Gaussian mixture model and Markov random fields for hyperspectral image classification

Hamid Ghanbari a, Saeid Homayouni b, Abdolreza Safari a and Pedram Ghamisi c

aSchool of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran; bDepartment of Geography, Environment, and Geomatics, University of Ottawa, Ottawa, Canada; cHelmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology, Exploration, Freiburg, Germany

ABSTRACT
This paper presents a novel method for reliable and efficient spatial-spectral classification of hyperspectral data. This algorithm is based on the Bayesian labelling by combining the results of the Gaussian mixture model (GMM) with spatial-contextual information extracted by Markov random fields (MRF). Moreover, a new fuzzy segmentation-based function was defined and incorporated into the spatial energy involved to improve the performance of MRF. To evaluate the proposed algorithm in real analysis scenarios, three benchmark hyperspectral datasets, i.e. Indian Pines, Pavia University and Salinas, were used. Experimental results demonstrated that the proposed method could considerably improve the classification's overall accuracies when compared to conventional MRF-based approaches.

INTRODUCTION

Over the past two decades, various hyperspectral image (HSI) classification methods have been developed for land cover mapping applications (Plaza et al., 2009). Most of those methods process each pixel independently without considering the correlations between spatially neighbouring pixels (Camps-Valls & Bruzzone, 2005). For HSI with excellent spatial resolution, it is very likely that two adjacent pixels belong to the same class of objects.

Gaussian mixture model (GMM) classifiers, as a popular classification tool, have gained significant attention in the remote sensing community; thanks to its capability of capturing non-Gaussian statistics of multivariate data and remarkable accuracy in classifying high-dimensional data (Benediktsson & Ghamisi, 2015). GMMs do not consider the spatial correlations and classify image data only based on their spectral information. Therefore, they do not employ the relevant information that can be extracted from the spatial domain.

Combining spectral and spatial information is beneficial for accurate image classification (J. Li, Bioucas-Dias, & Plaza, 2010). The primary goal here is to consider the spatial-contextual correlations of adjacent pixels, and combining them with spectral information, cooperatively in the classification scheme. This goal can be achieved by using various methods, such as extended attribute profiles (Ghamisi, Benediktsson, & Sveinsson, 2014), extinction profiles (Ghamisi et al., 2016), marker-based minimum spanning forest (Akbari, Homayouni, Safari, & Mehrshad, 2016), kernel-based methods (Camps-Valls, Gomez-Chova, Muñoz-Marí, Vila-Francés, & Calpe-Maravilla, 2006), flat zone area filtering (Jamshidpour, Safari, & Homayouni, 2017), morphological component analysis Multiple morphological component analysis based decomposition for remote sensing image classification and Markov random fields (MRFs) (Moser, Serpico, & Benediktsson, 2013). In particular, MRFs have received notable attention in the literature for incorporating contextual information into classification tasks. This is mainly because of their capability to work with many probability distributions in the sense that it can capture complex spatial dependencies between neighbouring pixels (Li, Bioucas-Dias, & Plaza, 2012; Moser & Serpico, 2013; Zhao, Zhang, Li, & Huang, 2007).

MRFs are a family of probabilistic models for integrating spatial-spectral information into image classification problems. They have been applied recently to exploit the spatial information, in which neighbouring pixels tend to be of the same class. The underlying assumption in the MRF is that, in HSIs, it is most likely that two adjacent pixels have the same class label. In the MRF framework, the maximum a posteriori (MAP) decision rule is usually formulated as the minimization of an appropriate energy function (Solberg, Taxt, & Jain, 1996). MRF-based classifiers are well-known and frequently used approaches in various remote sensing applications. In particular, Li, Prasad, and Fowler (2014) investigated the integration
of a GMM method within an MRF framework for accurate HSI classification. In this context, they have considered a GMM classifier based on local Fisher’s discriminant analysis (LFDA) and locality-preserving nonnegative matrix factorization (LPNMF)-induced feature subspace. Then, the GMM-based pixel-wise classification results were subsequently followed by an MRF-based technique that integrates the spatial information. Tarabalka, Fauvel, Chanussot, and Benediktsson (2010) developed a Bayesian-like processing results were derived from SVM-based discriminant functions and combined with MRF models of the contextual information. In another research, it was shown that the MRF model has the advantage of working with various probability distributions (Li, Bioucas-Dias, et al., 2012). In that work, the posterior class densities exploited by a subspace multinomial logistic regression (MLR) classifier and the corresponding spatial contextual information was integrated. This information is generated by MRF-based multilevel logistic prior to a combinatorial optimization problem and solves this MAP estimation problem by a graph cut algorithm (Li, Bioucas-Dias, et al., 2012). Recently, a novel and rigorous supervised method have been proposed for the classification of multi-resolution images (i.e. a high spatial resolution panchromatic channel and several coarse spatial resolution multispectral channels) using an MRF (Moser, De Giorgi, & Serpico, 2016). The method iteratively exploits a linear mixture model to characterize the relationships between data at different resolutions and uses a graph cut approach to optimize Markovian energy for the contextual classification.

Another group of spatial-spectral classification approaches is based on segmentation algorithms (Ghamisi, Benediktsson, & Ulfarsson, 2014; Tarabalka et al., 2010). In these algorithms, an HSI is first segmented into homogeneous areas based on its intensity or texture properties. Then, the pixels within the same areas can be considered as a spatial neighbourhood. Afterward, majority voting is used for integrating the pixel-wise classification results with a segmentation map obtained by an image segmentation algorithm. For instance, an automatic framework for the classification of HSIs was employed in Ghamisi, Benediktsson, and Ulfarsson (2014). The method was based on combining Hidden MRF segmentation with the Support Vector Machines (SVM) classifier. In another research (Tarabalka, Benediktsson, & Chanussot, 2009), the authors proposed a new spectral-spatial classification scheme for HSIs that combines the results of a pixel-wise SVM classification and a segmentation map obtained by partitional clustering using majority voting. However, various spatial-spectral classifiers can be combined for obtaining high accuracies.

