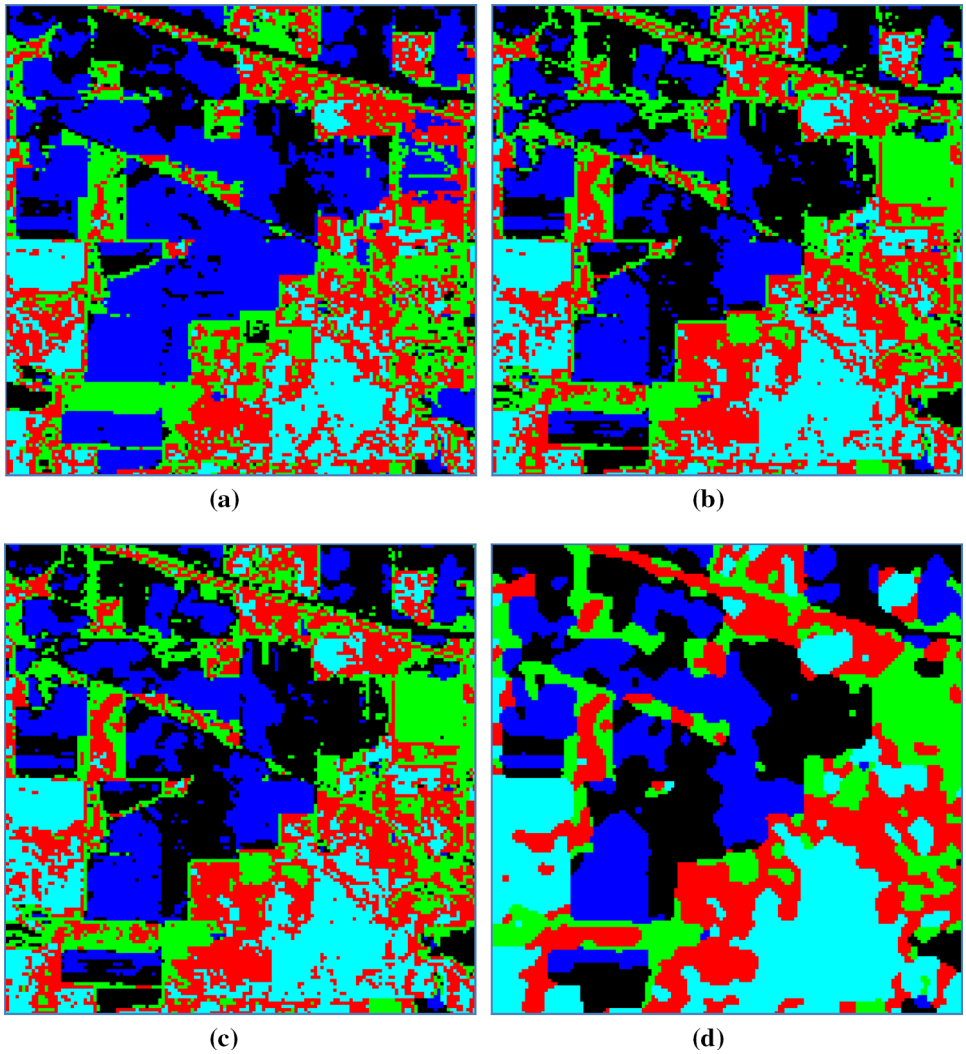


**Fig. 7** Classification results obtained for the Indiana hyperspectral dataset: **a** base partition; **b** MV method; **c** WMV method; and **d** MRF method

and AA) of (96.92 and 74.73 %). The (OA and AA) for the base partition, MV and WMV methods were equal to (76.31 and 66.05 %), (90.02 and 73.21 %) and (89.92 and 74.73 %), respectively. As can be seen, the MRF fusion method yielded significant gains with respect to MV and WMV methods. In addition, it is more stable since $\sigma_{\mathrm{OA}}$ is equal to 3.22, whereas it was equal to 6.18 and 5.87 for MV and WMV methods. From a qualitative point of view, Fig. 4 shows clearly the promising capabilities of the proposed method in increasing the quality of the classification result. It is worth noting that class 4 corresponds to the small blue circle in ground truth image of Fig. 2. This class has a small size, and hence the ratio of border pixels with respect to the non-border pixels is high. Since MRF technique gives inaccurate classification at the border pixels, the accuracy percentage for this class is low. Furthermore, any overlap between the distributions of this class with a much bigger class makes its class harder to separate from the bigger class.

