UNIVERSIDAD  DE EXTREMADURA

Tesis Doctoral

Nuevas técnicas de clustering como paso previo a la clasificacin de imgenes hiperespectrales

New Clustering Techniques prior to Hyperspectral Image Classification

Autor:
Yi Liu

Directores:
Antonio Plaza Miguel
Jun Li

DEPARTAMENTO DE TECNOLOGÍA DE LOS COMPUTADORES Y DE LAS COMUNICACIONES

Conformidad de los Directores:

Fdo:
Dr. Antonio Plaza Miguel

Fdo:
Dr. Jun Li

2017

Resumen

Las imágenes hiperespectrales proporcionan una rica fuente de información a la hora de reconocer objetos con características espectrales similares asociadas a materiales diferentes en la superficie de la Tierra. La elevada dimensionalidad espectral de las imágenes hiperespectrales introduce importantes desafíos a la hora de abordar su clasificación. En particular, estas imágenes proporcionan información espectral muy detallada que permite realizar una caracterización muy precisa, pero que también introduce gran redundancia que complica el análisis desde un punto de vista computacional y algorítmico. En este sentido, las técnicas tradicionales para selección de bandas y extracción de características normalmente seleccionan las bandas espectrales más representativas de la imagen o bien proyectan la imagen original en un subespacio, de forma que puede perderse información relevante para el proceso de clasificación, incluyendo información física. Con esta limitación en mente, en el presente trabajo de tesis doctoral hemos desarrollado un conjunto de nuevas técnicas de *clustering* que pueden utilizarse para particionar espectralmente la imagen con carácter previo al proceso de clasificación. Por un lado, consideramos características intrínsecas de los datos hiperespectrales, tales como la gran correlación espectral existente entre las bandas de la imagen o las características físicas del instrumento utilizado para obtener la imagen hiperespectral. Por otra parte, también hemos desarrollado un nuevo método de particionamiento espectral guiado por las clases presentes en la escena, el cual utiliza técnicas de selección de bandas para construir un sistema de clasificación múltiple que emplea múltiples particiones relevantes a las diferentes clases. Además de la redundancia espectral causada por la elevada dimensionalidad de las imágenes en este dominio, otro importante obstáculo en el proceso de clasificación es la presencia de píxeles mezcla. Estos píxeles suelen ubicarse en los bordes de los diferentes objetos presentes en la escena, introduciendo importantes problemas en el proceso de clasificación. La presencia de ruido y píxeles anómalos agrava la situación, impidiendo la generación de mapas de clasificación homogéneos y con contornos bien definidos como sería deseable. Para resolver este problema, en el presente trabajo de tesis doctoral proponemos nuevas técnicas capaces de particionar la imagen hiperespectral en el dominio espacial mediante la utilización del concepto de superpíxel, que ha sido ampliamente utilizado en la literatura para incrementar la precisión en la clasificación. Sin embargo, en la literatura no se ha propuesto hasta la fecha un método capaz de integrar de forma natural la información espacial en el proceso de clasificación mediante sobresegmentación, debido a que el proceso resulta altamente costoso desde el punto de vista computacional. Para abordar

este problema e incorporar de forma sencilla la información proveniente de múltiples particiones espaciales en el proceso de clasificación, hemos desarrollado un nuevo método que permite realizar este proceso (NP-complejo) mediante una técnica de relajación lineal, que aproxima el problema original en un dominio variable compacto. De esta forma, resulta mucho más sencillo construir funciones convexas con información *a priori*, que ofrecen un mayor grado de flexibilidad a la hora de incluir información espacial (proveniente de superpíxels) en el modelo de clasificación. Además, en la presente tesis doctoral también exploramos la posibilidad de explotar y validar esta nueva estrategia en el contexto de problemas que involucran datos obtenidos mediante múltiples sensores para la observación remota de la tierra, incluyendo tanto sensores hiperespectrales como sensores de tipo LiDAR.

Palabras Clave: Observación remota de la Tierra, imágenes hiperespectrales, clasificación, particionamiento espectral y espacial, *clustering*, superpíxels, múltiples sensores.

Abstract

Remotely sensed hyperspectral images provide a rich source of information for recognizing objects that exhibit representative spectral signatures associated with different materials on the surface of the Earth. The high spectral dimensionality of hyperspectral images poses important challenges for data interpretation. It provides detailed spectral information which can uncover relevant information about materials, but it also presents huge spectral redundancy complicating the analysis from a computational and algorithmic standpoint. Traditional feature selection/extraction techniques either select only the relevant bands that are believed to carry most of the information in the hyperspectral image or transform the original image into a feature space, thus discarding spectral bands and the associated physical information. Having this challenge in mind, in this thesis we have developed a set of new clustering techniques that can be used for spectral partitioning prior to hyperspectral image classification. On the one hand, we consider the intrinsic characteristics of the hyperspectral data, such as the high correlation among the spectral bands as well as the distinct characteristics that are driven by the imaging spectrometer. On the other hand, we have also designed a new class-oriented spectral partitioning framework which exploits band selection techniques in order to construct a multiple classifier system with partitioned band subgroups that are relevant to different classes. Apart from the aforementioned spectral redundancy that is caused by the high spectral dimensionality of hyperspectral images, another important challenge in the analysis of hyperspectral data is the presence of mixed pixels. These pixels are usually present along the boundaries of different objects, leading to huge challenges in labeling problems. The presence of noise and trivial, isolated pixels also hinders greatly the obtainment of smooth and well-contoured classification maps, which are usually desired. Taking these issues into account, in this thesis we deal with the problem via partitioning the three-dimensional hyperspectral cube data in the spatial (image) domain by superpixelizing techniques that have been proven successful in promoting classification performance. However, a straightforward using of over-segmented superpixels has been seldom explored, since it leads to a very complex (NP-hard) problem due to its discrete nature. In order to reinforce the information of spatial partitions into the classification model, we have attacked the aforementioned NP-hard problem via a linear relaxation technique, which approximates the original problem into a compact variable domain. Thus, it turns much easier to construct convex prior functions that provide more flexibly while including rich information (such as over-partitioned superpixels) into the classification model. Furthermore, in this thesis we have

also exploited and validated this strategy in the context of multi-source remote sensing data classification scenarios involving both hyperspectral and LiDAR data.

Keywords: Remote sensing, hyperspectral imaging, image classification, spectral and spatial partitioning, clustering superpixels, multi-source problems.

Acknowledgement

This thesis work has been developed with the support of the Chinese Scholarship Council (CSC) of the Ministry of Education of the People's Republic of China. The applicant was awarded a four years scholarship for the development of this thesis work. CSC is gratefully acknowledged for this outstanding support.

I would like to express my great appreciation and gratitude to my advisors, Profs. Antonio Plaza and Jun Li, for their continuous support, motivation and patience during these four years. This guidance helped me significantly in my research and in the writing of this thesis. Meanwhile, I also would like to express my gratitude to Prof. José Bioucas-Dias, for his generous support and terrific mentorship, for his patience, encouragement, and immense knowledge. These great supports, all together, play a formative role in my early academic years, helping and motivating me to focus and meanwhile to grasp a thorough scope of research, encouraging me to pursue the goals and career.

Besides, I would like to thank my fellow labmates and colleagues: Javier Plaza, Sergio Bernabé, Gabriel Martín, Jorge Sevilla, Inmaculada Dópido, Luis Ignacio Jiménez, Mahdi Khodadazadeh, Mario Haut, Mercedes Paoletti, and the wonderful lab visitors: Xiuping Jia, Alp Ertürk, Lianru Gao, Zebin Wu, Kun Tan, Dongmei Song, Hongyan Zhang, Xiong Xu, Yanli Sun, Victor Ayma, etc., for their great encouragement and discussions related to my research, and for all the fun we have had in the last four years. Also, I would like to thank all my friends in Cáceres and Lisbon, including David, Mayra, José, Milad, Lina, Zhiqi, Dongyan, Hongpan, Maria, Joshin, Miguel, Filipe, Alva, Christovao, etc. for their great companionship and support.

I would like to send my special appreciation and thanks to my family, my parents and my sisters for supporting me spiritually throughout writing this thesis, and to my girlfriend Hong Wang for her gentle, empathetic and beautiful company.

Contents

List of Figures

Table Index

Chapter 1

Introduction

1.1 Context and Motivations

The work developed in this thesis is part of the actual research lines of the Hyperspectral Computing Laboratory (HyperComp) research group at the Department of Technology of Computers and Communications, University of Extremadura, Spain. This work is focused on the development of new clustering techniques prior to hyperspectral image classification.

Hyperspectral imaging, a discipline related with both spectroscopy and photography, collects hundreds of images at different wavelength channels for the same area on the surface of the Earth. The pixel in a hyperspectral image is thus a high-dimensional vector, comprising observations spanning from the visible to the infra-red region of the electromagnetic spectrum. Common applications of hyperspectral image processing include target detection, material/class identification, information retrieval and extraction, etc. Physically, different materials and objects present distinct spectral reflectance characteristics, and they can be distinguished via their associated spectral signatures [2]. In fact, the ability to identify the spectral signatures of the observed image objects provides the potential to distinguish them in the same scene. Since hyperspectral images provide narrow spectral bands, they can provide information that is much more detailed than multispectral images (with typically tens of bands), very high spatial resolution images, etc. With hyperspectral images, a very accurate discrimination of different materials is hence possible, making it possible for instance to distinguish between the same at different growing seasons. A hyperspectral image is generally represented as a three-dimensional data cube, with two spatial dimensions and one spectral dimension (see Fig. 1.1). Hyperspectral images have been widely exploited in many different fields, including environmental monitoring, agriculture, disaster tracking, material recognition, disease diagnosis, among many others [3]. In Earth observation, hyperspectral images provide a rich source of information that has been exploited in many different contexts, including ecology, environmental science, geoinformatics, etc. Besides, hyperspectral data have also been exploited in the context of multi-temporal, multi-sensor and multi-resolution problems [4–6].

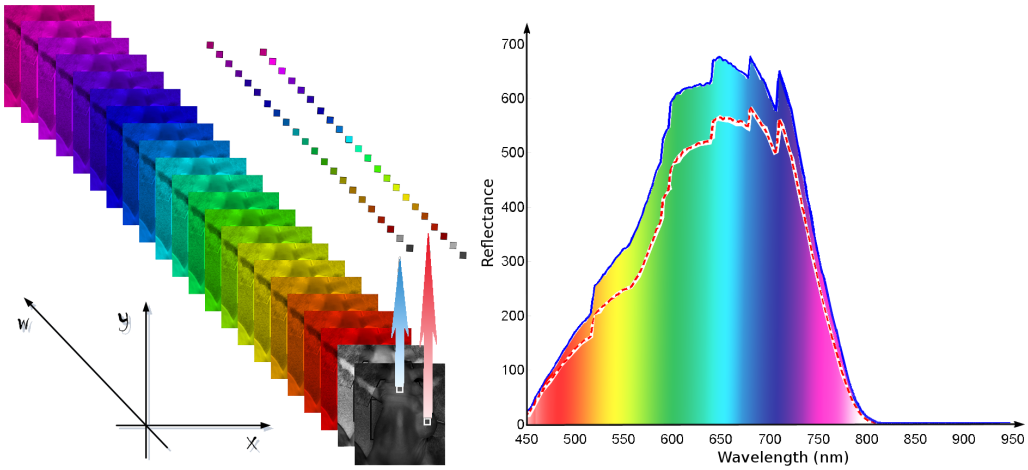


Figure 1.1: Graphical representation of a hyperspectral image as a three-dimensional image cube.

When it comes to specific applications, hyperspectral images have been widely used in (but not limited to) the following processing tasks:

- *Image labeling.* Image labeling has been one of the most important methods trying to interpret hyperspectral images via the assignment of different classes or categories (labels) to pixels. Given a set of hyperspectral pixel vectors, labeling consists of assigning a label to each pixel of the image by different criteria or models [7]. Several techniques have been developed for this purpose. Clustering, as an unsupervised method, takes the advantage of the intrinsic characteristics of the hyperspectral data itself to recognize and categorize the most similar or closest samples into multiple cliques. In this case, it is also called image spatial partitioning/segmentation when the pixel homogeneity in the spatial domain is also considered [8–10]. Supervised classification, on the other hand, intends to identify the land cover classes present in the entire image by using a set of previously available and highly representative training samples for each land cover type [3, 7]. Semi-supervised classification and active learning techniques are also often considered in order to deal with complicated situations, especially when the training samples are limited [11]. Finally, change detection can also be considered as a labeling problem (often in a binary scenario). This technique has been widely for multi-temporal data analysis [12, 13].

- *Object detection.* Automatic extraction of man-made objects, such as buildings, roads [14], anomalous targets, marine oil spills [15], civilian search and rescue (such as airplane wreckage) [16, 17], target monitoring after earthquakes [18], hurricane disasters [19], and so forth, is of paramount importance for supporting human activities. Object detection can often be performed with one-class classification, also known as *unary classification*, which can be viewed as a particular case of the labeling problem where only some parts of a hyperspectral image are concerned as

single, unified class. The classifier adopted in this case needs to learn from a training set containing only the objects of that class single class, which makes it a more difficult problem. Generally, the developed techniques and algorithms mainly focus on artificially generating the samples of the *background* class then performing a binary semi-supervised classification process. On the other hand, when the *a priori* information of a camouflaged object or background is unavailable, unsupervised object detection has also been widely explored and used [20–22]. In this scenario, feature selection techniques and images of multi-temporal observations play an important role [20, 22, 23].

- *Information retrieval.* Taking advantage of features such as spectral signatures, texture, locations, spatial adjacency and so forth, information retrieval is aimed at measuring and estimating the presence, significance and quantity of some specific materials or targets. For example, spectral unmixing has been one of the most widely used techniques for information retrieval in order to estimate the possible presence of several materials (along with their significance or abundance) in each pixel of the hyperspectral image [24]. Importantly, the application of information retrieval techniques greatly depends on the availability of pre-processing methods such as atmospheric correction [25], image deblurring and denoising [26] or feature extraction [27].

- *Multi-source data fusion.* Fusing multiple sources of remotely sensed data represents one of the most widely used techniques that aim to improve the interpretation of remote sensing imagery [28]. The complementary use of multi-spatial, multi-temporal and multi-sensor data exhibits the potential to broaden the horizons of remote sensing technology, while taking advantage of the increased availability of remotely sensed data sets collected from different sources. For the case of hyperspectral imagery, very high spatial resolution images [29], light detection and ranging (LiDAR) data [30] and passively-obtained images [31, 32] provide complementary information and features that can be exploited in the analysis of the data.

Other relevant techniques for hyperspectral image exploitation concern areas such as atmospheric radiation transfer modeling, compressive sensing and data storage, parallel analysis and computation, etc. These areas are very important but not addressed in detail in this document due to space considerations. Also, we must note that there are also other categorization of available hyperspectral data processing techniques [3]. In this regard, the aforementioned categories consider only a sample of available techniques while many others are available in both theoretical and practical scenarios. A common approach in all these techniques for hyperspectral data exploitation is the analysis of the three-dimensional structure of hyperspectral data data cubes (see Fig. 1.1), which presents some important challenges.

With the increase of spectral dimensionality, hyperspectral imagery provides the possibility to analyze in diagnostic detail the features associated with specific materials and thus easily recognize them and/or distinguish the differences among them. However,

the relevant information for such purpose may live in the features corresponding to a few limited bands [33]. This high dimensionality property also imposes great challenges in order to build linear and non-linear analysis models, since the useful features are often hidden by some irrelevant spectral bands. Computationally and theoretically, the complexity of many problems usually increases dramatically in such high dimensional spectral space. This problem is often called the *curse of dimensionality* [34], or the Hughes phenomenon. The presence of spectrally mixed pixels is also an important challenge for the successful recognition of land objects. Although the mixture problem can be seen as a sophisticated one that can be influenced by the scanning way of the imaging instrument (sample-wise, band-wise or block-wise), it can be tackled using approximations such as the linear mixture model. Another important concern in hyperspectral image analysis is the fact that acquiring labels of ground objects largely relies on field work. Hence, the collection of labeled data has been quite time-consuming and expensive, leading to an imbalance between the high spectral dimensionality of the data and the limited availability of labeled training samples [35]. As mentioned in previous contexts, this thesis work develops new techniques that intend to address some of the problems mentioned before, particularly in the context of classification problems. In order to deal with the aforementioned problems, we propose innovative solutions that can be categorized in the following main groups.

1.1.1 Spectral Partitioning

In order to deal with the high dimensionality of hyperspectral data, researchers are searching for ways to solve hyperspectral analysis problems in a much lower dimensional space. In a reduced spectral space, a number of traditional image analysis techniques can be adopted readily. For this purpose, many hyperspectral dimensionality reduction methods have been developed in recent years. Traditional dimensionality reduction is usually performed by screening and selecting the bands/features of much lower spectral dimension that are most relevant to specific problems. By this means, the redundant information is ruled out and discarded for the subsequent data analysis tasks. Dimensionality reduction techniques can be divided into two main categories: band selection and feature extraction, depending on whether the original information in the hyperspectral data cube is transformed into a feature space or not.

1.1.1.1 Feature extraction

In this context, feature extraction techniques are generally more flexible and widely used, including well-known approaches in the literature such as principal component analysis (PCA) [36], independent component analysis (ICA) [37], manifold learning (ML) [38] and subspace-based approaches [39]. Other techniques, such as wavelet transformations [40], best-bases feature extraction [41], kernel nonparametric weighted feature extraction [42], have also been widely explored and widely successful. Feature extraction takes advantage of the high correlation among the hyperspectral bands that exhibit redundancy of information [3] (see the correlation matrix map in Fig. 1.2). By transforming the original

data into a lower dimensional feature space, the original problems can be tackled in a more simple way by exploiting the much lower dimensionality. However, feature extraction generally transforms the original information after projecting the data into a certain feature space, which may be a challenge for certain applications that require meaningful spectral signatures according to their physical interpretation [43, 44].

1.1.1.2 Band selection

Band selection, on the other hand, takes advantage of the fact that a small number of the spectral bands often convey the most relevant information of the hyperspectral image. As a result, band selection indends to select the most useful bands from the original set of spectral bands, in supervised or unsupervised fashion. Here, the main concern is how to select the most informative, representative and significant bands from the initial set of hundreds of bands available. As a result, band selection is a tool of choice in cases in which some specific bands of the hyperspectral data are particularly relevant for the problem at hand [45]. Many algorithms and techniques have been developed for hyperspectral dimensionality reduction, such as the projection pursuit (PP) [46], principal component analysis (PCA) [47], independent component analysis (ICA) [48], spectral angle mapper (SAM) [49], minimum misclassification canonical analysis (or Fisher's discriminant analysis) (MMCA), etc. Other methods also use some specific criteria to perform band selection, such as using high-order moments for band ranking, divergence, decorrelation [50], using gene selection with support vector machine (SVM) methods based on recursive feature elimination (RFE) [51], and so forth. As opposed to feature extraction, band selection reduces the spectral dimensionality while, at the same time, retains their original physical meaning.

As a matter of fact, the discriminative information that a classifier can exploit in order to provide good discriminative performance may often live in weak features/bands that are usually discarded or lost after a certain feature transformation or the application of a band selection technique. The relevance of a spectral band for classification also highly varies depending on specific practical demands. Hence, it can be often challenging to decide which are the most appropriate bands for specific applications.

1.1.1.3 Alternative solutions explored in this thesis

With the aforementioned issues in mind, in this thesis we explore the role of spectral partitioning as an alternative to the aforementioned methods. In fact, spectral partitioning has been widely used in recent developments (including the ones presented in this work), aiming mainly at rearranging the original spectral bands of a hyperspectral image into multiple band subsets with lower dimensionality [52–54]. Spectral partitioning is thus similar to band selection in the sense that it can retain the original physical information of the original data. However, as opposed to band selection, spectral partitioning does not necessarily discard most of the original spectral bands to achieve lower dimensionality [54]. Instead, spectral partitioning generates several groups of band subsets from the original spectral bands, so that each band subset is a so-called spectral partition or cluster, containing a much lower number of spectral bands as compared with

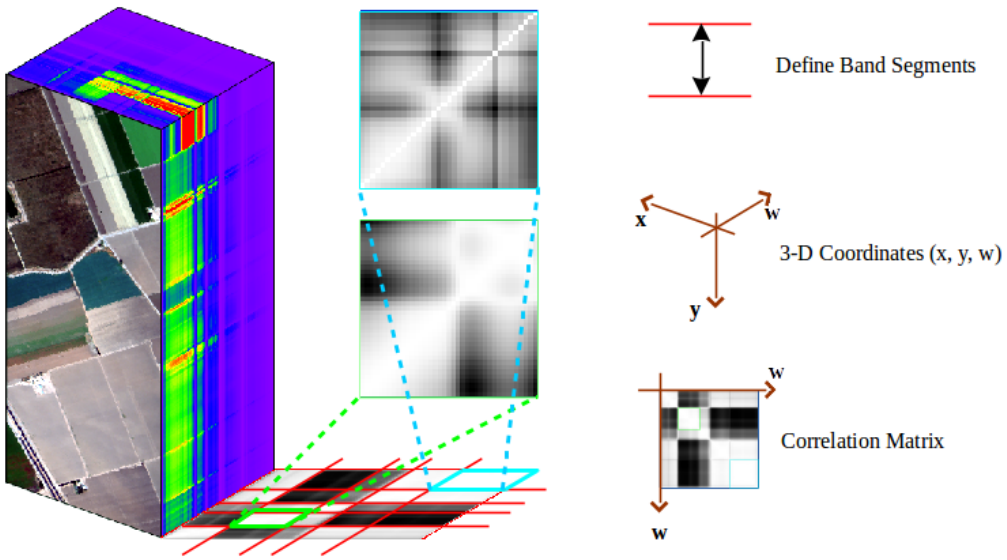


Figure 1.2: Correlation matrix and band segments of hyperspectral data.

the original hyperspectral image. As a result, the union of multiple subsets can make up to the full original image [55], which means that the full information present in the original scene is exploited in the process. As a result, the benefits of spectral partitioning can be highly relevant in the following contexts: 1) some specific applications require maintaining all the original features after dimensionality reduction, especially the original physical meaning of the spectral bands; 2) in supervised scenarios, such as classification, the class-dependent information often lives in weak features/bands that may be discarded by traditional dimensionality reduction techniques; and 3) spectral partitioning is designated to effectively provide multiple views of the original hyperspectral image by the obtained multiple band subgroups, which allow us to exploit the full spectral information in the original hyperspectral scene while circumventing the issues related with the curse of the dimensionality. It is also worth noting that spectral partitioning is particularly appealing for multi-classifier systems or ensemble learning scenarios. In other words, the spectrally partitioned band subgroups are capable of generating distinct perspectives or views that may potentially provide diversity to different classifiers for ensemble learning purposes [56, 57]. Band selection can also be viewed as a special case of spectral bi-partitioning, in which the unselected band partitions are discarded in the subsequent analysis tasks.

In the hyperspectral remote sensing community, the idea of spectral partitioning was first used in [52]. In order to improve the performance of PCA, the segmented principal component analysis technique described in the aforementioned work was developed in order to partition the dataset spectrally into a number of distinct spectral partitions, each of which contains a number of highly correlated bands. One of the characteristics of the partitioning strategy adopted in this work is that the extracted features are well defined from the principal components of each band subset, which are highly correlated. Hence,

researchers have adapted this segmented PCA for different tasks, such as parallelized compression of hyperspectral imagery [53] or mapping invasive plant species [58]. In addition, spectral partitioning has been also exploited in the framework of classification. In the work [59], a multi-hypothesis-prediction procedure is explored based on spectral partitioning to generate multiple hypotheses. The classification performance is shown to be significantly enhanced, especially under small training sample size constraints and noise corruption. Another advantage is that the analysis of the multiple partitions can be conducted in parallel. It should be noted, however, that in some cases [60] spectral partitioning techniques may not be appropriate for parallel implementation because the calculations made for each hyperspectral pixel would need to originate from several processing elements, thus requiring intensive inter-processor communication. Still, in other cases [61], the efficiency of image analysis can be enhanced via spectral-domain partitioning. In the work [62], band clustering-based spectral partitioning has been investigated for dimensionality reduction of hyperspectral image fusion and classification. In the classification scenario, the partitioning of the spectral data into smaller subspaces ensures that no relevant information is discarded, thus allowing for the interpretation of the data via multiple classifiers using limited training samples [63, 64]. Since the dimensionality of each subspace is much smaller than the dimensionality of the original feature space, each classifier is well conditioned.

Despite the successful application of spectral partitioning to different hyperspectral imaging problems, most of the techniques developed for this purpose focus on the advantages for dimensionality reduction purposes. In turn, a detailed exploration regarding the differences and connections among the partitioned multiple band subgroups have been seldom investigated in context of hyperspectral classification. In addition, the diversity provided by the different classifiers can be exploited by multiple classifier systems [57]. This intuitive rationale stems from the fact that the ensemble output of multiple classifiers will provide no substantial differences if all the involved classifiers obtain the same estimation. As a result, the diversity brought by the multiple perspectives provided by the spectral partitions is of great significance to explore and investigate alternative ways to improve classification performance. This is in fact one of the main ideas explored by the new spectral partitioning techniques presented in this thesis work.

1.1.2 Spatial Partitioning / Superpixelization

As mentioned previously, the imbalance between the high spectral dimensionality and the limited availability of labeled samples results in great challenges for hyperspectral image classification, often leading to ill-posed problems. In addition, the presence of mixed pixels in the local domain also makes it a difficult task to correctly label the image using only the spectral information. Spatial information, coming from the image domain, has been investigated and demonstrated to be highly useful in the task of improving hyperspectral image labeling tasks [65] (see Fig. 1.3).

To provide a simple example of the advantages that spatial information can bring to hyperspectral classification problems, we can think of a classification outcome in

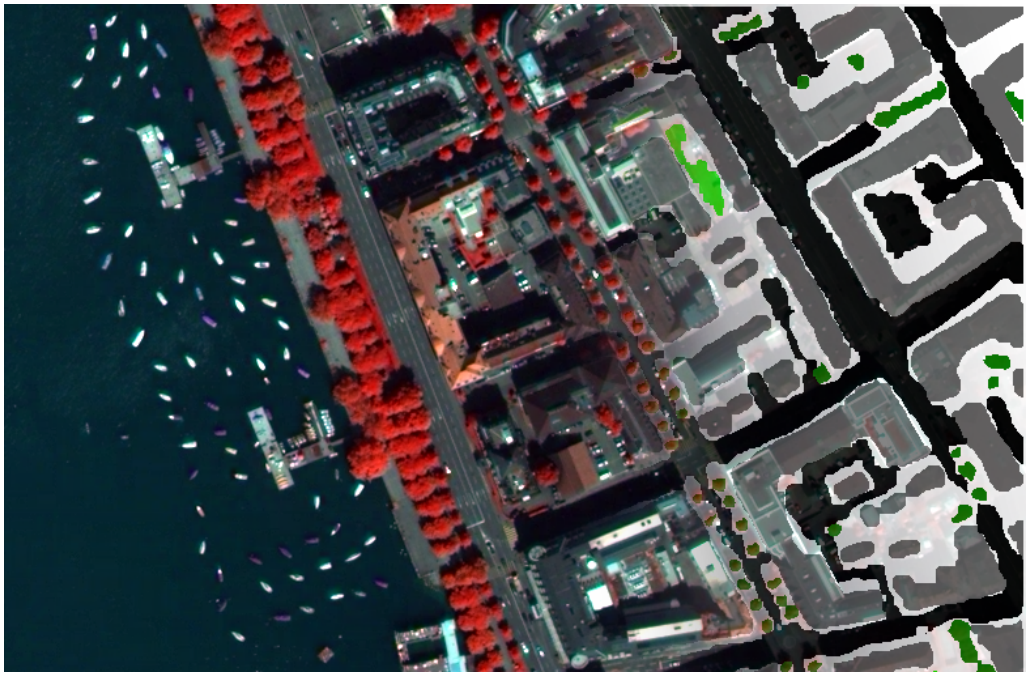


Figure 1.3: Graphical illustration of the advantages of spatial information in hyperspectral image classification.

which spurious points or cliques are present in the obtained result. With the help of spatial information, this situation can be potentially improved as illustrated graphically in Fig. 1.3. In order to exploit spatial and spectral information, spectral-spatial classification has been widely explored, resulting in a plethora of methods and techniques that jointly combine these two sources of information. The key aspect of these methods is to exploit the high correlation of neighboring pixels in hyperspectral images. Note that there already several types of techniques regarding how to include spatial information into the image labeling models.

1.1.2.1 Spatial pre-processing and post-processing

One paradigm for taking advantage of the spatial information is to directly make use of the fact the pixels in local neighboring areas share very high similarity/correlation and thus probably also share the same class labels. On the one hand, pre-processing [66] based on spatial filtering has been successfully exploited in spectral unmixing [67], classification [68] and change detection [69], etc. The spatial filtering operation is able to effectively decrease the influence of noise and outliers, thus uniforming and reshaping the spectral signatures of neighboring pixels. Besides traditional spatial filters such as the Sobel and Robert operators, morphological operators have also been developed considering the spatial characteristics of an object with respect to its shape and size [70–73]. By separating the image into targets and background, morphological operators are capable of better delineating the contour of certain land objects and

retaining their boundaries, and also allowing for a multiscale analysis of the objects based on structuring element filters of different size and shape [71]. Besides, others techniques in the spatial domain can be considered [74, 75]. On the other hand, post-classification is also an important strategy to improve the effectiveness of classification. Classical techniques include median value filtering or majority voting, among many others [76]. While pre-processing techniques mainly aim at reshaping the objects in order to better train a classifier, post-processing techniques mainly intend to refine the classification result after it has been already obtained. However, the pre-processing of the image may compromise the generalization ability of the classifier by oversmoothing the spectral signatures associated to the training samples. In turn, post-processing technique pursues a more naive and straightforward exploitation of spatial information after the classifier has obtained an output.

1.1.2.2 Segmentation

Unsupervised segmentation or spatial clustering techniques have also been widely used to improve hyperspectral image classification. The presence of well-contoured objects with precisely located boundaries may allow for a better classification without the need for changing the original spectral signatures of the training pixels. In line with this, several auhors have presented techniques for identifying spatial structures in the hyperspectral image by performing unsupervised segmentation such as watershed, partitional clustering and hierarchical segmentation (HSEG), and others [10, 77, 78]. From this standpoint, the image segmentation/partitioning techniques can also be treated as a special case of post-processing techniques.

1.1.2.3 Alternative solutions explored in this thesis

In recent years, the concept of superpixel (which refers to a *small* image segment) has inspired a family of new methods that exploit the spatial information for hyperspectral classification. The pixels comprising a superpixel are generally believed to share highly similar characteristics, which makes them good candidates to share the same label labeling problems. However, as opposed to normal segments, superpixels generally depict only a part of a land object instead of fully describing it. This greatly reduces the risk of misclassification.

Despite these important advantages, the inclusion of spatial information that comes from image spatial partitions has been mostly conducted in the form of post-classification processing [65, 79, 80]. The inclusion of spatial information by these means is usually not straightforward [81–83]. Consequentially, an over-usage of the segments or superpixels often leads to high accuracy but unsatisfactory classification maps. Although the classification statistics can be high (mainly driven by the large objects whose labels are collected far away from the boundaries), the boundaries of different objects may not be sharply delineated and small-sized objects may be undermined by the large ones in the final classification map. Intrinsically, a straightforward use of spatial information usually leads to a discrete optimization problem that is NP-hard due to its discrete nature. Many techniques and algorithms have been developed to deal with this problem.

Graph-cut [84] algorithms attack the discrete optimization labeling problem via cutting the graph of neighboring pixels while pursuing a minimum function energy. In contrast to the methods and techniques that belong to post-processing of classification, graph-cut algorithms also consider the cost of changing the image labels in the spatial domain. In the binary cases, an optimal solution can be obtained with graph-cuts, which have been widely used in applications like foreground/background recognition, object detection, etc. Apart from binary cases, this is a very difficult problem in multi-class labeling scenarios whose solution can only be approximated. In the works [68, 85, 86], graph-cut algorithms via α-expansion and α-β swap have exhibited significant advantages for promoting label consistency of adjacent pixels, minimizing the boundary length of objects and providing good approximation of optimal solutions. Methods based on graph-cut algorithms actually include the spatial information via the Potts model, which assumes a Gibbs distribution [87] towards the labels of neighboring pixels. Considering the discrete nature of the hyperspectral image labeling problem, it is still very difficult to include the spatial information that comes from superpixels, textures and gradients. In this context, techniques like the tree re-weighted scheme [88], belief propagation [89], robust higher order potential [90] in discrete optimization, as well as the primal-dual schema for Markov random field optimization [91, 92] in the compact domain, have been also widely investigated and shown to be effective in tackling these problems in recent years. In this thesis work, we intend to explore and attack the spectral-spatial hyperspectral imagery classification problem using superpixels, aiming to utilize the spatial information in a straightforward manner under a convex optimization scheme.

1.1.3 Synergistic Analysis of Multi-Source Remote Sensing Data

As explained in the previous subsection, the improvements in classification that can be gained through the use of superpixels strongly depend on the effectiveness of the superpixelization process for hyperspectral data. The characterization of spatial information is therefore of fundamental importance in this context [93]. However, delineation of object edges in hyperspectral images is a difficult task due to the presence of mixed pixels [93, 94].

To address this issue, several studies have been carried out in the literature [19, 30, 95, 96] analyzing the possibility of integrating data from multiple remote sensing instruments, including hyperspectral sensors as well as LiDAR sensors. These two kinds of sensors, with their different peculiarities and characteristics, can provide complementary information for specific problems, allowing for instance to tackle difficult problems such as the classification of different tree species or the estimation of biophysical parameters [30]. Assisted by the rich source of spatial information provided by the LiDAR data, challenging problems for hyperspectral imaging can now be tackled, including the acquisition of pigments of different plant species [97] as well as their bio-diversities [98], bio-masses [99] and so on. This is because the land objects with similar spectral signatures can often be distinguished by resorting to the LiDAR data by exploiting additional characteristics such as different elevations, number of point returns, density of points, etc. In addition, the LiDAR data are able to better depict the objects in three-dimensional

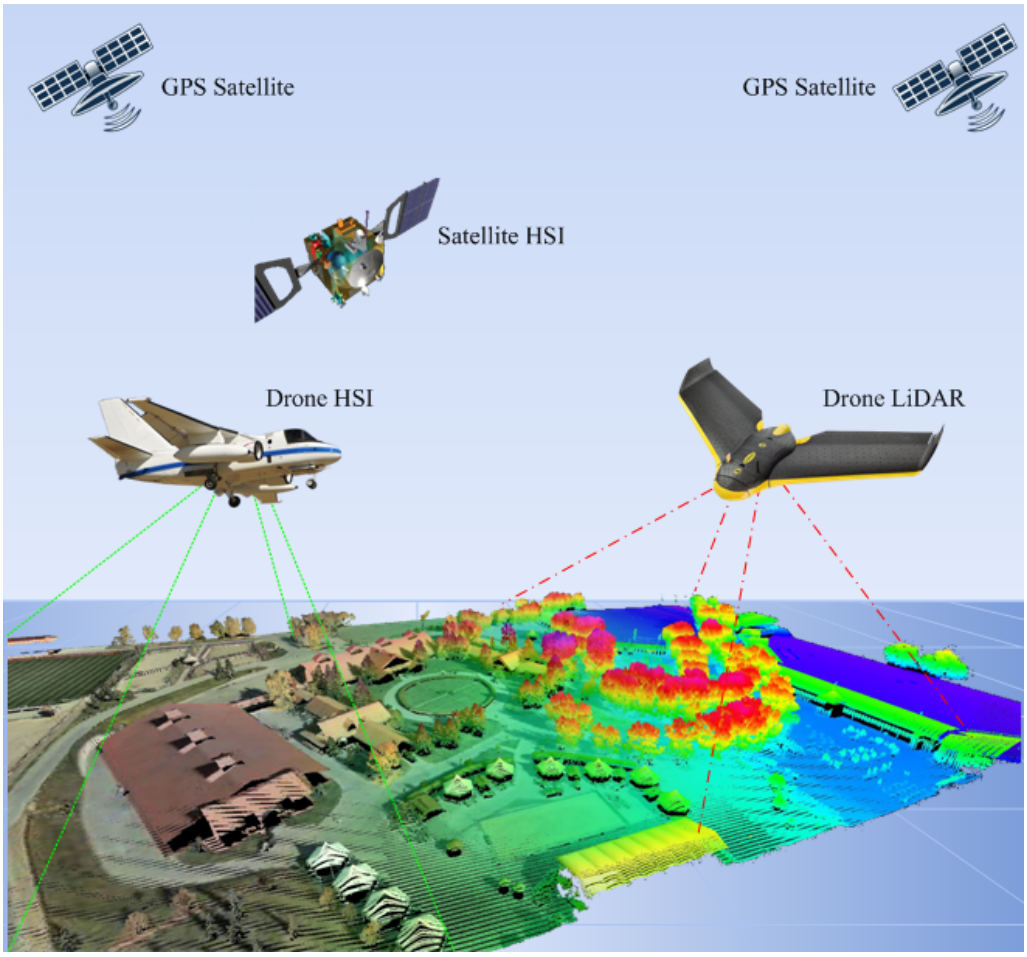


Figure 1.4: Synergistic remote sensing with LiDAR in addition to hyperspectral imaging.

space. Inspired by the aforementioned developments, in this thesis work we develop new strategies to complementarily take advantage of hyperspectral and LiDAR data, thus promoting the confidence of a classifier and the spatial regularization of the final classification map under a convex optimization mechanism.

In the literature, the synergistic use of LiDAR and optical data has been shown able to advance the state-of-the-art classification methods [100, 101]. In these studies, the LiDAR data have been used to improve image segmentation techniques [100], morphological filters [101], extended attribute profiles [102], among many others [103]. In general, the synergistic usage of these multiple-sensor datasets is performed by band-wise stacking of the rasterized elevation map of the LiDAR data (as well as other previously extracted textural and morphological features) onto the original hyperspectral image data. However, this strategy introduces some problems. First of all, the very large number of spectral bands of the hyperspectral data may easily dominate the (more limited) number of features provided by the LiDAR data. The distinguishing information of classes, that often lives in weak features coming from the LiDAR data, can be easily

ruled out by the largest available number of hyperspectral features (which may be redundant and/or irrelevant to the classification). On the other hand, an overuse of features coming from the LiDAR data may obscure the spectral details coming from the hyperspectral data, which are also required for a successful classification. To address these relevant issues, in this thesis work we propose a new strategy for balanced and synergistic exploiation of both hyperspectral and LiDAR data. Specifically, we use both hyperspectral and LiDAR data to better train a classifier using their complementary information, emphasizing the role of the LiDAR data in spatial information extraction in order to better delineate object boundaries, and adequately exploiting the rich spectral information coming from the hyperspectral data after noise reduction.

1.2 Objectives

The main goal of this thesis work is to develop new methods and techniques for spectral and spatial partitioning and/or clustering of hyperspectral images for classification purposes. The new spectral partitioning methods developed in this work consider that relevant features for classification may often live in weak features or bands that should not be discarded in the process. As opposed to traditional methods for dimensionality reduction like band selection and feature extraction, we do not select only a small number of the most informative features/bands while discarding all others. This is because it is very difficult to determine in advance which are the most relevant features/bands that can best train a classifier. As a result, one of our main contributions in this work is to use spectral partitioning as an efficient mechanism to generate multiple perspectives or views (via different spectral partitions of the original image), in order to take advantage of all the original spectral information present in the hyperspectral data. This is achieved by constructing multiple classifier systems based on intelligently derived spectral partitions, designed with the ultimate goal of achieving a more robust classification estimation. The diversity among the spectral partitions is thus mainly explored and discussed for this purpose in this thesis work. The thesis also intends to naturally include the spatial information that comes from superpixels into the classification scenario. This is also a highly innovative contribution, which involves the development of a convex optimization strategy that is specifically developed to involve multiple superpixelizations as a graph total variation that reinforces the label consistency in the spatial domain. Following this convex optimization process, an instance of spectral-spatial classification using multi-source remote sensing data (hyperspectral+LiDAR) is also developed as a novel contribution of this work. Overall, our newly developed strategies for spectral and spatial partitioning based hyperspectral image classification are expected to introduce significant contributions to the state-of-the-art in the field of hyperspectral imaging research. In order to achieve this general objective, we need to address a number of specific objectives which are listed below:

1. To acquire the background knowledge about hyperspectral analysis in remote sensing, specifically in the field of spectral partitioning in order to establish

innovative techniques for dimensionality reduction able to tackle the high-dimensional nature of hyperspectral data cubes in the spectral domain. We specifically focus on the understanding of spectral correlation and divergence of the data cube, as well as on the development of new techniques for spectral partitioning prior to classification.

2. To develop new classification techniques based on multiple classifier systems by exploring the diversity among different classifiers when applied to multiple spectral partitions that provide different views of the original hyperspectral data. The information relevance and the confidence of the classifier will be analyzed and evaluated experimentally with the ultimate goal of measuring the ensemble performance of multiple classifiers based on different spectral partitioning techniques.

3. To design new strategies for the inclusion of spatial information that comes from over-segmented superpixels in order to improve classification performance. We specifically focus on constructing a new convex optimization approach that considers the cost of modifying the class labels associated to neighboring pixels. We sidestep the discrete optimization labeling problem and include spatial information as a prior to the classifier.

4. To develop new and fast superpixelization schemes for hyperspectral imagery and exploit these techniques in different applications related to hyperspectral image classification.

5. To design an effective multi-source remote sensing data classification technique based on the integration of hyperspectral and LiDAR data to better exploit spatial information in combination of spectral information. The newly designed approach should be able to exploit the most relevant spectral information coming from the hyperspectral data and the spatial information coming from the hyperspectral+LiDAR data, in order to fully exploit the spectral and spatial partitioning techniques introduced by this thesis work.

