SVM BASED CLASSIFICATION USING MRF FEATURES FOR HYPERSPECTRAL IMAGES

K. Kavitha1, S. Arivazhagan2 & G. R. Breen Lentil3

Department of Electronics and Communication, Mepco Schlenk Engineering College, Sivakasi, India,
(1)breenlentil88@gmail.com, (2)vihadakavi@yahoo.com, (3)sarivu@mepcoeng.ac.in

Abstract: The extraction of land cover information from Satellite Images using Classifiers has been the subject of intense interest and research in the remote sensing community. Classifying the pixels in the Hyper Spectral Images depends on the Feature Extraction and Classifier selection processes. Statistical features such as the mean and standard deviation gives the statistical information of the pixels, while the textural features give the inter-relationship between the grey levels. To find the textural properties of the pixels, Markov Random Field features can be derived using the Markov random Field Models. Markov Random Fields measure the interdependence of neighbouring pixels to produce features. In this paper, neighbouring pixels are taken into account in a priority sequence according to their distance from the centre pixel and a step-by-step least squares method is proposed to extract a novel set of Markov Random Field features. Support Vector Machines (SVM) has recently attracted the attention of the remote sensing community. This project addresses the problem of classification of Hyper Spectral Images by using Support Vector Machines

Keywords: Markov Random Fields (MRF), Support Vector Machines (SVM), hyper spectral images, One-Against-All (OAA), RBF Kernel.

1. INTRODUCTION
Classifying the pixels in the Hyper Spectral Images and identifying their relevant class belongings depends on the feature extraction and classifier selection processes. Feature extraction is the akin process while classifying the images. Statistical features such as the mean and standard deviation gives the statistical information of the pixels, while the textural features give the inter-relationship between the grey levels. [1]. More over to find the textural properties of the pixels wavelet domain can be used [2]. Markov Random Field features can be derived using the Markov random Field Models. The MRF features are also texture based feature. The Markov Random Field feature for individual pixels in the spatial domain is proved as a promising technique in case of mono spectral and hyper spectral images.

The Markov Random Field Features are calculated from the Markov Random Field Models. Markov Random Field modeling replaces each pixel by its local conditional probability. The parameters of the Markov Random Field model is used as the feature for classification purpose. The supervised method such as Support Vector Machines (SVM) requires data for training and used as reference for automatically classifying new data.

An extensive literature is available on the pixel-level processing technique, i.e., techniques that assign each pixel to one of the classes based on its extracted features [3]. Yindi Zhao et al. (2007) explained the usage of Gaussian MRF features for the classification of the high spatial resolution images [4]. Thoonen. G. De Backer et al. (2008) used spatial features alone to classify Hyper spectral Data of Dune Vegetation along the Belgian coast [5]. M. Schroder and H. Rehrauer proposed methods for Spatial information retrieval from remote-sensing images using the Gibbs–Markov Random Felds [6].

Mahesh Pal and Giles M.Foody explained the feature selection for Classification of Hyperspectral Data by use of the SVM classifier [7]. Zoltan Kato and Ting-ChuenPong developed a Markov random field image segmentation model for color textured images [8]. Classification of hyper spectral images with nonlinear filtering and support vector machines is explained by M. Lennon et al. [9]. All the above mentioned works extracted Markov Random Field features such as beta and the variances of the different neighborhoods are
individually used to classify the images. In this proposed work, the analysis is made when the features are summed up and are used for classification. The usage of SVM classifier for hyper spectral images is shown by J. Gualtieri[10]. Multiclass classifier for hyper spectral images is explained by T, Joachims [11]. As classifiers are concerned, Spectral Angle Classifiers are classical one, used for hyper spectral image analysis in early works in which the input image spectra are compared with reference spectra for classification. The classes are decided by calculating angular separation between the input and reference spectra known as spectral signature. This is a time consuming work. For very complex boundaries of data this shows poor accuracy.

So, a classifier which is able to perform classification even with very complex boundaries was needed. As the Neural Network based classifiers and Support Vector Machine (SVM) are having that ability, they came into picture around 1990s. As the internal process of the Neural Network classifier is a hidden one and complex, a simple learning machine based on statistical theory became popular [12]. Unlike neural networks, SVM requires no pre processing on the data. Support Vector Machines advocate good results in the linear domain classification [27].

