UNIVERSIDAD **DE EXTREMADURA**

Tesis Doctoral

Nuevas Técnicas de Clasificación Probabilística de Imágenes Hiperespectrales

New Probabilistic Classification Techniques for Hyperspectral Images

Autor: Mahdi Khodadadzadeh

DEPARTAMENTO DE TECNOLOGÍA DE LOS COMPUTADORES Y DE LAS COMUNICACIONES

Conformidad del Director: Antonio Plaza Miguel

Fdo.:

Conformidad del Director: Jun Li

Fdo.:

2015

Resumen

Los sensores hiperespectrales permiten captar cientos de imágenes, correspondientes a diferentes canales espectrales, para un mismo área en la superficie terrestre. Dado que diferentes materiales tienen características espectrales diferentes, dichas imágenes ofrecen una herramienta muy efectiva para discriminar y clasificar diferentes objetos. Sin embargo, existen varias dificultades a la hora de abordar el proceso de clasificacion de imágenes hiperespectrales. Entre ellos, destaca la gran dimensionalidad de las imágenes, así como la presencia de ruido y píxeles mezcla (es decir, píxeles en los que cohabitan diferentes sustancias). El desarrollo de técnicas de clasificación probabilística basada en subespacios permite abordar estos problemas de forma eficiente, y constituye la principal contribución del presente trabajo de tesis doctoral. En concreto, proponemos nuevos métodos de clasificación probabilística basada en regresión logística múltiple para estimar las probabilidades de las diferentes clases. Además, para abordar el problema relacionado con la presencia de píxeles mezcla en la escena, proponemos una metodología innovadora que integra probabilidades estimadas de forma local y global utilizando un nuevo concepto de mapa de combinación de clases. Otra contribución novedosa de la tesis doctoral es la integración de información espacial-contextual en el proceso de clasificación espectral, utilizando un método robusto de relajación probabilística que incluye información estimada a partir de mapas de discontinuidad derivados a partir de la imagen hiperespectral original. Por último, la tesis doctoral introduce un nuevo método para el aprendizaje múltiple de características que no requiere parámetros de regularización. El nuevo método se aplica a la fusión y clasificación conjunta de imágenes hiperespectrales e imágenes LiDAR (light detection and ranging). La eficacia de los nuevos desarrollos propuestos se compara con otros métodos que han proporcionado resultados destacados en el estado del arte.

Palabras clave: imágenes hiperespectrales, clasificación espacial-espectral, métodos basados en subespacios, máquinas de vectores de soporte, regresión logística múltiple, métodos de relajación probabilística, aprendizaje múltiple de características.

Abstract

Hyperspectral sensors provide hundreds of images, corresponding to different wavelength channels, for the same area on the surface of the Earth. Since different materials show different spectral properties, hyperspectral imagery is an effective technology for accurately discriminating and classifying materials. However, issues such as the high dimensionality of the data and the presence of noise and mixed pixels in the data, present several challenges for image classification. Dealing with these issues, this thesis proposes several new techniques for hyperspectral image classification. Developing subspace-based techniques for probabilistic classification is the main focus of the thesis. Specifically, we propose subspace-based multinomial logistic regression methods for learning the posterior probabilities. Furthermore, in order to better characterize mixed pixels in the scene, we propose an innovative method for the integration of the global posterior probability distributions and local probabilities which result from the whole image and a set of previously derived class combination maps, respectively. Another contribution of the thesis is the integration of spatial-contextual information using a robust relaxation method, which includes the information from the discontinuity maps estimated from the original image cube. Finally, the thesis introduces a new multiple features learning method which does not require any regularization or weight parameters. We apply the proposed method for fusion and classification of hyperspectral and LiDAR (light detection and ranging) data. The effectiveness of the proposed techniques is illustrated by using several simulated and real hyperspectral data sets and comparing with state-of-the-art methods.

Keywords: Hyperspectral imaging, spectral-spatial classification, subspace-based approaches, support vector machine, multinomial logistic regression, relaxation methods, multiple feature learning.

Acknowledgement

My deepest gratitude and appreciation go first and foremost to my supervisors, Prof. Antonio Plaza and Prof. Jun Li, for their great patience, precious guidance and generous support throughout my doctoral work. All your constant encouragement helped me to focus on my research which I deeply thank you. It has been my honor to work with you.

I would like to express my sincere appreciation to Prof. José Bioucas-Dias for his collaboration and constructive comments in most of the developments presented in this thesis. I am always impressed by his mathematical vigor and sharp thinking. I also gratefully acknowledge and appreciate his great help and support during a research visit to the Instituto Superior Técnico, Universidade de Lisboa, Portugal.

I would also like to thank the collaborations of Profs Hassan Ghassemian, Lianru Gao, Saurabh Prasad and Xiuping Jia who also contributed to some of the developments presented in this thesis work.

Acronyms

AA	Average Accuracy
AD	Anisotropic Diffusion
AF	Attribute Filter
AP	Attribute Profile
AVIRIS	Airborne Visible/InfraRed Imaging Spectrometer
CR	Continuous Relaxation
CRF	Conditional Random Field
DLR	Deutsches Zentrum für Luft- und Raumfahrt e.V. (German Aerospace Center)
DMP	Differential Morphological Profile
DSM	Digital Surface Model
EAP	Extended Attribute Profile
ECHO	Extraction and Classification of Homogeneous Objects
EMAP	Extended Multi-Attribute Profile
EMP	Extended Morphological Profile
GLCM	Grey Level Co-occurrence Matrix
GRSS	Geoscience and Remote Sensing Society
HMRF	Hidden Markov Random Field
HYSENS	HYbrid molecule nanocrystal assemblies for photonic and electronic SENSing applications
Hysime	Hyperspectral subspace identification with minimum error
ISODATA	Iterative Self-Organizing Data Analysis Technique Algorithm
IVM	Import Vector Machine
LIBSVM	Library for Support Vector Machines
LiDAR	Light Detection And Ranging
LMM	Linear Mixing Model
LOGP	Logarithmic Opinion Pool
LORSAL	LOgistic Regression via Splitting and Augmented Lagrangian
MAP	Maximum A Posteriori
MCS	Multiple Classifier System
MF	Majority Filter
ML	Maximum Likelihood
MLL	Multilevel Logistic

MLR	Multinomial Logistic Regression
MLR*sub*	Subspace Multinomial Logistic Regression
MM	Mathematical Morphology
MP	Morphological Profile
MRF	Markov Random Field
MV	Majority Voting
NCALM	National science foundation funded Center for Airborne Laser Mapping
NRS	Nearest-Regularized-Subspace
OA	Overall Accuracy
OSP	Orthogonal Subspace Projection
PCA	Principal Component Analysis
PLR	Probabilistic Label Relaxation
PR	Probabilistic Relaxation
RBF	Radial Basis Function
RF	Random Forest
ROSIS	Reflective Optics Spectrographic Imaging System
RSC	Robust Subspace Clustering
SE	Structural Element
SMLR	Sparse Multinomial Logistic Regression
SpATV	Spatially Adaptive Total Variation
SSC	Sparse Subspace Clustering
SVM	Support Vector Machine
TV	Total Variation
USGS	United States Geological Survey
UTC	Coordinated Universal Time

Contents

List of Figures

Table Index

Chapter 1

Introduction

This chapter presents a general introduction of this thesis. First, the context of the research work and the existing literature is reviewed. Then, the thesis objectives and main contributions are described.

1.1 Context and Background

The work developed in this thesis is part of the actual research lines of the Hyperspectral Computing Laboratory (HyperComp) group coordinated by Prof. Antonio Plaza Miguel at the Department of Technology of Computers and Communications, University of Extremadura, who has served as advisor of this work together with Prof. Jun Li. This work focuses on the development of new methodologies for probabilistic classification of remotely sensed hyperspectral images, as well as on the integration of spectral and spatial information to improve the obtained classification maps.

1.1.1 Hyperspectral versus multispectral imaging

Remotely sensed data acquisition from the surface of the Earth is a timely and effective way for gathering information about the natural and built environments for planning and management of human resources. For this purpose, multispectral imagery has been used from airborne and satellite systems since the 1960s [1]. However, for multispectral images, up to tens of spectral bands are usually provided in the visible and near-infrared region of the electromagnetic spectrum, which limits their functionality for Earth observation purposes. The idea of hyperspectral imaging for remote sensing emerged at NASA's Jet Propulsion Laboratory in 1983, where the AVIRIS [2] was developed for delivering high-dimensional data cubes with hundreds of contiguous spectral channels (bands) covering the wavelength region from 400 to 2500 nanometers. Nowadays, advanced hyperspectral sensor systems acquire the detailed spectrum of reflected light throughout the visible, near-infrared, and mid-infrared portions of the electromagnetic spectrum [1].

As shown in Fig. 1.1, in a hyperspectral image every pixel represents as a high-dimensional vector containing values corresponding to reflectance spectrum so that the size of the vector is equal to the number of spectral bands. In other words, the hyperspectral image can be considered as a three dimensional hyperspectral data cube which stacks several gray scale images together corresponding to different spectral channels from the same scene. Typically, several hundreds of spectral bands are available for hyperspectral images. This amount of spectral information available for each pixel of a scene increases the possibility of accurately distinguishing different physical materials. This is possible because different materials exhibit different spectral signatures. Fig. 1.1 shows the spectral signatures of
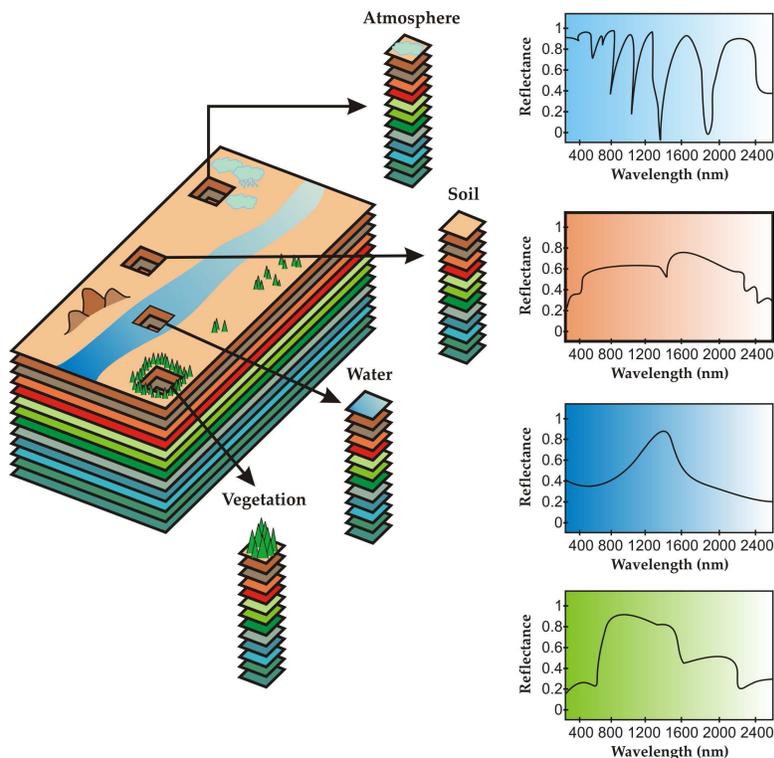
Figure 1.1: Hyperspectral data cube.

four different pixels from four materials, which show completely different behaviors in spectral domain. Besides these advanced remote sensing and Earth observation applications, the rich spectral attributes of hyperspectral data allow practical applications such as:

- Food Quality: Hyperspectral images can offer information about the chemical composition of food products. Therefore, the hyperspectral sensor systems are considered as powerful in-line inspection tools for increasing the quality and safety of food products such as poultry, fruits and vegetables [3, 4].

- Medical Sciences: For non-invasively diagnostic medical applications, hyperspectral images are used to get spectral information relating to the patient, tissue sample, or disease condition [5, 6].

- Mineralogy: Within the mining and mineral processing industries, hyperspectral imagery can be taken into account to identify several kinds of minerals. The in-line inspection systems composed of hyperspectral sensors increase the production capacity and efficiency [7, 8].

- Military Applications: Thanks to advances in hyperspectral sensor technology, the high spatial resolution hyperspectral data are considered for a wide range of military and defense applications such as target detection [9, 10].

1.1.2   Classification of hyperspectral images

Classification has been one of the most active areas of research and development in the field of hyperspectral image analysis [11]. Given a set of observations (*i.e.*, pixel vectors in a hyperspectral

image), the goal of classification is to assign a distinct class label to every pixel in the image [12]. Generally, the classification techniques are divided into two most common approaches, supervised and unsupervised as follows:

- Unsupervised classifiers: Based on a similarity criterion, unsupervised or clustering methods group pixels into sets of prototype groups called clusters, where the analyst merely identifies the number of clusters to generate in advance. Since multiple clusters may represent a single land cover class, generally merging clusters into a land cover type is necessary. There are different image clustering algorithms such as K-means [13] and ISODATA [14].

- Supervised classifiers: In these type of methods, the class labels and sets of representative samples for each class called training samples are specified *a priori* by a trained analyst, and each pixel of the image is labeled by taking information provided by the training samples. The training samples can be obtained based on maps and aerial photography or based on some on-site measurements. It is important to note that, in the remote sensing community, usually supervised classification is simply called classification.

Various methods have been suggested for classification of hyperspectral images. Although, it seems that hyperspectral images are not extremely different from multispectral images, in reality the analysis of hyperspectral images is more challenging and conventional classifiers such as the ML [15, 11, 12], nearest neighbors [16] and neural networks [17, 18, 19] may be not suitable for hyperspectral images. Because of high dimensionality and volume issues, the degradation mechanisms associated to the measurement process (e.g., noise and atmosphere), the high spatial and spectral redundancy, and their potential non-linear nature the classification of hyperspectral images has been proved to be a very complex task [20, 21]. For example, in the ML method, performance is strictly related to the quality and number of training samples. A sufficient number of training samples is required for proper parameter estimation of this classifier. It is often stated that the number of training samples for each class should comprise at least 10-30 times the number of wavebands [22], which for hyperspectral data (with hundreds of bands) is often unfeasible. This difficulty, i.e. the small ratio between the number of available training samples and the number of bands is called *curse of dimensionality* (i.e., the Hughes phenomenon [23]), which increases the noise and error factor of the classifier and makes it impossible to obtain reasonable estimates of the classifier parameters. Among conventional classifiers, the distribution-free neural network based classifiers of the 1990s have been broadly applied for classifying hyperspectral images. However, it is always expected to achieve sufficient overall image classification accuracy by using efficient and robust processing systems, thus the use of neural networks for hyperspectral image classification is not very common now, mainly due to their algorithmic and training complexity [24].

In recent years, SVMs have offered a popular classification framework for hyperspectral images, which has shown good performance with small training sample sizes [25, 26]. The SVM is a binary hard classifier that separates two classes by using a linear hyperplane. For finding the separating hyperplane, the training samples located near the class boundaries (i.e., the support vectors) that maximize the separation between the two classes are found, by using a constrained optimization process [27]. An important advantage of SVM algorithm is the possibility of using kernel trick in the formulation [28]. Because of several nonlinear phenomena present in the hyperspectral image, including such nonlinear ability in the kernel and thus constructing a nonlinear classifier is a very effective approach. In this way, the hyperspectral image may become linearly separable by projecting the pixel vectors into a higher dimensional space [24]. One of the main drawbacks of the SVM classifier is that it is originally a binary

3

hard classifier. The extension of SVM to the multi-class cases is usually done by combining several binary classifiers using the *one versus the rest* and the *one versus one* methods [28]. Most importantly, in [29] two techniques have been proposed for computing probabilities for the SVM classifier based on combining all pairwise comparisons. This method is implemented in the LIBSVM [30]. An extensive literature is available reporting techniques using the SVM classifier for classification of hyperspectral images (see, e.g. [25, 26, 31, 32]). IVM [33] is another kernel-based approach that was efficiently utilized for hyperspectral image classification [34, 35]. This classifier has the advantage of direct provision of a probabilistic output. Selection of an adequate kernel and the need of using techniques like cross-validation to estimate regularization parameters are the important disadvantages of the kernel-based methods SVM and IVM.

MCS (or classifier ensemble) methods are another effective approaches to process hyperspectral data when only a limited training samples are available [36, 37]. According to a certain combination approach (such as majority vote, Bayesian rule, etc.) or based on an iterative error minimization, MCS integrates the outputs of individual classifiers. In particular, RFs have shown remarkable performance in terms of dealing with high-dimensional datasets [38, 39]. In RF algorithm, the forest composed of many decision trees as binary classifiers, where each tree gives a unit vote for a particular class and the forest chooses the class that has the most votes. Very recently, rotation forests, have been successfully applied to hyperspectral remote sensing image classification [40]. This method consists in splitting the feature set into several subsets, running transformation algorithms separately on each subset and then reassembling a new extracted feature set. Different splits of the feature set lead to different rotations. Thus diverse classifiers are obtained. As reported in [40] this method works much better than other ensemble methods. The important drawback of the random and rotation forest methods is the need of a considerable amount of memory since these methods always store big matrices in memory [41].

Recently, the MLR algorithm has been adopted in hyperspectral image processing as a discriminative classifier [42, 43]. This method generalizes logistic regression to multiclass problems and exhibits a very important advantage of being able to model the posterior class distributions in a Bayesian framework. Logistic regression is a statistical binary classification method that fits data to a logistic function and tries to model the conditional probability of the class label given its observation [44]. In a simple perception, in the training phase of MLR algorithm (according to the number of class labels) the regression vectors are learnt by using the training samples. And then in the classification phase the posterior class probability of each test sample is estimated using all the regression vectors. The SMLR [45] adopts a Laplacian prior enforcing sparsity and therefore controlling the machine generalization capabilities. The LORSAL algorithm was introduced in [46], which is a faster algorithm than SMLR for processing of hyperspectral images with a very large number of classes.

1.1.3 Subspace-based methods

It has been proved that the original spectral features in a hyperspectral image contain high redundancy and there is a high correlation between adjacent bands [12]. Therefore, based on this assumption that the hyperspectral data may effectively live in a lower-dimensional subspace [47], several subspace-based methods have been proposed for classification [48, 49, 50, 51]. We can consider the following three main advantages to use subspace-based classification methods: 1) Reducing the dimensionality of hyperspectral data by projecting it to a precise subspace without losing the original spectral information; 2) Increasing the separability of the classes which are very similar in spectral sense; 3) Handling the effects of noise and the presence of heavily mixed pixels in a hyperspectral image. However, identification of subspaces

i.e. the number of subspaces and their dimensions, is a crucial step in subspace-based classification algorithms. A comprehensive discussion of hyperspectral subspace identification is available in [47], where a technique called Hysime was presented to identify an unique subspace in which the hyperspectral data live. The developers proposed the computationally efficient and automatic Hysime method using the estimation of the signal and the noise correlation matrices and then selecting the subset of eigenvalues that best represents the signal subspace in the least squared error sense. An extensive literature is available on the subspace-based classification methods. For instance, in [52] (later revisited in [48]), the OSP technique was introduced for hyperspectral image classification and dimensionality reduction based on projecting each pixel vector onto a subspace which is orthogonal to the undesired or interfering signatures. In [53], the OSP technique was revised and extended by three unconstrained least squares subspace projection approaches to estimate signature abundance and classify a target signature at subpixel scale. In [54], a kernel-based subspace projection technique was presented and evaluated in the context of an agricultural application. Perhaps, one of the most successful subspace-projection-based classifiers for processing of hyperspectral images has been MLR*sub* [51]. This classifier models the subspace associated with each specific class in an MLR framework. In other words, based on the basic assumption that the samples within each class can approximately lie in a lower dimensional subspace, the MLR*sub* uses a class dependent procedure integrated with the MLR classifier to represent each class using a subspace spanned by a set of basis vectors estimated from training samples. This approach has exhibited good classification performance using several hyperspectral scenes. More recently, in [55] NRS classifier was proposed, which couples nearest-subspace classification with the distance-weighted Tikhonov regularization [56] to seek an approximation of each testing sample via a linear combination of training samples within each class.

1.1.4 Integration of spatial and spectral information

The hyperspectral images classification maps resulting from conventional pixel-based classifiers using only spectral information show a large amount of salt and pepper effects, which decreases the classification accuracy significantly. In 1980's, Landgrebe and his research group proved that the methods for classifying multispectral remote sensing images provide more smoothed maps by integrating spectral properties of adjacent pixels [57, 58]. The basic assumption behind this approach was that a typical scene may consist several homogeneous regions which are large compared to the size of a pixel, and consequently, the pixels in the neighborhood are likely to assign the same class labels. Considering this fact that modern hyperspectral sensors provide very fine spatial resolution images to obtain more detailed information for the scene, integrating spectral and spatial-contextual information is a great advantage for improving the quality of the classification maps and as a result, classification accuracies. During recent years, various approaches have been discussed in the framework of spectral-spatial hyperspectral image classification [59].

1.1.4.1 Spatial preprocessing prior to classification

Performing image segmentation as a pre-classification step, so-called object-based classification, is a traditional and effective approach to use spatial information in order to improve classification accuracy. Segmentation can be defined as dividing a given image into homogeneous regions (*i.e.* objects) using appropriate homogeneity criterion [60]. The pixels of each region from a segmentation map would have spatial and spectral similarities. Techniques such as watershed [61, 62], petitional clustering [63, 64], and hierarchical segmentation [65, 66, 67] have been applied previously on remote sensing images. Several object-based classifier have been proposed in the literature. The famous ECHO algorithm uses an ML

method to classify mean vector of each region and then assign pixels belonging to the regions to label of their mean vectors [57]. Van der *et al.* also used segmentation but improved classification of mean vectors by using an SVM classifier [68]. More recently, in [63, 69, 70] Tarabalka *et al.* proposed to perform pixelwise classification and segmentation independently, and then, combine the results using the MV rule. While in [64] weighted MV rule was suggested for decision fusion, where pixels in the same segment contribute differently according to their distance to the spectral centroid.

1.1.4.2 Exploiting texture features

Another common strategy of spectral-spatial classification of hyperspectral images is extracting different kinds of spatial features and using them as complements of the spectral signatures in a classification procedure. For these methods two important issues need to be fulfilled with are: 1) adopting suitable spatial features and 2) developing an effective strategy for integrating spatial features into a classifier.

Generating texture profiles of hyperspectral image cubes has been widely used in the literature. For the texture information, the well-known way is to locally generate pixel based spatial measurements such as contrast, entropy or energy using a fixed neighborhood based procedure [12]. The GLCM is a traditional method for quantitative characterization of texture [71]. The GLCM is formed by calculating the occurrences of two neighboring pixels with specific values in an image by a given spatial relationship such as distance and direction. Based on GLCM, several statistical measures like Energy, contrast, correlation, entropy, and homogeneity can be extracted. The conventional 2-D GLCM is applied on a single band image, thus for hyperspectral images, the GLCM is usually calculated from the first principal component, which neglects the rich spectral information of hyperspectral images. In [72] a 3-D version of GLCM was proposed for hyperspectral images to deal with this issue. Moreover, the Gabor filters [73] and wavelet feature extraction methods [74, 75] have been recently proposed for 3-D texture analysis of hyperspectral data.

1.1.4.3 Exploiting morphological features

Morphological filters [76] based approaches have shown great potential for extracting spatial features from remotely sensed images. Opening and closing (which are the combinations of erosion and dilation operators) are the applicable operators for mathematical morphology which work using an SE of predefined size and shape. In a simple description, the opening and closing operators remove structures smaller than the size of the SE. By changing the size of the SE and repeating the opening and closing operations over a gray scale image, MP can be generated. Opening and closing by reconstruction [76] are more useful operations for extracting spatial information, since they allow removing of undesirable structures without considerably changing the shape of those structures of the image, and will be helpful for recovering structures which are not completely removed. Opening by reconstruction removes unconnected light structures of a gray scale image, whereas closing by reconstruction removes unconnected dark objects. In [77], using the residuals from opening and closing by reconstruction operations so-called DMP was suggested for feature extraction of very high resolution panchromatic images. In [78], the concept of MPs has been successfully extended to hyperspectral image classification. In this paper, they suggested EMPs by performing MP on the few principal components of the data resulted from PCA [12] feature extraction method. Respect to the advantages of using morphological AFs [79], recently, APs [80, 81, 82] were introduced as the generalization of the MP. Generating APs based on attributes such as area, volume, standard deviation, etc. has shown to be a relevant technique to model different kinds of the structural information available in remote sensing

(a)	(b)	(c)	(d)	(e)	(f)	(g)
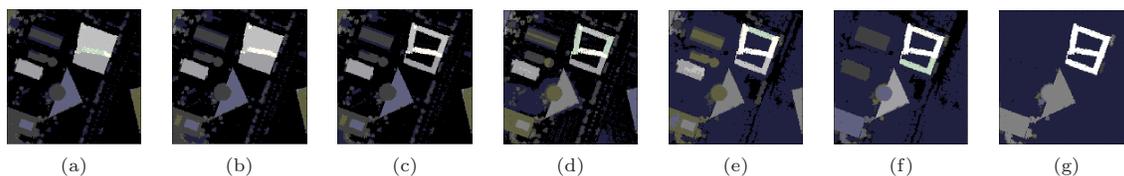
Figure 1.2: An example of an AP using attribute of area with conditions $\lambda_1 = 100$, $\lambda_2 = 500$ and $\lambda_3 = 1000$. The AP is composed of thickening images (a)-(c), the original image (d), and thinning images (e)-(g).

images. For generating an AP based on a specific attribute a predefined condition λ is needed. If the measured attribute of an object meets λ then that object is kept unchanged; otherwise, it is merged to the most similar surrounding object. So that the merging operation to the adjacent object of a lower and higher gray level are called thinning and thickening, respectively. Fig. 1.2 shows an example of generating AP based on attribute of area which can provide information about the size and the shape of objects in the image. Moreover, as illustrated in Fig. 1.3, EAPs are typically obtained by generating an AP on each of the first few PCA components (or any other features retained after applying feature selection on the original hyperspectral image) thus building a stacked vector using the AP on each feature. From the EAP definition, the consideration of multiple attributes leads to the concept of EMAP [83] which combines the EAPs by concatenating them in a single vector of features and improves the capability in extracting the spatial characteristics of the structures in the scene. Although generating EMAP is computationally heavy and time consuming, attribute filtering can be efficiently computed by applying a Max-tree algorithm [84]. Choosing suitable attributes and threshold values are the main difficulties of using the EMAP.

Besides spatial feature extraction issue, the other challenge in multiple feature learning is that how to adequately exploit the information (*i.e.* spectral and spatial) containing in these features. In a traditional way, the spatial features may be added as extra features on top of spectral bands for improving classification. For example in [85], spatial feature extraction was done using GLCM and then SVM classifier was applied on a stack vector of textural and spectral features. However, the simple concatenation or stacking of features such as morphological APs and spectral features may contain redundant information. In addition, as described before, a significant increase in the number of features may lead to high dimensionality issues that are in contrast with the limited number of training samples often available in remote sensing applications. In [86], applying feature extraction on both the EMP and the original hyperspectral data and concatenating the extracted feature vectors into one stacked vector was proposed. Exploiting spectral-spatial composite kernels is another strategy for classification purposes. Especially the composite kernels and multiple kernel learning methods in the framework of supervised SVMs have shown a significant capacity to integrate multiple types of features [87, 88, 89]. The main limitations of SVM-based composite kernel methods are the requirement of the convex combinations of kernels and optimizing the parameters. Recently, in [90] a successful framework has been introduced for the development of generalized composite kernel machines for spectral-spatial hyperspectral image classification, which equally balances the spectral and the spatial information contained in the hyperspectral data without any weight parameters. A distinguishing feature of the method in [90] is that it uses the MLR classifier, which naturally provides a probabilistic output and has a lot of flexibility in the construction of nonlinear kernels.
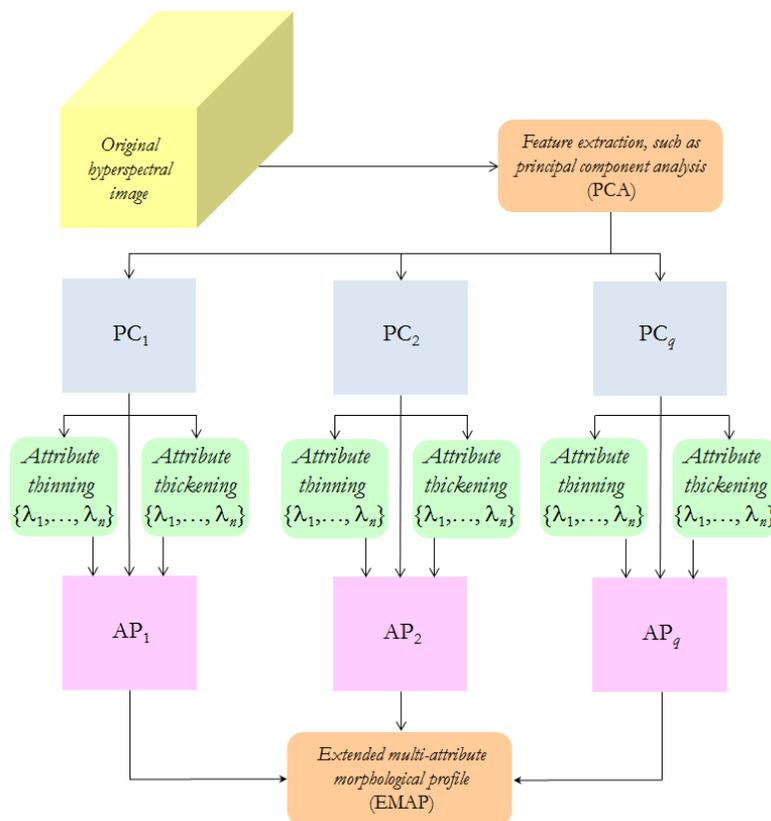
Figure 1.3: Graphical illustration of the procedure adopted in order to construct an EAP from a hyperspectral image. The EMAP is a combination of EAPs obtained with different attributes.

1.1.4.4 Spatial postprocessing after classification

Another well-known way to include spatial information in a per-pixel classification consists in performing postprocessing or postclassification [91]. This category includes traditional iterative methods, based on using MFs, which have been shown to be beneficial in removing most of the noise from the classification map [92]. The MF is a logical filter which relabels the center pixel of a neighborhood if it is not a member of majority class. Usually, using MFs cause the problem of removing linear features such as roads, sports tracks, buildings, etc., when trying to smooth the classification map [93]. On the other hand, by using probabilistic classification method, not only the estimated class label and consequently the classification map is known, but also the valuable membership information is available for each pixel in order to show how close the pixel is to the other classes. As a result, incorporating spatial-contextual information into the obtained probabilistic classification results is more effective. These normally iterative methods are broadly referred to as CR or PR [94]. In other words, after a probabilistic pixel-wise classification of the hyperspectral image, the postprocess of PR is applied to exploit the continuity, in probability sense, of neighboring labels. In recent decades, a considerable amount of research has been devoted for PR algorithms. One of the first approaches to include spatial-contextual information in probabilistic classification was PLR [95, 12]. PLR strategies use the probabilities of neighboring pixels iteratively to update the class probabilities for the center pixel based on a neighborhood function. A neighborhood function can be defined using the probabilities of the neighboring pixels, compatibility coefficients, and neighborhood weights. A *priori* compatibility coefficient describes numerically how compatible it is to
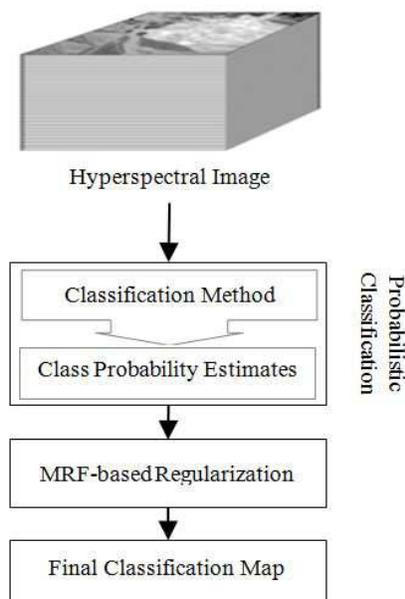
Figure 1.4: Flowchart of MRF-based spectral-spatial classification of hyperspectral image

have pixel m classified as class k and neighboring pixel n classified as $\acute{k}$ [12]. Recently, in order to correctly differentiate linear boundaries between classes, the PLR method was successfully extended for including of third-order statistics (*i.e.* using triplets of neighborhood pixels) within the probabilistic likelihood calculation [93]. The main limitation of PLR-based method is determining compatibility coefficients, which are normally calculated based on priori knowledge of the area under investigation and the sensor characteristics [96].

Perhaps the most popular PR strategy is based on the use of MRFs [97, 98]. MRF theory provides a convenient and consistent way of modeling spatial correlation of neighboring pixels in an image and widely used in remote sensing society to integrate spatial information into pixel-wise image classification problems. Practically, for the mathematically tractability, the statistical dependence among the labels in an MRF model is defined based on Hammersley-Clifford theorem about the equivalence between MRF and Gibbs distributions [97]. In a simple word, by using MRF model for classification of a pixel, the information from its neighbors is considered. Normally, MRF is used together with the MAP decision and estimation theory to formulate an objective function and in the MAP-MRF framework, the objective is the joint posterior probability of the MRF labels [97]. MRF-based approach can be implemented in two steps. First, a probabilistic pixelwise classification method is applied to learn the posterior probability distributions from the spectral information. Second, contextual information is included by means of an MRF regularization to refine the classification (see Fig. 1.4). Therefore, estimating class conditional probability distributions is an intrinsic issue when using conventional MRF. Typically ML classifier is applied to obtain an initial classification [99, 100]. However, because of poor estimates of the true mean vector and covariance matrix of each class, ML classifier is not efficient for hyperspectral data with limited training samples [11]. Unlike traditional methods, probabilistic SVM [29] is a distribution-free algorithm that can overcome the problem of poor statistical estimation and has shown good performance for classifying hyperspectral data. Several studies have been undertaken to integrate SVM within MRF

framework for accurate spectral-spatial classification of remote sensing images, and satisfactory results have been reported. For instance, in [101], class-conditional distributions were estimated by the Mean Field-based SVM regression algorithm. Spectral-spatial hyperspectral image classification was performed in [102], given the initial SVM classification map and a final MRF-based regularization process. For computing the MAP estimate of the true classification map, they adopted Metropolis algorithm based on stochastic relaxation and annealing [103, 104]. An adaptive MRF approach was proposed in [105], for hyperspectral image classification. They introduced a relative homogeneity index for each pixel to determine an appropriate weighting coefficient for the spatial contribution in the MRF regularization procedure. In [106], a novel and rigorous framework was proposed for contextual hyperspectral image classification, which combines SVMs and MRFs in a unique formulation. The authors in [107], proposed a strategy to fully exploit the spatial information by combining fuzzy SVM classifier and an MRF-based graph cut algorithm. Another recent tendency is to integrate MLR and MRF algorithms for spatial contextual classification. For example, very recently, combining SMLR algorithm [45] with an SpATV regularization was proposed, which showed significant performance [108]. In another effort to use MRF-based prior, in [109], the hyperspectral classification results were obtained by maximizing the marginal probability of the posterior distribution using the loopy belief propagation method, where the posterior class probability was modeled as an MLR classifier and an MRF. Particularly, in order to encourage neighboring pixels to have the same label, combining MRF-based multilevel logistic prior with MLRsub algorithm was proposed in [51]. For this purpose, they suggested to use an efficient min-cut optimization technique (α-Expansion algorithm) for computing a MAP segmentation [110], which showed significant performance. Similar strategy of using the graph-cut-based α-expansion algorithm has been employed in [111] and [112] for integrating spatial information with spectral classification results obtained by Gaussian mixture model and probabilistic sparse representation, respectively.

As a contextual framework, CRFs model the class posterior probability in a way that incorporates the spatial property in both the labeling field and the observation field, simultaneously [113]. There is some work in the literature using CRFs for spectral-spatial classification of hyperspectral images. Particularly, Zhong and Wang have done extensive research about the exploitation of CRFs for hyperspectral image classification [114, 115, 116]. However, for CRF-based classification, it is necessary to train several model parameters which is time consuming task and needs large number of training samples. Moreover, the reported experimental results show that, in most cases, the CRF-based methods do not demonstrate an improvement over the MRF-based methods. In another attempt for spectral-spatial classification of hyperspectral data, in [117], a generalization of MRF named HMRF was used. They proposed to combine the results of HMRF segmentation with SVM pixel-wise classification using majority voting within each object.

1.2 Objectives

The main objective of the thesis work is to develop new techniques for probabilistic spectral-spatial classification of hyperspectral images. Particularly, the focus is on the integration of the spectral information (coming from probabilistic pixel-wise classification) with spatial information exploited from probabilistic relaxation procedures. In order to achieve this general objective, we will address a number of specific objectives which are listed below:

1. It has been proved that projecting hyperspectral data to a precise subspace (or union of subspaces) can increase the separability of the classes. In addition to dealing with the problem of high

dimensionality, subspace-based classification methods can better handle the effects of noise and the presence of mixed pixels in a hyperspectral image. Consequently, this has been considered as an objective of this thesis work.

2. The probabilistic classification of hyperspectral images has several advantages over standard deterministic classification which simply obtains the "best" class label for each pixel. In other words, probabilistic algorithms allow us to assess classification uncertainty. Most importantly, because of the way the probabilities are generated, probabilistic classifiers can be more effectively incorporated into information fusion procedures. As described in the introduction, for refining the classification results using MRF-based regularization process, a probabilistic classifier is needed. In order to address this issue and increase the accuracy of spectral-spatial classification methods, better estimation of probabilities is crucial.

3. In order to reduce the effect of the degradation of the classification performance in the neighborhood of the class boundaries observed for existing relaxation strategies, as well as increasing the efficiency of those classification methods, developing a robust relaxation method is particularly important. As a result, we give special attention to the development of a framework able to include the information from the discontinuity maps estimated from the original image cube to a spatial relaxation procedure.

4. Efficiently exploiting the information coming from multiple features is a challenging issue in multiple feature learning approaches. Therefore, in this thesis, addressing this issue is a particular highlight. Specifically, we propose a robust technique for exploiting different types of features without any regularization parameters is considered.

5. The integration of hyperspectral data with other sources of data is also a very important aspect addressed in this thesis work. This is particularly critical in scenarios in which hyperspectral data alone cannot characterize the information classes (despite the exploitation of the very rich spatial and spectral information contained in hyperspectral images). As a result, in this thesis work we also deal particularly with the integration of hyperspectral data with other widely available sources such as LiDAR data.

1.3 Thesis Organization

1.3.1 Novel contributions of the thesis

The main contributions of the thesis are summarized in Fig. 1.5. As shown in this figure, all the newly developed techniques for probabilistic classification of hyperspectral images, are designed in a way that exploits the subspace projection concept. Moreover, one chapter of this thesis is specifically allocated for the relaxation procedures. In the following, we provide a description of the different chapters in which we have structured the present thesis work:

In Chapter 2, a new spectral-spatial classifier is presented for hyperspectral data that specifically addresses the issue of mixed pixel characterization. In this approach, the spectral information is characterized both locally and globally, which represents an innovation with regards to previous approaches for probabilistic classification of hyperspectral data.

In Chapter 3, a new MLR method is proposed for pixelwise hyperspectral classification, in which the feature vectors are formed by the energy of the spectral vectors projected on class-indexed subspaces.
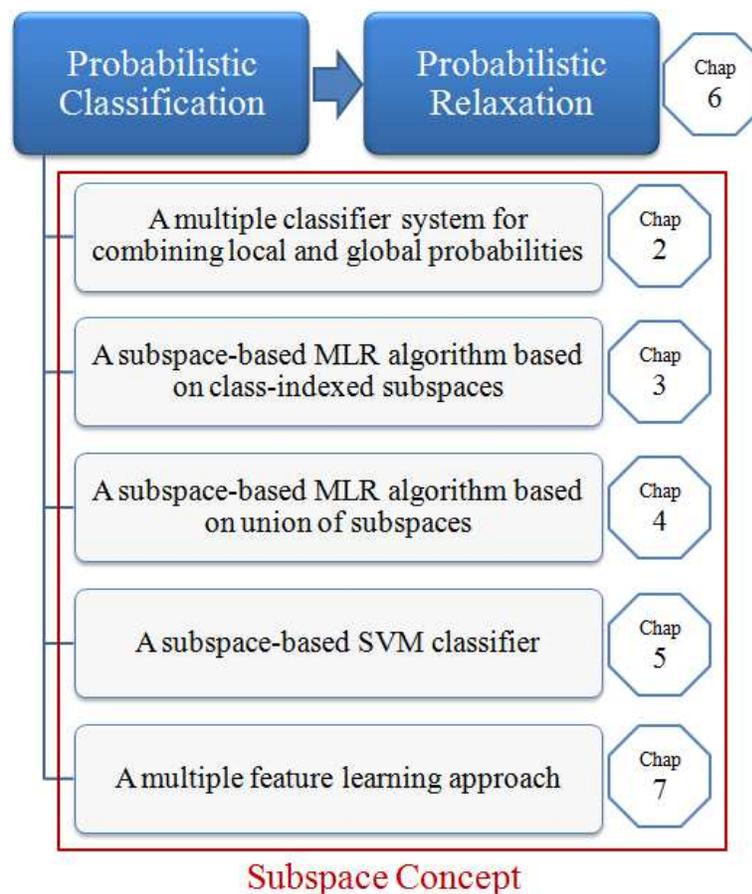
Figure 1.5: Flowchart illustrating the organization of this thesis.

