A GAUSSIAN APPROACH TO SUBSPACE BASED CLASSIFICATION OF HYPERSPECTRAL IMAGES

Mahdi Khodadadzadeh1, Lorenzo Bruzzone1, Jun Li2 and Antonio Plaza3

1Department of Information Engineering and Computer Science, University of Trento, Trento, Italy
2School of Geography and Planning, Sun Yat-Sen University, Guangzhou, China
3Hyperspectral Computing Laboratory, University of Extremadura, Cáceres, Spain

ABSTRACT

Supervised classification of hyperspectral images is a challenging task due to the relatively low ratio between the number of training samples and the number of spectral channels. Subspace-based classification methods deal with this difficulty by assuming that feature vectors lie in a low-dimensional subspace. Based on the fact that a class in a hyperspectral image may be composed of a number of different groups of materials and mixture of spectral features, we suggest to estimate several lower dimensional random subspaces for the samples within each class. For subspace learning and classification, we propose to exploit the union of random subspaces in a Gaussian Mixture Model. Experimental results, conducted on two real hyperspectral data sets, indicate that the proposed method provides competitive classification results in comparison with other state-of-the-art approaches.

Index Terms—Hyperspectral images, classification, subspace-based approaches, Gaussian mixture model, remote sensing

1. INTRODUCTION

Land cover classification is widely used for remotely sensed data exploitation. Hyperspectral sensors can be effectively used to discriminate between similar materials on the land surface. A hyperspectral image is composed of hundreds of continuous narrow spectral bands, which provide detailed and valuable spectral information. However, the dimensionality of a hyperspectral image is very high. Theoretically, a very high number of training samples is needed for the learning stage of a classification system, which is expensive and costly from both a practical and computational point of view [1]. In the literature, the use of feature selection and extraction methods before the classification has been suggested in order to reduce the dimensionality of hyperspectral data, and to circumvent this problem. Another approach that can provide competitive advantages is to use subspace-based classification methods, which assume that feature vectors lie in a low-dimensional subspace. More specifically, the subspace-based classifiers can efficiently exploit the characteristics of the groups of features (which sometimes, are hidden in the original hyperspectral space) and thus increase the accuracy of a classification system by separating the classes which are spectrally very similar. However, the task of recovering the underlying subspaces existing in the feature space (without losing the original information) is challenging. A typical approach is based on the estimation of a single lower dimensional subspace for the samples within each class. Based on this idea, recently, several subspace projection methods have been proposed for processing of hyperspectral images [2–5].

In this work, we concentrate on the Gaussian distribution and provide an analysis of the use of Gaussian models for subspace learning and classification of hyperspectral data. More importantly, we propose to use the Gaussian Mixture Model (GMM) to learn the union of different subspaces for each class. Since the spectral vectors in a given class may be composed of a number of different groups of materials and mixture of features, a more general model is based on the representation of each class with a collection of low-dimensional subspaces [6]. Another contribution of this work is to introduce a strategy inspired by bagging [7] in the framework of random subspace classification. Our method is based on the idea that each group of samples in the training set leads to a subspace estimate, and thus a group of random samples extracted from the training set of each class will lead to a random subspace which can be used to better characterize that class.

2. METHODOLOGICAL FRAMEWORK

Let $S \equiv \{1, 2, \ldots, n\}$ be the set of integers indexing the $n$ pixels of a hyperspectral image, $\mathbf{x} \equiv \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ denote the input hyperspectral image, where $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{id}]^T$ denotes a feature vector associated with an image pixel $i \in S$ and $d$ is the number of spectral bands. Let $\mathbf{y} \equiv \{y_1, \ldots, y_n\}$ be an image of labels, $y_i \equiv [y_i^{(1)}, y_i^{(2)}, \ldots, y_i^{(k)}]^T$, where $k$ is the number of classes. For $c = 1, \ldots, k$, if pixel $i$ belongs to class $c$, $y_i^{(c)} = 1$, otherwise, $y_i^{(c)} = 0$. Furthermore, let
\[ \mathcal{D}(c) = \{x(c)^{(1)}, \ldots, x(n(c))^{(c)}\} \]

represent the set of training samples with label \( c \) (\( n(c) \) is the number of training samples for the class \( c \)) and \( \mathcal{D} = \{ \mathcal{D}^{(1)}, \ldots, \mathcal{D}^{(k)} \} \) represent the complete training set.

