Title: A New Self-Training Based Unsupervised Satellite Image Classification Technique Using Cluster Ensemble Strategy

This paper appears in: IEEE Geoscience and Remote Sensing Letters

Date of Publication: 31 October 2014

Author(s): B. Banerjee, F. Bovolo, A. Bhattacharya, L. Bruzzone, S. Chaudhuri, B.K. Mohan

Volume: 12, Issue: 4

Page(s): 741 - 745

Product Type: Journals & Magazines

DOI: 10.1109/LGRS.2014.2360833

URL: http://dx.doi.org/10.1109/LGRS.2014.2360833

A New Self-Training Based Unsupervised Satellite Image Classification Technique Using Cluster Ensemble Strategy

Biplab Banerjee, Francesca Bovolo, *Senior Member, IEEE*, Avik Bhattacharya, *Member, IEEE*, Lorenzo Bruzzone, *Fellow, IEEE*, Subhasis Chaudhuri, *Fellow, IEEE*, B Krishna Mohan

***Abstract*—This letter addresses the problem of unsupervised land-cover classification of remotely sensed multi-spectral satellite images from the perspective of cluster ensembles and self-learning. The cluster ensembles combine multiple data partitions generated by different clustering algorithms into a single robust solution. A cluster ensemble based method is proposed here for the initialization of the unsupervised iterative expectation-maximization (EM) algorithm which eventually produces a better approximation of the cluster parameters considering a certain statistical model is followed to fit the data. The method assumes that the number of land-cover classes is known. A novel method for generating a consistent labeling scheme for each clustering of the consensus is introduced for cluster ensembles. A maximum likelihood (ML) classifier is henceforth trained on the updated parameter set obtained from the EM step and is further used to classify the rest of the image pixels. The self-learning classifier, though trained without any external supervision, reduces the effect of data overlapping from different clusters which otherwise a single clustering algorithm fails to identify. The clustering performance of the proposed method on a medium resolution and a very high spatial resolution image have effectively outperformed the results of the individual clustering of the ensemble.**

***Index Terms*—Image Segmentation, Clustering, Ensemble learning.**

I. Introduction

REMOTE sensing images are an important source of information regarding the Earth surface. For many applications, the underlying land-cover information from such images is needed. Supervised or unsupervised classification techniques exist in the literature to generate a reliable land-cover map of the geographical area captured in the image [1].

With respect to supervised classification, clustering is inherently an ill-posed problem. Given a set of data samples, each clustering solution is equally plausible with no prior knowledge about the underlying probability distribution of the data. The clustering algorithms assume some model to describe the data which in effect, is reflected in the corresponding clustering results. If the data model does not match with the actual distribution of the data, the clustering result becomes erroneous. Moreover, clustering algorithms require either an implicit or explicit initial estimation of the inherent cluster parameters (e.g., mean, variance). An improper initialization may lead to a unreliable clustering result.

In view of the above, it is evident that, in order to obtain a good clustering result, some information about the data distribution is needed. However this information is hard to obtain, even from a domain expert. The exploratory nature of the clustering task requires efficient and robust methods that would benefit by combining the strength of diverse clustering algorithms. Cluster ensemble techniques can be useful to highlight the common cluster subspace information to be adopted for the entire dataset by using some supervised classification strategies. This method can be termed as self-learning based clustering approach.

Application of ensemble-based clustering techniques are relatively new in remote sensing. [2] has proposed a land-cover clustering algorithm for Very High Resolution (VHR) images exploiting the advantages of the morphological attribute profiles and ensemble clustering. K-means has been used as the base clustering method and the diversity is introduced in the consensus with different initializations of the cluster centers for K-means. K-means with different cluster centroid initializations are grouped together using the concept of cluster alignment in [3]. [4] has proposed a cluster combination strategy based on Support Vector Machine (SVM) and mixture modeling with the aim of classifying land-covers in hyperspectral data.