The main rationale for this approach is that the hyperspectral measurements are often linear mixtures of the endmembers signatures and, then, each class corresponds to a given subset of the endmembers, thus defining a subspace.

In Chapter 4, a new probabilistic classification approach is presented based on union of subspaces for hyperspectral images. The proposed method integrates subspace clustering with MLR method for supervised hyperspectral image classification.

In Chapter 5, the subspace-projection-based concept is extended to SVMs, a very popular technique for remote sensing image classification by constructing the SVM nonlinear functions using the subspaces associated to each class.

In Chapter 6, a new discontinuity preserving relaxation strategy is developed which can be used for postprocessing of class probability estimates, as well as preprocessing of the original hyperspectral image. The newly proposed method is an iterative relaxation procedure which exploits spatial information such a way that it considers discontinuities existing in the data cube.

In Chapter 7, a new probabilistic classification strategy is proposed to efficiently exploit the information coming from multiple features. An important characteristic of the presented fusion approach is that it does not require any regularization parameters, so that different types of features can be integrated in a collaborative and flexible way. In this work, in addition to the spatial and the spectral

Table 1.1: List of mathematical notations used in this thesis.

Mathematical Notations	
n	Number of pixels of an image
$S = \{1, 2, ..., n\}$	Set of indexing integers
d	Number of spectral bands
$\mathbf{x}_i = [x_{i1}, x_{i2}, ..., x_{id}]^T$	Spectral vector with dimension of d for $i \in S$
$\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$	Hyperspectral image
k	Number of classes
$\mathcal{K} = \{1, 2, \ldots, k\}$	Set of k classes
$y_i \in \mathcal{K}$	Class label for $i \in S$
$\mathbf{y} = \{y_1, y_2, \ldots, y_n\}$	Labels of whole image
$n_{tr}^{(c)} \in \{n_{tr}^{(1)}, n_{tr}^{(2)}, \ldots, n_{tr}^{(k)}\}$	Number of training samples for class $c \in \mathcal{K}$
$\mathcal{D}^{(c)} \equiv \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_{n_{tr}^{(c)}}, y_{n_{tr}^{(c)}})\}$	Labeled training set for class $c \in \mathcal{K}$
$n_{tr} = \sum_{l=1}^{k} n_{tr}^{(l)}$	Total number of samples in the training set
$\mathcal{D} \equiv \{\mathcal{D}^{(1)}, \mathcal{D}^{(2)}, \ldots, \mathcal{D}^{(K)}\}$	Whole training set

features extracted from hyperspectral data (as a source of complementary information) we use LiDAR-based features to illustrate the proposed approach.

To conclude this chapter, the mathematical notations used throughout the thesis have been defined in Table 1.1.

Chapter 2

Spectral-Spatial Classification of Hyperspectral Data Using Local and Global Probabilities

2.1   Summary

Remotely sensed hyperspectral image classification is a very challenging task. This is due to many different aspects, such as the presence of mixed pixels in the data or the limited information available *a priori*. This has fostered the need to develop techniques able to exploit the rich spatial and spectral information present in the scenes while, at the same time, dealing with mixed pixels and limited training samples. In this chapter, we present a new spectral-spatial classifier for hyperspectral data that specifically addresses the issue of mixed pixel characterization. In our presented approach, the spectral information is characterized both locally and globally, which represents an innovation with regard to previous approaches for probabilistic classification of hyperspectral data. Specifically, we use the MLR*sub* method for learning the posterior probabilities and a pixel-based probabilistic SVM classifier as an indicator to locally determine the number of mixed components that participate in each pixel. The information provided by local and global probabilities is then fused and interpreted in order to characterize mixed pixels. Finally, spatial information is characterized by including an MRF regularizer. Our experimental results, conducted using both synthetic and real hyperspectral images, indicate that the proposed classifier leads to state-of-the-art performance when compared with other approaches, particularly in scenarios in which very limited training samples are available [1].

2.2   Introduction

Hyperspectral imaging instruments are now able to collect hundreds of images, corresponding to different wavelength channels, for the same area on the surface of the Earth [1]. Hyperspectral image classification has been a very active area of research in recent years [11]. Given a set of observations (*i.e.*, pixel vectors in a hyperspectral image), the goal of classification is to assign a unique label to each pixel vector so that it is well defined by a given class [12]. The availability of hyperspectral data with high spectral resolution

---

[1]Part of this chapter has been published in: M. Khodadadzadeh, J. Li, A. Plaza, H. Ghassemian, J. M. Bioucas-Dias and X. Li. **Spectral-Spatial Classification of Hyperspectral Data Using Local and Global Probabilities for Mixed Pixel Characterization**. IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 10, pp. 6298-6314, October 2014 [JCR(2013)=2.933].

has been quite important for many applications, such as crop mapping, environmental monitoring, and object identification for defense purposes [59].

Several techniques have been used to perform supervised classification of hyperspectral data. Classic techniques include ML [15, 11, 12], nearest neighbor classifiers [16], or neural networks [17, 18, 19], among many others [59]. The quality of these pixelwise classification methods is strongly related to the quality and number of training samples. In order to effectively learn the parameters of the classifier, a sufficient number of training samples are required. However, training samples are difficult and expensive to collect in practice [22]. This issue is quite problematic in hyperspectral analysis, in which there is often an unbalance between the high dimensionality of the data and the limited number of training samples available in practice, known as the Hughes effect [11].

In this context, kernel methods such as the SVM have been widely used in hyperspectral imaging to deal effectively with the Hughes phenomenon by addressing large input spaces and producing sparse solutions [25, 26, 31, 32]. Recently, the MLR [43] has been shown to provide an alternative approach to deal with ill-posed problems. This approach has been explored in hyperspectral imaging as a technique able to model the posterior class distributions in a Bayesian framework, thus supplying (in addition to the boundaries between the classes) a degree of plausibility for such classes [43]. A main difference between the MLR and other classifiers such as the probabilistic SVM is the fact that the former learns the posterior probabilities for the whole image. As a result, these classifiers exploit the probabilistic information in a different (possibly complementary) fashion, although this issue has never been explored in the literature in the past. Recently, the advantages of probabilistic SVM as a soft classification technique in discriminating between pure and mixed pixels, and automatically selecting endmember subsets were, respectively, investigated in [118] and [119]. These techniques pay particular attention to characterizing the number of mixtures participating in each pixel. A subspace-based version of the MLR classifier, called MLR*sub* [51], has also been recently developed. This method relies on the basic assumption that the samples within each class can approximately lie in a lower dimensional subspace and uses subspace projection methods to find this subspace. Since hyperspectral data are likely to be noisy and dominated by mixed pixels, the MLR*sub* has been shown to provide good performance (particularly, in the case of limited training samples) as normally classes live in a much lower space in comparison with the original data dimensionality.

Another strategy to deal with the limited number of training samples available in practice has been to efficiently exploit labeled information by using multiple classifier systems or classifier *ensembles* [120, 36, 121, 37]. This approach has been proved successful in different hyperspectral image classification applications [38, 39, 122, 123].

Finally, a well-known trend in order to alleviate the problem of insufficient number of training samples is to integrate the spatial-contextual information in the analysis. Many examples of spectral-spatial classifiers can be found in the hyperspectral imaging literature [57, 58, 59, 60, 124, 89, 68, 86]. In particular, approaches based on MRFs have been quite successful in hyperspectral imaging [97, 101, 102, 42, 43, 51]. In particular, [102] successfully combined a probabilistic SVM with an MRF regularizer for the classification of hyperspectral images. All of these methods exploit, in a way or another, the continuity (in probability sense) of neighboring labels. In other words, these methods exploit the likely fact that, in a hyperspectral image, two neighboring pixels may have the same label.

In this chapter, we propose a new spectral-spatial classifier in which the spectral information is characterized both locally and globally. Specifically, we use the MLR*sub* method to globally and locally learn the posterior probabilities for each pixel, where addressing the local probability is one of the main innovative contributions of this work. For local probability learning, we determine the number of mixed

components that participate in each pixel. For this purpose, we use a probabilistic SVM as an indicator to determine the number of mixed components. Finally, the spatial information is then characterized by exploiting an MRF regularizer.

When compared to the probabilistic SVM, the presented classifier considers mixtures in the model. This is very important since hyperspectral images are often dominated by mixed pixels. When compared to the MLR*sub*, which already addresses the presence of mixed pixels, the proposed classifier constrains the number of mixed components, thus improving its characterization since mixed pixels in hyperspectral images normally comprise only a few mixing components [125]. As a result, the presented approach provides two important contributions with regard to existing spectral-spatial approaches. The first one is the consideration of probabilistic information at both local and global levels. The second one is the characterization of the number of mixtures participating in each pixel, which is quite important since mixed pixels often dominate hyperspectral data.

The presented approach also observes two of the most pressing needs of current hyperspectral classifiers: the possibility to use very limited training sets (compensated by the multiple classifier flavor of our approach) and the need to integrate spatial information in the assessment (addressed by the inclusion of an MRF regularizer in the formulation). The resulting method, called SVM-MLR*sub*-MRF, achieves very good classification results which are competitive or superior to those provided by many other state-of-the-art supervised classifiers for hyperspectral image analysis.

The remainder of this chapter is organized as follows. Section 2.3 describes the different strategies used to implement the proposed spectral-spatial classifier. Section 2.4 describes the proposed approach. An important observation is that the presented approach should not be simply understood as a combination of existing approaches. Specifically, each of the processing algorithms described in Section 2.4 correspond to one out of many possible choices, selected based on their availability and also on the possibility to draw comparisons with other established techniques for spectral-spatial classification. However, it should be noted that other strategies for addressing local versus global information for mixed pixel characterization and spatial regularization could be used. In this regard, our selection should be strictly understood as a vehicle to demonstrate a new framework for classification of hyperspectral data and not merely as a combination of processing blocks. To the best of our knowledge, the presented framework addresses for the first time in the literature the aforementioned aspects in synergistic fashion. Section 2.5 presents extensive experiments using both simulated and real hyperspectral data designed in order to validate the method and provide comparisons with other state-of-the-art classifiers. Section 2.6 concludes with some remarks and hints at plausible future research lines.

2.3   Main components of the proposed method

In this section, we describe the different components that have been used in the development of the proposed method. First, we use probabilistic pixelwise classification methods to learn the posterior probability distributions from the spectral information. Here, we use two strategies to characterize spectral information: probabilistic SVM and MLR*sub*. Then, we use contextual information by means of an MRF regularization scheme to refine the classification results. As it can be observed from Fig. 1.4, estimating class conditional probability distributions is an intrinsic issue for the subsequent MRF-based classification. In the following, we outline the different strategies used to characterize spectral and spatial information, respectively, in the presented approach.

2.3.1 Characterization of spectral information

With the aforementioned definitions (Table 1.1) in mind, probabilistic pixelwise classification intends to obtain, for a given pixel $\mathbf{x}_i$, the class label y_i. This label can be obtained by computing the posterior probability $p(y_i|\mathbf{x}_i, \mathcal{D})$ as follows:

$$y_i = c, \quad if \quad p(y_i = c|\mathbf{x}_i, \mathcal{D}) > p(y_i = c_t|\mathbf{x}_i, \mathcal{D}) \\ \forall c_t \neq c \tag{2.1}$$

Various probabilistic classification techniques have been used to process hyperspectral data [12]. In this chapter, we use the probabilistic SVM [27] and the MLR*sub* classifiers [51] for probability estimation. SVMs and MLR*sub* rely, respectively, on discriminant functions and posterior class distributions which have shown good performance in hyperspectral data classification, particularly in scenarios dominated by small training samples. In the following, we describe these probabilistic classifiers in more details.

2.3.1.1 Probabilistic SVM algorithm

The SVM classifier is typically defined as follows [27]:

$$f(\mathbf{x}_j) = \sum_i \alpha_i y_i \Phi(\mathbf{x}_i, \mathbf{x}_j) + b, \tag{2.2}$$

where $\mathbf{x}_j \in \mathbf{x}$, $\mathbf{x}_i \in \mathcal{D}$, b is the bias, and $\{\alpha_i\}_{i=1}^{n_{tr}}$ represents lagrange multipliers which are determined by the parameter C (that controls the amount of penalty during the SVM optimization). Here, $y_i \in \{-1, 1\}$ and $\Phi(\mathbf{x}_i, \mathbf{x}_j)$ is a function of the inputs, which can be linear or nonlinear. In SVM classification, kernel methods have shown great advantage in comparison with linear methods [32]. In this chapter, we use a Gaussian radial basis function kernel $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2)$, whose width is controlled by parameter γ. Although the original SVM does not provide class probability estimates, different techniques can be used to obtain class probability estimates based on combining all pairwise comparisons [29]. In this chapter, one of the probabilistic SVM methods [126] included in the popular LIBSVM library [30] is used.

2.3.1.2 MLR*sub* algorithm

MLR-based techniques are able to model the posterior class distributions in a Bayesian framework. In these approaches, the densities $p(y_i|\mathbf{x}_i)$ are modeled with the MLR, which corresponds to discriminative model of the discriminative-generative pair for $p(\mathbf{x}_i|y_i)$ Gaussian and $p(y_i)$ multinomial. The MLR model is formally given by [44]:

$$p(y_i = c|\mathbf{x}_i, \boldsymbol{\omega}) = \frac{\exp\left(\boldsymbol{\omega}^{(c)^T}\mathbf{h}(\mathbf{x}_i)\right)}{\sum_{l=1}^{k} \exp\left(\boldsymbol{\omega}^{(l)^T}\mathbf{h}(\mathbf{x}_i)\right)}, \tag{2.3}$$

where $\mathbf{h}(\mathbf{x}_i) \equiv [h_1(\mathbf{x}_i), \ldots, h_m(\mathbf{x}_i)]^T$ is a vector of m fixed functions of the input data, often termed as features; $\boldsymbol{\omega}^{(c)} \equiv [\omega_1^{(c)}, \ldots, \omega_m^{(c)}]^T$ is the set of logistic regressors for class c, and $\boldsymbol{\omega} \equiv [\boldsymbol{\omega}^{(1)^T}, \ldots, \boldsymbol{\omega}^{(k)^T}]^T$. Recently, Li *et al.* [51] have proposed to combine MLR with a subspace projection method called MLR*sub* to cope with two main issues: the presence of mixed pixels in hyperspectral data and the availability of limited training samples. The idea of applying subspace projection methods to improve classification relies on the basic assumption that the samples within each class can approximately lie in a lower dimensional subspace. Thus, each class may be represented by a subspace spanned by a set of

basis vectors, while the classification criterion for a new input sample would be the distance from the class subspace [51]. In the MLR*sub* formulation, the input function $\mathbf{h}(\mathbf{x}_i)$ is class dependent and is given by

$$\mathbf{h}^{(c)}(\mathbf{x}_i) = [\|\mathbf{x}_i\|^2, \|\mathbf{x}_i^\mathrm{T}\mathbf{U}^{(c)}\|^2]^\mathrm{T}, \tag{2.4}$$

where $\mathbf{U}^{(c)} = \{\mathbf{u}_1^{(c)}, \ldots, \mathbf{u}_{r^{(c)}}^{(c)}\}$ is a set of $r^{(c)}$-dimensional orthonormal basis vectors for the subspace associated with class c ($r^{(c)} \ll d$).

2.3.2 Characterization of spatial information

In this section we describe the mechanism used to include spatial-contextual information in the presented method. For this purpose, we use MRF, which is a widely used contextual model and a classical probabilistic method to model spatial correlation of pixel neighbors. This approach has been successfully applied in the context of remote sensing problems [101, 126, 42]. In the MRF framework, the MAP decision rule is typically formulated as the minimization of a suitable energy function [97]. Normally, the MRF-based approach can be implemented in two steps in hyperspectral image analysis. First, a probabilistic pixelwise classification method (such as those described in the previous section) is applied to learn the posterior probability distributions from the spectral information. Second, contextual information is included by means of an MRF regularization to refine the classification, as already outlined in Fig. 1.4.

According to the MAP-MRF framework, a pixel belonging to a class c is very likely to have neighboring pixels belonging to the same class. By using the Hammersly-Clifford theorem [127], we can compute the MAP estimate of $\mathbf{y}$ as follows:

$$\widehat{\mathbf{y}} = \arg\min_{\mathbf{y}} \left(\sum_{i \in \mathcal{S}} -\log p(y_i|\mathbf{x}_i) - \mu \sum_{i \sim j} \delta(y_i - y_j) \right), \tag{2.5}$$

where the term $p(\mathbf{y}_i|\mathbf{x}_i)$ is the spectral energy function from the observed data, which needs to be estimated by probabilistic methods. In this chapter, we use the probabilistic SVM and MLR*sub* to learn the probabilities. Parameter μ is tunable and controls the degree of smoothness, and $\delta(y)$ is the unit impulse function, where $\delta(0) = 1$ and $\delta(y) = 0$ for $y \neq 0$. Notice that the pairwise terms, $\delta(y_i - y_j)$, attach higher probability to equal neighboring labels than the other way around. Minimization of expression (2.5) is a combinatorial optimization problem involving unary and pairwise interaction terms. A good approximation can be obtained by mapping the problem into the computation of a series of min-cuts on a suitable graphs [110]. This aspect has been thoroughly explored in the context of hyperspectral image classification in previous contributions [43].

2.4 Proposed approach

In this section, we present the proposed spectral-spatial classification approach called SVM-MLR*sub*-MRF. The full methodology is summarized by a detailed flowchart in Fig. 2.1. As shown in Fig. 2.1, the proposed approach mainly contains four steps: 1) generation of the class combinations map; 2) calculation of the local and global probabilities; 3) decision fusion; and 4) MRF regularization. In the following, we present the details of each individual steps.
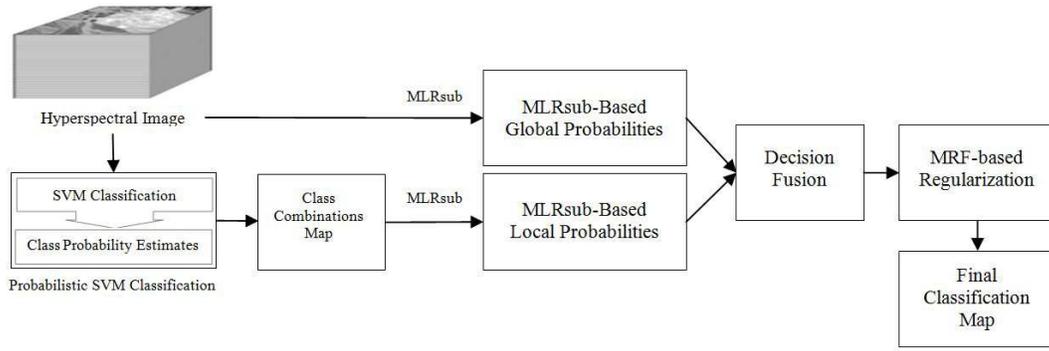
Figure 2.1: Flowchart of the proposed SVM-MLR*sub*-MRF method.

$$P_{SVM} = [0.3, 0.1, 0.6]$$
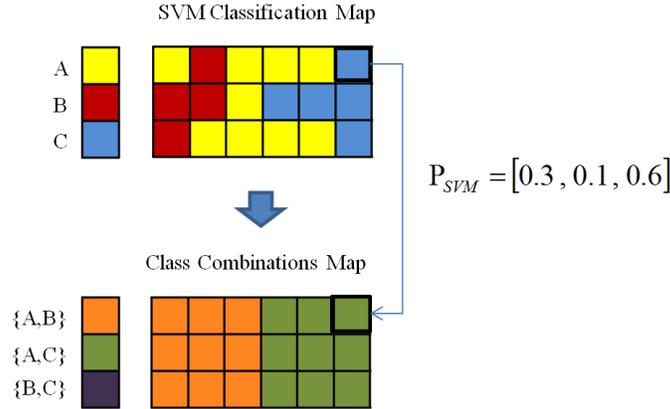
Figure 2.2: Example of the generation of a class combinations map, using threshold $M = 2$.

2.4.1 Generation of the Class Combination Map

The class combination map is generated from the probabilistic SVM classification results. Notice that the probabilistic SVM is only used as an indicator to determine the number of mixtures appearing in each pixel and does not contribute to the probability learning. For this purpose, a subset of the M most reliable class labels (mixed components) is chosen for each pixel as the possible class combinations for that pixel, and $M \leq k$. In case $M = k$, the local learning is equivalent to the global learning. It is also important to emphasize that, although in this work we use the probabilistic SVM for pixelwise classification due to its proved effectiveness in hyperspectral classification [32], other probabilistic classifiers could also be used as far as they are well suited to hyperspectral analysis. Furthermore, as a classifier, the probabilistic SVM has different characteristics in comparison with MLR*sub*, thus allowing for the possibility to use both classifiers in combined fashion in order to remove irrelevant class labels and to improve the efficiency of the class dependent subspace projection step in the MLR*sub* method, which will be described in the following section.

For illustrative purposes, Fig. 2.2 shows an example of how to generate a class combination map using the aforementioned strategy for a three class problem, where the classes are denoted as $\{A, B, C\}$ and the number of mixed components is set to $M = 2$. Using the probabilistic SVM, for each pixel we obtain a vector of three probabilities with respect to classes A, B, and C. As shown in Fig. 2.2, for the pixel at

the top-right corner of the image we assume that the probabilities are 0.3, 0.1, and 0.6 (for classes A, B, and C, respectively). Under these assumptions, the pixel would be assigned to the subset $\{A, C\}$ (from all possible combinations of the three classes). Notice that, in this example, there is no pixel assigned to the class combination $\{B, C\}$. Finally, it should be noted that the number of class combinations is given by $C(k, M)$, where, in this example, it is $C(3, 2) = 3$.

2.4.2 Calculation of the local and global probabilities

In this section, we describe the procedure used to calculate the local and global probabilities. Here, we use the MLR*sub* algorithm to learn the posterior probability distributions locally for the M classes selected in the previous step and globally for all classes. Let $\mathbf{p}_g$ and $\mathbf{p}_l$ denote the global and local posterior probabilities, respectively. For example, if we take the pixel used as an example in the previous section (i.e., the one located at the top-right corner in Fig. 2.2), in this case, compute the global and local probabilities as follows:

$$\mathbf{p}_g = \{p(y_i = c|\mathbf{x}_i, \boldsymbol{\omega}_g), c = A, B, C\}, \tag{2.6}$$

$$\mathbf{p}_l = \{p(y_i = A|\mathbf{x}_i, \boldsymbol{\omega}_l), 0, p(y_i = C|\mathbf{x}_i, \boldsymbol{\omega}_l)\}, \tag{2.7}$$

where (2.6) is the global learning step and $\boldsymbol{\omega}_g$ are the corresponding logistic regressors. On the other hand, (2.7) is the local learning step, and $\boldsymbol{\omega}_l$ represents the associated regressors. The global probability $\mathbf{p}_g$ in (2.6) is learned from the original data by the MLR*sub* algorithm [51]. Here, we exploit the good capabilities of MLR*sub* when dealing with mixtures. At this point, it is important to emphasize that we selected the MLR*sub* because, in real images, it is very likely that an observed pixel is a mixture of several components/classes. However, it is unlikely that the pixel is mixed by many components/classes [125]. Based on this observation, (2.7) uses MLR*sub* to locally learn the class posterior probabilities from the class combination map generated by the probabilistic SVM. Notice that, in the local estimation, only M classes are considered for each pixel, and we remove the remaining ones, *i.e.*, their probabilities are set to 0. For instance, in (2.7), $p(y_i = c|\mathbf{x}_i, \boldsymbol{\omega}_l) = 0$, which means that c is a removed class. In this way, by setting the probability of the irrelevant classes to zero, we remove those irrelevant classes from the combination set such that we eliminate the influence of the less relevant classes (or noise) in the local area. Therefore, considering the mixtures at a local (pixel) level is very important due to the following reasons. First, by eliminating the less relevant classes, the proposed approach locally eliminates noise, which greatly improves the separability of the features from noise, thus improving the performance of the MLR*sub* algorithm. Second, the class-dependent subspace reduces its dimensionality such that less training samples are required for learning.

2.4.3 Decision fusion

In this step, we combine the local and global probability distributions learned by the MLR*sub* algorithm to produce the final probabilities. For this purpose, we use the consensus theory [128] which follows a linear opinion pool [129]:

$$p(y_i = c|\mathbf{x}_i) = \sum_{j=1}^{N} \lambda_j p_j(y_i = c|\mathbf{x}_i), \tag{2.8}$$

where N is the number of data sources, $p_j(y_i = c|\mathbf{x}_i)$ is a source-specific posterior probability associated with data source j, and λ_j's are source-specific weights which control the relative influence of the data

sources, where $0 \leq \lambda_j \leq 1$ and $\sum_{j=1}^{N} \lambda_j = 1$. In this paper, we consider two data sources: global and local probability distributions. We simply combine these two data sources to compute the final class probability estimates as:

$$p(y_i = c|\mathbf{x}_i) \quad = \quad \lambda p_g(y_i = c|\mathbf{x}_i, \boldsymbol{\omega}_g) + (1 - \lambda)p_l(y_i = c|\mathbf{x}_i, \boldsymbol{\omega}_l), \quad (2.9)$$

where λ is a tunable parameter which controls the weights between the global and local probabilities and $0 \leq \lambda \leq 1$. It should be noted that, if $\lambda = 1$, only the global information is considered and the method remains as the original MLR*sub*. If $\lambda = 0$, only the local information is used. In our experimental results section, we will analyze the impact of parameter λ and discuss the relevance of global information and local information in the obtained classification results.

2.4.4   MRF-based spatial regularization

The last step of our proposed method consists of including the spatial-contextual information. As shown by Fig. 2.1, this stage is applied on the output of the decision fusion step. Although many strategies can be used for this purpose, we follow a commonly used strategy which relies on an MAP-MRF framework [51], as described in subsection 2.3.2. The strategy adopted in this work is similar to the one adopted by the SVM-MRF [101] or the MLR*sub*-MRF [43], which will be used for comparative purposes in the following section. Our strategy is also similar to the one used by the SVM-MRF in [35] but emphasizes that, in our comparisons, we do not exactly use the algorithm introduced in [35] but an implementation of SVM-MRF developed by ourselves and based on graph-cuts [110].

2.5   Experimental results

In this section, we use both synthetic and real hyperspectral data sets to evaluate the performance of the proposed SVM-MLR*sub*-MRF classification algorithm in different analysis scenarios. The main objective of the experimental validation with synthetic hyperspectral image is the assessment and characterization of the algorithm in a fully controlled environment, whereas the main objective of the experimental validation with real data sets is to compare the performance of the proposed method with other state-of-the-art methods in the literature. The remainder of this section is organized as follows. First, we describe the hyperspectral data sets (synthetic and real) used in experiments. Then, we describe the experimental setting. Next, we describe several experiments intended to address several important aspects of the presented method, such as the impact of parameters λ and M, a comparison with other standard methods, and an evaluation of the performance of the method in the presence of training sets with different numbers of samples.

2.5.1   Hyperspectral data sets

2.5.1.1   Synthetic data

A synthetic image has been generated with size of $n = 80 \times 120$, and the class distribution displayed in Fig. 2.3(a). The synthetic scene comprises eight classes which contain linear mixtures of a set of spectral signatures randomly selected from a digital spectral library compiled by the USGS and available online[2]. The USGS library contains spectral plots for nearly 500 materials (mostly minerals) in the $400 - 2500$ nm spectral range, where the bands have been convolved to the number of bands available for the AVIRIS [2]

---

[2]http://speclab.cr.usgs.gov/spectral-lib.html

(a)

(b)

class1　　　　　class2　　　　　class3　　　　　class4

class5　　　　　class6　　　　　class7　　　　　class8

(c)
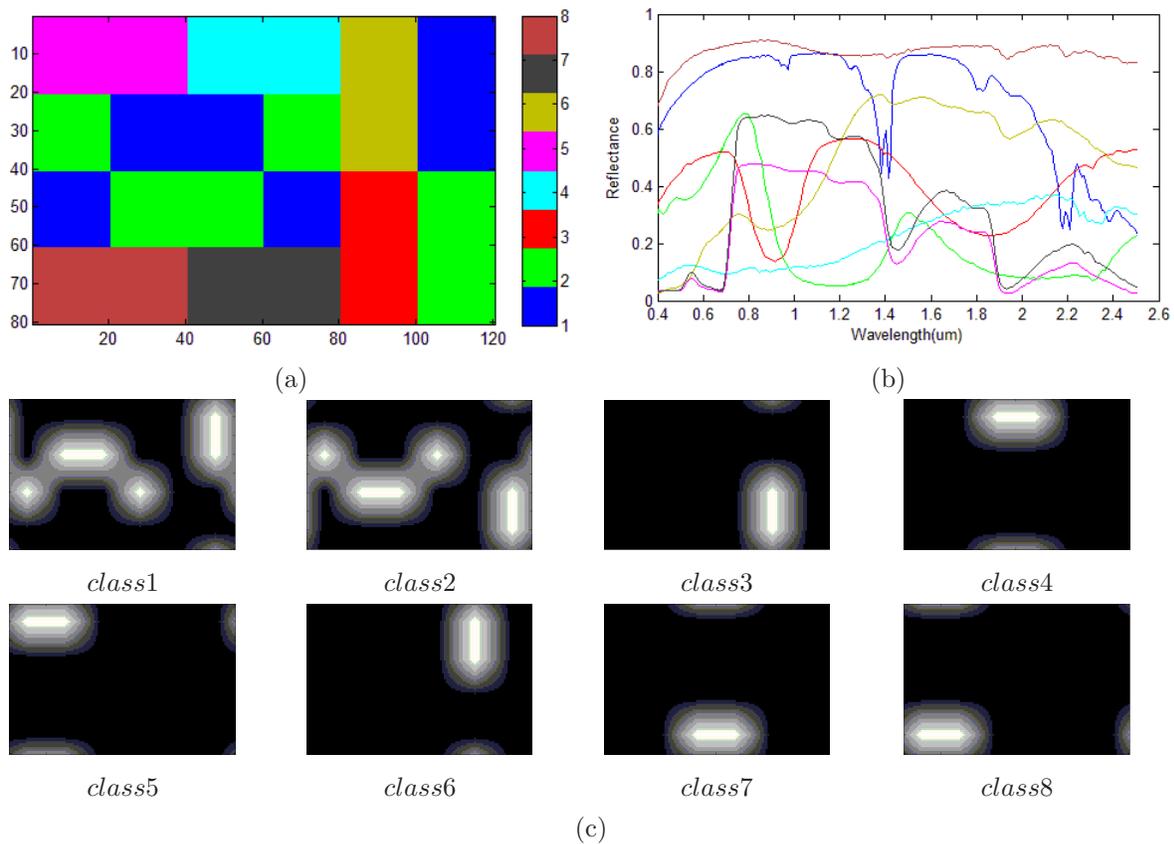
Figure 2.3: (a) Classes in a synthetic scene with $n = 80 \times 120$. (b) Spectral signatures of randomly selected materials from the USGS digital library used in the simulation (c) Fractional abundance distributions considered for generating mixed pixels using Gaussian filter of size $k = 25$ and standard deviation $\sigma = 30$.

that comprises 224 spectral bands. Fig. 2.3(b) shows the spectral signatures of eight randomly selected mineral signatures allocated to the main classes displayed in Fig. 2.3(a).

In order to simulate mixed pixels using linear mixtures, the eight class abundance maps are filtered by a $k \times k$ symmetric Gaussian lowpass filter with a fixed standard deviation σ. For illustrative purposes, Fig. 2.3(c) shows the abundance maps associated to the eight classes of the synthetic scene after applying a Gaussian filter of size $k = 25$ and standard deviation $\sigma = 20$. In each pixel of the scene, the fractional abundances vary from 0% (black color) to 100% (white color) and sum to unity. Using this procedure, signature abundance is not constant over class regions, and the pixels closer to the borders of the regions are more heavily mixed, as expected in real scenarios. Finally, zero-mean Gaussian noise is added to each band of the synthetic hyperspectral image so that the signal-to-noise ratio is equal to 20 dB according to the definition given in [130].

2.5.1.2　Real data

Two real hyperspectral data sets are used to evaluate the proposed approach. The first one is the well-known AVIRIS Indian Pines scene [see Fig. 2.4(a)], collected over Northwestern Indiana in June 1992 [11]. The scene is available online[3] and contains 145×145 pixels and 220 spectral bands between 0.4 and

---

[3]https://engineering.purdue.edu/~biehl/MultiSpec/

(a) False color composition	(b) Ground-truth
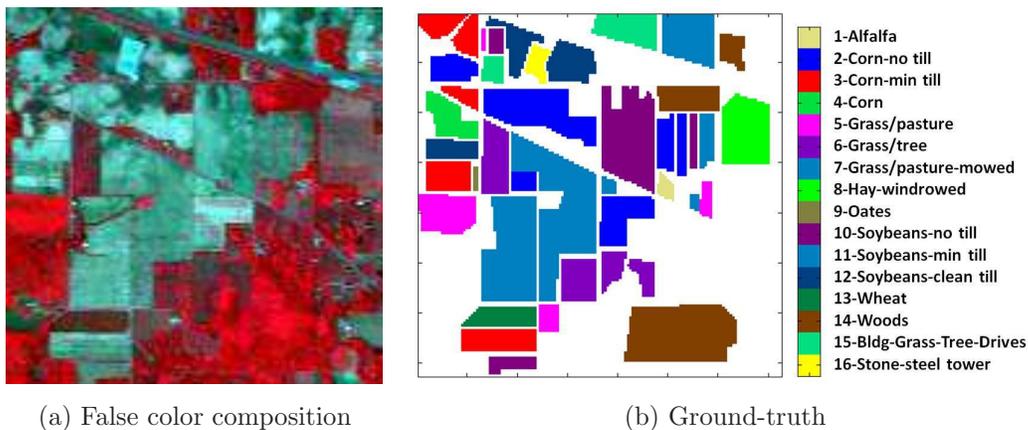
Figure 2.4: AVIRIS Indian Pines dataset.

2.5 micrometers. A total of 20 spectral bands was removed prior to experiments due to noise and water absorption in those channels. The ground-truth image displayed in Fig. 2.4(b) contains 10366 samples and 16 mutually exclusive classes having 20-2468 samples. This data is widely used as a benchmark for testing the accuracy of hyperspectral data classification algorithms, mainly because it constitutes a challenging classification problem due to the presence of mixed pixels in available classes and also because of the unbalanced number of available labeled pixels per class.

We have also used a scene collected by the ROSIS for evaluation purposes. This data was acquired over the urban area of the University of Pavia, Pavia, Italy. The flight was operated by the DLR in the framework of the HySens project, managed and sponsored by the European Commission. The image size in pixels is 610×340, with very high spatial resolution of 1.3 meters per pixel. The number of data channels in the acquired image is 103 (with spectral range from 0.43 to 0.86 micrometers). Fig. 2.5(a) shows a false color composite of the image, while Fig. 2.5(c) shows nine ground-truth classes of interest, which comprise urban features, as well as soil and vegetation features. In the original data set, out of the available ground-truth pixels, 3921 were used for training, and 42776 samples were used for testing [see Fig. 2.5(b)].

2.5.2 Experimental setting

Before describing our results, it is first important to discuss some considerations that define our experimental setting. For the synthetic image experiments, we considered two strategies. In our first strategy, we simulated some pure pixels for each class (using $k = 20$ and $\sigma = 30$ in the synthetic data simulation). In the second strategy, we increased the size of the filter to $k = 25$ (with the same $\sigma = 30$) so that all of the simulated pixels inside a class region were mixed with abundance fractions less than 80%, and the simulated image did not contain any pure pixels. In both cases, training samples were extracted from the purest available classes. All of the results reported in this chapter with synthetic data sets were obtained after 50 Monte Carlo runs in which we randomly select 8 different materials and also randomly select different training sets.

Concerning our real data experiments, the experimental setting can be briefly summarized as follows. For the experiments with the AVIRIS Indian Pines data, the training samples were randomly selected from the available ground truth, and the remaining samples are used for validation. For the smallest classes of this data set, if the total number of available labeled samples per class in the ground truth is

(a) False color composition   (b) Training data   (c) Ground-truth
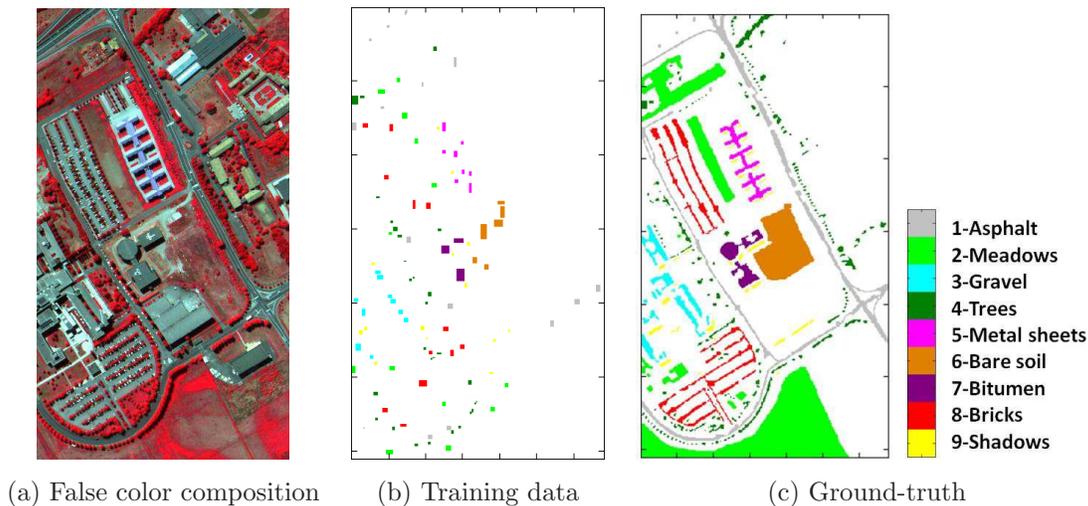
Figure 2.5: ROSIS Pavia University dataset.

smaller than the given number of training samples, we take half of the total samples for training. For the ROSIS Pavia University dataset, the training sets are composed of subsets of the original training samples, and the remaining test sets are used for validation. All of the results reported in this paper with the two considered real data sets were obtained after 30 Monte Carlo runs.

The classifiers compared in the presented study are the proposed SVM-MLRsub-MRF in addition to the standard SVM, MLRsub, and SVM-MRF. In all experiments for the MLRsub and MLRsub-MRF algorithms, we optimized the parameters as indicated in [51]. Concerning the probabilistic SVM classifier, we optimized the related parameters using tenfold cross validation. Finally, for the proposed approach, we use SVM-MLRsub and SVM-MLRsub-MRF to denote the algorithms with and without the MRF spatial regularizer, respectively.

2.5.3   Quantitative and comparative assessment

In this section, we conduct an experimental assessment of the presented approach using the simulated and real data sets described in Section 2.5.1 and bearing in mind the experimental setting described in Section 2.5.2. The experiments reported in this section can be summarized as follows. In a first experiment, we perform an assessment of the impact of parameter λ, which controls the degree of global information and local information used by the presented method. In a second experiment, we evaluate the impact of parameter M, which controls the number of class combinations for local estimation. In a third experiment, we compare the proposed method with other state-of-the-art methods. Finally, in a fourth experiment, we analyze the sensitivity of the considered method to different training sets and show the good performance of the proposed approach in the presence of limited training samples.