Hyperspectral data are highly correlated in adjacent spectral bands since the feature vectors are the discretization of continuous curves. Moreover, the impractically large size of resulting parameter space has hindered widespread adoption of GMM for HSI (Li et al., 2014). For example, if the HSI data are of d-dimensional, then the resulting dimensionality of the parameter space for a C-component GMM, assuming full covariance matrices, will be C(1 + d(1 + 1/2) + d). To deal with this problem, dimensionality reduction techniques have been suggested. It has been shown that the first projection of data might facilitate the learning of high-dimensional mixtures of Gaussians into a randomly chosen subspace of low dimension (Dasgupta, 2000). As a result, a random projection (RP) dimension reduction method has been used to represent the data in a lower-dimensional space. Dasgupta (2000) presented a comprehensive series of experiments intended to illustrate the benefits of this technique precisely. Moreover, LFDA, which has been introduced for hyperspectral data (Sugiyama, 2007), is an extension of FDA such that the class distributions are no longer restricted to be a unimodal Gaussian. Li, Prasad, Fowler, and Bruce (2012) demonstrated that appropriate LFDA preprocessing ensures that the GMM models learned in the HSI feature space do not have an unreasonably large number of free parameters to be estimated from the training data. LFDA obtains an excellent intra-class local structure in the projection while preserving the inter-class local structure at the same time.

This paper presents a novel GMM- and MRF-based (GMMRF) method for spectral-spatial classification of HSIs. In the first step, a GMM classifier, based on a low-dimensional feature space induced by either LFDA or RP, is applied to learn posterior probabilities from the spectral signatures. Then, spatial contextual information, which is achieved by using MRF regularization, is used for refining the classification results of the first step. An essential difference of the proposed method from previously proposed ones consists in applying a segmentation algorithm and defining a new “fuzzy segmentation-based” function into the spatial energy function involved in MRFs, aiming at improving classification efficiency while performing the spatial regularization.

**GMMRF classification scheme**

Figure 1 depicts an overview of the proposed classification method. The initial posterior probability that a pixel belongs to a specified class is estimated by performing the GMM classifier on the dimension reduced data. Taking spatial contextual information into account, the final classification map is obtained using the improved MRF regularization. The improved MRF regularization benefits from the “fuzzy segmentation-based” function, which is achieved by integrating the mean shift segmentation
and classification maps. In the following, the specific parts of the proposed framework will be discussed in detail.

As the input, an HSI, \( I \), made up of \( B \) two-dimensional arrays is given. This matrix is actually a set of \( n \) pixel vectors \( \mathbf{X} = \{ \mathbf{x}_i \in \mathbb{R}^B, i = 1, 2, \ldots, n \} \). Consider \( L = \{ l_1, l_2, \ldots, l_n \} \in \{ 1, 2, \ldots, C \} \) as a set of classes in the scene, where \( C \) is the number of classes. Classification consists in assigning each pixel to one of the \( C \) classes of interest. According to the Bayes rule, the posterior probability of the class labels \( L \) given data \( \mathbf{X} \)

\[
P(L|\mathbf{X}) = \frac{P(\mathbf{X}|L)P(L)}{P(\mathbf{X})},
\]

where \( P(\mathbf{X}) \) is a constant that does not affect the final result, \( P(L) \) can be viewed as the spatial prior and \( P(\mathbf{X}|L) \) is the likelihood function, which is derived from the data distribution. \( L \) minimizes \( P(L|\mathbf{X}) \) in Equation (1) is referred to as the MAP estimation of \( L \).

\[
\hat{L} = \arg\min_L \sum_{i=1}^{n} (-\log P(\mathbf{x}_i|l_i)) - \log P(l_i))
\]

The first and the second parts of the above equation are known as spectral energy and spatial energy, respectively. Equation (2) is a combinatorial optimization problem involving unary and pairwise information terms, which is difficult to compute. Several algorithms, such as iterated conditional modes (Besag, 1986), graph cuts (Boykov & Kolmogorov, 2004), tree-reweighted message passing (Kolmogorov, 2006) and \( \alpha \)-Expansion graph-cut-based (Li et al., 2012), have been proposed in the literature to solve this general optimization problem. In this paper, we used the Metropolis algorithm based on stochastic Simulated Annealing (SA) algorithm for computing the MAP estimate of the classification map (Geman & Geman, 1987). The Metropolis algorithm is guaranteed to converge to a global minimum and can significantly overcome the local minimum problem (Li, 2009).

**Local Fisher discriminant analysis and random projection**

LFDA provides supervised dimensionality reduction designed to handle non-Gaussian distributions (Sugiyama, 2007). LFDA combines the ideas of FDA and Locality preserving projection (LPP) (Li et al., 2012). FDA is a traditional linear approach for supervised dimensionality reduction, but it tends to give unexpected results if samples in some classes make several separate clusters (i.e. multimodal). LPP is an unsupervised dimensionality reduction method which can work well with multimodal data due to its locality preserving characteristic. LPP does not take the label information into account. Furthermore, LPP can make samples of different classes overlapped, if they are close to the original high-dimensional space. To overcome these problems, LFDA works in such a way that the local intra-class \( S^{(b)}(\mathbf{x}_i) \) and inter-class \( S^{(w)}(\mathbf{x}_i) \) scatter matrices are:

\[
S^{(w)} = \frac{1}{2} \sum_{i=1}^{n} A^{(w)}_{ij} (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_j - \mathbf{x}_i)^T,
\]

\[
S^{(b)} = \frac{1}{2} \sum_{i=1}^{n} A^{(b)}_{ij} (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_j - \mathbf{x}_i)^T,
\]

\( A^{(w)}_{ij} \) and \( A^{(b)}_{ij} \) are \( n \times n \), matrices which can be defined as

\[
A^{(w)}_{ij} = \begin{cases} A_{ij}/n_c & \text{if } l_i = l_j = c, \\ 0 & \text{if } l_i \neq l_j, \end{cases}
\]

\[
A^{(b)}_{ij} = \begin{cases} A_{ij}(1/n - 1/n_c) & \text{if } l_i = l_j = c, \\ 1/n & \text{if } l_i \neq l_j. \end{cases}
\]

where \( n_c \) is the number of available training samples for the \( c \)th class, \( \sum_{c=1}^{C} n_c = n \), and \( A \) is defined as the affinity matrix. Using \( S^{(w)}(\mathbf{x}_i) \) and \( S^{(b)}(\mathbf{x}_i) \), the LFDA transformation matrix \( T_{LFDA} \) is defined as:
\[ T_{LFDA} = \arg \min_{T_{LFDA} \in \mathbb{R}^d} \left( T_{LFDA}^T S^{(b)} T_{LFDA} \right)^{-1} \]

\[ T_{LFDA}^T S^{(b)} T_{LFDA}. \]  

LFDA has shown to yield better performance than FDA and LPP for dimensionality reduction for non-Gaussian or multimodal data.