In this letter, a cluster ensemble strategy is proposed for land-cover classification of multispectral images. The proposed method differs from the previous ones because:

- It introduces diversity in the initial clustering process by incorporating different clustering techniques which has to be extremely different from each other in the topology as well as the underlying theory. Thus it is expected that the common sub-space per cluster obtained after the ensemble method points with high confidence to the samples that are part of the same cluster.
- It applies a novel robust information theory based cluster mapping step to ensure consistency in the clustering results. As no single global method is followed by different clustering techniques to assign labels to different clusters, it is important to develop a consistency rule for the cluster labels over the consensus to avoid false cluster matching.

The proposed method can be summarized in four steps. The initial clusters are obtained independently for the given image by different clustering techniques assuming that the number of clusters is known. A novel cluster mapping technique is followed to identify the corresponding clusters in different clustering results. A set of reliable samples for each cluster is identified to be used for the initialization of an iterative EM-based retraining [5]. The EM algorithm approximates cluster parameters assuming that clusters are Gaussian distributed. The final classification is obtained by an ML classifier trained on the updated parameters produced by the EM algorithm.

The letter is organized as follows. Section II details the proposed unsupervised cluster ensemble based land-cover classification technique. Experimental details are mentioned in Section III. Section IV concludes the paper with discussion.

II. Proposed Unsupervised Land-Cover Classification Algorithm

A. Self- Training based Unsupervised Classification

Let $X = \{x_{1,1}, x_{1,2}, \ldots x_{R,S}\}$ represent a multi-spectral remotely sensed satellite image with $R \times S$ pixels where each pixel $x_{r,c} \in \mathbb{R}^d$ in spectral domain. Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ represent N land-cover classes characterizing the geographical area represented by image X. Let us assume that N is already known, whereas the class labels are not. Otherwise some iterative validation techniques from the literature [6] can be employed to estimate N given X.

In the context of the Bayes decision rule, a given pixel $x_{r,c}$ is assigned to a specific land-cover class ω_k according to:

$$x_{r,c} \in \omega_k \Leftrightarrow \underset{\omega_l \in \Omega}{\operatorname{argmax}} \{P(\omega_l) p(x_{r,c}|\omega_l)\} \tag{1}$$

$P(\omega_l)$ and $p(x_{r,c}|\omega_l)$ represent the prior probability and the conditional probability density function for the l^{th} land-cover class, respectively. The training of the Bayes classifier consists of estimating the true prior probability and conditional probability function that describe a given land-cover class. This requires highly reliable samples for each land-cover class to be identified for the estimation of the underlying statistical distribution of the classes. Since the true distribution of a given class is unknown, a common practice is to model it by a known distribution like Gaussian, Poisson function, etc. However, in the current scenario, no training data are initially available. The proposed method exploits cluster ensemble technique for identifying a set of samples belonging to each cluster with high confidence. These samples are used to initialize an unsupervised EM algorithm that estimates the true cluster statistical parameters.

The proposed unsupervised land-cover classification method is composed of four major steps:

- Cluster the image independently into N clusters by M clustering algorithms which are selected being drastically different from each other and being weak learners.
- A label matching for the clusters is followed in order to obtain a consistency in the cluster labels produced by different clustering methods.
- Assuming that a cluster can be modeled by a Gaussian function, initial estimate of its mean vector and the covariance matrix is obtained by identifying a set of reliable samples belonging to each cluster. The parameters are updated by using the iterative EM algorithm.
- A ML classifier is modeled on the updated parameters that is used for classifying the image.

B. Obtain the individual clustering results

The ensemble clustering technique requires first to cluster the dataset into N clusters by M diverse clustering methods. The diversity in the clustering techniques has been established here by selecting kernel K-means, normalized graph-cut, fuzzy c-means and K-medoid clustering techniques. They are weak learners in the sense that:

- K-means inherently considers that the clusters are hyper-spherical or hyper-ellipsoidal in shape.
- FCM is based on the weighted average of all the points that approximates the cluster mean. It eventually leads the local means to the global mean of all the samples.
- Graph based clustering methods depend on the underlying graph topology. If the topology changes, the graph Laplacian also changes which affects the clustering result.
- K-medoid is related to K-means but is based on the medoid-shift algorithm which is robust to outliers.