2.5.3.1   Experiment 1. Impact of parameter λ

In this experiment, we perform an analysis of the impact of parameter λ for the SVM-MLRsub-MRF algorithm by using the considered (synthetic and real) data sets with $M = 2$ this means that we set the number of class combinations to 2. The reason for this selection is to keep the number of mixtures per pixel low, as it is often the case in real scenarios. Table 2.1 shows the OA and AA (as a function of parameter λ) for the different scenes considered. For illustrative purposes, Fig. 2.6 provides a detailed

(a) AVIRIS Indian Pines (b) Pavia University
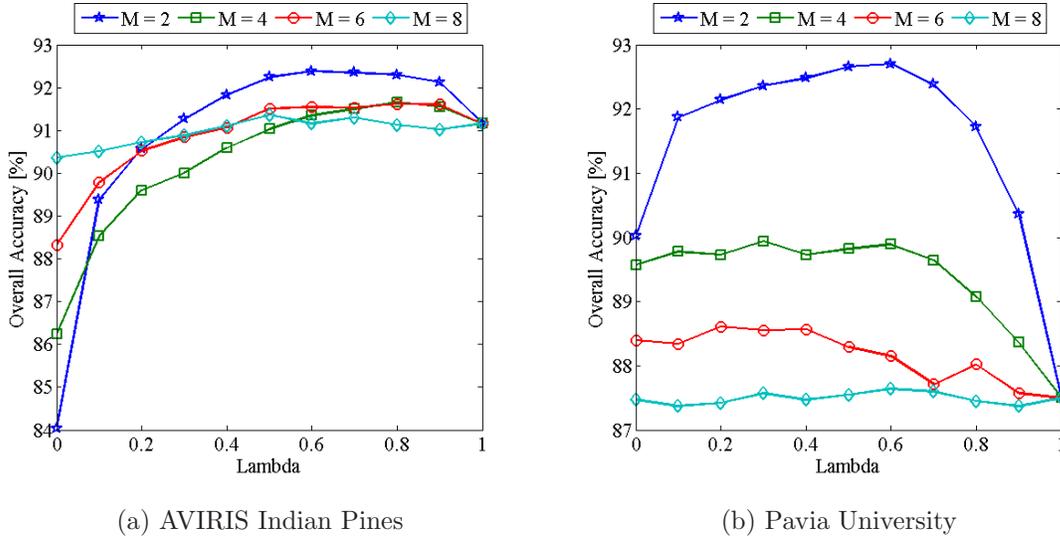
Figure 2.6: Classification results obtained by the proposed method after using different values of parameters λ and M for the (a) AVIRIS Indian Pines and (b) ROSIS Pavia University scenes.

analysis of the classification results obtained by the proposed method after using different values of parameters λ and M for the two considered hyperspectral scenes. In all experiments, we choose 50 random samples per class for training so that, in total, we have 400, 697, and 450 training samples for the synthetic data, AVIRIS Indian pines, and ROSIS Pavia University data sets, respectively (it should be noted that, for the AVIRIS Indian Pines scene, we may select a different number of samples for the small classes).

Several conclusions can be obtained from Table 2.1 and Fig. 2.6. First and foremost, it is remarkable that the proposed approach, which integrates the global information and local information, obtained the best performance in comparison with those results obtained only from the local ($\lambda = 0$) or global ($\lambda = 1$) information. While $\lambda \in [0.4, 0.6]$, the proposed approach obtained very good results for all considered data sets (hence, a reasonable setting is to assign equal weight to local information and global information, i.e., $\lambda = 0.5$). For other suboptimal values of λ, the obtained results are still better or comparable to those obtained by using the local or global information alone.

Furthermore, Fig. 2.6(a) reveals that the results obtained for the AVIRIS Indian Pines using the global information only ($\lambda = 1$) are better than those obtained using the local information alone ($\lambda = 0$). In turn, Fig. 2.6(b) reveals an opposite behavior for the ROSIS Pavia University data. However, it is clear from Fig. 7 that an intermediate value of λ(which is equivalent to considering both local and global probabilities) leads to good classification results in the two considered cases, particularly when the value of M is low (this is expected, since the number of mixtures in a given pixel is generally low). It can also be seen in Fig. 2.6 that the value of M is more relevant for the ROSIS Pavia University data than for the AVIRIS Indian Pines data, which is related with the different spatial resolutions of the considered scenes. From this experiment, we conclude that, in the considered case studies, low values of M and values of λ that ensure a good balance between local information and global information lead to good classification results.

Table 2.1: Overall (OA) and average (AA) classification accuracies [%] (as a function of parameter λ) obtained by the SVM-MLR*sub*-MRF method for the synthetic and real data sets considered in experiments. The best results are outlined in bold typeface.

Data	Accuracies	λ										
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Synthetic	OA	70.05	70.40	70.63	71.34	72.60	73.61	**73.99**	73.26	72.56	71.48	70.55
(pure training samples)	AA	68.15	68.47	68.74	69.48	71.15	72.84	**73.93**	73.45	73.13	72.19	71.33
Synthetic	OA	75.13	75.66	76.12	76.68	77.82	79.27	**79.40**	78.63	77.36	76.16	75.08
(mixed training samples)	AA	74.54	75.06	75.49	76.13	77.48	79.13	**79.75**	79.13	77.99	76.92	75.85
AVIRIS Indian Pines	OA	84.03	89.39	90.56	91.26	91.83	92.24	**92.38**	92.35	92.29	92.12	91.16
	AA	89.00	92.97	93.93	94.49	94.91	95.08	**95.21**	95.13	95.02	94.81	94.01
Pavia University	OA	90.02	91.78	92.14	92.36	92.48	92.66	**92.70**	92.38	91.72	90.36	87.50
	AA	90.65	91.53	91.62	91.68	91.79	**91.85**	91.78	91.44	90.87	90.18	88.23

Table 2.2: Overall (OA) and average (AA) classification accuracies [%] (as a function of parameter M, with fixed $\lambda = 0.5$) obtained by the SVM-MLRsub-MRF method for the synthetic and real data sets considered in experiments. The best results are outlined in bold typeface.

Data	Accuracies	M						
		2	3	4	5	6	7	8
Synthetic	OA	73.61	**74.26**	73.44	72.88	72.16	71.28	70.50
(pure training samples)	AA	72.84	**74.07**	73.43	73.30	72.76	71.89	71.19
Synthetic	OA	79.27	**79.73**	79.32	78.54	77.49	76.08	75.30
(mixed training samples)	AA	79.13	**79.95**	79.46	78.80	77.84	76.55	75.81
AVIRIS Indian Pines	OA	**92.24**	91.60	91.05	90.90	91.47	91.58	91.17
	AA	**95.08**	94.73	94.42	94.06	94.18	94.15	94.06
Pavia University	OA	**92.68**	90.64	89.82	88.93	88.19	87.89	87.58
	AA	**91.93**	90.17	89.45	88.42	88.29	88.11	88.03

2.5.3.2 Experiment 2. Impact of parameter M

In this experiment, we perform an evaluation of the impact of parameter M (controlling the number of class combinations) on the presented approach. Table 2.2 shows the classification results obtained by the proposed approach using different values of parameter M for all of the considered data sets. For the real data, the best results are obtained with $M = 2$, which means that most of the pixels are either pure or made up of two mixing components. This is a reasonable assumption since, in reality, most pixels are made up by a mixture of a limited number of materials, especially for images with high spatial resolution. However, in our synthetic image experiments, the mixed pixels were simulated in a way that pixels mixed by a higher number of materials are present in the scene. As a result, in the synthetic image experiments, better results can be obtained by using higher values of M.

2.5.3.3 Experiment 3. Comparison with other methods

Table 2.3 shows a comparison of the classification results obtained by the proposed approach (using different values of parameter M) with regard to those provided by other methods for the synthetic image data. Similarly, Tables 2.4 and 2.5 show a comparison of the presented approach to other methods using the AVIRIS Indian Pines and ROSIS Pavia University data sets, respectively. In all cases, different values of parameter M were considered for the proposed method, and parameter λ was set to 0.5.

Several conclusions can be obtained from the experiments reported in Tables 2.3, 2.4 and 2.5. First and foremost, it is noticeable that the MLRsub-MRF and SVM-MLRsub-MRF, which include spatial information and also perform mixed pixel characterization, outperform the SVM-MRF which does not characterize mixed pixels. For instance, Table 2.5 reveals that the proposed SVM-MLRsub-MRF approach obtained an OA of 92.68%, which contrasts with the OA of 83.96% achieved by SVM-MRF in the ROSIS Pavia University experiments. Similarly, Table 2.3 also reveals an OA of about 9.5% larger than that obtained by the SVM-MRF algorithm. However, the situation is different for the experiments with the AVIRIS Indian Pines reported in Table 2.4. Specifically, the MLRsub-MRF did not show a significant improvement with regard to the SVM-MRF, and consequently, the results obtained by the SVM-MLRsub-MRF method are not significant. The main reason for this behavior is the difference in the reference data for different classes in the AVIRIS Indian Pines scene, which varies from 20 to 2468 pixels

per class. For these data, we chose very limited training sets, and the samples are not evenly distributed among classes. For example, if we take one of the large classes such as *Soybeans-min till* (which contains 2418 samples), we only consider 50 samples for training, which is less than 0.02% of this class. This number of training samples is not enough to completely characterize the class, as it is also the case for other similar classes such as *Soybeans-no till* and *Soybeans-clean till*. Another problem observed in the classification of this image is the fact that class *Bldg-Grass-Tree-Drives* is a highly mixed class. Hence, it is not efficient to use the training samples from this class in the subspace projection procedure and, and consequently, the classification accuracy for the methods MLR*sub*-MRF and SVM-MLR*sub*-MRF are not significantly increased with regard to other methods as it was the case in the experiments with other data sets.

If we focus on the results reported in Table 2.3 in a fully controlled environment, we can conclude that the class probabilities estimated by the SVM classifier may not be fully reliable to be used in the MRF regularization procedure. This is because of the nature of the SVM method, which is a hard classifier. In turn, the MLR*sub*-MRF method better characterized noise and mixed pixels. However, the SVM-MLR*sub*-MRF method provided the highest classification accuracies in this experiment. For instance, in the synthetic experiment using pure training samples, the OA achieved by the presented method improved by 3.77% and 9.23% of the OA achieved by the MLR*sub*-MRF and SVM-MRF, respectively. When mixed training samples were used, the proposed SVM-MLR*sub*-MRF algorithm obtained an OA of 79.73%, which is the best result for this data set (4.40% and 10.05% higher than MLR*sub*-MRF and SVM-MRF, respectively). This is because mixed training samples are near the decision boundaries and can be very effective in class discrimination.

If we now focus on Table 2.5, we can observe that, in this experiment, the pixel-wise SVM classifier already provides high classification accuracies. However, including the spatial-contextual information significantly improves the classification accuracies as it can be particularly observed in the SVM-MLR*sub*-MRF method. Here, by using only 50 training samples per class (in total 450 samples, which is a very low number for this scene), the proposed algorithm obtained an OA of 92.68%, which is 5.18% and 8.72% higher than MLR*sub*-MRF and SVM-MRF, respectively.

In order to analyze the statistical significance of the results obtained by the different compared methods, we have used McNemar's test [131]. In this test, a value of $|Z| > 1.96$ indicates that there is a significant difference in accuracy between two classification methods. The sign of Z is also a criterion to indicate whether a first classifier is more accurate than a second one ($Z > 0$) or vice versa ($Z < 0$). Table 2.6 provides the results obtained for all of the considered (synthetic and real) data sets. As it can be seen from Table 2.6, the differences in classification accuracies between our proposed method (implemented with parameters $M = 2$ and $\lambda = 0.5$) and the MLRsub-MRF method are statistically significant. Compared to the SVMMRF method, SVM-MLRsub-MRF exhibits a statistically significant improvement in classification accuracies except for the AVIRIS Indian Pines scene, in which the McNemar's test indicates that the performances of the two methods are similar.

For illustrative purpose, Figs. 2.7 and 2.8 show some of the obtained classification maps for the AVIRIS Indian Pines and ROSIS Pavia University data sets, respectively. Each of the maps corresponds to one out of the 30 Monte Carlo experiments which were averaged to produce the results, respectively, reported in Tables 2.4 and 2.5. As shown in Figs. 2.7 and 2.8, good classification results with adequate class delineation and spatial consistency can be observed for the presented method (which was run in both cases using $M = 2$ and $\lambda = 0.5$) in comparison to other approaches.

Table 2.3: Overall (OA), average (AA), individual class accuracies [%] and kappa statistic (κ) obtained by different methods with the synthetic image data set, using a total of 450 training samples (50 per class).

Data	Accuracies	SVM	MLR*sub*	SVM-MLR*sub*			SVM-MRF	MLR*sub*-MRF	SVM-MLR*sub*-MRF		
				$M=2$	$M=3$	$M=4$			$M=2$	$M=3$	$M=4$
Synthetic	Overall	65.23	68.49	69.77	71.33	71.09	65.03	70.49	73.61	**74.26**	73.44
(pure	Average	66.32	68.97	68.44	70.54	70.65	62.74	71.19	72.84	**74.07**	73.43
training samples)	κ	59.38	62.89	64.09	66.02	65.78	58.23	65.21	68.64	**69.37**	68.57
Synthetic	Overall	69.33	72.18	74.68	75.95	75.70	69.68	75.33	79.27	**79.73**	79.32
(mixed	Average	69.84	72.58	74.08	75.52	75.57	68.68	75.83	79.13	**79.95**	79.46
training samples)	κ	63.88	67.04	69.86	71.35	71.07	63.78	70.72	75.31	**75.90**	75.34

Table 2.4: Overall (OA), average (AA), individual class accuracies and kappa statistic (κ) obtained by different methods with the AVIRIS Indian Pines data set, using a total of 697 training samples (50 per class, except for very small classes).

Class	Samples		SVM	MLR*sub*	SVM-MLR*sub*			SVM-MRF	MLR*sub*-MRF	SVM-MLR*sub*-MRF		
	Train	Test			$M=2$	$M=3$	$M=4$			$M=2$	$M=3$	$M=4$
Alfalfa	27	27	86.42	75.31	79.01	78.64	74.69	93.39	**98.10**	96.30	97.78	97.78
Corn-no till	50	1384	66.45	59.08	68.18	66.01	63.28	79.52	85.80	88.07	**89.46**	88.93
Corn-min till	50	784	68.76	61.27	67.73	61.55	58.78	87.05	89.75	**90.58**	89.13	88.68
Corn	50	184	84.13	79.96	84.20	84.40	83.70	**99.26**	98.88	98.99	98.82	98.89
Grass/trees	50	447	91.35	85.15	88.40	88.86	86.60	**96.73**	93.26	94.50	94.53	93.96
Grass/pasture	50	697	90.82	94.13	96.46	96.39	95.97	98.77	99.18	99.24	**99.27**	99.21
Grass/pasture-mowed	13	13	87.95	65.38	85.13	81.03	68.72	95.90	**98.35**	96.92	96.92	97.44
Hay-windrowed	50	439	95.87	98.73	98.37	98.07	98.56	99.34	99.48	99.45	99.46	**99.51**
Oats	10	10	85.33	83.00	90.33	88.67	87.67	94.52	92.38	**100**	97.00	95.33
Soybeans-notill	50	918	76.32	65.53	70.50	66.80	65.69	91.18	**94.21**	93.97	93.81	93.69
Soybeans-min till	50	2418	62.31	49.41	59.24	52.96	49.14	**90.23**	84.54	86.68	83.67	82.07
Soybeans-clean till	50	564	74.98	77.16	80.67	83.92	84.12	95.22	94.51	97.38	**97.42**	97.33
Wheat	50	162	97.59	99.63	98.93	99.36	99.51	98.90	99.86	**99.90**	**99.90**	99.84
Woods	50	1244	87.08	94.48	94.47	94.67	94.81	88.27	**99.52**	98.04	98.91	99.12
Bldg-Grass-Tree-Drives	50	330	70.87	41.45	53.84	49.10	46.64	**99.70**	77.85	81.78	79.99	79.42
Stone-Steel towers	47	48	98.26	93.89	89.17	92.36	93.54	**99.86**	99.03	99.17	99.58	99.51
OA			75.21	69.32	75.09	72.43	70.56	90.60	91.14	**92.24**	91.60	91.05
AA			82.78	76.47	81.54	80.17	78.21	94.24	94.04	**95.08**	94.73	94.42
κ			71.94	65.34	71.73	68.82	66.78	89.28	89.91	**91.12**	90.42	89.81

Table 2.5: Overall (OA), average (AA), individual class accuracies [%] and kappa statistic (κ) obtained by different methods with the AVIRIS Indian Pines data set, using a total of 450 training samples (50 per class).

Class	Samples		SVM	MLRsub	SVM-MLRsub			SVM-MRF	MLRsub-MRF	SVM-MLRsub-MRF		
	Train	Test			$M=2$	$M=3$	$M=4$			$M=2$	$M=3$	$M=4$
Asphalt	50	6631	73.67	43.99	65.73	59.55	54.52	**95.05**	80.13	91.05	89.79	85.62
Meadows	50	18649	63.79	66.31	75.95	72.55	70.90	69.63	89.71	**94.26**	93.00	93.16
Gravel	50	2099	71.34	63.80	67.07	63.70	63.55	72.61	71.56	**73.66**	68.33	68.42
Trees	50	3064	96.55	79.96	89.70	86.98	84.44	**97.93**	80.45	93.83	92.24	88.82
Meta sheets	50	1345	99.35	99.08	98.88	99.19	99.24	99.87	99.70	99.58	99.77	**99.78**
Bare soil	50	5029	87.06	63.96	65.97	61.81	60.83	**98.19**	91.09	91.55	84.34	84.25
Bitumen	50	1330	91.62	86.08	83.91	83.74	84.24	**95.86**	90.43	86.63	89.06	90.44
Bricks	50	3682	85.21	62.64	74.17	69.57	67.19	**97.92**	91.03	96.90	95.12	94.70
Shadows	50	947	98.90	98.52	99.10	98.61	98.50	99.13	99.72	**99.86**	99.85	99.83
OA			75.38	65.47	75.07	71.38	69.39	83.96	87.50	**92.68**	90.64	89.82
AA			85.28	73.81	80.05	77.30	75.94	91.80	88.20	**91.93**	90.17	89.45
κ			69.49	56.74	68.06	63.59	61.18	79.96	83.72	**90.40**	87.71	86.62

Table 2.6: Statistical significance of differences in classification accuracies for all data sets (with parameters $M = 2$ and $\lambda = 0.5$)

Data	Z (classification method 1/classification method 2)			
	SVM-MLR*sub*/SVM	SVM-MLR*sub*/MLR*sub*	SVM-MLR*sub*-MRF/SVM-MRF	SVM-MLR*sub*-MRF/MLR*sub*-MRF
Synthetic (Pure T.S.)	9.87	4.44	17.33	8.58
Synthetic (Mixed T.S.)	11.58	7.33	19.33	11.19
AVIRIS Indian Pines	-3.77	18.98	1.07	16.46
Pavia University	-1.19	48.72	44.10	33.69

SVM (75.22%) MLR*sub* (72.93%) SVM-MLR*sub* (76.65%)

SVM-MRF (89.75%) MLR*sub*-MRF (92.70%) SVM-MLR*sub*-MRF (93.61%)
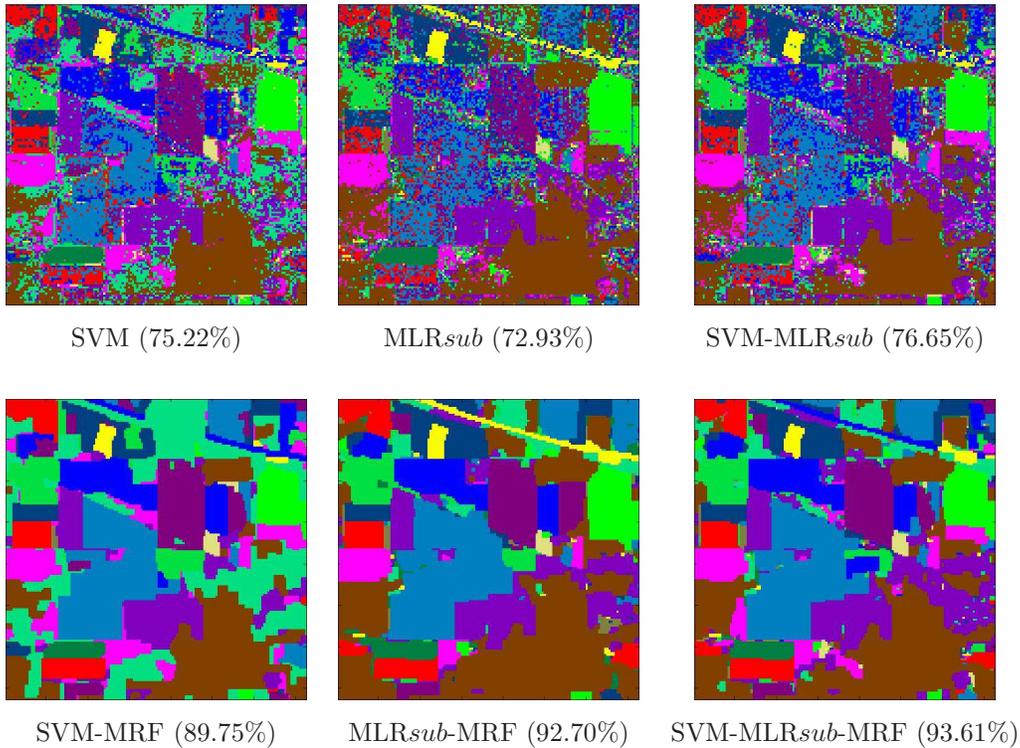
Figure 2.7: Classification results and overall classification accuracies (in the parentheses) obtained by different methods for the AVIRIS Indian Pines data set.

2.5.3.4 Experiment 4. Impact of the number of training samples

In this experiment, we first conduct an evaluation of the impact of the number of training samples on the proposed approach, using the two real data sets and fixing parameters $M = 2$ and $\lambda = 0.5$. Table 2.7 shows the classification accuracies obtained by the presented method as a function of the number of training samples per class (where the total number of training samples is given in the parentheses). In the case of the ROSIS Pavia University scene, we also performed an experiment using all of the available training samples (3921) in order to establish an upper bound to the presented results and to explore if, with a more limited training set, the results are indeed close to that bound. The results reported in Table 2.7 show that, for the AVIRIS Indian Pines data, we only need 40 training samples per class (for a total of 570) in order to achieve an OA of around 90.00% (and AA larger than 93%). For the ROSIS Pavia University data, we only need less than 40 samples per class in order to obtain an OA of 90%. This is remarkable, as sometimes it is very difficult to collect large training sets in practice. Table 2.7 also reveals that the presented method provided results which are comparable to those provided by the SVM-MRF and superior than those provided by MLR*sub*-MRF with the AVIRIS Indian Pines data. For the ROSIS Pavia University data, the proposed approach obtained an OA of 94.57% using only 80 training samples per class (for a total of 720 samples). This result is quite close to the upper bound result, obtained using all available (3921) training samples for this scene. In fact, this result is 4.32% higher than the results provided by MLR*sub*-MRF and 9.89% higher than the results provided by the SVM-MRF. This leads to two main observations. First, by including the local information, the SVM-MLR*sub*-MRF greatly improved the performance obtained by the MLR*sub*-MRF algorithm which

Figure 2.8: Classification results and overall classification accuracies (in the parentheses) obtained by different methods for the ROSIS Pavia University data set.

only considers the global information. A second observation is that the methods that characterize mixed pixels, i.e. MLR*sub*-MRF and SVM-MLR*sub*-MRF, can outperform the methods that do not incorporate mixed pixel characterization, even for scenes collected at high spatial resolution.

To conclude this section, Fig. 2.9 compares the performances of the methods: SVM-MRF, MLR*sub*-MRF, and SVM-MLR*sub*-MRF in 30 Monte Carlo runs conducted for different random sets of training sample sets for the two real hyperspectral scenes (AVIRIS Indian Pines, at the top of the figure, and ROSIS Pavia University, at the bottom of the figure). In order to establish a fair comparison, in each iteration, the same training set is used by all three methods. As Fig. 2.9 shows, when compared with MLR*sub*-MRF, the proposed method shows more uniform results and appears less sensitive to the quality of training samples. When compared with SVM-MRF, the proposed method shows slightly superior results for the AVIRIS Indian Pines scene and consistently better performance for the Pavia University scene. Again, we reiterate that the SVM-MLR*sub* method takes the advantages of both SVM and MLR*sub* and can compensate the situation in which one of the methods does not provide

Table 2.7: Overall (OA) and average (AA) accuracy [%] as a function of the number of training samples per class for the SVM-MLR*sub*-MRF method, where the total number of training samples are given in the parentheses.

Methods	Accuracies	AVIRIS Indian Pines				Pavia University				
		20 (303)	40 (570)	60 (817)	80 (1057)	20 (180)	40 (360)	60 (540)	80 (720)	All (3921)
SVM	OA	65.67	72.48	76.69	78.85	70.02	73.70	75.08	77.21	81.13
	AA	75.98	81.01	83.26	84.66	80.50	84.24	85.65	86.71	89.05
MLR*sub*	OA	65.36	68.16	69.72	70.07	64.84	64.89	64.70	65.10	70.61
	AA	75.10	76.10	75.65	75.10	72.72	74.08	73.97	74.16	73.92
SVM-MLR*sub*	OA	68.29	73.39	76.38	78.00	71.35	73.96	75.29	76.69	82.61
	AA	76.44	80.42	82.15	82.43	77.50	80.03	80.92	81.61	83.80
SVM-MRF	OA	82.10	88.65	91.17	92.01	79.56	82.93	82.54	84.68	86.17
	AA	89.54	93.52	94.26	94.36	86.13	91.12	91.68	92.36	92.56
MLR*sub*-MRF	OA	77.47	87.80	91.41	92.83	83.50	87.18	89.08	90.25	93.10
	AA	85.84	92.02	94.23	95.31	84.73	88.61	89.75	89.44	86.67
SVM-MLR*sub*-MRF	OA	82.70	89.79	93.10	93.86	85.88	90.97	92.10	94.57	95.56
	AA	88.89	93.40	95.52	96.14	87.06	91.17	92.70	93.06	90.69
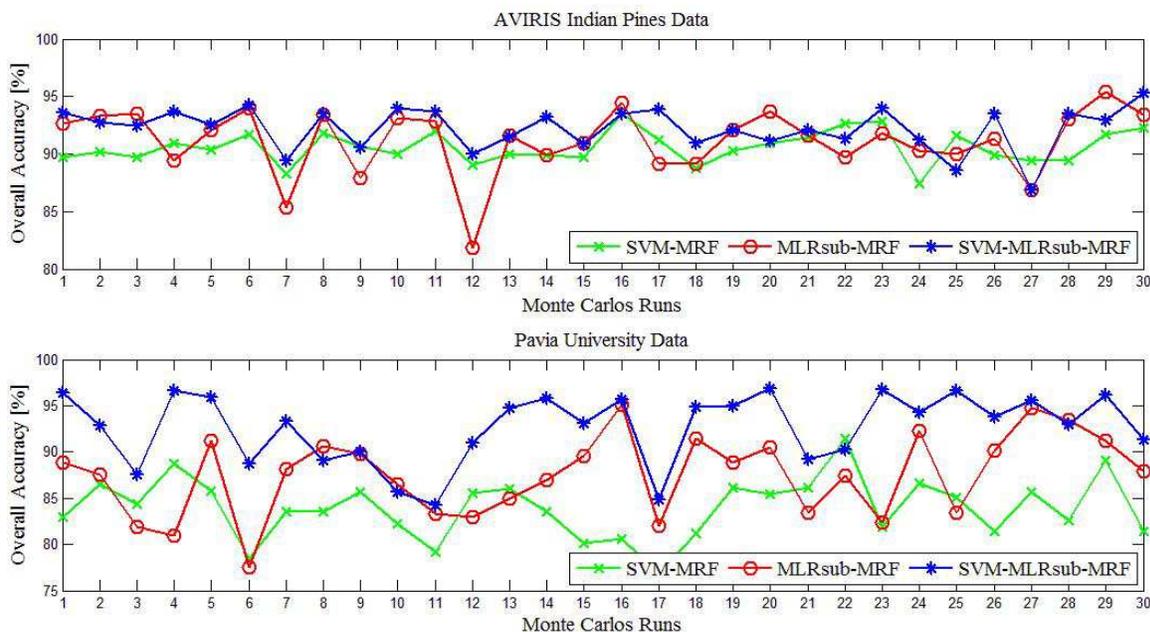
Figure 2.9: Comparison of the performance of the methods: SVM-MRF, MLRsub-MRF and SVM-MLRsub-MRF in 30 Monte Carlo runs conducted for different random sets of training samples sets for the two real hyperspectral scenes: AVIRIS Indian Pines (top), and ROSIS Pavia University (bottom). In each run the same training set is used by all three methods.

good performance by taking advantage of the other method. This is also the reason why the proposed SVM-MLRsub-MRF method can provide good performance in those cases in which none of the methods SVM-MRF and MLRsub-MRF exhibits good classification accuracies. This is the case, for instance, in iterations 6, 13, and 25 for the ROSIS Pavia University experiments reported in Fig. 2.9.

2.6 Summary and future directions

In this chapter, we introduce a novel spectral-spatial classifier for hyperspectral image data. The proposed method is based on the consideration of both global posterior probability distributions and local probabilities which result from the whole image and a set of previously derived class combination maps, respectively. The proposed approach, which intends to characterize mixed pixels in the scene and assumes that these pixels are normally mixed by only a few components, provides some distinguishing features with regard to other existing approaches. At the local learning level, the presented method removes the impact of irrelevant classes by means of a pre-processing stage (implemented using the probabilistic SVM) intended to produce a subset of M most probable classes for each pixel. This stage locally eliminates noise and enhances the impact of the most relevant classes. These aspects, together with the joint characterization of mixed pixels and spatial-contextual information, make our method unique and representative of a framework that, for the first time in the literature, integrates local and global probabilities in the analysis of hyperspectral data in order to constrain the number of mixing components used in the characterization of mixed pixels. This is consistent with the observation that, despite the presence of mixed pixels in real hyperspectral scenes, it is reasonable to assume that the mixing components in a given pixel are limited. Our experimental results, conducted using both synthetic

and real hyperspectral scenes widely used in the hyperspectral classification community, indicate that the proposed approach leads to state-of-the-art performance when compared with other approaches, particularly in scenarios in which very limited training samples are available.

As future research, we are currently developing a version of the presented algorithm in which parameter M is adaptively estimated for each pixel rather than set in advance as in the version of the algorithm reported in this chapter. Interestingly, we have empirically observed that the adaptive selection produces similar results to those obtained in this work with fixed parameter settings such as $M = 2$ or $M = 3$, which result in much lower computational cost than an adaptive estimation of the parameter on a per-pixel basis. As a result, we will continue exploring the possibility to select this parameter adaptively in order to improve the obtained classification results without increasing computational complexity, which currently stays on the same order of magnitude as the other methods used in the comparisons reported in this work.

In future developments, we will further explore the relationship between the parameters of our method and the spatial resolution, level of noise, and complexity of the analyzed scenes. We are also planning on exploring the applications of the presented method for the analysis of multitemporal data sets.

Chapter 3

Subspace-Based MLR Method for Hyperspectral Image Classification

3.1 Summary

In this chapter, we propose an MLR method for pixelwise hyperspectral classification. The feature vectors are formed by the energy of the spectral vectors projected on class-indexed subspaces. In this way, we model not only the linear mixing process that is often present in the hyperspectral measurement process but also the nonlinearities that are separable in the feature space defined by the aforementioned feature vectors. Our experimental results have been conducted using both simulated and real hyperspectral data sets, which are collected using NASA's AVIRIS and the ROSIS system. These results indicate that the proposed method provides competitive results in comparison with other state-of-the-art approaches [1].

3.2 Introduction

Hyperspectral sensors provide images in hundreds of continuous (narrow) spectral bands that can be used to discriminate different objects on the earth surface [1]. Recently, MLR has shown good performance in hyperspectral image classification. MLR is a discriminative approach that directly models the posterior class distributions [132, 42, 43, 51]. Recent examples on the use of MLR in hyperspectral classification problems can be found in [21, 133, 134, 135]. In this type of classifiers, we highlight the MLR*sub* method [51] that was specifically designed with the linear spectral mixing process in mind. In the MLR*sub* method, the classification of a pixel (with its associated spectral vector in a given class) corresponds to the largest projection of that vector onto the class indexed subspaces. In this work, and in order to model possible nonlinear mixing effects, we allow the MLR regression vectors to define arbitrary linear combinations of the projections of the subspaces learned from the training set. In comparison with the work in [51], which originally proposed MLR*sub*, the proposed subspace-based MLR (MLRsub_{mod}) introduces two main contributions:

- First, the newly developed method uses the projection of the original spectral vectors onto class-dependent subspaces in order to enhance class separability. At this point, we can mention two main reasons that support the use of these projections. One reason is that the hyperspectral

---

measurements are often linear mixtures of the endmembers signatures, and then, each class corresponds to a given subset of the endmembers, thus defining a subspace. The other reason is that we claim that a relevant number of nonlinear phenomena present in the hyperspectral image yields spectral vectors that are linearly separable in the feature space defined by the class indexed subspaces.

- Second, of the proposed method consists of including the class prior probabilities in the proposed model. This is expected to introduce advantages in scenarios in which the number of training samples per class depends on the area covered by that particular class in the scene.

The remainder of this chapter is structured as follows. Section 3.3 describes the newly proposed subspace based MLR method. Section 3.4 presents experimental results using both simulated and real hyperspectral scenes. The numerical results illustrate that the performance of the MLRsub classification algorithm can be significantly improved by using the proposed subspace-based projection feature vectors, and it incorporates the prior information from the known class proportions. Finally, Section 3.5 concludes this chapter with some remarks.

3.3 Class Dependent Subspace Based MLR (MLRsub_{mod})

In [51], it was shown that the posterior class density $p(y_i = c|\mathbf{x}_i)$ can be computed in the MLR framework by using the nonlinear functions $\mathbf{h}^{(c)}(\mathbf{x}_i) = [\|\mathbf{x}_i\|^2, \|\mathbf{x}_i^T\mathbf{U}^{(c)}\|^2]^T$, where $\mathbf{U}^{(c)}$ is a set of $r^{(c)}$-dimensional orthonormal basis vectors for the subspace associated with classes $c = 1, 2, \ldots, k$. Following [51], in this chapter $\mathbf{U}^{(c)}$ is computed as $\mathbf{U}^{(c)} = \{\mathbf{e}_1^{(c)}, \ldots, \mathbf{e}_{r^{(c)}}^{(c)}\}$, while $\mathbf{E}^{(c)} = \{\mathbf{e}_1^{(c)}, \ldots, \mathbf{e}_d^{(c)}\}$ is the eigenvector matrix computed from correlation matrix $\mathbf{R}^{(c)} = \mathbf{E}^{(c)}\mathbf{\Lambda}^{(c)}\mathbf{E}^{(c)^T}$. Here, $\mathbf{\Lambda}$ is the eigenvalue matrix with decreasing magnitude. In this chapter, following [51], we use a subspace projection accounting for 99.9% of the original spectral information in order to determine the size of $\mathbf{U}^{(c)}$. As shown in [51], the MLRsub method aims to deal with the problems defined by the linear mixing model. However, nonlinear mixing is very common in real scenarios. We claim that a number of nonlinearities present in the hyperspectral mixing process are, approximately, linearly separable in the feature space defined by the nonlinear functions $\mathbf{h}(\mathbf{x}_i) = [\|\mathbf{x}_i\|^2, \ \|\mathbf{x}_i^T\mathbf{U}^{(1)}\|^2, \ \ldots, \ \|\mathbf{x}_i^T\mathbf{U}^{(k)}\|^2]^T$, $i.e.$, the vector features containing as components the energy of the projections on all class subspaces plus the energy of the original vector. This claim will be supported in Section 3.4 with the experimental results.

Here, we use a nonlinear vector of regression functions $\mathbf{h}(\mathbf{x}_i) = [\|\mathbf{x}_i\|^2, \|\mathbf{x}_i^T\mathbf{U}^{(1)}\|^2, \ \ldots, \ \|\mathbf{x}_i^T\mathbf{U}^{(k)}\|^2]^T$ to compute the posterior class density $p(y_i = c|\mathbf{x}_i)$ for a given class c as follows:

$$p(y_i = c|\mathbf{x}_i, \boldsymbol{\omega}) = \frac{\exp\left(\boldsymbol{\omega}^{(c)^T}\mathbf{h}(\mathbf{x}_i)\right)p(y_i = c)}{\sum_{l=1}^k \exp\left(\boldsymbol{\omega}^{(l)^T}\mathbf{h}(\mathbf{x}_i)\right)p(y_i = l)}, \tag{3.1}$$

where, by assuming $p(y_i = c) = 1/k$, we exactly have an MLR classifier. However, in order to introduce the available prior knowledge, here, we include the estimation of the occurrence probabilities of each land-cover class from the training set. The prior probability for class c may be computed as $p(y_i = c) = \frac{n_{tr}^{(c)}}{n_{tr}}$ [136].

Notice that, if the data live in a class dependent subspace defined by the linear mixing model, the proposed approach (3.1) can be recovered by the conventional MLRsub in [51] by a setting of regressing parameters $\boldsymbol{\omega}^{(c)} = [\omega_1, 0 \ldots, 0, \omega_{c+1}, 0, \ldots, 0]^T$. Another important aspect is that, if the data do not

strictly live in a linear subspace and follow a nonlinear mixing model (which is a quite common scenario in practice), then as supported in Section 3.4, the proposed MLRsub_{mod} approach is able to separate the classes in the newly proposed feature space. Therefore, the proposed approach has the ability to handle both linear and nonlinear mixtures, which is the main contribution of this chapter. However, further work should be conducted in order to fully analyze how the assumed dependence between the classes handles the nonlinearity of the mixtures.

Under the present setup, we compute $\boldsymbol{\omega}$ in (3.1) by calculating the maximum a posteriori estimate as follows:

$$\widehat{\boldsymbol{\omega}} = \arg\max_{\boldsymbol{\omega}} \quad \ell(\boldsymbol{\omega}) + \log p(\boldsymbol{\omega}), \tag{3.2}$$

where $\ell(\boldsymbol{\omega}) \equiv \log \prod_{i=1}^{n_{tr}} p(y_i | \mathbf{x}_i, \boldsymbol{\omega})$ is the log-likelihood function. Similar to the MLRsub algorithm in [51], $p(\boldsymbol{\omega}) \propto e^{-\beta/2\|\boldsymbol{\omega}\|^2}$ ($\beta \geq 0$ is a regularization parameter controlling the weight of the prior) is a quadratic prior on $\boldsymbol{\omega}$ that is intended to cope with difficulties in learning the regression vector $\boldsymbol{\omega}$ associated with bad or ill-conditioning of the underlying inverse problem.

The optimization problem in (3.2) is convex, although the term $\ell(\boldsymbol{\omega})$ is non-quadratic. Following previous work in [44, 137, 51], we approximate this term by a quadratic lower bound, which leads to a sequence of quadratic problems that are easier to solve than the original problem.

3.4 Experimental Results and Discussion

In this section, we evaluate the proposed class subspace based MLR by using both simulated and real hyperspectral images. For the parameter settings, we follow the indications given in [51] and include a comparison with the MLRsub in [51]. It should be noted that, in this chapter, we only compare our MLRsub_{mod} with MLRsub. The main reason is that the proposed subspace based features yield better performance than those used in MLRsub. Another reason is that, in [51], there is already a comprehensive comparison with state-of-the-art methods.

3.4.1 Experiments with Simulated Data

In order to have an assessment in a fully controlled environment, we first used a simulated data set to evaluate the capability of the proposed approach for handling nonlinear mixtures. For this purpose, we generated a synthetic image with 50×50 samples for each of the eight classes simulated (see Fig. 3.1). We considered the following nonlinear mixture model for generating each simulated mixed pixel in class c:

$$\mathbf{x}_i^{(c)} = \sum_{j=0}^{M_l} \mathbf{m}^{(c+j)} \gamma_j + \alpha \prod_{j=0}^{M_{nl}} \mathbf{m}^{(c+j)} + \mathbf{n}_i, \tag{3.3}$$

where $\mathbf{m}^{(l)}, l = 1, ..., 10$ are different spectral signatures that were randomly selected from the USGS digital spectral library, γ_j and α are the parameters controlling the impact of the linear and nonlinear terms, respectively, and $\sum_{j=0}^{M_l} \gamma_j = 1 - \alpha$. In our simulation, γ_0 is the abundance of the objective class, i.e. the one that received the maximum abundance value in the simulation and that will define the label for the considered class. To have a comprehensive comparison using both linear and nonlinear mixtures, we used $\alpha = 0$ for classes $\{1, 3, 5, 7\}$, which means that these four classes stay in a linear subspace, and included nonlinear mixtures for classes $\{2, 4, 6, 8\}$. Furthermore, for each pixel we randomly chose a value over $\{1,2\}$ for parameters M_l and M_{nl}, which means that we set the number of mixtures in each pixel to
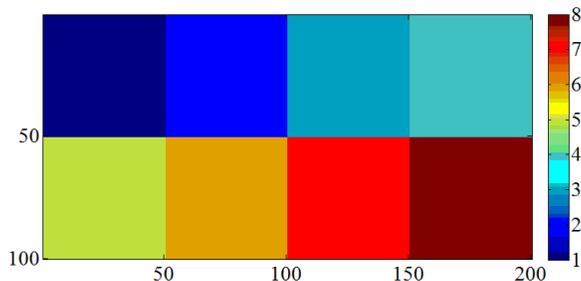
Figure 3.1: Simulated hyperspectral image comprising eight different classes.