RP, which has been introduced for hyperspectral data, is a general data reduction method (Dasgupta, 2000). In this method, the original high-dimensional data is projected into a low-dimensional subspace, without using the Eigen decomposition of matrices. As a result, it runs extremely fast and stable. The principal idea of RP comes from the Johnson-Lindenstrauss lemma (Frankl & Maehara, 1988). According to this lemma, if points in a vector space are projected onto a randomly selected subspace, the distances between these points are approximately preserved (Achlioptas, 2001). In RP, the \( B \)-dimensional data is projected into a \( d \)-dimensional \( (d < B) \) subspace. It can be done using a random matrix \( R_{d \times B} \) whose columns have unit lengths.

\[ X_{d \times N}^{RP} = R_{d \times B} X_{B \times N} \]  

In Equation (8), \( X_{B \times N} \) is the original set of \( B \)-dimensional observations. The choice of random matrix \( R \) is one of the key points of interest. The elements \( r_{ij} \) of \( R \) are often Gaussian distributed, but this does not need to be the case. Here, we used a sparse matrix that has been introduced by Achlioptas (Achlioptas, 2001) and distributed according to Equation (9).

\[ r_{ij} = \sqrt{3} \cdot \begin{cases} +1 & \text{with probability } \frac{1}{6} \\ 0 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{3} \end{cases} \]  

Bingham and Mannila (2001) showed that Achlioptas theoretical results have feasible significance. Moreover, RP is a data-independent and computationally efficient method that approximately preserves pairwise distances between the points of a set without introducing significant distortion. Hence, it is widely used in many dimension reduction applications that need to maintain the local structure of data and require immediate results.

**Gaussian mixture model classification**

GMM has been successfully used for HSI classification (Bingham & Mannila, 2001). It has also proved beneficial for a variety of classification tasks, such as speech and speaker recognition, clustering, etc. (Hassanpour, Shabahrami, & Wong, 2008). The multivariate finite mixture model can be obtained using the equation below:

\[ P(X_i | L) = \sum_{j=1}^{C} a_j \Phi \left( X_i; \mu_j, \sum_j \right), \]  

where \( a_j \) are the nonnegative mixing coefficients for the \( j \)th term, with sum-to-one constraint, i.e. \( a_1 + \ldots + a_C = 1 \). \( \Phi \left( X_i; \mu_j, \sum_j \right) \) denotes the normal probability density function with \( d \)-dimensional vector of means \( \mu_j \), and \( \sum_j \) as the covariance matrix. Note that in order to calculate the density estimate at a point \( x_i \), we only need to retain \( C-1 \) mixing coefficients, \( C \times d \) value for the means and \( (Cd(C+1))/2 \) value for the component’s covariance matrices. For estimating the parameters of GMM, the Expectation-Maximization (EM) algorithm is used.

**Expectation-Maximization algorithm**

EM is the process of finding an estimate of maximum likelihood when there are missing data (Figueiredo & Jain, 2002; Hasan & Gan, 2009). To use the EM algorithm, the number of terms \( C \) in the mixture must be known. This is usually obtained using a priori knowledge of the application, looking for groups of data using clustering methods or other methods of estimating the number of terms, such as adaptive mixtures approach (Priebe, 1994). In this paper, we already know the number of components of the GMM. Moreover, an initial guess for the values of the mean, variance and mixing coefficients is needed. The quality of the GMM model is sensitive to the initialization of the parameter space. For this purpose, the training samples which are randomly selected from the overall ground truth in different rates (1%, 5%, 10%, 15% and 20%) are used. After obtaining the initial estimation, the required parameters are updated using equations defined below. The posterior probabilities are given by:

\[ \hat{\pi}_j = \frac{\tilde{\alpha}_j \Phi \left( x_i; \tilde{\mu}_j, \sum_j \right) \hat{P}(x_i | L) }{ \hat{P}(x_i | L) }, \]

\[ \hat{P}(x_i | L) = \sum_{j=1}^{C} \tilde{\alpha}_j \Phi \left( x_i; \tilde{\mu}_j, \sum_j \right). \]

where, \( \hat{P}(x_i | L) \) is the finite mixture estimate at the point \( x_i \). Using the estimated posterior probability, the parameter of each component can be updated through Equations (13–15).

\[ \hat{\pi}_j = \frac{1}{n} \sum_{i=1}^{n} \hat{\pi}_j x_i \hat{\pi}_j \]  

\[ \hat{\mu}_j = \frac{1}{n} \sum_{i=1}^{n} \tilde{\pi}_j x_i \hat{\pi}_j \]
\[
\sum_{t} = \frac{1}{n} \sum_{i=1}^{n} \tilde{\tau}_{ij} \left( x_i - \mu_j \right) \left( x_i - \mu_j \right)^T a_j.
\] (15)

The procedure of EM algorithm for a finite mixture is as follows:

1. Determine the number of component densities \(C\).
2. Initialize the component parameters.
3. Calculate the posterior probability using Equation (11).
4. Update the component parameters using Equations (13–15).
   - Repeat steps (3) and (4) until the estimates converge.