But, the hyper spectral domain is a non-linear one. Non-linear domain can be converted into the linear domain by using kernel trick. SVMs with kernel functions are used [13]-[14]. Many types of kernels like linear, polynomial, radial Basis Function (RBF), Sigmoid etc., are available. Selection of proper kernel gives proper results. Formation of Composite kernel using SVM with Radial Basis Function (RBF) is a preferred combination which balances the computational complexity and accuracy.

From the literature, it is evident that the care must be shown towards the feature extraction and classifier selection. The Markov Random Field is very useful in extracting the features from the hyper spectral images. The features are extracted in priority basis so that low order features contain more details than the higher order features. This Priority order also makes the feature selection easier. The MRF model parameters are used for classification either by using them individually or by summing and concatenation. The classification is done by using the Support Vector Machines using the RBF kernel.

The rest of this paper is organized as follows. Section II describes the proposed method. Section III describes the novel feature extraction using the Markov Random Fields. Section IV briefly explains the classification using SVM classifier. Section V provides the results and discussions and conclusions are given in Section VI.

2. PROPOSED METHOD

The proposed work is given in Figure 1. The first stage is the selection of input image. For our experiment AVIRIS data set of INDIANA band has been selected. Among these total 220 bands, fifty bands have been selected from the wavelength 0.4 µm -1.8µm. The image contains 15 different classes. Among the fifteen classes nine classes are taken for classification. For individual bands Markov Random Field features are calculated.

From the extracted features summationfeatures are derived. Finally the classes are classified using the SVM classifier.

3. FEATURE EXTRACTION

Markovianity refers to the fact that the central pixel of an image interacts with only the neighbouring pixels, independent of other pixels. MRF model specifies spatial dependencies between the central pixel and its neighbouring pixels by defining a local conditional probability distribution, which is assumed to be Gaussian. The number and position of neighbouring pixels are determined by the
neighbour order defined using a Euclidean distance. Up to fifth order neighbours are depicted in Figure 2; wherein the number indicates the neighbour order to which the underlying site belongs. Note that one neighbour set with order \( n \) includes all the neighbours of sets with order one to \( n \).

A fourth-order MRF model centered on an image cell ‘s’, would include those cells marked one to four, and its corresponding neighbour set can be denoted as a set of shift vectors \( RN = \{(0, 1), (1, 0), (1, -1), (1, 1), (0, 2), (2, 0), (1, -2), (1, 2), (2, -1), (2, 1)\} \cup \{(-0, 1), (-1, 0), (-1, 1), (-1, -1), (0, -2), (-2, 0), (-1, 2), (-1, -2), (-2, 1), (-2, -1)\} \).

Figure 2: First to Fifth Order Neighbourhood of Site’s’

Suppose if the image is modeled as a fourth-order MRF model \( f(s) \), the grey-level intensity of a pixel ‘s’ has a local conditional probability density function,

\[
p(f(s)|fRN(s)) = \frac{1}{\sqrt{2\pi\nu}} \exp \left\{ -\frac{1}{2\nu} [f(s) - \mu]^2 \right\} \tag{1}
\]

where \( fRN(s) = \{f(s + r) \mid r \in RN\} \) stands for the set of values at the neighbouring site of the central pixel \( s \), \( \mu \) is the mean grey value of the whole image, \( \beta(r)'s \) are the model parameters, and \( \nu \) is the conditional variance.

The parameters \( \beta(r)'s \) and conditional variance \( \nu \) describe the MRF models and characterize textures. There are many existing methods for estimating those unknowns but none can guarantee consistency as well as stability. The choice of the least squares (LS) method is motivated by the simplicity, stability trade-off. Denote the set of the MRF parameters by \( \beta = \text{col} \left[ \beta(r) \mid r \in RN \right] \).