1.3 Main Contributions of the Thesis

The main contributions of this thesis work are summarized in Fig. 1.5, in which the relationship between the different chapters is highlighted. The thesis can be divided into three main parts according to the main goals summarized above. One part of the thesis is devoted to the introduction of spectral partitioning techniques prior to hyperspectral image classification, especially those considering the spectral correlation and spectrometer characteristics for dimensionality reduction (Chapters 2 and 3). Another part of the thesis is dedicated to the inclusion of spatial information that comes from over-segmented superpixels in image labeling problems (Chapter 4). Finally, another part of the thesis is devoted to the development of strategies for synergistically exploiting multi-sensor (hyperspectral and LiDAR) data for improved spatial and spectral

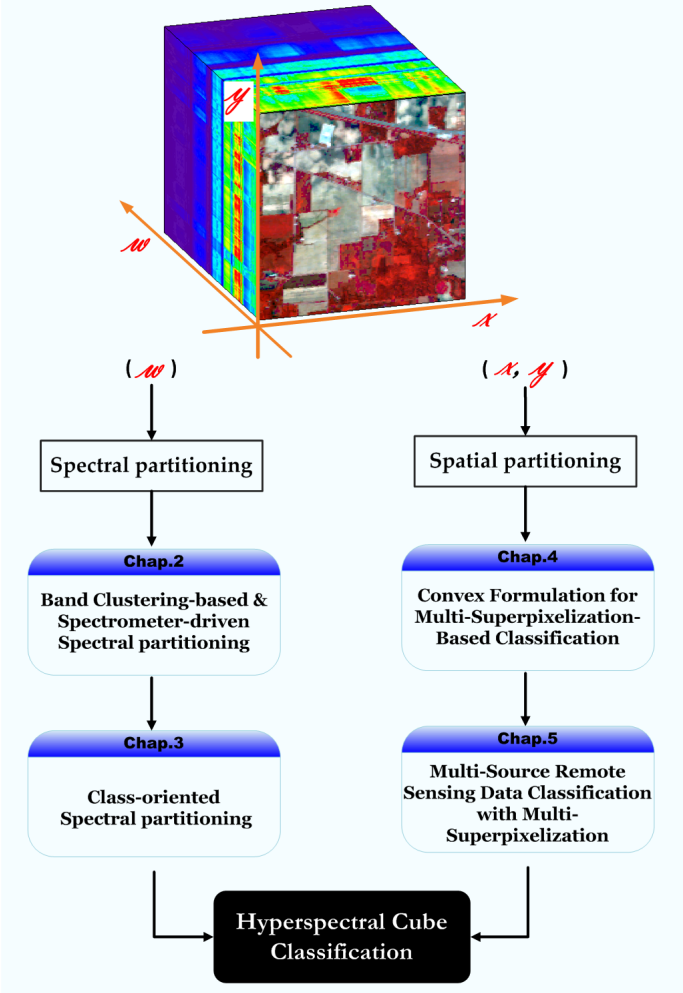


Figure 1.5: Flowchart illustrating the organization of this thesis.

characterization (Chapter 5). Chapter 6 concludes the thesis with remarks and hints at plausible future research lines. The individual contributions of each chapter are summarized below.

- Chapter 2 introduces two new spectral partitioning methods [54, 56] which create multiple views of the original hyperspectral data by considering band correlation and the characteristic of the imaging spectrometer, respectively. The proposed spectral partitioning methods are evaluated in classification problems using a variety of hyperspectral data collected by different spectrometers, demonstrating their capacity to outperform other state-of-the-art classifiers.

- Chapter 3 introduces a new class-oriented spectral partitioning strategy that considers the class-dependent information in hyperspectral band partitioning [55]. The new spectral partitioning strategy presented in this chapter is more classification oriented as compared with the ones introduced in Chapter 2. Again,

the performance of the newly introduced spectral partitioning method is validated using real hyperspectral datasets.

- Chapter 4 introduces an innovative convex formulation for hyperspectral image spectral-spatial classification based on superpixels [104]. In the proposed framework, the superpixel concept is formulated as a graph total variation regularizer, in order to naturally include the spatial information resulting from the adjacent pixels comprising a superpixel. By this means, an arbitrary number of superpixelizations can be simultaneously considered in the same classification process. This method is also shown experimentally to be able to produce state-of-the-art classification results when compared to other methods. In addition, we also show that the newly proposed method can also be used to provide prior information to a wide range of image labeling problems.

- Chapter 5 introduces an effective scheme that exploits both hyperspectral and LiDAR data in order to better model the classifier presented in Chapter 4 and provide refined superpixelizations with the support of the LiDAR data [105]. Our experiments, conducted with real hyperspectral and LiDAR data sets collected over the same area, indicate that the newly proposed method is highly effective in terms of classification performance.

- Chapter 6 summarizes the main contributions of this thesis work and the research lines that will be explored in the future.

To conclude this chapter, Table 1.1 provides a list of all the acronyms that have been used the preparation of this thesis document. The thesis also includes an appendix with the main publications resulting from this research, which include several journal citation reports (JCR) papers and presentations in prestigious international conferences in the field.

Table 1.1: List of acronyms used in this thesis.

Acronyms	
AAP	Adaptive Affinity Propagation
AAP-SP	AAP-Driven Spectral Partitioning
AP	Affinity Propagation
ADMM	Alternating Direction Method of Multipliers
ARES	Airborne Reflective Emissive Spectrometer
AVIRIS	Airborne Visible Infra-Red Imaging Spectrometer
BS	Band Selection
DAIS	Digital Airborne Imaging System
EAP	Extended Attribute Profile
FFT	Fast Furior Transformation
FPGA	Field Programmable Gate Arrays
GPU	Graphics Processing Unit
GTV	Graph Total Variation
HYDICE	Hyperspectral Digital Imagery Collection Experiment
HySime	Hyperspectral Subspace Identification by Minimum Error
HSI	Hyperspectral Image
ICA	Independent Component Analysis
LDA	Linear Discriminant Analysis
LiDAR	Light Detection and Ranging
LMM	Linear Mixture Models
LP	Linear Program
MAP	Maximum a Posteriori
MCS	Multiple Classifier System
MPO	Moreau Proximity Operators
MLR	Multinomial Logistic Regression
MMAP	Marginal Maximum a posteriori
MNF	Minimum Noise Fraction
MRF	Marcov Random Field
MSI	Multispectral Image
NP-hard	Non-Deterministic Polynomial Hard
PCA	Principal Component Analysis
ROSIS	Reflective Optics System Imaging Spectrometer
R-SP	Random Spectral Partitioning
SLIC	Simple Linear Iterative Clustering
SAD	Spectral Angle Distance
SALSA	Split Augmented Lagrangian Shrinkage Algorithm
SD-SP	Spectrometer-Driven Spectral Partitioning
SNR	Signal-to-Noise Ratio
SP	Spectral Partitioning
SUNSAL	Spectral Unmixing by Splitting and Argumented Lagrangian
SVD	Singular Value Decomposition
SVM	Support Vector Machine
VTV	Vectorial Total Variation
USGS	United States Geological Survey
VCA	Vertex Component Analysis

Chapter 2

Spectrometer-Driven Spectral Partitioning

2.1 Summary

Classification is an important and widely used technique for remotely sensed hyperspectral data interpretation. Although most techniques developed for hyperspectral image classification assume that the spectral signatures provided by an imaging spectrometer can be interpreted as a unique and continuous signal, in practice this signal may be obtained after the combination of several individual responses obtained from different spectrometers. In this chapter, we propose a new spectral partitioning strategy prior to classification which takes into account both the physical design of the imaging spectrometer system for partitioning the spectral bands collected by each spectrometer and the band correlation, and resamples them into different groups or partitions. The final classification result is obtained as a combination of the results obtained from each individual partition by means of a multiple classifier system. The proposed strategy not only incorporates the design of the imaging spectrometer into the classification process, but also circumvents problems such as the curse of dimensionality given by the unbalance between the high number of spectral bands and the generally limited number of training samples available for classification purposes. This concept is illustrated in this chapter using two different imaging spectrometers: the Airborne Visible Infra-Red Imaging Spectrometer, operated by NASA, and the Digital Airborne Imaging System, operated by the German Aerospace Center. Experimental results indicate the effectiveness of the proposed spectral partitioning strategy with respect to classification improvements on the order of overall accuracy when compared with state-of-the-art spatial-spectral classifiers with very limited training samples.

2.2 Introduction

Imaging spectroscopy (also called hyperspectral remote sensing) has experienced significant developments in recent years [3]. Currently, many sensors onboard airborne

Table 2.1: List of abbreviations used in this chapter

AVIRIS	Airborne Visible/infra-red Imaging Spectrometer
ARES	Airborne Reflective Emissive Spectrometer
DAIS	Digital Airborne Imaging System
HyMap	Hyperspectral Mapper
LDA	Linear discriminant analysis
HySime	hyperspectral subspace identification with minimum error
MNF	Minimum noise fraction
PCA	Principal component analysis
MLR	Multinomial logistic regression
SVM	Support vector machine
MRF	Markov random field
MLRsub	Subspace-based MLR
MCS	Multiple classifier system
AP	Affinity propagation
AAP	Adaptive affinity propagation
R-SP	Random spectral partitioning
AAP-SP	AAP-driven spectral partitioning
SD-SP	Spectrometer-driven spectral partitioning

and spaceborne platforms are available, and these instruments keep collecting data from different locations on the surface of the Earth [7]. Hyperspectral data have been useful in many applications, including disaster monitoring, natural resources exploitation, environmental applications, etc. [106–108]. With the increasing spatial, spectral and temporal resolutions of imaging spectrometers, the extremely high dimensionality and size of the data have become important concerns for hyperspectral data interpretation [109]. Among several techniques for hyperspectral image analysis, classification has been a very important research topic for interpreting hyperspectral data [110], in which the main challenges have been given by the unbalance between the high dimensionality of the data and the limited number of training samples generally available *a priori* [111].

In the following, we provide a description of the state-of-the-art in classification and spectral partitioning. Although supervised classification techniques such as the SVM [112] or MLR [86] have been shown to be quite successful for the interpretation of hyperspectral data (even in the presence of limited training samples), some techniques have taken advantage of dimensionality reduction [52, 113, 114] or subspace projection [115] prior to classification. The high existing correlation between bands has been exploited to design new methods for reducing data dimensionality, including methods that have found great popularity such as PCA [116], LDA [117], or the MNF [118]. Subspace projection techniques such as HySime [115] have also been used for this purpose. In fact, it has been reported in previous works that classification after dimensionality reduction, subspace projection or band/feature selection generally outperforms classification based on the full original hyperspectral data [7]. By reducing the number of bands, unsupervised feature selection [119, 120], semi-supervised feature

selection [121] and supervised feature selection [122] have been reported to be able to achieve similar or better classification accuracies than using all available bands.

Spectral partitioning, which is a form of dimensionality reduction, provides an alternative approach to deal with the high dimensionality of hyperspectral data. In comparison with traditional band reduction/selection approaches, a distinguishing feature of spectral partitioning is that all spectral bands of the input image can be used for the subsequent analysis process by creating multiple views of the original hyperspectral image. This can be achieved by selecting, for a given partition, a subset of bands that effectively subsample the original spectral signature in the original hyperspectral scene while retaining its main characteristics. The selection of multiple disjoint subsamples provides multiple views of the original hyperspectral signature that can be combined for classification purposes. As reported in [123], this strategy can be implemented by using techniques such as the AAP [124]), which first automatically generates spectral band clusters in which the spectral bands are correlated with each other, and then the bands are reassigned from the clusters into new groups (namely, spectral partitions) which constitute a subsampling of the original spectral signatures. Similarly, utilizing a bisection spectral partitioning strategy, the work in [125] reported a significant gain in classification performance for hyperspectral data. In [52], a spectral partitioning based on band correlation segmentation was also employed to achieve better results from the PCA transformation by performing the PCA on each partition (spectral segment). Besides, in [126] spectral and spatial partitioning were incorporated into a PCA-based compressive projection for hyperspectral image reconstruction. These works indicate that spectral partitioning can lead to improved classification results for hyperspectral data.

An important characteristic of spectral partitioning techniques such those mentioned above is that they do not consider the physical characteristics of the imaging spectrometer, which may be critical in the process of creating the spectral band clusters. In fact, most classification techniques assume that the spectral signatures provided by an imaging spectrometer system for a given pixel can be considered as a unique and continuous signal. However, it is well-known in hyperspectral imaging that different parts of the signal are subject to the different characteristics of the physical instruments after the combined individual responses from several different spectrometers. For instance, the AVIRIS[1] system [1] is formed by four different spectrometers, covering the nominal spectral ranges: 400–700 nm, 700–1300 nm, 1300–1900 nm, and 1900–2500 nm, respectively. As shown in Fig. 2.1, these four spectrometers (called A, B, C and D, respectively) have completely different characteristics. For example, in the D spectrometer, a carefully specified linear variable blocking filter is used to limit the wavelengths of light for each element in the detector array. This specialized filter is intended to separate different spectral orders. For the A spectrometer, a 32-element silicon detector array with a blue enhanced response is used. The B, C and D spectrometers each have 64-element arrays instead. As a result, our main contribution in this chapter is a new partitioning framework that is driven by the characteristics of the imaging spectrometer and its physical configuration, instead of assuming that the

[1] http://www.nasa.gov/centers/dryden/research/AirSci/ER-2/aviris.html

spectral signature is a continuous signal. The methodology that we propose in this chapter offers a new perspective that gives a new flavour to the problem of spectral partitioning. Given the fact that different spectrometers have different characteristics, our speculation in this chapter is that the conventional assumption that the spectral signatures can be interpreted as unique and distinct signals may not hold in all cases (at least, given the current status of hyperspectral technology), and that the use of a spectrometer-driven partitioning strategy prior to classification may lead to improved hyperspectral data interpretation results, in particular, when spectral partitioning is used for dimensionality reduction purposes. Similar observations can be made for other widely used imaging spectrometers, like DAIS 7915[2] [127], ARES [128], Hyperion [129], or HyMap [130], among several others [7].

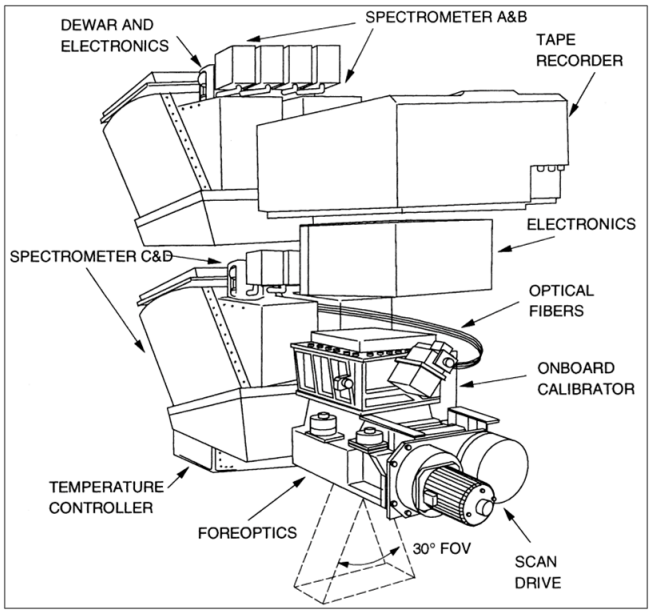


Figure 2.1: Four spectrometers in the AVIRIS system (reproduced from [1]).

In this chapter, we develop a new spectrometer-driven partitioning strategy that considers the differences between the individual spectrometers in the spectral partitioning step. An important consideration of our approach is that, as the dimensionality of each spectrometer signal is much smaller than the whole spectral signal, it allows for the design of a subsequent classification system that is efficient in scenarios in which limited training samples are available *a priori*, as the system uses multiple views of the original spectral signatures but with reduced dimensionality, which allows for the classification of the original hyperspectral data without discarding any information collected by the imaging spectrometer. In other words, instead of extracting features or removing bands, we reduce the dimensionality by reassigning the original bands into multiple views of the original data (according to the properties of the considered imaging spectrometer) and then using an MCS that considers all the information in the final classification.

[2]http://www.uv.es/leo/daisex/Sensors/DAIS.htm

For comparative purposes, the proposed spectrometer-driven partitioning strategy is compared with random partitioning and with another previously developed method (the AAP) which relies on statistical principles rather than physical principles related to the imaging spectrometer. The concept is illustrated using two different systems: AVIRIS and DAIS 7915.

The remainder of this chapter is organized as follows. In section 2.3, we introduce our spectral partitioning and classification framework, which is implemented using both the AAP clustering technique and our newly developed spectrometer-driven spectral partitioning strategy. This section also provides a comparison of these two spectral partitioning strategies to the simple random selection case, to illustrate the differences obtained between using a statistical-driven method (the AAP) and a more physically-driven method. Section 2.4 provides experimental results using representative data sets collected by the AVIRIS and DAIS 7915 systems. Finally, section 2.5 concludes with some remarks and hints at plausible future research lines.

2.3 Proposed Framework

In this section, we describe the proposed framework for hyperspectral image classification after spectral partitioning. Our framework is illustrated in Fig. 2.2. First, the original hyperspectral image is partitioned using spectral band clustering. For this purpose, we use two different strategies: AAP [124] (driven by statistical principles) and a newly-proposed spectrometer-driven approach (which is more driven by physical principles related to the design of the imaging spectrometer). The resulting q partitions provide multiple views of the original spectral signatures in the hyperspectral data. These different views are classified using an MCS which is based on a probabilistic classifier: the subspace multinomial logistic regression (MLRsub) classifier in [131]. In the following, we first describe the two spectral band clustering techniques considered (providing a comparison between them) and then provide details about the considered MCS system.

2.3.1 Spectral Band Clustering using Adaptive Affinity Propagation Spectral Partitioning (AAP-SP)

The AP method [8] was originally presented to automatically search for the optimal number of clusters in the analyzed data. A main contribution of the algorithm with regards to traditional unsupervised clustering approaches such as k-means [132] or ISODATA [133] is that the AP is less sensitive to initialization as it takes every sample as a potential cluster center. An adaptive version of AP (AAP) was proposed to automatically search for the optimal cluster number while removing inconsistencies [124]. As opposed to k-means or Isodata clustering, in AAP all data points are simultaneously considered as potential exemplars, but exchange deterministic messages until a good set of exemplars gradually emerges. In other words, using AAP we can perform band clustering without knowing in advance the number of exemplars. Let c be the number of band clusters and let ρ be the number of spectral partitions. In our previous work [123], we have adopted the AAP with a nested recursive iteration process for hyperspectral

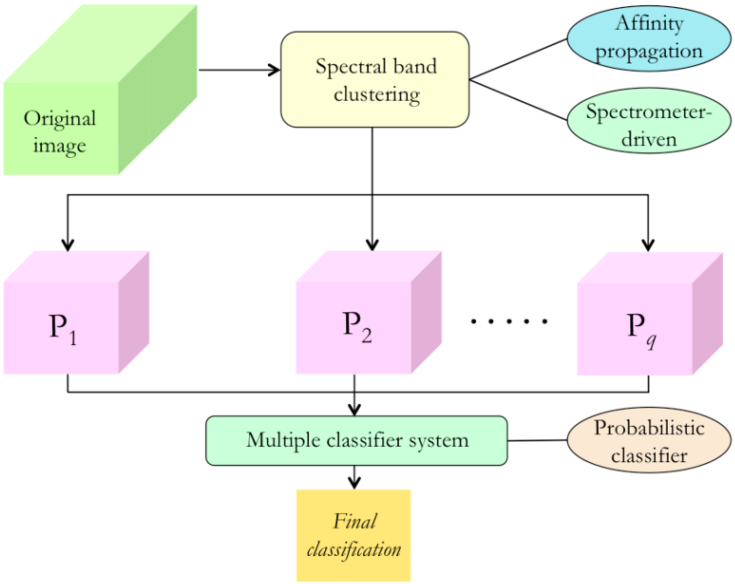


Figure 2.2: Proposed framework for spectral partitioning and classification of hyperspectral data.

band clustering. Resulting from this process, the original hyperspectral data is divided into c band clusters/groups, each of which contains a group of spectral bands with high similarity/correlation. After obtaining the c band clusters, we then reassign the bands with high similarity/correlation to spectral partitions by a band reassignment algorithm described in Section 2.3.2 such that each partition contains the same number of bands by using the AAP-SP approach.

2.3.2 Spectral Band Clustering using Spectrometer-Driven Spectral Partitioning (SD-SP)

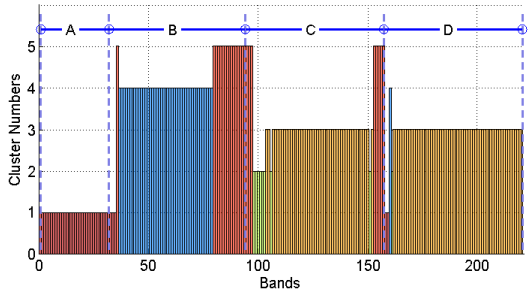
Imaging spectrometers generally collect spectral signatures as a combination of more than one spectrometer, each with different characteristics, signal to noise ratio (SNR), and etc. As a result, the conventional interpretation of spectral signatures as unique and distinct signals may not hold in all cases. In fact, the optical and physical characteristics of a group of spectral bands collected by the same spectrometer are expected to be quite similar and correlated, considering the electronic status and optical fibres of different spectrometers, as well as the different optical spectral ranges themselves. All these differences support the observation that the wavelengths corresponding to each spectrometer share high similarity and correlation, which provides an alternative source for partitioning the spectral signatures using a more physically driven approach than the aforementioned AAP, which is driven by statistical principles. Inspired by the aforementioned observations, in this chapter we have developed a new band partitioning strategy based on information conveyed by physical spectrometers to resample the spectral bands to different so-called partitions. The goal is to exploit the spectral

information by reducing spectral similarities while increasing the diversities among partitions.

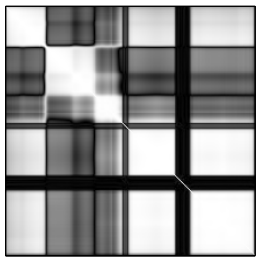
We have designed a simple band reassignment approach (which can be actually used for both AAP-SP and SD-SP). In the case of SD-SP, the procedure takes into account the information provided by different spectrometers. Let s be the number of spectrometers in the hyperspectral system. The spectral groups are defined by the spectral wavelengths available in each of the physical spectrometers and known in advance. Notice that s in SD-SP is the same as c for AAP-SP. Our goal is to reassign the spectral bands of the different groups into ρ partitions using equal interval sampling. If the number of bands in a group is divisible by ρ, it is easy to generate the ρ partitions by using equal interval sampling. Otherwise, in case that the number of bands in a group is smaller than ρ, we increase the number of bands by random selection from the current bands, until it reaches ρ bands. If the number of bands in a group is bigger than ρ, we remove the spectral bands having high correlation and similarity with the others from the group until it is divisible by ρ. It should be noted that this band reassignment strategy is driven by the physical information about the spectrometers of the system. Note also that the same reassignment strategy can be used to generate partitions from the AAP-SP derived band groups. For simplicity, here we used equal interval sampling due to the following reasons: 1) in this way, the spectral reflectance signature of each partition follows the basic shape of the original image. This allows to use reduced signatures but consistent with the spectral shapes originally collected by the sensor; 2) another reason is that, in our experiments, we have observed that competitive classification results are obtained as compared with regards to other strategies; 3) last but not least, the bigger the value of ρ (number of partitions), the less the similarity that exists among the bands in the same partition. As a result, by using equal interval sampling we can effectively manage the spectral similarities and correlations of the bands in the partitions.

2.3.3 Comparison of AAP-SP and SD-SP

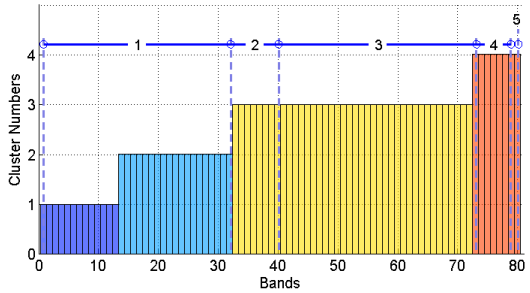
In this subsection, we compare the two spectral partitioning strategies described in previous subsections: AAP-SP and SD-SP. Figs. 2.3(a) and 2.3(c) respectively show the band clusters obtained by both methods for the AVIRIS Indian Pines data set and the DAIS Pavia City Center data set (both explained in detail in subsection 2.4.1). Specifically, the clustering results obtained by AAP were found using the same parameters for both systems: AVIRIS and DAIS 7915. In order to represent the partitions obtained by AAP in Figs. 2.3(a) and 2.3(c), the spectral bands that belong to the same cluster are denoted by bars of the same height and color. Here, the x-axis denotes the spectral band numbers and the y-axis denotes the corresponding cluster labels for each band. On the other hand, the letters: A, B, C and D represent the four spectrometers of the AVIRIS system in Fig. 2.3(a), while the horizontal blue lines between the vertical dashed lines denote the bands associated to each spectrometer in the AVIRIS data. Similarly, the numbers 1, 2, 3, 4 and 5 represent the five spectrometers of the DAIS 7915 system in Fig. 2.3(c). It can be seen that spectrometer 5 is composed by a single band, hence there is no horizontal blue line associated to this spectrometer. Tables 2.2 and 2.3 respectively show



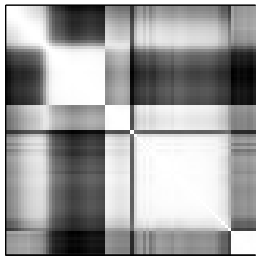
(a) AAP-SP vs SD-SP (AVIRIS)



(b) Correlation matrix (AVIRIS)



(c) AAP-SP vs SD-SP (DAIS 7915)



(d) Correlation matrix (DAIS 7915)

Figure 2.3: Comparison between the two spectral partitioning strategies: AAP-SP and SD-SP, using the AVIRIS Indian Pines and the DAIS 7915 Pavia City Center datasets. In plots (a) and (c), the x-axis denotes the wavelengths in micrometers, while the y-axis is the cluster labels obtained by AAP clustering. Each band cluster is denoted by a unique color and bar height. The letters: A, B, C and D represent the four spectrometers of the AVIRIS system, denoting the spectral partitions adopted by SD-SP. The horizonal blue lines between the vertical dashed lines denote the boundaries of band groups. Similarly, the numbers: 1, 2, 3, 4 and 5 represent the five spectrometers of the DAIS 7915 system.

Table 2.2: Description of the imaging spectrometers in the AVIRIS system.

Spectrometer	Wavelength range (nm)	Band numbers	Bandwidth
A	400–700	1–31	10 nm
B	700–1300	32–94	10 nm
C	1300–1900	95–157	10 nm
D	1900-2500	158–220	10 nm

the wavelength ranges, band numbers and bandwidth associated to each spectrometer in the AVIRIS and DAIS 7915 data. Finally, Figs. 2.3(b) and 2.3(d) respectively display the observed spectral correlation matrix for the AVIRIS Indian Pines and DAIS 7915 data.

Several observations can be made from the comparison given in Fig. 2.3. First and foremost, it is remarkable that both the AAP-SP and SD-SP clustering results present high consistency with the respective spectral correlation matrices. In addition, it can be seen that the band clusters found by AAP-SP (a statistical approach) exhibit some correspondence with regards to the band clusters found by SD-SP (a physically

Table 2.3: Description of the imaging spectrometers in the DAIS 7915 system. The third spectrometer covers two different spectral ranges. The fifth spectrometer is given by a single band (at 20000 nanometers) which provides the value of celsius temperature multiplied by 10.

Spectrometer	Wavelength range (nm)	Band numbers	Bandwidth
1	400 – 1000	1–32	15-30 nm
2	1500 – 1800	33–40	45 nm
3	2000 – 2500	41–72	20 nm
3	3000 – 5000	73	2 μm
4	8000 – 12600	74–79	0.9 μm
5	20000	80	–

driven approach). For the AVIRIS data, the first cluster found by AAP-SP corresponds almost exactly to the spectrometer A, while spectrometers C and D (both in the infra-red region) also correspond to the same cluster found by AAP-SP, with some noise in the areas between the clusters which correspond to boundary areas between different spectrometers. For the DAIS 7915 data, we can observe similar correspondences. Specifically, the first spectrometer in DAIS 7915 (corresponding to the visible part of the spectrum) is associated with two clusters identified by AAP-SP. The second and third spectrometers correspond to the infra-red region and are both associated with the same cluster by the AAP-SP. The fourth spectrometer is located in a completely different region and hence it is detected by a different cluster by AAP-SP. Finally, the fifth spectrometer collects only one band that provides the value of celsius temperature multiplied by 10 and is detected by a single independent cluster by AAP-SP. Since the information about the different imaging spectrometers is known in advance, it seems more practical to use the SD-SP to establish the spectral band groups instead of AAP-SP, as the latter relies on input parameters that require hand-tuning, but in both cases the identified band clusters appear meaningful in terms of the wavelength regions they describe. In the following subsection, we describe the MCS system used to obtain a final classification result from the partitions generated by the two considered clustering strategies (AAP-SP and SD-SP) using the band reassignment strategy described in the previous subsection.

2.3.4 Multiple Classifier System (MCS)

In this section, we describe the MCS system used to provide the final classification result from the ρ partitions of the original hyperspectral image obtained after band clustering and reassignment. The classifier system adopted in this chapter is the MLR*sub* method [131], which has been shown to be an effective probabilistic classifier that takes advantage from subspace projection to better separate the classes. Specifically, in the MLR*sub* 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. Since we are dealing with different partitions (or views) on the original hyperspectral data, we need a decision rule to fuse the individual classifications obtained by the MLR*sub*

from the different partitions. Let $\mathbf{p}_m(i)$ be the probability obtained by the MLR*sub* classifier for a given pixel i and partition m. In this chapter, we use a simple majority voting strategy to combine the results obtained from all the partitions. Specifically, the probabilities resulting from all the different partitions in a given pixel are modeled by:

$$\widehat{\mathbf{p}}(i) = \quad \frac{1}{\rho}\sum_{m=1}^{\rho}\mathbf{p}_m(i), \tag{2.1}$$

where ρ is the number of partitions. The final class label for pixel i is determined by majority voting as follows:

$$\widehat{y}(i) = \quad \arg\max_{k\in\{1,\ldots,K\}} \widehat{p}^{(k)}(i), \tag{2.2}$$

where K is the number of classes, $\widehat{p}^{(k)}(i)$ is the probability corresponding to class k for a given pixel i, and $\widehat{\mathbf{p}}(i) = \{\widehat{p}^{(1)}(i), \ldots, \widehat{p}^{(K)}(i)\}$. It should be noted that, in order to include spatial information in the classification stage, we can optionally perform a final spatial regularization by means of the Markov rand field (MRF) [134], as indicated by the MLR*sub*MRF method in [131]. The final classification thus combines the spectral information (obtained from different partitions or views of the original spectral data, which can be derived using both AAP-SP or SD-SP) and the spatial information, included by means of an MRF regularizer after combining the individual classification results for the different partitions using majority voting. In the following section, we evaluate the classification accuracy of the proposed framework (with and without MRF-based postprocessing) using the well-known AVIRIS Indian Pines and DAIS 7915 data sets.

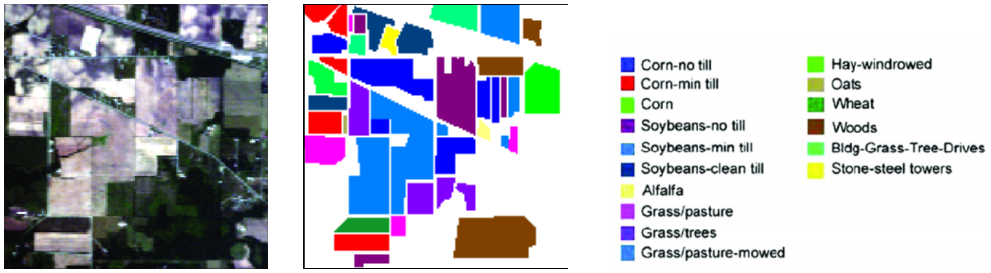
2.4 Experimental Results

In this section, we evaluate the proposed classification framework using two real hyperspectral data sets. Before reporting our experiments with these two scenes, we emphasize that we have optimized the parameter settings in order to obtain the best performance from each individual method in the chain. Specifically, we have empirically set the number of spectral partitions to $\rho = 5$. For the MLR*sub* classifier, we have optimized all the involved parameters following the procedure described in [131]. For the optional MRF spatial regularization process, the smoothing parameter is defined according to the indications in [135]. In order to study the impact of the MRF-based spatial regularization, we will present results with and without MRF spatial postprocessing for clarity.

The remainder of this section is organized as follows. In subsection 2.4.1 we describe the considered hyperspectral scenes in detail. Subsection 2.4.2 discusses and compares the spectral partitioning results obtained using random spectral partitioning (R-SP), AAP-SP and SD-SP. Subsections 2.4.3 and 2.4.4 respectively illustrate the performance of the proposed approach (implemented with both AAP-SP and SD-SP) with the AVIRIS Indian Pines and DAIS 7915 data sets.

2.4.1 Hyperspectral Data Sets

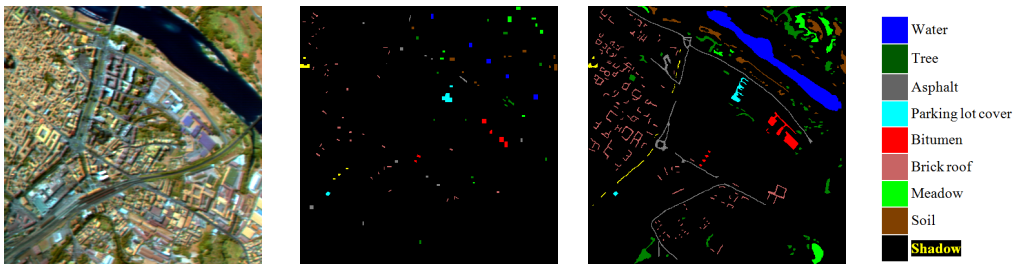
The AVIRIS Indian Pines data set[3], displayed in Fig. 2.4(a), comprises 145 × 145 pixels and was collected over Northwestern Indiana in June 1992. As shown by Fig. 2.4(b), a total of 10366 pixels are available in the labeled ground-truth, including 16 mutually exclusive classes. Although some of the bands are considered to be corrupted by water absorption features and noise, we will use all of them since the MLR*sub* classifier considered in this chapter has the capability to manage noise and mixtures by projecting the data to the class indexed subspaces.



(a) False color representation (b) Reference

Figure 2.4: The AVIRIS Indian Pines data set collected over Northwestern Indiana in June 1992

The DAIS 7915 Pavia City Center data set is illustrated in Fig. 2.5(a). It comprises 400 × 400 pixels and was captured over the central area of Pavia City in Italy, in 2001. A set of training samples is distributed with the data [see Fig. 2.5(b)] covering 9 mutually exclusively classes, while the total number of available labeled samples comprises 13275 pixels that do not overlap with the training samples [see Fig. 2.5(c)]. Similarly to the case of AVIRIS data, in our experiments, we use all the available spectral bands. Bearing in mind that the thermal spectrometer captures spectral bands with much higher quantization levels than other spectrometers, we normalize the data values to the scale [0,1] before processing.



(a) False color representation (b) Training samples (c) Test samples

Figure 2.5: The DAIS 7915 Pavia City Center data set collected over the city of Pavia, italy, in 2001.

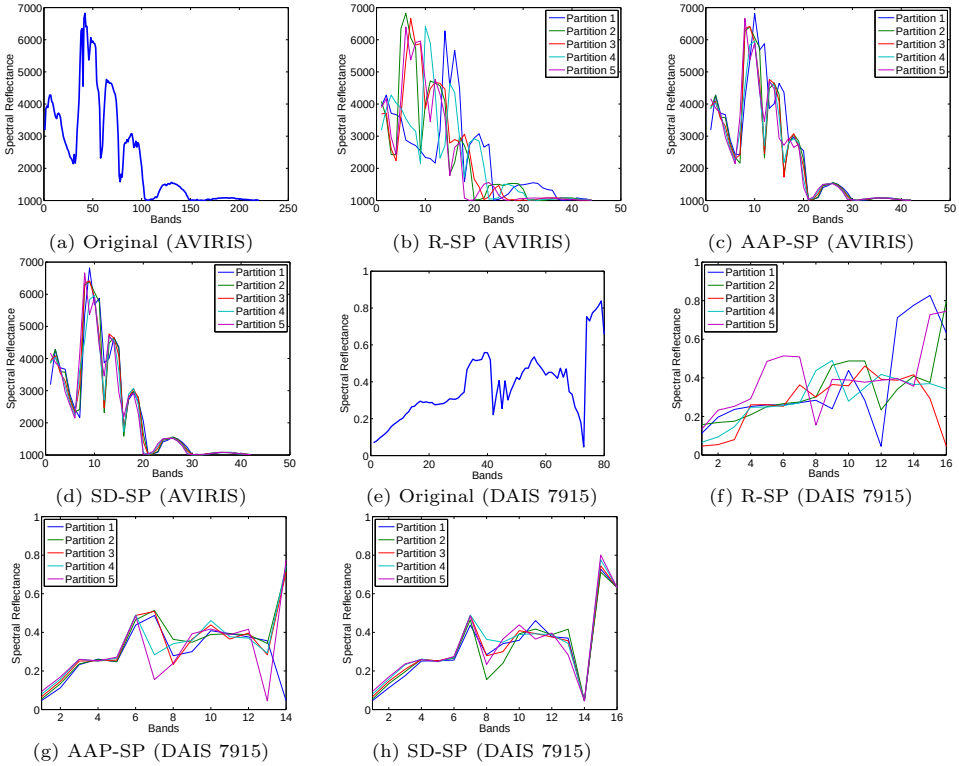
[3]https://engineering.purdue.edu/biehl/MultiSpec/hyperspectral.html

2.4.2 Comparison of Spectral Partitioning Results

Before providing a comparison between our presented methods, we first introduce R-SP as a baseline method for assessment. First, the spectral bands of the original hyperspectral image are randomly permuted. With the randomly ordered spectral bands as input, we employ the band reassigning method described 2.3.2 to generate final random spectral partitions. Specifically, we treat all the randomly ordered bands as one band group. By setting the number of partitions to $\rho = 5$, we perform band reassignment on both data sets: AVIRIS Indian Pines and DAIS Pavia City Center. The obtained partitions had 44 (R-SP), 41 (AAP-SP), 42 (SD-SP) spectral bands for the AVIRIS data, and 16 (R-SP), 14 (AAP-SP) and 16 (SD-SP) spectral bands for the DAIS 7915 data. Fig. 2.6 illustrates the (dimensionally reduced) spectral signatures obtained after spectral partitioning and band reassignment for a random pixel [located at spatial coordinates (100,100)] in the AVIRIS Indian Pines and DAIS 7915 Pavia City Center data sets. From Fig. 2.6 we can conclude that the spectral signatures obtained after spectral partitioning and band reassignment retain the shape of the original spectral signature in the pixel, which means that the partitions are able to capture the spectral characteristics of the whole spectral signature while reducing its dimensionality (this property cannot be kept by the R-SP). On the other hand, the different spectral signatures obtained for the same pixel exhibit some variability, which means that we can generate multiple views of the original spectral data which are all realistic since they are obtained by reassigning the spectral information already comprised by the original signature. For R-SP, the spectral partitions showed great variability due to the random permutation of band orders. If we compare the AAP-SP and SD-SP partitioning strategies we can observe similar results, but the AAP-SP uses a statistically-guided framework while the SD-SP uses a more physically-driven framework. Based on these properties, we conclude that the considered partitioning strategies are able to exploit the different characteristics of the spectral signature in each of the considered partitions. In the following subsections, we evaluate the differences obtained in the final classification when considering the two aforementioned partitioning strategies.

2.4.3 Experiments with the AVIRIS Indian Pines data

In this section, we describe the results obtained after applying the proposed classification framework (using the original spectral information and the multiple views provided by R-SP, AAP-SP, and SD-SP spectral partitioning) for the AVIRIS Indian Pines scene. Table 2.4 shows the overall (OA), average (AA) and κ statistic, as well as the individual classification accuracies, obtained after using 320 randomly selected samples (around 20 samples per class) for training and the rest of the labeled samples [see Fig. 2.4(c)] for testing. If the number of samples available in the ground truth image is smaller than 20, we take half of the total samples in that class for training and the other half for testing, which is the same for the case of the DAIS 7915 Pavia City Center data set. In our experiments, we considered the classification results obtained by the MLR*sub* and also by the MLR*sub*MRF which applies MRF-based spatial postprocessing after the spectral-



(a) Original (AVIRIS) (b) R-SP (AVIRIS) (c) AAP-SP (AVIRIS)

(d) SD-SP (AVIRIS) (e) Original (DAIS 7915) (f) R-SP (DAIS 7915)

(g) AAP-SP (DAIS 7915) (h) SD-SP (DAIS 7915)

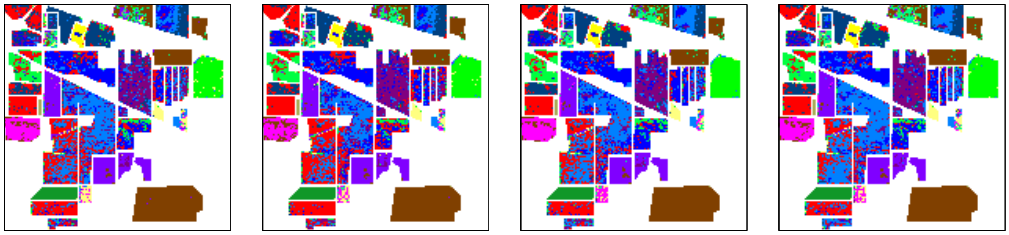
Figure 2.6: Spectral signatures of the pixel at spatial location (100,100) in the original AVIRIS Indian Pines data (a), and spectral signatures obtained after spectral partitioning with R-SP (b), AAP-SP (c) and SD-SP (d) and band reassignment for the AVIRIS Indian Pines data. Spectral signature of the pixel at spatial location (100,100) in the original DAIS 7915 Pavia City Center data (e), and spectral signatures obtained after spectral partitioning with R-SP (f), AAP-SP (g) and SD-SP (h) and band reassignment for the DAIS 7915 Pavia City Center data.

based classification. All the experiments were repeated 50 times and the results reported on Table 2.4 are in fact the average scores after 50 Monte Carlo runs. As shown by Table 2.4, both the AAP-SP and the SD-SP obtained very similar results in this experiment. It is also clear that the classification results after spectral partitioning were better than those obtained using the full spectral signatures in the original hyperspectral data, which indicates that the multiple views provided on the original spectral information were useful in terms of reducing the curse of the dimensionality, which was particularly useful given the limited number of training samples used for classification purposes. In addition, when the MRF-based postprocessing was used, the classification accuracies increased significantly. Both AAP-SP and SD-SP outperformed R-SP. In comparison with AAP-SP, SD-SP obtained quite similar classification results when the MLR*sub* was used, while the incorporation of spatial information by means of MRF-based postprocessing led to more significant differences between the two methods.

The experimental results reported on Table 2.4 suggest that even when compared with the MLR*sub* approach which was shown in [131] to be able to handle high-dimensional

Table 2.4: Overall, average and individual class accuracies [%] and κ statistic obtained by the proposed classification framework (with the original spectral information and with R-SP, AAP-SP and SD-SP spectral partitioning) for the AVIRIS Indian Pines scene data. In all cases, only 320 training samples were used.