2 or 3. Pure spectral signatures are considered for the first four classes, whereas the remaining classes are entirely made up of mixed pixels (see Table 3.1). Finally, zero-mean Gaussian noise with covariance $\sigma^2 \mathbf{I}$, i.e. $\mathbf{n}_i \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$, was added to the generated synthetic image. Here, the noise standard deviation is $\sigma = 0.4$.

The classification experiments using the simulated data set have been conducted as follows. For each class, we randomly chose 250 samples from the available ground truth for training, and the remaining samples were used for testing. Table 3.1 tabulates the accuracy obtained by the proposed method as a function of the value of parameter α that controls the impact of the nonlinear features, in comparison with the conventional MLRsub, and all the values of the OA reported in this section correspond to the average of the accuracy values obtained after 100 Monte Carlo runs. From the results reported in Table 3.1, we can conclude that the results achieved by the proposed MLRsub_{mod} algorithm are superior to those obtained by the MLRsub algorithm for all the considered values of parameter α. However, the improvement is more significant for low values of α. This is because, when parameter α increases, the value of γ_0 (i.e., the dominant class) decreases. If we compare the OA obtained for linear classes $\{1,3,5,7\}$ with regard to the OA obtained for nonlinear classes $\{2,4,6,8\}$, we can observe that the proposed approach has very good improvements. This indicates that the proposed method can efficiently handle nonlinear mixtures. Furthermore, if we compare the OA obtained for classes $\{1,2,3,4\}$ (which contain pure pixels) with the OA obtained for classes $\{5,6,7,8\}$ (which contain mixed pixels), it is apparent that the improvement in OA is more significant for the classes without pure pixels. In other words, the proposed method can better manage mixed pixels instead of pure pixels as mixed pixels stay in the boundaries of the subspaces so that they are more difficult for subspace identification.

3.4.2 Experiments with Real Hyperspectral Data

In this chapter, three different real hyperspectral images were used in our experiments: ROSIS Pavia University, AVIRIS Indian Pines and AVIRIS Salinas. The first two images have been described in subsection (2.5.1.2). Whereas in this chapter, for the AVIRIS Indian Pines image, we discarded four classes: *Alfalfa*, *grass/pasture-mowed*, *oats* and *stone-steel towers* which contain less that 100 labeled pixels. The considered ground truth map is shown in Fig. 3.2.

The third image considered in experiments is the AVIRIS Salinas image, collected over the Valley of Salinas, Southern California, in 1998. It contains 217×512 pixels and 204 spectral bands and is characterized by 3.7 m/pixel spatial resolution. Fig. 3.3(a) shows a false color composite of the image and Fig. 3.3(b) shows the ground-truth map with 16 mutually exclusive classes. Due to the spectral similarity of most classes, this data set also represents a very challenging classification problem.

Table 3.1: Overall classification accuracies [%] for different values of parameter α (with the noise standard deviation set to $\sigma = 0.4$) obtained by the MLRsub and MLRsub_{mod} methods for the simulated data set in Fig. 3.1.

Simulated classes			$\alpha = 0$		$\alpha = 0.3$		$\alpha = 0.4$		$\alpha = 0.5$	
Class	Mixtures	Pure pixels	MLRsub	MLRsub_{mod}	MLRsub	MLRsub_{mod}	MLRsub	MLRsub_{mod}	MLRsub	MLRsub_{mod}
C1	Linear	Yes	79.78	86.85	75.77	81.06	75.37	79.43	72.08	74.71
C2	Nonlinear	Yes	74.70	83.80	75.59	82.83	74.56	80.15	71.82	77.31
C3	Linear	Yes	75.92	86.34	73.26	79.31	69.09	74.78	66.15	70.26
C4	Nonlinear	Yes	76.17	85.34	77.24	84.30	75.01	81.67	72.65	79.39
C5	Linear	No	75.41	83.91	70.34	78.15	65.58	71.97	60.58	66.13
C6	Nonlinear	No	74.86	83.77	69.25	79.05	67.66	76.33	65.44	73.92
C7	Linear	No	74.19	82.51	66.53	75.48	62.84	71.12	58.18	65.06
C8	Nonlinear	No	79.58	86.23	77.19	83.59	74.04	80.23	71.38	77.28
OA{1,2,3,4}	-	Yes	76.64	85.58	75.46	81.87	73.50	79.01	70.67	75.42
OA{5,6,7,8}	-	No	76.01	84.10	70.83	79.07	67.53	74.91	63.90	70.60
OA{1,3,5,7}	Linear	-	76.32	84.90	71.48	78.50	68.22	74.33	64.25	69.04
OA{2,4,6,8}	Nonlinear	-	76.33	84.78	74.82	82.44	72.82	79.59	70.32	76.97
Overall accuracy (OA)			76.33	84.84	73.15	80.47	70.52	76.96	67.28	73.01
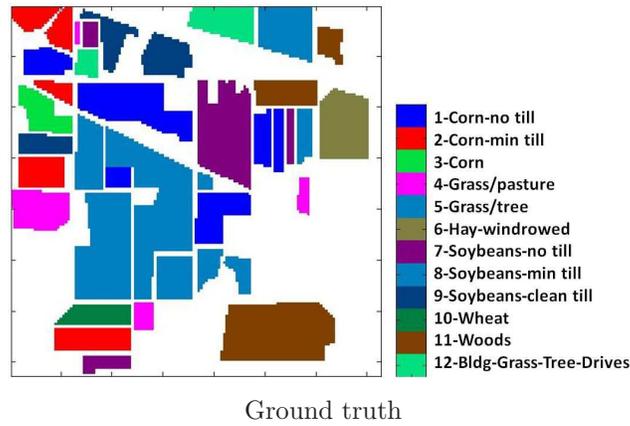
Ground truth

Figure 3.2: Twelve class Ground truth for the AVIRIS Indian Pines image

(a) False color composition   (b) Ground-truth
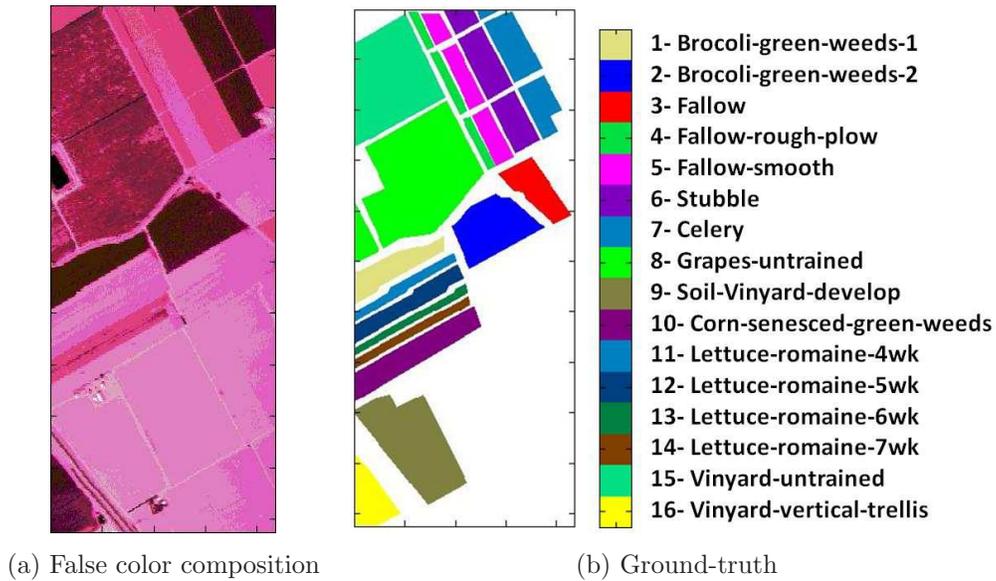
Figure 3.3: AVIRIS Salinas data set.

In our experiments with real hyperspectral scenes, we designed two strategies to choose the training set. In our first strategy, we choose a constant number of training samples per class. In the second strategy, we choose a number of training samples per class that is proportional to the number of available labeled samples. The classification accuracy reported for the real scenes was obtained after 30 Monte Carlo runs.

Tables 3.2, 3.3 and 3.4 summarize the OA, AA, kappa coefficient (κ) and the class-specific accuracy values for the two considered images, respectively. If we focus on the results reported for the classes in which different numbers of training samples are selected for different classes, it is noticeable that the MLRsub_{mod}, which includes class dependent information and integrates the prior distribution of classes in the scene, significantly improves the classification accuracy provided by the MLRsub. For instance, Table 3.3 shows that the proposed MLRsub_{mod} approach obtained an OA of 76.71% and an AA of 74.97%, which contrasts with an OA of 73.51% and an AA of 68.30% achieved by MLRsub in the AVIRIS Indian

MLR*sub*, OA=68.42%, AA=72.52%   MLR*sub$_{mod}$*, OA=76.35%, AA=81.87%

Figure 3.4: Classification results obtained for the ROSIS Pavia University data set with 781 training samples

Pines scene. Furthermore, the results reported in Tables 3.2 and 3.4 reveal more significant improvements for the Pavia University and Salinas images. Using the proposed method for the Pavia University image, the OA and the AA are improved by 5.57% and 8.09%, respectively, compared to the MLR*sub*. More significantly, for the AVIRIS Salinas image, the improvements in OA an AA are 9.46 % and 19.1%, respectively.

If we focus on the results reported for the classes in which a constant number of training samples is selected for all classes, we can see that the class specific accuracy values for MLR*sub$_{mod}$* are higher compared with those of MLR*sub* in most of the classes. This reveals that the proposed projection based feature vectors provide a more consistent estimation of the posterior probability distributions. For illustrative purposes, some classification maps are shown in Figs. 3.4, 3.5 and 3.6. These maps correspond to one of the 30 Monte Carlo runs conduced for each scene. Effective results can be observed in these figures.

3.5 Summary and future directions

In this chapter, we have developed a subspace-based MLR method for pixelwise hyperspectral classification. The proposed approach assumes that the observed vectors live in subspaces constructed by the classes and represents an extension of a previous methodology in which class independence was assumed. An important contribution of the proposed approach lies in its ability to deal with both linear and nonlinear mixtures. Our experimental results, which are conducted using both simulated and real hyperspectral data sets collected using NASA's AVIRIS and the ROSIS system, indicate that the proposed algorithm accurately performs in different hyperspectral image classification scenarios, particularly with limited training samples.

Table 3.2: Overall (OA), average (AA), κ statistic and individual classification accuracies [%] obtained for the ROSIS Pavia University image.

Training set	100 samples per class (900 in total)				20% of the 3921 training set (781 in total)			
	Samples		Classification methods		Samples		Classification methods	
Class	Train	Test	MLRsub	MLRsub_{mod}	Train	Test	MLRsub	MLRsub_{mod}
Asphalt	100	6531	44.61	67.00	109	6522	61.69	67.00
Meadows	100	18549	63.33	73.06	108	18541	72.89	76.67
Gravel	100	1999	65.91	66.94	78	2021	46.48	66.56
Trees	100	2964	76.68	94.36	104	2960	78.08	93.62
Metal sheets	100	1245	98.89	99.04	53	1292	98.45	98.87
Bare soil	100	4929	69.21	62.59	106	4923	65.37	64.25
Bitumen	100	1230	86.46	85.75	75	1255	66.60	85.56
Bricks	100	3582	64.22	78.07	102	3580	73.12	78.11
Shadows	100	847	99.29	99.71	46	901	94.73	99.57
OA			64.91	74.35	OA		70.46	76.03
AA			74.29	80.72	AA		73.04	81.13
κ			56.27	67.34	κ		62.20	69.31

Table 3.3: Overall (OA), average (AA), κ statistic and individual classification accuracies [%] obtained for the AVIRIS Indian Pines image.

Training set	100 samples per class (1200 in total)				10% of available samples per class (1011 in total)			
Class	Samples		Classification methods		Samples		Classification methods	
	Train	Test	MLRsub	MLRsub_{mod}	Train	Test	MLRsub	MLRsub_{mod}
Corn-no till	100	1334	60.63	64.90	143	1291	60.80	71.04
Corn-min till	100	734	65.56	65.40	83	751	39.61	60.04
Corn	100	134	89.07	86.89	23	211	24.44	46.57
Grass/trees	100	397	88.67	89.71	49	448	81.57	84.19
Grass/pasture	100	647	94.29	93.86	74	673	92.42	94.31
Hay-windrowed	100	389	99.38	99.07	48	441	99.24	98.34
Soybeans-notill	100	868	72.98	67.36	96	872	48.00	64.90
Soybeans-min till	100	2368	55.07	58.55	246	2222	89.68	78.33
Soybeans-clean till	100	514	84.01	79.26	61	553	58.09	64.73
Wheat	100	112	99.65	99.65	21	191	98.73	99.56
Woods	100	1194	91.47	91.63	129	1165	99.51	95.93
Bldg-Grass-Tree-Drives	100	280	66.53	70.54	38	342	27.49	41.70
OA			73.59	74.32	OA		73.51	76.71
AA			80.61	80.57	AA		68.30	74.97
κ			70.20	70.98	κ		68.87	73.23

Table 3.4: Overall (OA), average (AA), κ statistic and individual classification accuracies [%] obtained for the AVIRIS Salinas image.

Training set	100 samples per class (1600 in total)				2% of available samples per class (1076 in total)			
	Samples		Classification methods		Samples		Classification methods	
Class	Train	Test	MLRsub	MLRsub_{mod}	Train	Test	MLRsub	MLRsub_{mod}
Brocoli-green-weeds-1	100	1909	98.83	99.66	40	1969	86.39	99.37
Brocoli-green-weeds-2	100	3626	98.91	99.16	74	3652	99.04	99.13
Fallow	100	1876	59.76	95.64	39	1937	30.95	88.80
Fallow-rough-plow	100	1294	64.67	99.44	27	1367	6.33	99.19
Fallow-smooth	100	2578	99.28	98.15	53	2625	99.81	96.77
Stubble	100	3859	99.69	99.43	79	3880	99.70	99.36
Celery	100	3479	99.86	99.64	71	3508	99.81	99.59
Grapes-untrained	100	11171	58.52	66.85	225	11046	89.00	76.46
Soil-vinyard-develop	100	6103	99.70	98.59	124	6079	99.83	99.05
Corn-senesced-green-weeds	100	3178	92.04	91.64	65	3213	90.67	90.51
Lettuce-romaine-4wk	100	968	91.73	97.65	21	1047	8.72	93.13
Lettuce-romaine-5wk	100	1827	77.68	99.51	38	1889	55.54	98.35
Lettuce-romaine-6wk	100	816	99.59	98.89	18	898	98.94	97.02
Lettuce-romaine-7wk	100	970	96.00	98.00	21	1049	90.54	92.29
Vinyard-untrained	100	7168	76.40	72.61	145	7123	36.25	68.89
Vinyard-vertical-trellis	100	1707	98.53	98.25	36	1771	98.16	97.32
OA			84.02	88.17	OA		79.37	89.01
AA			88.20	94.57	AA		74.35	93.45
κ			82.27	86.87	κ		76.82	87.77

MLRsub, OA=72.04%, AA=66.43%   MLRsub_{mod}, OA=76.88%, AA=73.76%

Figure 3.5: Classification results for the AVIRIS Indian Pines data set using 1076 training samples
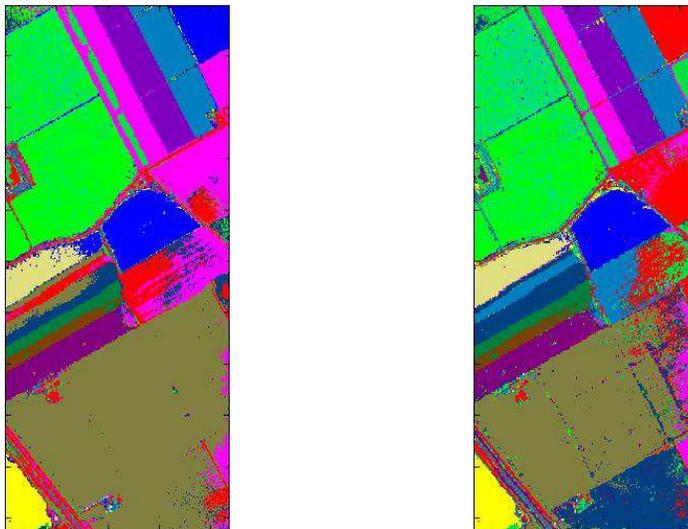
MLRsub, OA=80.81%, AA=78.01%   MLRsub_{mod}, OA=88.22%, AA=92.18%

Figure 3.6: Ground truth and classification results obtained for the AVIRIS Salinas data set with 1076 training samples

Chapter 4

Hyperspectral Image Classification Based on Union of Subspaces

4.1 Summary

Characterizing mixed pixels is an important topic in the analysis of hyperspectral data. Recently, a subspace-based technique in a MLR framework called MLR*sub* has been developed to address this issue. MLR*sub* assumes that the training samples of each class live in a single low-dimensional subspace. However, having in mind that materials in a given class tend to appear in groups and the (possible) presence on nonlinear mixing phenomena, a more powerful model is a union of subspaces. This chapter presents a new approach based on union of subspaces for hyperspectral images. The proposed method integrates subspace clustering with MLR method for supervised classification. Our experimental results using a newly released urban hyperspectral image indicate that the proposed method exhibits state-of-the-art classification performance [1].

4.2 Introduction

The discrimination of different objects on the earth surface can be achieved by processing of the hundreds of continuous narrow spectral bands collected by hyperspectral sensors [1]. Supervised classification in hyperspectral image processing is defined as the task of assigning a unique label to each pixel vector of the image under consideration using the information extracted from labeled training samples a priori. Several techniques have been used to perform supervised classification of hyperspectral data. In particular, using the class probability estimates resulting from a probabilistic classifier in an MRF framework, allows us to have more accurate classification results by integrating spectral and spatial information [138, 51, 102]. Recently, the MLR has shown good performance in hyperspectral image classification which models the posterior class distributions in a Bayesian framework [139, 140, 42, 43]. Specifically, the integration of a subspace projection method with the MLR algorithm (called MLR*sub*) has shown significant classification results [139, 51]. The assumption that hyperspectral vectors live in a lower-dimensional subspace is strongly linked with the LMM [51, 141]. In essence, if each class is associated with a group of materials, then the spectral vectors of this class are convex combinations of the spectral signatures from that class

---

[1]Part of this chapter has been published in: M. Khodadadzadeh, J. Li, A. Plaza and J. M. Bioucas-Dias. **Hyperspectral Image Classification Based on Union of Subspaces**. IEEE Joint Urban Remote Sensing Event (JURSE'15), Lausanne, Switzerland, 2015.
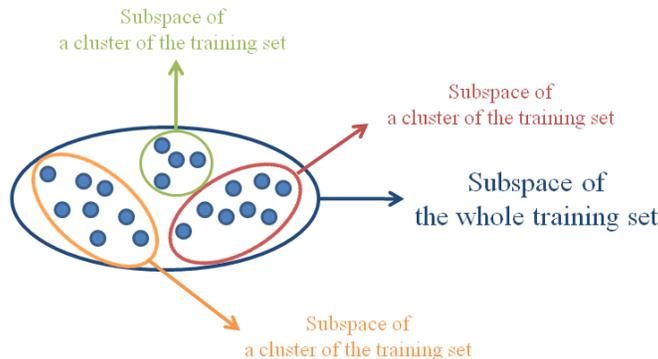
Figure 4.1: Graphical illustration of the concept of union of subspaces

and thus they live in the subspace spanned by those spectral signatures.

There are number of factors which degrade the modeling power of the subspace model. Among these factors, we highlight the possible presence of nonlinear mixing phenomena, and the typical low spatial resolution of hyperspectral images, which increases the likelihood of having mixed pixels from a number of different groups (clusters) of materials. Under these degrading factors, a better model is that the spectral vectors in given class lie in unions of subspaces [142]. This concept is illustrated in Fig. 4.1. Under the single subspace model, the class would be represented by the set in blue (a subspace). Under the union of subspaces, the class is represented by the union of red, green, and orange sets (subspaces). Notice that the single subspace model contains subsets that are not representative of the training set, which is not the case with the union of subspaces.

Exploiting the union of subspaces in an MLR framework for supervised hyperspectral image classification is the main contribution of this chapter. For this purpose, we suggest to use a subspace clustering method before the classification in order to divide training samples of each class into multiple subsets regarding to existing subspaces. Subspace clustering refers to the task of finding a multi-subspace representation that best fits high dimensional data samples, i.e. finding the number of subspaces and their dimensions and simultaneously clustering the data into multiple subspaces [143]. In this chapter, we introduce a new hyperspectral image classification methodology based on RSC [143] which, so far as we are aware, has not been applied for hyperspectral image analysis.

This chapter is organized as follows. In the next section we discuss methodological framework of the proposed algorithm, including subspace clustering and subspace-based MLR classifier. The experimental results are presented in Section 4.4. Finally, the proposed method is concluded and discussed in Section 4.5.

4.3   Methodological Framework

The proposed approach mainly comprises two main steps: 1) subspace clustering of training samples set; 2) subspace projection and probabilistic classification using MLR algorithm. In the following, we present the details of each step for the proposed approach.

4.3.1 Subspace Clustering

The first step of the proposed procedure consists in performing subspace clustering to find lower-dimensional subspaces of the available training set of the hyperspectral image. Recently, several subspace clustering algorithms have been developed. However, most of them are working under restrictive conditions [143]. Here, we suggest to use RSC [143] algorithm, which, as our experiments showed, is well suited for clustering of hyperspectral data. Based on the ideas from geometric functional analysis, the RSC method can accurately recover the underlying subspaces under minimal requirements on their orientation, and on the number of samples per subspace [143].

The RSC method is the extension of the SSC method [144] to cluster noisy data, that is always an important issue in hyperspectral image clustering. The SSC method applies spectral clustering to an adjacency matrix, obtained by sparsely representing each data point in terms of all the other data points through l_1-minimization; whereas, the RSC algorithm replaces the l_1-minimization step in SSC by an l_1-penalized least squares, i.e., Lasso, step and successfully performs subspace clustering under Gaussian noise.

4.3.2 Subspace-based MLR classifier

In [139, 51], the subspace for the class c, for $c = 1, 2, \ldots, k$, is estimated via eigen-analysis of the spectral vectors available in the set $\mathcal{D}^{(c)}$. The respective subspace is represented by the orthogonal matrix $\mathbf{U}^{(c)}$ holding on its columns an orthogonal basis computed from the sample correlation matrix of those vectors.

As already stated, in these work we assume the spectral vectors $\mathbf{x}_i \in \mathcal{D}^{(c)}$ live in an union of subspace to be learnt using the RSC method. The output of RSC consists in a partition of $\mathcal{D}^{(c)}$ into $L^{(c)}$ subsets. That is, for each c, we obtain the collection of sets $\mathcal{D}_i^{(c)}$ such that $\mathcal{D}_i^{(c)} \cap \mathcal{D}_j^{(c)} = \emptyset$ for $i \neq j$ and $\mathcal{D}^{(c)} = \mathcal{D}_1^{(c)} \cup \mathcal{D}_2^{(c)} \cup \ldots \cup \mathcal{D}_{L^{(c)}}^{(c)}$.

The obtained collection of subspaces is exploited by including the norms of the projection of the spectral vectors onto the subspaces estimated by RSC. More concretely, we propose the following feature vector:

$$\mathbf{h}(\mathbf{x}_i) = [\|\mathbf{x}_i\|^2, \ \|\mathbf{x}_i^T \mathbf{U}_1^{(1)}\|^2, \ \ldots, \ \|\mathbf{x}_i^T \mathbf{U}_{L^{(1)}}^{(1)}\|^2, \ \ldots, \|\mathbf{x}_i^T \mathbf{U}_1^{(k)}\|^2, \ \ldots, \ \|\mathbf{x}_i^T \mathbf{U}_{L^{(k)}}^{(k)}\|^2]^T, \qquad (4.1)$$

where, $\mathbf{U}_l^{(c)}$, $c = 1, \ldots, k$ and $l = 1, \ldots, L^{(c)}$ are orthogonal matrices holding basis for the subspaces spanned by the sets $\{\mathbf{x}_i \in \mathcal{D}_l^{(c)}\}$.

Similarly to [43], we apply the LORSAL algorithm [46] to estimate the regressors for the proposed subspace-based MLR classifier. The pseudocode of the proposed algorithm, referred as MLR*usub*, is shown in Algorithm 1.

As shown in Algorithm 1, for the input, we have the hyperspectral data $\mathbf{x}$ and the training set $\mathcal{D}$. The classification objective is the image of classes labels $\mathbf{y}$. In steps 3 and 4, we use the RSC method to cluster the class dependent training sets, and estimate the individual subspaces for every cluster. Then in step 5, the union of subspaces is obtained and in step 6, we use LORSAL for classification as LORSAL is able to manage linear/nonlinear features. Finally, in step 7, we obtain the image of class labels following model (2.3).

Algorithm 1 MLR*usub* (union of subspaces MLR)

1: **Input:** $\mathcal{D}$, $\mathbf{x}$
2: **Output:** $\mathbf{y}$
for $c := 1$ to k **do**
   3: $\left\{\mathcal{D}_l^{(c)}\right\}_{l=1}^{L^{(c)}} = \mathbf{RSC}(\mathcal{D}^{(c)})$ ($*$ subspace clustering $*$)
   for $l := 1$ to $L^{(c)}$ **do**
     4: $\mathbf{U}_l^{(c)} = \mathbf{sub}(\mathcal{D}_l^{(c)})$ ($*$ subspace computation $*$)
   end for
end for
5: $\mathbf{U} \equiv \left\{\mathbf{U}_1^{(1)}, \ldots, \mathbf{U}_{L^{(1)}}^{(1)}, \ldots, \mathbf{U}_1^{(k)}, \ldots, \mathbf{U}_{L^{(k)}}^{(k)}\right\}$
6: $\hat{\boldsymbol{\omega}} = \mathbf{LORSAL}(\mathbf{U}, \mathcal{D})$
7: $\mathbf{y} \equiv \mathbf{MLR}(\mathbf{x}, \hat{\boldsymbol{\omega}})$

4.4 Experimental Results

In this section, use a newly released urban hyperspectral image, University of Houston, to evaluate the proposed approach.

This data set was acquired on June 23, 2012 between the times 17:37:10 to 17:39:50 UTC by the NCALM over the University of Houston campus and its neighboring area. The average height of the sensor above ground was 5500ft. The hyperspectral image has 144 bands in the 380-1050 nm spectral region and spatial resolution 2.5 m. The image size in pixels is 349×1905. Fig. 4.2(a) shows a false color composite of the image, while Fig. 4.2(b) shows 15 classes of interest. In the original data set, 2832 were used for training and 12197 samples were used for testing [see Fig. 4.2]. Table 4.1 details the classes and the number of available training and test samples for each class. This data was distributed for the 2013 Data Fusion Contest of the IEEE GRSS. Detailed information about this image can be found in [145].

To have a complete comparison, the classification results are reported for the three probabilistic classifiers: SVM with RBF kernel [30], MLR*sub* [51] and MLR*sub*$_{mod}$ [139]. Concerning the classifiers, we optimized the related parameters. Figs. 4.3 (a)-(d) shows the obtained classification maps. As can be seen, using union of subspaces the MLR*usub* classifier gives higher classification accuracies than the other subspace based MLR classifiers. For instance, the proposed approach obtained an overall accuracy of 82.94% which is 11.44% higher than the result obtained by the MLR*sub* algorithm. More importantly, in the right part of the image where a large cloud shadow is present, the performance improvements reported for the proposed method are quite significant. For example for the class "Highway" in the cloud-covered region, we can see a significant improvement in the obtained classification result.

4.5 Summary and future directions

In this chapter a new classification method based on union of subspaces was proposed for characterizing mixed (linear and nonlinear) pixels in hyperspectral images. For this purpose, we exploited a subspace clustering method to partition training samples obtained for each class to several subsets and then the MLR algorithm was used to learn the posterior probability distributions from the spectral information of each subset, using a subspace projection. Our experimental results with a new urban hyperspectral image collected by the NCALM over the University of Houston campus showed that the proposed method exhibits state-of-the-art classification performance as compared to other widely used methods.

(a) False color composition
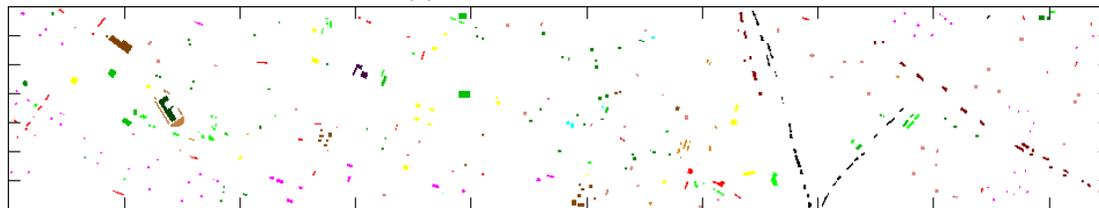
(b) Test samples
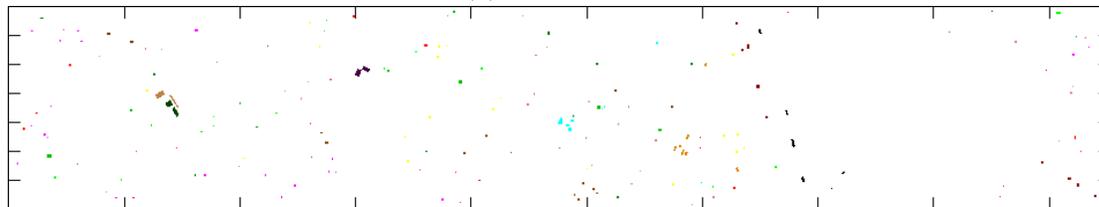
(c) Training samples
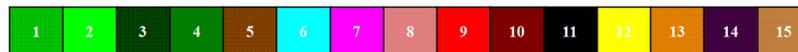
(d) Labels color: 1-Healthy grass, 2-Stressed grass, 3-Synthetic grass, 4-Trees, 5-Soil, 6-Water, 7-Residential, 8-Commercial,

9-Road, 10-Highway, 11-Railway, 12-Parking Lot 1, 13-Parking Lot 2, 14-Tennis Court, 15-Running Track

Figure 4.2: University of Houston data set, classification maps and overall and average classification accuracies (in the parentheses) obtained by different methods.

(a) SVM (80.49%,83.37%)

(b) MLR*sub* (71.50%,74.44%)

(c) MLR*sub*$_{mod}$ (79.86%,82.51%)

(d) MLR*usub* (82.94%,85.19%)

(e) Labels color: 1-Healthy grass, 2-Stressed grass, 3-Synthetic grass, 4-Trees, 5-Soil, 6-Water, 7-Residential, 8-Commercial, 9-Road, 10-Highway, 11-Railway, 12-Parking Lot 1, 13-Parking Lot 2, 14-Tennis Court, 15-Running Track

Figure 4.3: University of Houston data set, classification maps and overall and average classification accuracies (in the parentheses) obtained by different methods.

Table 4.1: Information classes and training-test samples.

Class		Samples	
Number	Name	Train	Test
1	Healthy grass	198	1053
2	Stressed grass	190	1064
3	Synthetic grass	192	505
4	Trees	188	1056
5	Soil	186	1056
6	Water	182	143
7	Residential	196	1072
8	Commercial	191	1053
9	Road	193	1059
10	Highway	191	1036
11	Railway	181	1054
12	Parking Lot 1	192	1041
13	Parking Lot 2	184	285
14	Tennis Court	181	247
15	Running Track	187	473
Total		2832	12197

Table 4.2: Overall (OA), average (AA), κ statistic and individual classification accuracies [%] obtained for the University of Houston image.

Class	Samples		Classification methods			
	Train	Test	SVM	MLRsub	MLRsub_{mod}	MLR$usub$
Healthy grass	198	1053	82.34	82.05	82.72	80.72
Stressed grass	190	1064	80.55	81.39	81.95	82.14
Synthetic grass	192	505	99.80	100.00	100.00	100.00
Trees	188	1056	92.80	98.30	96.31	91.29
Soil	186	1056	98.11	99.53	98.67	97.63
Water	182	143	95.10	82.52	95.10	99.30
Residential	196	1072	74.91	90.58	74.07	81.53
Commercial	191	1053	35.04	34.85	39.13	45.30
Road	193	1059	75.92	66.10	77.71	76.39
Highway	191	1036	58.30	46.72	67.57	89.77
Railway	181	1054	80.46	54.46	85.77	84.72
Parking Lot 1	192	1041	71.37	31.80	73.87	83.29
Parking Lot 2	184	285	2.46	51.93	69.47	71.58
Tennis Court	181	247	100.00	99.60	100.00	98.38
Running Track	187	473	97.89	96.83	95.35	95.77
OA			75.99	71.50	79.86	82.94
AA			76.34	74.44	82.51	85.19
κ			74.16	69.06	78.19	81.51

Chapter 5

Subspace-Based SVM Method for Hyperspectral Image Classification

5.1 Summary

Hyperspectral image classification has been a very active area of research in recent years. It faces challenges related with the high dimensionality of the data and the limited availability of training samples. In order to address these issues, subspace-based approaches have been developed to reduce the dimensionality of the input space in order to better exploit the (limited) training samples available. An example of this strategy is a recently developed MLR*sub* technique able to characterize mixed pixels, which are also an important concern in the analysis of hyperspectral data. In this chapter, we extend the subspace-projection-based concept to SVMs, a very popular technique for remote sensing image classification. For that purpose, we construct the SVM nonlinear functions using the subspaces associated to each class. The resulting approach, called SVM*sub*, is experimentally validated using a real hyperspectral data set collected using the National Aeronautics and Space Administration's Airborne Visible/Infrared Imaging Spectrometer. The obtained results indicate that the proposed algorithm exhibits good performance in the presence of very limited training samples [1].

5.2 Introduction

Hyperspectral image classification has been a very active area of research in recent years [91]. Given a set of observations (*i.e.*, pixel vectors in a hyperspectral image), the goal of classification is to assign a unique label to each pixel vector so that it is well-defined by a given class [11]. Although techniques for unsupervised classification and/or clustering have also been used in the literature [12], supervised classification has been more widely used [59] but it also faces challenges related with the high dimensionality of the data and the limited availability of training samples [11]. Even though hyperspectral images are characterized by their high spectral resolution, which allows capturing fine details of the spectral characteristics of materials in a wide range of applications [146], it has been demonstrated that the original spectral features contain high redundancy [12]. Specifically, there is a high correlation between adjacent bands and the number of the original spectral features may be too high

---

for classification purposes [59]. In addition, the original spectral features may not be the most effective ones to separate the objects of interest from others, since the hyperspectral data may effectively live in a lower-dimensional subspace [47]. These observations have fostered the use of subspace-based techniques for hyperspectral image classification, aimed at reducing the dimensionality of the input space in order to better exploit the (often limited) training samples available *a priori*.

In the hyperspectral imaging literature, subspace-based techniques have been widely used in spectral unmixing problems [147, 141], which interpret mixed pixels in a hypespectral scene in terms of a collection of pure spectral signatures (*endmembers* [148]) and their corresponding abundance fractions [149]. The connections between spectral unmixing and subspace projection were first explored in [150]. In [47], a technique called HySime was presented to identify the subspace in which the hyperspectral data lives, which is related with the estimation of the number of endmembers in a given scene. However, subspace-based techniques have also been used for detection and classification purposes, mainly due to their capacity to deal with mixed pixels and interferers. For instance, in [52] (later revisited in [48]), an OSP technique was introduced for hyperspectral image classification and dimensionality reduction. In [53] a least squares subspace projection approach to mixed pixel classification was discussed. In [54], a kernel-based subspace projection technique was presented and evaluated in the context of an agricultural application. In [151], a technique for hyperspectral signal subspace identification in the presence of rare signal components was explored. More recently, subspace-projection-based MLR classifier, called MLR*sub* [51], has been presented. This classifier models the subspace associated with each specific class. In other words, the MLR*sub* uses a class dependent procedure integrated with the MLR classifier to represent each class using a subspace spanned by a set of basis vectors. This approach exhibited good classification performance using several hyperspectral scenes. A general conclusion from the aforementioned studies is that subspace projection methods are useful for the separation of classes which are very similar in spectral terms due to spectral mixing and other phenomena.

Inspired by the previous development of MLR*sub*, this chapter presents a new methodology that combines a class indexed subspace projection technique integrated with the SVM classifier [31, 32], which has been widely used in order to deal effectively with the Hughes phenomenon. This phenomenon is related with the imbalance between the high dimensionality (in spectral sense) of hyperspectral data and the (generally limited) number of training samples available, which often compromises the performance of supervised classification techniques [11]. The SVM was first investigated as a binary classifier [28]. Given a training set mapped into a space by some mapping, the SVM separates the data by an optimal hyperplane. If the data are linearly separable, we can select two hyperplanes in a way that they separate the data and there are no points between them, and then try to maximize their distance. The region bounded by them is called the margin [152]. If the data are not linearly separable, soft margin classification with slack variables can be used to allow mis-classification of difficult or noisy cases. However, the most widely used approach in SVM classification is to combine soft margin classification with a kernel trick that allows separation of the classes in a higher dimensional space by means of a nonlinear transformation. In other words, the SVM used with a kernel function is a nonlinear classifier, where the nonlinear ability is included in the kernel and different kernels lead to different types of SVMs. The extension of SVMs to multi-class problems is usually done by combining several binary classifiers [28]. In this chapter, our main contribution is to incorporate a subspace-projection based approach to the classic SVM formulation, with the ultimate goal of having a more consistent estimation of the class distributions. The resulting classification technique, called SVM*sub*, is shown in this work to be robust to the presence of noise, mixed pixels and limited training samples.

The remainder of the chapter is organized as follows. Section 5.3 presents the proposed SVM*sub*

classification technique. Section 5.4 evaluates the classification performance of the SVM*sub* method in comparison with another established subspace-based classifiers such as the MLR*sub*, using a hyperspectral data set collected by the AVIRIS over the Indian Pines region in Indiana. Our experimental results indicate that the proposed SVM*sub* algorithm provides competitive classification results in comparison with other approaches. Section 5.5 concludes this chapter with some remarks and hints at plausible future research lines.

5.3 Class Dependent Subspace-Based SVM (SVM*sub*)

In chapter (3), a nonlinear function $\mathbf{h}(\mathbf{x}_i) = [\|\mathbf{x}_i\|^2, \ \|\mathbf{x}_i^T \mathbf{U}^{(1)}\|^2, \ \ldots, \ \|\mathbf{x}_i^T \mathbf{U}^{(k)}\|^2]^T$ was used to learn the logistic regressors. With the aforementioned notation in mind, we now extend the class dependent subspace concept to the SVM classifier, and also define the implementation of the nonlinear function $\mathbf{h}(\mathbf{x}_i)$ for the SVM model. Therefore, after the nonlinear transformation, the hyperspectral data $\mathbf{x}$ turns to $\mathbf{h}(\mathbf{x}) \equiv \{\mathbf{h}(\mathbf{x}_1), \ldots, \mathbf{h}(\mathbf{x}_n)\}$ in the SVM implementation.

The SVM is a supervised non-parametric statistical learning technique which learns from a set of labeled data instances, thus trying to find an optimal hyperplane that separates the dataset into a discrete (and predefined) number of classes in a way that is consistent with the training examples [32]. Here, the notion of optimal separation hyperplane refers to the decision boundary that minimizes misclassifications, which is obtained after the training step. In our context, this would be equivalent to assuming that the original spectral features are linearly separable in the input space. In practice, linear separability is difficult as the basic linear decision boundaries are often insufficient to properly model the data. For this purpose, a kernel trick is used to solve the inseparability problem by mapping the nonlinear correlations into a higher-dimensional space [153].

Obviously, the choice of the input function can have a significant impact on the obtained results [59]. In our approach, by introducing the class dependent subspace based nonlinear function $\mathbf{h}(\mathbf{x})$, the decision rule for a pixel u can be obtained as:

$$y_u = \mathrm{sgn}\left(\sum_{i=1}^{n_{tr}} y_i \alpha_i (\mathbf{h}(\mathbf{x}_i)^T \mathbf{h}(\mathbf{x}_u)) + b\right), \tag{5.1}$$

where n_{tr} is the number of labeled samples, $y_i \in \{-1, 1\}$, and $0 \leq \alpha_i \leq C$, with C being the soft margin parameter. For simplicity, sometimes, it is required that the hyperplane passes through the origin of the coordinate system, whereas general hyperplanes not necessarily passing through the origin can be enforced by setting $b = 0$ in (5.1) [152]. As shown in (5.1), in our approach the parameters involved in the calculation of the input function are k-dimensional, where k is the number of classes, independently of the size of the training set. This brings a significant advantage from a computational complexity viewpoint since, for some conventional kernel functions such as the Gaussian RBF or polynomial, the size of the kernel depends on the training set n_{tr}. This generally results in a prohibitive calculation for problems with large training sets. Therefore, the proposed approach (which adopts a linear SVM after a nonlinear transformation) significantly reduces the computational complexity and allows the SVM to manage problems with large training sets.