We begin with a simple model for a class dependent subspace, in which the observation for class \( c \) can be written as:

\[
x(c)^{(c)} = U(c) z_i,
\]

where \( U(c) = \{u^{(c)}_1, \ldots, u^{(c)}_{r(c)}\} \) is a set of \( r(c) \)-dimensional orthonormal basis vectors for the subspace associated with class \( c \) (\( r(c) \ll d \)). Note that for simplicity we do not consider the effect of noise in the proposed model. With this in mind, if we assume a Gaussian distribution with zero mean and diagonal covariance matrices for \( z_i \), i.e. \( z_i \sim \mathcal{N}(0, \sigma I) \), then we can compute the likelihood as \( p(x(c)^{(c)} | y_i^{(c)} = 1) \sim \mathcal{N}(0, \sigma U(c) U(c)^T) \).

As mentioned before, we may extract several sets of lower dimensional orthonormal basis vectors for a given class, i.e. \( U^{(c)} = \{u^{(c)}_1, \ldots, u^{(c)}_{m(c)}\} \) for \( j = 1, \ldots, m \). In order to exploit the information coming from the union of \( m \) subspaces, we propose the following generative model based on GMM:

\[
p(x(c)^{(c)} | y_i^{(c)} = 1) \sim \sum_{j=1}^{m} \alpha_j^{(c)} \mathcal{N}(0, \sigma U^{(c)} j U^{(c)} j^T),
\]

where \( \alpha_j \) is the proportion of the single subspace \( j \). Using the aforementioned models, the posterior class density \( p(y_i^{(c)} = 1 | x_i) \) can be computed as follows:

\[
p(y_i^{(c)} = 1 | x_i) = \frac{p(x_i | y_i^{(c)} = 1)p(y_i^{(c)} = 1)}{\sum_k p(x_i | y_i^{(c)} = 1)p(y_i^{(c)} = 1)},
\]

where the prior probability \( p(y_i^{(c)} = 1) \) can be calculated using prior information about the distributions of the classes in the scene. If this information is not available, one can use the simplifying assumption of equiprobable classes, i.e., \( p(y_i^{(c)} = 1) = 1/k \) for \( c = 1, \ldots, k \).

### 2.1. Subspace generation

The eigenanalysis of the correlation matrix or the covariance matrix of a set of samples can be used to compute a subspace base. For example, if \( R \) is the correlation matrix associated with a set of samples, by computing the eigenvalue decomposition, we have:

\[
R = EE^T = E \text{diag}(\lambda_1, \ldots, \lambda_d) E^T,
\]

where the orthogonal matrix \( E = \{e_1, \ldots, e_d\} \) contains the eigenvectors of \( R \) and the diagonal matrix \( A = \text{diag}(\lambda_1, \ldots, \lambda_d) \) contains the eigenvalues of \( R \) with decreasing magnitude, i.e., \( \lambda_1 \geq \cdots \geq \lambda_d \). Thus, an estimate of a subspace can be obtained by constructing a set of \( r \)-dimensional orthonormal basis vectors \( U = \{e_1, \ldots, e_r\} \) (\( r < d \)).

As detailed in Algorithm 1, for generating a collection of bases for the class dependent random subspaces, we randomly select sets of labeled samples out of training samples of each class and then generate the related subspace base using the aforementioned method. Note that, in this work, we consider the same dimension for the all computed subspaces. Moreover, we use a noise free model. Therefore, a denoising technique should be used before the subspace learning and image classification. For this purpose, we apply the fast noise estimation method which has been proposed in [8]. This method shows good performance based on the fact that there is a high correlation between neighboring spectral bands in hyperspectral images.