Class	# samples	MLR*sub*				MLR*sub*MRF			
		Original	R-SP	AAP-SP	SD-SP	Original	R-SP	AAP-SP	SD-SP
C1	54	84.99	88.76	90.00	89.82	97.76	97.82	97.76	97.88
C2	1434	47.86	52.29	54.38	55.96	61.92	65.66	71.51	71.46
C3	834	66.53	65.91	64.65	64.25	84.53	81.35	80.47	79.92
C4	234	72.98	76.49	78.14	79.21	95.48	97.87	97.17	97.42
C5	497	77.46	78.34	81.76	82.15	87.25	88.84	89.01	88.99
C6	747	91.78	91.72	91.83	92.00	97.48	97.70	97.65	97.70
C7	26	93.67	96.33	94.67	95.00	98.00	100	98.67	99.00
C8	489	94.89	95.85	95.45	95.39	99.56	99.62	99.51	99.51
C9	20	98.00	98.20	98.00	98.20	100	100	100	100
C10	968	58.30	65.57	67.92	67.50	74.33	85.73	88.36	88.13
C11	2468	49.09	47.04	50.23	49.18	61.96	60.73	67.96	64.99
C12	614	63.18	68.08	70.32	71.03	79.59	84.07	85.87	86.63
C13	212	99.64	99.50	99.49	99.50	99.92	99.87	99.84	99.87
C14	1294	98.26	98.54	98.25	98.17	99.53	99.52	99.35	99.48
C15	380	18.83	21.30	26.84	27.75	30.37	42.84	49.65	51.57
C16	95	93.91	95.24	95.32	95.40	99.01	99.60	99.57	99.49
Overall accuracy		65.75	67.10	68.77	68.78	77.04	78.94	82.01	82.15
Average accuracy		75.59	77.45	78.58	78.78	85.42	87.58	88.90	89.75
κ statistic		61.40	62.99	64.81	64.85	74.15	76.36	79.72	79.78



(a) Original (b) R-SP (c) AAP-SP (d) SD-SP

(e) Original (MRF) (f) R-SP (MRF) (g) AAP-SP (MRF) (h) SD-SP (MRF)

Figure 2.7: Some of the classification maps obtained by the proposed classification framework (with the original spectral information (a) and with R-SP (b), AAP-SP (c) and SD-SP (d) spectral partitioning) for the AVIRIS Indian Pines scene. Spatial smoothness of labels are also regularized by MRF (e-h) in order to promote the classification performance. In all cases, only 320 training samples were used as input of the basic classifier MLR*sub*.

problems in hyperspectral classification with limited training samples, our proposed classification framework exhibits the potential to improve the classification results by

Table 2.5: McNemar's test for the different classification results obtained from the AVIRIS Indian Pines data set.

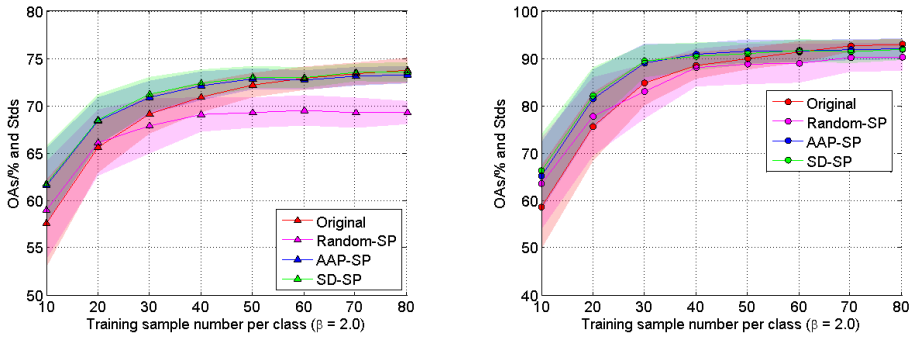
	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) MLR*sub*	492.85	126.33	1941.81	163.53	1576.03	4.22	462.13
(2) MLR*sub*MRF		151.86	1044.75	121.25	707.45	573.96	**0.07**
(3) MLR*sub*-AAPSP			1507.08	**3.15**	1077.04	180.82	153.14
(4) MLR*sub*MRF-AAPSP				1450.98	221.63	2095.29	965.60
(5) MLR*sub*-SDSP					1083.57	211.06	116.47
(6) MLR*sub*MRF-SDSP						1986.49	835.47
(7) MLR*sub*-RSP							619.09
(8) MLR*sub*MRF-RSP							

feeding multiple views of the original hyperspectral data into an MCS. At this point, it is also worth noting that all the original spectral information is used when the AAP-SP and the SD-SP partitioning approaches are used. In fact, these approaches just provide a mechanism to reassign the original spectral information into partitions that are intelligently created (according to statistical or physical principles) and which provide multiple views of the original spectral information that are then used to improve the final classification results. As expected, the inclusion of spatial information by means of an MRF-based spatial postprocessing can improve even further the obtained classification results.

For illustrative purposes, Fig. 2.7 shows some of the classification maps obtained in this experiment with the AVIRIS Indian Pines scene. Each map corresponds to one of the 50 Monte Carlo runs conducted in each case. As shown in Fig. 2.7, the use of AAP-SP and SD-SP leads to smoother classification maps as compared to the case in which the original spectral information is used, and the inclusion of MRF-based spatial postprocessing improves the spatial consistency of the final classification maps.

In order to further test the statistical significance of the differences in classification results reported in Table 2.4, we performed McNemar's Test[136] to compare the misclassification degree of Fig. 2.7 with significance level $\alpha = 0.05$ (meaning the chi-squared distribution $\chi^2_{\alpha,1} = 3.841459$). Table 2.5 illustrates the values obtained in the McNemar's test. Here, the values in the first and second row are larger than $\chi^2_{\alpha,1}$, meaning that our presented methods obtained different results with regards to those found by the original classifiers MLR*sub* and MLR*sub*MRF. On the other side, the classification differences between the R-SP and the original classifiers turned to be relatively lower. This indicates that, in comparison with the R-SP, our presented methods could obtain improved enhanced classification results.

To conclude this section, Fig. 2.8 evaluates the OA of our proposed framework as a function of the number of training samples. Here, we used both the MLR*sub* classifier and MLR*sub* plus MRF-based postprocessing (MLR*sub*MRF) to refine the final classification results. Several conclusions can be obtained from Fig. 2.8. First and foremost, the SD-SP spectral partitioning leads to slightly better classification accuracies than those obtained when using the AAP-SP. For instance, when the number of training samples was below 30 per class, both AAP-SP and SD-SP obtained about 4% increase in classification accuracy with regards to using the original spectral information. As the number of training samples per class was increased, the advantages of spectral partitioning over



(a) MLR*sub* (b) MLR*sub*MRF

Figure 2.8: Overall classification accuracies as a function of the number of training samples obtained by the proposed classification framework (with the original spectral information and with R-SP, AAP-SP and SD-SP spectral partitioning) for the AVIRIS Indian Pines scene. The solid lines represent the average of 50 Monte Carlo runs, while the colored area around the lines represent the standard deviation around the mean.

using the original spectral information were less significant. Meanwhile, when the number of training samples increased to 60 per class, the advantages of SD-SP and AAP-SP were less relevant for MLR*sub* and MLR*sub*MRF. Again, the differences between using the AAP-SP and SD-SP were not significant in this particular example. In all cases, both AAP-SP and SD-SP outperformed R-SP, particularly as the number of training samples increased. Finally, the standard deviations obtained when using both the SD-SP and AAP-SP are smaller than those obtained using the original spectral information and the R-SP. This indicates that our spectral partitioning framework generally leads to more stable and robust classification results.

We also note that the computational complexity of the spectral partitioning stage depends mainly on the number of spectral partitions. For example, if the user sets the number of spectral partitions to ρ, the approximate computational cost of our presented approaches AAP-SP and SD-SP would be about ρ times the cost of the original MLRsub classifier. And when involving spatial information with the MRF post-processing, additional computational time is needed for both the original MLR*sub*MRF classifier and our presented AAP-SP, SD-SP methods. Although the classification of a partition is similar to the classification of the original data set, we can perform the classification of different partitions in parallel exploiting multi-core architectures of co-processors such as GPUs, hence the complexity of the method can be kept within similar levels as the original classification using straightforward implementation.

2.4.4 Experiments with the DAIS7915 Pavia City Center scene

In this section, we conduct experiments with a data set collected by a different spectrometer. Table 2.6 displays the OA, AA, κ and individual class accuracies obtained by the proposed classification framework (with the original spectral information and with R-SP, AAP-SP and SD-SP spectral partitioning) using the DAIS 7915 Pavia City

Table 2.6: Overall, average and individual class accuracies [%] and κ statistic obtained by the proposed classification framework (with the original spectral information and R-SP, with AAP-SP and SD-SP spectral partitioning) for the DAIS 7915 Pavia City Center scene data. In all cases, only 90 training samples (10 per class) were used.

Class	# samples	MLR*sub*				MLR*sub*MRF			
		Original	R-SP	AAP-SP	SD-SP	Original	R-SP	AAP-SP	SD-SP
Water	4083	99.90	100	100	100	99.99	100	100	100
Tree	2302	89.48	90.99	88.86	89.80	93.40	92.71	92.74	92.83
Meadow	1136	83.55	90.35	89.37	88.60	89.88	95.63	94.26	93.90
Soil	1301	69.97	70.20	74.70	74.22	70.10	67.30	72.55	72.50
Asphalt	1630	81.95	85.31	87.32	86.64	86.74	87.93	89.45	88.91
Parking lot	132	81.15	77.56	82.06	82.25	83.87	73.72	77.68	78.68
Brick roof	2041	89.34	96.54	94.90	96.33	93.38	98.80	97.58	97.80
Bitumen	491	87.12	87.75	86.18	90.84	88.90	87.01	85.73	91.27
Shadow	159	87.96	100	89.72	91.63	92.50	94.90	93.40	94.75
Overall accuracy		89.13	91.53	91.58	91.91	91.76	92.70	93.06	93.24
Average accuracy		85.60	87.85	88.13	88.92	88.77	88.68	89.20	90.07
κ statistic		86.79	89.70	89.69	90.15	89.97	91.10	91.54	91.76

Table 2.7: McNemar's Test for different classification results with the DAIS 7915 Pavia City Center scene data

	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) MLR*sub*	365.37	240.33	557.80	460.05	741.48	135.02	420.25
(2) MLR*sub*MRF		3.86	115.01	20.56	242.27	34.60	42.44
(3) MLR*sub*-AAPSP			184.01	59.88	265.45	18.30	59.59
(4) MLR*sub*MRF-AAPSP				33.50	33.45	239.73	25.94
(5) MLR*sub*-SDSP					138.18	151.30	7.703
(6) MLR*sub*MRF-SDSP						403.98	141.21
(7) MLR*sub*-RSP							168.55
(8) MLR*sub*MRF-RSP							

Center data. In our experiment we used only 90 training samples (around 10 pixels per class) randomly selected from the available training samples in Fig. 2.5(b), and then we used the mutually exclusive labeled samples in Fig. 2.5(c) for testing. In our experiments, we considered the classification results obtained by the MLR*sub* and also by the MLR*sub*MRF which applies MRF-based spatial postprocessing after the spectral-based classification. All the experiments were repeated 100 times and the results reported on Table Table 2.6 are in fact the average scores after 100 Monte Carlo runs. As shown by Table 2.6, the use of SD-SP spectral partitioning resulted in the highest OA, AA and κ statistic values among all considered strategies, considering both MLR*sub* and MLR*sub*MRF.

At this point, we emphasize that the results reported on Table 2.6 were obtained using a very limited number of training samples per class (much smaller than the full training set reported on Fig. 2.5(b)). The high classification accuracies reported in this case can be partly attributed to the influence of the water class, which has a large number of samples and was very accurately classified in all cases. For illustrative purposes, Fig. 2.9 shows some of the classification maps obtained in these experiments. Each map corresponds to one of the 100 Monte Carlo runs conducted in each case respecting to Table 2.6. As shown in Fig. 2.9, both AAP-SP and SD-SP lead to smoother classification maps and, when the spatial information was included, the final maps were of great quality, in particular, when the SD-SP spectral partitioning strategy was used. In addition, Table2.7 illustrates the statistical differences between the classification reported on Table2.6, by

measuring the McNemar's test for the results in Fig. 2.9 (with significance level $\alpha = 0.05$ and the chi-squared distribution $\chi^2_{\alpha,1} = 3.841459$). From Table2.7, we can conclude that our presented methods (AAP-SP and SD-SP) obtained similar classification accuracies but higher than those obtained by R-SP.

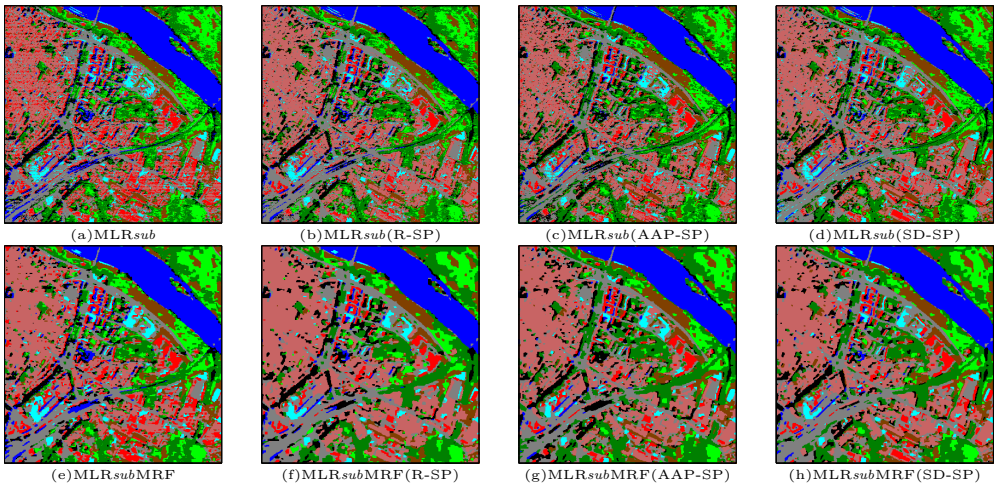
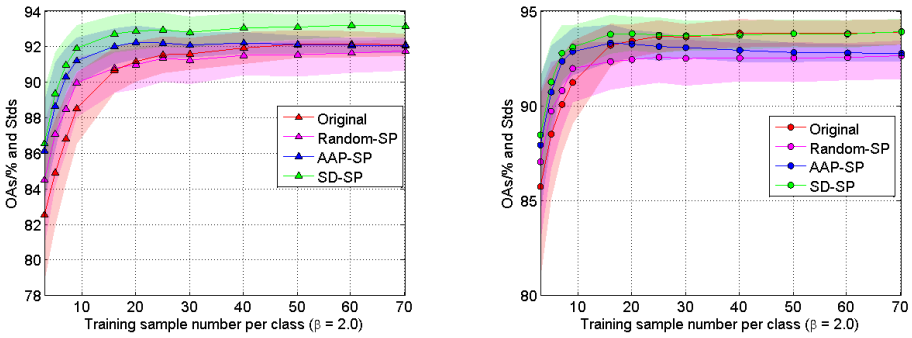


Figure 2.9: Some of the classification maps obtained by the proposed classification framework (with the original spectral information and with R-SP, AAP-SP and SD-SP spectral partitioning) for the DAIS 7915 Pavia City Center scene. Spatial smoothness of labels are also regularized by MRF (e-h) in order to promote the classification performance. In all cases, only 90 training samples (10 per class) were used.

To conclude this section, Fig. 2.10 evaluates the OA of our proposed framework as a function of the number of training samples. Here, we again used the MLR*sub* classifier and MRF-based spatial postprocessing to refine the final classification results. From Fig. 2.10, it is remarkable that the proposed framework (implemented with SD-SP) provides better results when compared with the same approach using AAP-SP and R-SP for spectral clustering or the original spectral information. This was particularly the case as the number of training samples was increased, in which the standard deviation also indicates less variability when spectral clustering approaches were used. Both SD-SP and AAP-SP obtained better results than R-SP, especially as the training samples were increased. Similar results were reported in our experiments with the AVIRIS Indian Pines data. The proposed framework based on spectral partitioning is particularly useful when very limited training sets are available (this is the most common case in remote sensing applications, due to the cost and effort associated with laborious ground campaigns).

2.5 Summary and future directions

In this chapter, we have developed a new framework for spectral partitioning of hyperspectral data which uses band similarity grouping and then band reassignment to reduce dimensionality and provide multiple views of the hyperspectral data while keeping the relevant information contained in the original spectral signatures. In this framework,



(a) MLRsub (b) MLRsubMRF

Figure 2.10: Overall classification accuracies as a function of the number of training samples obtained by the proposed classification framework (with the original spectral information and with R-SP, AAP-SP and SD-SP spectral partitioning) for the DAIS 7915 Pavia City Center scene. The solid lines represent the average of 100 Monte Carlo runs, while the colored area around the lines represent the standard deviation around the mean.

two different spectral partitioning strategies have been tested: one driven by statistical clustering, and another one driven by physical principles related to the design of the imaging spectrometer. The latter approach is quite innovative, as takes into account the physical characteristics of the system for partitioning the spectral bands collected by each spectrometer, and resampling them into different groups or partitions. Instead of ideally considering the spectral signatures as unique continuous signals, our proposed approach considers each spectral signature as a signal that is contributed from different instrument sources, and thus partition the data accordingly prior to conducting a probabilistic classification step that includes the possibility to include a spatial regularization at the end of the process. We have evaluated the methodology by using the AVIRIS and DAIS 7915 systems. Our experimental results indicate that the proposed strategy can exploit the information contained in the different spectrometers in order to improve the classification results using limited training samples, meaning that we can exploit the band groups corresponding to different spectrometers as an effective source of information for spectral partitioning and band reassignment prior to classification. In the future, we will use other imaging spectrometers and classification techniques in order to analyze the generality of the presented approach. We are also planning on using different numbers of bands for each partition (in the present configuration all partitions have the same number of bands) in order to better model the contributions from each spectrometer.

Chapter 3

Class-Oriented Spectral Partitioning

3.1 Summary

Remotely sensed hyperspectral images exhibit very high dimensionality in the spectral domain. As opposed to band selection techniques, which extract a subset of the original spectral bands in the image, spectral partitioning techniques reassign the original bands into subgroups that are then processed separately. From a classification perspective, this strategy has the advantage that all the original information in the hyperspectral data can be retained while addressing the curse of dimensionality given by the Hughes phenomenon. Even if spectral partitioning prior to classification has been widely used, the strategies adopted to perform such partitioning did not consider the diversity of spectral classes in the scene. In other words, available techniques are not driven by the information contained in the classes of interest, which can be very useful to perform the spectral partitioning in a more effective manner for classification purposes. To address this issue, in this chapter we present a new class-oriented spectral partitioning technique that exploits prior information about the classes by automatically ranking the spectral bands that are more useful for each specific class (instead of considering the hyperspectral image as a whole). The resulting multiple subgroups of bands with lower dimensionality are then fed to a multiple classifier system. Our experimental results, conducted with three different hyperspectral airborne images, suggest that the presented method leads to competitive results when compared to other state-of-the-art approaches in the field.

3.2 Introduction

Hyperspectral image analysis has developed significantly during the past two decades [7]. Hyperspectral data have been used in many different areas, including disaster monitoring, natural resources exploitation, environmental applications, etc. [106–108]. However, contrary to the rapid development of remote sensing technologies, the availability of training samples and labeled data has been quite limited [7]. As a result, many

hyperspectral image analysis methods (particularly, for supervised and semi-supervised classification) have focused on addressing the existing imbalance between high spectral dimensionality and limited labeled samples [137, 138]. In order to deal with the issue, dimensionality reduction technique has been widely developed and used [139–142]. Classic dimensionality reduction techniques can be separated into two categories: 1) band selection and 2) feature extraction. Considering the diversity of available algorithms and techniques, feature extraction turns out to be more flexible and widely used, including well-known approaches in the literature such as principal component analysis (PCA) [143], independent component analysis (ICA) [37], manifold learning (ML) [37, 144] and subspace-based approaches [39, 145]. However, feature extraction generally transforms the original information after projecting the data into a certain feature space [146], which may be a challenge for certain applications that require meaningful spectral signatures according to their physical interpretation [147, 148]. In turn, band selection algorithms are more effective in preserving the original information due to their capacity for selecting the most informative spectral bands among hundreds or even thousands of bands with great correlation and redundancy with supervised or unsupervised ways [149–151] while, at the same time, retaining their original physical meaning (i.e., even though some bands are discarded, there is generally no transformation of the retained bands into a different feature space). As a matter of fact, the discriminative information that allows a classifier to provide good performance is usually class-dependent and the relevant information may live in weak features/bands that are usually discarded or lost after subspace transformation or band selection. As a result, in practice, it is challenging to use either feature extraction or band selection for classification purposes.

Many techniques have been developed in order to address this problem. In [152], an efficient model selection procedure based on kernel alignment was developed for hyperspectral image classification. Resulting from this process, a weight (learned from the data) is assigned to each kernel so that both relevant and meaningless image features automatically emerge after training the model. In [153], an unsupervised band selection method called multi-objective optimization band selection (MOBS) was developed. The objective functions were optimized by a multi-objective evolutionary algorithm to find the best trade-off solutions. The final selected bands evolve from multiple possible instances of band selection during an iterative process, until the good stable performance of classification is achieved for different data sets. In order to facilitate the subsequent automatic interpretation image objects, high-level features are jointly integrated to infer the spatial and structural information encoded in the low-level and middle-level features [154]. Classic ensemble learning techniques such as bootstrap aggregation/bagging [155], AdaBoost [156] or cross-validated committees [157] can also be interpreted as feature fusion approaches that can take advantage of the relevant information provided by each classifier [158–160].

An alternative strategy used in recent developments is spectral partitioning, which can be seen as a kind of band selection approach that aims mainly at rearranging the original spectral bands in the hyperspectral image. However, as opposed to band selection,

spectral partitioning does not necessarily discard most of the original spectral bands to achieve lower dimensionality [56]. Instead, spectral partitioning generates several groups of band subsets from the original spectral bands, so that each band subset is a so-called spectral partition containing a much lower number of spectral bands as compared with the original hyperspectral image, meanwhile, the union of multiple subsets can make up to the full original image [54]. Therefore, spectral partitioning effectively provides multiple views of the original hyperspectral image by obtaining several subgroups that can simultaneously exploit most of the original spectral information in the hyperspectral scene without discarding a large proportion. In other words, different subgroups of spectral bands can be used to provide different classification results [54]. The diversity of classifiers constructed with the subgroups of bands provides the possibility to obtain a very robust classifier ensemble, which can be achieved by combining the classification results obtained from each of the subgroups using different ensemble learning strategies [161], such as bagging and boosting [156, 162], decision combination via majority voting [163], and multiple classifier systems (MCS) [158, 160]. The diversity obtained in the generation of the multiple views is one of the keys for successful spectral partitioning prior to classification [159, 160, 164].

3.2.1 Related work and motivations

Even if spectral partitioning prior to classification has already been used, the strategies adopted to perform such partitioning generally do not consider the information about the spectral classes in the scene. In other words, available techniques are not driven by the information contained in the classes of interest, which can be very useful to perform the spectral partitioning in a more effective manner for classification purposes [54, 165]. In high spatial resolution images, land-use classification can be successfully performed by interpreting the multiple spectral bands according to common classes in the scene [166]. With such class-oriented interpretation, one can also automatically perform target detection [167, 168]. However, in hyperspectral image analysis, it is generally difficult to anticipate which bands play a more relevant role for classification in a specific scene [169,170], as this information is generally class-dependent. In [152,171], the authors developed multiple kernel frameworks for hyperspectral image classification. The weight assigned to different kernel features was carefully established and interpreted. However, the specific role of spectral bands is more difficult to substantiate since it is often application-dependent. As a result, it is important that different spectral partitions derived from the original image represent different perspectives or views on the data [54, 141, 172, 173]. At this point, we clarify that these multiple views refer to the lower-dimensional partitions that are obtained by the proposed spectral partitioning framework.

Bearing in mind the aforementioned issues, our motivations to introduce a new spectral partitioning technique in this chapter can be summarized as follows: 1) to achieve the benefits of dimensionality reduction without actually discarding most of the spectral bands in the original image; 2) to facilitate multiple-classifier feature learning by considering different views (spectral partitions) of the original data; and 3) to explore the relevance of class-dependent features, which are often discarded by feature

extraction/selection techniques, all in the context of hyperspectral image classification. With the aforementioned ideas in mind, our main goal in this chapter is to develop a new strategy for spectral partitioning that can effectively exploit the spectral bands that are more relevant to each class of interest in the scene.

The newly presented spectral partitioning strategy consists of three steps. First of all, we rank the spectral bands that are more useful for each specific class (instead of considering the hyperspectral image as a whole). As a result, multiple subgroups of bands with lower dimensionality are selected in this step. Then, the obtained class-oriented spectral partitions are used as the input of a classifier ensemble strategy, MCS. For this purpose, we use two state-of-the-art classifiers: the support vector machine (SVM) [174] and the multinomial logistic regression (MLR) [175] to validate our presented method. In addition, we incorporate into the MCS the concept of classification with rejection to evaluate and improve our presented classifier [176,177]. This approach effectively merges the classification output from the spectral partitions with the classification obtained from the original image, providing a robust classification output. The proposed approach is compared to other available spectral partitioning methods, such as the one presented in [56].

The remainder of this chapter is organized as follows. Section 3.3 describes our proposed approach. In Section 3.4, we discuss experimental results obtained using three well-known hyperspectral data sets: the Reflective Optics System Imaging Spectrometer (ROSIS) Pavia University scene, the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Indian Pines scene, and the hyperspectral digital image collection experiment (HYDICE) Washington DC Mall scene. The experiments suggest that our presented method leads to competitive results when compared to other state-of-the-art approaches in the field. Conclusions and hints at plausible future research lines are given in section 3.5.

3.3 Class-oriented spectral partitioning method

3.3.1 Proposed spectral partitioning strategy

Let us denote by $\mathbf{x} \equiv \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \in \mathbb{R}^{d \times n}$ the original spectral signatures of the hyperspectral image $\mathbf{B}^d$ with n pixels indexed by $\mathcal{S} : \{1, 2, \cdots, n\}$ and d wavebands Ω ($|\Omega| = d$). Classification is the process that assigns each pixel $\mathbf{x}_i = \{x_{i1}, x_{i2}, \cdots, x_{id}\}, i \in \{1, \cdots, n\}$ with a label $y_i \in \mathcal{L} : \{1, 2, \cdots, N_c\}$, where N_c is the number of classes of interest in the scene. Spectral partitioning aims at separating and reassigning the d spectral bands Ω into a group of band subsets, namely spectral partitions $\{SP_i\}$, $i \in \{1, 2, \cdots, M\}$, so that:

$$\bigcup_{i \in \Omega} \{SP_i\} \subseteq \Omega. \tag{3.1}$$

This means that usually the union of the spectral partitions gives the whole set of

original spectral bands of the hyperspectral image. We also note that:

$$\begin{aligned} \emptyset \subseteq & \bigcap_{\forall i \neq j,\ i,j \in \Omega} \{SP_i, SP_j\}, \\ & \bigcup_{\forall i \neq j,\ i,j \in \Omega} \{SP_i, SP_j\} \supset \{SP_i\}, \end{aligned} \tag{3.2}$$

which means that two different spectral partitions may share mutual bands or not, while any given two spectral partitions should not be equivalent (note that one spectral partition might be the subset of another). Eqs. (3.1) and (3.2) generally describe the relationship of spectral partitions, meanwhile they also induce to finite number of solutions of the spectral partitioning problem. We generally need the spectral partitions to be different enough among each other in order to provide adequate diversity [161,162]. This brings different perspectives of the original hyperspectral image [54, 56]. If we let $|SP_i|$ be the number of bands in partition SP_i, we can infer from Eqs. (3.1) and (3.2) that $\sum_{i=1}^{M} |SP_i| \geq d$.

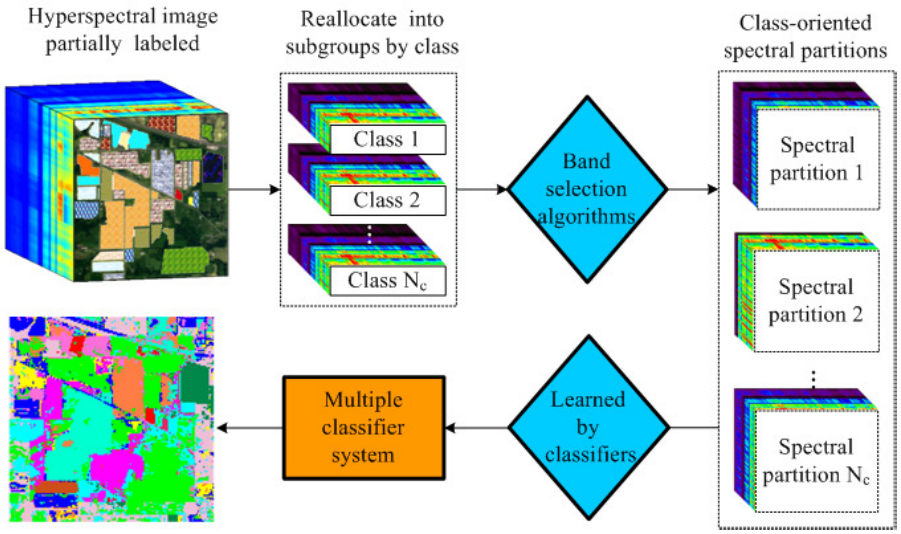


Figure 3.1: Flowchart of the proposed class-oriented spectral partitioning prior to classification approach.

With the aforementioned ideas in mind, Fig. 3.1 shows a general flowchart of our proposed class-oriented spectral partitioning approach prior to classification. First, training samples with labels $G = \{G_1, G_2, \cdots, G_{N_c}\}$ are separated into a certain number of subgroups G_i, where $G_i = \{\mathbf{x}_{G_i}, \mathbf{y}_{G_i}\}, i \in \{1, \cdots, N_c\}$ and the number of subgroups is given by the number of classes in the scene ($Nc = M$). Then, each subgroup G_i is used as input to a band selection algorithm that selects the most relevant bands using a specific criterion that is driven by the information contained in each class. The result is a set of Nc class-oriented spectral partitions $\{SP_i\}(i \in \{1, \cdots, N_c\})$ with much lower dimensionality as compared to the original hyperspectral image ($|SP_i| << d$). The spectral partitions are fed to an MCS that combines multiple SVM classifiers and

generates a final classification result.

Herein, we remark that the presented spectral partitioning strategy excludes by a large degree the influence of noise, outliers, and anomalous classes, and relies on the effectiveness of the band selection algorithms that are used in the spectral partitioning scheme. In the following, we describe the band selection techniques adopted in this chapter and the MCS used to generate the final classification output. In addition, we also introduce the concept of classification with rejection, as well as describe how we combine different classifiers to further improve the classification results using this concept.

3.3.2 Band selection algorithms

Band selection techniques intend to select an appropriate band subset from the original data set to represent the data according to some optimality criterion [114]. Generally, band selection can be understood as an exhaustive searching process for all possible cases: $(L_{|\Omega_{BS}|}) = \frac{d!}{(d-|\Omega_{BS}|)!|\Omega_{BS}|!}$, with Ω_{BS} being the selected bands and L being the number of all possible subset of selected bands, given a number of bands to be selected $|\Omega_{BS}|$ [114]. A general way to perform the searching process is to solve the following optimization problem:

$$\Omega^*_{BS} = \arg \max_{\Omega_{BS} \subset \Omega, |\Omega_{BS}|=n_{BS}} J(\Omega_{BS}), \tag{3.3}$$

where $n_{BS} = |\Omega_{BS}|$ is the number of selected bands in subset Ω_{BS} and $J(\Omega_{BS})$ establishes the relative importance of a given spectral band in Ω_{BS}. We have to note that, given the large number of spectral bands in a hyperspectral image, it is almost impossible to try all possible band combinations. We also note that, for different scenes, it is difficult to decide which spectral bands play a more relevant role, or to anticipate which combination is more useful. As a result, a general strategy has been to find a group of bands that are both of high quality (e.g., with low noise or not located in the water absorption region) while exhibiting high variance [170]. Consequentially, many available band selection algorithms are focused on defining target functions that calculate the priority score of a given spectral band [49, 178, 179].

In this chapter, we rely on two popular band selection algorithms: signal-to-noise ratio (SNR) [180] and band dependence minimization-based linearly constrained minimum variance (BDM-LCMV) [181] for evaluation purposes. We have selected these algorithms because their strategies for band selection are widely used, which makes the algorithms quite representative of existing band selection techniques. In this chapter, both algorithms are used to perform class-oriented band selection prior to classification. In the case of the SNR band selection algorithm, $J(\Omega_{BS})$ is defined with a noise-adjusted principal component based priority score, calculated for the l-th band of the image, $\mathbf{B}_l$, as follows:

$$J(\mathbf{B}_l) = \sum_{k=1}^{d} \mathbf{r}^2_{l,k} = \rho^{\mathrm{SNR}}_l, \tag{3.4}$$

where $\mathbf{r}^2_{l,k} = \sqrt{\lambda_{l,k}} \times \mathbf{v}_{l,k}$, for $l, k = 1, 2, \ldots, d$ and $\mathbf{B}_l$ is the $l-$th spectral data. Herein $\lambda_{l,k}$ is the set of eigenvalues of the noise-adjusted covariance matrix, and $\mathbf{v}_{l,k}$ denotes

their associated orthonormal eigenvectors. Here, we also use d to denote the number of wavelengths of the original hyperspectral image. Given a number of bands to be selected, Ω_{SB}, the SNR algorithm selects the first Ω_{SB} spectral bands with greater $J(\mathbf{B}_l)$ values from all the available bands. At this point, we emphasize that the SNR is one of the most widely used criteria to establish band priorities. The main assumption of our spectral partitioning method based on SNR (hereinafter SP-SNR) is that the spectral bands with higher SNR lead to better classification performance. This simple criterion is based on minimizing the impact of noise on the performance of the classifier. Consequentially, the SP-SNR is intended to minimize the impact of noise in classification performance by generating class-oriented spectral partitions with high SNR values.

SNR is a criterion related to single bands, while BDM-LCMV also includes a second constraint intended to minimize the similarity between the bands to be selected. From the viewpoint of classification, BDM-LCMV follows the assumption that a more informative band combination is generally given by a set of bands that are more distinct between each other. In the case of BDM-LCMV algorithm, the band priority score can be calculated as follows:

$$J(\mathbf{B}_l) = (\tilde{\mathbf{v}}_l^{\text{LCMV-CBS}})^T \widetilde{\sum}^{-1} \tilde{\mathbf{v}}_l^{\text{LCMV-CBS}}, \tag{3.5}$$

where:

$$\tilde{\mathbf{v}}_l^{\text{LCMV-CBS}} = \widetilde{\sum}^{-1} \mathbf{B}_l (\mathbf{B}_l^T \widetilde{\sum}^{-1} \mathbf{B}_l)^{-1} \mathbf{1}_N. \tag{3.6}$$

Here, $\tilde{\mathbf{v}}_l^{\text{LCMV-CBS}}$ represents the weight vectors associated to each spectral band and $\widetilde{\sum} = (1/d) \sum_{l=1}^{d} \mathbf{B}_l \times \mathbf{B}_l^T$ denotes the sample band correlation matrix, in which the problem described is referred to as LCMV-based constrained band selection[181]. In Eq. (3.6), $\mathbf{1}_N$ is an N-dimensional column vector with all 1's in its N components, and N denotes the number of columns of the spectral band $\mathbf{B}_l$. As a popular band selection algorithm, BDM-LCMV is also utilized in this chapter due to its capacity to select informative band combinations containing bands that are distinct between each other. With this issue in mind, we hold a reasonable expectation that BDM-LCMV may outperform SNR in terms of classification accuracies.

An important advantage of supervised band selection techniques such as those described before is that they can reduce the interference introduced by noise, outliers and unwanted pixels in a hyperspectral image scene. Another important advantage is that they are driven by the information contained in the classes of interest, turning the band selection process into a class-oriented one. An additional advantage in the case of BDM-LCMV is that it generally decreases the size of the used samples, which makes the inverse calculation faster and more accurate.

3.3.3 Classifier ensemble strategy

In this section, we describe the MCS system used to provide the final classification result from the Nc partitions of the original hyperspectral image, obtained by our newly

presented strategy for class-oriented spectral partitioning. The base classifier used for demonstration purposes in this chapter is the SVM [182], which has been a state-of-the-art classifier for remotely sensed hyperspectral data. Since we are dealing with different partitions (or views) of the original hyperspectral data, we need a decision rule to fuse the individual classifications obtained by the SVM from the different spectral partitions derived by our method. Let $\mathbf{p}_m(i)$ be the probability obtained by an SVM classifier for a given pixel i and partition m. $\mathbf{p}_m(i)$ provides the degree of membership of a pixel to different classes of interest. In this chapter, we use a soft ensemble strategy to combine the results obtained from all the partitions. Specifically, the probabilities resulting from all the different partitions in a given pixel are modelled by:

$$\widehat{\mathbf{p}}(i) = \frac{1}{N_c} \sum_{m=1}^{N_c} \mathbf{p}_m(i), \tag{3.7}$$

where m indexes the spectral partitions ($m \in \{1, \cdots, N_c\}$) and Nc is the number of partitions, which is equal to the number of classes, according to our interpretation in Fig.3.1. The final class label for pixel i is determined by maximum probabilistic voting as follows:

$$y_i = \arg \max_{k \in \{1, \cdots, N_c\}} \widehat{p}^{(k)}(i), \tag{3.8}$$

where $\widehat{p}^{(k)}(i)$ is the probability corresponding to class k for a given pixel i, and $\widehat{\mathbf{p}}(i) = \{\widehat{p}^{(1)}(i), \cdots, \widehat{p}^{(N_c)}(i)\}$.

3.3.4 Classification with rejection

In order to evaluate a set of classification outcome, four statistics are widely used: overall (OA), average (AA) and individual class accuracies, as well as the κ statistic. However, these statistics are unable to describe how confident a classifier is in the classification of a certain pixel. With limited training samples, the parametric optimization of a classifier sometimes results in high individual accuracies only for large classes (i.e., those with a significant number of pixels in the scene). However, the need for correctly classified pixels usually surpasses the need for high overall accuracies [176]. Hence, in this chapter, we exploit two additional statistical metrics to further refine the classification results provided by our proposed method.

Assuming that $\mathcal{R}$ is the set of rejected samples ($\bar{\mathcal{R}}$ is the set of non-rejected samples) and $\mathcal{C}$ is the set of correctly classified samples ($\bar{\mathcal{C}}$ is the set of incorrectly classified samples), we consider the *nonrejected accuracy* $\mathcal{A}$ and the *classification quality* $\mathcal{Q}$ [177], which are given as follows:

$$\mathcal{A} = \frac{|\mathcal{C} \bigcap \bar{\mathcal{R}}|}{|\bar{\mathcal{R}}|}, \tag{3.9}$$

and

$$\mathcal{Q} = \frac{|\mathcal{C} \bigcap \bar{\mathcal{R}} + \bar{\mathcal{C}} \bigcap \mathcal{R}|}{|\bar{\mathcal{R}} \bigcup \mathcal{R}|}. \tag{3.10}$$

In this chapter, we adopt a non-parametric way to simplify the rejection process when calculating the statistics $\mathcal{A}$ and $\mathcal{Q}$ as a function of the rejected fraction, following previous developments in [177]. By sorting the maximum probabilities of all pixels' probabilistic output, we can easily decide the rejections by accepting pixels with higher maximum probabilities and rejecting the remaining pixels. When different classifiers exhibit different confidence in the classification of the same pixels, this means that a rejected pixel by one classifier can hold higher confidence by another classifier. This makes it possible to further improve the classification results by combining the pixels with confident classifications by different techniques. In order to obtain a combined classification $\widehat{\mathbf{Y}}$, we define the acceptance of a classifier m as:

$$\mathcal{S}_{acc} = \text{where}(\mathbf{p}_m(i) > \lambda), \tag{3.11}$$

where λ is a constant decided by how many pixels we would like to reject. Then the rejection set from classifier n can be defined as:

$$\mathcal{S}_{rej} = \text{where}(\mathbf{p}_n(i) < \lambda). \tag{3.12}$$

Instead of using lambda, the rejection fraction (ratio of rejection) was recommended by the works [176,177]. The main reason is that it is easier to compare different methods under the rejection fraction. As a result, in our work, we used the ratio of rejection to illustrate the obtained results. As for the selection of the rejection fraction, it is generally defined by the user according to the desired accuracy after rejection. With the above formulation in mind, two classifiers can be simply combined as follows:

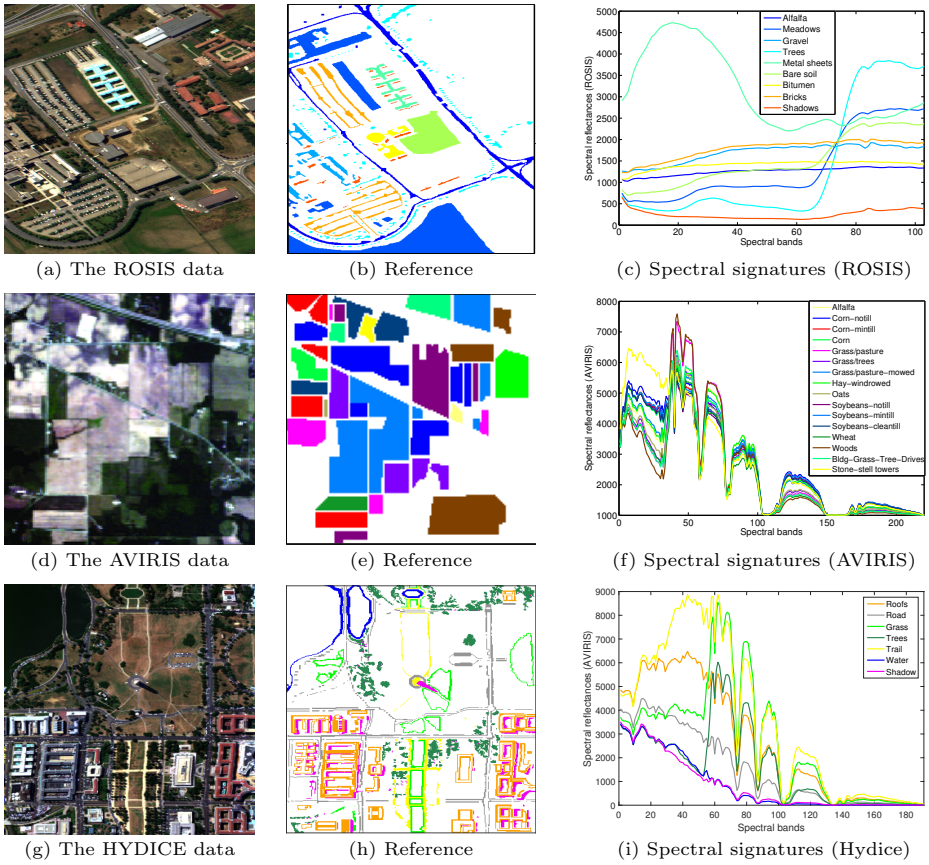
$$\widehat{\mathbf{Y}} = \arg\max \ \mathbf{p}_m(\mathcal{S}_{acc}) \bigcup \mathbf{p}_n(\mathcal{S}_{rej}), \tag{3.13}$$

where $\widehat{\mathbf{Y}}$ is the final classification map obtained after combining the output provided by the two classifiers. In practice, we just need to select a value for the rejection fraction when combining two classification results. To do this, we search the whole solution space of rejection fractions [0.01, 0.99] with the interval of 0.01 and then select the combination that reaches the highest overall accuracy. In this chapter, we use this simple concept to effectively merge the classification output of the aforementioned spectral partitioning strategy with the classification obtained from the original image, thus exploiting both strategies (i.e., class-dependent and image-dependent) to provide our final classification output.