Both K-means and FCM are sensitive to outliers. Hence, these methods are not likely to perform well in all the scenarios, however a proper combination of them is expected to improve their individual performance. Clustering techniques other than the ones listed above can be considered. The value of M has to be selected as a trade-off between the improvement of the accuracy of the final clustering results and the complexity of the system. Commonly small values of M satisfies the criterion.

Given X and N and set $M = 3$, let $\{\alpha_1, \alpha_2, \ldots, \alpha_N\}$, $\{\beta_1, \beta_2, \ldots, \beta_N\}$ and $\{\gamma_1, \gamma_2, \ldots, \gamma_N\}$ represent the cluster label sets of X. The cluster labels are inconsistent in the sense that for a given i, α_i, β_i and γ_i may not represent the same land-cover class. A cluster mapping step is necessary to solve this label ambiguity.

C. Cluster mapping

The cluster correspondence problem is solved here by a novel approach based on the Kullback-Leibler (KL) divergence cluster similarity and majority voting rule. The cluster mapping algorithm (Algorithm 1) is executed thrice considering each of the $\{\alpha_i\}_{i=1}^N$, $\{\beta_j\}_{j=1}^N$ and $\{\gamma_k\}_{k=1}^N$ independently as the base. In each case, the clustering results of the non-base methods are compared to the base one and cluster triplets are formed according to cluster similarity. A majority voting is applied to select for each cluster, the winner triplet of the consensus. For two discrete probability distributions R and Q representing the marginal data distributions of two different clusters and assuming that both of them are Gaussian distributed, the non-symmetric KL-divergence from R to Q is measured according to (2). KL divergence is a measure of relative entropy of two distributions over the same random variable. It further takes into account the degree of overlapping of the underlying datasets. The non-symmetrical divergence measure e.g. KL divergence is employed as the goal is to calculate the similarity of each clustering result with the results of the base clustering method. A small KL divergence signifies high similarity between the corresponding data distributions (clusters). A high value means the clusters are significantly different from each other.

$$KL(R,Q) = \sum_i \ln \left| \frac{R(i)}{Q(i)} \right| R(i) \quad (2)$$

For instance, let us consider $\{\alpha_i\}_{i=1}^N$ as the base. The best possible and unique β_j and γ_k $(1 \leq j,k \leq N)$ for each α_i $(1 \leq i \leq N)$ are identified using the KL-divergence measure. Given $\{\alpha_i\}_{i=1}^N$ and $\{\beta_j\}_{j=1}^N$, the algorithm proposed for the one-to-one mapping between the corresponding individual clusters is mentioned in Algorithm 2. Given D, a $N \times N$ matrix storing the pairwise KL-divergence among the clusters in $\{\alpha_i\}_{i=1}^N$ and $\{\beta_j\}_{j=1}^N$, the proposed algorithm tries to identify the matching cluster pair with the minimum KL divergence value at each iteration. Those pairs are omitted and the algorithm continues for the remaining clusters until all the assignments are performed.

Hence, for a given α_i, two pairs (α_i, β_j) and (α_i, γ_k) are formed and the triplet $(\alpha_i, \beta_j, \gamma_k)$ is defined by the union of(α_i, β_j) and (α_i, γ_k). A total of $3N$ triplets are formed by considering $\{\alpha_i\}_{i=1}^N$, $\{\beta_j\}_{j=1}^N$ and $\{\gamma_k\}_{k=1}^N$ as the base independently. Among them, for each α_i, $(1 \leq i \leq N)$, the winner triplet containing α_i is selected by applying the majority voting rule among all the triplets containing α_i. In case of the tie situation, for a given α_i, the triplet with the best average intra-cluster similarity is selected. This stage can alternatively be carried out with β_j, $(1 \leq j \leq N)$ or γ_k, $(1 \leq k \leq N)$.