5.4 Experimental Results

In this section, we evaluate the proposed SVM*sub* using a real hyperspectral image. A comparison with the conventional SVM, implemented with a Gaussian RBF kernel, is provided in order to evaluate

the proposed formulation with regard to a widely used one in the remote sensing community. We emphasize that our selection of the RBF kernel has been done after extensive experiments with other types of kernels. In addition, we also perform comparisons with the MLR*sub* in [51]. Although it is well-known that these methods can be combined with a spatial regularization technique such as the Markov random field [154] in order to enhance classification accuracies, we have decided to focus only on spectral-based classification in this contribution for clarity. We would also like to emphasize that, due to space considerations, we could not include a quantitative comparison to other widely used techniques for hyperspectral image classification. However, an extensive comparison of (classic and new) techniques for hyperspectral image classification has been recently presented in [21] and can be used as reference. The parameters of the different classification methods tested in this work have been carefully optimized by means of fivefold cross-validation. As for the parameters involved in the subspace estimation stage, we have also carefully optimized them following the procedure described in [51]. In all our experiments, we randomly select training samples from the reference data and report the overall, average, individual classification accuracies and the κ statistic obtained after 50 Monte Carlo runs.

In our first experiment with the AVIRIS Indian Pines scene, we analyze the performance of the SVM*sub* method under different noise conditions. Table 5.1 shows the classification results obtained by the different methods tested using only 320 training samples (20 samples per class). It should be noted that this is a very limited number of labeled samples, which we have set on purpose in order to address the fact that it is very common in practical scenarios that limited training sets are available. As shown by Table 5.1, the proposed SVM*sub* obtained the best classification results in both scenarios (i.e., with all the bands and with the noisy bands removed). Furthermore, MLR*sub* is also quite robust in noisy conditions, while the classic SVM appears to be more sensitive to noise. In addition, it is worth noting that the results obtained by subspace-based classifiers are comparable (or even slightly better) in the case in which all spectral bands (including noisy ones) are considered. This is because, even with noisy bands, the class dependent subspace can be better estimated as the dimensionality increases. This is an important observation, which reveals that subspace-based techniques are important in order to fully exploit the information present in the original hyperspectral data.

For illustrative purposes, Fig. 5.1 shows some of the obtained classification maps provided by the different tested methods for the AVIRIS Indian Pines scene. These maps correspond to one of the 50 Monte Carlo runs conducted in each case. As shown by Fig. 5.1, the classification accuracies provided by the SVM*sub* are significantly higher than those provided by the classic SVM formulation. Also, we emphasize that these results were obtained with a very limited number of training samples, which indicates that SVM*sub* can properly deal with the imbalance between the high dimensionality of the input data and the limited availability of training information. The proposed method can also deal with noise and mixed pixels, which dominate the AVIRIS Indian Pines since the agricultural features were very early in their growth cycle when the image was acquired.

In our second experiment with the AVIRIS Indian Pines scene, we evaluate the performance of the compared methods using different numbers of training samples. Here, we focus on the results obtained for the full hyperspectral image (with 220 spectral bands), as this is generally a more difficult problem due to noise and higher data dimensionality. Table 5.2 shows the overall classification accuracies (indicating the standard deviation) and the κ statistic obtained by the different methods tested, as a function of the number of training samples. The computational cost, including both the training and testing time, is also reported. In our experiments, we approximately used the same number of training samples per class (except for those classes which are very small). As shown by Table 5.2, the proposed SVM*sub* obtained better results than the other tested methods in all cases. For instance, when a total of 560 labeled samples

Table 5.1: Overall, average, individual class accuracies [%] and κ statistic obtained by the different tested methods for the AVIRIS Indian Pines scene. In all cases, only 320 training samples in total (20 per class) were used.

Class	# samples	Results using 220 spectral bands			Results using 200 spectral bands		
		SVM*sub*	SVM	MLR*sub*	SVM*sub*	SVM	MLR*sub*
Alfalfa	54	85.23	75.40	84.23	84.34	86.17	83.34
Corn-no till	1434	55.78	32.35	54.34	54.42	47.22	53.25
Corn-min till	834	54.45	39.18	59.92	57.32	50.42	59.47
Corn	234	75.94	66.55	74.95	73.39	71.36	73.03
Grass/pasture	497	83.21	69.94	83.91	84.02	79.54	83.68
Grass/tree	747	90.97	71.10	91.86	90.97	80.03	90.57
Grass/pasture-mowed	26	89.87	86.87	91.53	90.20	89.53	92.20
Hay-windrowed	489	93.07	77.59	94.69	91.78	89.11	93.36
Oats	20	95.76	65.02	98.60	94.96	87.51	95.78
Soybeans-no till	968	58.45	51.00	60.27	56.88	64.74	56.82
Soybeans-min till	2468	55.77	45.97	47.35	54.45	53.86	45.20
Soybeans-clean till	614	72.67	46.25	67.50	72.61	52.83	67.55
Wheat	212	98.88	91.22	99.65	98.95	93.91	99.58
Woods	1294	87.52	69.90	95.19	87.07	80.55	92.06
Bldg-Grass-Tree-Drives	380	63.84	45.22	30.83	61.94	52.30	37.22
Stone-steel towers	95	88.75	93.13	92.11	89.68	94.50	91.82
Overall accuracy		68.47	53.22	66.51	67.84	63.29	65.09
Average accuracy		78.13	64.17	76.68	77.69	73.35	75.93
κ statistic		64.39	47.69	62.31	63.72	58.76	60.78

Results using 220 spectral bands:

Reference map   SVM*sub* (66.73%)   SVM (54.78%)   MLR*sub* (64.45%)

Results using 200 spectral bands:

SVM*sub* (67.02%)   SVM (64.80%)   MLR*sub* (64.94%)
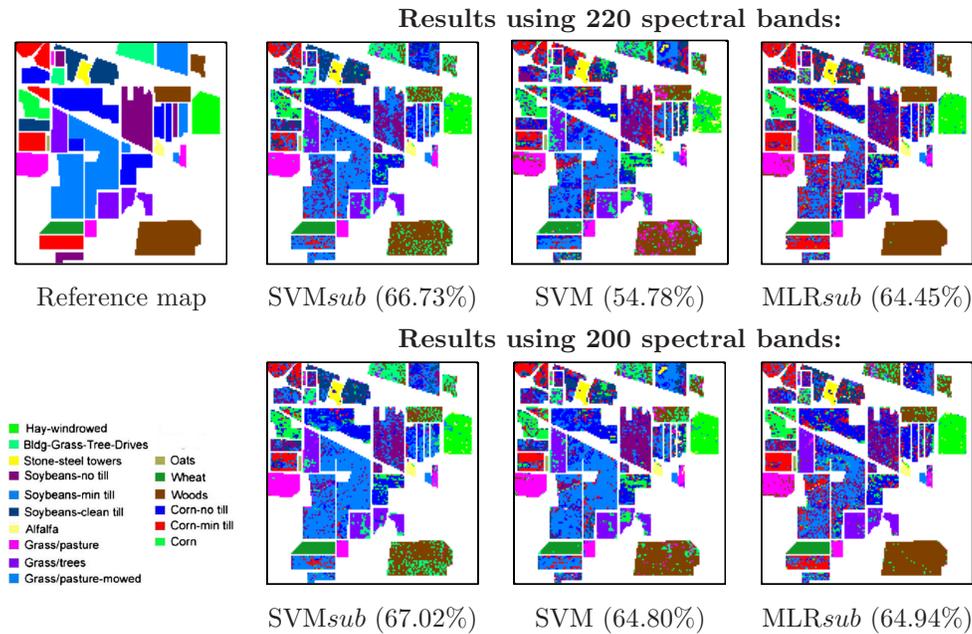
Figure 5.1: Classification maps obtained by the different tested methods for the AVIRIS Indian Pines scene. In all cases, only 320 training samples in total (20 per class) were used. The overall classification accuracies are given in the parentheses.

were used (approximately 35 samples per class), the proposed SVM*sub* obtained an overall accuracy of 78.92%, which is 19.23% higher than the one obtained by the traditional SVM and 8.94% higher than the one obtained by the MLR*sub*. Another important observation is that the subspace based methods, *i.e.*, SVM*sub* and MLR*sub*, are very fast even using a large number of training samples, while the conventional SVM method needs much more time when the number of training samples increases.

5.5 Summary and future directions

In this chapter, we have developed a new subspace-based SVM classifier called SVM*sub*. The main innovation of this classifier is that, in the construction of the SVM nonlinear function, we learn the subspace associated to each class. This formulation allows us to better cope with several phenomena that are quite important in hyperspectral image classification, such as the imbalance between the (high) dimensionality of the input data and the (limited) availability of training samples, as well as with the presence of noise and mixed pixels in the input data. The proposed method has been compared with the classic SVM formulation and also with a previously developed subspace-based technique based on the MLR classifier (MLR*sub*), obtaining good classification results with very limited training samples. In future work, we will explore the impact of including spatial information in the proposed formulation and also test the method under different conditions and analysis case studies.

Table 5.2: Overall classification accuracies [%] (plus/minus the standard deviation) and κ statistic (in the parentheses) obtained by the different tested methods for the AVIRIS Indian Pines scene, using different numbers of training samples. Computational cost (including both the training time and testing time) is also included. Both the total number of samples used and the (approximate) number of training samples per class are given (in the parentheses).

# Samples (per class)		Classification methods tested		
		SVMsub	SVM	MLRsub
160 (10)	Accuracy	61.16±2.82 (56.54)	39.23±3.91 (33.08)	61.37±2.52 (56.76)
	Time (seconds)	0.73	2.67	2.54
240 (15)	Accuracy	65.74±2.20 (61.45)	44.74±4.30 (39.15)	64.61±2.30 (60.24)
	Time (seconds)	0.77	3.44	3.27
320 (20)	Accuracy	68.47±2.06 (63.83)	53.22±2.15 (47.86)	66.51±1.85 (62.42)
	Time (seconds)	0.79	4.40	3.08
400 (25)	Accuracy	70.28±2.10 (66.36)	56.24±1.67 (50.89)	67.62±1.72 (63.46)
	Time (seconds)	0.81	5.51	3.30
480 (30)	Accuracy	72.01±1.57 (68.22)	59.03±1.73 (53.88)	69.29±1.81 (65.28)
	Time (seconds)	0.83	6.78	3.50
560 (35)	Accuracy	72.80±1.32 (69.08)	60.69±1.40 (55.64)	69.98±1.41 (66.02)
	Time (seconds)	0.85	8.07	3.82
640 (40)	Accuracy	73.69±1.66 (70.06)	62.05±1.00 (57.15)	71.04±1.35 (67.11)
	Time (seconds)	0.89	9.57	3.90
720 (45)	Accuracy	74.00±1.46 (70.37)	63.25±1.16 (58.44)	71.65±1.49 (67.73)
	Time (seconds)	0.91	11.53	4.22
800 (50)	Accuracy	74.36±1.58 (70.77)	64.16±0.96 (59.44)	71.47±1.34 (67.57)
	Time (seconds)	0.96	14.24	4.48

Chapter 6

A Discontinuity Preserving Relaxation scheme for Spectral-Spatial Hyperspectral Classification

6.1 Summary

In remote sensing image processing, relaxation is defined as a method that uses the local relationship among neighboring pixels to correct spectral or spatial distortions. In recent years, relaxation methods have shown great success in classification of remotely sensed data. Relaxation, as a preprocessing step, can reduce noise and improve the class separability in the spectral domain. On the other hand, relaxation as a post processing approach, works on the label image or class probabilities obtained from pixel-wise classifiers. In this work, we develop a discontinuity preserving relaxation strategy, which can be used for postprocessing of class probability estimates, as well as preprocessing of the original hyperspectral image. The newly proposed method is an iterative relaxation procedure which exploits spatial information in such a way that it considers discontinuities existing in the data cube. Our experimental results indicate that the proposed methodology leads to state-of-the-art classification results when combined with probabilistic classifiers for several widely used hyperspectral data sets, even when very limited training samples are available [1].

6.2 Introduction

Remotely sensed hyperspectral image classification has been a very active area of research in recent years [21]. Although techniques for unsupervised classification and/or clustering have also been used in the literature [12], supervised classification has been more popular in many applications [11]. Still, there are several important challenges when performing supervised hyperspectral image classification [59], such as the unbalance between high dimensionality and limited training samples, or the presence of mixed pixels in the data (which may compromise classification results for coarse spatial resolutions). Another

---

[1]Part of this chapter has been submitted to: J. Li, M. Khodadadzadeh, A. Plaza, X. Jia and J. M. Bioucas-Dias. **A Discontinuity Preserving Relaxation scheme for Spectral-Spatial Hyperspectral Image Classification**, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015 [JCR(2013)=2.827].

relevant challenge is the need to integrate the spatial and spectral information to take advantage of the complementarities that both sources of information can provide. Such integration can reduce the negative impact of the aforementioned challenges.

According to the principle that, in remote sensing images, neighboring pixels are likely to have the same contextual properties, spectral-spatial techniques can be effectively exploited to improve the classification accuracy [12]. For example in [87], simply adding mean of neighboring pixel values for each band to the original spectral feature vector of central pixel has been shown better classification performance than conventional spectral methods. In [89], it is proposed to extract textural features from the hyperspectral image using efficient image enhancement algorithms and then combine them with spectral information via kernels in a semisupervised graph-based framework for classification. In other approaches, modelling different kinds of the structural information contained in hyperspectral images by using morphological filters and integrating with spectral information have been successfully used for hyperspectral image classification [155, 81, 90].

The important category of spectral-spatial techniques comprises relaxation methods which are defined as methods that use the local relationship among neighboring pixels to correct spectral or spatial distortions. As preprocessing, spatial smoothing over the hyperspectral data can remove noise and enhance spatial texture information [124, 156, 157]. For example in [156], in order to classify land cover mathematical morphology based noise reduction filter has been used before the ML classification algorithm. In [124], authors showed that anisotropic diffusion algorithm can reduce the spatial and spectral variability of the image, while preserving the edges of objects, which will improve the classification accuracy of hyperspectral imagery. On the other hand, as a postprocessing method, relaxation-based approaches can be an effective tool to improve classification accuracies [12]. These normally iterative methods are broadly referred to as CR or PR [158, 94, 138, 159], which incorporate spatial-contextual information into the obtained probabilistic classification results. In other words, after a probabilistic pixel-wise classification of the hyperspectral image, the process of PR is applied to exploit the continuity, in probability sense, of neighboring labels. Perhaps the most popular PR strategy is based on the use of MRFs [98, 97, 12, 99]. Specifically, the MRF has been shown to be a very successful technique for refining the classification results provided by classifiers such as the probabilistic SVM [101, 102, 105] or, more recently, the MLR [42, 43, 51]. However, one of the first approaches to include spatial-contextual information in probabilistic classification was PLR [93, 95, 12]. PLR strategies use the probabilities of neighboring pixels iteratively to update the class probabilities for the center pixel based on a neighborhood function [12].

It has been observed that, quite often, the use of spatial information as relaxation, although, on one hand, it clearly improves the classification accuracy in smooth image areas, on the other hand, it degrades the classification performance in the neighborhood of the class boundaries. Fundamentally, this is a consequence of enforcing smoothness across the boundaries. Based on this observation, in this work, we develop a new relaxation strategy for hyperspectral image classification which aims at introducing spatial relaxation while, at the same time, accurately preserving the edges of class boundaries. This edge preserving strategy relies on discontinuity maps estimated from the original image cube. These maps are accurate because they are inferred from the many image bands, usually on the order of hundreds, with aligned discontinuities.

The proposed approach can be also used as a preprocessing step to logically relax the original spectral vectors by considering discontinuities from the data cube. This step is able to reduce noise and improve the class separability while preserving discontinuities by including edge information. However, as a postprocessing, the proposed approach is therefore based on the main principles of PLR-based
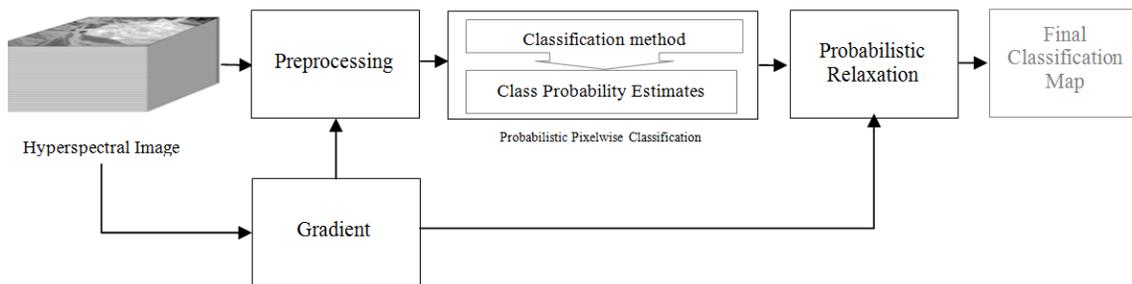
Figure 6.1: Flowchart of the proposed method.

methods, which can be considered as a form of probabilistic relaxation since they iteratively improve the probabilistic output of the considered classifier by naturally imposing spatial consistency in the final classified image. This is important, as some spatial post-processing strategies tend to generate an undesired blob-like effect in the final classification results. In this regard, our experimental results indicate that the proposed methodology leads to state-of-the-art classification results when compared with other widely used PR-based methods (e.g. PLR and MRF). The probabilistic outputs and the fact that the presented method does not require prior information about the scene are other important features of the proposed approach.

The remainder of the chapter is organized as follows. Section **6.3** describes the main stages of the proposed classification framework, including preprocessing, classification and edge-preserving probability relaxation. Section **6.4** presents an experimental validation of the method, conducted using three well-known hyperspectral data sets collected by the AVIRIS over the Indian Pines, Indiana, and Salinas Valley, California, and by the ROSIS over the city of Pavia, Italy. Section **6.5** concludes the chapter with some remarks and hints at plausible future research.

6.3   Proposed Framework

In this section, we first present probabilistic pixelwise classification method which is applied in this work and then we describe the proposed relaxation approach which is used in both preprocessing and postprocessing methods. The flowchart of proposed method has been shown in Fig. **6.1**.

6.3.1   Probabilistic Pixel-Wise Classification

Various probabilistic classification techniques have been successfully used for hyperspectral data [91, 27, 51]. In this work, we consider an MLR algorithm. MLR-based techniques exhibit the advantage of modelling directly the posterior class distributions. In this context, the densities $p(y_i|\mathbf{x}_i)$ can be modelled by the MLR, which corresponds to a discriminative model of the discriminative-generative pair for $p(\mathbf{x}_i|y_i)$ (Gaussian) and $p(y_i)$ (multinomial). In section **2.3.1.2**, the MLR algorithm has been explained in detail.

Recently, [51] proposed to combine the MLR with a subspace projection method called MLR*sub*. The idea of applying subspace projection methods to improve classification relies on the basic assumption that the samples within each class can approximately lie in a lower dimensional subspace. Thus, each class may be represented by a subspace spanned by a set of basis vectors, while the classification criterion for a new input sample is the distance from the class subspace [51]. In chapter **3**, a modified version of

MLRsub was proposed, which uses the input function $\mathbf{h}(\mathbf{x}_i) = [\|\mathbf{x}_i\|^2, \ \|\mathbf{x}_i^T \mathbf{U}^{(1)}\|^2, \ \ldots, \ \|\mathbf{x}_i^T \mathbf{U}^{(k)}\|^2]^T$. The fact that hyperspectral vectors tend to live in unions of subspaces, underlies this input function. In the following, we simply refer to the MLRsub classifier (see chapter 3) adopted in this work as MLR for simplicity.

6.3.2 Discontinuity Preserving Relaxation

In this work, we introduce a new relaxation method to logically smooth the classification results or the original hyperspectral image using both spatial and spectral information while preserving the discontinuities extracted from the data cube.

Let $\mathbf{p} = [\mathbf{p}_1, \ldots, \mathbf{p}_n] \in \mathbb{R}^{k \times n}$, $\mathbf{p}_i = [p_i(1), \ldots, p_i(k)]^T$ for $i \in \mathcal{S}$ be the k-dimensional multivariate vector of probabilities defined on site i. Let $\mathbf{u} = [\mathbf{u}_1, \ldots, \mathbf{u}_n] \in \mathbb{R}^{n \times k}$, for $i \in \mathcal{S}$, $\mathbf{u}_i = [u_i(1), \ldots, u_i(k)]^T$ be the final vectors of probabilities obtained from the relaxation process. In this work, we implement a relaxation scheme that is the solution of the following optimization problem

$$
\begin{aligned}
&\min_{\mathbf{u}} \quad (1-\lambda)\|\mathbf{u} - \mathbf{p}\|^2 + \lambda \sum_i \sum_{j \in \partial_i} \varepsilon_j \|\mathbf{u}_j - \mathbf{u}_i\|^2 \\
&\text{s.t.:} \quad \mathbf{u}_i \geq 0, \ \ \mathbf{1}^T \mathbf{u}_i = 1,
\end{aligned}
\tag{6.1}
$$

where the constraints are justified by the fact that the vectors $\mathbf{u}_i$ represent probabilities $\mathbf{1}$ is a vector column of k 1s, λ ($0 \leq \lambda \leq 1$) is a weight parameter controlling the relative impact of the both terms in the objective function, ∂_i denotes the 8-neighborhood of pixel i (other types of neighborhood can be applied), and ε_j is a value in the site $j \in S$ of edge image ε given by:

$$
\varepsilon = \exp\left(-\sum_{i=1}^{d} \mathrm{sobel}(\mathbf{X}^{(i)})\right),
\tag{6.2}
$$

where sobel() denotes the Sobel filter, which detects the discontinuities in an image and the output at each pixel is 0 or 1. The Sobel filter is applied on each spectral channel in a specific direction and $\mathbf{X}^{(i)}$ denotes the ith band of the original data cube $\mathbf{X}$. Note that here, to have a better interpretation of the edges, we considered the average of the results obtained by applying sobel() in two vertical and horizontal directions.

In the proposed relaxation scheme (6.1), the fist term in the objective function measures the data misfit and the second term promotes smooth solutions weighted by the parameter ε_j, which, according to its definition, is large when there are no discontinuities between the neighboring pixels it connects and small when there are discontinuities. The solution of (6.1) corresponds therefore to tradeoff between adjustment to the "noisy" classification, imposed by the first term, and smoothness imposed by the second term. We stress, however, that due to the presence of map ε, the smoothness is not applied across the discontinuities.

At this point we would like to make reference to edge preserving image restoration methods such as those based on TV [159] or based on AD [160]. In both cases (i.e. TV and AD) the objective is similar to ours: apply strong smoothing in areas away from edges and avoid smoothing the edges. However, in our case we know the edges in advance, which is not the case of those methods. This is a considerable advantage, which results from the availability of many hyperspectral bands.

Problem (6.1) is strictly convex and therefore is has a unique solution. Herein, we apply a projected iterative Gauss Seidel scheme which consists in iteratively minimizing the objective function in (6.1) with

respect to each optimization variable $u_i(k)$ and, after a complete sweep, project on the probabilities at each pixel onto the probability simplex. The obtained algorithm is shown in Algorithm 2, where *iters* is the number of maximum iterations defined in advance, $Err^{(t+1)} = \frac{\|\mathbf{u}^{t+1} - \mathbf{u}^t\|}{\|\mathbf{u}^t\|}$ is an error parameter and τ is the error threshold parameter controlling the degree of convergence.

Algorithm 2 Discontinuity Preserving Relaxation (Probabilistic Relaxation)

Input: p, ε, λ, *iters*, $Err^{(1)} = \|\mathbf{p}\|$, τ
Output: u
$t := 1$
while $Err^{(t+1)} - Err^{(t)} \leq \tau$ **or** $t \leq$ *iters* **do**
   for $c := 1$ to k **do**
      $u_i^{(t+1)}(c) = \frac{(1-\lambda)p_i(c) + \lambda \sum_{j \in \partial_i} \varepsilon_j u_j^{(t)}(c)}{(1-\lambda) + \lambda \sum_{j \in \partial_i} \varepsilon_j}$
   end for
   $u_i^{(t+1)} = u_i^{(t+1)} / \sum_{c=1}^{k} u_i^{(t+1)}(c)$
   $Err^{(t+1)} = \frac{\|\mathbf{u}^{t+1} - \mathbf{u}^t\|}{\|\mathbf{u}^t\|}$
end while

At this point, we would like to call attention to the fact that, apart from the constraints used in (6.1) linked to the fact that we estimating probabilities, the ratione used to carry out PR can be used to denoise the original bands of the hyperspectral image $\mathbf{X}$ ensuring the preservation of the discontinuities. The correspondent algorithm, which may be used as a preprocessing step, is shown in Algorithm 3, where $Err_{:m}$ denotes the error parameter for the mth band, $\tilde{\mathbf{x}}_{:m}$ is the processed image of the mth band which corresponds to the original mth band, *i.e.*, $\mathbf{x}_{:m} = [x_{1m}, \ldots, x_{nm}]$. Finally, we empirically find out that both algorithms converge very fast, say, less than 20 iterations.

Algorithm 3 Discontinuity Preserving Relaxation (Preprocessing)

Input: X, ε, λ, *iters*, $Err^{(1)} = \|\mathbf{X}\|$, τ
Output: $\tilde{\mathbf{X}}$
for $m := 1$ to d **do**
   $t := 1$
   $Err_{:m}^{(1)} = Err^{(1)}$
   while $Err_{:m}^{(t+1)} - Err_{:m}^{(t)} \leq \tau$, **or** $t \leq$ *iters* **do**
      $\tilde{x}_{im}^{(t+1)} = \frac{(1-\lambda)x_{im} + \lambda \sum_{j \in \partial_i} \varepsilon_j \tilde{x}_{jm}^{(t)}}{(1-\lambda) + \lambda \sum_{j \in \partial_i} \varepsilon_j}$
      $Err_{:m}^{(t+1)} = \frac{\|\tilde{\mathbf{x}}_{:m}^{t+1} - \tilde{\mathbf{x}}_{:m}^t\|}{\|\tilde{\mathbf{x}}_{:m}^t\|}$
   end while
end for

6.4   Experimental Results and Discussion

In this section, we use both simulated and real hyperspectral data to evaluate the proposed approach. The main goal of using simulated data set is to evaluate the performance of the algorithm in a fully controlled environment, while the experiments with real experiments are intended to provide a quantitative evaluation of the method in real analysis scenarios. For simplicity, in this section we refer to spatial preprocessing as "pp", while "MLL" and "pr" denote MLL-based [110, 161] and PR-based spatial relaxation, respectively.

6.4.1   Experiments with Simulated Data

In our first experiment, we use a simulated image with eight classes and 100×100 pixels, in which the spatial distribution is extracted from a real image and the spectral signatures are selected from the USGS digital spectral library[2]. The ground truth image and the spectral signatures of eight randomly selected mineral signatures allocated to the main classes are shown in Fig. 6.2. We considered the following linear mixture model for generating a simulated mixed pixel:

$$\mathbf{x}_i = \sum_{j \in \partial_i} \mathbf{m}^{(j)} \gamma_j + \mathbf{n}_i, \tag{6.3}$$

where, $\mathbf{m}^{(l)}, l = 1, ..., 8$ are spectral signatures obtained randomly from the USGS spectral library, and γ_j, which follows a random distribution with $0 \leq \gamma_j \leq 1$ and $\sum_{j \in \partial_i} \gamma_j = 1$, determines the abundance of the signatures which contribute to the mixture model. Note that here, the maximum abundance value of γ_j is assigned to the objective class according to the groundtruth image. ∂_i is a neighborhood with a specific size around the central pixel i over the considered ground truth image. ∂_i determines a set of class labels to contribute in the mixture. So that the pixels near the borders of the regions are generated by mixtures of different class labels and the pixels far from the borders are considered pure. In our simulations we set the size of the neighborhood to 9×9 pixels. For illustrative purposes, Fig. 6.2(c) shows an example of the abundance maps associated to the eight classes of the simulation image. In each pixel of the scene, the fractional abundances vary from 0% (black color) to 100% (white color) and sum to unity. Note that, using the suggested procedure, signature abundance is not constant over class regions and the pixels closer to the discontinuities are more heavily mixed, as expected in real scenarios. Zero-mean Gaussian noise with covariance $\sigma^2 \mathbf{I}$ i.e. $\mathbf{n}_i \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ is finally added to the generated synthetic image. For each class, we randomly chose 10 samples (in total 80 samples) from the ground truth image in Fig. 6.2(a) for training purposes.

We have conducted different experiments with the simulated hyperspectral image described earlier. These experiments have been carefully designed in order to analyze several relevant aspects of our proposed method in a fully controlled environment. All of the results reported in this chapter with the simulated data sets were obtained after 30 Monte Carlo runs in which we randomly select 8 different materials and also randomly select different training sets.

6.4.2   Impact of Parameter λ:

In our first experiment, we analyze the impact of the tunable parameter λ intended to control the relative impact of the both terms in the proposed relaxation scheme. It should be noted that, if $\lambda = 0$, only the first term is considered and the method remains as the original MLR algorithm. If $\lambda = 1$, only the smoothing term is used. Fig. 6.3(a) plots the obtained OA results as a function of λ, with $\sigma = 0.1$ and the maximum number of iterations as 20. From Fig. 6.3(a), we can conclude that the relaxation performance indeed depends on the setting of λ. However, even with $0.7 \leq \lambda \leq 0.9$, the proposed relaxation method leads to significant classification results for the considered problem. Fig. 6.3(b) shows convergence of the proposed PR method with different values of λ parameter. As can be observed, the proposed approach converged very fast, *i.e.*, less than 20 iterations, for all cases with different value of λ. Hence, in this chapter, we set the parameter $\lambda = 0.9$ and the maximum number of iterations as 20 for the remaining simulated experiments.

---

[2]http://speclab.cr.usgs.gov/spectral-lib.html

(a)

(b)

*class*1

*class*2

*class*3

*class*4

*class*5

*class*6

*class*7

*class*8

(c)
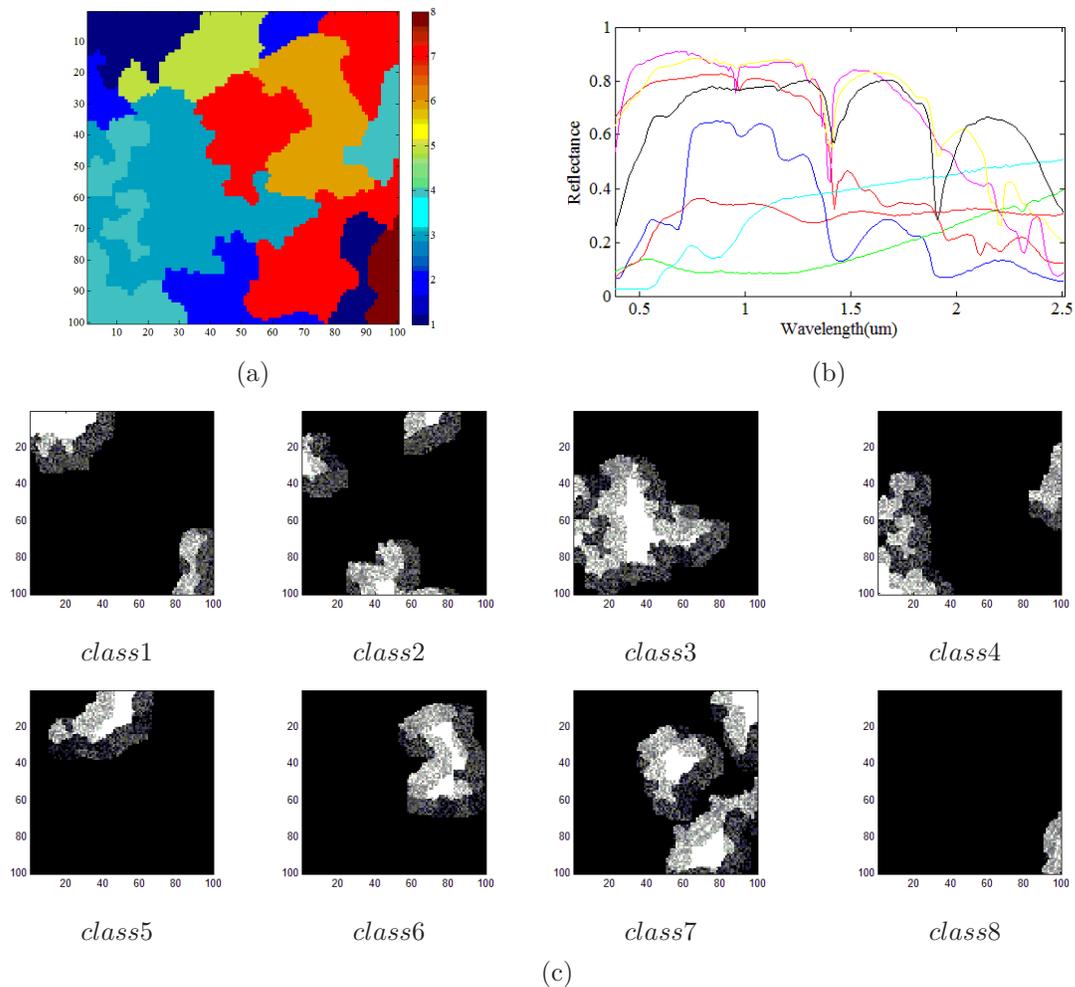
Figure 6.2: (a) Classes in a synthetic scene with $n = 100 \times 100$. (b) Spectral signatures of randomly selected materials from the USGS digital library used in the simulation (c) Fractional abundance distributions considered for generating mixed pixels using a fixed window of size 9×9 pixels.

6.4.3   Impact of Noise:

In the other experiment with simulated data, we evaluate the impact of noise on the proposed relaxation method. Table 6.1 shows the classification results obtained by the proposed approach using different values of noise standard deviation σ. Several conclusions can be obtained from Table 6.1. First and foremost, it is remarkable that the proposed approach, which carefully uses the local relationship among neighboring pixels, has been improved the performance of MLR-based classification accuracy. Clearly, the performance of the proposed relaxation method decreases as σ increases. When the noise is low, using the proposed method as probabilistic relaxation shows better performance than preprocessing, however, in high noise images, relaxation method as preprocessing shows significant improvements. Note that, the results obtained using both preprocessing and probabilistic relaxation i.e. ppMLRpr are always superior.

(a)   (b)
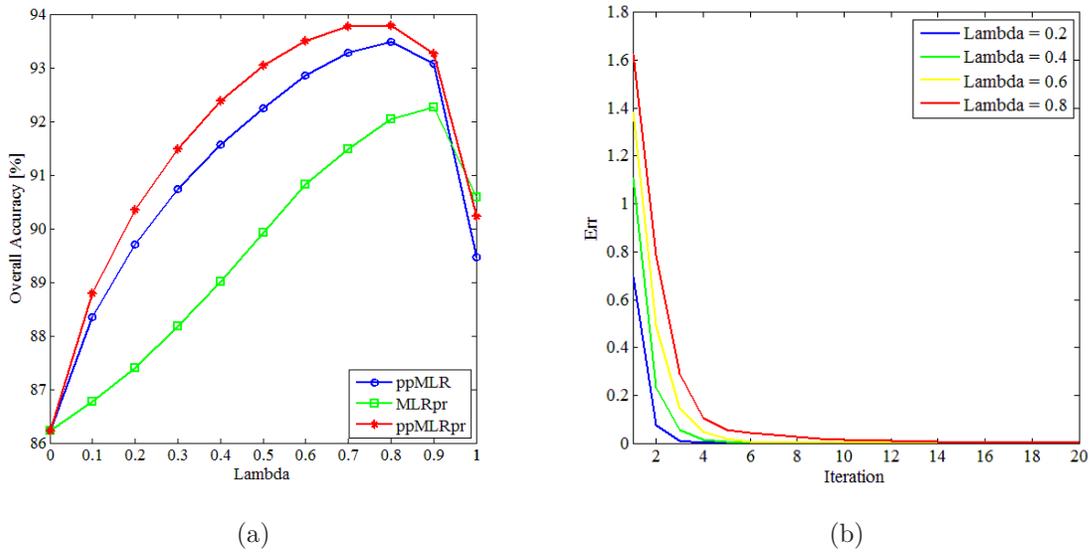
Figure 6.3: Impact of parameter λ.

Table 6.1: Overall (OA) and average (AA) classification accuracies [%] (as a function of parameter σ) noise

Methods	Accuracies	σ				
		0	0.05	0.10	0.15	0.20
MLR	OA	92.63	91.09	83.02	69.92	56.16
	AA	93.77	92.09	85.07	72.71	59.74
ppMLR	OA	93.08	95.50	91.50	79.19	65.74
	AA	94.09	96.05	92.46	81.17	68.49
MLRpr	OA	95.70	95.77	87.76	72.69	57.82
	AA	96.23	96.19	89.57	75.63	61.64
ppMLRpr	OA	94.57	95.90	91.80	79.62	65.90
	AA	95.43	96.44	92.76	81.76	68.94

6.4.4 Real experiments

Three different hyperspectral images were used for the experiments: AVIRIS Indian Pines image described in subsection 2.5.1.2, AVIRIS Salinas image described in subsection 3.4.2 and ROSIS Pavia University image described in subsection 2.5.1.2. These data sets have different characteristics and contexts (two agricultural areas and an urban area, with different spectral and spatial resolutions).

Moreover, for the three considered hyperspectral images, the discontinuities maps were generated using equation (6.2) which have been shown in Figs 6.4.

6.4.4.1 Experimental setup

Before describing our results, it is first important to report the parameters and main considerations in our experiments. For the experiments with the AVIRIS Indian Pines and Salinas images, the training

(a) AVIRIS Indian Pines   (b) AVIRIS Salinas   (c) ROSIS Pavia University
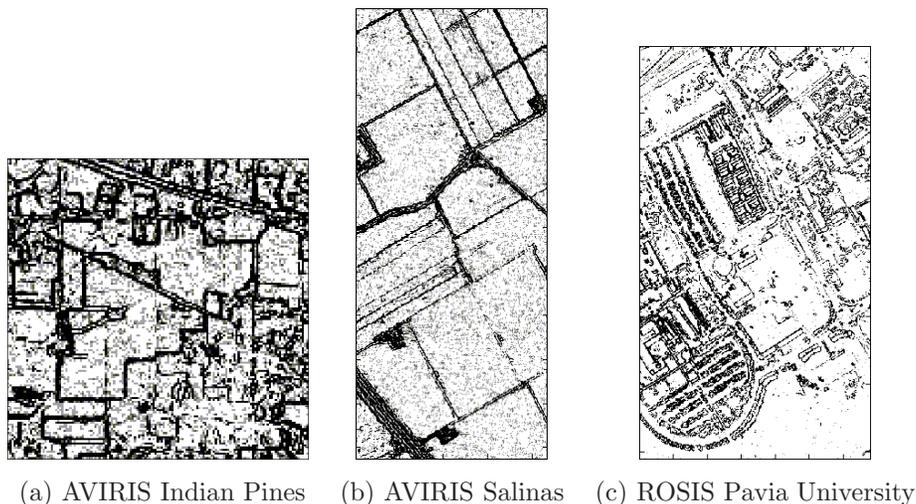
Figure 6.4: Discontinuity maps for the considered data sets.

samples were randomly selected from the available ground truth and the remaining samples are used for validation. However, for the ROSIS Pavia University image, small subsets of the original training samples were used. Concerning the λ parameter of the proposed relaxation methods, we considered $\lambda = 0.9$. For the stopping, the maximum number of iterations in all experiments was set to 20. These settings, although suboptimal, lead to very good classification performance. Note that, in all the experiments, the results reported correspond to the average of the results obtained after 20 Monte Carlo runs.

6.4.4.2 Experiments for AVIRIS Images

Tables 6.2 and 6.3 report the obtained classification accuracies for the AVIRIS Indian Pines and Salinas images, respectively. The metrics reported are the individual classification accuracies, as well as the OA, AA and κ statistic. These tables provide the results for each step of the proposed spectral-spatial relaxation method. Moreover, the results have been compared with the recently proposed spectral-spatial classification method MLR*sub*MLL [51]. From the results reported in Tables 6.2 and 6.3, we can conclude that our proposed method exhibits state-of-the-art. For instance, Table 6.2 reveals that the proposed relaxation method i.e. ppMLRpr obtained an OA of 91.05% for the AVIRIS Indian Pines image, which contrasts with the OA of 64.30% obtained by the MLR-based classifier. Compared to MLRMLL the OA achieved by the presented method improved by about 16% the OA obtained by this method. For the AVIRIS Salinas image, we obtained comparable results.