3.4 Experimental results

In this section, we evaluate the presented spectral partitioning method by using three well-known hyperspectral data sets: a reflective optics system imaging Spectrometer (ROSIS) data set collected over Pavia University, Italy, an airborne visible infra-red imaging spectrometer (AVIRIS) data set collected over the Indian Pines region, and a hyperspectral digital image collection experiment (HYDICE) data set collected over Washington DC Mall. We first describe the three hyperspectral data sets. Then, we discuss the results obtained by using the proposed classification framework based on spectral partitioning, using the aforementioned data sets.

3.4.1 Hyperspectral images used in experiments



(a) The ROSIS data (b) Reference (c) Spectral signatures (ROSIS)

(d) The AVIRIS data (e) Reference (f) Spectral signatures (AVIRIS)

(g) The HYDICE data (h) Reference (i) Spectral signatures (Hydice)

Figure 3.2: Experimental hyperspectral data sets along with their ground-truth and average spectral signatures per class.

The first hyperspectral image used in our experiments [see Fig. 3.2 (a)] was collected by ROSIS over the University of Pavia, Italy. The data set consists of 115 spectral bands between 0.4 and 1.0 μm, with size of 610×340 pixels. The noisy bands had been removed, yielding 103 spectral bands. The ground-truth image in Fig. 3.2(b) contains 9 ground-truth classes, 3921 fixed training samples distributed together with the data, and 40002 test samples [3, 183].

The second hyperspectral image used in our experiment is the well-known AVIRIS Indian Pines[1] data set, displayed in Fig. 3.2(d). It comprises 145 × 145 pixels and was collected over Northwestern Indiana in June 1992. As shown by Fig. 3.2(e), a total of 10366 pixels are available in the labeled ground-truth, including 16 mutually exclusive classes. In the following experiments with this data set, we use 640 randomly selected training samples in total to illustrate the performance of the method using limited training sets. We also conduct 20 Monte Carlo runs to ensure statistical significance.

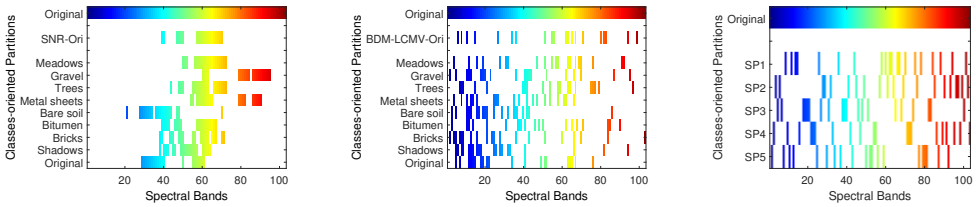
[1] https: //engineering.purdue.edu/biehl/MultiSpec/hyperspectral.html

Although some of the bands are considered to be corrupted by water absorption features and noise, we will use all of them since the considered band selection algorithms have the capacity to automatically screen and select the most useful spectral bands.

The third hyperspectral image used in experiments was collected by HYDICE over the Washington DC Mall. It comprises 191 bands with 3 m spatial resolution. This image originally contains 1280 lines with 307 pixels in each line. The test image used in experiments and its ground-truth are shown in Figs. 3.2(g) and 3.2(h), respectively. The image is available online [2]. In our experiments, we use 26981 labeled samples collected from 7 ground-truth classes.

For the three data sets, we plot the average spectral signatures of each class in Figs. 3.2 (c), 3.2(f) and 3.2(i) to give an idea of the spectral similarity of the classes. Using the aforementioned hyperspectral data sets, we statistically evaluate the sensitivity of our methods to different percentages of training samples after 20 Monte Carlo runs. The processing times are also reported and discussed. All our experiments have been conducted by using Matlab R2013a on a desktop PC equipped with an Intel Core i7 CPU (at 3.6 GHz) and 32 GB of RAM.

3.4.2 Experiments with real hyperspectral data



(a) SP-SNR (b) SP-BDM-LCMV (c) SP-AAP

Figure 3.3: Spectral partitions obtained by our presented spectral partitioning (SP) method from the ROSIS Pavia University data. Our method incorporates two band selection algorithms: SNR (a) and BDM-LCMV (b), respectively. The spectral partitions obtained by SP-AAP are displayed in (c). In all the plots, the x-axis denotes the set of original spectral bands, while the y-axis represents the group of selected bands (each horizontal line displays one group of selected bands).

Before reporting our experimental results, we emphasize that we have optimized the parameter settings in order to obtain the best performance from each individual method involved in the classification framework. The band selection algorithms used in this chapter have a single input parameter that is the number of spectral bands to be selected. We use the same number of bands for all individual spectral partitions in the framework of the presented method for simplicity. In order to decide this number, we need to bear in mind that the number of spectral bands should be enough to provide discriminative spectral details with less band overlapping among the partitions, to preserve the diversity of the views provided. After several empirical experiments, we decided to use 20 bands for the ROSIS Pavia University data set, 60 bands for the AVIRIS Indian Pines data set,

[2] https://engineering.purdue.edu/ landgreb/Hyperspectral.Ex.html

and 40 bands for the HYDICE Washington DC Mall data set. For the two considered classifiers: SVM and MLR, we used the Gaussian radial basis function (RBF) kernel. This has been one of the most successful and widely used kernel functions in the literature [184, 185]. In our experiments, we empirically found that this kernel provides superior performance to other kernels. For the MLR classifier, the logistic regressors (assumed to be random vectors with independent Laplacian components) are learned using the LORSAL algorithm [11, 175].

3.4.2.1 Experiments with ROSIS Pavia University data

In this subsection, our presented spectral partitioning method is firstly tested with the ROSIS Pavia University data. Two well-established band selection algorithms, SNR and BDM-LCMV, are used to construct two spectral partitioning methods called SP-SNR and SP-BDM-LCMV, respectively. Our method is compared with three other strategies based on 1) conducting classification on the original hyperspectral image, 2) conducting classification on the band subsets selected by the same two band selection algorithms and 3) conducting classification after applying the recently developed adaptive affinity propagation based spectral partitioning method (SP-AAP) in [56]. Figs. 3.3(a) and 3.3(b) respectively illustrate the multiple groups of selected bands obtained using the SNR and the BDM-LCMV criteria, while Fig. 3.3(c) shows the bands in the spectral partitions obtained after using SP-AAP. In all the plots, the x-axis denotes the set of original spectral bands while the y-axis represents the selected bands for each class. For example, from up to down along the y-axis, the first horizontal line of the plots shows all the original spectral bands of the ROSIS data and the second horizontal line (SNR-Ori) shows the selected bands by the SNR algorithm with all labeled training samples as input. Then, from the third line on, the figure lists the class-oriented spectral partitions (multiple views) consisting of the selected bands that correspond to the different classes in the scene, e.g. *Meadows*, *Gravel*, *trees*, and so forth. In the case of SP-AAP, five spectral partitions (SPs) are identified in Fig. 3.3(c).

Several observations can be made for the plots displayed in Fig. 3.3. First and foremost, in contrast to the original band selection algorithms, our presented method is able to make use of more of the original spectral bands via different multiple partitions of selected bands. The union of the spectral partitioned bands gives a much larger band set in comparison to applying a band selection algorithm over the original image. Another important observation is that there is great diversity in the selected bands among the spectral partitions that are obtained by using our presented spectral partitioning method. We remark here that possible reason is that most of the classes in the Pavia University scene are dissimilar resulting in different band selections. And this is crucial for generating different views for the classification ensemble process [159, 164]. We also observed that the union of the obtained partitions does not give the full original hyperspectral image. This is expected, as the original hyperspectral image exhibits high redundancy and not all the bands may be useful for classification purposes. Quite opposite, the SP-AAP method generates more equally sampled spectral bands in each partition. In this case, we are still considering the full spectral information in the scene

but voluntarily deciding not to use all of the bands for the construction of the multiple views. Finally, we note that the two considered band selection algorithms obtained different spectral partitions, which will lead to different classification results. The key to our approach is to exploit these partitions synergistically by means of the presented MCS. In the following, we discuss the experimental results acquired after using the obtained spectral partitions for classification purposes. In order to provide a fair comparison with the SP-AAP, we set its number of spectral partitions to five with the ultimate goal of obtaining the same number of spectral bands in each partition. This is mainly because the number of bands of each spectral partition has a strong impact on the final classification performance.

Table 3.1 shows the classification results obtained by the different tested methods for the ROSIS Pavia University data. When the SNR is used for band selection prior to classification, the results remain almost the same, while 1.55% increase in OA is observed when the BDM-LCMV band selection algorithm is used, as compared to using the original spectral information. In comparison with the single band selection strategy, the classification results are improved by the considered multiple classifier-based feature learning strategy based on multiple views provided by different spectral partitions. On the other hand, our spectral partitioning method leads to improvements in classification accuracy. For example, in the case of using BDM-LCMV for spectral partitioning, the OA improved by 5.33% for the SVM classifier and by 5.07% for the MLR classifier, respectively. When the SNR is used for spectral partitioning, the observed increase in OA amounts to 1.64% (SVM) and 1.74% (MLR). This is consistent with the observations in Figs. 3.3(a) and 3.3(b), where we can see that the BDM-LCMV obtained more diversity in the spectral partitions than the SNR criterion. Also, our BDM-LCMV obtained better results than the SP-AAP with 2.09% (SVM) and 0.13% (MLR) increase in OA, respectively.

In order to further evaluate the potential of our presented method, Fig. 3.4 displays the classification maps after the rejection of some unconfident pixels to obtain a 90% accuracy of $\mathcal{A}$ [see (3.9)] of the remaining pixels. It is remarkable that 96% pixels of the classification obtained using the original spectral information have to be discarded to get a more confident result of 90% accuracy, while our presented method (using the SVM classifier) only needs to discard 39% (SP-SNR) and 20% (SP-BDM-LCMV) of the pixels. It is also remarkable that the presented spectral partitioning method shows a great advantage over standard band selection algorithms in the case of the ROSIS Pavia University data. Similar observations can be made for the MLR classifier. For illustrative purposes, Fig. 3.5 plots the nonrejected accuracy ($\mathcal{A}$) as well as the classification quality ($\mathcal{Q}$) as a function of the rejected fractions. From Fig. 3.5, we can also see that our presented method has more biased confidence towards the individual pixels to label, in comparison with the original spectral information. As a matter of fact, the partitions obtained by the proposed method appear to be more informative than those obtained using traditional band selection algorithms. This opens the door to combining several classifiers when they have different preferred pixel subsets in terms of confidence, especially when both subsets share a limited common set of pixels. In this

Table 3.1: Overall, average and individual class accuracies [%] and κ statistic obtained by the presented classification framework implemented using the SVM and MLR with the band selection algorithms: SNR and BDM-LCMV, for the ROSIS Pavia University scene. The SP-AAP spectral partitioning method is also included for comparison. Here, use the accepted pixel subset of the spectral partition approaches while using the rejected pixel subset of the original classifier. The results obtained using the original spectral information and the spectral bands selected by using the SNR and BDM-LCMV are also included. In all cases, 3921 training samples have been used.

Class	**SVM**	**Band Selection (BS)**			**Spectral Partitioning (SP)**			**BS + Original**		**SP + Original**		
	Train/test	All	SNR	BDM-LCMV	SNR	BDM-LCMV	AAP	SNR	BDM	SNR	BDM-LCMV	AAP
Alfalfa	548/6304	84.63	83.98	83.57	77.92	84.76	83.55	84.58	84.66	84.20	84.33	84.41
Meadows	540/18146	66.04	66.04	70.69	74.80	78.61	76.32	66.85	71.45	77.23	78.89	77.36
Gravel	392/1815	73.39	68.54	67.99	56.36	69.31	62.75	71.40	70.41	72.89	72.07	67.71
Trees	524/2912	97.49	98.18	98.11	97.32	97.97	94.06	98.45	98.15	97.80	98.18	96.67
Metal sheets	265/1113	99.46	99.37	99.28	99.28	99.46	99.37	99.46	99.46	99.46	99.46	99.46
Bare soil	53/4572	94.27	92.94	92.17	90.44	91.69	90.66	94.07	93.83	93.83	92.67	92.72
Bitumen	375/981	91.13	86.54	89.91	85.32	89.50	87.36	89.09	91.34	91.34	91.03	89.91
Bricks	514/3364	92.12	91.91	93.25	86.92	93.40	91.83	92.69	93.16	92.21	92.78	92.72
Shadows	231/795	99.37	97.74	98.74	95.85	98.87	99.75	99.25	99.12	99.25	99.50	99.62
Overall accuracy		79.22	78.63	80.77	80.26	84.55	82.46	79.53	81.63	84.19	84.86	83.85
Average accuracy		88.66	87.25	88.19	84.91	89.29	87.29	88.43	89.06	89.80	89.88	88.95
κ statistic		73.82	73.12	75.57	74.71	80.05	77.40	74.21	76.63	79.65	80.45	79.18

Class	**MLR**	**Band Selection (BS)**			**Spectral Partitioning (SP)**			**BS + Original**		**SP + Original**		
	Train/test	All	SNR	BDM-LCMV	SNR	BDM-LCMV	AAP	SNR	BDM	SNR	BDM-LCMV	AAP
Alfalfa	548/6304	77.27	71.39	79.52	75.97	80.50	82.73	80.08	80.93	78.39	81.99	82.73
Meadows	540/18146	75.79	74.14	80.31	83.03	84.45	82.55	81.97	82.88	83.33	84.85	82.55
Gravel	392/1815	65.84	64.81	78.27	75.25	75.68	81.60	78.83	78.95	78.40	81.17	81.60
Trees	524/2912	93.68	86.68	92.31	92.60	91.65	94.70	91.06	92.93	91.94	91.98	94.70
Metal sheets	265/1113	99.46	99.35	99.46	98.80	99.78	99.13	99.35	99.35	98.80	99.78	99.13
Bare soil	53/4572	84.38	78.11	79.46	80.30	87.11	88.83	83.68	84.28	81.38	87.36	88.83
Bitumen	375/981	85.63	85.88	91.22	92.11	93.77	93.26	89.06	91.98	91.98	93.64	93.26
Bricks	514/3364	86.21	72.48	73.65	69.42	83.43	80.12	71.92	75.54	70.97	75.70	80.12
Shadows	231/795	97.48	100	100	100	99.83	100	100	100	100	99.83	100
Overall accuracy		80.06	76.01	81.50	81.80	85.13	85.00	82.56	83.69	82.67	85.20	85.00
Average accuracy		85.08	81.43	86.02	85.28	88.47	89.21	86.22	87.43	86.13	88.48	89.21
κ statistic		74.42	68.48	75.55	75.70	79.97	79.86	76.64	78.18	76.78	80.04	79.68

52

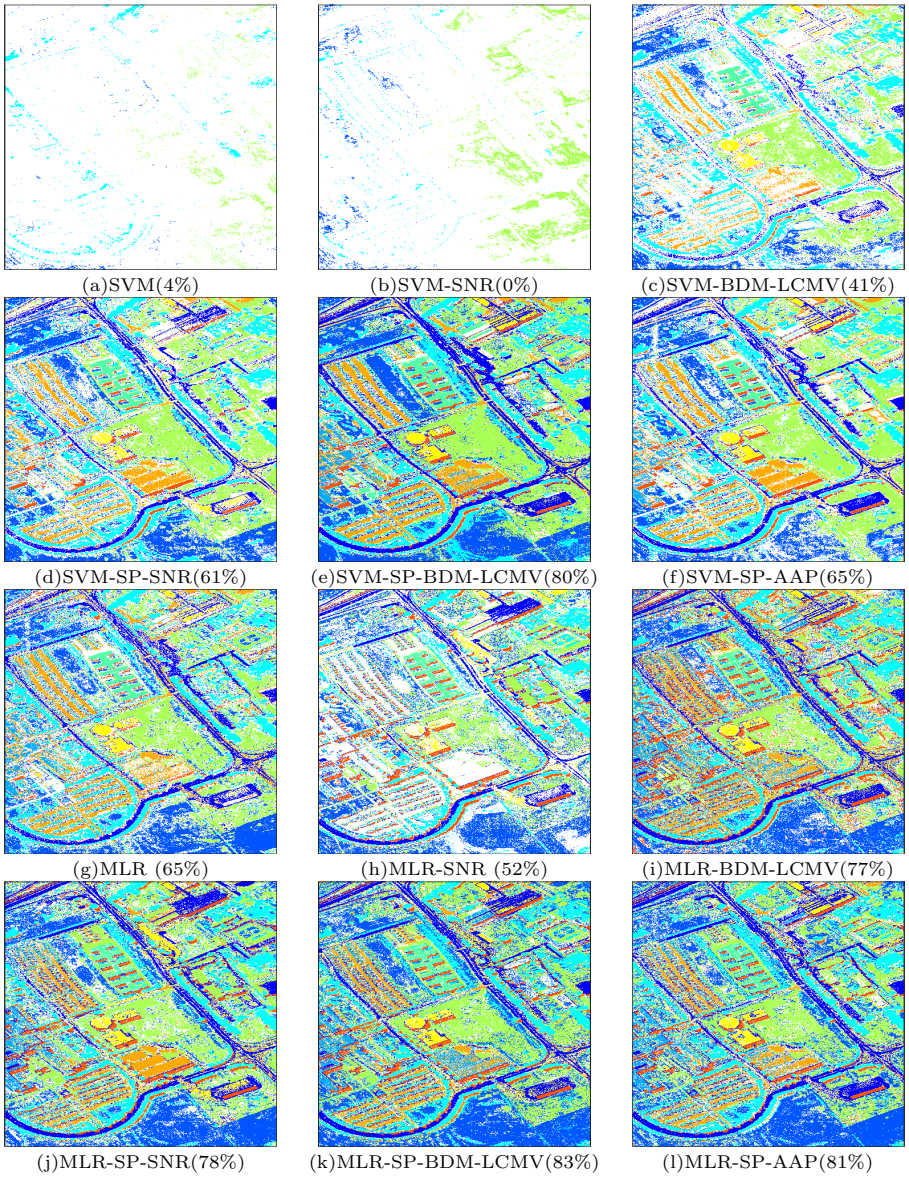
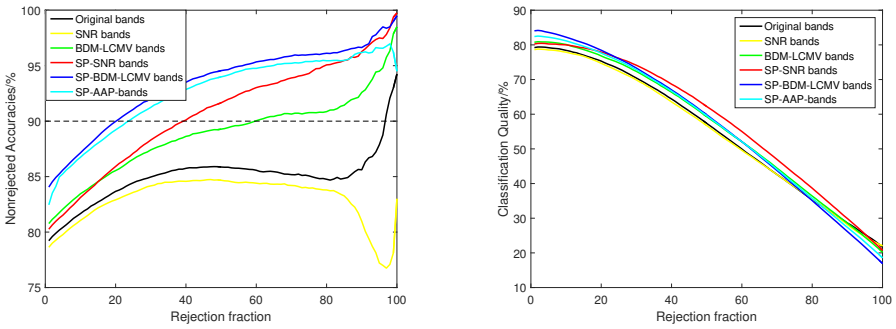
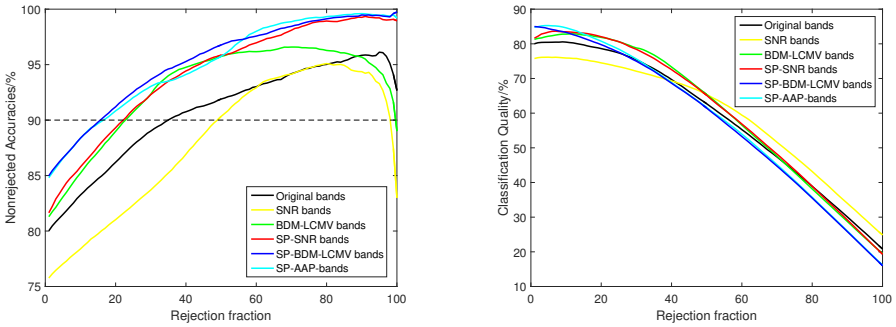


Figure 3.4: Classification maps (with rejection) obtained by the proposed classification framework, using the original spectral information (a,g), with the 20 selected bands by using SNR (b,h) and BDM-LCMV (c,i) algorithms, and with the spectral partitions obtained by our proposed spectral partitioning (SP) approach implemented with SNR (d,h) and BDM-LCMV (e,k and the SP-AAP method (f,l) for the ROSIS Pavia University data. The percentage in the parenthesis denotes the proportion of pixels remaining after rejection. Maps (a-f) are obtained with the SVM classifier, while (g-l) are obtained with the MLR classifier. In all cases, a total of 3921 randomly selected training samples were used for training and the rest used for testing. The maps are displayed with partial pixels rejected in order to obtain a 90% classification accuracy for the remaining ones. Note that the overall accuracies for the unrejected pixels of each map are calculated by considering the labeled test samples, with the training samples excluded.

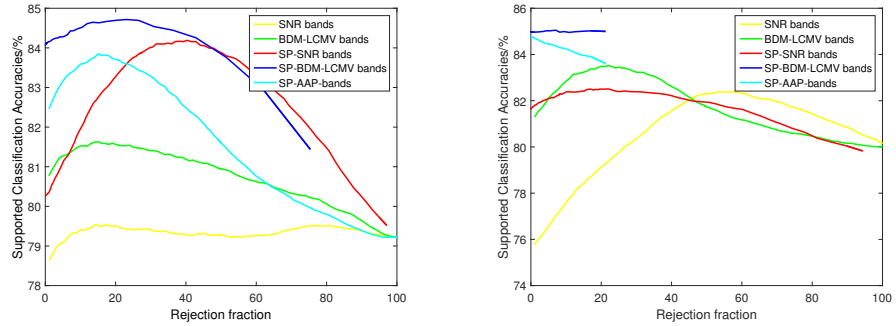


(a) Nonrejected accuracies $\mathcal{A}$ (SVM) (b) Classification qualities $\mathcal{Q}$ (SVM)



(c) Nonrejected accuracies $\mathcal{A}$ (MLR) (d) Classification qualities $\mathcal{Q}$ (MLR)

Figure 3.5: Nonrejected accuracies ($\mathcal{A}$) and classification qualities ($\mathcal{Q}$) as a function of rejected fractions with the ROSIS Pavia University data set. These plots correspond to the results in Table 3.1.



(a) SVM (b) MLR

Figure 3.6: Classification OAs (as a function of rejections) obtained by our proposed classification framework with the ROSIS Pavia University data, after being supported by the rejected pixels of the classifier using the original spectral information.

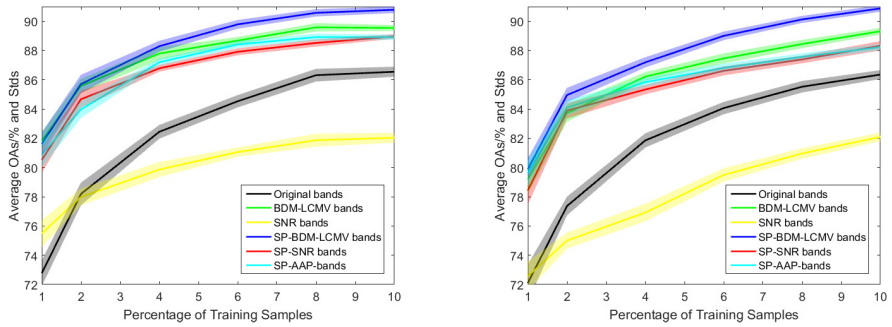
case, the SP-AAP obtained slightly lower but competitive performance when compared to our proposed methods, SP-SNR and SP-BDM-LCMV.

With the aforementioned observations in mind, we try to further improve the classification results of our presented method by using Eq. (3.13). Since more confidence exists in the results of our presented method, we use the accepted pixel subset of SP-

Table 3.2: Processing times of different methods for ROSIS Pavia University scene.

Time/sec	SVM (different percentages of training samples [%])					
	1	2	4	6	8	10
Original	1.66 ± 0.16	2.66 ± 0.32	4.65 ± 0.34	6.60 ± 0.35	8.62 ± 0.58	10.58 ± 0.6
SNR	1.48 ± 0.13	2.34 ± 0.30	4.15 ± 0.32	5.89 ± 0.34	7.61 ± 0.56	9.58 ± 0.59
BDM-LCMV	1.36 ± 0.15	2.25 ± 0.31	3.88 ± 0.32	5.28 ± 0.34	7.11 ± 0.55	9.07 ± 0.58
SP-SNR	9.99 ± 0.45	15.27 ± 0.64	27.25 ± 0.92	40.99 ± 1.48	55.80 ± 1.69	71.00 ± 2.63
SP-BDM-LCMV	8.49 ± 0.54	12.81 ± 0.72	23.75 ± 0.84	36.64 ± 1.46	51.21 ± 2.02	66.17 ± 2.46
SP-AAP	5.43 ± 0.46	8.54 ± 0.75	16.32 ± 0.80	23.01 ± 1.44	31.81 ± 1.79	39.98 ± 2.66
Time/sec	MLR (different percentages of training samples [%])					
	1	2	4	6	8	10
Original	1.20 ± 0.13	3.48 ± 0.37	10.41 ± 1.34	21.66 ± 2.07	35.60 ± 2.69	50.13 ± 3.15
SNR	1.08 ± 0.13	3.15 ± 0.38	9.38 ± 1.33	19.63 ± 2.04	32.03 ± 2.35	45.13 ± 2.95
BDM-LCMV	1.00 ± 0.12	3.07 ± 0.39	9.28 ± 1.34	19.49 ± 2.05	32.11 ± 2.46	44.57 ± 3.13
SP-SNR	10.57 ± 0.30	30.25 ± 1.08	91.63 ± 3.53	192.27 ± 5.18	311.27 ± 6.66	466.48 ± 6.90
SP-BDM-LCMV	10.23 ± 0.24	29.74 ± 0.94	91.27 ± 3.24	190.92 ± 3.67	310.65 ± 3.89	462.65 ± 4.15
SP-AAP	5.49 ± 0.28	16.43 ± 1.00	50.43 ± 3.29	104.93 ± 4.09	172.10 ± 4.73	257.27 ± 5.33

SNR (or SP-BDM-LCMV) while using the rejected pixel subset of the classifier applied to the original hyperspectral image. In this case, we are effectively combining the results obtained using the original spectral information and the spectral partitions derived by our proposed approach. The results are shown in Table 3.1. In other words, the results in this table have been optimized to find the best combination of classifiers bearing in mind the rejection fractions. The effect of this combination on the OAs% is shown in Fig. 3.6, as a function of the rejection fractions. Remarkably, the classification result of SP-SNR is improved by about 4% (from 80.26% to 84.19%) when using the SVM and by 6.66% (from 76.01% to 82.56%) when using the MLR.



(a) SVM (b) MLR

Figure 3.7: Overall classification accuracies (as a function of the number of training samples) obtained by the proposed classification framework (with the original spectral information and with SP-SNR, SP-BDM-LCMV and SP-AAP) for the ROSIS Pavia University scene. The solid lines represent the average of 20 Monte Carlo runs, whereas the colored area around the lines represent the standard deviation around the mean. Plots (a) is obtained using the SVM classifier and (b) is obtained using the MLR classifier.

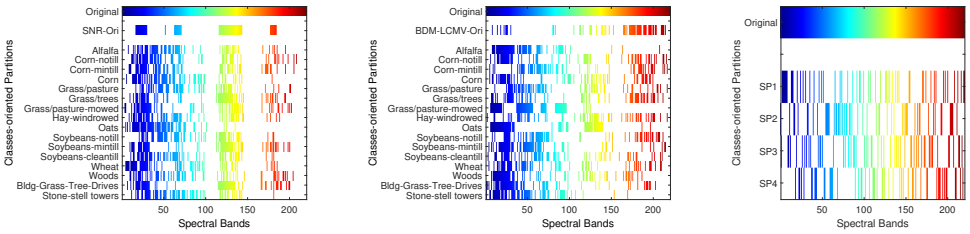
In order to evaluate the sensitivity of our methods to different numbers of training samples, Fig. 3.7 plots the OA achieved by different methods as a function of the percentage of training samples used to train the classifier. From Fig. 3.7, it is remarkable that the proposed framework (implemented with SP-BDM-LCMV) provides better results when compared with the other methods. When the number of training

samples increases, the performance also increases, implying robustness with respect to different training sample sets. Also, the SP-SNR obtained results which are close to those obtained by the SP-AAP method, even though the single SNR band selection method performed comparatively worst among all methods. On the other hand, we also calculated the processing times in the considered computing environment (see Table 3.2). As shown by Table 3.2, our presented spectral partitioning methods are time-consuming as their computational time is approximately 7 to 10 times that of using the original spectral bands of the hyperspectral image. However, the presented spectral partitioning framework performs very fast in the task of obtaining the spectral partitions, as it only considers the labeled samples as input. Also, the spectral partitions are independently learned by the classifiers. As a result, it is very feasible to perform the classification step in parallel by resorting to multi-core architectures of exploiting co-processors such as graphics computing units (GPUs). With these implementation enhancements, the complexity of the method can be kept within similar levels as the classification of the original image.

3.4.2.2 Experiments with AVIRIS Indian Pines data

In this subsection, we report the experimental results obtained by our spectral partitioning method with the AVIRIS Indian Pines data. The same band selection algorithms: SNR and BDM-LCMV are also used in this case to generate spectral partitions, which are respectively referred to as SP-SNR and SP-BDM-LCMV. For simplicity, we conduct the experiments using the same conditions for all groups. Considering the fact that most classes of the AVIRIS Indian Pines scene are given by vegetation features with spectrally similar signatures, a larger number of spectral bands is needed in order to retain enough spectral details to distinguish different classes. On the other hand, increasing the number of bands may be detrimental to the diversity for the MCS, since a larger number of bands in the partitions means higher partition overlaps [172]. Consequentially, and for simplicity purposes, we empirically set the number of selected bands to 60 for our spectral partitioning methods as well as for the SP-AAP method. First of all, we perform the experiments by using 640 randomly selected labeled samples for both the SVM [182] and MLR-LORSAL classifiers [11, 175]. Each group of results is obtained after 20 Monte Carlo runs. The average values and standard variations are reported. Then, we also reported the performance of our presented methods with regards to different percentages of labeled training samples, along with the corresponding processing times.

Fig. 3.8 displays the spectral bands selected by the SNR and BDM-LCMV algorithms, as well as the spectral partitions obtained by using our SP-SNR and SP-BDM-LCMV methods and the SP-AAP method. In the case of SP-SNR, it can be seen that each spectral partition consists of more diverse bands as compared to using SNR with all the training samples. It is also interesting to note that noisy and water absorption bands have low probability to be selected by SP-SNR and SP-BDM-LCMV, even though we used all of the original bands as input to the algorithms in order to allow them to select the most informative bands automatically. The SP-AAP spectral partitioning method is



(a) SP-SNR (b) SP-BDM-LCMV (c) SP-AAP

Figure 3.8: Spectral partitions obtained by our presented spectral partitioning (SP) method from the AVIRIS Indian Pines data. Our method incorporates two band selection algorithms: SNR (a) and BDM-LCMV (b). The spectral partitions obtained by SP-AAP are displayed in (c). In all plots, the x-axis denotes the set of original spectral bands, while the y-axis represents the group of selected bands (each horizontal line displays one group of selected bands).

Table 3.3: Overall, average and individual class accuracies [%] and κ statistic obtained by the presented classification framework implemented using the SVM with the band selection algorithms: SNR and BDM-LCMV, for the AVIRIS Indian Pines scene. The results obtained using the original spectral information and the spectral bands selected by using the SNR and BDM-LCMV are also included. SP-AAP is also included for comparison. In all cases, only 640 randomly selected training samples have been used.

Class	Band Selection (SVM)			Spectral Partitioning (SP) (SVM)		
	All	SNR	BDM-LCMV	SNR	BDM-LCMV	SP-AAP
C1	94.56 ± 5.15	96.4 ± 5.46	95.00 ± 6.99	96.76 ± 5.44	97.12 ± 5.41	93.93 ± 6.25
C2	75.66 ± 3.43	70.85 ± 6.29	71.07 ± 4.49	76.62 ± 4.27	76.82 ± 3.88	75.44 ± 3.74
C3	73.05 ± 4.94	70.96 ± 5.70	68.46 ± 5.32	75.51 ± 5.35	75.31 ± 5.36	72.90 ± 4.54
C4	90.78 ± 3.03	88.66 ± 3.56	88.31 ± 3.70	90.86 ± 2.95	91.27 ± 3.04	91.00 ± 3.94
C5	92.44 ± 2.33	90.36 ± 3.01	89.45 ± 3.19	92.12 ± 2.93	91.62 ± 3.18	91.28 ± 3.26
C6	93.73 ± 2.25	91.84 ± 2.94	90.88 ± 2.10	93.67 ± 2.30	93.77 ± 2.14	94.21 ± 1.88
C7	92.31 ± 6.11	91.54 ± 7.85	91.54 ± 6.06	91.92 ± 6.82	91.54 ± 7.01	90.32 ± 7.89
C8	96.52 ± 1.13	96.00 ± 1.49	95.84 ± 1.20	97.21 ± 1.30	97.27 ± 1.08	96.58 ± 1.98
C9	92.50 ± 14.1	83.50 ± 19.81	84.00 ± 14.65	92.50 ± 13.33	93.50 ± 12.26	90.94 ± 16.83
C10	79.84 ± 5.21	78.04 ± 4.84	74.55 ± 5.68	83.06 ± 5.42	82.96 ± 4.63	81.85 ± 5.04
C11	72.15 ± 3.13	71.24 ± 3.57	67.26 ± 3.86	75.55 ± 3.82	75.43 ± 3.58	72.19 ± 4.89
C12	87.74 ± 4.37	87.33 ± 4.10	85.69 ± 4.73	90.08 ± 3.81	89.28 ± 3.71	87.03 ± 3.61
C13	99.36 ± 0.50	98.66 ± 1.08	99.01 ± 0.79	99.21 ± 0.64	99.24 ± 0.54	99.36 ± 0.37
C14	90.37 ± 2.71	91.82 ± 3.41	90.24 ± 4.32	92.62 ± 2.42	92.77 ± 1.94	91.80 ± 2.87
C15	75.73 ± 6.12	68.31 ± 5.48	67.00 ± 5.39	73.40 ± 6.80	73.17 ± 5.76	72.22 ± 5.75
C16	94.34 ± 3.08	92.97 ± 2.79	93.98 ± 2.58	93.61 ± 2.41	94.53 ± 2.64	94.97 ± 3.19
OA	81.50 ± 1.20	79.83 ± 1.44	77.88 ± 1.29	83.34 ± 1.20	83.28 ± 1.23	83.05 ± 1.06
AA	87.57 ± 1.35	85.53 ± 1.85	84.52 ± 1.37	88.42 ± 1.53	88.48 ± 1.42	88.73 ± 1.49
κ	78.98 ± 1.33	77.05 ± 1.61	74.87 ± 1.44	81.04 ± 1.34	80.98 ± 1.37	80.52 ± 1.16

not able to detect and screen the noisy and water absorption bands in this case. Finally, we can also see from Fig. 3.8 that the spectral partitions obtained exhibit diverse band selections but with similar patterns across the classes. It is reasonable that the band selections share a degree of similarity considering the fact that the classes in the scene are mostly related to vegetation features [see also Fig. 3.2(f)].

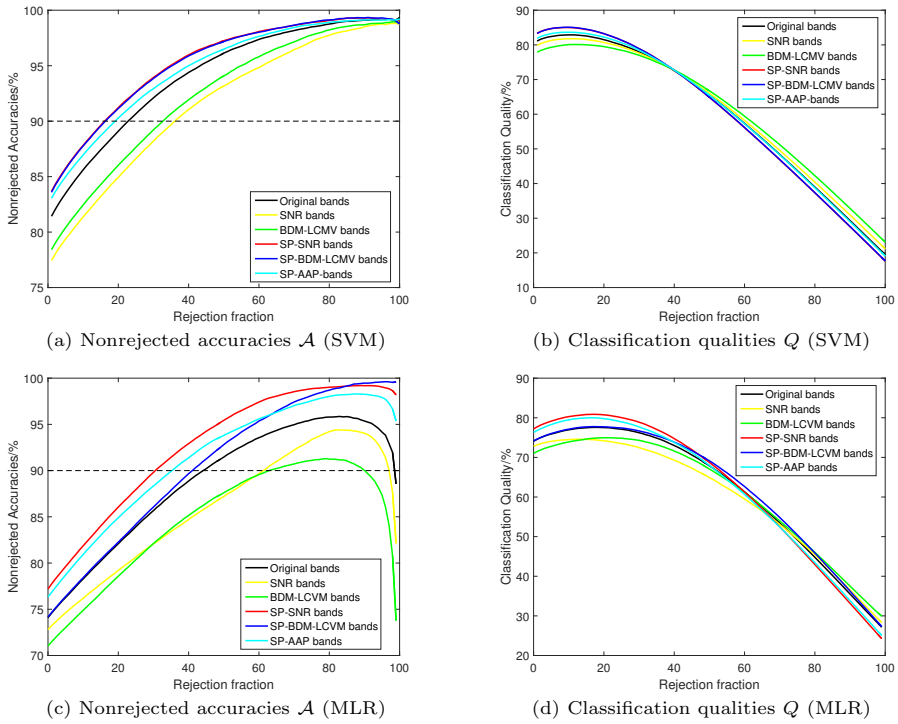
Tables 3.3 and 3.4 shows the OA, AA, κ and individual classification accuracies

Table 3.4: Overall, average and individual class accuracies [%] and κ statistic obtained by the presented classification framework implemented using the MLR with the band selection algorithms: SNR and BDM-LCMV, for the AVIRIS Indian Pines scene. The results obtained using the original spectral information and the spectral bands selected by using the SNR and BDM-LCMV are also included. SP-AAP is also included for comparison. In all cases, only 640 randomly selected training samples have been used.

Class	Band Selection (MLR)			Spectral Partitioning (SP) (MLR)		
	All	SNR	BDM-LCMV	SNR	BDM-LCMV	SP-AAP
C1	86.43 ± 6.91	86.79 ± 8.44	83.57 ± 8.39	89.21 ± 7.43	90.00 ± 8.16	82.50 ± 9.41
C2	70.86 ± 3.81	68.11 ± 3.74	67.54 ± 3.65	71.87 ± 3.44	71.50 ± 3.27	69.01 ± 3.56
C3	58.05 ± 3.84	57.40 ± 4.68	59.01 ± 4.21	62.95 ± 3.72	58.26 ± 3.86	61.39 ± 4.44
C4	78.60 ± 5.38	78.42 ± 3.46	73.20 ± 6.59	78.71 ± 6.39	77.85 ± 4.50	73.97 ± 6.16
C5	83.77 ± 3.03	82.48 ± 3.49	83.16 ± 2.96	85.42 ± 3.07	87.53 ± 3.48	84.72 ± 3.31
C6	93.53 ± 1.82	92.67 ± 1.99	88.19 ± 2.50	93.16 ± 2.21	92.39 ± 1.96	89.05 ± 2.57
C7	82.24 ± 8.67	88.46 ± 7.69	75.71 ± 10.91	83.24 ± 9.02	84.55 ± 7.90	83.40 ± 9.09
C8	96.56 ± 1.19	93.90 ± 1.95	91.78 ± 2.97	94.95 ± 2.55	92.96 ± 3.19	93.16 ± 3.08
C9	82.62 ± 16.34	84.29 ± 15.70	74.85 ± 12.25	82.79 ± 13.03	90.40 ± 8.25	69.75 ± 14.00
C10	69.26 ± 5.81	71.07 ± 4.41	68.31 ± 4.46	76.25 ± 4.58	70.26 ± 4.55	77.06 ± 3.76
C11	58.35 ± 5.19	56.52 ± 5.08	56.61 ± 4.05	66.67 ± 4.48	59.67 ± 4.51	67.64 ± 4.63
C12	78.48 ± 5.38	76.94 ± 4.04	66.24 ± 5.39	79.07 ± 3.85	74.91 ± 5.35	76.83 ± 3.94
C13	99.56 ± 0.26	99.53 ± 0.31	98.89 ± 0.50	99.33 ± 0.44	99.12 ± 0.75	99.06 ± 0.58
C14	90.49 ± 2.50	87.66 ± 4.40	85.16 ± 4.03	90.83 ± 3.67	89.94 ± 2.63	90.14 ± 4.15
C15	71.74 ± 4.26	67.66 ± 3.44	63.99 ± 3.48	69.75 ± 4.60	68.42 ± 3.45	64.09 ± 5.85
C16	92.22 ± 4.18	93.23 ± 3.60	93.13 ± 3.56	94.70 ± 3.39	87.92 ± 3.70	92.41 ± 5.34
OA	73.80 ± 1.24	72.25 ± 1.12	70.45 ± 1.12	77.37 ± 1.20	73.83 ± 1.15	76.31 ± 1.34
AA	80.80 ± 1.77	80.32 ± 1.41	76.83 ± 1.75	81.47 ± 1.43	80.98 ± 1.22	79.20 ± 1.24
κ	70.33 ± 1.34	68.57 ± 1.21	66.62 ± 1.22	73.17 ± 1.33	70.30 ± 1.26	72.29 ± 1.48

(obtained by the SVM and MLR, respectively) after using 640 randomly selected pixels (about 40 pixels per class on average) for training and the rest of the labeled samples for testing. If the number of samples available in the ground-truth image is less than 20, we take half of the total samples in that class for training and the other half for testing. Several observations can be drawn from the results reported on Tables 3.3 and 3.4. First of all, the presented approach leads to an increase of about 1% to 4% in classification accuracy (regardless of the band selection method used) in comparison with using all the spectral bands in the original image. Although the advantage seems moderate, the individual accuracies of 13 out of 16 classes (with the SVM) and 12 out of 16 classes (with the MLR) in the scene are improved by our presented spectral partitioning method. Secondly, when compared with standard band selection (SNR and BDM-LCMV) on the original image, our presented method can further improve the classification accuracy by about 5% by providing multiple views from the original hyperspectral image into the MCS. This suggests that our presented spectral partitioning method is able to enhance the overall statistics by improving most of the individual classes via multiple views of the original input features that are provided by different spectral partitions.