### 3.2 Experiments on Real Data

The real hyperspectral dataset represents a section of a scene acquired in 1992 over the Indiana test Pines site in Northwestern Indiana by the AVIRIS sensor. The image has a spatial dimension of $145 \times 145$ pixels and a spatial resolution of 20 m per pixel. All available 220 spectral channels were used in the experiments. Figure 5a shows a single spectral channel of the Indiana Pine image. The ground-truth image (see Fig. 5b) used to assess the proposed approach consists of $c = 5$ land-cover classes, which are (wood = 2,171 pixels, corn = 2,502 pixels, grass = 959 pixels, hay = 489, and soy-bean = 4,050 pixels).

Figure 6 shows the computation time and the overall accuracy (OA) with respect to the size of the ensemble $P$ for real dataset. It is interesting to notice that we have executed the algorithm 35 times for every value of $P$ and computed the mean and standard deviations for the accuracy results. The spatial regularization parameter $\beta_{SP}$ is set to 1.5 as suggested

in [21,24] and the number of bands $N_b$ is selected each time randomly within the range [5,20] to reduce the Hughes effect.

As can be seen, the overall accuracy is in general stable in the range between $P=20$ and $P=40$. Thus, $P=20$ is a good compromise between accuracy and computational time. Table 2 shows the results obtained for $P=20$. The base partition (selected using the maximum entropy principle) resulted in (OA and AA) of (54.31 and 54.61 %). The MV, WMV and MRF methods yielded (OA and AA) of (55.21 and 58.90 %), (55.23 and 59.09 %) and (58.64 and 63.48 %), respectively. The ensemble methods based on MV, WMV were able to improve slightly the classification accuracy obtained by the base partition. However, the proposed MRF method was the most significant. The gain in accuracy with respect to the base partition, MV and WMV was equal to 4.33, 3.43, 3.41 %, respectively. Furthermore, the partitions shown in see Fig. 7 confirm again the capabilities of the proposed MRF fusion scheme.

## 4 Conclusions

This paper has presented a novel ensemble method for classifying hyperspectral images based on FCM and MRF. The FCM algorithm was used to generate an ensemble of partitions. Then, the MRF method was adopted to fuse efficiently the obtained partitions by taking into account the kinds of contextual information, which are the spatial and inter-partition contextual information. The experimental results obtained on synthetic and real datasets confirm the promising capability of the proposed fusion scheme in improving the classification accuracy with respect to state of the art methods based on majority and weighted MV. On average, the proposed method produced a gain in accuracy between 3 and 4 % as compared to the state-of-the-art weighted MV fusion technique. One limitation of the proposed method is that it does not work well at class borders, because the MRF technique starts to make incorrect estimates. One solution could be to change the fusion model at or near borders using edge detection methods. Furthermore, in the case of significantly unbalanced classes, FCM clustering tends to merger small classes with larger neighboring ones. In future works, one can adopt cost-sensitive clustering algorithms as a possible solution to this issue. Another investigation that can be left for future work is choosing a better method to select the most discriminative spectral channels for clustering. Indeed, instead of using random selection, one can look for suitable criteria to select the most informative bands or to transform the bands into a different space.

**Acknowledgments** This work was supported by a research Grant from the National Plan for Science and Technology (NPST). King Saud University. Project Identifier: 12-SPA-1192-02.