A more detailed investigation of individual class accuracies is important to assess quantitatively the impact of the proposed method on class separability. As indicated on Tables 6.2 and 6.3, the improvement is quite significant for the sets of similar class labels. For example, the classification accuracies obtained by the MLR method with preprocessing for the classes: *Corn-no till*, *Corn-min till* and *Corn* in the AVIRIS Indian Pines scene were 82.49%, 86.60% and 94.94% respectively, which are 32.13%, 27.98% and 25.24% higher than those obtained by the MLR algorithm. It is also remarkable that the accuracies for these classes increased in 3.05%, 3.30% and 2.50%, respectively, when the proposed MLRpr method with preprocessing was used. The same conclusion can be obtained after comparing the individual class accuracies for the sets of {*Grass/trees*, *Grass/pasture*, *Grass/pasture-mowed*} and {*Soybeans-no till*, *Soybeans-min till*, *Soybeans-clean till*}. For the AVIRIS Salinas image, it is also possible to consider other

Table 6.2: Overall (OA) and average (AA) classification accuracies [%] obtained by different methods for the AVIRIS Indian Pines data set.

Class	Train/Test	Methods				
		MLR	MLRMLL	Proposed Relaxation Methods		
				ppMLR	MLRpr	ppMLRpr
Alfalfa	15 / 54	86.67 (4.06)	92.96 (5.29)	98.61 (1.33)	97.31 (0.95)	97.96 (1.02)
Corn-no till	15 / 1434	50.36 (7.87)	61.98 (9.16)	82.49 (6.39)	70.25 (8.89)	85.54 (6.27)
Corn-min till	15 / 834	58.62 (10.93)	70.85 (13.01)	86.60 (6.57)	74.03 (14.26)	89.90 (7.09)
Corn	15 / 234	69.70 (10.81)	90.49 (8.71)	94.94 (4.96)	95.60 (9.13)	97.44 (4.48)
Grass/trees	15 / 497	82.27 (5.68)	88.16 (5.32)	91.18 (4.04)	90.99 (7.57)	91.64 (5.11)
Grass/pasture	15 / 747	88.80 (5.06)	96.06 (2.68)	97.48 (1.15)	99.65 (0.42)	98.25 (0.75)
Grass/pasture-mowed	15 / 26	96.73 (2.87)	98.65 (2.26)	99.81 (0.86)	97.31 (4.69)	94.23 (22.34)
Hay-windrowed	15 / 489	90.26 (4.35)	97.29 (3.14)	99.16 (0.71)	99.64 (1.36)	99.98 (0.06)
Oats	15 / 20	99.75 (1.12)	100.00 (0.00)	100.00 (0.00)	43.50 (28.75)	92.00 (11.17)
Soybeans-notill	15 / 968	53.36 (9.30)	68.16 (12.31)	87.57 (6.85)	74.83 (12.18)	88.57 (5.76)
Soybeans-min till	15 / 2468	49.59 (9.57)	60.65 (10.74)	80.68 (5.78)	68.07 (14.52)	86.29 (5.72)
Soybeans-clean till	15 / 614	58.36 (10.92)	76.69 (12.36)	88.88 (5.32)	84.45 (14.14)	93.20 (6.16)
Wheat	15 / 212	97.81 (2.31)	99.46 (0.41)	99.58 (0.34)	99.98 (0.11)	99.62 (0.42)
Woods	15 / 1294	86.81 (7.63)	94.11 (7.10)	93.96 (3.72)	95.89 (7.05)	95.24 (4.56)
Bldg-Grass-Tree-Drives	15 / 380	44.30 (12.02)	57.25 (15.18)	94.99 (6.40)	72.33 (18.72)	95.96 (5.84)
Stone-Steel towers	15 / 95	92.16 (4.29)	96.58 (4.19)	98.63 (2.11)	97.95 (2.01)	96.79 (1.95)
OA		64.30 (2.29)	75.09 (2.86)	88.36 (1.67)	80.67 (3.12)	91.05 (1.87)
AA		75.35 (1.60)	84.33 (1.36)	93.41 (1.03)	85.11 (2.52)	93.91 (2.27)
κ		60.03 (2.45)	72.03 (3.10)	86.88 (1.86)	78.22 (3.36)	89.87 (2.09)

Table 6.3: Overall (OA) and average (AA) classification accuracies [%] obtained by different methods for the AVIRIS Salinas data set.

Class	Train/Test	Methods				
		MLR	MLRMLL	Proposed Relaxation Methods		
				ppMLR	MLRpr	ppMLRpr
Brocoli-green-weeds-1	15 / 2009	99.31 (0.55)	99.93 (0.40)	99.88 (0.27)	99.98 (0.13)	99.90 (0.24)
Brocol-green-weeds-2	15 / 3726	98.42 (1.14)	99.12 (0.09)	99.41 (1.18)	99.79 (0.37)	96.67 (15.90)
Fallow	15 / 1976	91.81 (6.20)	94.21 (5.43)	98.85 (2.85)	95.65 (6.10)	96.08 (18.27)
Fallow-rough-plow	15 / 1394	98.35 (2.99)	98.70 (2.75)	99.79 (0.32)	99.94 (0.13)	98.90 (5.62)
Fallow-smooth	15 / 2678	95.88 (2.70)	98.71 (1.53)	97.93 (1.00)	98.72 (0.22)	97.42 (6.96)
Stubble	15 / 3959	98.62 (1.08)	99.31 (0.37)	99.69 (0.36)	99.82 (0.22)	98.49 (7.43)
Celery	15 / 3579	98.78 (0.62)	99.22 (0.20)	99.76 (0.17)	99.91 (0.01)	98.25 (9.13)
Grapes-untrained	15 / 11271	66.18 (9.09)	74.39 (24.50)	81.76 (5.80)	77.46 (12.22)	84.68 (6.42)
Soil-vinyard-develop	15 / 6203	97.30 (0.90)	98.58 (1.04)	98.91 (0.76)	100.00 (0.01)	97.46 (9.05)
Corn-senesced-green-weeds	15 / 3278	81.72 (5.92)	85.67 (2.08)	90.69 (3.24)	88.77 (8.35)	92.17 (3.51)
Lettuce-romaine-4wk	15 / 1068	93.12 (3.61)	95.55 (1.89)	97.18 (2.80)	99.46 (0.77)	99.58 (0.52)
Lettuce-romaine-5wk	15 / 1927	97.62 (2.98)	99.15 (0.00)	98.82 (1.97)	99.60 (1.28)	98.94 (1.98)
Lettuce-romaine-6wk	15 / 916	98.70 (0.86)	99.01 (0.80)	98.73 (1.05)	98.86 (0.34)	98.73 (0.64)
Lettuce-romaine-7wk	15 / 1070	94.20 (3.01)	96.81 (3.41)	97.47 (2.25)	97.21 (1.77)	97.13 (2.25)
Vinyard-untrained	15 / 7268	63.90 (9.33)	71.53 (41.58)	87.32 (9.06)	78.01 (14.28)	90.55 (8.65)
Vinyard-vertical-trellis	15 / 1807	94.16 (2.89)	96.32 (1.64)	96.71 (3.39)	98.36 (0.78)	96.93 (3.69)
OA		85.28 (1.51)	89.02 (6.54)	93.30 (1.70)	91.26 (2.30)	93.79 (4.46)
AA		91.76 (0.76)	94.14 (2.60)	96.43 (0.81)	95.72 (1.05)	96.37 (4.48)
κ		83.67 (1.66)	87.80 (7.28)	92.56 (1.89)	90.29 (2.55)	93.11 (4.91)

MLR (64.15%) MLR-MLL (73.11 %)

ppMLR (86.66%) MLRpr (80.13 %) ppMLRpr (90.98 %)
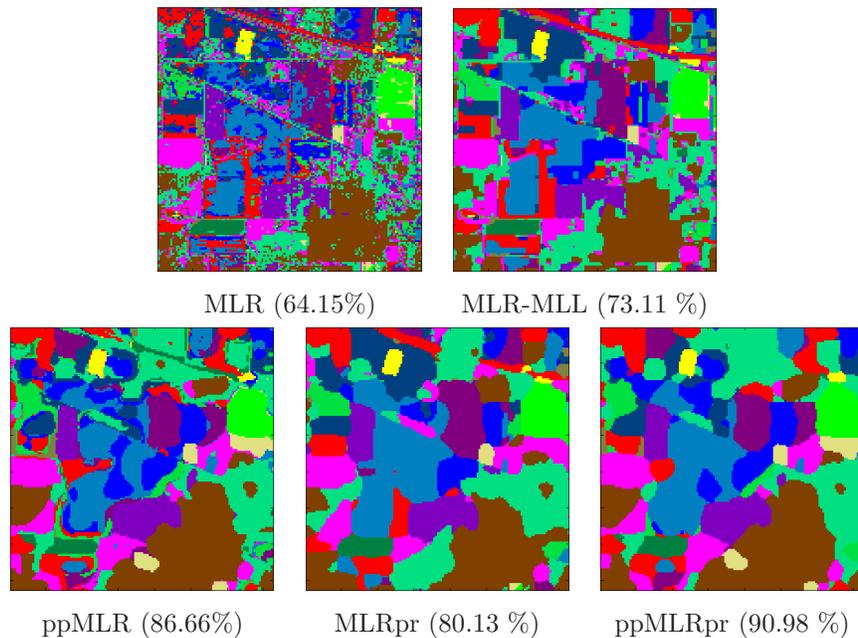
Figure 6.5: Classification maps obtained by different methods for the AVIRIS Indian Pines scene (PP refers to preprocessing and the overall accuracies are reported in the parentheses).

sets of similar classes and obtain the same conclusion. For instance, pixel-wise classifier MLR obtained low accuracies for class *Vinyard-untrained*, i.e. 63.90%. However, after applying the preprocessing method the accuracies for this class were increased by 23.42%. This improvement is significant because, for example, the MLR+MLL method obtained 71.53% accuracy for this class, which is just 7.63% higher than MLR result. It is also noticeable that the accuracy obtained by the proposed method ppMLRpr for the class *Vinyard-untrained* is 90.55%, which is 26.65% higher than the result obtained by the MLR algorithm.

For illustrative purposes, Figs. 6.5 and 6.6 show the obtained classification maps for the AVIRIS Indian Pines and Salinas data sets. Each of the maps corresponds to one out of the 30 Monte Carlo experiments which were averaged to produce the results reported in Tables 6.2 and 6.3. From Figs. 6.5 and 6.6, it can be seen that using spatial information (both at the preprocessing and postprocessing level) can lead to more homogeneous regions in classification maps, when compared to the pixel-wise classification maps. Most importantly, the proposed method exhibits very good performance in the task of delineating the borders of classes of interest.

Figs 6.7 and 6.8 illustrate the performance of the proposed relaxation method in detailed. For example Fig 6.7 shows the changes of probabilities of class *Soybean min-till* for the all pixels. We can conclude that our proposed method preserve discontinuities during relaxation process. Similarly for preprocessing(Fig 6.8), the proposed method obviously smooth the original hyperspectral image while it considered edge information.

6.4.5   Experiments for the ROSIS Pavia University Image

Table 6.4 details the classification results obtained for the ROSIS Pavia University scene. Several conclusions can be obtained from this table. First and foremost, it is remarkable that the proposed relaxation approach exhibited very good performance using very limited number of training samples.

MLR (86.97%)   MLR-MLL (90.67%)

ppMLR (95.87%)   MLRpr (94.38%)   ppMLRpr (96.80%)

Figure 6.6: Classification maps obtained by different methods for the AVIRIS Salinas scene (PP refers to preprocessing and the overall accuracies are reported in the parentheses).

For instance, our proposed method obtained an OA of 85.05%, which is 18.05% higher than the one obtained by the MLR algorithm, whereas the MLRMLL obtained an OA of 76.50%, which is 9.5% higher than the result obtained by the MLR algorithm. For illustrative purposes, Fig. 6.9 shows the obtained classification maps for the Pavia University data set.

6.5 Summary and future directions

In this work, we have developed a new methodology for spectral-spatial classification of remotely sensed hyperspectral scenes. The main features of our proposed approach can be summarized as follows. First, it provides spatially homogeneous regions after probabilistic classification, thus exploiting the intrinsic correlation which exists between neighboring pixels to improve the final classification results. Second, it specifically models the pixels at the borders of the regions to provide a better delineation of the classified

Table 6.4: Overall (OA) and average (AA) classification accuracies [%] obtained by different methods for the ROSIS Pavia University scene.

Class	Train/Test	Methods				
		MLR	MLRMLL	Proposed Relaxation Methods		
				ppMLR	MLRpr	ppMLRpr
Asphalt	15 / 6631	61.32 (5.93)	75.38 (7.74)	76.81 (6.95)	89.03 (6.09)	78.14 (8.38)
Meadows	15 / 18649	66.96 (9.98)	79.36 (12.78)	90.87 (5.14)	80.74 (12.94)	90.36 (6.37)
Gravel	15 / 2099	57.06 (10.19)	57.39 (17.56)	79.47 (7.42)	71.59 (18.08)	81.55 (9.32)
Trees	15 / 3064	89.64 (9.79)	91.72 (12.68)	86.11 (8.44)	84.68 (15.58)	83.49 (8.49)
Metal sheets	15 / 1345	96.02 (5.15)	98.35 (4.05)	97.17 (3.11)	97.99 (3.55)	97.73 (4.44)
Bare soil	15 / 5029	45.45 (8.64)	42.48 (18.93)	68.79 (12.56)	59.58 (17.55)	70.61 (14.96)
Bitumen	15 / 1330	78.92 (9.50)	89.55 (13.16)	81.48 (13.29)	96.42 (11.89)	80.09 (12.51)
Bricks	15 / 3682	70.67 (8.09)	90.10 (5.13)	87.65 (5.50)	92.25 (6.35)	88.88 (4.33)
Shadows	15 / 947	98.50 (2.50)	99.50 (1.26)	91.38 (6.36)	99.90 (0.11)	92.41 (5.13)
OA		67.00 (3.55)	76.50 (5.35)	84.83 (2.70)	81.81 (5.25)	85.05 (2.87)
AA		73.84 (1.95)	80.42 (3.49)	84.41 (3.00)	85.80 (4.12)	84.80 (2.69)
κ		58.23 (3.72)	69.46 (6.27)	80.06 (3.44)	76.40 (6.34)	80.36 (3.62)

The Soybean min-till class MLR (66.13 %)

ppMLR (75.77 %) MLRpr (65.72 %) ppMLRpr (81.20 %)
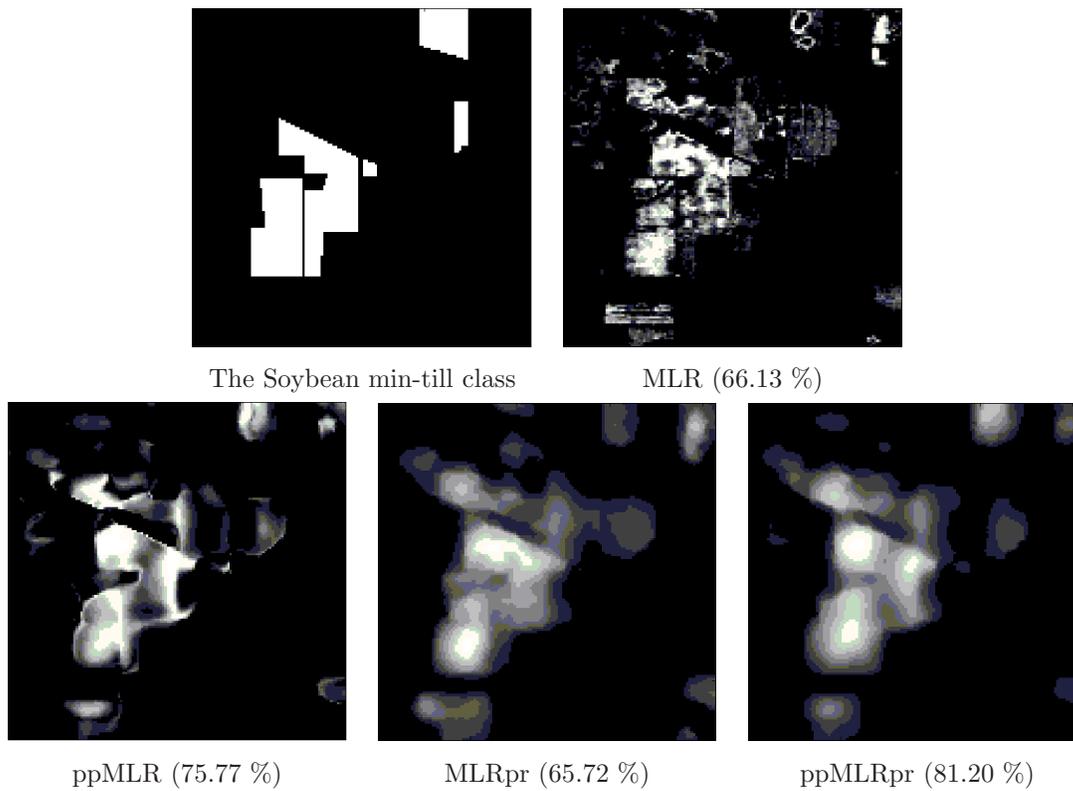
Figure 6.7: The Probability image of the class *Soybean min-till* resulted from the proposed methods

objects. In other words, our proposed approach is able to provide accurate spectral-spatial classification while preserving the edges and the boundaries between classes, which is quite important as the inclusion of spatial regularizer tends to blur the class boundaries and provide non-smooth delineations.

Our experimental results, conducted using a variety of (simulated and real) hyperspectral scenes and spectral-spatial classification strategies, indicate that the proposed approach provides state-of-the-art classification results. particularly, the proposed method provides high classification accuracies when very limited training samples are used.
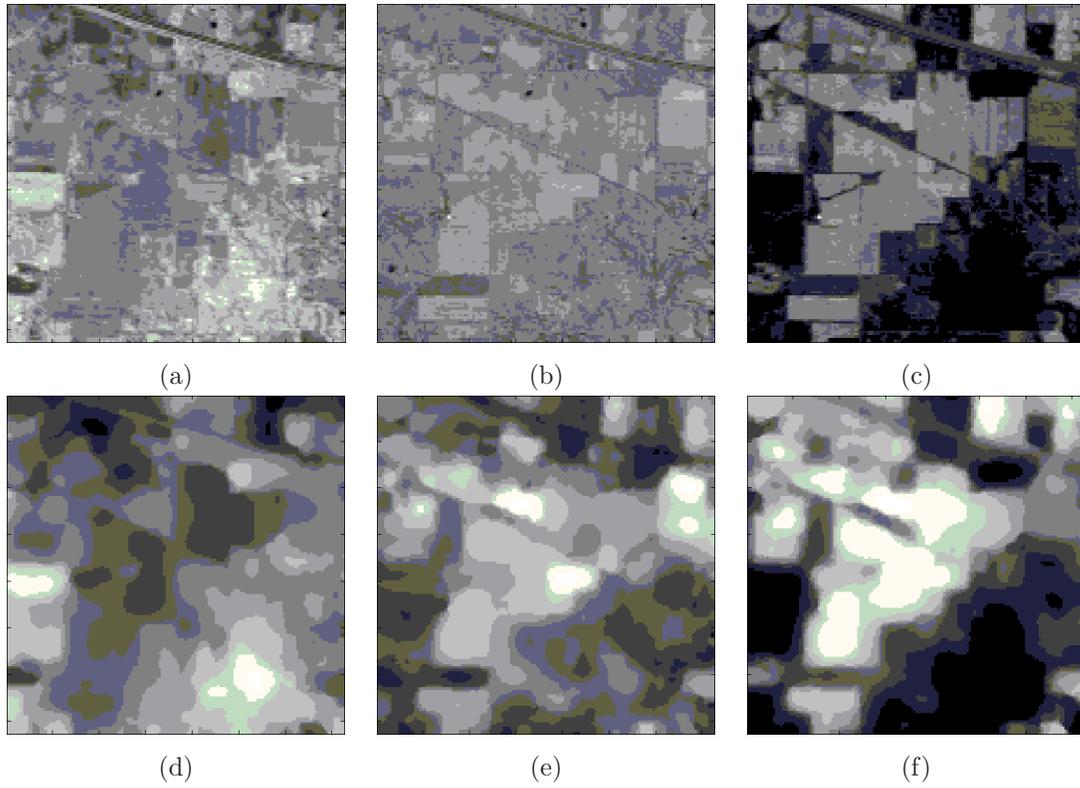
Figure 6.8: The bands of numbers 50, 100 and 150 of the original hyperspectal image before (a-c) and after (d-f) preprocessing)

MLR (65.52%)   MLR-MLL (75.54%)
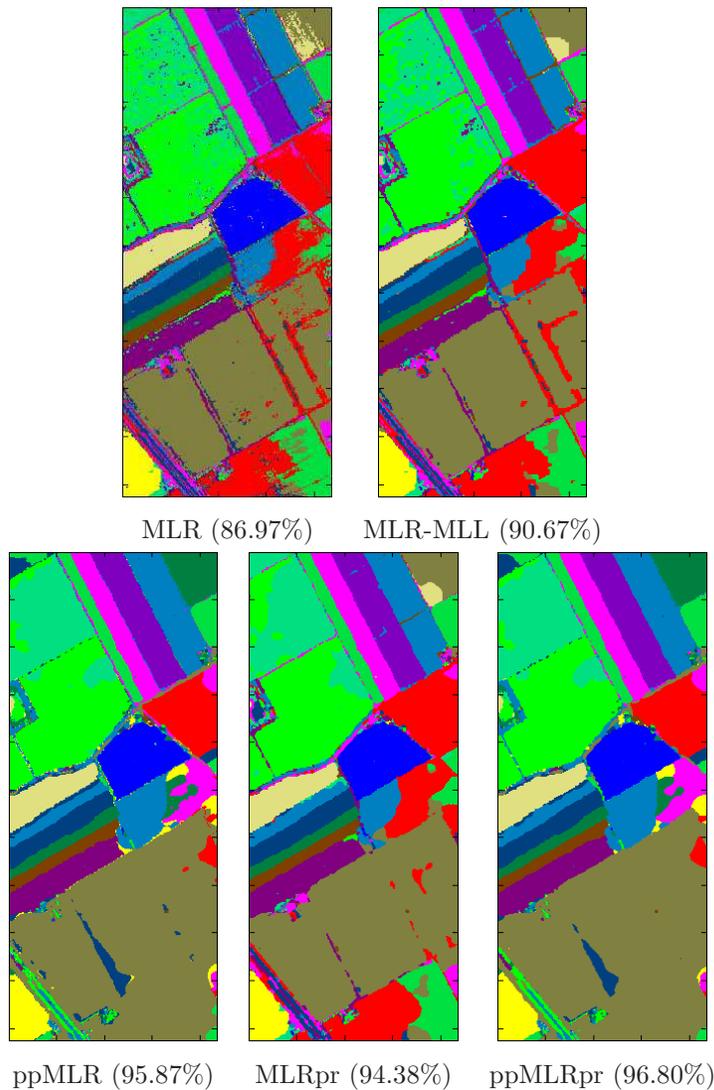
ppMLR (84.77%)   MLRpr (81.41%)   ppMLRpr (82.53%)
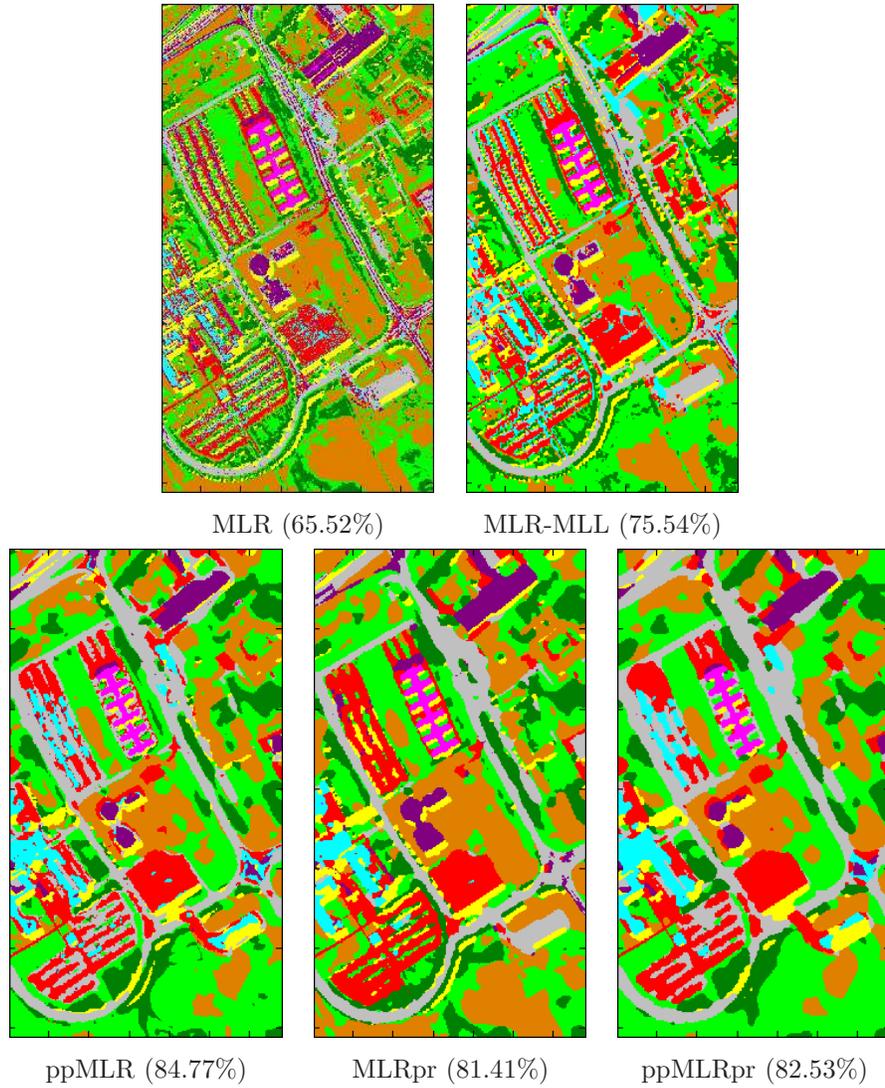
Figure 6.9: Classification maps obtained by different methods for the ROSIS Pavia University scene (PP refers to preprocessing and the overall accuracies are reported in the parentheses).

Chapter 7

Fusion of Hyperspectral and LiDAR Remote Sensing Data Using Multiple Feature Learning

7.1 Summary

Hyperspectral image classification has been an active topic of research. In recent years, it has been found that LiDAR data provides a source of complementary information that can greatly assist in the classification of hyperspectral data, in particular when it is difficult to separate complex classes. This is because, in addition to the spatial and the spectral information provided by hyperspectral data, LiDAR can provide very valuable information about the height of the surveyed area that can help with the discrimination of classes and their separability. In the past, several efforts have been investigated for fusion of hyperspectral and LiDAR data, with some efforts driven by the morphological information that can be derived from both data sources. However, a main challenge for the learning approaches is how to exploit the information coming from multiple features. Specifically, it has been found that simple concatenation or stacking of features such as morphological attribute profiles may contain redundant information. In addition, a significant increase in the number of features may lead to very high-dimensional input features. This is in contrast with the limited number of training samples often available in remote sensing applications, which may lead to the Hughes effect. In this work, we develop a new efficient strategy for fusion and classification of hyperspectral and LiDAR data. Our approach has been designed to integrate multiple types of features extracted from these data. An important characteristic of the presented approach is that it does not require any regularization parameters, so that different types of features can be efficiently exploited and integrated in a collaborative and flexible way. Our experimental results, conducted using a hyperspectral image and a LiDAR derived DSM collected over the University of Houston campus and the neighboring urban area, indicate that the proposed framework for multiple feature learning provides state-of-the-art classification results [1].

7.2   Introduction

Hyperspectral imaging is concerned with the extraction of information from objects or scenes lying on the Earth surface, using hundreds of (narrow) spectral bands typically covering the visible and near infra-red domains [21]. In hyperspectral imaging, also termed imaging spectroscopy [162], the sensor acquires a spectral vector with hundreds or thousands of elements from every pixel in a given scene. The result is the so-called hyperspectral image or hyperspectral data *cube*. It should be noted that hyperspectral images are spectrally smooth and spatially piece-wise smooth; this means that the values in neighboring locations and wavelengths are often highly correlated [91].

Hyperspectral image classification has been a very active area of research in recent years [59]. Given a set of observations (*i.e.*, pixel vectors in a hyperspectral image), the goal of classification is to assign a unique label to each pixel vector so that it is well-defined by a given class. The wider availability of hyperspectral data with high spatial resolution has been quite important for classification techniques. However, in some cases the spatial resolution of the hyperspectral data is not enough to separate complex classes such as those present in urban environments [59]. This aspect, together with the expected (linear or nonlinear) mixing happening at sub-pixel scales [141], complicates the classification process significantly. In some cases, the data coming from other sources can be used to improve and/or refine the results of classification. A good example is the use of LiDAR data [163], which can provide information about the height of the same surveyed area. LiDAR has been shown to be a very useful source of data for classification purposes [164].

In the literature, many techniques have been developed for fusion of hyperspectral and LiDAR data for classification purposes [165]. In 2013, the Data Fusion Technical Committee of the IEEE GRSS organized a contest[2] involving two data sets: a hyperspectral image and a LiDAR derived DSM, both at the same spatial resolution (2.5m), and two parallel competitions were established in order to devise advanced methods for fusion and classification of hyperspectral and LiDAR data[3]. Many other examples can be found in classification of urban areas [166], but also in classification of complex forest areas [167]. Techniques based on morphological features have been quite successful in the literature. For instance, the methodology in [168] jointly considered the features extracted by morphological attribute profiles [80] computed on both the hyperspectral and LiDAR data, and then fused the spectral, spatial and elevation information in a stacked architecture. In [169] it was pointed out that the simple concatenation or stacking of features such as morphological attribute profiles may contain redundant information. The main challenge in multiple feature learning is that how to adequately exploit the information containing in these features. In addition, a significant increase in the number of features may lead to high dimensionality issues that are in contrast with the limited number of training samples often available in remote sensing applications [170], which may lead to the Hughes effect. To address these issues, decision fusion techniques have been applied [171].

In this chapter, we develop a new strategy for fusing hyperspectral and LiDAR data for classification purposes. The main contribution of our newly proposed approach is its capacity to integrate multiple types of features extracted using spatial and spectral information. For the LiDAR data, the DSM provides rich information in spatial sense, while for the hyperspectral data the very rich spectral resolution provides detailed spectral signatures that can be very useful for classification purposes. A main characteristic of the presented approach is that it can adaptively exploit information from both spatially and spectrally derived features, thus being able to address practical scenarios in which different sources of information

---

[2]http://hyperspectral.ee.uh.edu/?page_id=459
[3]http://hyperspectral.ee.uh.edu/?page_id=795

(spatial or spectral) may be useful to separate different types of classes. In order to achieve this goal, the proposed approach has been designed in a way that it exhibits great flexibility to combine different types of features without any regularization parameters, thus taking advantage of the complementarity that the features can provide without any *a priori* restrictions. Our presented approach is thus aimed at exploiting the different properties that both spatial and spectral features can provide. In order to achieve the desired spectral-spatial integration that is normally expected in advanced classification problems, we consider morphological features as an important part of our framework, which also exploits the original spectral information contained in the hyperspectral scene. The integration is achieved by a multiple feature learning approach based on the MLR*sub* [51] classifier.

The remainder of this chapter is organized as follows. Section 7.3 describes the methodological framework developed in this work for the fusion of hyperspectral and LiDAR data. Section 7.4 describes our experimental results, conducted using a hyperspectral image and a LiDAR derived DSM collected over the University of Houston campus and the neighboring urban area. This scene was recently used in the IEEE GRSS contest and, therefore, the results obtained for this scene will be compared with other several approaches already tested with these data. Finally, section 7.5 concludes the chapter with some remarks and hints at plausible future research lines.

7.3   Methodological Framework

In this section, we introduce the proposed approach for the integration of the hyperspectral and LiDAR data. The proposed approach comprises the following main steps. In the first step, we use morphological attribute profiles [80] to extract the spatial features in the hyperspectral and LiDAR data. Then in the second step, we perform classification over all the obtained spatial features and the original spectral features by using the MLR*sub* classifier [51]. We have selected this classifier as it provide great flexibility for multiple feature learning based on the LOGP rule [172] while, at the same time, offering great ability for learning ill-posed problems by projecting the data into its class indexed subspace. In this way, by working in a subspace the proposed method can adequately handle the unbalance between the increased dimensionality of the data (expanded even more after including the morphological attribute profiles) and the very limited availability of training samples in practice. In a final step, we use MRFs [97] for spatial regularization in order to promote spatial smoothness in the final classification result, as indicated in [51]. In the following, we present in detail each step of the proposed approach for classification of hyperspectral and LiDAR data.

7.3.1   Feature Extraction by Morphological Attribute Profiles

Let $\mathbf{X}^L \equiv (x_1^L, x_2^L, ..., x_n^L)$ be the DSM derived from the LiDAR data, where n is the number of pixels in $\mathbf{X}^L$. Similarly, let us denote the hyperspectral image as $\mathbf{X}^h \equiv (\mathbf{x}_1^h, \mathbf{x}_2^h, ..., \mathbf{x}_n^h)$, where $\mathbf{x}_i \in \mathbb{R}^d$, for $i = 1, 2, \ldots, n$, denotes a spectral vector, n is the number of pixels in $\mathbf{X}^h$, and d is the number of spectral bands.

Mathematical morphology was described in subsection 1.1.4.3 as an effective approach for modeling the spatial characteristics of the objects in remotely sensed images. In this chapter we suggest to use morphological APs for feature extraction from the both hyperspectral and LiDAR DSM images. In our specific context, the multiple features available $\mathcal{X}$ are defined as follows: $\mathcal{X} \equiv (\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \tilde{\mathbf{X}}_3, \tilde{\mathbf{X}}_4)$, where $\tilde{\mathbf{X}}_1 = \mathbf{X}^h$ (i.e., the hyperspectral data), $\tilde{\mathbf{X}}_2 = \mathbf{X}^L$ (i.e., the LiDAR data), $\tilde{\mathbf{X}}_3 = \text{EMAP}(\mathbf{X}^h)$ (i.e., an EMAP built on the first few PCA components of the original hyperspectal data), and $\tilde{\mathbf{X}}_4 = \text{AP}(\mathbf{X}^L)$

(i.e., the AP of the LiDAR data). Notice that, for the LiDAR data, we could only generate its APs as it is a single-channel DSM image.

7.3.2 Probabilistic Classification Using Spectral Features

In the literature, various probabilistic techniques have been suggested for classification of remote sensing data. In this chapter we focus on two techniques: the SVM [27, 26, 31] and MLRsub (see chapter 3), which have shown good performance in hyperspectral data classification.

7.3.2.1 SVM Classification Technique

The SVM is originally a binary hard classifier that separates two classes by using a linear hyperplane. For finding the separating hyperplane, the training samples located near the class boundaries (i.e., the support vectors) that maximize the separation between the two classes are found, by using a constrained optimization process [27]. An important advantage of SVM algorithm is the possibility of using a kernel trick in the formulation [32]. In this chapter, we use the multi-class probabilistic SVM method [29, 126] implemented in the popular LIBSVM library [30], considering Gaussian RBF kernel $K((\tilde{\mathbf{x}}_i)_m, (\tilde{\mathbf{x}}_j)_m) = \exp(-\gamma \|(\tilde{\mathbf{x}}_i)_m - (\tilde{\mathbf{x}}_j)_m\|^2)$.

7.3.2.2 MLRsub Classification Technique

Using the notations in this chapter, the MLR classifier is given by:

$$p_m(y_i = c | (\tilde{\mathbf{x}}_i)_m, \boldsymbol{\omega}_m) = \frac{\exp\left(\boldsymbol{\omega}_m^{(c)} \mathbf{h}((\tilde{\mathbf{x}}_i)_m)\right)}{\sum_{c=1}^{k} \exp\left(\boldsymbol{\omega}_m^{(c)} \mathbf{h}((\tilde{\mathbf{x}}_i)_m)\right)}, \tag{7.1}$$

where $\mathbf{h}((\tilde{\mathbf{x}}_i)_m) \equiv [h_1((\tilde{\mathbf{x}}_i)_m), \ldots, h_l((\tilde{\mathbf{x}}_i)_m)]^T$ is a vector of l fixed functions of the input data, often termed as features; $\boldsymbol{\omega}_m^{(c)}$ is the set of logistic regressors for class c, and $\boldsymbol{\omega}_m \equiv [\boldsymbol{\omega}_m^{(1)^T}, \ldots, \boldsymbol{\omega}_m^{(k)^T}]^T$. The MLR$sub$ combines the classic MLR formulation with a subspace projection method in order to cope with highly mixed hyperspectral data using limited training samples. The idea of applying subspace projection methods to improve classification relies on the basic assumption that the samples within each class can approximately lie in a lower dimensional subspace. Thus, each class may be represented by a subspace spanned by a set of basis vectors, while the classification criterion for a new input sample would be the distance from the class subspace [51]. In this chapter, we use a modified version of MLRsub algorithm which was proposed in chapter 3. This method use the following input function $\mathbf{h}(\tilde{\mathbf{x}}_i)$ in (7.1) as:

$$\mathbf{h}((\tilde{\mathbf{x}}_i)_m) = [\|(\tilde{\mathbf{x}}_i)_m\|^2, \|(\tilde{\mathbf{x}}_i)_m^{\mathrm{T}} \mathbf{U}_m^{(1)}\|^2, \ldots, \|(\tilde{\mathbf{x}}_i)_m^{\mathrm{T}} \mathbf{U}_m^{(k)}\|^2]^{\mathrm{T}}, \tag{7.2}$$

where $\mathbf{U}_m^{(c)} = \{(\mathbf{u}_1^{(c)})_m, \ldots, (\mathbf{u}_{r^{(c)}}^{(c)})_m\}$ is a set of $r^{(c)}$-dimensional orthonormal basis vectors for the subspace associated with class c ($r^{(c)} \ll d_m$, and d_m is the dimensionality of feature $\tilde{\mathbf{X}}_m$).

7.3.3 Probabilistic Classification Using Multiple Spectral and Spatial Features

According to the LOGP rule [172], which is a decision fusion scheme that is commonly applied to combine information from multiple features, for any pixel $i = 1, \ldots, n$ we have:

$$p_{\text{LOGP}}(y_i = c | (\tilde{\mathbf{x}}_i)_1, \ldots, (\tilde{\mathbf{x}}_i)_4, \boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_4, \alpha_1, \ldots, \alpha_4) = \frac{\prod_{m=1}^{4} p_m(y_i = c | (\tilde{\mathbf{x}}_i)_m, \boldsymbol{\omega}_m)^{\alpha_m}}{\sum_{c=1}^{k} \prod_{m=1}^{4} p_m(y_i = c | (\tilde{\mathbf{x}}_i)_m, \boldsymbol{\omega}_m)^{\alpha_m}}, \quad (7.3)$$

where $\{\alpha_m | 0 \leq \alpha_m \leq 1, \sum_{m=1}^{4} \alpha_m = 1\}$ is a tunable parameter which controls the impact of each feature vector on the final decision probability. Notice that, for the multiple feature learning problem in (7.3), we have two different types of parameters: 1) the logistic regressors $\boldsymbol{\omega}_m$, and 2) the weight parameter α_m. These parameters are respectively associated with the classifier and with the features. This leads to difficulties from the viewpoint of both their optimization and the overall computational cost. Learning the weight parameter α_m is generally the most difficult task in the LOGP framework.

In order to relax the difficulty of learning the two different types of parameters, in this work we propose to use the MLR*sub* classifier in chapter 3 to model the posterior density $p_m(y_i = c | (\tilde{\mathbf{x}}_i)_m, \boldsymbol{\omega}_m)$. This provides the possibility of learning the regressors and the weight parameters in combined fashion, so that we only need to learn one type of parameter. This greatly lightens the computational cost while, at the same time, relaxing the optimization problem.

