A NOVEL GRAPH BASED FUZZY CLUSTERING TECHNIQUE FOR UNSUPERVISED CLASSIFICATION OF REMOTE SENSING IMAGES

Biplab Banerjee*, B. Krishna Mohan

Satellite Image Analysis Lab,
Center of Studies in Resources Engineering (CSRE),
Indian Institute of Technology Bombay,
Mumbai, India.
{biplab.banerjee,bkmohan}@iitb.ac.in

KEY WORDS: Clustering, Image segmentation, Maximum likelihood classifier, Expectation maximization

ABSTRACT:

This paper addresses the problem of unsupervised land-cover classification of multi-spectral remotely sensed images in the context of self-learning by exploring different graph based clustering techniques hierarchically. The only assumption used here is that the number of land-cover classes is known a priori. Object based image analysis paradigm which processes a given image at different levels, has emerged as a popular alternative to the pixel based approaches for remote sensing image segmentation considering the high spatial resolution of the images. A graph based fuzzy clustering technique is proposed here to obtain a better merging of an initially over-segmented image in the spectral domain compared to conventional clustering techniques. Instead of using Euclidean distance measure, the cumulative graph edge weight is used to find the distance between a pair of points to better cope with the topology of the feature space. In order to handle uncertainty in assigning class labels to pixels, which is not always a crisp allocation for remote sensing data, fuzzy set theoretic technique is incorporated to the graph based clustering. Minimum Spanning Tree (MST) based clustering technique is used to over-segment the image at the first level. Furthermore, considering that the spectral signature of different land-cover classes may overlap significantly, a self-learning based Maximum Likelihood (ML) classifier coupled with the Expectation Maximization (EM) based iterative unsupervised parameter retraining scheme is used to generate the final land-cover classification map. Results on two medium resolution images establish the superior performance of the proposed technique in comparison to the traditional fuzzy c-means clustering technique.

1. INTRODUCTION

Satellite image analysis has attained extensive popularity in the recent past with the advent of several high performance new-age sensors with high spatial and spectral properties. These images are of great importance in diverse application domains including Environmental Monitoring, Urban Planning, Extraction of Regions of Interest (ROI) from the Earth Surface etc. These applications require the proper extraction of the land-cover information from the images for further analysis. Image segmentation is a useful tool in this respect (Pal and Pal, 1993) which can extract image regions employing learning based or computer vision based techniques. Clustering is the most popular unsupervised land-cover classification technique cited in the literature (Jain et al., 1999) .

Clustering is inherently an ill-posed problem in the sense that, given a set of data points sampled from many groups, different clustering solutions are equally plausible with no prior knowledge about the underlying probability distribution of the data. The clustering algorithms assume some model to describe the data. If the data model does not match with the actual distribution of the data, the clustering result becomes erroneous. Moreover, many clustering algorithms require some initial estimations of some of the inherent cluster parameters (mean, variance, etc.) implicitly or explicitly. An improper initialization may lead to a non-reliable clustering outcome.

Clustering methods like K-means, Fuzzy c-means, density based clustering etc. have been used successfully in segmenting remote sensing data in the past (Saha et al., 2012) (Rekik et al., 2006). Graph based clustering techniques (Felzenszwalb and Huttenlocher, 2004) are better than these traditional clustering algorithms as graph topology can capture the spatial distribution of the data well which is important in clustering non-linearly separable datasets including remote sensing images. Given a graph of the input image pixels, the graph based pixel clustering is posed as the graph-cut problem which aims at removing the set of inconsistent edges from the graph which span different clusters. The only disadvantage of the graph based clustering method is its high resource utilization in handling large volume of data. The size of the graph Laplacian matrix grows rapidly and it subsequently makes the graph-cut problem even more difficult. In object based framework, graph-cut based techniques are preferred as the merging step to merge the regions of an over-segmented image which is much less in number compared to the number of image pixels.

Spanning tree is a reduced, acyclic version of a given graph which is minimally connected in the sense that the removal of any edge from a spanning tree leaves the tree disconnected. Clustering using MST is simpler compared to the approximation algorithms like the spectral clustering or brute-force approaches used to cluster a general graph as it is easy to identify the set of inconsistent edges in a spanning tree (Banerjee et al., 2014).