Algorithm 1 Input: $\{\alpha_i\}_{i=1}^N$, $\{\beta_j\}_{j=1}^N$ and $\{\gamma_k\}_{k=1}^N$
Output: The unique consistent label assignment among $\{\alpha_i\}_{i=1}^N$, $\{\beta_j\}_{j=1}^N$ and $\{\gamma_k\}_{k=1}^N$

1: $\text{Temp} = \{\{\alpha_i\}_{i=1}^N, \{\beta_j\}_{j=1}^N$ and $\{\gamma_k\}_{k=1}^N\}$
2: **for** $m \in \text{Temp}$ **do**
3: Compare m and the remaining two clustering results from Temp independently using Algorithm 2.
4: For each cluster in m, a triplet with two clusters from the other two clustering results are formed.
5: **end for**
6: Considering $\{\alpha_i\}_{i=1}^N$ as the base, for all the individual α_i's, all the winner triplets of a majority voting on all the triplets are identified. Ties are resolved on the higher similarity measure basis.

Algorithm 2 Input: $\{\alpha_i\}_{i=1}^N$, $\{\beta_j\}_{j=1}^N$
Output: The unique label mappings between $\{\alpha_i\}_{i=1}^N$ and $\{\beta_j\}_{j=1}^N$

1: Compute the KL distance for each possible $\{\alpha\}$, $\{\beta\}$ pair and store it into $D_{N \times N}$.
2: Remove from D the element corresponding to the minimum KL distance. Let i_1, i_2 denote the row and column indices of the minimum value.
3: $\alpha(i_1)$ and $\beta(i_2)$ are declared to form an identical pair.
4: Repeat steps 2 and 3 until D is empty.

Once the cluster mapping is performed, the next step is to identify the common set of samples per triplet. These samples are the highly reliable candidates of the corresponding unknown land-cover which the triplet represents, as they have been recognized to belong to the same cluster.

D. Identification of the set of reliable samples per triplet

Given a triplet $(\alpha_i, \beta_j, \gamma_k)$, the common set of samples which belong to all the three individual clustering results are obtained by performing the set intersection operation on the samples with labels α_i, β_j and γ_k, independently. The specific samples which are very close to the centroid of the common set represent the set of highly reliable samples for the cluster. To select these highly reliable set of samples, the maximum pairwise Euclidean distance within the common set of samples is calculated. The specific subset of samples lying within the sphere rooting at the centroid and having a radius of $\delta\%$ of the maximum pairwise Euclidean distance have high memberships of belonging to that cluster. Euclidean distance is considered here as it is well-known in finding the distance between two feature vectors. Same process is repeated for all the triplets found from the Algorithm 1. A small δ provides more reliable samples. Let $\mathrm{Tr} = \{\mathrm{Tr}_1, \mathrm{Tr}_2, \ldots, \mathrm{Tr}_N\}$ denote the set of reliable samples for each cluster found in this step.

E. Classification by ML classifier and EM retraining

This step produces the clustering of X using an ML classifier retrained with the EM algorithm. The training of ML classifier requires the estimation of the class prior and the class conditional probabilities. As it is a very common practice in the remote sensing community, here we consider that each class is modeled with a Gaussian density function. This is a reasonable assumption when dealing with multispectral images acquired by passive sensors as the pixels follow the rule of large numbers and group in Gaussian clusters [5]. Under this assumption, the estimation of the class-wise probability terms of (1) reduces to the estimation of the mean, covariance matrix and the class prior probabilities. Let us consider a given cluster label ω_i representing the triplet $(\alpha_i, \beta_j, \gamma_k)$ where $\omega_i \approx \alpha_i \approx \beta_j \approx \gamma_k$ correspond to the common land-cover class given by the clustering algorithms. $\theta_i = \{\mu_i, \Sigma_i, P(\omega_i)\}$, $(1 \leq i \leq N)$ represents the set of cluster parameters (mean, covariance matrix, class prior) initialized from Tr. The values of the parameters in θ can be updated using the iterative EM algorithm considering the image X as a mixture of N Gaussian functions using the equations:

$$P_i^{l+1}(\omega_i) = \frac{1}{R \times S} \sum_{x_{r,s} \in X} \frac{P^l(\omega_i) p^l(x_{r,s}|\omega_i)}{P^l(x_{r,s})} \quad (3)$$