**Mean shift segmentation**

Image segmentation is of crucial importance to early vision tasks where pixels with similar features are assembled into analogous regions. In recent years, researchers have developed many useful image segmentation algorithms (Zuva, Olugbara, Ojo, & Ngwira, 2011). In this paper, the mean shift algorithm as a low-level segmentation technique is used. The Mean Shift method segments an image into many visual homogenous areas. Based on these areas, more features can be involved which facilitates the subsequent image interpretation. Mean shift, as a simple iterative method which shifts each pixel to the average of pixels in its neighbourhood, was introduced by Fukunaga and Hostetler (1975), and generalized by Cheng (1995). Since then, this segmentation method has become a widely-used approach in the vision community. It is one of many techniques under the heading of “feature space analysis.” The mean shift algorithm consists of two main steps: a mean shift filtering of the original image data in the feature space and a succeeding clustering of the filtered data points. Mean shift algorithm is a popular segmentation algorithm because of its simplicity and flexibility. As with all clustering problems, there is no correct clustering. Instead, correct is usually defined by what seems reasonable given the problem domain and application. Mean shift provides one nice knob (the kernel bandwidth parameter) that can easily be tuned appropriately for different applications. There are many different types of kernels, but the most popular one is the Gaussian kernel. So, as it was iteratively suggested in the literature, we used the Gaussian kernel in our method. We tested the results of the experiments with different number of bandwidth, and the results of the proposed method showed to be less sensitive to the bandwidth parameter. The results showed that the optimal value of bandwidth is \([4–8]\). A detailed description of this approach can be found in Comaniciu and Meer (2002).

**MRF-based regularization**

In the field of machine vision, image analysis is usually defined as a modelling problem. MRF provides a consistent and convenient modelling algorithm with contextual-based entities or intrinsically linked with features. To improve the classification performance achieved by using only spectral information, we integrated the spatial contextual information with spectral information, by using an MRF-based regularization technique. In the MRF, a clique is defined either as a single term or a set of pixels that are neighbours. Here, an eight-neighbourhood clique is assumed (let \(N_j\) be the set of neighbours for a given pixel \(x_i\)). The most common form of the inference over the posterior MRF, in vision and image processing problems, is MAP estimation which is computed through Equation (2). In this paper, we used an MRF model based on (Tarabalka et al., 2010), thus, the MRF computes the MAP estimation via a Metropolis algorithm. This algorithm is based on the stochastic relaxation and annealing for computing the MAP estimation of the true classification image given the GMM classification map. The proposed method is based on the Bayesian approach and aims at minimizing the global energy in an image, by the iterative minimization of local spatial and spectral energies. These energies are associated with randomly chosen image sites, i.e. pixels. The local energy of a randomly chosen site associated with a pixel \(x_i\) was computed using Equation (16).

\[
E(x_i) = E_{\text{spectral}}(x_i) + E_{\text{spatial}}(x_i),
\] (16)

where, \(E_{\text{spectral}}(x_i)\) is the spectral energy and \(E_{\text{spatial}}(x_i)\) is the spatial energy function computed over the local neighbourhood \(N_i\). The spectral energy is defined as:

\[
E_{\text{spectral}}(x_i) = -\log P(x_i|l_i),
\] (17)

where, \(P(x_i|l_i)\) is the membership value of the pixel \(x_i\) in each of the classes’ fraction outputs. First, we consider the standard spatial energy, which is computed as:

\[
E_{\text{spatial}}(x_i) = \sum_{x_j \in N_i} \beta (1 - \delta(l_i - l_j)),
\] (18)

where, \(\delta(L)\) (i.e. \(\delta(0) = 1\) and \(\delta(L) = 0\) for \(L \neq 0\)) is the unit impulse function. \(\beta\) is the parameter which controls the level of smoothness. The spatial energy is suitable for an image with large spatial structures, such as agricultural lands. However, if a small object is presented in the image, this function tends to assign this object to the class of adjacent objects.

A reliable segmentation map can improve the spatial energy of the MRF regularization and help to overcome the problem mentioned above. The second expression of the spatial energy
\[ E_{\text{spatial}}^{S} \text{ is used to mitigate this drawback and to preserve the small structures and remove the noisy effects in the classification map. For this purpose, we use the result of the Mean Shift segmentation map. The output of this segmentation method is composed of some objects. Each object consists of several pixels with the same label. In other words, pixels in each object share the same characteristics. In order to further improve the MRF regularization, we investigate the benefits of integrating the segmentation map with the MRF model. Hence, we defined a new “fuzzy segmentation-based” function which is derived from the segmentation map and calculated as follows:} \]

\[ \varepsilon(x_i) = 1 - \frac{\rho_i}{N}, \quad (19) \]

where \( N \) is the number of pixels in each object and \( \rho_i \) is the number of pixels with the same label pixel of the \( x_i \). To employ the fuzzy segmentation function in the MRF regularization process, first, the object that a given pixel \( x_i \) fall into, is assigned using the output of the Mean Shift segmentation map. Then, the total number of pixels with any labels (\( N \)) and the set of pixels with the same label of that given pixel (\( \rho \)) in the object are counted. Thus, the fuzzy segmentation map is acquired through the Equation (19). The described technique leads to considerable improvement in terms of classification accuracy because it uses the local neighbourhoods by taking into account the spatial information in the classification process. Figure 2 represents an illustration of computing the “fuzzy segmentation-based” function. From here, the following spatial energy function is defined:

\[ E_{\text{spatial}}^{S}(x_i) = \sum_{x_j \in N_i} \beta \varepsilon(x_j) (1 - \delta(l_i - l_j)). \quad (20) \]

The superscript “\( S \)” means that the segmentation information is taken into account. In the following, we refer to two different methods, namely, GMMMRF and GMMMRF-S, when Equations (17) and (20) are used for computing the spatial energy, respectively. In each iteration, a pixel was randomly chosen. The local energy of this pixel \( E(x_i) \) was computed by Equation (16). Then a new class label \( l_{\text{new}} \) is selected for the point. This label is the most repetitive label in the corresponding segment, except the old label. If there is not any different label, the new label will be assigned randomly. A new label energy \( E_{\text{new}}(x_i) \) is also computed. If the variation of energy meets \( \Delta E = E_{\text{new}}(x_i) - E(x_i) < 0 \), the new class label will be assigned to \( x_i \) otherwise, the new class label assignment is accepted with probability \( p = \exp(-\Delta E) \).