In order to evaluate the classification results in Tables 3.3 and 3.4 from the viewpoint of individual pixels, Fig. 3.9 displays the corresponding classification statistics: nonrejected accuracy ($\mathcal{A}$) and classification quality ($\mathcal{Q}$) as a function of the rejected fractions. It can be seen from Figs. 3.9(a) and 3.9(c) that all discussed methods exhibit



(a) Nonrejected accuracies $\mathcal{A}$ (SVM)

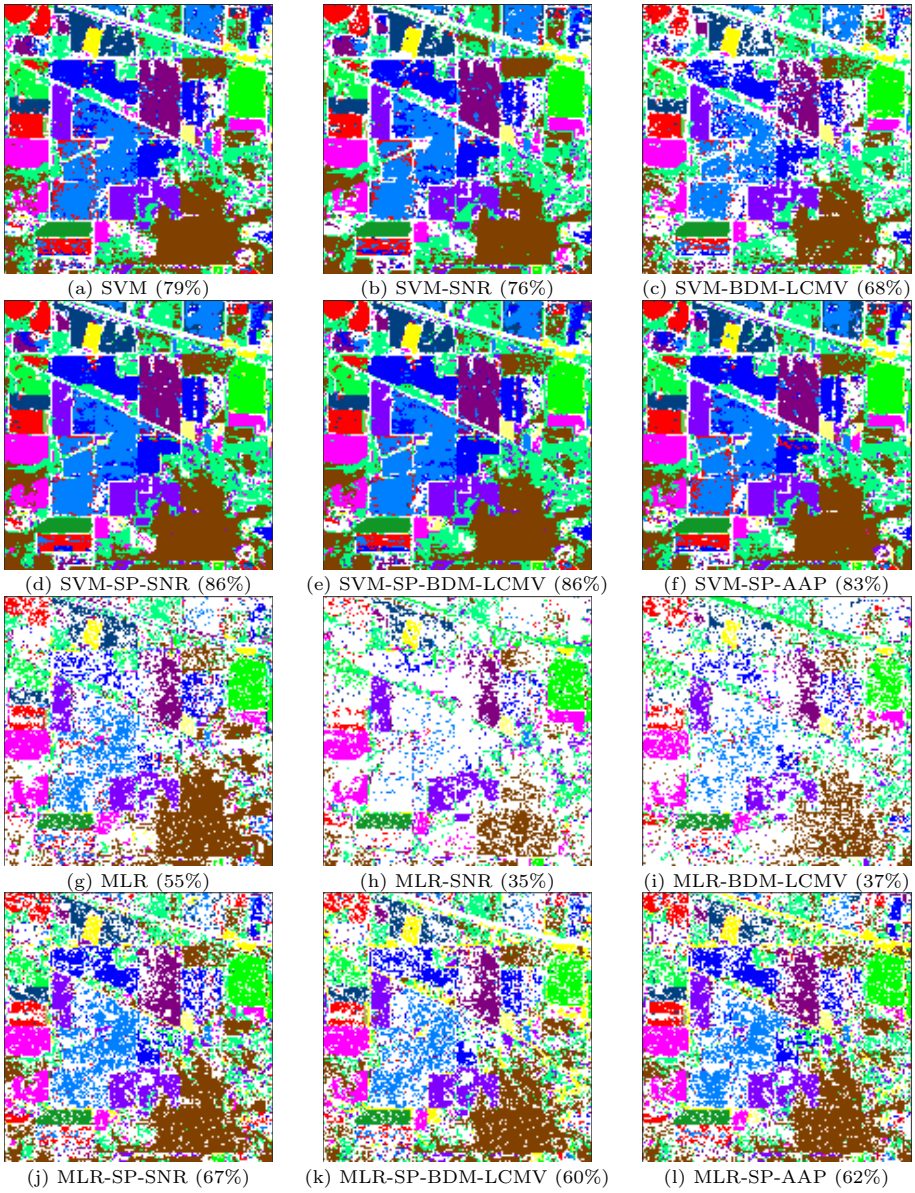
(b) Classification qualities Q (SVM)

(c) Nonrejected accuracies $\mathcal{A}$ (MLR)

(d) Classification qualities Q (MLR)

Figure 3.9: Statistics of nonrejected accuracies $\mathcal{A}$ and classification qualities $\mathcal{Q}$ of the classifications as a function of rejected fractions with the AVIRIS Indian Pines data. These plots correspond to the results in Tables 3.3, 3.4.

a similar trend, with the rejection fraction increasing. A similar pattern can be observed in Figs. 3.9(b) and 3.9(d), which implies that similar confidence distributions towards the pixels are provided by all the methods considered in this experiment. This observation is illustrated by the classification maps with 90% of $\mathcal{A}$ after rejection in Fig. 3.10. Even though only about 2% to 4% increase in OA is observed after applying our proposed spectral partitioning framework, about 7% and 30% increase of nonrejected accuracy $\mathcal{A}$ is observed when compared with the classifier applied to the original hyperspectral image, and also after using the considered band selection algorithms. In comparison with the SP-AAP, our methods also provide higher classification accuracy and rejection confidence when using different rejection fractions. On the other hand, if we compare the five classification maps with 90% of nonrejected accuracy $\mathcal{A}$ in Fig. 3.10, we can observe that all the methods involved in this case hold similar confidence towards the pixels in the scene. This is consistent with our discussion of the results in Fig. 3.9. Despite the presence of similar confidence distribution in the pixels, we plot the combined classification result against the classifications obtained using SVM with all the original spectral bands and our presented spectral partitioning method to validate the discussions above. From Fig. 3.11 we can see that, although the proposed spectral partitioning strategy obtains better results than standard band selection algorithms and SP-AAP spectral partitioning method, other extra advantages can be achieved in the views of $\mathcal{A}$



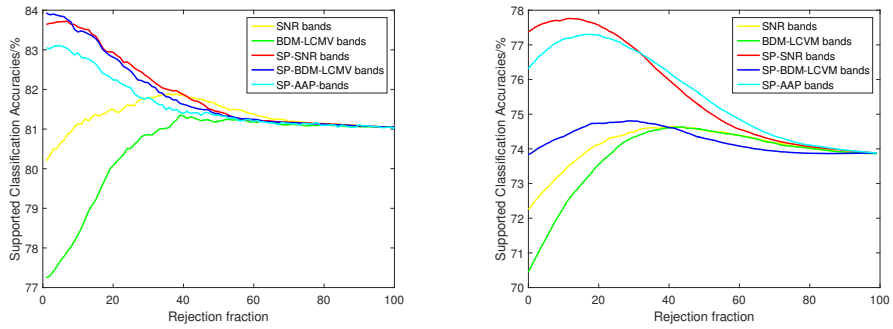
(a) SVM (79%) (b) SVM-SNR (76%) (c) SVM-BDM-LCMV (68%)

(d) SVM-SP-SNR (86%) (e) SVM-SP-BDM-LCMV (86%) (f) SVM-SP-AAP (83%)

(g) MLR (55%) (h) MLR-SNR (35%) (i) MLR-BDM-LCMV (37%)

(j) MLR-SP-SNR (67%) (k) MLR-SP-BDM-LCMV (60%) (l) MLR-SP-AAP (62%)

Figure 3.10: Classification maps (with rejection) obtained by the proposed classification framework, using the original spectral information (a,g), with the 60 selected bands by using SNR (b,h) and BDM-LCMV (c,i) algorithms, and with the spectral partitions obtained by our proposed spectral partitioning (SP) approach implemented with SNR (d,h) and BDM-LCMV (e,k) and the SP-AAP method (f,l) for the AVIRIS Indian Pines data. The percentage in the parenthesis denotes the proportion of pixels remaining after rejection. In all cases, a total of 640 randomly selected training samples were used for training and the rest used for testing. The maps are displayed with partial pixels rejected in order to obtain a 90% classification accuracy for the remaining ones. Note that the overall accuracies for the unrejected pixels of each map are calculated by considering the labeled test samples, with the training samples excluded.

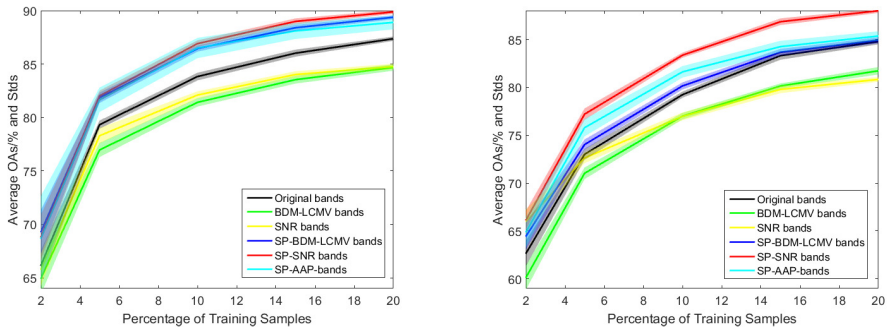


(a) SVM (b) MLR

Figure 3.11: Classification OAs (as a function of rejections) obtained by our proposed classification framework with the AVIRIS Indian Pines data, after being supported by the rejected pixels of the classifier using the original spectral information. Note that plot (a) is obtained from one Monte Carlo run in Table 3.3, while plot (b) is obtained from one Monte Carlo run in Table 3.4. In both cases, the selected run is close to the statistical average.

and $\mathcal{Q}$ with rejections. The improvement is also observed in the case of the ROSIS Pavia University data, after supported by the classifier using all the original spectral bands as a whole. This is consistent with the aforementioned observations and discussions, in the sense that the results of our presented spectral partitioning method (embedded with band selection algorithms) are further improved after being supported by the classifier using all the original spectral bands, a phenomenon that is observed for both the ROSIS Pavia University scene and the AVIRIS Indian Pines scene. Also, the similarity of the confidence distributions of the two classifications obtained for each pixel suggests that similar confidences towards individual pixels can be obtained by the different methods tested in the case of the AVIRIS data scene. This is reasonable, considering the fact that most of the classes in this scene comprise vegetation features that share very similar spectral signatures, leading to low classification diversity that spectral partitioning is able to generate by providing multiple views of the original hyperspectral data.

In order to evaluate the sensitivity of our methods to different numbers of training samples, we also plot the overall accuracy statistics (mean value and standard deviation) of different methods as a function of the percentage of training samples as input (see Fig. 3.12). From Fig. 3.12, it is remarkable that our presented methods outperform those methods that only use the spectral bands of the original image or the bands selected by single-band selection algorithms. This is the case for different percentages of training samples. Finally, it can also be seen that for both classifiers (SVM and MLR), our class-oriented spectral partitioning methods generally obtained higher OAs when compared with the SP-AAP method. Meanwhile, we also report the processing times of the experiments reported in Fig. 3.12 on Table 3.5. From Table 3.5, we can observe that our presented spectral partitioning methods required 5 to 12 times more computation than the original group and 3 times more than the SP-AAP. As explained before, these times can be significantly reduced by resorting to simple parallel implementation frameworks.



(a) SVM (b) MLR

Figure 3.12: Overall classification accuracies (as a function of the number of training samples) obtained by the proposed classification framework (with the original spectral information and with SP-SNR, SP-BDM-LCMV and SP-AAP) for the AVIRIS Indian Pines scene. The solid lines represent the average of 20 Monte Carlo runs, whereas the colored area around the lines represent the standard deviation around the mean. Plots (a) is obtained using the SVM classifier and (b) is obtained using the MLR classifier.

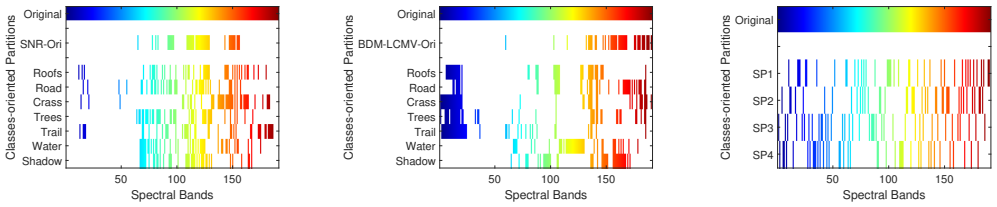
Table 3.5: Processing times of different methods for the AVIRIS Indian Pines scene.

Time/sec	SVM (different percentage of training samples [%])				
	2	5	10	15	20
Original	1.87 ± 0.05	7.39 ± 0.24	20.96 ± 0.54	45.01 ± 1.12	77.80 ± 1.00
SNR	1.68 ± 0.05	6.37 ± 0.24	18.82 ± 0.57	40.88 ± 1.12	70.05 ± 1.00
BDM-LCMV	1.59 ± 0.05	6.29 ± 0.24	18.59 ± 0.54	40.08 ± 1.21	69.72 ± 0.99
SP-SNR	9.09 ± 0.36	37.69 ± 0.95	108.89 ± 2.04	238.03 ± 4.30	409.95 ± 3.20
SP-BDM-LCMV	8.67 ± 0.05	36.77 ± 0.80	106.92 ± 2.19	237.15 ± 2.98	417.06 ± 2.21
SP-AAP	2.91 ± 0.10	12.83 ± 0.27	34.82 ± 1.50	78.24 ± 2.11	128.57 ± 2.33
Time/sec	MLR (different percentage of training samples [%])				
	2	5	10	15	20
Original	0.81 ± 0.11	1.42 ± 0.09	2.36 ± 0.16	3.44 ± 0.27	4.48 ± 0.29
SNR	0.70 ± 0.07	1.22 ± 0.10	2.10 ± 0.16	2.97 ± 0.28	3.89 ± 0.27
BDM-LCMV	0.73 ± 0.08	1.23 ± 0.10	2.15 ± 0.12	3.06 ± 0.28	4.04 ± 0.28
SP-SNR	9.14 ± 0.24	15.19 ± 0.38	24.79 ± 0.58	36.28 ± 0.99	47.59 ± 1.30
SP-BDM-LCMV	9.21 ± 0.17	15.57 ± 0.36	25.98 ± 0.64	38.88 ± 1.12	51.44 ± 1.47
SP-AAP	3.20 ± 0.17	3.89 ± 0.24	6.99 ± 0.54	12.06 ± 0.79	17.33 ± 1.40

3.4.2.3 Experiments with the HYDICE Washington DC Mall data set

In this subsection, we tested our spectral partitioning method with the HYDICE Washington DC Mall data set. First of all, we used 2% randomly selected training samples from the labeled reference data for training (and the rest for testing). For simplicity, the experimental results are obtained by using the same conditions for all groups. Similar to the experiments with the previous two hyperspectral data sets, we set empirically the number of bands to 40 for single band selection approaches, for the partitions generated by our spectral partitioning methods, and for the SP-AAP.

Figs. 3.13(a) and 3.13(b) respectively plot the multiple spectral partitions obtained by using the SP-SNR and SP-BDM-LCMV methods. The spectral partitioning results obtained by SP-AAP are shown in Fig.3.13(c). In all the plots, the x-axis represents



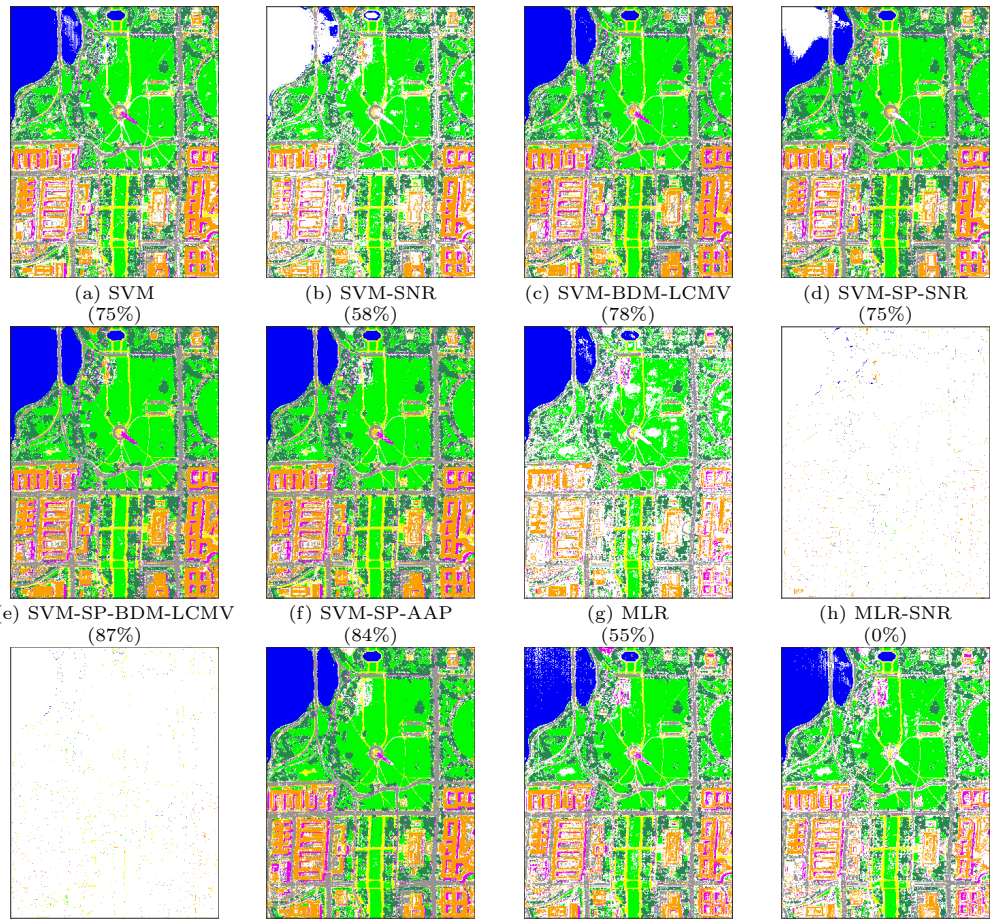
(a) SP-SNR (b) SP-BDM-LCMV (c) SP-AAP

Figure 3.13: Spectral partitions obtained by our presented spectral partitioning (SP) method from the HYDICE DC Mall data set. Our method incorporates two band selection algorithms: SNR (a) and BDM-LCMV (b). The spectral partitions obtained by SP-AAP are displayed in (c). In all plots, the x-axis denotes the set of original spectral bands, while the y-axis represents the group of selected bands (each horizontal line displays one group of selected bands).

Table 3.6: Overall, average and individual class accuracies (OA, AA, CAs)[%] and κ statistic obtained by the presented classification framework implemented using the SVM with the band selection algorithms: SNR and BDM-LCMV, for the HYDICE Washington DC Mall scene. The results obtained using the original spectral information and the spectral bands selected by using the SNR and BDM-LCMV are also included. SP-AAP is also included for comparison. In all cases, only 2% randomly selected training samples from the labeled reference data have been used.

Class	Band Selection (SVM)			Spectral Partitioning (SP) (SVM)		
	All	SNR	BDM-LCMV	SNR	BDM-LCMV	SP-AAP
C1	81.34 ± 2.67	83.02 ± 2.83	82.24 ± 3.26	86.78 ± 2.81	88.14 ± 2.25	85.47 ± 2.47
C2	90.48 ± 1.80	90.09 ± 1.38	92.55 ± 1.42	90.57 ± 2.09	93.56 ± 1.43	92.89 ± 1.23
C3	94.09 ± 1.52	95.00 ± 2.37	94.86 ± 1.51	94.72 ± 2.28	95.83 ± 1.91	95.13 ± 1.77
C4	93.16 ± 0.92	93.65 ± 1.44	93.18 ± 0.80	94.14 ± 1.51	93.94 ± 1.05	93.68 ± 0.86
C5	90.96 ± 1.92	87.59 ± 3.22	92.17 ± 1.21	89.48 ± 2.74	93.35 ± 1.66	92.88 ± 1.74
C6	94.40 ± 1.14	90.55 ± 1.69	96.14 ± 0.89	91.78 ± 1.27	93.82 ± 1.73	94.54 ± 1.32
C7	97.58 ± 0.47	94.57 ± 1.58	97.36 ± 0.84	94.66 ± 1.70	97.19 ± 0.62	97.61 ± 0.45
OA	89.77 ± 0.54	89.56 ± 0.74	90.65 ± 0.74	91.00 ± 0.64	92.71 ± 0.54	91.76 ± 0.64
AA	91.72 ± 0.35	90.64 ± 0.57	92.64 ± 0.31	91.73 ± 0.57	93.69 ± 0.44	93.17 ± 0.42
κ	87.49 ± 0.65	87.22 ± 0.89	88.57 ± 0.88	88.96 ± 0.77	91.06 ± 0.65	89.91 ± 0.77
Class	Band Selection (MLR)			Spectral Partitioning (SP) (MLR)		
	All	SNR	BDM-LCMV	SNR	BDM-LCMV	SP-AAP
C1	77.51 ± 2.92	79.81 ± 2.13	81.28 ± 1.91	90.36 ± 1.95	90.01 ± 2.34	84.05 ± 3.82
C2	85.89 ± 1.49	90.49 ± 1.29	89.02 ± 1.61	91.73 ± 1.48	90.1 ± 1.58	87.26 ± 1.46
C3	91.78 ± 1.76	94.99 ± 2.02	94.83 ± 2.07	95.4 ± 1.75	94.56 ± 1.76	92.67 ± 1.75
C4	93.4 ± 1.43	92.33 ± 1.93	90.43 ± 1.77	94.53 ± 1.32	93.92 ± 1.01	93.95 ± 1.47
C5	93.95 ± 1.51	91.31 ± 1.79	91.91 ± 1.42	93.72 ± 0.94	92.6 ± 1.37	93.76 ± 1.32
C6	99.05 ± 0.47	84.99 ± 7.47	70.56 ± 5.45	96.42 ± 0.51	84.28 ± 7.2	99.16 ± 0.34
C7	96.68 ± 1.29	95.22 ± 1.34	82.67 ± 3.34	97.08 ± 0.72	87.98 ± 4.84	95.37 ± 1.52
OA	87.95 ± 0.75	88.56 ± 0.77	86.19 ± 0.68	93.19 ± 0.65	90.86 ± 0.81	90.10 ± 1.09
AA	91.18 ± 0.50	89.88 ± 1.12	85.81 ± 0.73	94.18 ± 0.44	90.50 ± 0.93	92.32 ± 0.64
κ	85.35 ± 0.90	86.05 ± 0.93	83.18 ± 0.81	91.65 ± 0.79	88.79 ± 0.98	87.90 ± 1.31

the set of indices of the original spectral bands while the y-axis represents the selected bands for the original image and for each class. From Fig. 3.13, we can observe that each class-oriented partition takes a band subset of much lower dimensionality than the original number of bands. In all cases, the selection of partitions allows us to obtain



(a) SVM (75%) (b) SVM-SNR (58%) (c) SVM-BDM-LCMV (78%) (d) SVM-SP-SNR (75%)

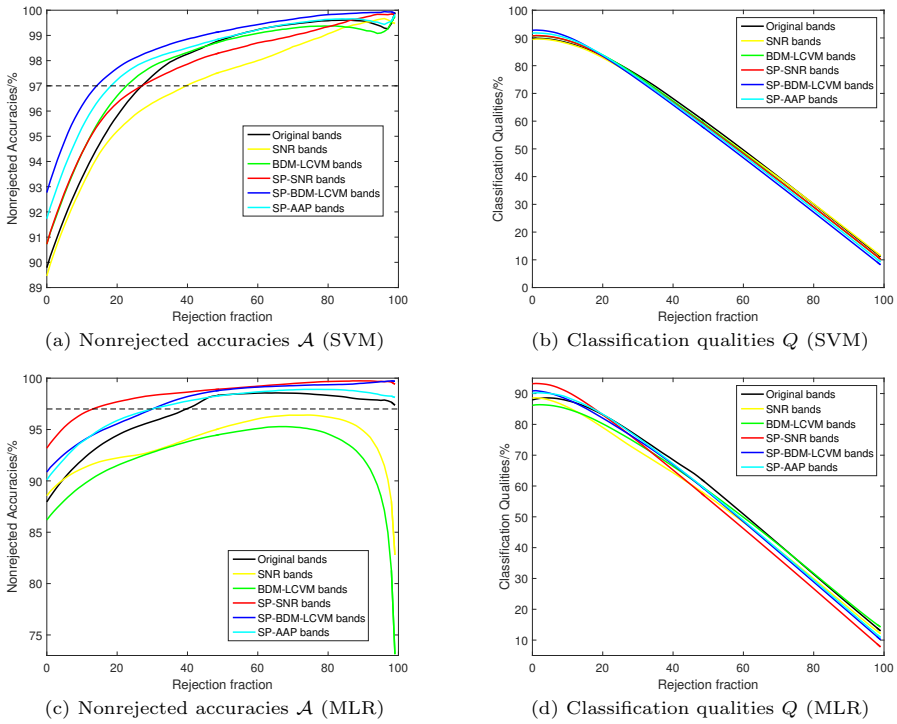
(e) SVM-SP-BDM-LCMV (87%) (f) SVM-SP-AAP (84%) (g) MLR (55%) (h) MLR-SNR (0%)

(i) MLR-BDM-LCMV (0%) (j) MLR-SP-SNR (85%) (k) MLR-SP-BDM-LCMV (70%) (l) MLR-SP-AAP (65%)

Figure 3.14: Classification maps (with rejection) obtained by the proposed classification framework using the original spectral information (a,g), the 40 selected bands by using SNR (b,h) and BDM-LCMV (c,i) algorithms, the spectral partitions obtained by our proposed approach implemented with SNR (d,j) and BDM-LCMV (e,k), and the SP-AAP method (f,l) for the HYDICE Washington DC mall data. The number in the parenthesis denotes the proportion of pixels remaining after rejection. In all cases, a total of 2% randomly selected training samples were used for training. The maps are displayed with partial pixels rejected in order to obtain a 97% classification accuracy for the remaining ones. Note that the overall accuracies for the unrejected pixels of each map are calculated by considering the labeled test samples, with the training samples excluded.

multiple views of the original data.

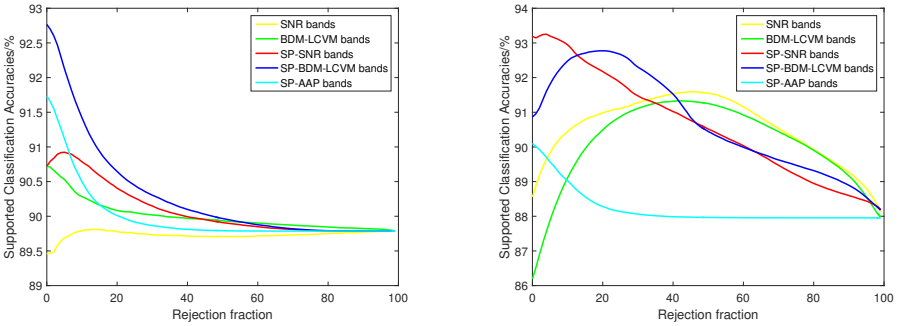
Table 3.6 displays the OA, AA, κ, and individual accuracies obtained after using only 2% randomly selected samples for training and the rest for testing [see Fig. 3.2(i)]. Several observations can be drawn from Table 3.6. First of all, an increase of about 1% to 6% in classification OA is obtained by our presented SP-SNR and SP-BDM-LCMV methods in comparison with those using all the spectral bands, and also with the



(a) Nonrejected accuracies $\mathcal{A}$ (SVM) (b) Classification qualities Q (SVM)

(c) Nonrejected accuracies $\mathcal{A}$ (MLR) (d) Classification qualities Q (MLR)

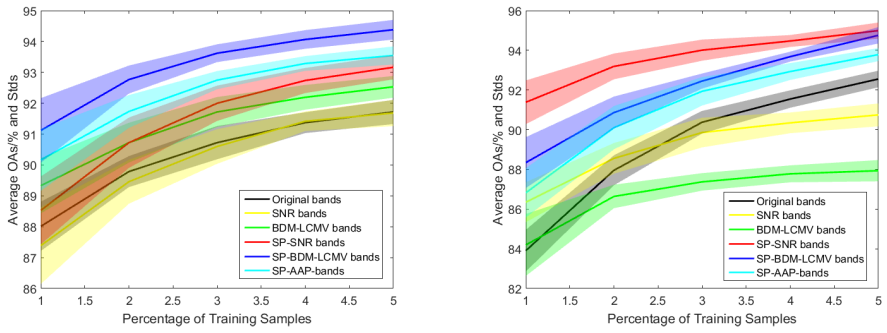
Figure 3.15: Statistics of nonrejected accuracies $\mathcal{A}$ and classification qualities $\mathcal{Q}$ of the classifications as a function of rejected fractions with the HYDICE Washington DC Mall scene. These plots correspond to the results in Table 3.6.



(a) SVM (b) MLR

Figure 3.16: Classification OAs (as a function of rejections) obtained by our proposed classification framework with the HYDICE Washington DC mall data, after being supported by the rejected pixels of the classifier using the original spectral information. Note here that the plots of both figures (a),(b) are obtained, respectively, from the same one Monte Carlo run of Table 3.6 that is close to the statistical average.

single band selection approaches. This observation is consistent with those made for the two previous hyperspectral data sets, suggesting the effectiveness of our newly proposed methods for multiple classifier-based feature learning based on multiple views that are



(a) SVM (b) MLR

Figure 3.17: Overall classification accuracies (as a function of the number of training samples) obtained by the proposed classification framework (with the original spectral information and with SNR, BDM-LCMV band selection methods and the AAP spectral partitioning method) for the HYDICE Washington DC mall scene. The solid lines represent the average of 20 Monte Carlo runs, whereas the colored areas around the lines represent the standard deviation around the mean. Plots (a) and (b) are respectively obtained by using SVM and MLR classifiers.

Table 3.7: Processing times of different methods with HYDICE Washington DC mall scene.

Time/sec	SVM (different percentage of training samples [%])				
	1	2	3	4	5
Original	0.91 ± 0.19	1.74 ± 0.18	3.30 ± 0.15	5.08 ± 0.15	7.12 ± 0.33
SNR	0.80 ± 0.03	1.47 ± 0.07	2.70 ± 0.10	4.29 ± 0.21	6.12 ± 0.19
BDM-LCMV	0.79 ± 0.04	1.49 ± 0.08	2.71 ± 0.11	4.27 ± 0.16	6.14 ± 0.27
SP-SNR	5.32 ± 0.22	10.06 ± 0.52	18.80 ± 0.59	29.76 ± 1.02	42.35 ± 1.03
SP-BDM-LCMV	5.24 ± 0.23	9.96 ± 0.52	18.76 ± 0.72	29.53 ± 0.76	42.16 ± 1.39
SP-AAP	3.10 ± 0.16	5.88 ± 0.28	10.95 ± 0.42	17.30 ± 0.49	24.48 ± 0.66
Time/sec	MLR (different percentage of training samples [%])				
	1	2	3	4	5
Original	5.09 ± 0.15	6.29 ± 0.24	7.47 ± 0.24	8.49 ± 0.30	9.09 ± 0.22
SNR	3.96 ± 0.05	4.29 ± 0.08	4.67 ± 0.08	4.94 ± 0.07	5.24 ± 0.07
BDM-LCMV	3.96 ± 0.04	4.27 ± 0.07	4.55 ± 0.06	4.82 ± 0.08	5.11 ± 0.06
SP-SNR	27.27 ± 0.27	29.47 ± 0.27	31.66 ± 0.44	33.50 ± 0.61	35.52 ± 0.53
SP-BDM-LCMV	27.02 ± 0.30	28.99 ± 0.35	30.76 ± 0.43	32.31 ± 0.47	34.17 ± 0.51
SP-AAP	15.65 ± 0.16	16.82 ± 0.20	17.94 ± 0.21	18.92 ± 0.27	19.96 ± 0.32

provided by spectral partitions. Secondly, in comparison with the SP-AAP method, both SP-SNR and SP-BDM-LCMV obtained improved accuracies. Specifically, relevant improvements are obtained for the individual classes. This is also consistent with the observations made for the results obtained with the previous hyperspectral data sets.

In order to better illustrate this, Fig. 3.14 displays the classification maps obtained for different methods after rejection to pursue 97% OA of the nonrejected pixels. It is remarkable that, by using our presented spectral partitioning methods (SP-SNR and SP-BDM-LCMV), an increase of 12% (SVM) and 30% (MLR) in OA are achieved in comparison with using the spectral bands of the original image, and 3% to 20% improvements in OA are obtained as compared with the SP-AAP. In order to better

demonstrate this, Fig. 3.15 plots the classification accuracy $\mathcal{A}$ and quality $\mathcal{Q}$ as functions of the rejection fractions. In Fig. 3.15, it is remarkable that our presented method obtains more confidence in the classification of individual pixels than the other methods. From the previous experiments with the ROSIS and AVIRIS data sets, we could already observe that different nonrejected accuracies lead to improvements in classification when combined with the classification results obtained using the original image. Hence, in the case of the HYDICE Washington DC mall data set, we plot the supported classification results in Fig. 3.16. From Fig. 3.16, we can see that the supported classification accuracies are improved up to 5%. It is also remarkable that, after combining two classifications, the results can be further improved, especially in the case of using the MLR classifier.

With the aforementioned experimental observations in mind, we conclude that our presented spectral partitioning approaches (SP-SNR and SP-BDM-LCMV) generally obtain better classification accuracies when compared with the SP-AAP method. The obtained results are dependent on the learning rules of the classifiers as well as on the considered hyperspectral data sets. For example, the SP-BDM-LCMV generally outperforms other methods, while the SP-SNR also provides competitive performance in all cases. In turn, the SP-AAP provides comparatively worst classification accuracies when using the SVM classifier with the ROSIS Pavia University data set. However, when the MLR classifier is used, the SP-AAP performs better than the SP-SNR and worse than the SP-BDM-LCMV (see Table 3.1 and Fig. 3.6).

To conclude this section, we evaluate the sensitivity of the compared methods to different percentages of training samples. Fig. 3.17 plots the overall accuracies of different methods when using different proportions of training samples as input. First of all, it is remarkable that our presented spectral partitioning methods provide highly competitive results in this experiment. The processing times, reported on Table 3.7, indicate that our methods need about 4 to 6 times more computation than the classification of the original image and about 2 times more than the SP-AAP. These results can be improved by straightforward parallelization, as discussed in previous experiments.

3.5 Summary and future directions

In this chapter, we presented a new class-oriented spectral partitioning method for hyperspectral image classification. The proposed method is shown to be effective in the task of exploiting the information contained in the specific classes by rearranging the original spectral information. This is mainly done to address the Hughes phenomenon by means of a multiple classifier system, while avoiding the elimination of relevant spectral bands in the original hyperspectral image that may be useful for the discrimination of the classes. Inspired by the idea of classification with rejection, we also designed a strategy that combines different classifiers for further improving the classification result. In our experiments, two spectral partitioning methods are constructed based on two well-established band selection algorithms: SNR and BDM-LCMV. The obtained spectral partitions keep most of the relevant spectral bands from the original image and provide different views (understood as low-dimensional partitions) for multiple feature

learning, thus addressing the potential problems associated to the limited availability of training samples. Our experiments illustrated reasonable advantages in classification accuracy achieved by our presented spectral partitioning framework with three well-known hyperspectral images: the ROSIS Pavia University data, the AVIRIS Indian Pines data and the HYDICE Washington DC mall data. Furthermore, the experimental classification results indicate that the key for a successful spectral partitioning lies in the capability to generate a group of spectral partitions with diverse views of the original hyperspectral image. Besides, we also observed the effectiveness of combining different classifiers with rejection (including those resulting from the multiple classifier views generated by spectral partitioning and the original hyperspectral image). In our future work, we will focus on exploring the potential of generating a group of spectral partitions by multiple criteria, given by an extensive set of band selection algorithms instead of using only one criterion. The potential of employing different classifiers will also be considered to enlarge the diversity of the multiple views generated through spectral partitions. We will also work towards the parallel implementation of the presented methods, which will allow us to increase their computational performance due to their inherently parallel nature.

Chapter 4

Convex Formulation for Remote Sensing Image Classification with Multi-Superpixelizations

4.1 Summary

Superpixels are a powerful device to characterize spatial-contextual information in remotely sensed hyperspectral images interpretation. However, the exploitation of superpixels is not straightforward, as it leads to a nonconvex, NP-hard discrete integer optimization problem. In this chapter, we attack this problem in a compact domain which allows us to build convex functions, thus opening the door to the incorporation of over-segmented superpixels. Specifically, this chapter first develops a new method for generating over-segmented superpixels. Then, we introduce a convex expression for over-segmented superpixels that adopts the form of graph total variation. This chapter also considers the vectorial total variation in order to promote piecewise smoothness and align discontinuities along the boundaries. An approximated optimal solution of the resulting problem is provided with an instance of the split augmented Lagrangian shrinkage algorithm (SALSA). Experiments on hyperspectral remote sensing images demonstrate that the integration of vectorial total variation and superpixel-based graph total variation is capable of improving the labeling results with much precise boundaries and inner consistency inside over-segmented superpixels, leading to high classification accuracies.

4.2 Introduction

Supervised classification is an important task for hyperspectral remotely sensed data exploitation, which assigns a set of class labels to each pixel in the scene given an available training reference [3]. In this context, spatial information has been shown

capable of greatly improving the classification performance from the viewpoint of statistical accuracy and mapping effectiveness [135, 183, 186–189]. The inclusion of spatial information is often tackled by means of image segmentation. As a discrete problem, segmentation aims to partition an image into multiple segments, which consist of a set of pixels that share some common characteristics(i.e. they belong to the same object or may have the same surface orientation). As an important source of spatial information, image segments or superpixels play a significant role in remote sensing image analysis [3], as well as machine vision [190], medical imaging [191], etc. The pixels comprising the same superpixel (especially the over-segmented ones) are generally believed to share highly similar characteristics, such as class labels [192]. At the superpixel level, hyperspectral image classification can also be processed much faster as compared to pixel-wise techniques [187, 192].

Many techniques and methods have been developed to deal with the image segmentation problem in the spatial domain, such as thresholding [190], clustering, compression/histogram/edge-based algorithms [193, 194], region growing [195], integer optimization via graph-cuts [88, 196], variational methods [197, 198] as well as Bayesian theory based algorithms such as Markov random fields (MRFs), etc. [134, 199]. However, image segmentation usually leads to an integer optimization problem that is NP-hard and thus hard to be solved exactly. This is because the label image is naturally a discrete representation of the original image. Actually, in the context of supervised image segmentation, apart from a few examples almost all functions associated with a realistic model are non-convex and even NP-hard [200]. This means that they are hard to solve and hence a direct minimization usually leads to poor local minima. A popular and well-established paradigm for modelling these problems is function or energy minimization, where the spatial information is tackled with the Potts model in the Markov Random Field (MRF) community or the minimal partition problem in the partial differential equations (PDE) community. More interesting details regarding binary labeling problems are experimentally surveyed in the work of Klodt et al. [201], and the more general case of multi-label problems is reported in the work of Nieuwenhuis [200].

In the MRF community, it is often assumed that labels of neighbouring pixels follow a Gibbs distribution [199]. In this context, graph-cut algorithms have been developed to model the resulting integer optimization problem. In the work [84], Boykov and Jolly firstly present the optimal solution of image binary segmentation via graph-cut algorithms that solves max-flow problems. Recent efforts [84, 202–204] attack this problem under a discrete optimization framework, by introducing prior regularizers to promote the spatial patterns of the label image and approximate the solutions via graph-cut algorithms. Kohli et al. designed a novel model for enforcing label consistency, which is able to combine multiple image segmentations in a principled manner based on the higher order conditional random fields (CRFs) [90]. Following this line, an optimal solution can be found for the binary case and can only be approached for the multi-label case. Other techniques also attack the labeling problem via block coordinate descent (BCD) [205] in a dual objective, such as MPLP [206], MSD [207], TRW-S [88], quadratic pseudo-boolean optimization (QPBO) [208, 209], among many others.

In the hyperspectral remote sensing image classification literature, the concept of superpixel has been widely explored and utilized. Superpixels, especially over-segmented ones, play a significant role in promoting contextual consistency of hyperspectral classification, which is usually limited by the imbalance between high spectral dimensionality and limited training samples [187]. Using image segments, Tarabalka et al. developed several post-processing techniques for improving hyperspectral image classification [10, 85, 186]. Fang et al. addressed this problem under a sparse model with superpixels [188]. In a more straightforward way, Li et al. embedded the graph-cut algorithm associated with spatial information into a novel active learning scheme that iteratively updates the data term to remarkably improve the labeling process [135]. The work by Zhang [187] et al. tackled this problem by following a Bayesian framework also called super pixel-based graphical model. In the work [189], the discontinuity information provided by boundaries has been formulated to reinforce the label consistency for hyperspectral image classification. In polarimetric synthetic aperture radar (PolSAR) image analysis, Xu et al. combine statistical information with spatial-contextual information using the stochastic expectation maximization (SEM) algorithm [210]. However, in spite of its great success, the NP-hardness of the integer optimization problems renders little flexibility with respect to including superpixels as a spatial prior, mainly due to its discrete nature. In addition, it is generally very difficult to decide which is the best available image superpixelization considering both the variety of image segmentation techniques and the parametric adjustments. Thus, it remains very challenging to naturally exploit superpixels in supervised labeling scenarios for hyperspectral image classification. This problem becomes even more complicated when we consider multiple superpixelizations.

4.2.1 Contributions

As mentioned before, the Bayesian framework is widely used in order to exploit spatial-contextual information. Under the Bayesian perspective, spatial or contextual information can both be viewed as priors to the conditional probabilities. However, the maximum a posteriori (MAP) segmentation leads to integer optimization problems that are hard to solve due to their discrete nature [211]. In order to deal with this issue, the linear program (LP) [207, 212] or convex relaxation has been used to relax the discrete labeling problem into a compact domain. It is then much easier to convexly model the prior regularizers over the compact set, which opens the door to the inclusion of different priors resulting from real-world knowledge. The solution of the original problem is then approximated (or even obtained) under a primal-dual scheme [205–207, 209] using a linearly relaxed approach in polynomial time.

Based on the convex relaxation program, and in the spirit of [91, 211, 213], this chapter introduces a new image labeling mechanism which is extremely flexible with respect to the inclusion of spatial information coming from superpixels, in the form of a spatial regularizer. In this regularizer, each superpixel is formulated as a constraint of graph total variation (GTV) that reinforces the pairwise label consistency in between its comprising pixels. In addition to the graph total variation associated with superpixels,

this chapter also utilizes a second spatial regularizer, the vectorial total variation (VTV), which promotes piecewise smoothness and aligns discontinuities along the edges in the image domain [211], thus improving the boundary recall of the resulting classification. Besides, a new framework based on the sigular value decomposition (SVD) is designed for the purpose of adapting the superpixelization or superpixel generation method to hyperspectral images, considering their high spectral dimensionality. Here, we specifically generate over-segmented superpixels with the fast simple linear iterative clustering (SLIC) algorithm [214] considering that the over-segmented ones are more likely to be homogeneous.

The main contributions of this chapter can be summarized as follows: 1) development of a new strategy to relax the NP-hard integer optimization problem related to image labeling into a compact domain, and characterization of over-segmented superpixels as a GTV regularizer under a Bayesian image segmentation perspective, 2) introduction of a VTV as a second spatial regularizer for boundary recalling purposes and development of a new algorithm based on the SALSA [215] method to solve the resulting problem, 3) adaptation of the SLIC algorithm [214] to hyperspectral image superpixelization, and 4) provision of experimental evidences illustrating the potential of the proposed methodology in the context of hyperspectral remote sensing image classification.

4.2.2 Related work

In order to tackle the aforementioned integer optimization problem associated with image segmentation, the work by Marroquin et al. [216] extended the Bayesian segmentation framework with a hidden MRF paradigm, which transforms the NP-hard optimization problem to a continuous domain. Under this paradigm, one can include additional information as a prior to the maximum likelihood function, such as MRF [216] or a wavelet-based prior [217]. In the context of convex optimization, some NP-hard problems associated with integer optimization problems like shortest path, max-flow etc, are often first relaxed and then solved or approximated as LP or SDP problems. The hidden fields paradigm can actually be viewed as a statistical interpretation of the relaxation technique. In the work [213] Condessa et al. sidestep the discrete nature of image segmentation by formulating the problem in a Bayesian framework with a hidden set of real-valued random fields. Then the segmentation via the constrained split augmented Lagrangian shrinkage algorithm (SegSALSA) is introduced to infer the hidden fields. In turn, the labels can be obtained via marginalized MAP (MMAP). By this means, prior information such as structure tensor regularization (STR) [218] and VTV [219, 220] are incorporated under a convex scheme.