$$\mu_i^{l+1} = \frac{\sum_{x_{r,s} \in X} \frac{P^l(\omega_i) p^l(x_{r,s}|\omega_i)}{P^l(x_{r,s})} x_{r,s}}{\sum_{x_{r,s} \in X} \frac{P^l(\omega_i) p^k(x_{r,s}|\omega_i)}{P^l(x_{r,s})}} \quad (4)$$

$$\Sigma_i^{l+1} = \frac{\sum_{x_{r,s} \in X} \frac{P^l(\omega_i) p^l(x_{r,s}|\omega_i)(x_{r,s} - \mu_i^{l+1})^2}{P_{l(x_{r,s})}}}{\sum_{x_{r,s} \in X} \frac{P^l(\omega_i) p^l(x_{r,s}|\omega_i)}{P^l(x_{r,s})}} \quad (5)$$

where l represents the l^{th} iteration. In EM, at each iteration, the estimated new values of the parameters provide an increase of the log-likelihood function until a local maxima is reached. Once the updated θ is obtained for each class ω_i, the Bayes rule of (1) is used to classify all the remaining samples of X to produce the final classification map.

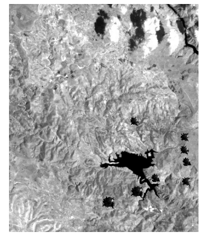
III. Experimental Results

The effectiveness of the proposed unsupervised land-cover classification technique has been analyzed on two datasets. The land-cover classification accuracy of the proposed self-training based unsupervised classifier has been compared with the ones produced by the clustering methods used for the ensemble independently, i.e., the proposed method has been applied in the limit case by considering single clustering method in the EM initialization phase (Section II-B). None of the four clustering methods resulted in reliable training samples because of cluster overlapping, thus the performance of the ML classifier is poor. In addition the proposed method has been compared with the cluster ensemble strategy proposed in [3]. 5 different versions of the traditional K-means with different cluster centroid initializations have been considered in [3]. In order to select a set of highly reliable samples for each cluster, δ was set to 25.

In order to assess the robustness of the cluster mapping process, M was set to 3 and all the combinations of three clustering algorithms out of four have been tested. The effectiveness and robustness of the proposed method is proven by reporting the average classification accuracy obtained after EM+ML classification over all possible trials. A good classification accuracy implies that the cluster mapping step performs well, on failure of which poor accuracy would be recorded, instead.

A. Medium Resolution Sardinia Dataset

The first study area considered in the experiments was acquired by the Thematic Mapper (TM) sensor of the LandSat 5 satellite in September 1995. Though the image consists of 7 bands but in the experiments conducted, band 6 has been neglected due to its lower geometrical resolution. The selected test site is a section of 412×493 pixels of a scene including the area surrounding the Lake Mulargia on the Island of Sardinia (Italy). Figure 1(a) depicts the band 4 of the image. 5 natural land-cover classes can be identified from the image, i.e. Pasture, Forest, Urban, Water and Vineyard. A burned area class has additionally been simulated in the image to increase the complexity [7]. Test samples have been selected by photo-interpretation for all the classes and the corresponding reference map is used to assess the clustering accuracy. Radial Basis Function (RBF) kernel function has been used along with Kernel K-means. The kernel neighborhood parameter has been set empirically. A typical value of the kernel parameter found is 7.5×10^{-4} which provides the best separation of the clusters in term of the inter-cluster distance. The cluster centroids and medoids for Kernel K-means, FCM and K-medoids and the membership matrix for FCM have been initialized heuristically. The proposed mapping algorithm is efficient in properly identifying clusters which are heavily overlapped with each other, i.e. Pasture, Vineyard and Urban.