Traditional graph based clustering methods like min-cut, normalized cut, MST based clustering etc. are predominantly crisp clustering techniques where a given data point is assigned a class label with a membership degree of either 0 or 1. This assumption is many times vague in clustering remote sensing image pixels as it is difficult to properly estimate the class label of a given pixel in remote sensing environment considering the differences in the spatial resolution of the sensors. Hence, in segmenting satellite images, a proper combination of the graph topology with fuzzy

*Corresponding author.

techniques is expected to enhance the results.

Object based segmentation techniques for remote sensing images (Buddhiraju and Rizvi, 2010) first generates an over-segmented version of the original image. In the second level, these large number of small objects are merged properly to obtain a stable segmentation. This hierarchical segmentation scheme allows to explore different regional properties of the small regions in merging them which is not possible in pixel based techniques. It also reduces the uncertainty related to the pixel label assignment which many pixel based techniques are unable to handle properly.

Furthermore, in spite of processing the images at different levels using object based approach, another problem in properly segmenting remote sensing images arises considering the fact that the spectral signature of many land-cover classes overlap drastically. It is difficult for any clustering technique to cope with this situation. As a remedy to this problem, self-learning classifiers can be employed to refine the segmentation result to some extent. Self-learning essentially means that the training points for modeling the classifier are selected automatically and the true classifier free parameters are obtained by a sophisticated parameter retraining algorithm like EM (Bruzzone and Prieto, 2001).

With this background, the proposed object-based unsupervised land-cover classification technique of remote sensing images can be summarized using the following steps:

- Obtain an initial over-segmentation of the input multi-spectral image using minimum spanning tree (MST) based clustering using spectral features.
- Perform the proposed graph based fuzzy clustering technique for merging the regions. It gives an estimation of the class-labels of the image pixels.
- Select a set of highly reliable samples per cluster. Considering that a given land-cover class can be modeled by a Mixture of Gaussian (GMM) functions, the iterative EM algorithm is used to obtain a true estimate of the model parameters. EM is initialized from the selected highly reliable samples per cluster.
- The final land-cover map is obtained by an ML classifier built based on the updated parameter sets obtained from the EM stage.

The letter is organized as follows. Section II and Section III describe the self-learning based classification problem in the current context and the proposed solution respectively. Experimental results are discussed in Section IV. The letter concludes in Section IV with references to the future endeavor based on this work.

2. UNSUPERVISED LAND-COVER CLASSIFICATION OF MULTI-SPECTRAL REMOTE SENSING IMAGES IN THE CONTEXT OF SELF-LEARNING

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 $x_{r,c} \in R^d$, i.e., each pixel can be represented using d spectral bands. Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ represent N land-cover classes characterizing the geographical area represented by image X. Let us also assume that N is known a priori, whereas the class labels are not. 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:

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

$P(\omega_l)$ 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 phase of the Bayes classifier consists of estimating the true prior probability and conditional probability function that describe each land-cover class. This requires some highly reliable samples for each land-cover class to be identified for the estimation of the underlying statistical distribution of the class. Since the true distribution of a given class is unknown, a common practice is to fit some known distribution like Gaussian, Poisson function, etc. to characterize the density function of the class.

In remote sensing literature, it is well-accepted to represent each land-cover with a multi-variate Gaussian function. Hence, for each land-cover class ω_i, the set of parameters to be estimated is $\theta_i = \{\mu_i, \Sigma_i, P(\omega_i)\}$ where μ_i and Σ_i represent the mean and the covariance matrix of the pixels of ω_i and $P(\omega_i)$ is the prior probability of ω_i.

In the current setup, no labeled training data are available. These samples are selected automatically following the application of two graph based clustering techniques hierarchically to the input image. The initial values of the model parameters for the PDF's are estimated from these automatically selected samples and adopted to the entire image space by using the iterative EM algorithm. The ML classifier modeled with the updated parameters of (1) is further used to generate final unsupervised classification result of X.

3. THE PROPOSED METHODOLOGY

The proposed unsupervised land-cover classification algorithm of multi-spectral remote sensing images has four major steps:

1. Apply MST based clustering to the pixels of X in the spectral domain into K groups where $K >> N$. K is selected to be very large with respect to N to ensure the over-segmentation of X.
2. A complete graph G is formed considering the regions found in the previous step as nodes. A novel graph-fuzzy clustering algorithm is used further to perform merging of those regions. The output of this step is an approximation of the land-cover classification of X.
3. A set of highly reliable samples per cluster obtained in the previous stage are identified. Considering X as the Mixture of Gaussians, the mean, co-variance matrix and the class prior probabilities for each land-cover class are initialized from this set of samples which are further adapted to the entire X by iterative EM algorithm.
4. An ML based classifier modeled on the updated parameters is used to produce the final classification.