**Image analysis**

**Hyperspectral datasets**

The proposed GMMMRF-S and GMMMRF approaches are experimentally evaluated using three real benchmark hyperspectral data sets recorded by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and the Reflective Optics System Imaging Spectrometer (ROSIS).

- **University of Pavia image** is an urban area that was collected by the ROSIS sensor with a spatial resolution of 1.3 m. The image is of 610 by 340 pixels and 103 spectral channels. The reference data contain nine classes of interest.

- **Indian Pines** dataset was acquired by AVIRIS sensor with a spatial resolution of 20 m/pixel. The image contains 145 × 145 pixels in 202 bands. This dataset contains 16 mutually exclusive classes and 10,366 labelled pixels. Sixteen classes of interest are considered, which are detailed in Table 1.

- **AVIRIS Salinas**, this well-known dataset was captured over Salinas Valley, California, which is investigated as the third hyperspectral data for the experimental tests. The image has 204 spectral bands and a spatial resolution of 3.7 m. The ground truth data contains 16 separated classes and a total of 54,129 labelled pixels.

**Figure 2.** The procedure of computing the \( \varepsilon(x_i) \) coefficient through the “fuzzy segmentation-based” function.
**AVIRIS Salinas**, this well-known dataset was captured over Salinas Valley, California, which is investigated as the third hyperspectral data for the experimental tests. The image has 204 spectral bands and a spatial resolution of 3.7 m. The ground truth data contains 16 separated classes and a total of 54,129 labelled pixels.

**Experimental results**

In all experiments, the EM algorithm was iterated 1000 times. As shown in Figure 3, we have investigated the performances of the proposed methods for different values of the context weight parameter $\beta$. The optimal value of $\beta$ for GMMRF is $\beta \in [1, 2]$ and for the GMMMRF-S is $\beta \in [2, 4]$.

LFDA and RP can be viewed as preprocessing for the proposed methods, which implies that reduced dimensionality is an essential parameter for LFDA-GMMMRF and RP-GMMMRF classifiers. The first experiment is performed on the Indian Pines dataset. To avoid any biases, all the experiments were repeated 10 times, and we have reported the averaged classification accuracies. Table 1 summarizes the global (Overall Accuracy (OA), Average Accuracy (AA) and a kappa coefficient of agreement) and class-specific classification accuracies for the Indian Pines image. In addition, for each data set, the standardized McNemar’s test was applied separately to the map of each method to assess whether the difference in accuracy between this map and the result of RP-GMMMRF-S is statistically significant.

$$Z = \frac{f_{12} - f_{11}}{\sqrt{f_{12} + f_{11}}}$$

where $f_{12}$ indicates the number of test samples that are correctly labelled by RP-GMMMRF-S and simultaneously not by the comparison method. Using the commonly accepted 5% level of significance, the difference between the results of RP-GMMMRF-S and comparison method is statistically significant if $|Z| > 1.96$ (Dietterich, 1998). If this condition is met, a positive or negative value of $Z$ indicates that the RP-GMMMRF-S or the compared method, is more accurate on the test set.

|                         | OA  | AA  | Kappa Coefficient | McNemar’s test | Alfalfa  | Corn-no till | Corn-min till | Corn | Grass-pasture | Grass-trees | Grass-pasture-mowed | Hay-windrowed | Oats | Soybean-no till | Soybean-min till | Soybean-clean | Wheat | Woods | Bldg-Grass-Trees-Drives | Stone-Steel-Towers |
|-------------------------|-----|-----|-------------------|----------------|----------|--------------|---------------|------|---------------|-------------|---------------------|---------------|------|----------------|----------------|--------------|--------|--------|----------------------|-------------------|
| LFDA-GMMRF              | 74.20 | 75.53 | 90.71             | 85.33          | 87.54    | 80.05        |
| RP-GMMRF               | 79.17 | 79.29 | 89.44             | 84.89          | 85.52    | 81.60        |
| RP-GMMMRF-S            | 73.71 | 74.32 | 87.41             | 82.22          | 83.28    | 74.64        |
| RP-GMMMRF              | 71.95 | 72.91 | 84.31             | 80.37          | 78.31    | 76.31        |
| LFDA-GMMMRF-S          | 64.05 | 67.10 | 77.35             | 72.29          | 76.39    | 69.29        |
| LFDA-GMMMRF            | 57.31 | 56.25 | 83.58             | 77.31          | 67.13    | 58.78        |

Table 1. Classification accuracy results for Indian Pines dataset with 15% training samples.

Figure 3. Accuracies for different values of the $\beta$ parameter with 15% training samples for methods (a) GMMRF and (b) GMMMRF-S.
The performance of the above methods with different numbers of training samples for Indian Pines, Salinas and Pavia University datasets is illustrated in Figure 4. The training samples are randomly selected from the overall ground truth in different rates (1%, 5%, 10%, 15% and 20%) for the three data sets. The remaining samples in each rate are used as test samples to evaluate the classification accuracies. As can be seen from Figure 4, the classification accuracies of all the classification methods share a similar trend as the number of training samples increases. The classification performances of all the methods gradually increase with the increase in training number. The classification methods incorporating MRF models (RP-GMMRF, RP-GMMRF-S, LFDA-GMMRF and LFDA-GMMRF-S) always outperform traditional RP-GMM and LFDA-GMM, in all cases, with a different rate of training samples. Furthermore, RP-GMMRF-S and LFDA-GMMRF-S methods incorporating fuzzy segmentation-based function alongside the MRF model perform better than RP-GMMRF and LFDA-GMMRF, respectively. The RP-GMMRF and LFDA-GMMRF-S share similar trends in term of overall accuracy and LFDA-GMMRF-S outperforms the RP-GMMRF method for Pavia University and Indian Pines data sets, but in the case of “Salinas,” RP-GMMRF performs better than the LFDA-GMMRF-S method. The RP-GMMRF-S outperforms other methods, in all conditions, with a

![Figure 4. Performance analysis for different rates of training samples, with the three data sets. (a) Indian Pines; (b) Pavia University; (c) Salinas data set.](image)

![Figure 5. Classification maps for Indian Pines dataset.](image)
different number of training and test samples and all three data sets.