Under the present setup, by embedding the MLR*sub* model in (7.1) into the LOGP framework for multiple feature learning in (7.3), we can now obtain:

$$p_{\text{LOGP}}(y_i = c | (\tilde{\mathbf{x}}_i)_1, \ldots, (\tilde{\mathbf{x}}_i)_4, \boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_4, \alpha_1, \ldots, \alpha_4) = \frac{\exp\left(\sum_{m=1}^{4} \alpha_m \boldsymbol{\omega}_m^{(c)} \mathbf{h}((\tilde{\mathbf{x}}_i)_m)\right)}{\sum_{c=1}^{k} \exp\left(\sum_{m=1}^{4} \alpha_m \boldsymbol{\omega}_m^{(c)} \mathbf{h}((\tilde{\mathbf{x}}_i)_m)\right)}. \quad (7.4)$$

Notice that, in (7.4), we still have two different types of parameters α_m and $\boldsymbol{\omega}_m$. In general, in order to learn the joint density (7.3) or (7.4), learning of the weight parameter α_m is essential. In [172], several strategies were provided in order to learn the weight parameters. However, as shown in [172] and also as it was mentioned before, it is very time-consuming and difficult to search for the optimal value of α_m. However, the MLR*sub* has the potential to overcome these difficulties as it provides a different look to the problem by associating the weight with the regressors instead of the features, *i.e.*, by letting:

$$\tilde{\boldsymbol{\omega}}_m^{(c)} = \alpha_m \boldsymbol{\omega}_m^{(c)}, \quad (7.5)$$

the regressors $\boldsymbol{\omega}_m$ and weight parameters α_m are now combined into a new set of regressors $\tilde{\boldsymbol{\omega}}_m$ associated with the MLR*sub* classifier. In this way, we avoid the weight parameters for the features by introducing them into the regressors. This means that, under this transformation, the weight parameters for the features will be learned together the classifier. This provides important advantages from the viewpoint of learning: by transferring the weight parameter used for the features (observations) to the weight parameter used for the classifier, the proposed approach joins two different problems (parameter and classifier learning) into a single one, which represents a main contribution of this work from the viewpoint of simplifying and making more natural the process of learning from multiple features (i.e., those derived by the hyperspectral and the LiDAR data). By introducing the transformation (7.5), problem (7.4) changes to:

$$p_{\text{LOGP}}(y_i = c | (\tilde{\mathbf{x}}_i)_1, \ldots, (\tilde{\mathbf{x}}_i)_4, \tilde{\boldsymbol{\omega}}_1, \ldots, \tilde{\boldsymbol{\omega}}_4) = \frac{\exp\left(\sum_{m=1}^{4} \tilde{\boldsymbol{\omega}}_m^{(c)} \mathbf{h}((\tilde{\mathbf{x}}_i)_m)\right)}{\sum_{c=1}^{k} \exp\left(\sum_{m=1}^{4} \tilde{\boldsymbol{\omega}}_m^{(c)} \mathbf{h}((\tilde{\mathbf{x}}_i)_m)\right)}. \quad (7.6)$$
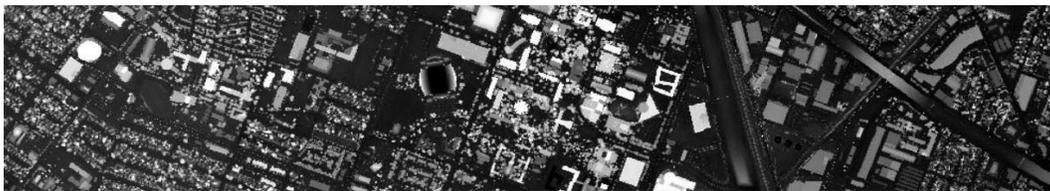
Figure 7.1: LiDAR derived DSM for the Houston data.

Notice that problem (7.6) is able to handle multiple features with the advantage of having a similar structure as the original MLRsub method. Therefore, problem (7.6) can be solved using the same approach presented in [51], which provides a detailed optimization framework for learning the logistic regressors. As the goal of this work is to adequately exploit the information coming from hyperspectral and LiDAR data, the proposed framework in (7.6) provides a great flexility in handling features from these two different sources of information, thus being able to naturally integrate spatial and spectral features without the need for *ad hoc* weights or regularization parameters. Although the proposed framework has the ability to integrate multiple types of features, in this work we constrain ourselves to the set $\mathcal{X} \equiv (\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \tilde{\mathbf{X}}_3, \tilde{\mathbf{X}}_4)$ with $\tilde{\mathbf{X}}_1 = \mathbf{X}^h$, $\tilde{\mathbf{X}}_2 = \mathbf{X}^L$, $\tilde{\mathbf{X}}_3 = \text{EMAP}(\mathbf{X}^h)$, and $\tilde{\mathbf{X}}_4 = \text{AP}(\mathbf{X}^L)$ as described in the previous subsection, for the validation of our approach (although additional features can be included in future developments of the method).

7.4 Experimental Results

In our experiments we have considered University of Houston data set consists of a hyperspectral image and a LiDAR derived DSM, both at the same spatial resolution (2.5m). The hyperspectral image was described in subsection (4.4). The corresponding co-registered DSM consists of elevation in meters above sea level (per the Geoid 2012A model). The data were acquired by the National Science Foundation NCALM over the University of Houston campus and the neighboring urban area. The LiDAR data was acquired on June 22, 2012, between the time 14:37:55 to 15:38:10 UTC. The average height of the sensor above ground was 2000ft. The false color composition of the hyperspectral data, the available ground-truth and the training set have shown in Fig. 4.2. Fig. 7.1 shows the LiDAR derived DSM. Moreover, detailed information about the class labels and the number of available training and test samples has been reported in Table 4.1. The hyperspectral/LiDAR data and the ground-truth available online for public use, which allows for detailed inter-comparisons among different algorithms.

Tables 7.1 and 7.2 show the classification results obtained by the SVM and MLRsub classifiers, respectively, and the different individual features considered in this work: the original hyperspectral image (i.e., $\mathbf{X}^h$), the AP built on the LiDAR DSM data [i.e., $\text{AP}(\mathbf{X}^L)$] and the EMAP built on the hyperspectral data [i.e., $\text{EMAP}(\mathbf{X}^h)$]. In order to build the $\text{EMAP}(\mathbf{X}^h)$ we fixed the number of principal components to be retained to the number of components that contain more than 98% of the total variance of the original hyperspectral data, where the obtained components are scaled to the range $[0, 1000]$ and converted to integer in order to build the attribute filters. Specifically, the EMAPs are built using the area (related to the size of the regions) and standard deviation (which measures the homogeneity of the pixels enclosed by the regions) attributes. The threshold values are chosen in the range $\{50, 500\}$ with a stepwise increment of 50 for the area attribute. For the standard deviation, attribute values ranging from 2.5% to 20% of the mean of the feature with a stepwise increment of 2.5% are chosen [168]. Finally,

in order to build the $AP(\mathbf{X}^L)$ we also used the area and standard deviation attributes, with the same configuration indicated above, but this time applied to a single component given by the LiDAR DSM data. Concerning the probabilistic SVM and MLR*sub* classifiers, we optimized the related parameters.

In Tables 7.1 and 7.2, we display the classification results obtained for individual features, with and without the MRF-based post-processing. As shown by Tables 7.1 and 7.2, the original spectral information contained in the hyperspectral image seems to be the most useful type of feature for classification purposes, while the AP built on the LiDAR DSM is not discriminative enough to separate many of the classes. This is expected, since the LiDAR DSM alone provides information about height that is not expected to be able to discriminate between the different urban classes in the considered data. Interestingly, the EMAP built on the original hyperspectral data is also not as discriminative as the original spectral information in this example. This also comes at no surprise, since the EMAP is mainly based on the spatial features of the objects but the classes in the considered problem are all difficult to discriminate based on spatial properties only, hence the use of spectral properties is quite important. Ideally, the separability of the classes could improve if we considered multiple features simultaneously.

Additionally, if we compare the results obtained by SVM with the results obtained by the MLR*sub* algorithm, we may conclude that SVM shows better performance for classification of the different individual features. However, the performance of the two classifiers for classification of the original hyperspectral image is comparable. Even the MLR*sub*-MRF has shown slightly better results in comparison with SVM-MRF for classification of individual spectral features. This is expected, since the MLR*sub* algorithm was originally developed for the classification of hyperspectral images based on the assumption that the hyperspectral features lie in a lower dimensional subspace. The other reason is that SVM is originally a hard classifier and the estimated class probabilities used in the MRF-based relaxation procedure are often not reliable. At this point, it is also important to reiterate that the main purpose of this chapter is developing a robust technique for improving the classification of hyperspectral images using other complementary features, and the MLR*sub* method shows good potential for this purpose.

On the other hand, Table 7.3 shows the classification results obtained by the proposed framework for multiple feature learning when different types of features are considered for the classification. Here, we also reported the results obtained by the MLR*sub* and by the MLR*sub*-MRF with spatial post-processing. The cases considered are: the original hyperspectral image plus the APs built on the LiDAR DSM [i.e., $\mathbf{X}^h$+$AP(\mathbf{X}^L)$], the original hyperspectral image plus the EMAPs built on the same image [i.e., $\mathbf{X}^h$+$EMAP(\mathbf{X}^h)$], the APs built of the LiDAR DSM plus the EMAPs built on the original hyperspectral image [i.e., $AP(\mathbf{X}^L)$+$EMAP(\mathbf{X}^h)$] and all the features available [i.e., $\mathbf{X}^h$+$AP(\mathbf{X}^L)$+$EMAP(\mathbf{X}^h)$]. As shown by Table 7.3, the combination of multiple features always increased the classification results with regards to the single-feature cases reported on Table 7.2. In particular, the classification results improved significantly when the three considered features: $\mathbf{X}^h$, $AP(\mathbf{X}^L)$ and $EMAP(\mathbf{X}^h)$ were used simultaneously. Interestingly, the combination of $AP(\mathbf{X}^L)$ and $EMAP(\mathbf{X}^h)$ without using the original spectral information in $\mathbf{X}^h$ also provided good classification results, which is interesting since the AP and EMAP are spatially-guided features. However, the combination of $\mathbf{X}^h$ and $EMAP(\mathbf{X}^h)$ resulted in the lowest classification accuracies reported on Table 7.3. This observation is also interesting, since both $\mathbf{X}^h$ and $EMAP(\mathbf{X}^h)$ are derived from the original hyperspectral image without including any information about the LiDAR DSM. In turn, the inclusion of the LiDAR-based $AP(\mathbf{X}^L)$ always resulted in an increase in the classification accuracies reported which indicates that, under the proposed framework, the LiDAR information represents a source of complementary information that can improve the analysis of the hyperspectral data alone.

Table 7.1: Class specific and overall classification accuracies [%] obtained by the SVM and the SVM-MRF for different types of combined features. The best results for each class are outlined in bold typeface.

#	Class	SVM			SVM-MRF		
		$\mathbf{X}^h$	AP($\mathbf{X}^L$)	EMAP($\mathbf{X}^h$)	$\mathbf{X}^h$	AP($\mathbf{X}^L$)	EMAP($\mathbf{X}^h$)
1	Healthy grass	82.43	42.55	82.15	**83.00**	35.71	80.82
2	Stressed grass	82.05	41.54	83.08	83.83	50.56	**84.02**
3	Synthetic grass	99.80	88.51	**100.00**	**100.00**	88.12	**100.00**
4	Trees	**92.80**	72.16	83.71	93.37	70.17	84.94
5	Soil	98.48	70.36	99.81	**100.00**	75.57	**100.00**
6	Water	95.10	66.43	93.71	**95.80**	66.43	79.02
7	Residential	75.47	71.74	75.00	**81.16**	67.63	77.61
8	Commercial	46.91	**89.08**	48.24	45.87	85.75	49.76
9	Road	77.53	59.02	77.05	**86.02**	66.01	81.78
10	Highway	60.04	63.90	47.49	**65.64**	64.00	48.55
11	Railway	81.02	99.91	91.08	83.21	**100.00**	91.37
12	Parking Lot 1	85.49	71.09	79.35	**98.46**	74.83	87.42
13	Parking Lot 2	75.09	69.47	72.63	**84.21**	72.28	76.49
14	Tennis Court	**100.00**	99.19	**100.00**	100.00	100.00	100.00
15	Running Track	98.31	98.94	**100.00**	99.15	100.00	100.00
	OA	80.49	70.80	79.19	**84.05**	71.67	80.77
	AA	83.37	73.59	82.22	**86.65**	74.47	82.79
	κ	78.98	68.41	77.41	**82.83**	69.34	79.12

In order to evaluate the statistical significance of the difference in accuracy between two classifications, the McNemar's test has been widely used in the remote sensing community [131]. In this test, a value of $|Z| > 1.96$ indicates that there is a significant difference in accuracy between two classification results. The sign of Z is also a criterion to indicate whether the first classifier compared is more accurate than the second one ($Z > 0$) or vice versa ($Z < 0$). As it can be seen from Table 7.4, the differences in classification accuracies between the case of using all considered features and the other combinations are statistically significant. Moreover, in a different scenario, in order to better evaluate the statistical significance of the differences, the McNemar's test was used for each object class and Bonferroni correction [173] was applied for McNemar test values in order to control the family wise type-I error rate. We considered each P value to indicate a significant difference if P was less than or equal to 0.001. The significant differences in accuracy between the case of using all considered features and EMAP($\mathbf{X}^h$) were the most for the MLR method and the least for the MLR-MRF method. Furthermore, differences in accuracy between the case of using all considered features and $\mathbf{X}^h$ were the most significant for the MLR-MRF method and the least significant for the MLR method.

At this point, it is important to reiterate that the proposed framework automatically integrates the different sources of information (spatial and spectral) automatically, and without the need to set a parameter to control their relative weight in the final classification result. This is a very important characteristic, as it allows for the integration of multiple features in a very natural way. In fact, additional

Table 7.2: Class specific and overall classification accuracies [%] obtained by the MLR*sub* and the MLR*sub*-MRF for different types of combined features. The best results for each class are outlined in bold typeface.

#	Class	MLR*sub*			MLR*sub*-MRF		
		$\mathbf{X}^h$	AP($\mathbf{X}^L$)	EMAP($\mathbf{X}^h$)	$\mathbf{X}^h$	AP($\mathbf{X}^L$)	EMAP($\mathbf{X}^h$)
1	Grass Healthy	82.53	60.97	80.53	**83.10**	57.83	80.72
2	Grass Stressed	81.02	12.12	77.16	**84.21**	12.69	79.42
3	Grass Synthetic	99.41	88.12	**100**	**100**	78.81	**100**
4	Tree	97.44	50.47	65.25	**100**	45.45	66.57
5	Soil	96.12	22.35	96.40	**99.53**	30.59	99.15
6	Water	94.41	67.83	78.32	**95.80**	66.43	79.02
7	Residential	73.51	76.77	58.68	81.53	**83.77**	52.71
8	Comercial	62.68	**80.72**	40.46	57.55	77.68	41.03
9	Road	71.39	24.27	75.83	**86.87**	30.88	83.00
10	Highway	86.87	56.85	48.17	**88.51**	59.75	46.81
11	Railway	67.65	93.26	94.97	79.89	88.24	**100**
12	Parking Lot 1	59.85	61.38	81.65	72.72	53.41	**97.79**
13	Parking Lot 2	57.19	67.02	66.32	**84.21**	69.47	82.46
14	Tennis Court	97.57	99.60	96.76	**100**	**100**	**100**
15	Running Track	95.77	89.64	97.46	98.52	**100**	98.73
	OA	79.60	58.08	74.53	**85.18**	58.26	77.44
	AA	81.56	63.43	77.20	**87.50**	63.67	80.49
	Kappa	77.86	54.67	72.35	**83.92**	54.76	75.52

features such as texture, border-related features, etc. could be integrated in the proposed framework in order in enhance the obtained classification results. As shown by our experiments, the classification results were obtained using a relatively low number of training samples, and the fact that we increased the number of features did not decrease but rather increase the classification accuracies for the same number of training samples. This is due to the subspace-based nature of the MLR*sub* and MLR*sub*-MRF classifiers used to obtain the final classification maps. For illustrative purposes, Figs. 7.2 and 7.3 respectively show some of the classification maps obtained by using the MLR*sub* and the MLR*sub*-MRF for the considered Houston hyperspectral/LiDAR data. Effective classification results can be observed in the final maps reported in these figures.

Moreover, using the described data set, we have conducted another experiment to more carefully analyze the validation set and the effectiveness of the selected features in the proposed method. In this experiment we have randomly selected 10% of the available labeled samples of each class from the validation set for training purposes. It is important to mention that in the original training set no training samples were selected from the right part of the image where a large cloud shadow is present. However, in the validation set, there are a significant number of samples from this part. Table 7.5 details the average of the results for this experiment obtained after 30 Monte Carlo runs. Several conclusions can be obtained from Table 7.5. First and foremost, it is remarkable that the proposed multiple feature Learning method, showed better performances in comparison with using single kind of features which

Table 7.3: class specific and overall classification accuracies [%] obtained by the MLR*sub* and the MLR*sub*-MRF for multiple feature combinations. The best results for each class are outlined in bold typeface.

#	Class	MLR*sub*				MLR*sub*-MRF			
		$\mathbf{X}^h$ +AP($\mathbf{X}^L$)	$\mathbf{X}^h$ +EMAP($\mathbf{X}^h$)	AP($\mathbf{X}^L$) +EMAP($\mathbf{X}^h$)	$\mathbf{X}^h$+AP($\mathbf{X}^L$) +EMAP($\mathbf{X}^h$)	$\mathbf{X}^h$ +AP($\mathbf{X}^L$)	$\mathbf{X}^h$ +EMAP($\mathbf{X}^h$)	AP($\mathbf{X}^L$) +EMAP($\mathbf{X}^h$)	$\mathbf{X}^h$+AP($\mathbf{X}^L$) +EMAP($\mathbf{X}^h$)
1	Grass Healthy	83.00	80.91	80.63	82.91	83.00	80.91	80.82	**83.10**
2	Grass Stressed	77.44	81.48	74.34	81.48	79.79	**84.30**	74.25	83.08
3	Grass Synthetic	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
4	Tree	97.92	86.55	86.17	95.83	**99.62**	88.73	87.69	95.93
5	Soil	98.20	98.58	96.69	99.05	99.72	**100**	98.20	**100**
6	Water	91.61	95.10	79.02	91.61	**95.80**	**95.80**	79.02	**95.80**
7	Residential	86.29	76.77	88.15	87.59	82.56	81.72	**90.67**	84.79
8	Commercial	88.03	64.10	77.97	84.14	**88.32**	57.45	77.30	83.57
9	Road	85.65	82.34	93.77	91.78	92.92	88.86	**97.92**	96.51
10	Highway	69.11	92.28	71.43	86.20	69.69	**99.90**	73.26	86.78
11	Railway	98.10	74.95	98.86	98.58	99.53	82.92	**99.91**	99.72
12	Parking Lot 1	85.88	91.64	89.53	92.32	94.91	**100**	96.73	99.81
13	Parking Lot 2	70.53	70.53	76.49	76.84	75.79	**86.32**	78.25	85.26
14	Tennis Court	99.60	98.79	**100**	99.60	**100**	**100**	**100**	**100**
15	Running Track	98.73	98.10	98.73	98.73	98.94	98.73	98.94	98.73
	OA	87.91	84.40	86.86	90.65	89.85	87.86	88.56	**92.05**
	AA	88.67	86.14	87.45	91.11	90.71	89.71	88.86	**92.87**
	Kappa	86.87	83.10	85.74	89.85	88.99	86.85	87.58	**91.37**

Table 7.4: Statistical significance of differences in classification accuracies.

Method	Value of z calculated by the McNemar's test					
				$\{\mathbf{X}^h+\text{AP}(\mathbf{X}^L)+\text{EMAP}(\mathbf{X}^h)\}/$		
	$\mathbf{X}^h$	$\text{AP}(\mathbf{X}^L)$	$\text{EMAP}(\mathbf{X}^h)$	$\{\mathbf{X}^h+\text{AP}(\mathbf{X}^L)\}$	$\{\mathbf{X}^h+\text{EMAP}(\mathbf{X}^h)\}$	$\{\text{AP}(\mathbf{X}^L)+\text{EMAP}(\mathbf{X}^h)\}$
MLR	30.78	59.70	42.42	13.61	22.32	16.42
MLR-MRF	24.87	61.73	41.23	11.51	20.11	15.42
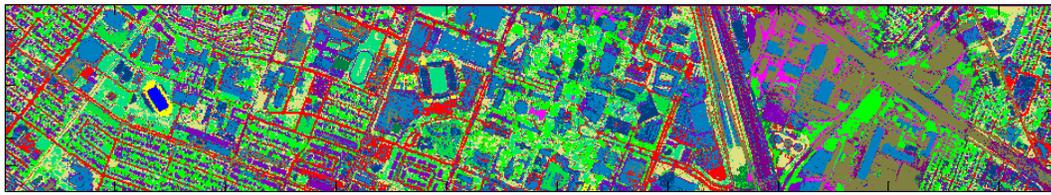
(a) MLR*sub* classification (79.60%) $\mathbf{X}^h$

(b) MLR*sub* classification (58.08%) using AP($\mathbf{X}^L$)

(c) MLR*sub* classification (74.53%) using EMAP($\mathbf{X}^h$)
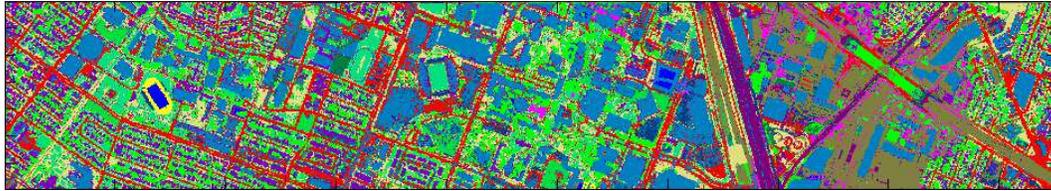
(d) MLR*sub* classification (87.91%) using $\mathbf{X}^h$+AP($\mathbf{X}^L$)

(e) MLR*sub* classification (84.40%) using $\mathbf{X}^h$+EMAP($\mathbf{X}^h$)

(f) MLR*sub* classification (86.86%) using AP($\mathbf{X}^L$)+EMAP($\mathbf{X}^h$)
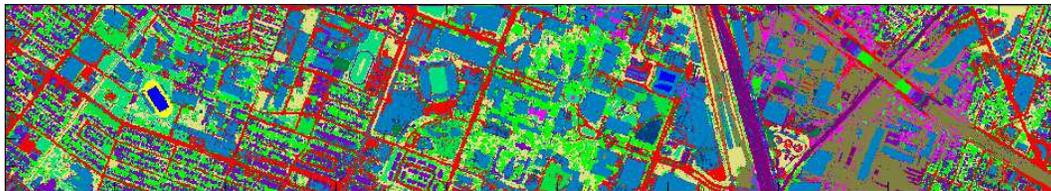
(g) MLR*sub* classification (90.65%) using all the available features: $\mathbf{X}^h$+AP($\mathbf{X}^L$)+EMAP($\mathbf{X}^h$)

Figure 7.2: Classification maps obtained by the MLR*sub* for different features: (a) $\mathbf{X}^h$, (b) AP($\mathbf{X}^L$), (c) EMAP($\mathbf{X}^h$), (d) $\mathbf{X}^h$+AP($\mathbf{X}^L$), (e) $\mathbf{X}^h$+EMAP($\mathbf{X}^h$), (f) AP($\mathbf{X}^L$)+EMAP($\mathbf{X}^h$), (g) $\mathbf{X}^h$+AP($\mathbf{X}^L$)+EMAP($\mathbf{X}^h$).

(a) MLR*sub*-MRF classification (85.18%) using $\mathbf{X}^h$
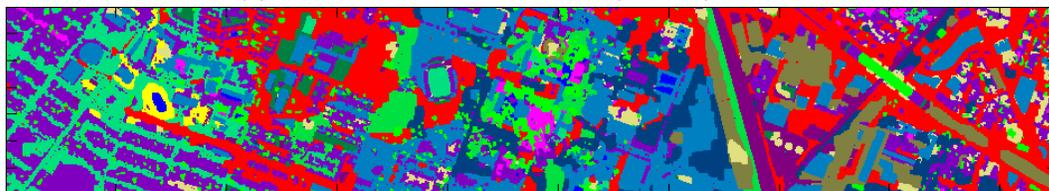
(b) MLR*sub*-MRF classification (58.26%) using AP($\mathbf{X}^L$)

(c) MLR*sub*-MRF classification (77.44%) using EMAP($\mathbf{X}^h$)

(d) MLR*sub*-MRF classification (89.85%) using $\mathbf{X}^h$+AP($\mathbf{X}^L$)

(e) MLR*sub*-MRF classification (87.86%) using $\mathbf{X}^h$+EMAP($\mathbf{X}^h$)

(f) MLR*sub*-MRF classification (88.56%) using AP($\mathbf{X}^L$)+EMAP($\mathbf{X}^h$)
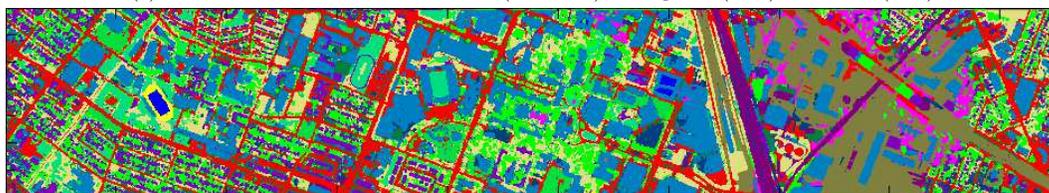
(g) MLR*sub*-MRF classification (92.05%) using all the available features: $\mathbf{X}^h$+AP($\mathbf{X}^L$)+EMAP($\mathbf{X}^h$)

Figure 7.3: Classification maps obtained by the MLR*sub*-MRF for different features: (a) $\mathbf{X}^h$, (b) AP($\mathbf{X}^L$), (c) EMAP($\mathbf{X}^h$), (d) $\mathbf{X}^h$+AP($\mathbf{X}^L$), (e) $\mathbf{X}^h$+EMAP($\mathbf{X}^h$), (f) AP($\mathbf{X}^L$)+EMAP($\mathbf{X}^h$), (g) $\mathbf{X}^h$+AP($\mathbf{X}^L$)+EMAP($\mathbf{X}^h$).

means that our proposed method efficiently exploits the information contained in both data sources. More importantly, the performance improvements reported for the proposed method using all kinds of suggested features are quite significant. For example for the classes *Parking Lot 2* corresponded to parked vehicles and *Highway* in the cloud-covered region, we can see a significant improvement in the obtained classification result.

7.5   Summary and future directions

In this chapter, we have developed a new efficient strategy for fusion and classification of hyperspectral and LiDAR data. Our approach effectively integrates multiple types of features extracted from these data without the need for any regularization or weight parameters, so that different types of features can be efficiently exploited and integrated in a collaborative and flexible way. In this work we have considered several types of spatial and spectral features derived from the original hyperspectral image and from the LiDAR derived DSM, including the full original spectral information and different types of morphological profiles calculated for the hyperspectral and the LiDAR data. Our experimental results, conducted using a hyperspectral image and a LiDAR derived DSM collected over the University of Houston campus and the neighboring urban area, indicate that the information provided by LiDAR can effectively complement the spectral and the spatial information that can be extracted from the original hyperspectral data, providing an increase in the classification accuracies when the LiDAR data is used as a complementary source of information with regards to the original hyperspectral data. Although our analysis in this work has been constrained to a reduced number and type of features, the proposed framework is completely open and flexible in its capacity to integrate additional types of (spatial and spectral) features. As a result, future work will be directed towards the inclusion of additional types of features such as texture, border-oriented features, etc. Although our experiments have been reported for a data set that has been widely used in the recent literature as it was distributed as part of the IEEE GRSS Data Fusion contest in 2013, in the future we will also conduct further experiments using additional hyperspectral and LiDAR image pairs and perform comparisons of our approach with regards to other techniques recently presented for fusing hyperspectral and LiDAR data.

Table 7.5: Class specific and overall classification accuracies [%] obtained by the MLR*sub* for different selection of feature combinations, using 10% of validation samples for each class. The best results for each class are outlined in bold typeface.

#	Class	Samples		MLR*sub*						
		Train	Test	$\mathbf{X}^h$	AP($\mathbf{X}^L$)	EMAP($\mathbf{X}^h$)	$\mathbf{X}^h$ +AP($\mathbf{X}^L$)	$\mathbf{X}^h$ +EMAP($\mathbf{X}^h$)	AP($\mathbf{X}^L$) +EMAP($\mathbf{X}^h$)	$\mathbf{X}^h$+AP($\mathbf{X}^L$) +EMAP($\mathbf{X}^h$)
1	Healthy grass	105	1053	95.95 (2.09)	38.73 (6.51)	83.16 (3.75)	94.90 (3.01)	**96.22 (2.59)**	87.59 (2.72)	95.23 (3.62)
2	Stressed grass	106	1064	**97.03 (1.37)**	26.21 (4.78)	89.51 (4.07)	96.34 (1.31)	96.85 (1.36)	91.92 (3.65)	96.66 (1.40)
3	Synthetic grass	50	505	99.93 (0.11)	86.66 (7.92)	**100.00 (0.00)**	**100.00 (0.00)**	**100.00 (0.00)**	**100.00 (0.00)**	**100.00 (0.00)**
4	Trees	105	1056	94.13 (2.19)	82.24 (2.87)	88.72 (2.14)	98.91 (0.55)	94.25 (2.46)	98.45 (1.07)	**99.42 (0.57)**
5	Soil	105	1056	98.00 (0.71)	61.49 (2.80)	97.53 (1.34)	98.32 (0.79)	99.23 (0.35)	98.38 (0.94)	**99.51 (0.31)**
6	Water	14	143	93.54 (3.01)	69.72 (5.19)	77.55 (4.89)	92.73 (6.08)	**93.99 (5.34)**	80.58 (7.04)	93.45 (7.30)
7	Residential	107	1072	83.21 (3.35)	82.62 (2.81)	62.22 (5.52)	96.48 (1.61)	89.42 (3.81)	96.45 (1.51)	**98.37 (1.19)**
8	Commercial	105	1053	67.20 (7.34)	90.63 (2.81)	50.05 (3.05)	95.19 (2.08)	76.92 (6.37)	93.75 (2.92)	**95.48 (2.55)**
9	Road	105	1059	72.51 (3.08)	59.35 (7.53)	78.75 (2.47)	89.12 (2.31)	87.07 (2.31)	93.50 (1.62)	**94.65 (1.57)**
10	Highway	103	1036	86.82 (2.22)	65.16 (5.42)	90.00 (3.75)	93.99 (2.55)	96.57 (2.67)	97.70 (0.76)	**99.07 (0.59)**
11	Railway	105	1054	77.75 (2.88)	90.95 (2.64)	90.28 (3.29)	96.39 (0.86)	87.91 (2.88)	**97.44 (0.81)**	97.33 (0.81)
12	Parking Lot 1	104	1041	76.96 (4.22)	57.86 (10.11)	85.76 (5.36)	86.42 (4.43)	91.52 (3.58)	88.21 (4.39)	**92.36 (2.90)**
13	Parking Lot 2	28	285	48.97 (6.41)	58.48 (7.90)	73.81 (4.59)	73.10 (4.34)	75.57 (4.32)	83.92 (3.84)	**84.30 (3.72)**
14	Tennis Court	24	247	96.92 (1.81)	97.91 (0.82)	92.47 (3.28)	99.55 (0.33)	98.18 (1.20)	98.92 (0.50)	**99.69 (0.27)**
15	Running Track	47	473	98.09 (0.84)	91.16 (5.01)	97.76 (1.09)	**99.34 (0.58)**	98.91 (0.66)	99.07 (0.60)	99.27 (0.57)
	OA			85.60 (1.17)	67.93 (1.54)	82.94 (0.79)	94.60 (0.51)	92.01 (1.08)	94.45 (0.46)	**96.77 (0.34)**
	AA			85.80 (1.03)	70.61 (1.41)	83.84 (0.67)	94.05 (0.63)	92.17 (0.99)	93.73 (0.50)	**96.32 (0.49)**
	κ			84.38 (1.27)	65.31 (1.65)	81.49 (0.85)	94.14 (0.55)	91.33 (1.17)	93.97 (0.49)	**96.49 (0.37)**

Chapter 8

Conclusions and Future Research Lines

8.1   Conclusions

This thesis presented novel techniques and methodologies for classification of hyperspectral data. The main emphasis was given to developing probabilistic classification approaches, considering the fact that subspace-based methods can efficiently cope with the presence of noise and mixed pixels in the hyperspectral images. Furthermore, we tried to exploit other complementary sources of information such as spatial-contextual information available in hyperspectral data and height information coming from LiDAR data in order to improve classification accuracies. As the main concluding remarks, the following points can be mentioned:

- In the second chapter of the thesis, in order to address the issue of mixed pixel characterization, a new spectral-spatial classifier was presented. The MLR*sub* classifier was used to learn the probabilities, where in addition to considering global posterior probability distributions which resulted from the whole image, a new strategy for locally learning the classification probabilities using previously derived class combination map was proposed. Since the mixed pixels in a hyperspectral image are normally mixed by only a few components, locally removing the impact of irrelevant classes by means of a class combination map could better characterize mixed pixels. The class combination map, that determines the number of mixtures appearing in each pixel as a pre-processing stage, was generated from the probabilistic SVM classification results. Integrating local and global probabilities in the analysis of hyperspectral data in order to constrain the number of mixing components used in the characterization of mixed pixels, represents an innovation with regard to previous approaches for probabilistic classification of hyperspectral data. Moreover, in the final step of the proposed strategy and after fusing the information provided by local and global probabilities, spatial information was characterized by including an MRF regularizer.

- In the third chapter of the thesis, we developed a subspace-based MLR method for pixelwise hyperspectral image classification. As an extension of a previous methodology in which class indexed input functions were used in the MLR algorithm for the subspace projection, in our newly proposed approach the integration of class-indexed subspaces constructed by the information coming from all the classes was explored. We showed that incorporating a general input function instead of using several class indexed input functions in the MLR classifier for projecting

hyperspectral data into union of subspaces can better handle the effects of noise and the presence of mixed pixels in a hyperspectral image. The experimental results using both simulated and real hyperspectral data sets demonstrated that our proposed method models not only the linear mixing process that is often present in hyperspectral images but also the nonlinearities that are more separable in the feature space defined by the union of class-indexed subspaces .

- In the forth chapter of the thesis, following the work in chapter 3, a new classification method based on union of subspaces was proposed for characterizing mixed (linear and nonlinear) pixels in hyperspectral images. In chapter 3, we showed that including subspace idea in the MLR classifier using a general input function shows better performance than several class indexed input functions. In addition to this strategy, in this chapter we proposed to use the assumption that the training samples of each class may lie in a union of subspaces rather than a single lower-dimensional subspace. For this purpose, we exploited a subspace clustering method to partition training samples of each class to multiple subsets regarding to existing subspaces and then the MLR algorithm was used to learn the posterior probability distributions from the spectral information of each subset, using a subspace projection. For subspace clustering, we suggested to use a new concept of RSC which, so far as we are aware, has not been applied for hyperspectral image analysis.

- In the fifth chapter of the thesis, we extended the subspace-projection-based concept to SVM classifier, a very popular technique for hyperpsectral image classification. For this purpose, we suggested to learn the subspace associated to each class in the construction of the SVM nonlinear function. The resulted formulation for the SVM classifier allows us to better cope with several phenomena that are quite important in hyperspectral image classification, such as the imbalance between the (high) dimensionality of the input data and the (limited) availability of training samples, as well as with the presence of noise and mixed pixels in the input data.

- In the sixth chapter of the thesis, we proposed a new methodology for spectral-spatial classification of hyperspectral images. In order to use local relationship among neighboring pixels, we developed a discontinuity preserving relaxation strategy, which can be used for postprocessing of class probability estimates, as well as preprocessing of the original hyperspectral image. The proposed approach showed significant performance in enforcing smoothness while, at the same time, accurately preserving the edges of class boundaries. We suggested to estimate the discontinuity map from the original image cube since it contains many image bands that allows consistent estimation.

- In the seventh chapter of the thesis, we developed a robust strategy for exploiting the information coming from multiple features without the need for any regularization or weight parameters. The proposed method was based on the integration of LOGP rule, which is a decision fusion scheme that is commonly applied to combine information from multiple features, and MLR*sub* algorithm in chapter 3 of this thesis, which is used for modeling the posterior probabilities. An important contribution of this work was transferring the weight parameter used for the features (observations) to the weight parameter used for the classifier. In this chapter we considered several types of spatial and spectral features derived from the original hyperspectral image and from the LiDAR derived DSM, including the full original spectral information and different types of morphological profiles calculated for the hyperspectral and the LiDAR data.

8.2   Perspectives

As with any new work, there are many open avenues for future research that deserve attention and which will be explored in our future developments. In the following, we list the most relevant of these perspectives for future work:

- It would be interesting to further improve the estimation of local probabilities by adaptively estimating the number of mixed components for each pixel rather than set them in advance. This is expected to improve the class separability and consequently the obtained classification results. This aim should be fulfilled carefully without imposing high computational complexity.

- The integration of techniques for spectral unmixing and hyperspectral image classification is also another future possibility which can be explored for integrating local and global probabilities in the analysis of hyperspectral data in order to constrain the number of mixing components used in the characterization of mixed pixels.

- We showed that performing subspace clustering of training samples as a preliminary step before subspace based MLR algorithm may better characterize mixed pixels and consequently, improve classification results. As an interesting future research line, we are planning to develop a unified framework for probabilistic classification of hyperspectral images based on union of subspaces.

- We are currently developing probabilistic sparse representation approach using MLR algorithm to estimate the class conditional distribution. The sparse representation-based classification method looks for the sparsest representation of a test sample in a dictionary composed of all training samples and thus, it can be considered for the problem of classification in the union of subspaces setting.

- For the purpose of classification of hyperspectral images, future work will be directed towards the inclusion of additional types of features such as texture, border-oriented features and performing further experiments for the fusion of hyperspectral and LiDAR images.

- Another topic of research deserving future attention is the computationally efficient implementations of the new techniques developed in this thesis using high performance computing architectures, such as clusters of computers (possibly, with specialized hardware accelerators such as graphics processing units).

Apendix A

Publications

The results of this thesis work have been published in several international journal papers, peer-reviewed international conference papers and peer-reviewed national conference papers. Specifically, the candidate has co-authored 6 journal citation reports (JCR) papers, 9 peer-review international conference papers.

The candidate has been a pre-doctoral researcher in the Hyperspectral Computing Laboratory (HyperComp), Department of Technology of Computers and Communications, University of Extremadura, Spain. Below, we provide a description of the publications achieved by the candidate providing also a short description of the journal or workshop where they were presented.

A.1   International journal papers

1. J. Li, **M. Khodadadzadeh**, A. Plaza, X. Jia and J. M. Bioucas-Dias. **A Discontinuity Preserving Relaxation scheme for Spectral-Spatial Hyperspectral Image Classification**, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, accepted for publication subject to minor revisions, 2015 [JCR(2013)=2.827].

   This paper submitted to the journal *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, which is a very important journal in the first quarter of the remote sensing and electrical and electronic engineering areas of JCR. This paper develops an iterative relaxation procedure which exploits spatial information in such a way that it considers discontinuities existing in the hyperspectral data cube, and constitutes the basis for the sixth chapter of the thesis.

2. **M. Khodadadzadeh**, J. Li, S. Prasad and A. Plaza. **Fusion of Hyperspectral and LiDAR Remote Sensing Data Using Multiple Feature Learning**, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, accepted for publication, 2015 [JCR(2013)=2.827].

   This paper was published in the journal *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, which is a very important journal in the first quarter of the remote sensing and electrical and electronic engineering areas of JCR. This paper proposes a robust strategy for fusion and classification of hyperspectral and LiDAR data, and constitutes the basis for the seventh chapter of the thesis.

3. L. Gao, J. Li, **M. Khodadadzadeh**, A. Plaza, B. Zhang, Z. He, and H. Yan. **Subspace-Based Support Vector Machines for Hyperspectral Image Classification**. IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 2, pp. 349-353, February 2015 [JCR(2013)=1.809].

This paper was published in the journal *IEEE Geoscience and Remote Sensing Letters*, which is one of the main journals of the remote sensing category of JCR. It is also in the second quarter of the electrical and electronic engineering category of JCR. The paper extends the subspace-projection-based concept to SVM classifier, and constitutes the basis of the fifth chapter of this thesis.

4. **M. Khodadadzadeh**, J. Li, A. Plaza, and J. M. Bioucas-Dias. **A Subspace-Based Multinomial Logistic Regression for Hyperspectral Image Classification**. IEEE Geoscience and Remote Sensing Letters, vol. 11, no. 12, pp. 2105-2109, December 2014 [JCR(2013)=1.809].

   This paper was published in the journal *IEEE Geoscience and Remote Sensing Letters*, which is one of the main journals of the remote sensing category of JCR. It is also in the second quarter of the electrical and electronic engineering category of JCR. Based on the idea of projection on class-indexed subspaces, the paper proposes a subspace-based MLR method for pixelwise hyperspectral classification, and constitutes the basis of the third chapter of this thesis.

5. **M. Khodadadzadeh**, J. Li, A. Plaza, H. Ghassemian, J. M. Bioucas-Dias and X. Li. **Spectral-Spatial Classification of Hyperspectral Data Using Local and Global Probabilities for Mixed Pixel Characterization**. IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 10, pp. 6298-6314, October 2014 [JCR(2013)=2.933].