## References

1. Bazi, Y.; Melgani, F.: Toward an optimal SVM classification system for hyperspectral remote sensing images. IEEE Trans. Geosci. Remote Sens. **44**, 3374–3385 (2006)
2. Shi, H.; Shen, Y.; Liu, Z.: Hyperspectral bands reduction based on rough sets and Fuzzy c-means clustering. 20th IEEE Instrum. Meas. Technol. Conf. (IMTC'03) **2**, 1053–1056 (2003)
3. Gda, A.; Filho, S.; Frery, A.C.; de Araujo, C.C.; Alice, H.; Cerqueira, J.; Loureiro, J.A.; de Lima M.E.; Mdas Oliveira, G.S.; Horta, M.M.: Hyperspectral images clustering on reconfigurable hardware using the K-means algorithm. 16th Symp. Integr. Circuits Syst. Design (SBCCI'03) 99–104 (2003)
4. Lee, S.; Crawford, M.M.: Hierarchical clustering approach for unsupervised image classification of hyperspectral data. 2004 IEEE Geosc. Remote Sens. Symp. (IGARSS'04) Anchorage **2**, 941–944 (2004)
5. Kersten, P.R.; Lee, J.S.; Ainsworth, T.L.: Unsupervised classification of polarimetric synthetic aperture radar images using Fuzzy clustering and EM clustering. IEEE Trans. Geosci. Remote Sens. **43**, 519–527 (2005)
6. Marcal, A.R.S.; Castro, L.: Hierarchical clustering of multispectral images using combined spectral and spatial criteria. IEEE Geosci. Remote Sens. Lett. **2**, 59–63 (2005)
7. Tran, T.N.; Wehrens, R.; Hoekman, D.H.; Buydens, L.M.C.: Initialization of Markov random field clustering of large remote sensing images. IEEE Trans. Geosci. Remote Sens. **43**, 1912–1919 (2005)
8. Zhong, Y.; Zhang, L.; Huang, B.; Li, P.: An unsupervised artificial immune classifier for multi/hyperspectral remote sensing imagery. IEEE Trans. Geosci. Remote Sens. **44**, 420–431 (2006)
9. Bandyopadhyay, S.; Maulik, U.; Mukhopadhyay, A.: Multiobjective genetic clustering for pixel classification in remote sensing imagery. IEEE Trans. Geosci. Remote Sens. **45**, 1506–1511 (2007)
10. Liu, X.; Li, X.; Zhang, Y.; Yang, C.; Xu, W.; Li, M.; Luo, H.: Remote sensing image classification based on dot density function weighted FCM clustering algorithm. 2007 IEEE Geosci. Remote Sens. Symp. (IGARSS'07) Barc. **2**, 2010–2013 (2007)
11. Xia, G.S.; He, C.; Sun, H.: A rapid and automatic MRF-based clustering method for SAR images. IEEE Geosci. Remote Sens. Lett. **4**, 596–600 (2007)
12. Bilgin, G.; Ertürk, S.; Yildirim, T.: Unsupervised classification of hyperspectral-image data using Fuzzy approaches that spatially exploit membership relations. IEEE Geosci. Remote Sens. Lett. **5**, 673–677 (2008)
13. Paoli, A.; Melgani, F.; Pasolli, E.: Clustering of hyperspectral images based on multiobjective swarm optimization. IEEE Trans. Geosci. Remote Sens. **47**, 4175–4188 (2009)
14. Ayed, H.G.; Kamel, M.S.: On voting based consensus of cluster ensembles. Pattern Recognit. **43**, 1043–1953 (2010)
15. Ayed, H.G.; Kamel, M.S.: Cumulative voting consensus method for partitions with a variable number of clusters. IEEE Trans. Pattern Anal. Mach. Intell. **30**, 160–173 (2008)
16. Kuncheva, L.I.; Vetrov, D.P.: Evaluation of stability of k-means cluster ensembles with respect to random initialization. IEEE Trans. Pattern Anal. Mach. Intell. **28**, 1798–1808 (2006)
17. Fred, A.; Jain, A.K.: Combining multiple clustering's using evidence accumulation. IEEE Trans. Pattern Anal. Mach. Intell. **27**, 835–850 (2005)
18. Kuhn, H.: The Hungarian method for the assignment problem. Nav. Res. Logist. Quart. **2**, 83–97 (1955)
19. Nock, R.; Nielsen, F.: On weighting clustering. IEEE Trans. Pattern Anal. Mach. Intell. **28**, 1223–1235 (2006)
20. Dubes, R.C.; Jain, A.K.: Random field models in image analysis. J. Appl. Stat. **16**, 131–163 (1989)

21. Melgani, F.; Bazi, Y.: Markovian fusion approach to robust unsupervised change detection in remotely sensed imagery. IEEE Geosci. Remote Sens. Lett. **3**, 457–461 (2006)
22. Melgani, F.; Serpico, S.B.: A Markov random field approach to spatio-temporal contextual image classification. IEEE Trans. Geosci. Remote Sens. **41**, 2478–2487 (2003)
23. Serpico, S.B.; Moser, G.: Weight parameter optimization by the Ho-Kasyap algorithm in MRF models for supervised image classification. IEEE Trans. Geosci. Remote Sens. **44**, 3695–3705 (2006)
24. Bazi, Y.; Melgani, F.; Bruzzone, L.; Vernazza, G.: A genetic expectation maximization method for unsupervised change detection in multitemporal SAR imagery. Int. J. Remote Sens. **30**, 6591–6610 (2009)
25. Picco, M.; Palacio, G.: Unsupervised classification of SAR images using Markov random fields and $g_I^0$ models. IEEE Geosci. Remote Sens. Lett. **8**, 350–353 (2011)