Figure 5 shows the reference data and the classification maps for LFDA-GMM, RP-GMM, LFDA-GMMRF, LFDA-GMMRF-S, RP-GMMRF and RP-GMMRF-S for Indian Pines data set. As can be seen from Figure 5, the corresponding GMMRF-S and GMMRF classification maps are comparable and contain more homogeneous regions when compared to the GMM classification map. Since the considered image contains mixed pixels, the advantage of using the segmentation map in MRF process versus not using it, is evident here.

As it can be seen from Table 1, the spatial-spectral approaches yielded higher classification accuracies compared to the pixel-wise method. For instance, the overall accuracies of RP-GMMRF-S and LFDA-GMMRF-S were improved by almost 5.2% and 7.5%, respectively, when compared to RP-GMMRF and LFDA-GMMRF. Moreover, the proposed GMMRF-S technique demonstrated the best class-specific accuracies for most of the available classes. This amount of improvement in accuracy is because of the significant presence of mixed pixels in all available classes in the Indian Pines image that makes the outputs of pixel-wise classification methods noisy. However, by using the spatial information especially using the segmentation map in the MRF regularization, the noisy effects and mixed pixels was noticeably reduced.

Table 2 contains the global classification accuracies for the LFDA-GMM, RP-GMM, RP-GMMRF-S, RP-GMMRF, LFDA-GMMRF-S and LFDA-GMMRF applied to the Pavia University dataset.

As can be seen from Table 2, using neighbourhood pixel information with MRF can improve classification accuracies. The RP-GMMRF-S gives the best performance regarding classification accuracies when compared to other methods. RP-GMMRF-S improved the classification accuracies of RP-GMM and RP-GMMRF by almost 3.5% and 2.5%, respectively. In the same way, the LFDA-GMMRF-S improved the classification accuracies by 4% and 2.7% in comparison with LFDA-GMM and LFDA-GMMRF, respectively. It is worth mentioning that when the image consists of small structures, such as trees, buildings’ roof or mixed pixels, the inclusion of the segmentation map information in the context-based regularization improves the classification performances. For instance, in Indian Pines and Pavia University images, the differences between GMMRF-S and GMMRF is significant (see Figure 6). In overall, the improvements for Indian Pines and Pavia University were about 6.4% and 2.6% respectively in term of overall accuracy. Consequently, the advantage of the segmentation-based GMMRF-S method for the classification of urban images and scenarios with mixed pixels is confirmed.

Table 3 summarizes the global and class-specific classification accuracies for the Salinas data set. For the Salinas data set, which represents a vegetation classification scenario and contains large spatial structures, reference data do not comprise region edges. As a result, the advantages of the GMMRF-S method versus the GMMMRF method are not evident. In this case, the differences between GMMRF-S and GMMRF are insignificant, and the improvement value is 0.5% in overall accuracy. Therefore, the proposed method efficient only in the case where there is no large misclassified region in the initial pixel-wise classification map.

The overall results of McNemar’s test confirm that the differences between the accuracies of RP-GMMRF-S and the above methods were statistically significant. For the Indian Pines and Salinas data sets $Z>8.5$ and $Z>10$ were obtained, respectively (Tables 1 and 3). In the case of “Pavia University”, $Z>3$ was obtained when comparing with each of these methods except for the RP-GMMRF, while the difference in accuracy between RP-GMMRF-S and RP-GMMRF was not significant (Table 2).

In order to compare the performances of the proposed method with other recently proposed classification techniques, MLR-MRF (Li, Bioucas-Dias, et al.,

---

**Table 2. Classification accuracy results for Pavia University dataset with 15% training samples.**

| Classes                  | LFDA-GMM | RP-GMM | RP-GMMRF-S | RP-GMMRF | LFDA-GMMRF-S | LFDA-GMMRF |
|--------------------------|----------|--------|------------|----------|--------------|------------|
| OA                       | 89.14    | 90.70  | 94.20      | 92.74    | 93.13        | 91.40      |
| AA                       | 89.10    | 91.23  | 94.84      | 93.29    | 93.28        | 92.24      |
| Kappa Coefficient        | 0.87     | 0.88   | 0.93       | 0.91     | 0.90         | 0.89       |
| McNemar’s test           | 9.84     | 6.68   | -          | 1.08     | 3.01         | 7.16       |
| Water                    | 94.23    | 93.78  | 98.99      | 97.65    | 95.33        | 95.22      |
| Asphalt                  | 93.43    | 94.45  | 96.43      | 95.22    | 96.47        | 94.21      |
| Meadows                  | 91.45    | 92.30  | 97.41      | 94.14    | 94.20        | 93.37      |
| Gravel                   | 92.53    | 91.52  | 98.50      | 98.50    | 97.66        | 95.51      |
| Trees                    | 93.25    | 93.90  | 97.55      | 96.50    | 95.53        | 95.36      |
| Painted metal sheets     | 84.33    | 85.30  | 88.31      | 87.31    | 88.12        | 87.90      |
| Bare Soil                | 82.66    | 90.05  | 93.65      | 92.00    | 92.95        | 93.35      |
| Bitumen                  | 81.67    | 84.55  | 87.54      | 85.64    | 86.65        | 84.48      |
| Self-Blocking Bricks     | 92.13    | 97.75  | 99.36      | 97.15    | 98.56        | 94.78      |
| Shadows                  | 85.28    | 88.70  | 90.75      | 89.75    | 91.33        | 88.23      |
SVM-MRF (Tarabalka et al., 2010), edge-based SVMMRF-E method (Tarabalka et al., 2010) and LPNMF-GMM-MRF (W. Li et al., 2014) were also implemented. Classification accuracies of the proposed method, as well as those for current state-of-the-art techniques, are shown in Table 4 for the Pavia University and Salinas dataset. All classification techniques are implemented using optimal parameters, and the training samples are randomly selected from the overall ground truth in a 10% rate. To avoid any bias, the experiments are repeated 10 times, and we reported the average classification accuracy. As can be seen from Table 4, for the Pavia University and Salinas dataset, the proposed RP-GMMMRF-S and LFDA-GMMMRF-S significantly outperform other classification techniques, and RP-GMMMRF-S gives the highest accuracy. In the case of Indian Pines dataset, the LPNMF-GMM-MRF and RP-GMMMRF-S methods with a slight difference to each other performed better than the other methods.