   This paper was published in the *IEEE Transactions on Geoscience and Remote Sensing*, which is a top scholarly journal in the field of remote sensing. This paper develops a new spectral-spatial classifier for hyperspectral image based on the consideration of both global posterior probability distributions and local probabilities, and constitutes the basis of the second chapter presented in this thesis.

A.2   International journal papers submitted

1. **M. Khodadadzadeh**, J. Li, A. Plaza and J. M. Bioucas-Dias. **Hyperspectral Image Classification Based on Union of Subspaces**. submitted to IEEE Transactions on Geoscience and Remote Sensing, 2015 [JCR(2013)=2.933].

   This paper submitted to the journal *IEEE Transactions on Geoscience and Remote Sensing*, which is a top scholarly journal in the field of remote sensing. This paper proposes a new probabilistic classification approach based on union of subspaces for hyperspectral images.

A.3   Peer-reviewed international conference papers

1. **M. Khodadadzadeh**, A. Cuartero, J. Li, A. Felicisimo, A. Plaza. **Fusion of Hyperspectral And LiDAR Data Using Generalized Composite Kernels: A Case Study in Extremadura, Spain**. This work was presented as an oral presentation in the *IEEE International Geoscience and Remote Sensing Symposium* (IGARSS) held in Milan, Italy, in 2015. This is the most important international workshop in the remote sensing field. The paper proposes to exploit composite kernels for the integration of hyperspectral and LiDAR data for classification purposes.

2. **M. Khodadadzadeh**, J. Li, A. Plaza and J. M. Bioucas-Dias. **Hyperspectral Image Classification Based on Union of Subspaces**. This work was presented as an oral presentation

in the *IEEE Urban Remote Sensing Joint Event* (JURSE) held in Lausanne, Switzerland, in 2015. JURSE is one of the most important international workshops that specialize in urban hyperspectral remote sensing. Exploiting the union of subspaces in an MLR framework for supervised hyperspectral image classification is the main contribution of this paper.

3. **M. Khodadadzadeh**, J. Li, A. Plaza, P. Gamba, J. A. Benediktsson and J. M. Bioucas-Dias. **A New Framework for Hyperspectral Image Classification Using Multiple Spectral and Spatial Features**. This work was presented as an oral presentation in the *IEEE International Geoscience and Remote Sensing Symposium* (IGARSS) held in Quebec City, Canada, in 2014. This is the most important international workshop in the remote sensing field. The paper develops a new multiple feature learning approach for integrating spectral and spatial information in classification of hyperspectral images.

4. **M. Khodadadzadeh**, J. Li, A. Plaza, H. Ghassemian and J. M. Bioucas-Dias. **Spectral-Spatial Classification for Hyperspectral Data Using SVM and Subspace MLR**. This work was presented as an oral presentation in the *IEEE International Geoscience and Remote Sensing Symposium* (IGARSS) held in Melbourne, Australia, in 2013. This is the most important international workshop in the remote sensing field. The paper presents a new multiple-classifier approach for accurate spectral-spatial classification of hyperspectral images.

5. R. Rajabi, **M. Khodadadzadeh** and H. Ghassemian. **Graph Regularized Nonnegative Matrix Factorization for Hyperspectral Data Unmixing**. This work was presented as an oral presentation in the *Iranian Conference on Machine Vision and Image Processing* (MVIP) held in Tehran, Iran, in 2011. This is the most important iranian conference specialized in the fields of machine vision, image processing and their applications. This paper examines the applicability of the graph regularized nonnegative matrix factorization algorithm for spectral unmixing.

6. **M. Khodadadzadeh**, R. Rajabi and H. Ghassemian. **A Novel Approach for Spectral-Spatial Classification of Hyperspectral Data Based on SVM-MRF Method**. This work was presented as a poster in the *IEEE International Geoscience and Remote Sensing Symposium* (IGARSS) held in Vancouver, Canada, in 2011. This is the most important international workshop in the remote sensing field. The paper presents an innovative approach for integrating contextual information only for uncertain pixels which are the pixels near the borders (spatial boundaries) to improve the classification performance.

7. **M. Khodadadzadeh**, R. Rajabi and H. Ghassemian. **Segmentation and Classification of Remote Sensing Images Using Confident Marker Selection**. This work was presented as a poster in the *International Symposium on Artificial Intelligence and Signal Processing* (AISP) held in Tehran, Iran, in 2011. This is an important international event specialized in the fields of artificial intelligence and signal processing. This paper proposes a straightforward spectral-spatial method to choose the most reliable classified pixels in order to define suitable markers.

8. **M. Khodadadzadeh** and H. Ghassemian. **A Novel Contextual Classification of Hyperspectral Data Using Probabilistic Label Relaxation Process**. This work was presented as an oral presentation in the *Iranian Conference on Electrical Engineering* (ICEE) held in Tehran, Iran, in 2011. This is the most important and flagship conference in electrical and computer engineering in Iran. This paper proposes to use techniques that provide a multi-class

probability estimate for SVM classifier and then by using these class probability estimates, perform PLR after SVM classification to incorporate contextual information.

9. **M. Khodadadzadeh**, R. Rajabi, and H. Ghassemian. **Combination of Region-Based and Pixel-Based Hyperspectral Image Classification using Erosion Technique and MRF Model**. This work was presented as an oral presentation in the *Iranian Conference on Electrical Engineering* (ICEE) held in Isfahan, Iran, 2010. This is the most important and flagship conference in electrical and computer engineering in Iran. In this paper a new classification approach is proposed for hyperspectral images based on combining region-based and pixel-based methods

Bibliography

[1] C.-I. Chang. *Hyperspectral Data Exploitation: Theory and Applications*. New York: Wiley-Interscience, 2007. [Cited in pags. 1, 15, 39 and 51]

[2] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis, et al. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). 65(3):227–248, 1998. [Cited in pags. 1 and 22]

[3] D.-W. Sun. *Hyperspectral imaging for food quality analysis and control*. Elsevier, 2010. [Cited in pag. 2]

[4] D. Lorente, N. Aleixos, J. Gómez-Sanchis, S. Cubero, O. L. García-Navarrete, and J. Blasco. Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment. *Food and Bioprocess Technology*, 5(4):1121–1142, 2012. [Cited in pag. 2]

[5] G. Mooradian, M. Weiderhold, A. E. Dabiri, and C. Coyle. Hyperspectral imaging methods and apparatus for non-invasive diagnosis of tissue for cancer, July 1998. US Patent 5,782,770. [Cited in pag. 2]

[6] L. Zhi, D. Zhang, J.-q. Yan, Q.-L. Li, and Q.-l. Tang. Classification of hyperspectral medical tongue images for tongue diagnosis. *Computerized Medical Imaging and Graphics*, 31(8):672–678, 2007. [Cited in pag. 2]

[7] T.J. Cudahy, L.B. Whitbourn, P.M. Connor, P. Mason, and R.N. Phillips. Mapping surface mineralogy and scattering behavior using backscattered reflectance from a hyperspectral midinfrared airborne co 2 laser system (miraco2las). *IEEE Transactions on Geoscience and Remote Sensing*, 37(4):2019–2034, July 1999. [Cited in pag. 2]

[8] F. V. D. Meer, H. V. D. Werff, F. V. Ruitenbeek, C. A. Hecker, W. H. Bakker, M. F. Noomen, M. V. D. Meijde, E. J. M. Carranza, J. B. D. Smeth, and T. Woldai. Multi- and hyperspectral geologic remote sensing: A review. *International Journal of Applied Earth Observation and Geoinformation*, 14(1):112 – 128, 2012. [Cited in pag. 2]

[9] M. T. Eismann, C. R. Schwartz, J. N. Cederquist, J. A. Hackwell, and R. J. Huppi. Comparison of infrared imaging hyperspectral sensors for military target detection applications. In *Proceedings of the SPIE*, volume 2819, pages 91–101, 1996. [Cited in pag. 2]

[10] K. Heesung and N.M. Nasrabadi. Kernel rx-algorithm: a nonlinear anomaly detector for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 43(2):388–397, February 2005. [Cited in pag. 2]

[11] D. A. Landgrebe. *Signal Theory Methods in Multispectral Remote Sensing*. John Wiley & Sons, Ltd, 2003.   [Cited in pags. 2, 3, 9, 15, 16, 23, 59, 60 and 67]

[12] J. A. Richards and X. Jia. *Remote Sensing Digital Image Analysis*. Springer-Verlag Berlin Heidelberg, 4th edition, 2006.   [Cited in pags. 3, 4, 6, 8, 9, 15, 16, 18, 59, 67 and 68]

[13] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297. University of California Press, 1967.   [Cited in pag. 3]

[14] G. H. Ball and D. J. Hall. ISODATA. A novel method of data analysis and pattern classification. Technical report, Menlo Park: Stanford Research Institute, 1965.   [Cited in pag. 3]

[15] R. A. Schowengerdt. *Remote Sensing: Models and Methods for Image Processing*. Acadamice Press, 3rd edition, 2007.   [Cited in pags. 3 and 16]

[16] L. Samaniego, A. Bardossy, and K. Schulz. Suppervised classification of remote sensed imagery using a modified k-nn technique. *IEEE Transactions on Geoscience and Remote Sensing*, 46(7):2112–2125, July 2008.   [Cited in pags. 3 and 16]

[17] S. Subramanian, N. Gat, M. Sheffield, J. Barhen, and N. Toomarian. Methodology for hyperspectral image classification using novel neural network. In *Proceedings of SPIE Algorithms Multispectral Hyperspectral Imagery III*, volume 3071, pages 128–137, August 1997.   [Cited in pags. 3 and 16]

[18] H. Yang, F. V. D. Meer, W. Bakker, and Z. J. Tan. A back-propagation neural network for mineralogical mapping from aviris data. *International Journal of Remote Sensing*, 20(1):97–110, 1999.   [Cited in pags. 3 and 16]

[19] C. Hernandez-Espinosa, M. Fernandez-Redondo, and J. Torres-Sospedra. Some experiments with ensembles of neural networks for classification of hyperspectral images. In *Proceedings of the ISNN*, volume 1, pages 912–917, 2004.   [Cited in pags. 3 and 16]

[20] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson. Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. *IEEE Signal Processing Magazine*, 31(1):45–54, January 2014.   [Cited in pag. 3]

[21] J.M. Bioucas-Dias, A Plaza, G. Camps-Valls, P. Scheunders, N.M. Nasrabadi, and J. Chanussot. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and Remote Sensing Magazine*, 1(2):6–36, June 2013.   [Cited in pags. 3, 39, 62, 67 and 86]

[22] G. M. Foody and A. Mathur. The use of small training sets containing mixed pixels for accurate hard image classification: training on mixed spectral responses for classification by a SVM. *Remote Sensing of Environment*, 103(7):179–189, April 2006.   [Cited in pags. 3 and 16]

[23] G. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14(1):55–63, January 1968.   [Cited in pag. 3]

[24] J. A. Richards. Analysis of remotely sensed data: the formative decades and the future. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):422–432, March 2005.   [Cited in pag. 3]

[25] J. A. Gualtieri and R. F. Cromp. Support vector machines for hyperspectral remote sensing classification. In *Proceedings of SPIE*, volume 4584, pages 506–508, 1998. [Cited in pags. 3, 4 and 16]

[26] C. Huang, L. S. Davis, and J. R. Townshend. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4):725–749, February 2002.
[Cited in pags. 3, 4, 16 and 88]

[27] V. Vapnik and A. Chervonenkis. The necessary and sufficient conditions for consistency in the empirical risk minimization method. *Pattern Recognition and Image Analysis*, 1(3):283–305, 1991.
[Cited in pags. 3, 18, 69 and 88]

[28] B. Scholkopf and A. Smola. *Learning With Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press Series, Cambridge, MA, 2002. [Cited in pags. 3, 4 and 60]

[29] T.-F. Wu, C.-J. Lin, and R. C. Weng. Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, 5:975–1005, December 2004.
[Cited in pags. 4, 9, 18 and 88]

[30] C. Chang and C. Lin. LIBSVM: A library for support vector machines, 2009.
[Cited in pags. 4, 18, 54 and 88]

[31] F. Melgani and L. Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8):1778–1790, August 2004. [Cited in pags. 4, 16, 60 and 88]

[32] G. Camps-Valls and L. Bruzzone. Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 43(6):1351–1362, June 2005.
[Cited in pags. 4, 16, 18, 20, 60, 61 and 88]

[33] J. Zhu and T. Hastie. Kernel logistic regression and the import vector machine. *Journal of Computational and Graphical Statistics*, 14(1):185–205, March 2005. [Cited in pag. 4]

[34] R. Roscher, B. Waske, and W. Forstner. Incremental import vector machines for classifying hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 50(9):3463–3473, September 2012. [Cited in pag. 4]

[35] S. Suess, S. V. D. Linden, P.J. Leitao, A. Okujeni, B. Waske, and P. Hostert. Import vector machines for quantitative analysis of hyperspectral data. *IEEE Geoscience and Remote Sensing Letters*, 11(2):449–453, February 2014. [Cited in pag. 4]

[36] J. A. Benediktsson, J. Chanussot, and M. Fauvel. Multiple classifier systems in remote sensing: From basics to recent developments. *Mult. Classif. Syst.*, 4472:501–512, 2007. [Cited in pags. 4 and 16]

[37] P. Du, J. Xia, W. Zhang, K. Tan, Y. Liu, and S. Liu. Multiple classifier system for remote sensing image classification: A review. *Sensors*, 12:4764–4792, 2012. [Cited in pags. 4 and 16]

[38] J. Ham, Y. Chen, and M. M. Crawford. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):492–501, March 2005. [Cited in pags. 4 and 16]

[39] J. C.-W. Chan and D. Paelinckx. Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, 112(6):2299–3011, June 2008. [Cited in pags. 4 and 16]

111

[40] J. Xia, P. Du, X. He, and J. Chanussot. Hyperspectral remote sensing image classification based on rotation forest. *IEEE Geoscience and Remote Sensing Letters*, 11(1):239–243, January 2014.
[Cited in pag. 4]

[41] N. Chehata, L. Guo, and C. Mallet. Airborne LiDAR feature selection for urban classification using random forests. *International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences*, 38(3):207–2011, June 2009. [Cited in pag. 4]

[42] J. Li, J. Bioucas-Dias, and A. Plaza. Semi-supervised hyperspectral image segmentation using multinomial logistic regression with active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4085–4098, 2010. [Cited in pags. 4, 16, 19, 39, 51 and 68]

[43] J. Li, J. Bioucas-Dias, and A. Plaza. Hyperspectral image segmentation using a new Bayesian approach with active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 49(10):3947–3960, October 2011. [Cited in pags. 4, 16, 19, 22, 39, 51, 53 and 68]

[44] D. Böhning. Multinomial logistic regression algorithm. *Annals of the Institute of Statistical Mathematics*, 44(1):197–200, 1992. [Cited in pags. 4, 18 and 41]

[45] B. Krishnapuram, L. Carin, M.A.T. Figueiredo, and A.J. Hartemink. Sparse multinomial logistic regression: fast algorithms and generalization bounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(6):957–968, June 2005. [Cited in pags. 4 and 10]

[46] J. Bioucas-Dias and M. Figueiredo. Logistic regression via variable splitting and augmented lagrangian tools. Technical report, Instituto Superior Tecnico, Lisbon, Portugal, 2009.
[Cited in pags. 4 and 53]

[47] J. Bioucas-Dias and J. Nascimento. Hyperspectral subspace identification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(8):2435–2445, 2008. [Cited in pags. 4, 5 and 60]

[48] C.-I. Chang. Orthogonal subspace projection (OSP) revisited: A comprehensive study and analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):502–518, 2005.
[Cited in pags. 4, 5 and 60]

[49] H. Bagan, Y. Yasuoka, T. Endo, Xiaohui Wang, and Z. Feng. Classification of airborne hyperspectral data based on the average learning subspace method. *IEEE Geoscience and Remote Sensing Letters*, 5(3):368–372, July 2008. [Cited in pag. 4]

[50] J.-M. Yang, B.-C. Kuo, P.-T. Yu, and C.-H. Chuang. A dynamic subspace method for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 48(7):2840–2853, July 2010. [Cited in pag. 4]

[51] J. Li, J. Bioucas-Dias, and A. Plaza. 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, 2012.
[Cited in pags. 4, 5, 10, 16, 18, 19, 21, 22, 25, 39, 40, 41, 51, 53, 54, 60, 62, 68, 69, 75, 87, 88 and 90]

[52] J. C. Harsanyi and C.-I. Chang. Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection. *IEEE Transactions on Geoscience and Remote Sensing*, 32(4):779–785, 1994. [Cited in pags. 5 and 60]

112

[53] C. Chang, X. Zhao, M. L. G. Althouse, and J. J. Pan. Least squares subspace projection approach to mixed pixel classification for hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 36(3):898–912, 1998.   [Cited in pags. 5 and 60]

[54] R. Larsen, M. Arngren, P. Hansen, and A. Nielsen. Kernel based subspace projection of near infrared hyperspectral images of maize kernels. *Image Analysis*, pages 560–569, 2009.
[Cited in pags. 5 and 60]

[55] W. Li, E.W. Tramel, S. Prasad, and J.E. Fowler. Nearest regularized subspace for hyperspectral classification. *IEEE Transactions on Geoscience and Remote Sensing*, 52(1):477–489, January 2014.   [Cited in pag. 5]

[56] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems.* V. H. Winston & Sons, Washington, DC, USA, 1977.   [Cited in pag. 5]

[57] R. Kettig and D. Landgrebe. Classification of multispectral image data by extraction and classification of homogenous objects. *IEEE Transactions on Geoscience and Remote Sensing*, 14:19–26, 1976.   [Cited in pags. 5, 6 and 16]

[58] H. Ghassemian and D. Landgrebe. Object-oriented feature extraction method for image data compaction. *IEEE Control System Magazine*, 8(3):42–46, June 1988.   [Cited in pags. 5 and 16]

[59] A. Plaza, J. A. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, J. A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni. Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment*, 113:110–122, 2009.   [Cited in pags. 5, 16, 59, 60, 61, 67 and 86]

[60] S. M. Jong and F. V. D. Meer. *Remote Sensing Image Analysis: Including the Spatial Domain.* Kluwer Academic Publishers, 2004.   [Cited in pags. 5 and 16]

[61] D. Li, G. Zhang, Z. Wu, and L. Yi. An edge embedded marker-based watershed algorithm for high spatial resolution remote sensing image segmentation. *IEEE Transactions on Image Processing*, 19(10):2781–2787, October 2010.   [Cited in pag. 5]

[62] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson. Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recognition*, 43(7):2367 – 2379, 2010.   [Cited in pag. 5]

[63] Y. Tarabalka and J. A. Benediktsson. Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques. *IEEE Transactions on Geoscience and Remote Sensing*, 47(8):761–776, March 2009.   [Cited in pags. 5 and 6]

[64] H. Yang, Q. Du, and B. Ma. Decision fusion on suppervised and unsuppervised classifiers for hyperspectral imagery. *IEEE. Geoscience and Remote Sensing Letters*, 7(4):875–879, October 2010.   [Cited in pags. 5 and 6]

[65] L. Bruzzone and L. Carlin. A multilevel context-based system for classification of very high spatial resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, 44(9):2587–2600, September 2006.   [Cited in pag. 5]

[66] H.G. Akcay and S. Aksoy. Automatic detection of geospatial objects using multiple hierarchical segmentations. *IEEE Transactions on Geoscience and Remote Sensing*, 46(7):2097–2111, July 2008. [Cited in pag. 5]

[67] Y. Tarabalka, J.C. Tilton, J. A. Benediktsson, and J. Chanussot. A marker-based approach for the automated selection of a single segmentation from a hierarchical set of image segmentations. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(1):262–272, February 2012. [Cited in pag. 5]

[68] V. D. Linden, A. Janz, B. Waske, M. Eiden, and P. Hostert. Classifying segmented hyperspectral data from a heterogeneous urban environment using support vector machines. *Journal of Applied Remote Sensing*, 1, 2007. [Cited in pags. 6 and 16]

[69] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson. Segmentation and classification of hyperspectral images using minimum spaning forest growing from automatically selected markers. *IEEE. Transactions on System, Man, and Cybernetics*, 40(5):1267–1279, December 2010. [Cited in pag. 6]

[70] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton. Multiple spectral-spatial classification approach for hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4122–4132, November 2010. [Cited in pag. 6]

[71] R.M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3(6):610–621, November 1973. [Cited in pag. 6]

[72] F. Tsai and J.-S. Lai. Feature extraction of hyperspectral image cubes using three-dimensional gray-level cooccurrence. *IEEE Transactions on Geoscience and Remote Sensing*, 51(6):3504–3513, June 2013. [Cited in pag. 6]

[73] T.C. Bau, S. Sarkar, and G. Healey. Hyperspectral region classification using a three-dimensional gabor filterbank. *IEEE Transactions on Geoscience and Remote Sensing*, 48(9):3457–3464, September 2010. [Cited in pag. 6]

[74] Y. Qian, M. Ye, and J. Zhou. Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features. *IEEE Transactions on Geoscience and Remote Sensing*, 51(4):2276–2291, April 2013. [Cited in pag. 6]

[75] X. Guo, X. Huang, and L. Zhang. Three-dimensional wavelet texture feature extraction and classification for multi/hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 11(12):2183–2187, December 2014. [Cited in pag. 6]

[76] P. Soille. *Morphological Image Analysis, Principles and Applications*. Berlin, Germany: Springer-Verlag, 2nd edition, 2003. [Cited in pag. 6]

[77] M. Pesaresi and J. A. Benediktsson. A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(2):309 –320, 2001. [Cited in pag. 6]

[78] J. A. Benediktsson, J. A. Palmason, and J.R. Sveinsson. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):480 – 491, 2005. [Cited in pag. 6]

114

[79] E. J. Breen and R. Jones. Attribute openings, thinnings, and granulometries. *Computer Vision and Image Understanding*, 64(3):377 – 389, 1996.   [Cited in pag. 6]

[80] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone. Morphological attribute profiles for the analysis of very high resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, 48(10):3747 –3762, 2010.   [Cited in pags. 6, 86 and 87]

[81] P. Ghamisi, J. A. Benediktsson, and J.R. Sveinsson. Automatic spectral-spatial classification framework based on attribute profiles and supervised feature extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 52(9):5771–5782, September 2014.   [Cited in pags. 6 and 68]

[82] B. Song, J. Li, M. D. Mura, P. Li, A. Plaza, J.M. Bioucas-Dias, J. A. Benediktsson, and J. Chanussot. Remotely sensed image classification using sparse representations of morphological attribute profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 52(8):5122–5136, August 2014.   [Cited in pag. 6]

[83] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone. Extended profiles with morphological attribute filters for the analysis of hyperspectral data. *International Journal of Remote Sensing*, 31(22):5975–5991, 2010.   [Cited in pag. 7]

[84] P. Salembier, A. Oliveras, and L. Garrido. Antiextensive connected operators for image and sequence processing. *IEEE Transactions on Geoscience and Remote Sensing*, 7(4):555 –570, 1998.   [Cited in pag. 7]

[85] X. Huang and L. Zhang. A comparative study of spatial approaches for urban mapping using hyperspectral ROSIS images over Pavia city, northern Italy. *International Journal of Remote Sensing*, 30(12):3205–3221, 2009.   [Cited in pag. 7]

[86] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson. Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 46(11):3804–3814, November 2008.   [Cited in pags. 7 and 16]

[87] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. CalpeMaravilla. Composite kernels for hyperspectral image classification. *IEEE. Geoscience and Remote Sensing Letters*, 3(1):93–97, January 2006.   [Cited in pags. 7 and 68]

[88] M. Fauvel, J. Chanussot, and J. A. Benediktsson. Adaptive pixel neighborhood definition for the classification of hyperspectral images with support vector machines and composite kernel. In *Proceedings of the 15th IEEE International Conference on Image Processing*, pages 1884–1887, October 2008.   [Cited in pag. 7]

[89] S. Velasco-Forero and V. Manian. Improving hyperspectral image classification using spatial preprocessing. *IEEE. Geoscience and Remote Sensing Letters*, 6(2):297–301, April 2009.   [Cited in pags. 7, 16 and 68]

[90] J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson. Generalized composite kernel framework for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 51(9):4816–4829, 2013.   [Cited in pags. 7 and 68]

[91] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton. Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE*, 101(3):652–675, March 2013.
[Cited in pags. 8, 59, 69 and 86]

[92] J. O. Eklundh, H. Yamamoto, and A. Rosenfeld. A relaxation method for multispectral pixel classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2(1):72–75, January 1980.
[Cited in pag. 8]

[93] B. R. Cuenca, J. A. Malpica, and M. C. Alonso. A spatial contextual postclassification method for preserving linear objects in multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 51(1):395–406, January 2013.
[Cited in pags. 8, 9 and 68]

[94] L. Pelkowitz. A continuous relaxation labeling algorithm for Markov random fields. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(3):709–715, 1990.
[Cited in pags. 8 and 68]

[95] A. Rosenfeld, R. Hummel, and S. Zucker. Scene labelling by relaxation algorithms. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-6(6):420–433, June 1976.
[Cited in pags. 8 and 68]

[96] A. Reigber, M. Jäger, M. Neumann, and L. Ferro-Famil. Classifying polarimetric SAR data by combining expectation methods with spatial context. *International Journal of Remote Sensing*, 31(3):727–744, 2010.
[Cited in pag. 9]

[97] S. Z. Li. *Markov Random Field Modeling in Image Analysis.* Springer-Verlag London, 3rd edition, 2009.
[Cited in pags. 9, 16, 19, 68 and 87]

[98] R.C. Dubes, A.K. Jain, S.G. Nadabar, and C.C. Chen. MRF model-based algorithms for image segmentation. In *Proceedings of the 10th International Conference on Pattern Recognition*, volume i, pages 808–814, 1990.
[Cited in pags. 9 and 68]

[99] X. Jia and J. A. Richards. Managing the spectral-spatial mix in context classification using Markov random fields. *IEEE Geoscience and Remote Sensing Letters*, 5(2):311–314, 2008.
[Cited in pags. 9 and 68]

[100] Q. Jackson and D. Landgrebe. Adaptive Bayesian contextual classification based on Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 40(11):2454–2463, November 2002.
[Cited in pag. 9]

[101] A. Farag, R. Mohamed, and A. El-Baz. A unified framework for map estimation in remote sensing image segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 43(7):1617–1634, July 2005.
[Cited in pags. 10, 16, 19, 22 and 68]

[102] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson. SVM- and MRF-based method for accurate classification of hyperspectral images. *IEEE. Geoscience and Remote Sensing Letters*, 7(4):736–740, October 2010.
[Cited in pags. 10, 16, 51 and 68]

[103] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.
[Cited in pag. 10]

[104] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, November 1984. [Cited in pag. 10]

[105] B. Zhang, S. Li, X. Jia, L. Gao, and M. Peng. Adaptive Markov random field approach for classification of hyperspectral imagery. *IEEE. Geoscience and Remote Sensing Letters*, 8(5):973–977, September 2011. [Cited in pags. 10 and 68]

[106] G. Moser and S.B. Serpico. 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, May 2013. [Cited in pag. 10]

[107] J. Bai, S. Xiang, and C. Pan. A graph-based classification method for hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 51(2):803–817, February 2013. [Cited in pag. 10]

[108] L. Sun, Z. Wu, J. Liu, L. Xiao, and Z. Wei. Supervised spectral-spatial hyperspectral image classification with weighted Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 53(3):1490–1503, March 2015. [Cited in pag. 10]

[109] J. Li, J.M. Bioucas-Dias, and A. Plaza. Spectral-spatial classification of hyperspectral data using loopy belief propagation and active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 51(2):844–856, February 2013. [Cited in pag. 10]

[110] Y. Boykov, O. Veksler, and R. Zabih. Efficient approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1222–1239, November 2001. [Cited in pags. 10, 19, 22 and 71]

[111] L. Xu and J. Li. Bayesian classification of hyperspectral imagery based on probabilistic sparse representation and Markov random field. *IEEE Geoscience and Remote Sensing Letters*, 11(4):823–827, April 2014. [Cited in pag. 10]

[112] W. Li, S. Prasad, and J.E. Fowler. Hyperspectral image classification using Gaussian mixture models and Markov random fields. *IEEE Geoscience and Remote Sensing Letters*, 11(1):153–157, January 2014. [Cited in pag. 10]

[113] J. Lafferty. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. pages 282–289. Morgan Kaufmann, 2001. [Cited in pag. 10]

[114] P. Zhong and R. Wang. Learning conditional random fields for classification of hyperspectral images. *IEEE Transactions on Image Processing*, 19(7):1890–1907, July 2010. [Cited in pag. 10]

[115] P. Zhong and R. Wang. Modeling and classifying hyperspectral imagery by crfs with sparse higher order potentials. *IEEE Transactions on Geoscience and Remote Sensing*, 49(2):688–705, February 2011. [Cited in pag. 10]

[116] P. Zhong and R. Wang. Jointly learning the hybrid crf and mlr model for simultaneous denoising and classification of hyperspectral imagery. *IEEE Transactions on Neural Networks and Learning Systems*, 25(7):1319–1334, July 2014. [Cited in pag. 10]

[117] P. Ghamisi, J. A. Benediktsson, and M.O. Ulfarsson. Spectral-spatial classification of hyperspectral images based on hidden Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 52(5):2565–2574, May 2014. [Cited in pag. 10]

[118] A. Villa, J. Chanussot, J. A. Benediktsson, and C. Jutten. Spectral unmixing for the classification of hyperspectral images at a finer spatial resolution. *IEEE Journal of Selected Topics in Signal Processing*, 5(3):521–533, June 2011. [Cited in pag. 16]

[119] X. Jia, C. Dey, D. Fraser, L. Lymburner, and A. Lewis. Controlled spectral unmixing using extended support vector machines. In *Proceedings of the 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Reykjavik, Iceland*, pages 1–4, June 2010. [Cited in pag. 16]

[120] P. C. Smits. Multiple classifier systems for supervised remote sensing image classification based on dynamic classifier selection. *IEEE Transactions on Geoscience and Remote Sensing*, 40:801–813, 2002. [Cited in pag. 16]

[121] G. M. Foody, D. S. Boyd, and C. Sanchez-Hernandez. Mapping a specific class with an ensemble of classifiers. *International Journal of Remote Sensing*, 28:1733–1746, 2007. [Cited in pag. 16]

[122] S. Kumar, J. Ghosh, and M. M. Crawford. Hierarchical fusion of multiple classifiers for hyperspectral data analysis. *Pattern Analysis & Applications*, 5:210–220, 2002. [Cited in pag. 16]

[123] X. Ceamanos, B. Waske, J. A. Benediktsson, J. Chanussot, M. Fauvel, and J. R. Sveinsson. A classiler ensemble based on fusion of support vector machines for classifying hyperspectral data. *International Journal of Image and Data Fusion*, 1:293–307, 2010. [Cited in pag. 16]

[124] Y. Wang, R. Niu, and X. Yu. Anisotropic diffusion for hyperspectral imagery enhancement. *IEEE Sensors Journal*, 10(3):469–477, March 2010. [Cited in pags. 16 and 68]

[125] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza. Sparse unmixing of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 49(6):2014–2039, 2011. [Cited in pags. 17 and 21]

[126] C.-J. Lin H.-T. Lin and R. C. Weng. A note on Platt's probabilistic outputs for support vector machines. *Department of Computer Science, National Taiwan University, Taipei, Taiwan*, 2003. [Cited in pags. 18, 19 and 88]

[127] P. Clifford. Markov random fields in statistics. In Geoffrey Grimmett and Dominic Welsh, editors, *Disorder in Physical Systems: A Volume in Honour of John M. Hammersley*, pages 19–32. Oxford University Press, Oxford, 1990. [Cited in pag. 19]

[128] J. A. Benediktsson and P. H. Swain. Consensus theoretic classification methods. *IEEE Transactions on Systems, Man, and Cybernetics*, 22:688–704, 1992. [Cited in pag. 21]

[129] J. A. Benediktsson and I. Kanellopoulos. Classification of multisource and hyperspectral data based on decision fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3):1367–1377, 1999. [Cited in pag. 21]

[130] J. Nascimento and J. Bioucas-Dias. Vertex Component Analysis: A Fast Algorithm to Unmix Hyperspectral Data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4):898–910, 2005. [Cited in pag. 23]

[131] G. M. Foody. Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering & Remote Sensing*, 70(5):627–633, May 2004. [Cited in pags. 29 and 92]

[132] Q. Cheng, P.K. Varshney, and M.K. Arora. Logistic regression for feature selection and soft classification of remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 3(4):491–494, October 2006. [Cited in pag. 39]

[133] M. Pal and G.M. Foody. Evaluation of SVM, RVM and SMLR for accurate image classification with limited ground data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(5):1344–1355, October 2012. [Cited in pag. 39]

[134] J.S. Borges, J.M. Bioucas-Dias, and A. R. S. Marcal. Bayesian hyperspectral image segmentation with discriminative class learning. *IEEE Transactions on Geoscience and Remote Sensing*, 49(6):2151–2164, June 2011. [Cited in pag. 39]

[135] P. Zhong, P. Zhang, and R. Wang. Dynamic learning of SMLR for feature selection and classification of hyperspectral data. *IEEE Geoscience and Remote Sensing Letters*, 5(2):280–284, April 2008. [Cited in pag. 39]

[136] M. A. T. Figueiredo. Bayesian image segmentation using Gaussian field priors. In *Energy Minimization Methods Computer Vision and Pattern Recognition*, pages 74–89, 2005. [Cited in pag. 40]

[137] B. Krishnapuram, L. Carin, M. Figueiredo, and A. Hartemink. Sparse multinomial logistic regression: Fast algorithms and generalization bounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(6):957–968, 2005. [Cited in pag. 41]

[138] J. Bioucas-Dias, F. Condessa, and J. Kovacevic. Alternating direction optimization for image segmentation using hidden Markov measure field models. In *Proceedings of the IS&T/SPIE Electronic Imaging*, pages 90190P–90190P. International Society for Optics and Photonics, 2014. [Cited in pags. 51 and 68]

[139] M. Khodadadzadeh, J. Li, A. Plaza, and J.M. Bioucas-Dias. A subspace-based multinomial logistic regression for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 11(12):2105–2109, December 2014. [Cited in pags. 51, 53 and 54]

[140] I. Dopido, J. Li, P.R. Marpu, A Plaza, J.M. Bioucas Dias, and J.A Benediktsson. Semisupervised self-learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 51(7):4032–4044, July 2013. [Cited in pag. 51]

[141] J. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2):354–379, 2012. [Cited in pags. 51, 60 and 86]

[142] M. Soltanolkotabi and E. J. Candès. A geometric analysis of subspace clustering with outliers. *The Annals of Statistics*, 40(4):2195–2238, August 2012. [Cited in pag. 52]

[143] M. Soltanolkotabi, E. Elhamifar, and E. J. Candes. Robust subspace clustering. *The Annals of Statistics*, 42(2):669–699, 2014. [Cited in pags. 52 and 53]

[144] E. Elhamifar and R. Vidal. Sparse subspace clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2790–2797, June 2009. [Cited in pag. 53]

[145] C. Debes, A Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. van Kasteren, Wenzhi Liao, R. Bellens, A Pizurica, S. Gautama, W. Philips, S. Prasad, Qian Du, and F. Pacifici. Hyperspectral and lidar data fusion: Outcome of the 2013 grss data fusion contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):2405–2418, June 2014. [Cited in pag. 54]

[146] J. Chanussot, M. M. Crawford, and B.-C. Kuo. Foreword to the special issue on hyperspectral image and signal processing. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):3871–3876, 2010. [Cited in pag. 59]

[147] A. Plaza, Q. Du, J. Bioucas-Dias, X. Jia, and F. Kruse. Foreword to the special issue on spectral unmixing of remotely sensed data. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11):4103–4110, 2011. [Cited in pag. 60]

[148] A. Plaza, P. Martinez, R. Perez, and J. Plaza. A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 42(3):650–663, 2004. [Cited in pag. 60]

[149] N. Keshava and J. F. Mustard. Spectral unmixing. *IEEE Signal Processing Magazine*, 19(1):44–57, 2002. [Cited in pag. 60]

[150] J. J. Settle. On the relationship between spectral unmixing and subspace projection. *IEEE Transactions on Geoscience and Remote Sensing*, 34:1045–1046, 1996. [Cited in pag. 60]

[151] M. Diani N. Acito and G. Corsini. Hyperspectral signal subspace identification in the presence of rare signal components. *IEEE Transactions on Geoscience and Remote Sensing*, 48(4):1940–1954, 2010. [Cited in pag. 60]

[152] V. Vapnik. *Statistical Learning Theory*. John Wiley, New York, 1998. [Cited in pags. 60 and 61]

[153] B. Scholkopf, A. J. Smola, and K.-R Muller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10:1299–1319, 1998. [Cited in pag. 61]

[154] M. C. Zhang, R. M. Haralick, and J. B. Campbell. Multispectral image context classification using stochastic relaxation. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(1):128–140, January/February 1990. [Cited in pag. 62]

[155] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone. Extended profiles with morphological attribute filters for the analysis of hyperspectral data. *International Journal of Remote Sensing*, 31(22):5975–5991, December 2010. [Cited in pag. 68]

[156] I. Yildirim, O. K. Ersoy, and B. Yazgan. Improvement of classification accuracy in remote sensing using morphological filter. *Advances in Space Research*, 36(5):1003–1006, 2005. [Cited in pag. 68]

[157] P.-F. Hsieh and D. Landgrebe. *Classification of high dimensional data*. PhD thesis, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, 1998. [Cited in pag. 68]

[158] J. A. Richards, D.A. Landgrebe, and P.H. Swain. Pixel labeling by supervised probabilistic relaxation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-3(2):188–191, 1981. [Cited in pag. 68]

[159] L. Sun, Z. Wu, J. Liu, and Z. Wei. Supervised hyperspectral image classification using sparse logistic regression and spatial-tv regularization. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1019–1022, July 2013. [Cited in pags. 68 and 70]

[160] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, July 1990. [Cited in pag. 70]

[161] S. Bagon. Matlab wrapper for graph cut, December 2006. [Cited in pag. 71]

[162] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock. Imaging spectrometry for Earth remote sensing. *Science*, 228:1147–1153, 1985. [Cited in pag. 86]

[163] H.S. Lee and N.H. Younan. DTM extraction of LiDAR returns via adaptive processing. *IEEE Transactions on Geoscience and Remote Sensing*, 41(9):2063–2069, September 2003. [Cited in pag. 86]

[164] J. Jung, E. Pasolli, S. Prasad, J. C. Tilton, and M. M. Crawford. A framework for land cover classification using discrete return LiDAR data: Adopting pseudo-waveform and hierarchical segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(2):491–502, February 2014. [Cited in pag. 86]

[165] H. Wang, C. Glennie, and S. Prasad. Voxelization of full waveform LiDAR data for fusion with hyperspectral imagery. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 3407–3410, July 2013. [Cited in pag. 86]

[166] D. Lemp and U. Weidner. Segment-based characterization of roof surfaces using hyperspectral and laser scanning data. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 7, pages 4942–4945, July 2005. [Cited in pag. 86]

[167] M. Dalponte, L. Bruzzone, and D. Gianelle. Fusion of hyperspectral and LiDAR remote sensing data for classification of complex forest areas. *IEEE Transactions on Geoscience and Remote Sensing*, 46(5):1416–1427, May 2008. [Cited in pag. 86]

[168] M. Pedergnana, P.R. Marpu, M. D. Mura, J. A. Benediktsson, and L. Bruzzone. Classification of remote sensing optical and LiDAR data using extended attribute profiles. *IEEE Journal of Selected Topics in Signal Processing*, 6(7):856–865, November 2012. [Cited in pags. 86 and 90]

[169] W. Liao, R. Bellens, A. Pizurica, S. Gautama, and W. Philips. Graph-based feature fusion of hyperspectral and LiDAR remote sensing data using morphological features. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 7, pages 4942–4945, July 2013. [Cited in pag. 86]

[170] S. Prasad and L.M. Bruce. Overcoming the small sample size problem in hyperspectral classification and detection tasks. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 5, pages V – 381–V – 384, July 2008. [Cited in pag. 86]

[171] G. Licciardi, F. Pacifici, D. Tuia, S. Prasad, T. West, F. Giacco, J. Inglada, E. Christophe, J. Chanussot, and P. Gamba. Decision fusion for the classification of hyperspectral data: Outcome of the 2008 GRS-S data fusion contest. *IEEE Transactions on Geoscience and Remote Sensing*, 47(11):3857–3865, 2009. [Cited in pag. 86]

[172] J. A. Benediktsson, J.R. Sveinsson, and P.H. Swain. Hybrid consensus theoretic classification. *IEEE Transactions on Geoscience and Remote Sensing*, 35(4):833–843, July 1997. [Cited in pags. 87 and 89]

[173] C. E. Bonferroni. Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8:3–62, 1936. [Cited in pag. 92]