The LS estimate of the model parameter vector \( \beta \) is given by

\[
\hat{\beta} = \left[ \sum_{s \in S} Q(s)Q^T(s) \right]^{-1} \left[ \sum_{s \in S} Q(s)(f(s) - \mu) \right] \tag{2}
\]

where \( Q(s) = \text{col}[(f(s + r) - \mu) + (f(s - r) - \mu) \mid r \in RN] \). The estimate \( \nu \) of the conditional variance \( \nu \) is calculated by

\[
\nu = \frac{1}{MN} \sum_{s \in S} \left[ (f(s) - \mu) - Q^T(s)\beta \right]^2 \tag{3}
\]

MRF-based approach employs \( \eta = [\beta, \nu] \) as a texture feature vector. The classical MRF based method thinks that all of the neighbouring pixels, which are treated equally, interact on the center pixel simultaneously. However, it is reasonable that neighbouring pixels have influence on the center pixel in a priority sequence, i.e., the closer the neighbouring pixel to the center pixel, the higher the priority. To study this situation, a step-by-step LS method is proposed, to derive a different set of MRF texture features. For this purpose, the parameters of the MRF model can be divided into a certain number of groups and then estimated step-by-step.

For a fourth-order MRF model, \( \beta \) is divided into four groups \( \beta^{(1)}, \beta^{(2)}, \beta^{(3)} \) and \( \beta^{(4)} \). Each group \( \beta^{(k)} \) corresponds to a group of neighbours, denoted by RN that is \( RN^{(1)} = \{(0, 1), (1, 0)\}, RN^{(2)} = \{(1, -1), (1, 1)\}, RN^{(3)} = \{(0, 2), (2, 0)\}, \) and \( RN^{(4)} = \{(1, -2), (1, 2), (2, -1), (2, 1)\} \).

3.1. Feature Extraction for \( k^{th} \) Order Neighborhood

For \( 1 < k \leq m \)

\[
\beta^k = \left[ \sum_{s \in S} Q^k(s)(Q^k(s))^T \right]^{-1} \left[ \sum_{s \in S} Q^k(s)(f(s) - \mu + R) \right] \tag{4}
\]

\[
\nu^k = \frac{1}{MN} \sum_{s \in S} \left[ (f(s) - \mu) - \sum_{j=1}^{k} ((Q^j(s))^T) \beta^j \right]^2 \tag{5}
\]

From (4)–(5), it is obvious that the parameters in the low level group are computed independent of those in the higher level groups, however the parameters in the high-level groups are estimated based on those in the lower level groups.

A MRF feature vector was constructed using the results of each estimating step, namely, \( \psi = [\beta, \nu(k) \mid 1 \leq k \leq m] \). By means of step by step LS, the
lower order parameters are independent of those of higher orders.

4. CLASSIFICATION METHODOLOGY

This classification system is based on statistical learning theory. Support Vector Machines (SVMs) are a relatively new supervised classification technique to the land cover mapping community. They have their roots in Statistical Learning Theory and have gained prominence because they are robust, accurate and effective even when using a small training sample. In simple words, given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether a new example falls into one category or the other. Intuitively, an SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

The SVM can be seen as a new way to train polynomial, radial basis function, or multilayer perceptron classifiers, in which the weights of the network are found by solving a Quadratic Programming (QP) problem with linear inequality and equality constraints using structural risk minimisation rather than by solving a non-convex, unconstrained minimisation problem, as in standard neural network training technique using empirical risk minimisation. Empirical risk minimises (ERM) the misclassification error on the training set, whereas structural risk minimises (SRM) the probability of misclassifying a previously unseen data point drawn randomly from a fixed but unknown probability distribution. The name SVM results from the fact that one of the outcomes of the algorithm, in addition to the parameters for the classifiers, is a set of data points (the “support vectors”) which contain, in a sense, all the information relevant to the classification problem. Linearly separable classes are the simplest case on which to train a support vector machine. In the situations where it is not possible to have a decision surface (a hyper plane) defined by the linear equations on the training data, SVM propose that a feature vector, \( x \in \mathbb{R}^n \), is mapped into a higher dimensional Euclidean space, Figure 3, \( F \) via a non-linear vector function \( \Phi: \mathbb{R}^n \rightarrow F \). The optimal margin problem in the space \( F \) can be written by replacing \( x_i, \Phi (x) \) with \( \Phi (x_i) \cdot \Phi (x) \) then solving the optimization problem for \( \lambda_i \) in the transformed feature space by association with the \( \lambda > 0 \). By using this mapping, the solution of the SVM has the form:

\[
f(x) = \text{sign} \sum \lambda_i y_i \phi (x) \phi (x) + b \quad (6)
\]

As suggested by equation 10 the only quantity that one need to compute is the scalar product, of the form \( \phi (x). \phi (y) \). It is therefore convenient to introduce the concept of the kernel function \( K \) such that:

\[
K(x_i, x_j) = \phi (x_i). \phi (x_j) \quad (7)
\]

A number of kernel functions are used within SVM. They are the simple dot product, polynomial kernel of degree ‘d’, radial basis function (RBF), two-layer neural network, A linear spline with an infinite number of points.