### Algorithm 1 Random subspace generation

1. **Input:** \( \mathcal{D}, m, r \)
2. **Output:** \( U \)
3. **for** \( c := 1 \) to \( k \) **do**
   1. **do**
      1. **for** \( j := 1 \) to \( m \) **do**
         1. \( r^{(c)} = \text{rand}(\mathcal{D}^{(c)}) \) (* random selection of samples *)
         2. \( U^{(c)} = \text{sub}(r^{(c)}, r) \) (* subspace generating *)
         3. \( U \leftarrow \{U^{(c)} \cup U\} \)
      2. **end for**
   2. **end for**
3. **end for**

### 2.2. Expectation-Maximization estimation of GMM parameters

As stated before, we assume that the observation in a given class lies in the union of random subspaces and we model this by using GMM. In this section we briefly present the Expectation-Maximization (EM) [9] algorithm to estimate the mixture parameters. Starting from some initial estimates, the EM algorithm updates the estimates iteratively until a specified convergence based on a log-likelihood function is reached.

According to (2), for a given class \( c \), there are two types of parameters to estimate, i.e., mixture weights \( \alpha_j^{(c)} \) and \( \sigma U^{(c)} j U^{(c)} j^T \) for \( j = 1, \ldots, m \). Since \( \theta_j^{(c)} = \sigma U^{(c)} j U^{(c)} j^T \) can be computed in advance and are unchanged during the EM iterations, the goal is to estimate mixture weights \( \alpha_j^{(c)} \), \( j = 1, \ldots, m \), which determine the impact of each single subspace in the union of subspaces. Given \( \Theta^{(c)} = \{\alpha_1^{(c)}, \ldots, \alpha_m^{(c)}, \theta_1^{(c)}, \ldots, \theta_m^{(c)}\} \), the GMM model is defined as:

\[
p(x_i^{(c)} | \Theta^{(c)}) = \sum_{j=1}^{m} \alpha_j^{(c)} p_j(x_i^{(c)} | z_j^{(c)}, \theta_j^{(c)}),
\]

where \( z^{(c)} = [z_1^{(c)}, \ldots, z_m^{(c)}] \) is a vector of \( m \) binary indicator variables. Clearly, each component in the GMM is a multivariate Gaussian distribution, i.e., \( p_j(x_i^{(c)} | z_j^{(c)}, \theta_j^{(c)}) \sim \mathcal{N}(0, \sigma U^{(c)} j U^{(c)} j^T) \)
After initialization (e.g., \( \alpha_j^{(c)} = \frac{1}{m} \)), each iteration of the EM algorithm consists of an E-step and M-step, as follows:

- **E-Step**: Computing the membership weights \( w_{ij}^{(c)} \) for all \( i = 1, \ldots, n^{(c)} \) and \( j = 1, \ldots, m \), using the current parameter values \( \Theta^{(c)} \) as:
  \[
  w_{ij}^{(c)} = \frac{p(z_{ij}^{(c)} = 1|x_{ij}^{(c)}, \Theta^{(c)})}{\sum_{c=1}^k \alpha_j^{(c)} p(x_{ij}^{(c)}|z_{ij}^{(c)}, \Theta^{(c)})} \tag{6}
  \]

- **M-Step**: Updating the parameter values as:
  \[
  (\alpha_j^{(c)})_{\text{new}} = \frac{1}{n^{(c)}} \sum_{i=1}^{n^{(c)}} w_{ij}^{(c)} \quad \forall j = 1, \ldots, m. \tag{7}
  \]

After updating the parameters in the M-step, the algorithm goes back to the E-step. Generally, the algorithm stops when the value of log-likelihood does not change significantly or when a predefined number of iterations is reached.