3.1 Obtain the initial over-segmentation of X

Given $R \times S$ pixels of X, a minimum spanning tree $T\{V_T, E_T\}$ is constructed considering the pixels as the nodes and the Euclidean distance between the pixels as the edge weight connecting two nodes using Prims algorithm (Graham and Hell, 1985). T is undirected, acyclic and the nodes in T are minimally connected to each other in the sense that the removal of an edge from T makes T disconnected. In order to generate K clusters of X from T, the following steps are followed:

- Sort the edges of T in descending order of the corresponding edge weights.
- Delete top K edges. It will generate $K+1$ sub-trees where the pixels (nodes) of each sub-tree correspond to a cluster.
- X is now over-segmented into K clusters. Several objects of each cluster are likely to be present in the over-segmented image.
- Let us consider $R_{\text{over}} = \{R_1, R_2, \ldots, R_M\}$to represent the regions found from this step.

As $K >> N$, a proper merging step is needed to alleviate the problem due to over-segmentation. The proposed graph based fuzzy clustering technique performs this task.

3.2 Region merging using a novel graph based fuzzy clustering

A complete graph $G(V, E)$ is built considering $V = \{R_i\}_{i=1}^{M}$ where a given R_i represents the set of pixels belonging to that particular region. The weight of the edge between a pair of nodes (R_i, R_j) which are Gaussian distributed and (μ_i, Σ_i) and (μ_j, Σ_j) defining the mean vectors and the covariance matrices of them, is specified using the Euclidean distance between their centroids. A small distance value indicates better matching between the corresponding regions, i.e. both the regions have high probability of belonging to the same land-cover class. Because it is already assumed that X contains N land-cover classes, the goal of this step is to cluster G into N groups. The proposed clustering scheme is discussed below.

3.2.1 Initialization of the fuzzy membership co-efficient for each node in G Given $G(V, E)$, the proposed method first initializes the fuzzy membership of each node (region) in V in the range $[0, 1]$ randomly. Let $U_{N\times M} = \{\alpha_{ki}\}$ $(1 \leq k \leq N, 1 \leq i \leq M)$ represent the membership of each node(region) in each of the N clusters.

3.2.2 Clustering step The k^{th} cluster centroid $(1 \leq k \leq N)$ is approximated as:

$$C_k = \frac{Mi=1\sum \alpha_{ki}^m \mu_i}{Mi=1\sum \alpha_{ki}^m} \tag{2}$$

where μ_i is the centroid of the pixels in R_i and m is the degree of fuzziness. Subsequently, the new values in U are calculated as:

$$\alpha_{ki} = \frac{1}{Mj=1\sum (\frac{||\mu_i - C_k||}{||\mu_i - C_j||})^{\frac{2}{m-1}}} \tag{3}$$

The distances $||\mu_i - C_k||$ or $||\mu_i - C_j||$ are not the usual L_2-norm used in the traditional case. The distance is defined as the minimum sum of edge weights connecting the nodes representing R_i and the pseudo node representing the center C_k or C_j. G is modified to accommodate the centroids. The well-known Floyd-Warshall algorithm (Dreyfus, 1969) is used to find the shortest path between all the pairs of nodes in G. 1-nearest-neighbor rule is followed next to assign each region to one of the clusters.

These two steps are carried out iteratively until the convergence criteria (discussed below) is satisfied.

3.2.3 Convergence of the clustering algorithm The proposed method converges when a cost function (F) is minimized. F is composed of two parts:

- Sum of the mean edge weights of every cluster (F_1).
- Sum of mean weights of the edges spanning different clusters (F_2).

F is defined as:

$$F = \frac{F_1}{F_2} \tag{4}$$

In order to check for the minimization of F, the value of F is evaluated for each iteration of the clustering algorithm. if $F(\text{iter}+1) \geq F(\text{iter})$ for any iteration iter, then the algorithm stops and the clustering outcome of the iter^{th} is considered to be the final result.

Algorithm 1 describes the proposed graph based fuzzy clustering scheme.