(a)

(b)

Fig. 1: (a)The band 4 of the simulated Sardinia Dataset (b)The NIR band of the QuickBird Dataset

The performance of the aforementioned clustering techniques are depicted in Table I. It can be observed that, for all the clustering methods, the class-wise accuracies of Pasture and Vineyard classes are very low (44.12% to 62.77% and 52.47% to 64.10%, respectively). This is due to the fact that these two classes have similar spectral signatures, e.g. they are heavily overlapped. The Urban class has slight overlapping with these two classes in different spectral bands. The land-cover mapping accuracy produced by Kernel K-means is the best individual clustering result (79.04%). The proposed clustering technique produces an average overall accuracy of 86.45% which is far better than the one achieved by each clustering method independently. Sharp improvement in the clustering accuracies mainly for the Pasture and the Vineyard classes can be observed. Though some of the samples of the Urban class which are within the class boundaries of Pasture and Vineyard are wrongly misclassified, hence the clustering accuracy of the Urban class has degraded to a minor extent by the proposed method with respect to the individual clustering results. The proposed method also outperforms the ensemble method of [3] by a considerable margin ($> 4\%$). In a supervised scenario, where true labeled samples from each classes are considered to train a ML classifier [7], a generalization accuracy of 96.24% has been observed. The performance of the proposed self-training based ML classifier is very close to that of the supervised ML classifier. In addition, the proposed method demonstrated to be robust with respect to the use of different clustering algorithms since the maximum variation of both class-wise and overall accuracy is of about $\pm.03\%$ and $\pm.0029\%$ respectively.

B. High Resolution QuickBird Dataset

The second study area considered here is a high resolution 4-band QuickBird image of a typical suburban area of a city in India of size 2000×2000. The test site is an urban fringe consisting of Lake, Pool, Vegetation, Field, Road, Shadow, Bright Roof, Dark Roof and Mountain. Test samples have been collected for all the classes and the corresponding reference map is prepared for the validation purpose. Figure 1(b) depicts the near-infra-red band of the study area. The cluster mapping technique is successful in identifying similar clusters from the three clustering results given all the land-cover classes are more or less overlapped with each other in the spectral domain. The clustering accuracies are mentioned in Table II.

TABLE I: Comparison of the class-wise and overall accuracies in % (Sardinia Dataset).

Clusters	samples	KK-means	Ncut	FCM	K-medoid	Method of [3]	Proposed method
Cluster1 (Pasture)	470	53.80	56.65	44.12	62.77	54.55	72.37 $\pm$1.97
Cluster2 (Forest)	128	93.71	93.08	94.33	91.67	94.70	98.72 $\pm$0.00
Cluster3 (Urban)	408	92.43	95.69	90.91	88.90	87.29	90.14 $\pm$1.60
Cluster4 (Water)	804	100.00	100.00	100.00	100.00	100.00	100.00 $\pm$0.00
Cluster5 (Vineyard)	179	64.10	52.47	60.68	57.65	70.14	65.97 $\pm$1.55
Cluster6 (Burned area)	176	96.65	95.65	96.52	94.78	85.47	96.55 $\pm$0.00
Overall	2165	**79.04**	**74.32**	**72.85**	**74.26**	**81.62**	**86.45 $\pm$0.29**

TABLE II: Comparison of the class-wise and overall accuracies in % (QuickBird Dataset).

Land Cover	samples	KK-means	Ncut	FCM	K-medoid	Method of [3]	Proposed method
Cluster1 (Lake)	729	81.44	83.48	86.18	80.54	83.14	93.81 $\pm$0.00
Cluster2 (Pool)	891	95.89	96.88	91.33	93.70	92.25	96.31 $\pm$0.00
Cluster3 (Vegetation)	93	85.37	85.69	84.19	84.17	82.51	89.95 $\pm$0.00
Cluster4 (Field)	36	90.81	89.79	89.78	91.29	85.23	96.59 $\pm$0.00
Cluster5 (Road)	94	62.28	59.43	72.61	62.28	61.62	86.45 $\pm$0.00
Cluster6 (Shadow)	264	95.12	92.28	95.07	95.12	93.17	96.15 $\pm$0.00
Cluster7 (Bright Roof)	489	90.88	91.18	95.33	92.44	84.17	95.28 $\pm$0.00
Cluster8 (Dark Roof)	100	66.74	58.11	68.18	63.31	68.78	79.15 $\pm$0.00
Cluster9 (Mountain)	150	69.51	73.42	67.21	67.22	71.16	84.83 $\pm$0.00
Overall	2864	**82.27**	**81.41**	**84.12**	**81.39**	**80.22**	**91.41 $\pm$0.00**