This chapter also has strong connections with the work of Bioucas-Dias et al. [211] and the work of Condessa et al. [213]. There is, however, a major difference. The methodologies presented in [211, 213] compute the probabilities of labelings to use elsewhere, namely, in statistical inference problems. Our objective, more in line with [91, 221], is to use convex relaxation to approximate the original discrete problem. Similar to the work [213, 222], our resulting algorithm is also convex, time-efficient and highly parallelizable. The remainder of this chapter is organized as follows: Section 4.3

introduces the problem and the objective function associated with the vectorial total variation and graph total variation regularizers under the MAP famework. Section 4.4 describes our newly developed algorithm, that is an instance of the split augmented Lagrangian shrinkage (SALSA), to solve the resulted problem in Section 4.3. Section 4.5 presents experimental evidence of the performance of our proposed method in the context of hyperspectral remote sensing classification. Section 4.6 concludes the chapter with some remarks and hints at plausible future lines.

4.3 Problem Formulation

The mathematical terms of the image segmentation problem are formulated with the following notations. Let $\mathcal{S} \equiv \{1, 2, \cdots, n\}$ be a set of integers indexing the n pixels of an image, $\mathbf{x} \equiv [\mathbf{x}_1, \cdots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ a matrix of n vectors across d dimensions. Let $\mathbf{y} \equiv (y_1, \cdots, y_n) \in \mathcal{L}^n \equiv \{1, \cdots, K\}^n$ be an image of class labels, termed segmentation or labeling, such that $y_i = k$ if and only if the label of pixel i belongs to class k. Given$\mathbf{x}$, a supervised image segmentation aims to find a partition $P \equiv \{R_1, \cdots, R_K\}$ of $\mathcal{S}$ such that the features indiced by a given setR_i, for $i \in \{1, \cdots, K\}$, are *similar* in some sense. Similarly, an unsupervised image superpixelization can be represented by another partition $\mathcal{N} \equiv \{\mathcal{V}_1, \mathcal{V}_2, \cdots, \mathcal{V}_T\} = \mathbf{x}$ out of $\mathcal{S}$, where the t-th superpixel clique comprises of n_t pixels and thus $\sum n_t = n$. We remark that there exists a one-to-one correspondence between a superpixel and a graph. And the main goal of this chapter is to straightforwardly reinforce the spatial information provided by the superpixels into the supervised image segmentation problem.

4.3.1 Maximum *a posteriori* segmentation

We adopt a Bayesian perspective into the segmentation problem. Given the posterior probability $p_{\mathbb{Y}|\mathbb{X}}(\mathbf{y}|\mathbf{x})$, the observation model $p_{\mathbb{X}|\mathbb{Y}}(\mathbf{x}|\mathbf{y})$, and the prior probability $p_{\mathbb{Y}}(\mathbf{y})$ (often an MRF), the MAP segmentation is given by

$$\widehat{\mathbf{y}}_{MAP} = \arg\max_{\mathbf{y} \in \mathcal{L}^n} p_{\mathbb{Y}|\mathbb{X}}(\mathbf{y}|\mathbf{x}) = \arg\max_{\mathbf{y} \in \mathcal{L}^n} p_{\mathbb{X}|\mathbb{Y}}(\mathbf{x}|\mathbf{y}) p_{\mathbb{Y}}(\mathbf{y}). \tag{4.1}$$

Under the conditional independence assumption, we have

$$p_{\mathbb{X}|\mathbb{Y}}(\mathbf{x}|\mathbf{y}) = \prod_{i=1}^{n} p_{\mathbb{X}_i|\mathbb{Y}_i}(\mathbf{x}_i|y_i) = \prod_{k=1}^{K} \prod_{i \in R_k} p(\mathbf{x}_i), \tag{4.2}$$

where $p(\mathbf{x}_i) = p_{\mathbb{X}_i|\mathbb{Y}_i}(\mathbf{x}_i|y_i = k)$. In a supervised scenario, the class probabilities $p_{\mathbb{X}_i|\mathbb{Y}_i}(\cdot|y_i = k)$ for $k \in \mathcal{L}$ are already known or learned from a training set. Having into consideration (4.2), we may write

$$\begin{aligned} \widehat{\mathbf{y}}_{MAP} &= \arg\min_{\mathbf{y} \in \mathcal{L}^n} -\log\left(p_{\mathbb{X}|\mathbb{Y}}(\mathbf{x}|\mathbf{y}) p_{\mathbb{Y}}(\mathbf{y})\right) \\ &= \arg\min_{\mathbf{y} \in \mathcal{L}^n} \sum_{i=1}^{n} D_i(y_i) + \lambda U(\mathbf{y}), \end{aligned} \tag{4.3}$$

where $D_i(y_i) = -\log p_{\mathbb{X}_i|\mathbb{Y}_i}(\mathbf{x}_i|y_i)$ denotes the log-likelihood probability density (often called a data term in the bayesian image segmentation scenario), $\lambda U(\mathbf{y}) = -\log p_{\mathbb{Y}}(\mathbf{y})$ corresponds to the prior function, and $\lambda > 0$ is a tunable regularization parameter controlling the power of spatial prior that is often an MRF. The minimization of (4.3) is a nonconvex integer optimization problem over a discrete domain of $\mathbf{y}$. In the case of Potts [134] prior and $K = 2$, the problem has an exact solution obtained by mapping the problem into of a min-cut computation on a suitable graph [223]. However, for $K > 2$, the optimization (4.3) is proven NP-hard [88] and, therefore, usually only approximations may be computed. To further complicate the use of integer formulations, the class of regularizers U that may be used in (4.3) is quite narrow; for example, it is not a simple task to include prior coming from superpixels into U.

4.3.2 Convex relaxation

In order to utilize the linear program of convex relaxation, first we replace the constraint $\mathbf{y} \in \mathcal{L}^n$ in (4.3) by a more common integer constraint in convex optimization problems. Let $\mathbf{z}_i = [z_{1i}, \ldots, z_{Ki}]^T \in \{0,1\}^K$ be a "1-of-K" representation of y_i; that is, $(y_i = k) \Leftrightarrow [z_{il} = 0 \text{ for } l \neq k \text{ and } z_{ik} = 1]$. Using this representation, optimization (4.3) may be equivalently written as

$$\begin{aligned} \widehat{\mathbf{z}} \in \arg \min_{\mathbf{z}} & \sum_{i=1}^{n} \mathbf{q}_i^T \mathbf{z}_i + \lambda\phi(\mathbf{z}), \\ \text{s.t.:} \quad & \mathbf{1}_K^T \mathbf{z} = \mathbf{1}_n^T \\ & \mathbf{z} \in \{0,1\}^{K \times n} \end{aligned} \tag{4.4}$$

where $\mathbf{z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n]$ is the relaxed hidden field, $\mathbf{q}_i = [D(y_i = 1), \ldots, D(y_i = K)]^T$, $\phi(\mathbf{z}) = U(\mathbf{y})$, and $\mathbf{1}_p$ is a p-dimensional column vector of 1's.

As proposed in [91, 221], and also related with [211, 213], we relax the optimization (4.4) by replacing the discrete set $\{0,1\}$ to the interval $[0,1]$, obtaining the optimization

$$\begin{aligned} \widehat{\mathbf{z}} \in \arg \min_{\mathbf{z} \in \mathbb{R}^{K \times n}} & \sum_{i=1}^{n} \mathbf{q}_i^T \mathbf{z}_i + \lambda\phi(\mathbf{z}), \\ \text{s.t.:} \quad & \mathbf{1}_K^T \mathbf{z} = \mathbf{1}_n^T \\ & \mathbf{z} \geq 0 \end{aligned} \tag{4.5}$$

where $\lambda > 0$ is a tunable regularization parameter. This resulted (4.5) is not equivalent to (4.4) but, still, a very close approximation. The solution $\widehat{\mathbf{z}}$ to the relaxed linear program can be used to gain information about the solution to the original integer program. Although the solutions $\widehat{z}_{ki}$ yielded by (4.5) are mostly discrete [200], a few elements, mostly in the boundary of the classes, may not be in $\{0,1\}$. In order to recover a complete discrete solution, we can compute

$$\widehat{y}_i = \arg\max_k \ \widehat{z}_{ki}, \ \ i \in \mathcal{S}, \ \ k \in \mathcal{L}.$$

The formulation (4.5) yields excellent results when compared with the original integer formulation, as extensively illustrated in [200]. In addition, a proper tailoring to the

function $\phi(\cdot)$ allows to embrace a much larger group of prior information sources that comes from real-world knowledge of the scene, besides the Gibbs distribution assumption, the MRF or the CRF. This is exactly the possibility that we explore in this chapter.

4.3.3 Spatial regularizers

4.3.3.1 Vectorial Total Variation

Herein this chapter uses two spatial regularizers of the VTV $\phi_{VTV}(\cdot)$ and GTV $\phi_{GTV}(\cdot)$. Specifically, the ϕ_{VTV} is of a form of vectorial total variation (VTV) [220]

$$\phi_{VTV}(\mathbf{z}) \equiv \lambda_1 \sum_{n\in\mathcal{S}} \omega_n \sqrt{||\mathbf{D}_h\mathbf{z}[n]||^2 + ||\mathbf{D}_v\mathbf{z}[n]||^2}, \tag{4.6}$$

where $\mathbf{D}_h, \mathbf{D}_v : \mathbb{R}^{K\times n} \mapsto \mathbb{R}^{K\times n}$ are linear operators computing horizontal and vertical first order backward vector differences, respectively, ω_n can often be the magnitude at pixel $[n]$ weighting the n-th piece of the VTV regularizer, $||\cdot||$ is the standard Euclidean norm, and $\lambda_1 > 0$ is the regularizer parameter controlling the strengh of the spatial prior. The regularizer (4.6) is utilized in order to promote the piecewise smoothness of $\mathbf{z}$ and also preserve aligned edges across $\mathbf{z}$ in the image domain. We employ this isometric total variation considering its capacity to allow piecewise variability along different directions. The vectorial total variation is convex, although not strictly, allowing optimization via proximal methods relying on Moreau proximity operators [213, 224].

4.3.3.2 Graph Total Variation

Given a graph $\mathcal{N}_j$, for $j = 1, \cdots, C$, associated with the j-th superpixelization, its corresponding graph total variation (GTV) is formulated as the sum of the total variation over each subgraph $\mathcal{V}_{j,t}$, for $t = 1, \cdots, n_T$, in the following

$$\phi_{\text{GTV}}(\mathbf{z}) = \sum_{\mathcal{V}_{j,t}\in\mathcal{N}_j} \omega_t \sum_{(m,l)\in\mathcal{V}_{j,t}} \omega_{ml}||\mathbf{z}_m - \mathbf{z}_l||^2,$$

where ω_{ml} denotes the pairwise weight between nodes m and l in the subgraph $\mathcal{V}_{j,t}$. Minimizing this graph total variation thus leads to the promoted consistency of the hidden layers $\mathbf{z}$ for each over-segmented superpixel partition. Without providing extra prior information to each node, we can simply set $\omega_{ml} = \frac{1}{n_{j,t}}$ and formulate the GTV associated with multiple superpixelizations as

$$\phi_{\text{GTV}}(\mathbf{z}) = \sum_{j=1}^{C} \omega_c||(\mathbf{A}_c - \mathbf{I})\mathbf{z}||_F^2 \quad = \sum_{j=1}^{C} \sum_{\mathcal{V}_t\in\mathcal{N}_j} \omega_t||(\mathbf{A}_{j,t} - \mathbf{I})\mathbf{z}_{[\mathcal{V}_t]}||_F^2, \tag{4.7}$$

where j indices the j-th graph/superpixelization, $\mathbf{A}_j$ is a blockwise disjoint union of fully connected subgraphs, $\mathbf{A}_{j,t} = \mathbf{E}/n_{j,t}$ is a normalized adjacency square matrix of the subgraph $\mathcal{V}_{j,t}$, $n_{j,t}$ is the node number of sugbraph $\mathcal{V}_{j,t}$, $\mathbf{E}$ is a square 1-matrix of size $n_{j,t}$, and $||\cdot||_F$ is the Frobenius norm. Note that there is a one-to-one correspondence between graph and superpixelization, subgraph and partition element (superpixel), and node and pixel. This means that each node (pixel) is connected to all other nodes

belonging to the same partition element, and to no node belonging to a different partition element. Thus, the minimization over $\mathbf{z}$ of the graph Laplacian $(\mathbf{A}_j - \mathbf{I})$ promotes the minimization of the total variation of the vectors $\mathbf{z}_i$ across the pixels i belonging to the same partition element, for all partition elements belonging to the j-th superpixelization. Therefore, this regularizer promotes constant vectors $\mathbf{z}_i$ within the partition elements. Under the perspectives of both Bayesian labeling framework and convex optimization, this formulation is based on the definition of the GTV prior as sums of convex functions and associated linear operators. In this chapter, we use the fast superpixel clustering algorithm SLIC [214] aiming to obtain multiple over-segmented spatial partitions.

By taking the Frobenius norm, we can decouple (4.7) pixelwisely and superpixelwisely, thus opening the possibility to flexibly weight specific superpixelizations, objects or classes for more practical purposes. On the other side, note that here we are allowed to combine multiple superpixelizations/segmentations at the same time, which avoids the dilemma of selecting "the best" segmentation.

4.4 Optimization Algorithm

Having in mind the data term $-(\log \mathbf{p}_i^T)\mathbf{z}_i$ or $\mathbf{q}_i^T\mathbf{z}_i$ ($\mathbf{p}$ is the probability of vectors that are often known or learned by supervised classifiers), the VTV (4.6) and the GTV (4.7), the resulting optimization problem turns out to be

$$\begin{aligned} \widehat{\mathbf{z}} \in \ \arg \min_{\mathbf{z}\in\mathbb{R}^{K\times n}} & \sum_{i=1}^{n} -(\log \mathbf{p}_i^T)\mathbf{z}_i \\ & + \lambda_1 \sum_{n\in\mathcal{S}} \omega_n \sqrt{||\mathbf{D}_h\mathbf{z}[n]||^2 + ||\mathbf{D}_v\mathbf{z}[n]||^2} \\ & + \lambda_2 \sum_{c=1}^{C} \omega_c ||(\mathbf{A}_c - \mathbf{I})\mathbf{z}||_F^2, \\ \text{s.t. :} \quad & \mathbf{z} \geq 0,\ \mathbf{1}_K^T\mathbf{z} = \mathbf{1}_n^T,\ \lambda_1 \geq 0,\ \lambda_2 \geq 0 \end{aligned} \tag{4.8}$$

where the constraint $\mathbf{z} \in [0,1]^{K\times n}$ is removed as it is equivalent to $\mathbf{z} \geq 0$ and $\mathbf{1}_K^T\mathbf{z} = \mathbf{1}_n^T$, C is the number of graphs/superpixelizations of the original image In (4.8) the first part, often called the data term, is a linear combination of variable $\mathbf{z}_i$, which is thus a convex term. Together with the convex (4.6) and (4.7), (4.8), we obtain a convex optimization problem, which is exactly what we originally target in this chapter.

The optimization (4.8) has a solution as it is a convex problem defined on a compact set. The resulting optimization algorithm is very similar to SALSA [215] and SegSALSA [211,213]. To solve the problem given by (4.8), we start with the formulation of SALSA introduced in [215]:

$$\min_{\mathbf{z}\in\mathbb{R}^{K\times n}} \sum_{j=1}^{3} f_j(\mathbf{z}) + \sum_{j=1}^{m+2} g_j(\mathbf{H}_j^g\mathbf{z}) \tag{4.9}$$

where the convex functions f_j, for $j = 1, \cdots, 3$, correspond to the data term, sum-to-one constraint and nonnegativity constraints, respectively, and g_j , for $j = 1, \cdots, C+2$ are a

set of closed, proper and convex functions, $\mathbf{H}_j^g$, for $j = 1, \cdots, C+2$ are linear operators, and $\sum_{j=1}^{C+2} g_j(\mathbf{H}_j^g \mathbf{z})$ corresponds to a prior which is a summation of $C+2$ terms. Note here that C is the number of graphs/superpixelizations, while the number "2" corresponds to the vectorial total variation along the horizontal and vertical directions. We define the convex functions f_i as follows

$$\begin{aligned} f_1(\zeta) &= \sum_{i \in \mathcal{S}} -(\log \mathbf{p}_i^T)\zeta, \\ f_2(\zeta) &= \iota_+(\zeta), \\ f_3(\zeta) &= \iota_1(\zeta), \end{aligned} \tag{4.10}$$

where ζ are dummy variables whose dimensions depend on functions f_j for $j = 1, \cdots, 3$. $\iota_+(\cdot)$ is the indicator function defined over the set $\zeta \in \mathbb{R}_+^{K \times n}$ with: $\iota_+(\zeta) = 0$ if $\zeta \in \mathbb{R}_+^{K \times n}$ and $\iota_+(\zeta) = +\infty$, otherwise. Likewise, $\iota_1(\zeta)$ is the indicator in set $\{\mathbf{1}_n\}$, with $\iota_1 = 0$ if and only if $\zeta \in \{\mathbf{1}_n\}$, and $+\infty$ otherwise.

By introducing in (4.9) variable splitting $\mathbf{u}_j^f = \mathbf{z}$, for $j = 1, \cdots, 3$ and $\mathbf{u}_j^g = \mathbf{H}_i^g \mathbf{z}$, for $j = 1, \cdots, C+2$, we convert the original optimization into the equivalent constrained form

$$\begin{aligned} &\min_{\mathbf{u}^f, \mathbf{u}^g, \mathbf{z}} \sum_{j=1}^{3} f_j(\mathbf{u}_j^f) + \sum_{j=1}^{C+2} g_j(\mathbf{u}_j^g) \\ &\text{s.t.} \quad \begin{bmatrix} \mathbf{u}^f \\ \mathbf{u}^g \end{bmatrix} = \begin{bmatrix} \mathbf{G}^f \\ \mathbf{G}^g \end{bmatrix} \mathbf{z}, \end{aligned} \tag{4.11}$$

where $\mathbf{u}_1^f, \mathbf{u}_2^f, \mathbf{u}_3^f \in \mathbb{R}^{K \times n}$, and the dimension of $\mathbf{u}_j^g$, for $j = 1, \cdots, C+2$, being dependent on the prior selection. In (4.11), we stack columnwise the identity operators $\boldsymbol{I}$ into the single operator $\mathbf{G}^f : \mathbb{R}^{K \times n} \to \mathbb{R}^{K \times 3n}$ and define $\mathbf{G}^g$ as a columnwise stacking of the operators $\mathbf{H}_j^g$ associated with the prior term.

4.4.1 SALSA Formulation

Taking the formulation (4.11), we apply the C-SALSA methodology [215]. We denote the scaled Lagrange multipliers associated with the constraints $\mathbf{u}^f = \mathbf{G}^f \mathbf{z}$ and $\mathbf{u}^g = \mathbf{G}^g \mathbf{z}$ as $\mathbf{d}^f = [\mathbf{d}_1^{fT}, \ldots, \mathbf{d}_3^{fT}]$ and $\mathbf{d}^g = [\mathbf{d}_1^{gT}, \ldots, \mathbf{d}_m^{gT}]$, respectively. We thus have the following C-SALSA based formulation for (4.11),

$$\mathbf{z}^{k+1} = \arg\min_{\mathbf{z}} \left\| \begin{bmatrix} \mathbf{G}^f \\ \mathbf{G}^g \end{bmatrix} \mathbf{z} - \begin{bmatrix} \mathbf{u}^{f,k} \\ \mathbf{u}^{g,k} \end{bmatrix} - \begin{bmatrix} \mathbf{d}^{f,k} \\ \mathbf{d}^{g,k} \end{bmatrix} \right\|_F^2, \tag{4.12}$$

$$\begin{aligned} \mathbf{u}_j^{f,k+1} &= \arg\min_{\mathbf{u}_j^f} f_j(\mathbf{u}_j^f) + \frac{\mu}{2} \|\mathbf{G}^f \mathbf{z}^{k+1} - \mathbf{u}_j^f - \mathbf{d}_j^{f,k}\|_F^2, \\ \mathbf{u}_j^{g,k+1} &= \arg\min_{\mathbf{u}_j^g} g_j(\mathbf{u}_j^g) + \frac{\mu}{2} \|\mathbf{G}^g \mathbf{z}^{k+1} - \mathbf{u}_j^g - \mathbf{d}_j^{g,k}\|_F^2, \end{aligned} \tag{4.13}$$

$$\begin{aligned} \mathbf{d}^{f,k+1} &= \mathbf{d}^{f,k} - \left[\mathbf{G}^f \mathbf{z}^{k+1} - \mathbf{u}^{f,k+1}\right], \\ \mathbf{d}^{g,k+1} &= \mathbf{d}^{g,k} - \left[\mathbf{G}^g \mathbf{z}^{k+1} - \mathbf{u}^{g,k+1}\right], \end{aligned} \tag{4.14}$$

where $\mu > 0$ is a parameter of the optimization controlling the variable splitting.

4.4.2 Optimization with respect to z

The solution of the quadratic form of (4.12) is [211]

$$\mathbf{z}^{k+1} = (\mathbf{G}^*\mathbf{G})^{-1}\mathbf{G}^*(\mathbf{u}^k + \mathbf{d}^k) = \mathbf{F}^{-1}\left(3\mathbf{I}(\mathbf{u}_j^{f,k} + \mathbf{d}_j^{f,k}) + \sum_{j=1}^{C+2}(\mathbf{H}_j^g)^*(\mathbf{u}_j^{g,k} + \mathbf{d}_j^{g,k})\right),$$

where

$$\mathbf{F} = 3\mathbf{I} + \sum_{j=1}^{C+2}(\mathbf{H}_j^g)^*\mathbf{H}_j^g,$$

and $(\cdot)^*$ represents the adjoint operator with respect to the Frobenius norm. Having in mind that the operator $\mathbf{G}$ is the columnwise stacking of the operators $\mathbf{G^f}$ and $\mathbf{G^g}$, if the linear operators $\mathbf{H_j^g}$ can be represented through cyclic convolution operators, solving (4.12) with respect to relaxation variable $\mathbf{z}$ can be implemented through cyclic convolution operations, thus diagonalizable in the frequency domain and consequently easily performed using the fast Fourier transform (FFT) with $O(Kn\ln n)$ complexity.

4.4.3 Optimization of the split variables with Moreau proximity operators (MPO)

The optimization subproblems associated with (4.13) can be solved through proximal methods, by computing the associated Moreau proximity operators (MPO) [224] of each of the convex functions. We first present the closed form expressions of these operators for the data fit term, and sum-to-one and nonnegativity constraints.

4.4.3.1 Moreau proximity operator for f_1

The Moreau proximity operator for the data fit f_1 is

$$\psi_{f_1/\mu}(\boldsymbol{\nu}) = \arg\min_{\boldsymbol{\zeta}}\left(\sum_{i\in\mathcal{S}}\boldsymbol{q}_i^T\boldsymbol{\zeta}_i\right) + \frac{\mu}{2}\|\boldsymbol{\zeta} - \boldsymbol{\nu}\|_F^2,$$

where $\boldsymbol{\nu} \equiv [\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_n] \in \mathbb{R}^{K\times n}$, and $\boldsymbol{\zeta} \equiv [\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_n] \in \mathbb{R}^{K\times n}$. This optimization is decoupled (pixelwise) with respect to $\boldsymbol{\zeta}_i$ for $i \in \mathcal{S}$, meaning

$$\psi_{f_1/\mu}(\boldsymbol{\nu}) = (\psi_{f_1/\mu}(\boldsymbol{\nu}_1), \ldots, \psi_{f_1/\mu}(\boldsymbol{\nu}_n)),$$

such that

$$\psi_{f_1/\mu}(\boldsymbol{\nu}_i) = \arg\min_{\boldsymbol{\zeta_i}}\ \boldsymbol{q}_i^T\boldsymbol{\zeta}_i + \frac{\mu}{2}\|\boldsymbol{\zeta}_i - \boldsymbol{\nu}_i\|_F^2.$$

We find the proximal operator by locating the vertex of the paraboloid $\nabla\psi_{f_1/\mu}/\nabla\boldsymbol{\zeta}_i = 0$ with respect to $\boldsymbol{\zeta}_i$, leading to

$$\psi_{f_1/\mu}(\boldsymbol{\nu}_i) = \boldsymbol{\nu}_i + (\log \mathbf{q}_i)/\mu.$$

This operator has a complexity of $O(Kn)$, the number of classes across the number of pixels.

4.4.3.2 Moreau proximity operator for f_2

The Moreau proximity operator for the sum-to-one constraint f_2 is

$$\psi_{f_2/\mu}(\boldsymbol{\nu}) = \arg\min_{\boldsymbol{\zeta}} \iota_1(\boldsymbol{\zeta}) + \frac{\mu}{2}\|\boldsymbol{\zeta} - \boldsymbol{\nu}\|_F^2 = \left(\mathbf{I} - \frac{\mathbf{1}_K\mathbf{1}_K^T}{K}\right)\nu + \frac{\mathbf{1}_K\mathbf{1}_n^T}{K},$$

where $\nu, \boldsymbol{\zeta} \in \mathbb{R}^{K\times n}$. This operator is the projection in the probability simplex, and has a complexity of $O(Kn)$, the number of classes times the number of pixels.

4.4.3.3 Moreau proximity operator for f_3

The Moreau proximity operator for nonnegativity constraint f_3 is

$$\psi_{f_3/\mu}(\boldsymbol{\nu}) = \arg\min_{\boldsymbol{\zeta}} \iota_+(\boldsymbol{\zeta}) + \frac{\mu}{2}\|\boldsymbol{\zeta} - \boldsymbol{\nu}\|_F^2 = \max\{\mathbf{0}, \boldsymbol{\nu}\},$$

where $\boldsymbol{\nu}, \boldsymbol{\zeta} \in \mathbb{R}^{K\times n}$. This operator is the projection in the first orthant, and has a complexity of $O(Kn)$, the number of classes times the number of pixels.

4.4.4 Optimization of Vectorial Total Variation

The inclusion of the VTV prior in (4.8) introduces the term $\lambda_1 \sum_{i\in\mathcal{S}} \omega_i \sqrt{||(\mathbf{D}_h\mathbf{z})_i||^2 + ||(\mathbf{D}_v\mathbf{z})_i||^2}$ which means that our prior has a single term, $m = 1$.

We define the linear operator $\mathbf{H}^g : \mathbb{R}^{K\times n} \to \mathbb{R}^{2K\times n}$ as

$$\mathbf{H}^g = \begin{pmatrix} \mathbf{D}_h \\ \mathbf{D}_v \end{pmatrix}, \tag{4.15}$$

where $\mathbf{D}_h$ and $\mathbf{D}_v$ correspond to the circular horizontal difference operators previously defined. The corresponding convex function g is defined as

$$g(\boldsymbol{\zeta}) = \lambda_1 \sum_{i\in\mathcal{S}} \omega_i \sqrt{||\boldsymbol{\zeta}_i^h||^2 + ||\boldsymbol{\zeta}_i^v||^2}, \tag{4.16}$$

where $\boldsymbol{\zeta} = [\boldsymbol{\zeta}^h \ \boldsymbol{\zeta}^v] \in \mathbb{R}^{2K\times n}$, and $\boldsymbol{\zeta}^h$ and $\boldsymbol{\zeta}^v$ belong to the range of the horizontal and vertical difference operators $\mathbf{D}_h$ and $\mathbf{D}_v$, respectively.

The Moreau proximity operator for the VTV prior is thus

$$\psi_{g/\mu}(\boldsymbol{\nu}) = \arg\min_{\boldsymbol{\zeta}} \left(\sum_{i\in\mathcal{S}} \omega_i \sqrt{||\boldsymbol{\zeta}_i^h||^2 + ||\boldsymbol{\zeta}_i^v||^2} \right) + \frac{\mu}{2\lambda_1}||\boldsymbol{\zeta} - \boldsymbol{\nu}||_F^2, \tag{4.17}$$

where $\boldsymbol{\nu}, \boldsymbol{\zeta} \in \mathbb{R}^{2K\times n}$ and $\boldsymbol{\zeta}^h, \boldsymbol{\zeta}^v \in \mathbb{R}^{K\times n}$. This optimization can be pixelwise decoupled and can be solved with the vector soft thresholding operator [224]

$$\psi_{g/\mu}(\boldsymbol{\nu}_i) = \max\{0, ||\boldsymbol{\nu}_i|| - \lambda_1\omega_i/\mu\}\frac{\boldsymbol{\nu}_i}{||\boldsymbol{\nu}_i||}, \tag{4.18}$$

which is of complexity of $O(Kn)$. If the linear operators $\boldsymbol{H}_j^g$ can be represented through cyclic convolution operators, solving (4.12) with respect to the hidden field $\mathbf{z}$ can be implemented through cyclic convolution operations, thus diagonalizable in the frequency domain and consequently easily performed in the frequency domain [225].

4.4.5 Optimization of Graph Total Variation

As introduced in (4.8), the Moreau proximity operator of the j-th graph (superpixelization), for $j \in \{1, \cdots, C\}$, is

$$\psi_{g_j/\mu}(\boldsymbol{\nu}) = \arg\min_{\boldsymbol{\zeta}} \lambda_2 \omega_j ||(\mathbf{A}_j - \mathbf{I})\boldsymbol{\zeta}||_F^2 + \frac{\mu}{2}||\boldsymbol{\zeta} - \boldsymbol{\nu}||_F^2, \tag{4.19}$$

where $\boldsymbol{\nu}, \boldsymbol{\zeta} \in \mathbb{R}^{K \times n}$, which is a quadratic optimization problem. The Moreau proximity operator is thus

$$\psi_{g_j/\mu}(\boldsymbol{\nu}) = \frac{\mu}{2\lambda_2\omega_j}\Big((\mathbf{A}_j - \mathbf{I})^T(\mathbf{A}_j - \mathbf{I}) + \frac{\mu}{2\lambda_2\omega_j}\mathbf{I}\Big)^{-1}\boldsymbol{\nu}. \tag{4.20}$$

To easily represent the matrix to calculate the inverse, let

$$\mathbf{F^{-1}} = \frac{\mu}{2\lambda_2\omega_j}\Big((\mathbf{A}_j - \mathbf{I})^T(\mathbf{A}_j - \mathbf{I}) + \frac{\mu}{2\lambda_2\omega_j}\mathbf{I}\Big)^{-1},$$

and

$$[\mathbf{A}_{j,t} - \mathbf{I}] = \begin{bmatrix} \frac{1}{n_t} - 1 & \frac{1}{n_t} & \cdots & \frac{1}{n_t} \\ \frac{1}{n_t} & \frac{1}{n_t} - 1 & \cdots & \frac{1}{n_t} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n_t} & \frac{1}{n_t} & \cdots & \frac{1}{n_t} - 1 \end{bmatrix},$$

$$[(\mathbf{A}_{j,t} - \mathbf{I})^T(\mathbf{A}_{j,t} - \mathbf{I})] = \begin{bmatrix} \frac{n_t - 1}{n_t} & -\frac{1}{n_t} & \cdots & -\frac{1}{n_t} \\ -\frac{1}{n_t} & \frac{n_t - 1}{n_t} & \cdots & -\frac{1}{n_t} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n_t} & -\frac{1}{n_t} & \cdots & \frac{{}_tn - 1}{n_t} \end{bmatrix}$$
$$= -\frac{1}{n_t}\mathbf{E} + \mathbf{I}.$$

where $\mathbf{A}_{j,t}$ are diagonal blocks of $\mathbf{A}_j$, corresponding to a fully connected subgraph of the partition graph with n_t ($n_t = n_{j,t}$ to keep notation light) nodes, a partition element with n_t pixels, and is a $n_t \times n_t$ matrix and $\mathbf{E} = \{1\}$ denoting a $n_t \times n_t$ square matrix . Note that $\mathbf{A}_{j,t}$ is normalized by the pixel number of its corresponding superpixel and that $\mathbf{E} = \mathbf{v}\mathbf{v}^T$, with $\mathbf{v}$ being a column vector of 1's. Trivially, letting $\theta_j = \frac{\mu}{2\lambda_2\omega_j}$, there comes

$$\begin{aligned} \mathbf{F}^{-1} &= \theta_j\Big((\mathbf{A}_j - \mathbf{I})^T(\mathbf{A}_j - \mathbf{I}) + \theta_j\mathbf{I}\Big)^{-1} \\ &= \theta_j\Big((-\frac{1}{n_t}\mathbf{E} + \mathbf{I}) + \theta_j\mathbf{I}\Big)^{-1} \\ &= -\theta n_t\Big((-1 - \theta_j)n_t\mathbf{I} + \mathbf{v}\mathbf{v}^T\Big)^{-1}. \end{aligned}$$

Since $\exists\ ((-1-\theta_j)n_t\mathbf{I})^{-1}$ and $1 + \mathbf{v}^T((-1-\theta_j)n_t\mathbf{I})^{-1}\mathbf{v} = \frac{\theta_j}{1+\theta_j} \neq 0$ ($\theta_j \neq 0$), according to the Sherman Morrison Formula [1], trivially we can acquire $\mathbf{F}$ without the inverting

[1] https://en.wikipedia.org/wiki/Sherman%E2%80%93Morrison_formula

calculation

$$\begin{aligned}\mathbf{F} &= \Big(\frac{\theta_j}{1+\theta_j}\mathbf{I} + \frac{1}{(1+\theta_j)n_t}\mathbf{E}\Big),\ (\theta_j = \frac{\mu}{2\lambda_2\omega_j}) \\ &= \Big(\frac{\mu}{\mu+2\lambda_2\omega_j}\mathbf{I} + \frac{2\lambda_2\omega_j}{(\mu+2\lambda_2\omega_j)n_t}\mathbf{E}\Big).\end{aligned} \tag{4.21}$$

Hence, the Moreau Proximity Operator of the GTV is

$$\psi_{g_j/\mu}(\boldsymbol{\nu}) = \Big(\frac{\mu}{\mu+2\lambda_2\omega_j}\mathbf{I} + \frac{2\lambda_2\omega_j}{(\mu+2\lambda_2\omega_j)n_t}\mathbf{E}\Big)\boldsymbol{\nu}, \tag{4.22}$$

which can be pixelwise decoupled as

$$\psi_{g_j/\mu}(\boldsymbol{\nu}_i) = \frac{\mu}{\mu+2\lambda_2\omega_j}\boldsymbol{\nu}_i + \frac{2\lambda_2\omega_j}{(\mu+2\lambda_2\omega_j)n_t}\sum_{k=1}^{n_t}\boldsymbol{\nu}_k, \tag{4.23}$$

where $i, k \in \mathcal{S} \equiv \{1, \cdots, n\}$ index the pixels that belong to the same t-th partition as the i-th, k-th pixel. To be specific, the first term of (4.23) corresponds to the value on the i-th node itself, the value of the i-th pixel, and the second term corresponds to the mean of $\boldsymbol{\nu}$ on the fully connected subgraph that the i-th node belongs to, such that this operator has a complexity of $O(CKn)$.

4.4.6 SegSALSA-VTV-GTV

Algorithm 1 SegSALSA-VTV-GTV

initialization:
choose $(\boldsymbol{u}_j^{f,0}, \boldsymbol{d}_j^{f,0}) \in \mathbb{R}^{n\times K}$, $j = 1, \ldots, 3$
choose $(\boldsymbol{u}_j^{g,0}, \boldsymbol{d}_j^{g,0})$, $j = 1, \ldots, C+2$
define $\boldsymbol{F} = 3\mathbf{I} + \sum_{j=1}^{m} \boldsymbol{H}_j^{g*}\boldsymbol{H}_j^{g}$
set $\mu \in]0, +\infty[$
for $k = 0, 1, \ldots, k_{STOP}$ **do**
  $\boldsymbol{z}^{k+1} = \boldsymbol{F}^{-1}\big(\sum_{j=1}^{3}(\boldsymbol{u}_j^{f,k} + \mathbf{d}_j^{f,k}) + \sum_{j=1}^{C+2}(\boldsymbol{H}_j^{g})^*(\mathbf{u}_j^{g,k} + \mathbf{d}_j^{g,k})\big)$
  for $j = 1$ to 3 **do**
    $\boldsymbol{u}_j^{f,k+1} = \text{prox}_{f_j/\mu}(\boldsymbol{z}^{k+1} - \boldsymbol{d}_j^{f,k})$
    $\boldsymbol{d}_j^{f,k+1} = \boldsymbol{d}_j^{f,k} - (\boldsymbol{z}^{k+1} - \boldsymbol{u}_j^{f,k+1})$
  end for
  for $j = 1$ to $C+2$ **do**
    $\boldsymbol{u}_j^{g,k+1} = \text{prox}_{g_j/\mu}(\boldsymbol{z}^{k+1} - \boldsymbol{d}_j^{g,k})$
    $\boldsymbol{d}_j^{g,k+1} = \boldsymbol{d}_j^{g,k} - (\boldsymbol{z}^{k+1} - \boldsymbol{u}_j^{g,k+1})$
  end for
end for
return $\boldsymbol{z}^{k+1}$

Algorithm 1 displays the pseudocode for our proposed instance of the SALSA algorithm, termed Segmentation via Augmented Lagrangian Shrinkage Algorithm associated with vectorial total variation and graph total variation (SegSALSA-VTV-GTV). SegSALSA-VTV-GTV converges for any $\mu > 0$, while the convergence speed is highly dependent on the value of μ. Note that the computational complexities with

respect to $\mathbf{x}$ and MPOs for f_j, $j = 1, \cdots, 3$, and g_j, $j = 1, \cdots, C+2$, together lead to the conclusion that the computational complexity of SegSALSA-VTV-GTV per iteration is dominated by the term $O(Kn \ln n)$, associated to the computation of $\mathbf{z}^{k+1}$. Meanwhile, the algorithm can be straightforwardly parallelized for high performance computation via GPU and other architectures, due to the fact that optimizing of the algorithm actually decouples pixelwisely each term of (4.8). Also note that the parameter μ only influences the convergence speed of the algorithm, only if the algorithm converges. The parameters λ_1 and $\lambda_2 \times \omega_j$, for $j \in \{1, \cdots, C\}$ can be tuned to control the importance of the corresponding VTV and GTV regularizers. For simplicity purposes, we directly use one value for each production of $\lambda_2 \times \omega_j$, for $j \in \{1, \cdots, C\}$, since they are actually the values of λ_2 times of the weights ω_j.

4.5 Experiments

In this section, we evaluate the proposed method with remote sensing data sets that are obtained by different types of sensors, namely, hyperspectral images (HSI) and multispectral images (MSI). Before reporting our experimental results, we introduce our newly designed framework in Fig. 4.1.

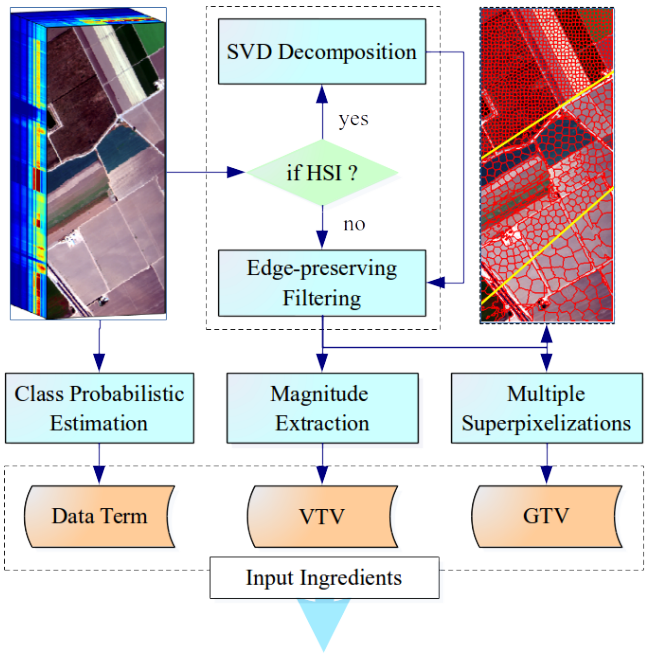
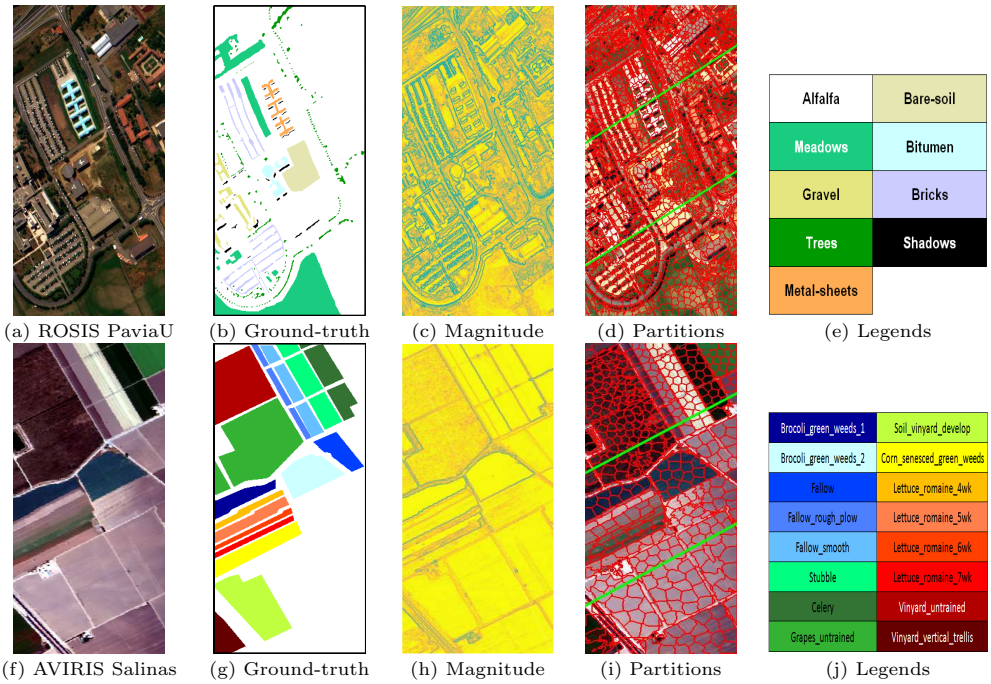


Figure 4.1: Experimental framework for our proposed method with respect to hyperspectral and multispectral remote sensing image data sets.

As shown in (4.8), there are three main ingredients in our method: the data term, the magnitude map for VTV, and the multiple over-segmented superpixelizations for GTV. First of all, a classifier multinomial logistic regression -MLR in our case, whose

regressors are learned by the LORSAL algorithm [175, 226]), is exploited to estimate the class probabilities of the image, in preparation for the data term. Before obtaining the remaining two terms (VTV and GTV), we perform the SVD for dealing with the high spectral dimensionality of the hyperspectral images. Then, in order to weaken the trivial textural details of the image, the anisotropic edge preserving filtering [227] is then utilized to the main transformed components and to the multispectral image. Finally, we extract the magnitude map with the Sobel operator and generate multiple over-segmented superpixelizations with the fast spatial clustering algorithm SLIC [214], available in the VLFeat toolbox [2], with varying parameters.