Algorithm 1 Input: N, $\{R_i\}_{i=1}^{M}$
Output: The grouping of $\{R_i\}_{i=1}^{M}$ into N clusters

1: Initialize $U_{N\times M}$ randomly in the range $[0, 1]$.
2: Construct a complete graph $G(V, E)$ considering the regions in $\{R\}_{i=1}^{M}$ as nodes and the Euclidean distance between the centroids of a pair of regions as the edge weight between them.
3: Update the cluster centers and the fuzzy membership matrix according to (2) and (3). The distance between a pair of nodes is calculated in accordance with the method discussed in Section 3.2.2.
4: Apply 1 nearest-neighbor ranking to associate each node with one of the clusters.
5: Repeat 3-4 until the cost function in (4) is minimized.

3.3 Identification of the set of reliable samples per cluster

The specific set of samples which are very close to the centroid of the each cluster represents the set of highly reliable samples for the cluster. To select these highly reliable set of samples, the maximum pairwise Euclidean distance among the samples of the cluster 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. Same process is repeated for all the clusters. A small δ provides more reliable samples. Let $\text{Tr} = \{\text{Tr}_1, \text{Tr}_2, \ldots, \text{Tr}_N\}$ denote the set of reliable samples for each cluster found in this step.

3.4 Final land-cover classification using EM and ML

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. 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:

$$P_i^{l+1}(\omega_i) = \frac{1}{R \times S} \sum_{x_{r,s} \in X} \frac{P^l(\omega_i) p^l(x_{r,s}|\omega_i)}{P^l(x_{r,s})} \tag{5}$$

$$\mu_i^{l+1} = \frac{\sum_{x_{r,s} \in X} \frac{P^l(\omega_i) p^l(x_{r,s}|\omega_i)}{P^l(x_{r,s})} x_{r,s}}{\sum_{x_{r,s} \in X} \frac{P^l(\omega_i) p^k(x_{r,s}|\omega_i)}{P^l(x_{r,s})}} \quad (6)$$

$$\Sigma_i^{l+1} = \frac{\sum_{x_{r,s} \in X} \frac{P^l(\omega_i) p^l(x_{r,s}|\omega_i)(x_{r,s} - \mu_i^{l+1})^2}{P_{l(x_{r,s})}}}{\sum_{x_{r,s} \in X} \frac{P^l(\omega_i) p^l(x_{r,s}|\omega_i)}{P^l(x_{r,s})}} \quad (7)$$

l represents the l^{th} iteration. In EM, at each iteration, the estimated new values of the parameters provide an increase of the negative log likelihood function until a local maxima is reached. Once the updated θ are 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.

4. EXPERIMENTAL RESULTS

$K = 50$ is considered to over-segment the image initially using the MST based clustering technique. $\delta = 25$ is considered Section 3.3 to point to the highly reliable set of samples for each cluster. Though, results on medium resolution images is exhibited here, the proposed technique can be extended to Very High Resolution (VHR) images without any modification.

4.1 Medium resolution Indian Remote Sensing Satellite (IRS) dataset

The first study area considered is a 1024×1024 image of Thane area, Mumbai, Maharashtra, India. The image was captured by Indian Remote Sensing Satellite 1C LISS III. The Near Infra-Red (NIR) band of the image is shown in Figure 1 and it has a spatial resolution of 23.8m $\times$23.8m. 7 land-cover classes are identified from the image, e.g. Water, Vegetation, Forest Vegetation, Settlements, Swamp, Hilly areas and some other classes. Test samples are collected from these classes for the experimental purpose and the reference class labels are used to assess the performance of the proposed technique. 100 samples per class are selected for this purpose. The result of the proposed method is compared with the framework where the proposed graph based fuzzy clustering is replaced by FCM.

The scatter plot of the reference test data along with the actual class labels are shown in Figure 2. The scatter plot for the proposed method and the proposed framework with FCM are depicted in Figures 3-4.

It can be seen from the scatter plots that, the classification result of the proposed method agrees well with the ground reference except for a few pixels of the Settlement class which are wrongly classified to Some other classes, , i.e.the producer accuracy of the proposed algorithm for the set of test samples is 100% for 6 classes. For the class tagged as some other classes, the Producer accuracy is 99%.

On the contrary, FCM is unable to detect the Water class. As observed from Figure 4, Water and Forest Vegetation classes are merged. This is the problem of using the Euclidean distance in the feature space. The proposed graph-fuzzy based clustering approach removes the problem entirely. The application of a supervised ML classifier on a set of reliable training samples produces a generalization accuracy of 100% on the test samples. The proposed method almost touches that upper bound.