In order to find the optimal decision surface, the support vector training algorithm tries to separate, as best as possible, the clouds of data points representing each class. Data points closer to the boundary between the classes are more important in the classification than are data points that are far away, since data points closer to the boundary are harder to classify. These data points help to shape and define better decision surface than other points. The support vector machine tries to find data points that are closest to the separating surfaces, therefore the support vectors are border points, and due to this reason support vectors are very few.

SVM was initially designed for binary (two-class) problems. When dealing with several classes, as in the case of land cover classification, an appropriate multi-class method is needed. Different possibilities for this include:
• Combine several binary classifiers: the “one against the rest” approach
• Combine several classifiers: the “one against one” approach.

One Against All (OAA) strategy is the simplest and is most commonly used while classifying the images. By which one class is separated from others. Thus the classes are separated hierarchically. A simple block diagram of OAA BHT with four classes is shown in Figure 4.

![OAA Binary Hierarchical Tree (BHT)](image)

**Figure 4: OAA Binary Hierarchical Tree (BHT)**

5. RESULTS AND DISCUSSION

For our experiment AVIRIS data set of INDIANA band has been selected. Among these total 220 bands, fifty bands have been selected from the wavelength 0.4µm -1.8µm. The image contains the 16 different classes. Among the sixteen classes nine classes are chosen for the classification purpose. For individual bands Markov random field features are calculated by defining a fifth order neighbourhood. From the extracted features summation features are derived.

The classifier produces the output, whether the pixel under test belongs to the interested trained class or not. Similarly other classes are also trained. Randomly selected pixels are tested against the training samples. By this way, classes are separated hierarchically.

After identifying the pixels of interested class it is labeled and indicated by white gray level. All other pixels are assigned black gray level. Then the pixels are displayed. The output is subtracted from the labeled ground truth and number of misclassified pixels is calculated. The accuracy of each class is calculated and shown in Table 1.

It is evident that the accuracy of classification is improved when the number of bands is increased. This is because that in each band the extracted features are different for the same object. Hence summation and concatenated features are used. Similarly the pixels which are very close to the central pixel contain more information than the pixels that are away from enter pixel. Thus the features are extracted in a hierarchical order.

![Input Image (First Band)](image)

![Ground Truth Image](image)

![Output Image](image)
Table 1
Accuracy Using SVM Classification

<table>
<thead>
<tr>
<th>Class</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corn-min</td>
<td>96.17</td>
</tr>
<tr>
<td>Building</td>
<td>85.66</td>
</tr>
<tr>
<td>Grass</td>
<td>96.22</td>
</tr>
<tr>
<td>Hay windrowed</td>
<td>81.77</td>
</tr>
<tr>
<td>Soy-clean</td>
<td>95.22</td>
</tr>
<tr>
<td>Soy-min</td>
<td>99.3</td>
</tr>
<tr>
<td>Soy-no till</td>
<td>99.25</td>
</tr>
<tr>
<td>Alfalfa</td>
<td>100</td>
</tr>
<tr>
<td>Woods</td>
<td>95.87</td>
</tr>
<tr>
<td>Average Accuracy</td>
<td>94.38</td>
</tr>
</tbody>
</table>

Soy-clean, Soy-min, Soy-no till are the classes which have similar tonal variations. But they can be properly distinguished by this proposed method. The fine texture 'Alfalfa', where the interdependency of pixels dominates exhibits 100% accuracy while using the MRF features. It is obvious that MRF outperforms Co-Occurrence, Run Length or any other features extraction techniques.

But the coarse textured classes like Building and Hay advocates poor accuracies, since the interdependencies are lower, and those classes are assimilated by the background. The proposed method produces good results, even without using the spectral features of the hyper spectral image. So this type of classifications is useful during the in availability of the spectral data.

6. CONCLUSION
The classification of hyper spectral remote sensing data using Support Vector Machines was investigated. SVM provided very accurate classification, even in the case of a very limited number of training samples and high dimensional data the proposed work can be repeated by using composite kernels.

REFERENCES