### Table 3. Classification accuracy results for Salinas dataset with 15% training samples.

| Classifier          | OA    | AA    | Kappa | OA    | AA    | Kappa | OA    | AA    | Kappa |
|---------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| LFDA-GMM            | 92.45 | 92.88 | 94.70 | 94.10 | 93.62 | 93.11 |
| RP-GMM              | 89.60 | 90.92 | 96.40 | 92.45 | 90.00 | 92.12 |
| Kappa               | 0.91  | 0.91  | 0.93  | 0.92  | 0.92  | 0.92  |
| McNemar's test      | 22.00 | 14.86 | 97.40 | 10.20 | 11.29 | 11.25 |
| Brocii_green_weeds_1| 91.27 | 95.90 | 97.82 | 94.11 | 94.40 | 94.04 |
| Brocii_green_weeds_2| 74.32 | 66.59 | 76.03 | 67.14 | 63.56 | 76.03 |
| Fallow              | 96.65 | 97.00 | 98.17 | 97.65 | 96.28 | 96.28 |
| Fallow_rough_plow   | 95.25 | 95.21 | 99.95 | 96.97 | 100.00 | 99.15 |
| Fallow_smooth       | 96.72 | 97.73 | 99.54 | 100.00 | 100.00 | 98.45 |
| Stubble             | 95.50 | 92.21 | 100.00 | 97.11 | 100.00 | 100.00 |
| Celery              | 87.60 | 94.10 | 94.21 | 94.06 | 88.83 | 90.22 |
| Grapes_untrained    | 93.11 | 98.11 | 99.59 | 99.31 | 99.98 | 98.56 |
| Soil_vineyard_develop| 82.35 | 91.05 | 89.80 | 89.98 | 83.45 | 80.00 |
| Corn_senesced_green_weeds | 83.59 | 75.25 | 92.34 | 70.59 | 86.53 | 83.41 |
| Lettuce_romaine_4wk | 93.55 | 85.43 | 98.48 | 83.47 | 98.82 | 97.50 |
| Lettuce_romaine_5wk | 78.05 | 89.31 | 98.81 | 92.96 | 78.33 | 76.83 |
| Lettuce_romaine_6wk | 81.21 | 91.15 | 98.67 | 96.31 | 82.98 | 90.61 |
| Lettuce_romaine_7wk | 99.80 | 94.54 | 99.85 | 98.16 | 100.00 | 99.80 |
| Vinyard_untrained   | 90.94 | 93.22 | 99.83 | 97.71 | 99.33 | 95.44 |
| Vinyard_vertical_trellis | 93.78 | 97.84 | 99.79 | 100.00 | 100.00 | 97.29 |

### Figure 6. Pavia University classification maps.

### Table 4. The comparison of the proposed method with the state-of-the-art methods including MLR-MRF, SVM-MRF, SVMMRF-E and LPNMF-GMM-MRF.

| Classifier          | Pavia University Dataset | Salinas Dataset | Indian Pines Dataset |
|---------------------|--------------------------|----------------|---------------------|
| OA      | AA     | Kappa | OA      | AA     | Kappa | OA     | AA     | Kappa |
| MLR-MRF | 87.34  | 91.23  | 0.85   | 89.65  | 90.11  | 0.88   | 86.21  | 85.32  | 0.86   |
| SVM-MRF | 87.87  | 91.82  | 0.86   | 91.78  | 92.27  | 0.90   | 88.76  | 89.45  | 0.88   |
| SVMMRF-E | 88.76  | 92.15  | 0.88   | 94.31  | 95.14  | 0.93   | 90.73  | 90.05  | 0.89   |
| LPNMF-GMM-MRF | 92.85  | 93.51  | 0.90   | 93.87  | 94.39  | 0.91   | 91.13  | 90.22  | 0.91   |
| RP-GMMMRF-S | 94.10  | 94.85  | 0.92   | 94.51  | 96.23  | 0.93   | 91.54  | 89.72  | 0.91   |
| LFDA_GMMMRF-S | 92.93  | 94.23  | 0.91   | 93.07  | 92.07  | 0.92   | 88.42  | 87.64  | 0.88   |
Finally, we compared the computational cost of the classification methods using the same experimental data. Experiments are conducted using MATLAB on a 2.4-GHz processor with 3.6-GB of RAM. The processing times (in seconds) are 455 for RP-GMM, 470 for LFDA-GMM, 461 for RP-GMM-MRF and 466 for RP-GMM-MRF-S respectively. Therefore, the MRF regularization amounts to only an average of 9 s for the experimental data.

Conclusion

In this paper, a novel efficient method based on the combination of GMM and MRF was proposed for spectral-spatial classification of hyperspectral data. In this method, GMM was used as a pixel-wise classifier. In parallel, MRF was applied to extract spatial information. One important motivation of this paper, for the consideration of MRF regularization, was the usefulness of highlighting additional information from image objects. This information allows involving the shape characteristics, as well as the neighbourhood relationships. For that reason, a new fuzzy segmentation-based function was defined and combined with the MRF model. The efficiency of the proposed method was tested in both scenarios of with and without considering the segmentation map. Results confirm that the integration of MRF extracted contextual information with GMM, as a pixel-wise classifier, can improve accuracies noticeably. In particular, when the large spatial structure exists in the scene, this approach can be beneficial. This attribute helps to mitigate the noisy effects of the pixel-wise classification. However, for the small structures, when the spatial information from neighbouring pixels is taken into account, these structures might likely be disappeared or merged with more prominent structures. The results of experiments showed that the proposed GMM-MRF classification strategy outperformed other traditional techniques over a wide range of operating conditions. Moreover, GMM-MRF-S was even superior to the GMM-MRF. Further research on this method also needs to cover more contextual information and advanced strategies for robustification of the spatial information used in the method.