### 3. EXPERIMENTAL RESULTS

In this section, we use the well known AVIRIS Indian Pines and ROSIS Pavia University images (which were captured from an agricultural area and an urban area, respectively) to evaluate the proposed approach. Detailed information about these images can be found in [4]. For the Indian Pines image, we discarded four small classes which contain less than 100 labeled pixels in the original ground-truth. In order to evaluate the proposed approach with limited training size, only 50 samples per class were randomly selected for training. For the Indian Pines image, the training samples were selected from the ground-truth and the rest of samples were used for validation. For the Pavia University image, the training samples were selected from the original training set and the original test set was used for validation. For the proposed method we fixed the dimension of the subspaces to 10 \((r = 10)\) and we considered the union of \( m = 50 \) different random subspaces for each class. We also fixed the value of \( \sigma = 1 \) in the calculation of the covariance matrices of Gaussian distributions.

To have a complete comparison, in addition to the proposed GMM-based method, the classification results are reported for two classifiers: the support vector machine (SVM) [10] and the MLR\(_{Sub} \) [4] (which combines the multinomial logistic regression algorithm with a subspace projection method). These methods have shown good performance for hyperspectral classification problems in small training samples size. All the experiments were repeated 20 times and the average classification accuracies are reported. It should also be noted that, for the SVM and MLR\(_{Sub} \) classifiers, the optimal parameters were chosen by fivefold cross validation. Moreover, we evaluate the classification results after incorporating spatial information and regularization of pixelwise classification maps. For this purpose, we applied a Markov random field (MRF) based method which uses graph cuts technique [11].

Table 1 shows the obtained classification results. Several conclusions can be derived from this Table. First and foremost, it is remarkable that the proposed approach, which exploits the union of several random subspaces for a given class, gives higher classification accuracies than the MLR\(_{Sub} \) classifier which uses a single lower dimensional subspace for the samples within each class. For instance, the proposed approach obtained an overall accuracy (OA) of 76.90\% for the Pavia University image, which is 11.51\% higher than the OA obtained by the MLR\(_{Sub} \) method. Interestingly, this OA is even 2.92\% higher than the OA obtained by the SVM method, which is one of the most effective pixel-wise classification methods for hyperspectral images. Furthermore, if we compare the results after MRF-based postprocessing, we can see that the accuracy improvements reported for the proposed method are quite significant. For example, for the Indian Pines image, the GMM-MRF obtained an OA of 91.86\%, which is 6.18\% and 2.65\% higher than the OAs obtained by MLR\(_{Sub-MRF} \) and SVM-MRF methods, respectively. For illustrative purposes, Fig. 1 shows the obtained classification maps for the Indian Pines image.

### 4. CONCLUSION AND REMARKS

In this paper, we proposed a novel supervised classification method for hyperspectral images based on subspace projection. We proposed a GMM based classifier to exploit the union of several random subspaces for representing each class. Each random subspace was estimated using a group of random samples extracted from the training set of each class. Our experimental results, conducted on two real hyperspectral data sets, showed that the proposed method provides very good classification results. More importantly, as compared to the approach based on using a single subspace for a class, the performance improvements reported for the proposed method were quite significant. In future developments, we will further explore the impact of the parameters of the proposed method on the classification performance. More specifically, in this paper, we fixed the dimension of the subspaces to 10, which can be estimated automatically and adaptively.

### 5. REFERENCES


Table 1. Overall and average classification accuracies [%] using 50 training samples per class.

<table>
<thead>
<tr>
<th>Data</th>
<th>Accuracy</th>
<th>Spectral Classification Algorithms</th>
<th>Spectral-Spatial Classification Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SVM</td>
<td>MLR_sub</td>
</tr>
<tr>
<td>Indian Pines</td>
<td>Overall</td>
<td>77.05</td>
<td>69.16</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>83.12</td>
<td>76.78</td>
</tr>
<tr>
<td>Pavia University</td>
<td>Overall</td>
<td>73.98</td>
<td>65.39</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>84.71</td>
<td>74.30</td>
</tr>
</tbody>
</table>

Fig. 1. Classification maps obtained by different methods for the AVIRIS Indian Pines scene (the overall and average accuracies are reported in the parentheses)