4.5.1 Experimental results with hyperspectral images (HSIs)



(a) ROSIS PaviaU (b) Ground-truth (c) Magnitude (d) Partitions (e) Legends

(f) AVIRIS Salinas (g) Ground-truth (h) Magnitude (i) Partitions (j) Legends

Figure 4.2: Experiments with hyperspectral data. First row lists the ROSIS Pavia University dataset: (a) HSI RGB composite, (b) ground-truth reference, (c) magnitude map, (d) multiple over-segmented partitions/superpixels, (e) class legends. The second row shows the results (f-j) of the AVIRIS Salinas dataset.

In this subsection, we evaluate our proposed algorithm with the ROSIS Pavia University data set as well as the AVIRIS Salinas data set. The first hyperspectral image used in our experiments [see Fig. 4.2 (a)] was collected by ROSIS over the University of Pavia, Italy. The data set consists of 115 spectral bands which covering the wavelength range from 0.43 to 0.86 μm, with size of 610×340 pixels. The noisy bands had been removed, yielding 103 spectral bands that are actually used in this chapter. The ground-truth image contains 9 ground-truth classes, including 42776 labeled samples. The second

[2] http://www.vlfeat.org/doc/api/slic.html

Table 4.1: Overall, average and individual class accuracies [%] and κ statistic obtained by the presented classification framework implemented using the MLR classifier in comparison with the state-of-the-art methods, majority voting (MVS), Graphcut and discontinuity preserving relaxation (DPR). In particular, we set the parameter of the proposed method $\lambda_1 = 5$, $\lambda_2 \times c_j = 2$, for $j = \{1, \cdots, 3\}$. The averages are and corresponding standard deviations are calculated under 20 monte carlo runs. In all cases, only 30 randomly selected training samples per class have been used for the ROSIS Pavia University data set.

Class	MLR	MVs	Graphcut	DPR	Proposed
Alfalfa	74.22 ± 3.15	87.03 ± 3.05	88.85 ± 4.34	**91.64 ± 3.60**	90.67 ± 4.35
Meadows	73.41 ± 3.96	84.47 ± 5.45	84.10 ± 5.92	85.03 ± 5.29	**86.66 ± 6.38**
Gravel	75.35 ± 5.05	93.09 ± 6.57	89.47 ± 7.06	88.67 ± 5.80	**95.76 ± 5.35**
Trees	91.99 ± 2.56	68.89 ± 4.88	**93.33 ± 2.83**	91.02 ± 2.89	92.22 ± 3.87
Metal sheets	99.82 ± 0.12	97.88 ± 2.45	**100 ± 0.02**	100 ± 0.00	99.93 ± 0.02
Bare soil	78.11 ± 4.29	95.18 ± 3.28	96.39 ± 4.60	98.16 ± 3.31	**98.69 ± 2.95**
Bitumen	89.34 ± 2.90	99.46 ± 0.00	97.63 ± 1.28	**99.92 ± 0.07**	99.28 ± 0.36
Bricks	73.69 ± 4.33	**90.21 ± 7.05**	78.23 ± 7.51	88.11 ± 5.98	79.92 ± 11.18
Shadows	**99.73 ± 0.34**	88.71 ± 3.29	99.69 ± 0.76	99.48 ± 0.04	99.16 ± 0.44
Overall accuracy	77.41 ± 1.78	86.89 ± 2.38	87.94 ± 2.53	89.70 ± 2.35	**90.02 ± 2.96**
Average accuracy	83.96 ± 0.88	89.43 ± 1.15	91.97 ± 1.28	93.56 ± 1.12	**93.59 ± 1.41**
κ statistic	71.51 ± 2.05	83.07 ± 2.90	84.53 ± 3.11	86.75 ± 2.92	**87.17 ± 3.65**

hyperspectral image used in our experiment is the well-known AVIRIS Salinas[3] data set, collected by the AVIRIS sensor of 224 bands in the wavelength range from 0.38 to 2.50 μm over Salinas Valley, California. As displayed in Fig. 4.2(f), it comprises 512 lines by 217 samples across 204 spectral bands after discarding 20 water absorption bands. For the reference collection, a total of 54129 pixels are available in the labeled ground-truth, including 16 mutually exclusive classes. For both hyperspectral data sets, we also prepared the magnitude map (Fig. 4.2 (c,h)) and three over-segmented superpixelizations (Fig. 4.2 (d,i)), respectively.

Before displaying the experimental results, we first introduce our experimental setup for the analysis of hyperspectral images. The class probabilities are estimated by an MLR, where the logistic regressors (assumed to be random vectors with independent Laplacian components) are learnt using the LORSAL algorithm [175, 226]. The MLR classifiers are learned with 30 training samples and 15 samples per class, respectively, for the ROSIS Pavia University data and the AVIRIS Salinas data. For simplicity, we use three over-segmented superpixelization representations (with SLIC algorithm) to construct the graph total variation. In order to obtain convergence in our proposed method, we empirically set 200 iterations. Meanwhile, some recently developed state-of-the-art methods, namely, majority voting approaches [85], graph-cut [84] algorithm and the discontinuity preserving relation scheme [189], are also employed in this chapter for evaluation and comparison purposes. Also, for statistical purposes, 20 Monte Carlo runs are performed for all the compared methods.

The obtained results are displayed in Tables 4.1, 4.2. Several observations can be made from these results. First, all the segmentation results obtain remarkable

[3]http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes#Salinas_scene

Table 4.2: Overall, average and individual class accuracies [%] and κ statistic obtained by the presented classification framework implemented using the MLR classifier in comparison with the state-of-the-art methods, majority voting (MVS), Graphcut and discontinuity preserving relaxation (DPR). In particular, we set the parameter of the proposed method $\lambda_1 = 5$, $\lambda_2 \times c_j = 5$, for $j = \{1, \cdots, 3\}$. The averages are and corresponding standard deviations are calculated under 20 monte carlo runs. In all cases, only 15 randomly selected training samples per class have been used for the AVIRIS Salinas data set.

Class	MLR	MVs	Graphcut	DPR	Proposed
Brocoli-green-weeds-1	99.12 ± 0.49	**100 ± 0.00**	99.97 ± 0.12	**100 ± 0.00**	**100 ± 0.00**
Brocoli-green-weeds-2	99.86 ± 0.21	99.92 ± 0.02	**100 ± 0.00**	**100 ± 0.00**	**100 ± 0.00**
Fallow	93.33 ± 4.42	95.28 ± 8.97	**99.96 ± 0.10**	99.27 ± 2.08	99.84 ± 0.15
Fallow-rough-plow	99.66 ± 0.15	98.98 ± 0.00	99.75 ± 0.09	**100 ± 0.00**	99.85 ± 0.00
Fallow-smooth	97.41 ± 0.81	86.34 ± 5.49	98.68 ± 0.32	**98.80 ± 0.34**	94.18 ± 4.73
Stubble	99.22 ± 0.23	99.92 ± 0.00	99.73 ± 0.17	**99.97 ± 0.01**	99.92 ± 0.00
Celery	99.88 ± 0.06	99.29 ± 0.08	99.91 ± 0.03	**100 ± 0.00**	99.92 ± 0.01
Grapes-untrained	67.57 ± 9.79	81.02 ± 13.41	78.88 ± 13.12	78.70 ± 11.84	**82.51 ± 18.94**
Soil-vinyard-develop	98.45 ± 0.73	98.66 ± 1.02	98.93 ± 0.82	**99.55 ± 0.55**	99.44 ± 0.58
Corn-senesced-green-weeds	85.74 ± 4.43	90.51 ± 4.58	91.45 ± 3.57	93.20 ± 3.15	**93.62 ± 2.03**
Lettuce-romaine-4wk	91.68 ± 2.89	96.81 ± 0.07	96.42 ± 2.77	98.74 ± 0.72	96.23 ± 1.32
Lettuce-romaine-5wk	99.61 ± 0.57	**100 ± 0.00**	**100 ± 0.00**	**100 ± 0.00**	**100 ± 0.00**
Lettuce-romaine-6wk	89.14 ± 4.81	71.20 ± 10.64	**95.16 ± 4.00**	93.31 ± 4.37	90.58 ± 8.34
Lettuce-romaine-7wk	92.29 ± 0.92	90.60 ± 6.34	97.95 ± 1.49	**99.70 ± 0.15**	97.27 ± 0.52
Vinyard-untrained	66.14 ± 8.10	78.94 ± 16.07	79.12 ± 13.37	78.52 ± 11.13	**84.34 ± 15.33**
Vinyard-vertical-trellis	96.11 ± 2.62	98.12 ± 0.04	**99.32 ± 0.74**	98.90 ± 0.91	98.43 ± 0.84
Overall accuracy	86.50 ± 1.43	90.75 ± 2.17	91.82 ± 1.96	91.93 ± 1.65	**93.14 ± 3.19**
Average accuracy	92.20 ± 0.48	92.85 ± 0.83	95.95 ± 0.65	**96.17 ± 0.54**	96.01 ± 1.03
κ statistic	85.02 ± 1.56	89.72 ± 2.41	90.92 ± 2.16	91.03 ± 1.81	**92.39 ± 3.50**

improvements compared with the fundamental MLR classifier, particularly after the inclusion of spatial information. Meanwhile, the proposed algorithm also outperforms the compared methods with respect to the overall accuracy for both HSI datasets. For illustrative purposes, Fig. 4.3 also displays the corresponding classification maps that are obtained from one of the 20 Monte Carlo runs. Remarkably, our obtained segmentation maps show stronger pixel consistency while keeping more precise contours for the land objects, which are exactly what this chapter explores via formulating the VTV and superpixel-related GTV regularizers. Specifically, when compared with majority voting and graph-cut methods, our proposed method (with VTV) preserves better the edge/boundary information, since majority voting does not consider the boundary discontinuity and the graph-cut does not consider the oblique (only the vertical and horizontal) discontinuities. On the other hand, the map acquired by the discontinuity preserving relaxation method misses some small scale details due to the fact that the DPR method relies greatly on the quality of edge extraction, which is a challenging task in HSI processing. Bearing these observations in mind, our proposed method turns out to be the-state-of-the-art.

The regularization parameters play a significant role in adjusting the performance of the whole machinery. In order to illustrate the effect of the two spatial regularizers, we display additional segmentation maps obtained by using different values of the regularizer

Table 4.3: Overall, average and individual class accuracies [%] and κ statistic obtained by the presented classification framework implemented using the MLR classifier in comparison with the state-of-the-art methods, majority voting (MVS), Graphcut and discontinuity preserving relaxation (DPR). In particular, we set the parameter of the proposed method $\lambda_1 = 4$, $\lambda_2 \times c_j = 2$, for $j = \{1, \cdots, 3\}$. The averages are and corresponding standard deviations are calculated under 20 monte carlo runs. In all cases, 200 randomly selected training samples per class have been used for the Zurich3 data set.

Class	MLR	MVs	Graphcut	DPR	Proposed
Roads	77.54 ± 1.36	**84.43 ± 1.24**	83.11 ± 1.29	79.01 ± 1.36	70.59 ± 4.78
Buildings	52.54 ± 3.51	50.03 ± 5.69	57.15 ± 4.75	60.39 ± 4.17	**72.50 ± 5.46**
Trees	84.96 ± 0.90	91.20 ± 0.81	88.83 ± 1.09	90.51 ± 1.10	**94.75 ± 1.02**
Grass	91.06 ± 1.19	94.67 ± 0.62	93.03 ± 1.17	93.09 ± 1.14	**95.43 ± 1.12**
Bare-Soil	95.02 ± 0.76	95.22 ± 0.96	95.70 ± 0.57	**95.80 ± 0.80**	92.87 ± 0.75
Water	99.56 ± 0.08	99.77 ± 0.01	99.61 ± 0.08	**99.84 ± 0.06**	99.78 ± 0.03
Railways	56.60 ± 1.91	69.77 ± 2.29	73.31 ± 2.65	78.79 ± 1.89	**94.32 ± 1.28**
Overall accuracy	87.87 ± 0.19	90.39 ± 0.48	90.62 ± 0.36	91.16 ± 0.34	**92.98 ± 0.50**
Average accuracy	79.61 ± 0.23	83.58 ± 0.78	84.39 ± 0.60	85.35 ± 0.55	**88.61 ± 0.80**
κ statistic	81.99 ± 0.27	85.71 ± 0.71	86.08 ± 0.53	86.87 ± 0.50	**89.58 ± 0.74**

parameters λ_1 and $\lambda_2 \times c_j$, for $j = \{1, \cdots, 3\}$. For simplicity, we set the same parametric values for three over-segmented superpixelizations. The obtained results of both HSI datasets are shown in Fig. 4.3, displaying a clue also on how to tune the values of the parameters by hand. Specifically, this means that only the superpixel-based regularizer (GTV) is utilized when setting $\lambda_1 = 0$. And, likewise, when $\lambda_2 \times c_j = 0$, only the VTV regularizer is considered. It is obviously observed that different scales of contours of the land-objects can be controlled by tuning the parameters. For both regularizers, small values of the parameters lead to more details of the land objects, especially the small ones while greater values tend to keep the main contours of the objects. As for the VTV and GTV regularizers, when both of them are considered, the performance improves greatly in comparison with the individual use of one regularizer, which is consistent with what we have anticipated, i.e., involving the VTV to both promote piecewise smoothness and align the discontinuities along the boundaries while incorporating GTV to reinforce the label consistency over the over-segmented superpixels.

4.5.2 Experimental results with high spatial resolution images (VHR)

In this section, we evaluate our proposed framework with two remotely sensed multispectral images (MSI). The "Zurich Summer v1.0" dataset [4] is a collection of 20 chips (crops), taken from a QuickBird acquisition of the city of Zurich (Switzerland) in August 2002. The QuickBird images are composed by 4 channels (NIR-R-G-B) and were pansharpened to the PAN resolution of 0.62 meters / pixel. In this collection, 8 different urban and periurban classes were manually annotated: Roads, Buildings, Trees, Grass, Bare Soil, Water, Railways and Swimming pools [228]. The cumulative number of

[4] https://sites.google.com/site/michelevolpiresearch/data/zurich-dataset

Table 4.4: Overall, average and individual class accuracies [%] and κ statistic obtained by the presented classification framework implemented using the MLR classifier in comparison with the state-of-the-art methods, majority voting (MVS), Graphcut and discontinuity preserving relaxation (DPR). In particular, we set the parameter of the proposed method $\lambda_1 = 4$, $\lambda_2 \times c_j = 2$, for $j = \{1, \cdots, 3\}$. The averages are and corresponding standard deviations are calculated under 20 monte carlo runs. In all cases, 200 randomly selected training samples per class have been used for the Zurich6 data set.

Class	MLR	MVs	Graphcut	DPR	Proposed
Roads	73.28 ± 2.01	82.03 ± 1.28	82.03 ± 1.89	83.12 ± 1.43	**84.20 ± 1.77**
Buildings	64.92 ± 2.61	63.10 ± 4.91	65.19 ± 4.10	66.99 ± 2.96	**67.44 ± 4.15**
Trees	76.60 ± 1.67	84.87 ± 2.23	83.53 ± 1.31	86.17 ± 1.52	**93.03 ± 1.91**
Grass	88.14 ± 1.22	**92.10 ± 1.01**	90.47 ± 1.18	90.66 ± 1.37	91.38 ± 2.24
Bare-Soil	72.49 ± 2.08	93.89 ± 3.68	83.42 ± 1.04	83.85 ± 1.96	**94.61 ± 3.68**
Water	84.61 ± 1.28	**100 ± 0.00**	98.10 ± 2.02	93.49 ± 1.80	99.95 ± 0.15
Swimming-Pools	92.78 ± 0.83	**97.55 ± 0.29**	93.39 ± 0.75	94.91 ± 0.92	96.22 ± 3.40
Overall accuracy	74.01 ± 0.54	78.98 ± 1.25	78.93 ± 0.89	80.36 ± 0.71	**82.36 ± 0.91**
Average accuracy	78.97 ± 0.31	87.65 ± 0.93	85.16 ± 0.51	85.60 ± 0.46	**89.55 ± 0.63**
κ statistic	65.91 ± 0.62	72.10 ± 1.63	72.05 ± 1.18	73.82 ± 0.92	**76.32 ± 1.21**

class samples is highly unbalanced, to reflect real world situations. And the purpose of distributing datasets is to encourage reproducibility of experiments. In this chapter, we employ the 3rd (926 × 943 pixels) and 6th (812 × 984 pixels) images, as an example, to test our proposed method. As illustrated in the framework (see Fig. 4.1), the QuickBird MSI images and their corresponding prepared ingredients are shown in Fig. 4.4. We extracted the magnitude maps for both used image and transformed the images into three over-segmented superpixel representations by the SLIC method (via the VLFeat toolbox [5]). The result of our proposed method is obtained after 200 iterations with convergence, while 20 Monte Carlo runs are employed for all the considered methods.

To start with, the accuracy statistics of 20 Monte Carlo runs are shown in Tables 4.3, 4.4 with the mean plus/minus the standard deviation reported in the tables. From both tables, we can see that the accuracies are remarkably increased after the spatial information is incorporated by different methods. In comparison with state-of-the-art methods, namely, MVs, graph-cut and discontinuity preserving relaxation, our proposed method achieves the highest values in terms of overall accuracy, κ statistic and average accuracy, which is consistent with our results with hyperspectral images. For illustrative purposes, we also display the classification maps that are obtained with the methods in Fig. 4.5. First of all, it is remarkable that our proposed method obtain classification maps in which both the label consistency and boundary discontinuity are well refined, in comparison with state-of-the-art methods. Also, the hand-tuned parameters of the VTV and GTV spatial regularizers provide an intuitive description of their effects on the classification performance of our proposed machinery (i.e., small values of the parameters preserve more trivial details of the land objects while greater values keep the basic contours or strong boundaries of large-scale land objects). This is quite consistent with

[5] http://www.vlfeat.org/doc/api/slic.html

what have been observed with the experiments on HSIs in Section 4.5.1 and meanwhile is what we originally explore in this chapter.

4.6 Summary and future directions

This chapter proposed a new method, SegSALSA-VTV-GTV, which provides a convex formulation to exploit the spatial information coming from multiple over-segmented superpixelizations into supervised hyperspectral image segmentation. With this method, we sidestep the NP-hardness of the original discrete integer optimization problem by using a relaxation technique based on linear programming, and successfully express the over-segmented superpixels in the form of a graphical spatial regularizer across the relaxed hidden field. In addition, we formulate a convex optimization problem and approximate the solution for the original NP-hard image labeling problem. Specifically, we design a framework to substantiate and validate the method. The experimental results, obtained with remotely sensed hyperspectral and multispectral images, demonstrate that our proposed method achieves state-of-the-art performance when compared with other methods, such as majority voting, graph-cut and discontinuity preserving relaxation techniques. Our proposed approach can also be viewed as a general framework for solving a range of similar problems, such as change detection, regression, etc. Also the proposed approach is highly parallele and pixelwise decoupled, thus it can be straightforwardly implemented in parallel using high performance computing architectures. Our future perspectives will focus on exploring and evaluating the potential of the proposed approach when dealing with remote sensing data coming from multiple sources.

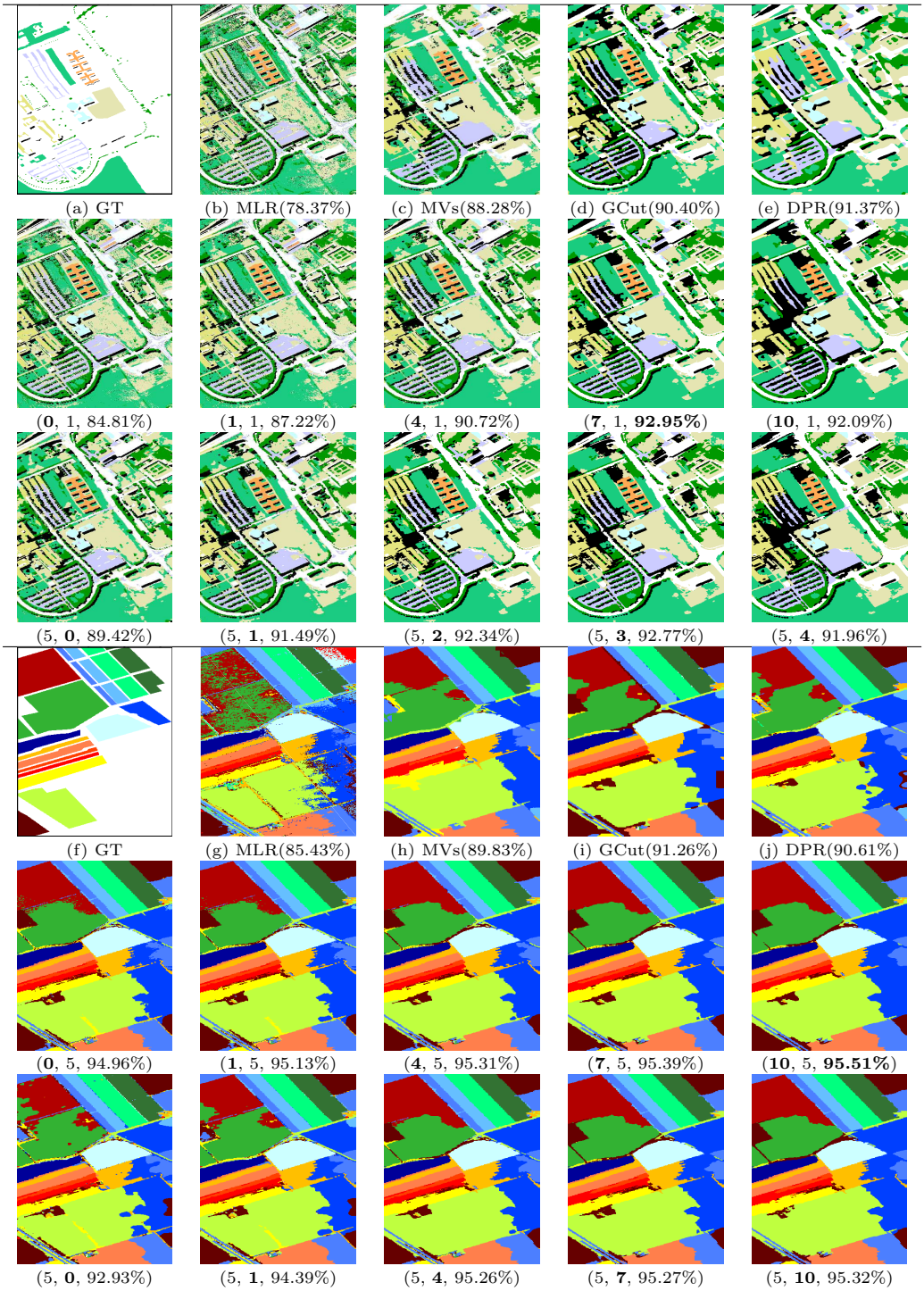
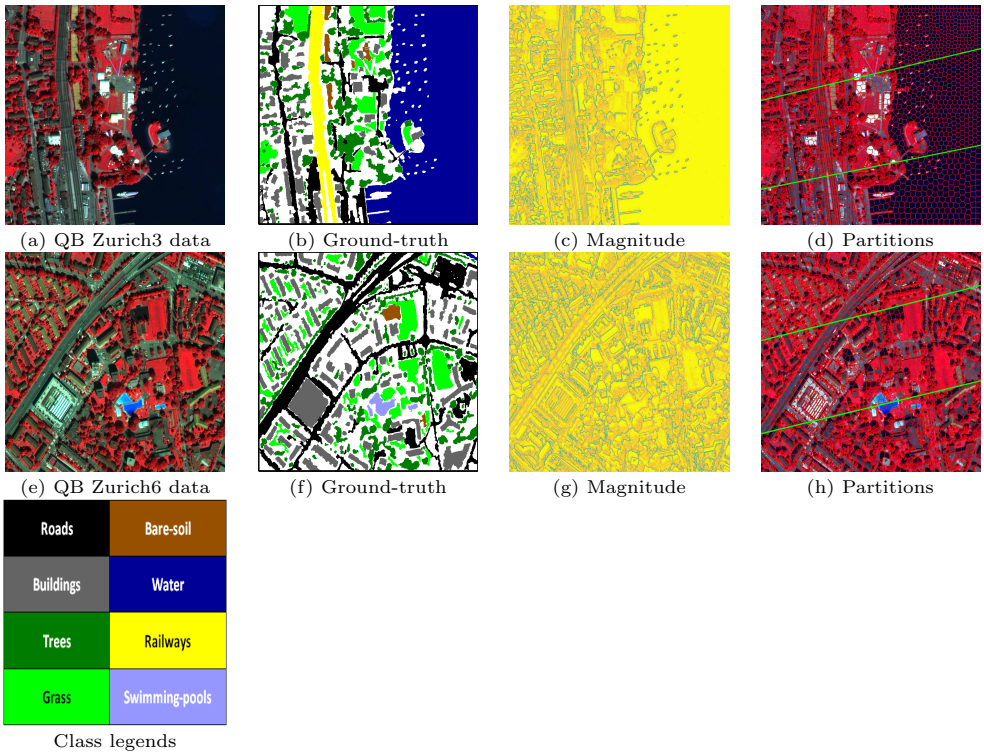


Figure 4.3: Classification maps of two HSI datasets: ROSIS Pavia University and AVIRIS Salinas. Top row (the ROSIS Pavia University dataset) shows: (a) ground-truth, (b) MLR classification, (c) majovity voting, (d) graph-cut and (e) discontinuity preserving relaxation. Extra results of the proposed method are displayed in the 2nd and 3rd rows with varying parameter values (λ_1, $\lambda_2 \times c_j$, for $j = \{1, \cdots, 3\}$, overall accuracy). In the 4th to 6th rows, corresponding results of the AVIRIS Salinas dataset are also showed.



(a) QB Zurich3 data (b) Ground-truth (c) Magnitude (d) Partitions

(e) QB Zurich6 data (f) Ground-truth (g) Magnitude (h) Partitions

Class legends

Figure 4.4: The MSI datasets. First row lists the QB Zurich3 dataset: (a) MSI RGB composite, (b) ground-truth reference, (c) magnitude map, (d) multiple over-segmented partitions/superpixels. Second row lists the corresponding maps(e-h) of the QB Zurich6 dataset.

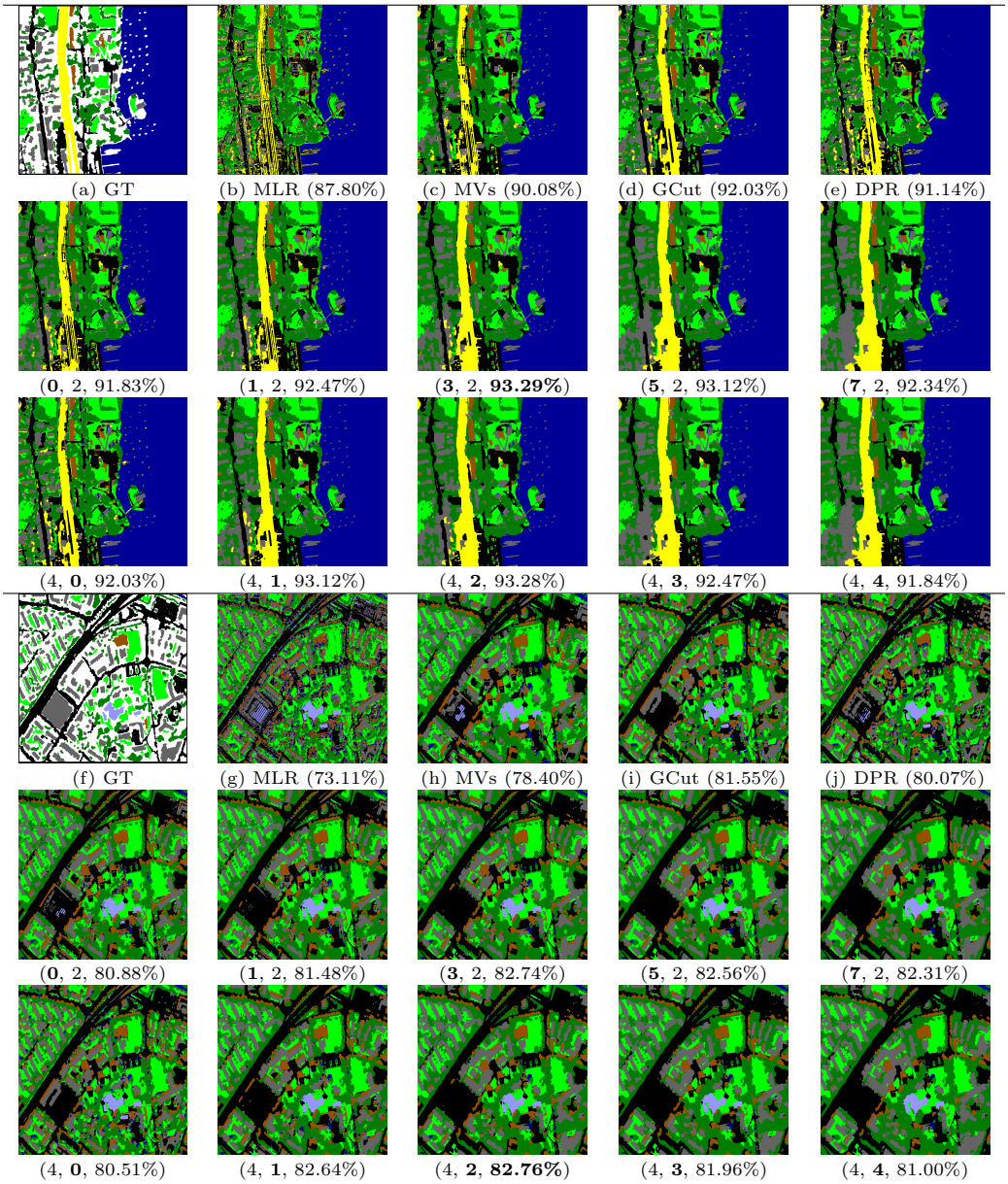


Figure 4.5: Segmentation maps of two QuickBird Zurich v1.0 MSI datasets: Zurich3 and Zurich6. Top row (the Zurich3 dataset) lists: (a) ground-truth, (b) MLR classification, (c) majovity voting, (d) graph-cutt and (e) discontinuity preserving relaxation. Results of the proposed method are displayed in the 2nd and 3rd rows with varying parameter values (λ_1, $\lambda_2 \times c_j$, for $j = \{1, \cdots, 3\}$, overall accuracy). In the 4th to 6th rows, corresponding results of the Zurich6 dataset are also showed.

Chapter 5

Multi-Superpixelization-based Convex Formulation for Joint Classification of Hyperspectral and LiDAR Data

5.1 Summary

The synergistic analysis of light detection and ranging (LiDAR) and hyperspectral data is attracting a significant interest in recent years due to the complementary nature of these two sources of remote sensing data. In this chapter, we propose a new spectral-spatial classification method able to jointly exploit these two kinds of data. Our work is based on three innovative components: 1) a superpixel generation method aimed at multivariate image spatial partitioning, 2) a multi-source framework for feature extraction, and 3) a convex framework used to approach the solutions of the resulted image labeling problem associated with vectorial total variation and superpixel-based graph total variation regularizers. Our experimental results, conducted with a hyperspectral data set collected by the Compact Airborne Spectrographic Imager (CASI) spectrometer over the city of Houston in 2013 and a corresponding LiDAR data set, illustrate the effectiveness of the proposed framework.

5.2 Introduction

Spectral-spatial classification techniques have been shown to be capable of greatly enhancing the classification performance as compared with techniques only using the spectral information in hyperspectral images [3], especially considering very high spatial resolution imagery. With the inclusion of spatial information, the existing spatial correlation between pixel signatures can be further exploited [67]. On the other hand, spatial information has also been helpful when dealing with the imbalance between the high spectral dimensionality of hyperspectral data and the generally limited availability

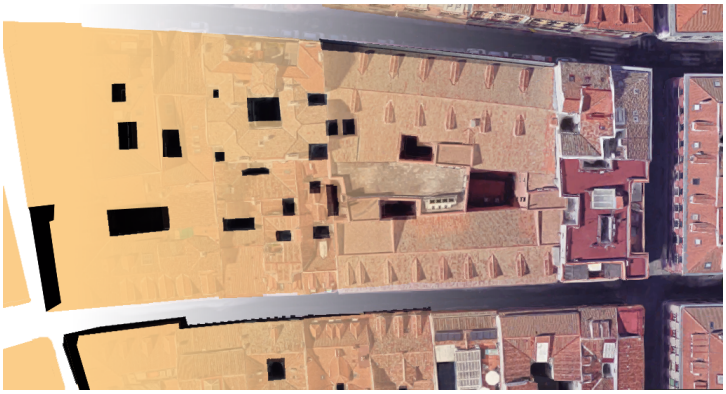


Figure 5.1: Importance of spatial information in a very high resolution image.

of labeled samples [10].

The techniques that exploit spatial information for hyperspectral image classification can generally be separated into two categories: pre-classification processing and post-classification processing. Techniques such as spatial filtering [10], morphological operators like attribute profiles (APs) [229], extended APs (EAPs) [70] allow a classifier to get rid of noise, contextual variations, outliers, and so on. However, the use of spatial information may also lead to overfitting, decreasing the generalization ability of the classificar. Post-classification techniques aim at including the spatial information into the classification model via regularizing the previously obtained label estimations. Traditionally, a wealth of techniques have been developed in the literature for spectral-spatial hyperspectral image classification [70, 104, 187, 211, 229]. The high spatial correlation turns out to be the key aspect for this task. Particularly, the image segments such as (over-segmented) superpixels have been shown to be very useful in promoting classification accuracy by taking advantage of the fact that the pixels comprising a superpixel generally share similar characteristics [104].

Compared with other image segmentation techniques, superpixelization can provide highly homogeneous segments. When using superpixels in classification, there usually exists a risk of modifying the sample labels in a superpixel [187] given the data term provided by a classifier, while the straightforward way usually leads to a discrete optimization problem that is NP-hard [104]. Apart from very few examples [223], a solution can only be approximated by graph-cut algorithms [88] in the discrete domain, or by other optimization techniques [200] in a relaxed or transformed compact domain. In our previous work [104], we have attacked this problem with a convex formulation with graph total variation (GTV) based on superpixels and vectorial total variation (VTV) after linearly relaxing the discrete optimization problem into a continuous domain. The spatial regularization of VTV and GTV is proven beneficial in promoting piecewise smoothness, aligning the discontinuities of object boundaries while enforcing the label consistency of pixels in a superpixel. Similar techniques have also been developed [211, 213]. However, the difficulty of performing accurate edge detection in hyperspectral images introduces great challenges to the design of accurate spatial

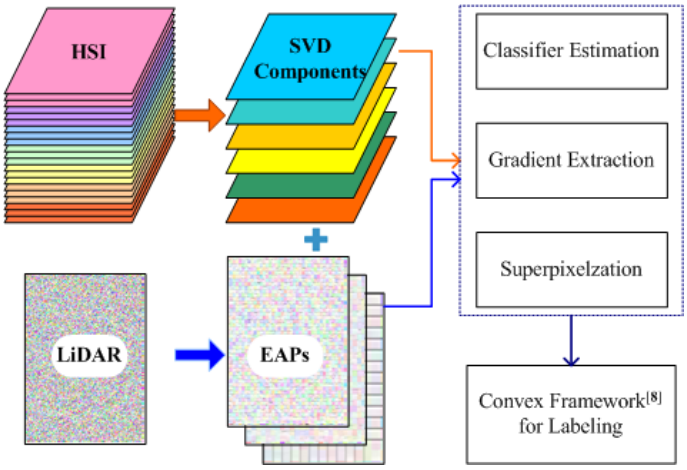


Figure 5.2: Block diagram of the proposed method.

partitioning techniques [93]. As a result, the extracted edges of superpixels may often fail to meet the demand of a classification mission to precisely discriminate some close land covers in the spatial domain, such as different plant species.

Motivated by the aforementioned issues, in this chapter we develop a new superpixel-based framework for spectral-spatial classification of hyperspectral and LiDAR data sets. The rich spatial information provided by LiDAR data is used in this chapter to complement the hyperspectral data. Considering these multi-source datasets, our main contributions include: 1) a feature extraction method with the multi-source datasets via singular value decomposition (SVD) and extended attribute profiles (EAP), 2) adapting the simple linear clustering algorithm (SLIC) [214] into multivariate image partitioning, and 3) approximating the solution of the resulted classification problem with a convex formulation [104] considering the vectorial total variation (VTV) and graph total variation (GTV) based on superpixels. Our method is experimentally validated and compared with state-of-the-art methods using a hyperspectral data set collected by the Compact Airborne Spectrographic Imager (CASI) spectrometer over the city of Houston in 2013 and a corresponding LiDAR data set [1].

5.3 Proposed Method

5.3.1 Convex formulation with spatial information

Let $\mathbf{x} \equiv [\mathbf{x}_1, \cdots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ be a feature matrix containing d-dimensional image feature vectors indexed by the integer set $\mathcal{S} \equiv \{1, 2, \cdots, n\}$. Let $\mathbf{y} \equiv (y_1, \cdots, y_n) \in \mathcal{L}^n$ be a label image of K classes, such that $y_i = k$ if and only if the label of pixel i is k. Given the posterior probability $p_{\mathbb{Y}\|\mathbb{X}}(\mathbf{y}|\mathbf{x})$, the conditional independence assumption of the observation model $p_{\mathbb{X}\|\mathbb{Y}}(\mathbf{x}|\mathbf{y})$, and the prior probability $p_{\mathbb{Y}}(\mathbf{y})$ (often an MRF), the maximum a posteriori (MAP) segmentation (or labeling) in a supervised scenario can be

[1] http://www.grss-ieee.org/community/technical-committees/data-fusion/2013-ieee-grss-data-fusion-contest/

written as:

$$\begin{aligned}\hat{\mathbf{y}}_{MAP} &= \arg\min_{\mathbf{y}\in\mathcal{L}^n} -\log(p_{\mathbb{X}|\mathbb{Y}}(\mathbf{x}|\mathbf{y})p_{\mathbb{Y}}(\mathbf{y})) \\ &= \arg\min_{\mathbf{y}\in\mathcal{L}^n} \sum_{i=1}^{n} D_i(y_i) + \lambda U(\mathbf{y}),\end{aligned} \tag{5.1}$$

where $D_i(y_i) = -\log p_{\mathbb{X}_i|\mathbb{Y}_i}(\mathbf{x}_i|y_i)$, $\lambda U(\mathbf{y}) = -\log p(\mathbf{y})$, and $\lambda > 0$ is a regularization parameter. The minimization of (5.1) is an integer optimization problem. In the case of Potts [134] prior and $K = 2$, the problem has exact solution obtained by mapping the problem into the computation of a min-cut on a suitable graph [223]. However, for $K > 2$, the optimization (5.1) is NP-hard [88] and, therefore, only approximations may be computed. In the work [104], the solution of this problem is approximated via the following convex formulation:

$$\begin{aligned}\widehat{\mathbf{z}} \in \arg \min_{\mathbf{z}\in[0,1]^{K\times n}} & \sum_{i=1}^{n} \mathbf{q}_i^T \mathbf{z}_i + \lambda_1 \sum_{n\in\mathcal{S}} \sqrt{||\mathbf{D}_h \mathbf{z}[n]||^2 + ||\mathbf{D}_v \mathbf{z}[n]||^2} + \lambda_2 \sum_{c}^{C} ||(\mathbf{A}_c - \mathbf{I})\mathbf{z}||_F^2 \\ \text{s.t.:} \quad & \mathbf{1}_K^T \mathbf{z} = \mathbf{1}_n^T,\end{aligned} \tag{5.2}$$

where $\mathbf{z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n]$ is the relaxed hidden field out of the discrete variable $\mathbf{y}$, $\mathbf{q}_i = [D(y_i = 1), \ldots, D(y_i = K)]^T$ is the data term often provided by a classifier, the second term represents the VTV regularizer with $\mathbf{D}_h, \mathbf{D}_v : \mathbb{R}^{K\times n} \mapsto \mathbb{R}^{K\times n}$ being linear operators computing horizontal and vertical first order backward vector differences, respectively, and the third term is the superpixel-based GTV [104]. Note that we are able to append an arbitrary number of superpixelizations by this means. In total, (5.2) is a convex formulation whose solutions can be obtained by the SALSA algorithm [211] as an instance of the ADMM method [200]. The solutions $\widehat{z}_{ki}$ yielded by (5.2) are mostly discrete [200] and a complete discrete solution is computed by $\widehat{\mathbf{y}}_i = \arg\max_k \widehat{z}_{ki}, \;\; i \in \mathcal{S}$.

5.3.2 Feature extraction and superpixelization

In this section, we describe the feature extraction methods with the multi-source data, followed by the preparation of the three terms in (5.2). Fig. 5.2 displays the presented framework including mainly three parts, feature extractions via SVD and EAP, probabilistic class estimation with the MLR classifier [211], and spatial regularizers associated with both gradient extraction and superpixels.

First of all, the EAP [70] and SVD are respectively performed on the LiDAR and hyperspectral data for feature extraction purposes. Specifically, via SVD transformation, we retain the acquired d_{SVD} major components $\mathbf{F}_{\text{SVD}} \in \mathbb{R}^{d_{\text{SVD}}\times n}$ [see Fig. 5.2 (e)] with the goal of relaxing the imbalance between limited training samples and high spectral dimensionality. As suggested in the work [70], d_{EAP} EAPs $\mathbf{F}_{\text{EAP}} \in \mathcal{R}^{d_{\text{EAP}}\times n}$ [see Fig. 5.2(f)], as a concatenation of many APs, are generated for the high resolution LiDAR data. As a result, the extracted features are collected by layer-wise stacking $\{\mathbf{F}_{\text{SVD}}, \mathbf{F}_{\text{EAP}}\}$, resulting in $\mathbf{F} \in \mathbb{R}^{d_{\text{F}}\times n}$, where $d_{\text{F}} = d_{\text{SVD}} + d_{\text{EAP}}$.

Once these features have been constructed, we train the MLR classifier with the features $\mathbf{f} \equiv [\mathbf{f}_1, \ldots, \mathbf{f}_n]$ where the high dimensionality is thus reduced to d_{F} or d_{SVD}

depending on the actual input. Hence, we estimate the data term of (5.2) by the MLR in the work [175]:

$$p(y_i = k|\mathbf{f}_i, \boldsymbol{\omega}) = \frac{\exp\left(\boldsymbol{\omega}^{(k)^T}\mathbf{h}(\mathbf{f}_i)\right)}{\sum_{k=1}^{K}\exp\left(\boldsymbol{\omega}^{(k)^T}\mathbf{h}(\mathbf{f}_i)\right)}, \tag{5.3}$$

where $\mathbf{h}(\mathbf{f}_i) \equiv [h_1(\mathbf{f}_i), \ldots, h_m(\mathbf{f}_i)]^T$ is a vector of m fixed functions, often termed as features; $\boldsymbol{\omega}^{(k)} \equiv [\omega_1^{(k)}, \ldots, \omega_m^{(k)}]^T$ is the set of logistic regressors for class k learnt by the LORSAL algorithm [175], and $\boldsymbol{\omega} \equiv [\boldsymbol{\omega}^{(1)^T}, \ldots, \boldsymbol{\omega}^{(K)^T}]^T$. Note that in this chapter, we train the classifier with $\mathbf{F}_{\text{SVD}}$ and $\mathbf{F}$, respectively, in order to also test the potential of LiDAR data in promoting a classifier.