Acknowledgments

The authors would like to thank Prof. P. Gamba from the University of Pavia for providing the ROSIS data sets.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Hamid Ghanbari http://orcid.org/0000-0002-9557-495X

References

Achlioptas, D. (2001). Database-friendly random projections. Paper presented at the Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of database systems.

Akbari, D., Homayouni, S., Safari, A., & Mehrshad, N. (2016). Mapping urban land cover based on spatial-spectral classification of hyperspectral remote-sensing data. International Journal of Remote Sensing, 37(2), 440–454.

Benediktsson, J.A., & Ghamisi, P. (2015). Spectral-spatial classification of hyperspectral remote sensing images. Norwood, MA: Artech House.

Besag, J. (1986). On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society. Series B (Methodological), 48(3), 259–302.

Bingham, E., & Mannila, H. (2001). Random projection in dimensionality reduction: Applications to image and text data. Paper presented at the Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining.

Boykov, Y., & Kolmogorov, V. (2004). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(9), 1124–1137.

Camps-Valls, G., & Bruzzone, L. (2005). Kernel-based methods for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 43(6), 1351–1362.

Camps-Valls, G., Gomez-Chova, L., Muñoz-Marí, J., Vila-Francés, J., & Calpe-Maravilla, J. (2006). Composite kernels for hyperspectral image classification. IEEE Geoscience and Remote Sensing Letters, 3(1), 93–97.

Cheng, Y. (1995). Mean shift, mode seeking, and clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(8), 790–799.

Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5), 603–619.

Dasgupta, S. (2000). Experiments with random projection. Paper presented at the Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence.

Dietterich, T.G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. Neural Computation, 10(7), 1895–1923.

Figueiredo, M.A.T., & Jain, A.K. (2002). Unsupervised learning of finite mixture models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(3), 381–396.

Frankl, P., & Maehara, H. (1988). The Johnson-lindenstrauss lemma and the sphericity of some graphs. Journal of Combinatorial Theory, Series B, 44(3), 355–362.

Fukunaga, K., & Hostetler, L. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Transactions on Information Theory, 21(1), 32–40.

Geman, S., & Geman, D. (1987). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. In Readings in Computer Vision (pp. 564–584). Amsterdam: Elsevier.
Sequential EM for unsupervised spectral-spatial classification of hyperspectral images based on hidden markov random fields. IEEE Transactions on Geoscience and Remote Sensing, 52(5), 2565–2574.

Hassanpour, R., Shahbahrami, A., & Wong, S. (2008). Adaptive Gaussian mixture model for skin color segmentation. World Academy of Science, Engineering and Technology, 41, 1–6.

Jamshidpour, N., Safari, A., & Homayouni, S. (2017). Spectral–spatial semisupervised hyperspectral classification using adaptive neighborhood. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(9), 4183–4197.

Kolmogorov, V. (2006). Convergent tree-reweighted message passing for energy minimization. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(10), 1568–1583.

Li, J., Bioucas-Dias, J.M., & Plaza, A. (2010). Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. IEEE Transactions on Geoscience and Remote Sensing, 48(11), 4085–4098.

Li, J., Bioucas-Dias, J.M., & Plaza, A. (2012). Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields. IEEE Transactions on Geoscience and Remote Sensing, 50(3), 809–823.

Li, S.Z. (2009). Markov random field modeling in image analysis. London: Springer-Verlag Science & Business Media.

Li, W., Prasad, S., & Fowler, J.E. (2014). Hyperspectral image classification using gaussian mixture models and markov random fields. IEEE Geoscience and Remote Sensing Letters, 11(1), 153–157.

Li, W., Prasad, S., Fowler, J.E., & Bruce, L.M. (2012). Locality-preserving dimensionality reduction and classification for hyperspectral image analysis. IEEE Transactions on Geoscience and Remote Sensing, 50(4), 1185–1198.

Moser, G., De Giorgi, A., & Serpico, S.B. (2016). Multiresolution supervised classification of panchromatic and multispectral images by Markov random fields and graph cuts. IEEE Transactions on Geoscience and Remote Sensing, 54(9), 5054–5070.

Moser, G., & Serpico, S.B. (2013). Combining support vector machines and Markov random fields in an integrated framework for contextual image classification. IEEE Transactions on Geoscience and Remote Sensing, 51(5), 2734–2752.

Moser, G., Serpico, S.B., & Benediktsson, J.A. (2013). Land-cover mapping by Markov modeling of spatial–Contextual information in very-high-resolution remote sensing images. Proceedings of the IEEE, 101(3), 631–651.

Plaza, A., Benediktsson, J.A., Boardman, J.W., Brazil, J., Bruzzone, L., Camps-Valls, G. … Gualtieri, A. (2009). Recent advances in techniques for hyperspectral image processing. Remote Sensing of Environment, 113, S110–S122.

Priebe, C.E. (1994). Adaptive mixtures. Journal of the American Statistical Association, 89(427), 796–806.

Solberg, A.H.S., T:t:xt, T., & Jain, A.K. (1996). A Markov random field model for classification of multisource satellite imagery. IEEE Transactions on Geoscience and Remote Sensing, 34(1), 100–113.

Sugiyama, M. (2007). Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. Journal of Machine Learning Research, 8(May), 1027–1061.

Tarabalka, Y., Benediktsson, J.A., & Chanussot, J. (2009). Spectral–spatial classification of hyperspectral imagery based on partitioned clustering techniques. IEEE Transactions on Geoscience and Remote Sensing, 47(8), 2973–2987.

Tarabalka, Y., Fauvel, M., Chanussot, J., & Benediktsson, J.A. (2010). SVM-and MRF-based method for accurate classification of hyperspectral images. IEEE Geoscience and Remote Sensing Letters, 7(4), 736–740.

Zhao, Y., Zhang, L., Li, P., & Huang, B. (2007). Classification of high spatial resolution imagery using improved gaussian markov random-field-based texture features. IEEE Transactions on Geoscience and Remote Sensing, 45(5), 1458–1468.

Zuza, T., Olugbara, O.O., Ojo, S.O., & Ngwira, S.M. (2011). Image segmentation, available techniques, developments and open issues. Canadian Journal on Image Processing and Computer Vision, 2(3), 20–29.