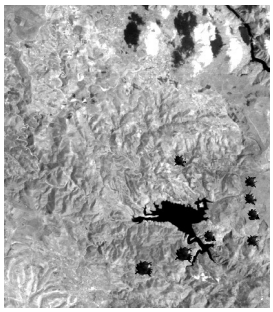

Figure 5: The band 4 of the the simulated Sardinia Dataset

4.2 Sardinia dataset

The second study area considered in the experiments is 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 5 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 (Bahirat et al., 2012). Test samples are available for all the classes and the corresponding reference map is used to assess the clustering accuracy. The scatter plot of the test samples with the reference map is shown in Figure 6.

It can be observed from the scatter plot that Pasture, Vineyard and Urban classes are highly overlapped in all the spectral bands. FCM with Euclidean distance is unable to produce a good clustering result in this respect. However, certain improvement in the overlapped classes in term of the Producer accuracy is noticed with the proposed technique. For Pasture, the proposed method has an enhancement of 13% in the classification accuracy compared to the result of FCM. Similar enhancement is also observed for Vineyard class (8%). The overall producer accuracy of the proposed technique is 83.32% which is better than the result with the FCM clustering (72.85%) and is close to the classification performance of a supervised ML classifier trained on some manually selected set of samples (87.45%).

Table 1: Comparison of the class-wise Producer accuracies on the Sardinia Dataset

Cluster (Underlying Land Cover)	# of samples	FCM based clustering (%)	Proposed method(%)
Cluster1 (Pasture)	470	44.12	67.61
Cluster2 (Forest)	128	93.71	93.08
Cluster3 (Urban)	408	92.43	90.69
Cluster4 (Water)	804	100.00	100.00
Cluster5 (Vineyard)	179	60.68	68.31
Cluster6 (Burned area)	176	96.65	95.65
Overall	2165	**72.85**	**83.32**

5. CONCLUSION

A hierarchical unsupervised land-cover classification technique for multi-spectral remote sensing images is proposed in this correspondence. The proposed method parses the given image at different level. An over-segmented version of the image is initially generated by using a MST based clustering technique. A graph based fuzzy clustering technique is proposed to merge the regions found in the previous stage efficiently. The proposed clustering method ensembles the advantages of fuzzy membership and graph cut together and the resulting algorithm is efficient in handling remote sensing images where it is difficult to

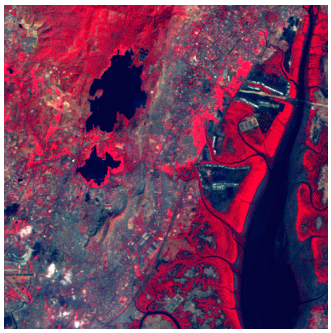

Figure 1: The NIR band of the IRS Mumbai image

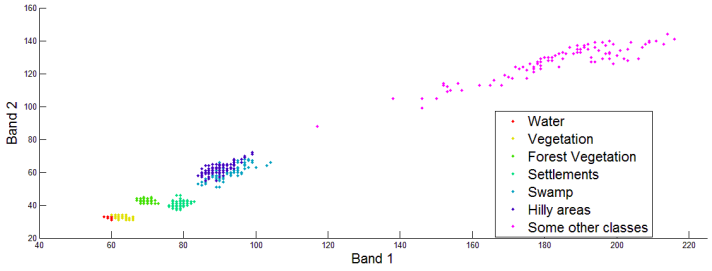

Figure 2: The scatter plot of the test samples with reference labels for the IRS data

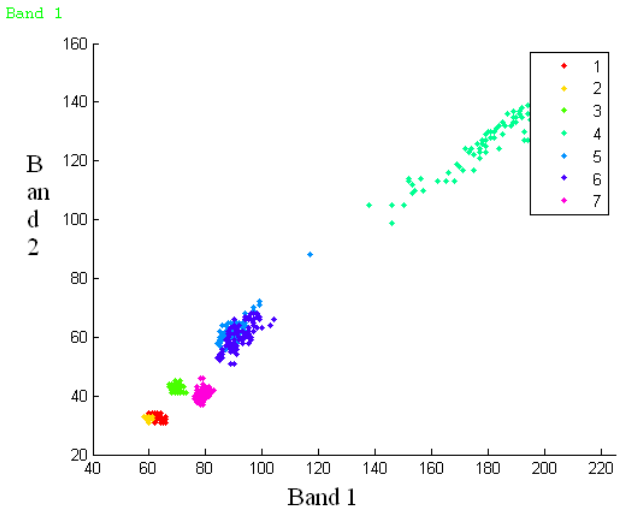

Figure 3: The scatter plot of the test samples for the proposed method for the IRS data