It can be observed from the results that the proposed method produces an improvement of more than 7% in the clustering accuracy (that becomes 91.41 % in average) compared to the best individual clustering result (i.e., 84.12%). Particularly, subtle enhancement can be observed for the Road, Dark Roof and the Mountain classes which have substantial overlapping in different spectral bands. The performance of the method of [3] is 80.22% due to the inability of K-means in handling overlapping data irrespective of centroid initializations. In addition, the proposed method demonstrated to be robust with respect to the use of different clustering algorithms since the overall accuracy is highly stable when the clustering ensemble changes with negligible variation in the class-wise accuracies.

The performance of a supervised ML classifier on a set of labeled samples per class produces a classification accuracy of 97.23%. The details of the training set used for this purpose can be obtained in [8]. The proposed unsupervised ML classifier comes close to the upper bound.

IV. Conclusion

This letter proposes an unsupervised classification method for the multispectral remotely sensed images. A novel consensus clustering technique is proposed to map the results of several clustering algorithms. The outcome of the cluster ensemble is further used to initialize the class-wise statistical parameters for a self-learning based ML classifier. Initially the image is clustered using three diverse clustering methods. The proposed cluster mapping technique imposes consistency to the different clustering outcomes. A set of highly reliable samples per cluster are selected to initialize the cluster parameters for a EM based parameter retraining method considering the image as a mixture of Gaussian functions. Experimental results prove that the proposed framework is invariant to the underlying clustering techniques and can correspond correct clusters given that the clustering algorithms are efficient in detecting the clusters substantially. The subsequent ML+EM based classification would go totally wrong if the cluster mapping fails to provide correct mappings for all the clusters. The performance of the proposed system comes very close to the one of the supervised ML classification but without the need of costly training data collection. It is scalable to larger datasets containing many clusters as the running time of the proposed ensemble technique is quadratic to the number of clusters (for finding the inter-cluster divergence). A graph based method for label matching for all the clustering results is now considered as a future endeavor of the current work.

Acknowledgment

The work was carried out under the framework of India Trento Program for Advance Research (ITPAR) during a stay of Mr. Biplab Banerjee at the University of Trento (Italy).

References

[1] M. Tyagi, F. Bovolo, A. K. Mehra, S. Chaudhuri, and L. Bruzzone, "A context-sensitive clustering technique based on graph-cut initialization and expectation-maximization algorithm," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 1, pp. 21–25, 2008.

[2] N. Alajlan, N. Ammour, Y. Bazi, and H. Hichri, "A cluster ensemble method for robust unsupervised classification of vhr remote sensing images," in *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*. IEEE, 2011, pp. 2896–2899.

[3] Z.-H. Zhou and W. Tang, "Clusterer ensemble," *Knowledge-Based Systems*, vol. 19, no. 1, pp. 77–83, 2006.

[4] M. Chi, Q. Qian, and J. A. Benediktsson, "Cluster-based ensemble classification for hyperspectral remote sensing images," in *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*, vol. 1. IEEE, 2008, pp. I–209.

[5] L. Bruzzone and D. F. Prieto, "Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 39, no. 2, pp. 456–460, 2001.

[6] V. Cherkassky and F. M. Mulier, *Learning from data: concepts, theory, and methods*. Wiley. com, 2007.

[7] K. Bahirat, F. Bovolo, L. Bruzzone, and S. Chaudhuri, "A novel domain adaptation bayesian classifier for updating land-cover maps with class differences in source and target domains," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 50, no. 7, pp. 2810–2826, 2012.

[8] I. A. Rizvi and B. K. Mohan, "Object-based image analysis of high-resolution satellite images using modified cloud basis function neural network and probabilistic relaxation labeling process," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 12, pp. 4815–4820, 2011.