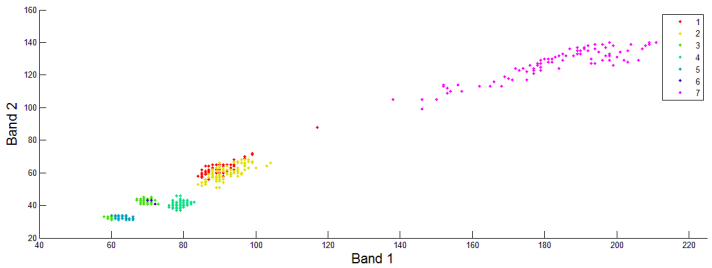

Figure 4: The scatter plot of the test samples for the graph-fuzzy clustering replaced by FCM for the IRS data

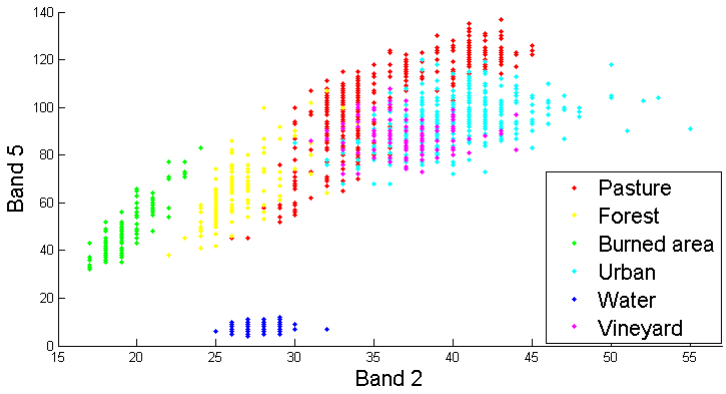

Figure 6: The scatter plot of the test samples with the reference labels for the Sardinia data

correctly predict the class labels of the pixels with crisp membership values. Considering that the spectral properties of many land-cover classes overlap significantly, a EM+ML based self-learning classifier is used in the post-processing step to generate the final classification map. The performance of the proposed technique is close to the one produced by the supervised classifier trained on the manually selected set of labeled samples without the need to work with costly training samples. The application of ensemble clustering technique in this framework is the future mode of research.

6. ACKNOWLEDGEMENT

The authors are grateful to Prof. Lorenzo Bruzzone, University of Trento for sharing the Sardinia dataset.

REFERENCES

Bahirat, K., Bovolo, F., Bruzzone, L. and Chaudhuri, S., 2012. 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 50(7), pp. 2810–2826.

Banerjee, B., Varma, S., Buddhiraju, K. M. and Eeti, L. N., 2014. Unsupervised multi-spectral satellite image segmentation combining modified mean-shift and a new minimum spanning tree based clustering technique. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of 7(3), pp. 888–894.

Bruzzone, L. and Prieto, D. F., 2001. Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images. Geoscience and Remote Sensing, IEEE Transactions on 39(2), pp. 456–460.

Buddhiraju, K. M. and Rizvi, I. A., 2010. Comparison of cbf, ann and svm classifiers for object based classification of high resolution satellite images. In: Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International, IEEE, pp. 40–43.

Dreyfus, S. E., 1969. An appraisal of some shortest-path algorithms. Operations research 17(3), pp. 395–412.

Felzenszwalb, P. F. and Huttenlocher, D. P., 2004. Efficient graph-based image segmentation. International Journal of Computer Vision 59(2), pp. 167–181.

Graham, R. L. and Hell, P., 1985. On the history of the minimum spanning tree problem. Annals of the History of Computing 7(1), pp. 43–57.

Jain, A. K., Murty, M. N. and Flynn, P. J., 1999. Data clustering: a review. ACM computing surveys (CSUR) 31(3), pp. 264–323.

Pal, N. R. and Pal, S. K., 1993. A review on image segmentation techniques. Pattern recognition 26(9), pp. 1277–1294.

Rekik, A., Zribi, M., Benjelloun, M. and Ben Hamida, A., 2006. A k-means clustering algorithm initialization for unsupervised statistical satellite image segmentation. In: E-Learning in Industrial Electronics, 2006 1ST IEEE International Conference on, IEEE, pp. 11–16.

Saha, I., Maulik, U., Bandyopadhyay, S. and Plewczynski, D., 2012. Svmefc: Svm ensemble fuzzy clustering for satellite image segmentation. Geoscience and Remote Sensing Letters, IEEE 9(1), pp. 52–55.