To fill up the spatial regularization terms, we obtain the gradient map for VTV with the stacked muulti-variate feature image $\mathbf{F}$ through $\mathbf{G} = \sum_{i=1}^{d_{\text{F}}} \omega_i \mathbf{D}[\mathbf{F}^i]$, where $\mathbf{D}$ computes the 1^{st} order backward differences horizontally or vertically and $\omega_i = var(\mathbf{F}^i)/\sum var(\mathbf{F}^i)$, for $i = \{1, \ldots, d_{\text{F}}\}$, weights each component $\mathbf{F}^i$ with the value of its variance that is normalized for all. As for the superpixelization, we measure the *color* differences via the spectral angle distance (SAD) and discard the transformation from RGB space to CEILAB space [214], in order to adapt the SLIC algorithm to the case of multivariate images $\mathbf{F}$ out from the original based on three RGB channels.

5.4 Experimental Results

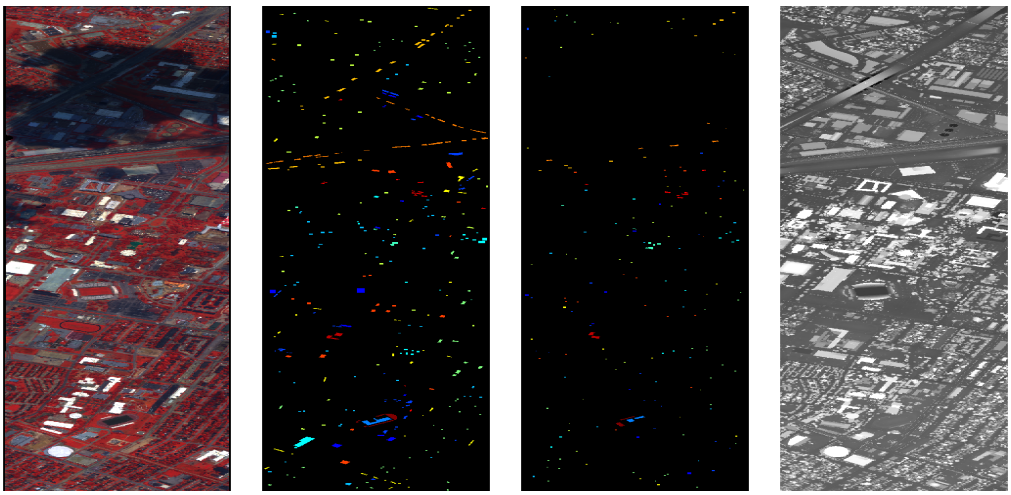
We evaluate our proposed method by testing its performance in classification with the Houston dataset that includes a hyperspectral data set and a rasterized LiDAR data set that are geographically co-registered, as shown in Figs. 5.3(a) and 5.3(d). The references comprise a fixed number of 2832 training samples and 12197 training samples. To set up the experiments, we choose the $1st$ 8 of the SVD major components ($d_{\text{SVD}} = 8$, [see Fig. 5.3(e)] from the hyperspectral image, and 3 extended attribute profiles [$d_{\text{EAP}} = 3$, see Fig. 5.3(f)] built from the LiDAR data. Hence, 11 components in total are exploited, which are later fed to the MLR classifier and used for the gradient map [see Fig. 5.3(g)] and superpixelization associated with the spatial regularizers [see (5.2)]. Note that here we utilize 3 superpixelizations for the graph total variation constraints. Specifically, graph-cut promotes the classification by reinforcing the label consistency for pixels in neighboring areas by considering the cost as the function energy [88]. On the other hand, the DPR takes advantage of the edges in the image objects to align the class labels following the discontinuities along the boundaries [189]. We thus selected these two state-of-the-art methods for comparison, in order to explore the effectiveness of our proposed method. Also, we compare the results with and without the inclusion of LiDAR data in preparing the data term provided by the MLR classifier.

The classification results are displayed in Fig. 5.4, from which several observations can be made. First of all, we can observe that the classification performance is dramatically improved with the inclusion of spatial information for all the compared methods. Also, the inclusion of the information of the LiDAR data promotes the classification results for

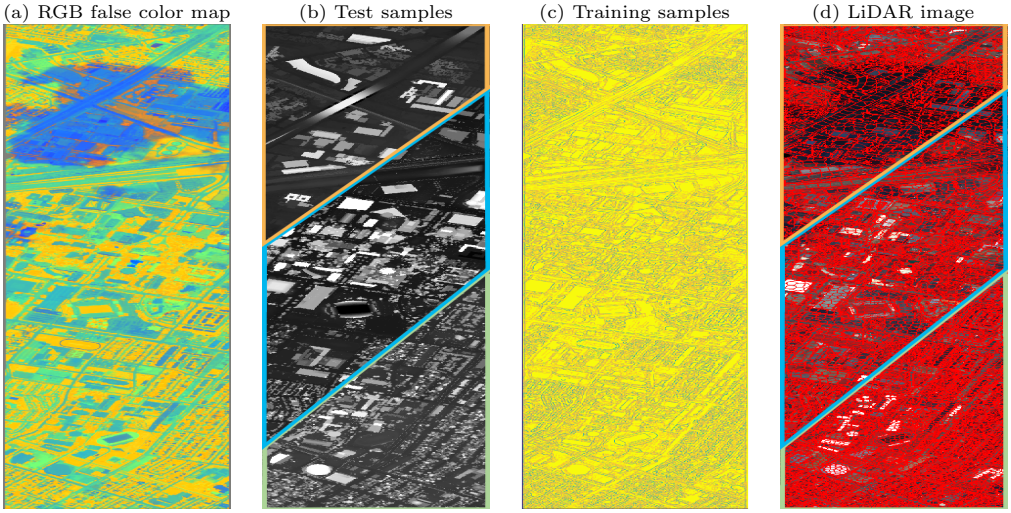
all cases, from Fig. 5.4(a) to Fig. 5.4(e), from Fig. 5.4(b) to Fig. 5.4(f), from Fig. 5.4(c) to Fig. 5.4(g), and from Fig. 5.4(d) to Fig. 5.4(h). Regarding our method based on [104], the results obtained improve those found by using the graph-cut and DPR algorithms, with and without using the LiDAR data.

5.5 Summary and future directions

This chapter proposes a new spectral-spatial classification method able to jointly exploit hyperspectral and LiDAR data. In addition to hyperspectral imagery, LiDAR data are also considered in order to provide improved classification estimations as well as well-defined spatial regularizers of GTV (based on superpixels) and VTV. Our experimental results with real hyperspectral and LiDAR data indicate the effectiveness of the newly proposed method when compared with the-state-of-the-art algorithms. Our future work will focus on a deeper and more extensive validation of the proposed method with additional hyperspectral and LiDAR datasets.



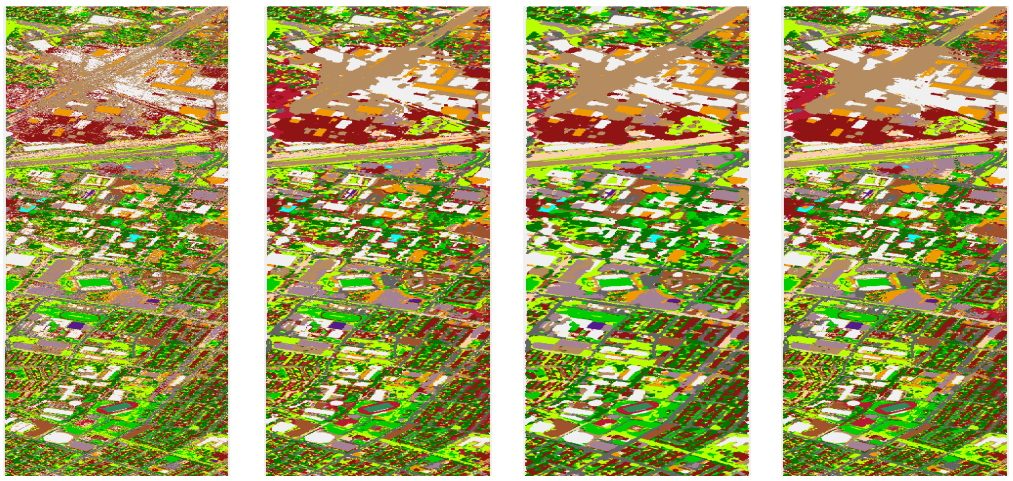
(a) RGB false color map (b) Test samples (c) Training samples (d) LiDAR image



(e) $\mathbf{F}_{\text{SVD}}$ of HSI (f) $\mathbf{F}_{\text{EAP}}$ of LiDAR (g) Manitude of $\mathbf{F}$ (h) Partitions of $\mathbf{F}$

Class	test - train	Class	test - train
Healthy grass	1053-198	Road	1059-193
Stressed grass	1064-190	Highway	1036-191
Synthetic grass	505-192	Railway	1054-181
Trees	1056-188	Parking lot 1	1041-192
Soil	1056-186	Parking lot 2	285-184
Water	143-182	Tennis court	247-181
Residential	1072-196	Running track	473-187
Commercial	1053-191		

Figure 5.3: The CASI hyperspectral data set collected over the city of Houston in 2013 and a corresponding LiDAR data collected over the same area.



(a) MLR_$\mathbf{F}_{\mathrm{SVD}}$ (82.73%) (b) GC_$\mathbf{F}_{\mathrm{SVD}}$(88.60%) (c) DPR_$\mathbf{F}_{\mathrm{SVD}}$(86.99%) (d) CF_$\mathbf{F}_{\mathrm{SVD}}$(89.68%)

(e) MLR_$\mathbf{F}$ (88.79%) (f) GC_$\mathbf{F}$(89.71%) (g) DPR_$\mathbf{F}$(89.07%) (h) CF_$\mathbf{F}$(89.89%)

Figure 5.4: Classification results obtained by different methods using the CASI hyperspectral data and the corresponding LiDAR data.

Chapter 6

Conclusions and Future Research Lines

In this thesis, we have addressed some challenges related to hyperspectral image classification. On the one hand, the high spectral dimensionality and high redundance of spectral information hinders greatly the development of relevant algorithms and techniques for hyperspectral image understanding. Especially, hyperspectral image classification has often been an ill-posed problem, due to the imbalance between the high spectral dimensions and limited availability of the labeled samples. On the other hand, the spatial information has seldom been exploited in a simple manner in image labeling problems, although proven to be very helpful in promoting classification performance of hyperspectral images. Taking the concept of superpixel, especially that resulting from over-segmented cases, the hyperspectral image labeling problem is usually NP-hard to be solved in polynomial time, due to its discrete nature. As a result, the work developed in this doctoral thesis is intended to attack the main challenges of hyperspectral image classification by considering both the high-dimensional nature and the discrete labeling problems involved in hyperspectral image analysis.

The first important original contribution of this thesis is the development of different spectral partitioning strategies that reassign the hyperspectral bands into multiple groups of band subset of much lower dimensionality, instead of discarding most of the bands or features as performed by traditional feature selection/extraction techniques. By this means, the original information (especially the physical-related information) remains intact for the subsequent analysis tasks. In detail, we mainly considered two aspects to achieve this goal. From the viewpoint of hyperspectral image analysis, we have developed a new band-clustering based spectral partitioning strategy which considers the high correlation/similarity among neighboring spectral bands, and also a new spectrometer-driven spectral partitioning strategy which considers the intrinsic characteristics imposed by the multi-spectrometers of the imaging device. On the other hand, we have also developed a new class-oriented spectral partitioning method based on band selecting techniques, taking into account the fact that the relevant information for classification purposes is class-dependent and may often live in weak bands or features that are

usually abandoned by traditional dimensionality reducing techniques. All these spectral partitioning strategies provide significant innovation in the context of dimensionality reduction and classification of hyperspectral images.

Another important contribution of this thesis work is the exploration of new strategies for utilizing spatial information in hyperspectral image classification. By means of a linear program, we have adopted the Bayesian maximum a posteriori (MAP) image segmentation perspective and then relaxed the original discrete optimization problem of image labeling into a compact domain, where the information from over-segmented superpixels is convexly formulated in the form of graph total variation. In this regard, the contribution introduced in this thesis is of great importance as it opens the door to flexibly append the information of real-world knowledge as priors in order to regularize the image labeling model. Also, it allows us to use an arbitrary number of spatial partitions of the hyperspectral images, which allows us to mitigate the dilemma related to the difficulty of searching for *the best* image partition from a huge number of image partitioning methods available, along with the varying parametric setups. Based on this newly developed algorithm, this thesis work has also conducted an instance of application with a multi-source data set, including a hyperspectral image and a rasterised LiDAR data that are obtained from the same scene and spatially co-registered. The advantages of our proposed method were demonstrated in comparison with other state-of-the-art spatial-spectral image classification techniques.

The future lines of this thesis work will focus on the development of novel ensemble learning techniques able to boost the final classification output resulting from the multiple classifiers that are learned from diverse perspectives of the spectral partitions. Another future line will focus on improving the graph total variation via the consideration of the inner structure of a subgraph/superpixel. This consideration has the potential to provide a more organic exploitation of superpixels with regards to minimizing risks caused by isolated or noisy pixels and, meanwhile, balancing the power between the superpixels that are of different sizes. The realization of these aforementioned works is in progress by the candidate. A final research line that will be pursued in the future is related to the combination of spectral partitioning and spatial partitioning as a whole in the scope of hyperspectral image classification.

Apendix A

Publications

The results of this thesis work have been published in several international journal papers and peer-reviewed international conference papers. Specifically, the applicant has co-authored 4 JCR journal papers and 8 international peer-reviewed conference papers directly related with this thesis work. This thesis work have been supported by the Chinese Scholarship Council (CSC) of the Ministry of Education of the People's Republis of China. The candidate has been a pre-doctoral researcher in the Department of Computer and Communications in the University of Extremadura, forming part of the Hyperspectral Computing Laboratory (HyperComp) research group of the University of Extremadura during four years. During the development of the thesis, the candidate has been a visitor (for a total of 14 months) with the Instituto de Telecomunicações, Instituto Superior Técnico, Universidade de Lisboa, Portugal. In the following, we describe the publications achieved by the candidate, providing also a short description of the journal or conference where it was presented and indication of the specific contributions of the candidate to each publication.

A.1 International Journal Papers

1. Y. Liu; J. Bioucas-Dias; J. Li; A. Plaza (2017). Multi-superpixelization-based convex formulation for joint classification of hyperspectral and LiDAR data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, submitted under review. [JCR(2015)=2.145].

   This paper was published in the journal *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, which is one of the main journals of the remote sensing category of JCR. This paper explores the application of the convex formulation work introduced in chapter four to multi-source data set, including a hyperspectral image and a co-registered LiDAR data set. To reach this goal, a framework based on SVD is developed for the multi-source image data superpixel generation and classification.

2. Y. Liu; F. Condessa; J. Bioucas-Dias; J. Li; A. Plaza (2017). Convex Formulation for Remote Sensing Image Classification with Multi-Superpixelizations. IEEE

Transactions on Geoscience and Remote Sensing, submitted under review. [JCR(2015)=3.360].

This paper was submitted to the journal *IEEE Transactions on Geoscience 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 provides a straightforward to make use of the spatial information from over-segmented superpixels. With the linear program, this work convexly relaxes the original NP-hard discrete optimization problem of image labeling into the compact domain, which opens the door to freely forge the real-world knowlege as priors to regularize the image labeling process. Armed with this machinery, the over-segmented superpixels are formulated into the form of graph total variation that reinforces the label consistency of the pixels comprising the same superpixel. The solution of the formulated problem is obtained via a developed instance of the SALSA algorithm.

3. Y. Liu; J. Li; A. Plaza (2017). Class-Oriented Spectral Partitioning for Remotely Sensed Hyperspectral Image Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14(2), 409-416. [JCR(2015)=2.145].

   This paper was published in the journal *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, which is one of the main journals of the remote sensing category of JCR. This paper develops a new class-oriented spatial partitioning strategy which obtains the class-oriented spectral partitions, considering that the relevant information for classification may often live in weak features/bands and be class-dependent. Several traditional band selection algorithms are used, for this purpose, to generate the spectral partitions for each class in a supervised manner.

4. Y. Liu; J. Li; A. Plaza (2015). Spectrometer-Driven Spectral Partitioning for Hyperspectral Image Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(2), 668-680. [JCR(2015)=2.145].

   This paper was published in the journal *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, which is one of the main journals of the remote sensing category of JCR. This paper is focused on developing a spectrometer-driven spatial partitioning strategy which exploits the characteristics of the multiple spectrometers and high band correlation. The band similarity provided by the spectrometers and band clusters are utilized to generate multiple band subsets, namely, spectral partitions that have much lower spectral dimensions, which provides diverse perspectives into the original hyperspectral images. Instead of mostly discarded by traditional dimensionality reduction techniques, the band subsets are fed to a multiple classifier system for ensemble learning purposes to obtain a final classification output.

A.2 Peer-reviewed International Conference Papers

The work presented in this doctoral thesis, along with others performed during the time period it was done, was also presented in several conferences and workshops around the world. Each one of them is detailed as follows:

1. Y. Liu; J. Bioucas-Dias; J. Li; A. Plaza (2017). A Cloud Shadow Removal Method Based On Linear Unmixing Model For Hyperspectral Image Classification. In Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International (submitted and under review). IEEE.

   This work has been accepted for oral presentation at the *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* that will be held in Fort Worth, Texas, USA, in July 2017. This is the most important international workshop in the remote sensing field. This work develops a novel cloud shadow effect removal method based on linear unmixing model, where the unmixing is only performed once on the well-illuminated pixels instead of performed twice on both the shadowed and well-illuminated pixels that is usually conducted by the traditional methods.

2. Y. Liu; J. Bioucas-Dias; J. Li; A. Plaza (2017). Multi-superpixelization-based convex formulation for joint classification of hyperspectral and LiDAR data. In Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International (submitted and under review). IEEE.

   This work has been accepted for oral presentation at the *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* that will be held in Fort Worth, Texas, USA, in July 2017. This is the most important international workshop in the remote sensing field. The work conducted in this manuscript is a application of convex formulation for image classification with superpixels to the case of multiple data source, which in our work, includes a hyperspectral data and a co-registered LiDAR data.

3. Y. Liu; F. Condessa; J. Bioucas-Dias; J. Li; A. Plaza (2016). Convex formulation for hyperspectral image classification with superpixels. In Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International (pp. 3294-3297). IEEE.

   This work was presented as an oral presentation in the *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* held in Beijing City, China, in 2016. This is the most important international workshop in the remote sensing field. This work develops convex formulation to deal with the NP-hard problem of image labeling with superpixels. The linear program is used to relax the original discrete optimization problem into a compact domain, which opens the door to flexibly append regularizing information from multi-sources as priors. Under this machinery, the superpixels are formulated as graph total variation and a corresponding solution is approached by the developed instance of the SALSA algorithm.

Convex formulation for hyperspectral image classification with superpixels

4. Y. Sun; X. Zhang; A. Plaza; J. Li; I. Dópido; Y. Liu (2016). A new semi-supervised classification strategy combining active learning and spectral unmixing of hyperspectral data. In SPIE Remote Sensing (pp. 1000708-1000708). International Society for Optics and Photonics.

   This work was presented as an oral presentation in the *In SPIE Remote Sensing* held in Edingburgh City, UK, in 2016. This work explores the possibility and feasible clue of utilizing the spectral unmixing technology into a semi-supervised classification process. In this work, the abundance information and class probabilistic estimation are simultaneously considered in the screening of candidate training samples.

5. Y. Liu; J. Li; A. Plaza; Y. Sun (2016). A multiple criteria-based spectral partitioning method for remotely sensed hyperspectral image classification. In SPIE Remote Sensing (pp. 100070C-100070C). International Society for Optics and Photonics.

   This work was presented as an oral presentation in the *In SPIE Remote Sensing* held in Edingburgh City, UK, in 2016. This work aims at diversity generating with using different band selection algorithms in the scenario spectral partitioning of hyperspectral images. The diversiy is generated in order to boost the ensemble learning process with the multiple classifiers that are learned with the spectral partitions.

6. Y. Liu; J. Li; A. Plaza (2015). Class-oriented spectral partitioning for hyperspectral image classification. In Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International (pp. 4983-4986). IEEE.

   This work was presented as an oral presentation in the *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* held in Milan City, Italy, in 2015. This is the most important international workshop in the remote sensing field. This work develops a class-oriented spectral partitioning method for hyperspectral image classification. For each spectral partition, the corresponding class-dependent bands are selected by using traditional band selection algorithms.

7. Y. Liu; J. Li; A. Plaza (2014). Spectrometer-Driven Spectral Partitioning for Hyperspectral Image Classification. Workshop on Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS), 2014.

   This work was presented as an oral presentation in the *Workshop on Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS)* held in Lausanne, Switzerland, in 2014. This work presents a spectrometer-driven spectral partitioning method for hyperspectral image classification. The band similarities imposed by the spectrometer is considered to provide the base band groups for spectral partitioning.

8. Y. Liu; J. Li; A. Plaza; J. M. Bioucas-Dias; P.G. Rodriguez; A. Cuartero (2014). Spectral Partitioning for Hyperspectral Remote Sensing Image Classification. In Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International (pp. 3434-3437). IEEE.

   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. This work introduces a new spectral band clustering based spectral partitioning method for dimensionality reduction of hyperspectral image in the scenario of classification.

Bibliography

[1] R. Green, M. Eastwood, C. Sarture, et al. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). *Remote Sensing of Environment*, 65(3):227–248, 1998. [Quoted on pages xi, 21 and 22]

[2] R. Green, M. Eastwood, C. Sarture, et al. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (aviris). *Remote sensing of environment*, 65:227–248, 1998. [Quoted on page 3]

[3] A. Plaza, J. Benediktsson, J. Boardman, et al. Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment*, 113:S110–S122, 2009. [Quoted on pages 3, 4, 5, 6, 19, 48, 69, 70 and 93]

[4] T. Sakamoto, M. Yokozawa, H. Toritani, et al. A crop phenology detection method using time-series modis data. *Remote sensing of environment*, 96(3):366–374, 2005. [Quoted on page 3]

[5] C. Berger, M. Voltersen, R. Eckardt, et al. Multi-modal and multi-temporal data fusion: Outcome of the 2012 grss data fusion contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(3):1324–1340, 2013. [Quoted on page 3]

[6] A. Vrieling, J. De, M. Steven, et al. Timing of erosion and satellite data: A multi-resolution approach to soil erosion risk mapping. *International Journal of Applied Earth Observation and Geoinformation*, 10(3):267–281, 2008. [Quoted on page 3]

[7] J. Bioucas-Dias, A. Plaza, G. Camps-Valls, and otherss. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and Remote Sensing Magazine*, 1:6–36, 2013. [Quoted on pages 4, 20, 22 and 39]

[8] B. Frey and D. Dueck. Clustering by passing messages between data points. *Science*, 315(5814):972–976, 2007. [Quoted on pages 4 and 23]

[9] G. Martín and A. Plaza. Spatial preprocessing for endmember extraction using unsupervised clustering and orthogonal subspace projection concepts. In *2010 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 959–962. IEEE, 2010. [Quoted on page 4]

[10] Y. Tarabalka, J. Benediktsson, and J. Chanussot. Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques. *IEEE Transactions on Geoscience and Remote Sensing*, 47(8):2973–2987, 2009. [Quoted on pages 4, 11, 71 and 94]

[11] J. Li, J. Bioucas-Dias, and A. Plaza. Semi-supervised hyperspectral classification using active label selection. In *SPIE Europe Remote Sensing*, pages 74770F–74770F. International Society for Optics and Photonics, 2009. [Quoted on pages 4, 50 and 56]

[12] L. Bruzzone and D. Prieto. An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images. *IEEE Transactions on image processing*, 11(4):452–466, 2002. [Quoted on page 4]

[13] A. Ertürk, M.-D. Iordache, and A. Plaza. Sparse unmixing-based change detection for multitemporal hyperspectral images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(2):708–719, 2016. [Quoted on page 4]

[14] S. Valero, P. Salembier, and J. Chanussot. Object recognition in urban hyperspectral images using binary partition tree representation. In *igarss*, pages 4098–4101. IEEE, 2013. [Quoted on page 4]

[15] M. Fingas and C. Brown. Review of oil spill remote sensing. *Spill Science & Technology Bulletin*, 4(4):199–208, 1997. [Quoted on page 4]

[16] W. Marcus, C. Legleiter, R. Aspinall, et al. High spatial resolution hyperspectral mapping of in-stream habitats, depths, and woody debris in mountain streams. *Geomorphology*, 55(1):363–380, 2003. [Quoted on page 4]

[17] M. Eismann, A. Stocker, and N. Nasrabadi. Automated hyperspectral cueing for civilian search and rescue. *Proceedings of the IEEE*, 97(6):1031–1055, 2009. [Quoted on page 4]

[18] D. Tralli, R. Blom, V. Zlotnicki, et al. Satellite remote sensing of earthquake, volcano, flood, landslide and coastal inundation hazards. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59(4):185–198, 2005. [Quoted on page 4]

[19] T. Lillesand, R. Kiefer, and J. Chipman. *Remote sensing and image interpretation*. John Wiley & Sons, 2014. [Quoted on pages 4 and 12]

[20] S-S. Chiang, C-I. Chang, and I. Ginsberg. Unsupervised target detection in hyperspectral images using projection pursuit. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7):1380–1391, 2001. [Quoted on page 5]

[21] L. Jiménez, J. Rivera-Medina, E. Rodríguez-Díaz, et al. Integration of spatial and spectral information by means of unsupervised extraction and

classification for homogenous objects applied to multispectral and hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4):844–851, 2005. [Quoted on page 5]

[22] S. Kim. Unsupervised spectral-spatial feature selection-based camouflaged object detection using vnir hyperspectral camera. *The Scientific World Journal*, 2015, 2015. [Quoted on page 5]

[23] H. Akçay and S. Aksoy. Automatic detection of geospatial objects using multiple hierarchical segmentations. *IEEE Transactions on Geoscience and Remote Sensing*, 46(7):2097–2111, 2008. [Quoted on page 5]

[24] J. Bioucas-Dias, A. Plaza, N. Dobigeon, et al. 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. [Quoted on page 5]

[25] L. Bernstein, S. Adler-Golden, R. Sundberg, et al. A new method for atmospheric correction and aerosol optical property retrieval for vis-swir multi-and hyperspectral imaging sensors: Quac (quick atmospheric correction). Technical report, DTIC Document, 2005. [Quoted on page 5]

[26] J. Moreno. Spectral/spatial integration effects on information extraction from multispectral data: multiresolution approaches. In *Satellite Remote Sensing*, pages 324–338. International Society for Optics and Photonics, 1995. [Quoted on page 5]

[27] M. Datcu, H. Daschiel, A. Pelizzari, et al. Information mining in remote sensing image archives: system concepts. *IEEE Transactions on Geoscience and Remote Sensing*, 41(12):2923–2936, 2003. [Quoted on page 5]

[28] J. Zhang. Multi-source remote sensing data fusion: status and trends. *International Journal of Image and Data Fusion*, 1(1):5–24, 2010. [Quoted on page 5]

[29] M. Dalponte, L. Bruzzone, and D. Gianelle. Tree species classification in the southern alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and lidar data. *Remote sensing of environment*, 123:258–270, 2012. [Quoted on page 5]

[30] 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, 2008. [Quoted on pages 5 and 12]

[31] J. 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. [Quoted on page 5]

[32] C.-M. Chen, G. Hepner, and R. Forster. Fusion of hyperspectral and radar data using the ihs transformation to enhance urban surface features. *ISPRS Journal of photogrammetry and Remote Sensing*, 58(1):19–30, 2003. [Quoted on page 5]

[33] Liu T, J. Li, A. Plaza, et al. Class-oriented spectral partitioning for hyperspectral image classification. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 4983–4986. IEEE, 2015. [Quoted on page 6]

[34] J. Plaza and A. Plaza. Spectral mixture analysis of hyperspectral scenes using intelligently selected training samples. *IEEE Geoscience and Remote Sensing Letters*, 7:1–5, 2010. [Quoted on page 6]

[35] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Marí, et al. Composite kernels for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 3(1):93–97, 2006. [Quoted on page 6]

[36] S. Wold, K. Esbensen, and P. Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987. [Quoted on page 6]

[37] T.W. Lee. Independent component analysis. In *Independent Component Analysis*, pages 27–66. Springer, 1998. [Quoted on pages 6 and 40]

[38] M. Crawford, L. Ma, and W. Kim. Exploring nonlinear manifold learning for classification of hyperspectral data. In *Optical Remote Sensing*, pages 207–234. Springer, 2011. [Quoted on page 6]

[39] J. Bioucas-Dias and J. Nascimento. Hyperspectral subspace identification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(8):2435–2445, 2008. [Quoted on pages 6 and 40]

[40] L. Bruce, C. Koger, and J. Li. Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE Transactions on geoscience and remote sensing*, 40(10):2331–2338, 2002. [Quoted on page 6]

[41] S. Kumar, J. Ghosh, and M. Crawford. Best-bases feature extraction algorithms for classification of hyperspectral data. *IEEE Transactions on Geoscience and remote sensing*, 39(7):1368–1379, 2001. [Quoted on page 6]

[42] B. Kuo, C. Li, and J. Yang. Kernel nonparametric weighted feature extraction for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 47(4):1139–1155, 2009. [Quoted on page 6]

[43] M. Schlerf and C. Atzberger. Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing data. *Remote sensing of environment*, 100(3):281–294, 2006. [Quoted on page 7]

[44] R. Huang and M. He. Band selection based on feature weighting for classification of hyperspectral data. *IEEE Geoscience and Remote Sensing Letters*, 2(2):156–159, 2005. [Quoted on page 7]

[45] B.-C. Gao, P. Yang, W. Han, et al. An algorithm using visible and 1.38-μm channels to retrieve cirrus cloud reflectances from aircraft and satellite data. *IEEE Transactions on Geoscience and Remote Sensing*, 40(8):1659–1668, 2002. [Quoted on page 7]

[46] L. Jimenez and D. Landgrebe. Hyperspectral data analysis and supervised feature reduction via projection pursuit. *IEEE Transactions on Geoscience and Remote Sensing*, 37(6):2653–2667, 1999. [Quoted on page 7]

[47] C-I. Chang, Q. Du, T. Sun, et al. A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 37(6):2631–2641, 1999. [Quoted on page 7]

[48] H. Du, H. Qi, X. Wang, et al. Band selection using independent component analysis for hyperspectral image processing. In *Applied Imagery Pattern Recognition Workshop, 2003. Proceedings. 32nd*, pages 93–98. IEEE, 2003. [Quoted on page 7]

[49] N. Keshava. Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries. *IEEE Transactions on Geoscience and remote sensing*, 42(7):1552–1565, 2004. [Quoted on pages 7 and 44]

[50] D. Brady, K. Choi, D. Marks, et al. Compressive holography. *Opt. Express*, 17:13040–13049, 2009. [Quoted on page 7]

[51] R. Archibald and G. Fann. Feature selection and classification of hyperspectral images with support vector machines. *IEEE Geoscience and remote sensing letters*, 4(4):674–677, 2007. [Quoted on page 7]

[52] X. Jia and J.A. Richards. Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification. *IEEE Transactions on Geoscience and Remote Sensing*, 37(1):538–542, 1999. [Quoted on pages 7, 8, 20 and 21]

[53] Q. Du, W. Zhu, H. Yang, et al. Segmented principal component analysis for parallel compression of hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 6(4):713–717, 2009. [Quoted on pages 7 and 9]

[54] Y. Liu, J. Li, and A. Plaza. Spectrometer-driven spectral partitioning for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(2):668–680, 2016. [Quoted on pages 7, 16, 41 and 43]

[55] Y. Liu, J. Li, P. Du, et al. Class-oriented spectral partitioning for remotely sensed hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2016. [Quoted on pages 8 and 16]

[56] Y. Liu, J. Li, A. Plaza, et al. Spectral partitioning for hyperspectral remote sensing image classification. In *2014 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 3434–3437. IEEE, 2014. [Quoted on pages 8, 16, 41, 42, 43 and 50]

[57] L. Kuncheva and C. Whitaker. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine learning*, 51(2):181–207, 2003. [Quoted on pages 8 and 9]

[58] F. Tsai, E-K. Lin, and K. Yoshino. Spectrally segmented principal component analysis of hyperspectral imagery for mapping invasive plant species. *International Journal of Remote Sensing*, 28(5):1023–1039, 2007. [Quoted on page 9]

[59] C. Chen, W. Li, E.W. Tramel, et al. Spectral–spatial preprocessing using multihypothesis prediction for noise-robust hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(4):1047–1059, 2014. [Quoted on page 9]

[60] A. Plaza, J. Plaza, A. Paz, et al. Parallel hyperspectral image and signal processing [applications corner]. *IEEE Signal Processing Magazine*, 28(3):119–126, 2011. [Quoted on page 9]

[61] J. Plaza, A. Plaza, R. Pérez, et al. Parallel classification of hyperspectral images using neural networks. In *Computational Intelligence for Remote Sensing*, pages 193–216. Springer, 2008. [Quoted on page 9]

[62] Y. Zhao, L. Zhang, and S.G. Kong. Band-subset-based clustering and fusion for hyperspectral imagery classification. *IEEE Transactions on Geoscience and Remote Sensing*, 49(2):747–756, 2011. [Quoted on page 9]

[63] S. Prasad and L.M. Bruce. Hyperspectral feature space partitioning via mutual information for data fusion. In *2007 IEEE International Geoscience and Remote Sensing Symposium*, pages 4846–4849. IEEE, 2007. [Quoted on page 9]

[64] S. Prasad and L.M. Bruce. Overcoming the small sample size problem in hyperspectral classification and detection tasks. In *IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium*, volume 5, pages V–381. IEEE, 2008. [Quoted on page 9]

[65] M. Fauvel, Y. Tarabalka, J. Benediktsson, et al. Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE*, 101(3):652–675, 2013. [Quoted on pages 9 and 11]

[66] M. Vidal and J. Amigo. Pre-processing of hyperspectral images. essential steps before image analysis. *Chemometrics and Intelligent Laboratory Systems*, 117:138–148, 2012. [Quoted on page 10]

[67] G. Martín and A. Plaza. Spatial-spectral preprocessing prior to endmember identification and unmixing of remotely sensed hyperspectral data. *IEEE journal of selected topics in applied earth observations and remote sensing*, 5(2):380–395, 2012. [Quoted on pages 10 and 93]

[68] G. Rellier, X. Descombes, F. Falzon, et al. Texture feature analysis using a gauss-markov model in hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 42(7):1543–1551, 2004. [Quoted on pages 10 and 12]

[69] S. Phinn, C. Roelfsema, A. Dekker, et al. Mapping seagrass species, cover and biomass in shallow waters: An assessment of satellite multi-spectral and airborne hyper-spectral imaging systems in moreton bay (australia). *Remote Sensing of Environment*, 112(8):3413–3425, 2008. [Quoted on page 10]

[70] M. Dalla Mura, A. Villa, J. Benediktsson, et al. Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis. *IEEE Geoscience and Remote Sensing Letters*, 8(3):542–546, 2011. [Quoted on pages 10, 94 and 96]

[71] P. Marpu, M. Pedergnana, M. Dalla Mura, et al. Automatic generation of standard deviation attribute profiles for spectral–spatial classification of remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 10(2):293–297, 2013. [Quoted on pages 10 and 11]

[72] J. Li, H.Zhang, and L. Zhang. Supervised segmentation of very high resolution images by the use of extended morphological attribute profiles and a sparse transform. *IEEE Geoscience and Remote Sensing Letters*, 11(8):1409–1413, 2014. [Quoted on page 10]

[73] P. Ghamisi, J. Benediktsson, and J. 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, 2014. [Quoted on page 10]

[74] R. Pande-Chhetri and A. Abd-Elrahman. Filtering high-resolution hyperspectral imagery in a maximum noise fraction transform domain using wavelet-based de-striping. *International journal of remote sensing*, 34(6):2216–2235, 2013. [Quoted on page 11]

[75] X. Tang, W. Pearlman, and J. Modestino. Hyperspectral image compression using three-dimensional wavelet coding. In *Proceedings of SPIE*, volume 5022, pages 1037–1047, 2003. [Quoted on page 11]

[76] X. Huang, Q. Lu, L. Zhang, et al. New postprocessing methods for remote sensing image classification: A systematic study. *IEEE Transactions on Geoscience and Remote Sensing*, 52(11):7140–7159, 2014. [Quoted on page 11]

[77] Y. Tarabalka, J. Chanussot, J. Benediktsson, et al. Segmentation and classification of hyperspectral data using watershed. In *IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium*, volume 3, pages III–652. IEEE, 2008. [Quoted on page 11]

[78] Y. Tarabalka, J. Benediktsson, J. Chanussot, et al. Classification of hyperspectral data using support vector machines and adaptive neighborhoods. In *Proc. 6th EARSeL SIG IS Workshop*, pages 1–6, 2009. [Quoted on page 11]

[79] D.R. Thompson, L. Mandrake, M.S. Gilmore, et al. Superpixel endmember detection. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4023–4033, 2010. [Quoted on page 11]

[80] D.B. Cavanaugh, M. Dombrowski, and B. Catanzaro. Spatially corrected full-cubed hyperspectral imager, October 7 2008. US Patent 7,433,042. [Quoted on page 11]

[81] G. Zhang, X. Jia, and N. Kwok. Spectral-spatial based super pixel remote sensing image classification. In *2011 4th International Congress on Image and Signal Processing (CISP)*, volume 3, pages 1680–1684. IEEE, 2011. [Quoted on page 11]

[82] J. Li, H. Zhang, and L. Zhang. Efficient superpixel-level multitask joint sparse representation for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 53(10):5338–5351, 2015. [Quoted on page 11]

[83] Y. Liu, F. Condessa, J. Bioucas-Dias, et al. Convex formulation for hyperspectral image classification with superpixels. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 3294–3297. IEEE, 2016. [Quoted on page 11]

[84] Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in nd images. In *Eighth IEEE International Conference on Computer Vision, 2001. Proceedings.*, volume 1, pages 105–112. IEEE, 2001. [Quoted on pages 12, 70 and 84]

[85] Y. Tarabalka, J. Chanussot, and J. Benediktsson. Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 40(5):1267–1279, 2010. [Quoted on pages 12, 71 and 84]

[86] J. Li, J. Bioucas-Dias, and A. Plaza. Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4085–4098, 2010. [Quoted on pages 12 and 20]

[87] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *Journal of Applied Statistics*, 20(5-6):25–62, 1993. [Quoted on page 12]

[88] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE transactions on pattern analysis and machine intelligence*, 28(10):1568–1583, 2006. [Quoted on pages 12, 70, 74, 94, 96 and 97]

[89] P. Felzenszwalb and D. Huttenlocher. Efficient belief propagation for early vision. *International journal of computer vision*, 70(1):41–54, 2006. [Quoted on page 12]

[90] P. Kohli, P. Torr, et al. Robust higher order potentials for enforcing label consistency. *International Journal of Computer Vision*, 82(3):302–324, 2009. [Quoted on pages 12 and 70]

[91] T. Pock, A. Chambolle, D. Cremers, et al. A convex relaxation approach for computing minimal partitions. In *CVPR 2009. IEEE Conference on Computer Vision and Pattern Recognition, 2009.*, pages 810–817. IEEE, 2009. [Quoted on pages 12, 71, 72 and 74]

[92] B. Glocker, N. Komodakis, G. Tziritas, et al. Dense image registration through mrfs and efficient linear programming. *Medical image analysis*, 12(6):731–741, 2008. [Quoted on page 12]

[93] R. Maini and H. Aggarwal. Study and comparison of various image edge detection techniques. *International journal of image processing*, 3(1):1–11, 2009. [Quoted on pages 12 and 95]

[94] W. Bakker and K. Schmidt. Hyperspectral edge filtering for measuring homogeneity of surface cover types. *ISPRS Journal of Photogrammetry and Remote Sensing*, 56(4):246–256, 2002. [Quoted on page 12]

[95] G. Asner, D. Knapp, J. Boardman, et al. Carnegie airborne observatory-2: Increasing science data dimensionality via high-fidelity multi-sensor fusion. *Remote Sensing of Environment*, 124:454–465, 2012. [Quoted on page 12]

[96] T. Jones, N. Coops, and T. Sharma. Assessing the utility of airborne hyperspectral and lidar data for species distribution mapping in the coastal pacific northwest, canada. *Remote Sensing of Environment*, 114(12):2841–2852, 2010. [Quoted on page 12]

[97] G. Blackburn. Hyperspectral remote sensing of plant pigments. *Journal of experimental botany*, 58(4):855–867, 2007. [Quoted on page 12]

[98] W. Turner, S. Spector, N. Gardiner, et al. Remote sensing for biodiversity science and conservation. *Trends in ecology & evolution*, 18(6):306–314, 2003. [Quoted on page 12]

[99] B. Koch. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6):581–590, 2010. [Quoted on page 12]

[100] Y. Chen, W. Su, J. Li, et al. Hierarchical object oriented classification using very high resolution imagery and lidar data over urban areas. *Advances in Space Research*, 43(7):1101–1110, 2009. [Quoted on page 13]

[101] M. Pedergnana, P. Marpu, M. Dalla, et al. Classification of remote sensing optical and lidar data using extended attribute profiles. *IEEE Journal of Selected Topics in Signal Processing*, 6(7):856–865, 2012. [Quoted on page 13]

[102] M. Khodadadzadeh, J. Li, S. Prasad, et al. Fusion of hyperspectral and lidar remote sensing data using multiple feature learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):2971–2983, 2015. [Quoted on page 13]

[103] C. Debes, A. Merentitis, R. Heremans, et al. 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, 2014. [Quoted on page 13]

[104] Y. Liu, F. Condessa, J. Bioucas-Dias, et al. Convex formulation for hyperspectral image classification with multi-superpixelizations (under review). *IEEE Transactions on Geoscience and Remote Sensing*, 2016. [Quoted on pages 17, 94, 95, 96 and 98]

[105] Y. Liu, J. Bioucas-Dias, J. Li, et al. A convex formulation framework for superpixel-based multi-source remote sensing image classification (under review). *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2016. [Quoted on page 17]

[106] M. Govender, K. Chetty, and H .Bulcock. A review of hyperspectral remote sensing and its application in vegetation and water resource studies. *Water SA (Pretoria)*, 33:145–151, 2007. [Quoted on pages 20 and 39]

[107] A. Plaza and C.-I. Chang. *High performance computing in remote sensing*. Taylor & Francis: Boca Raton, FL, 2007. [Quoted on pages 20 and 39]

[108] F. Dell'Acqua, P. Gamba, A. Ferrari, et al. Exploiting spectral and spatial information in hyperspectral urban data with high resolution. *IEEE Geoscience and Remote Sensing Letters*, 1(4):322–326, 2004. [Quoted on pages 20 and 39]

[109] J. Bioucas-Dias, A. Plaza, G. Camps-Valls, et al. Hyperspectral semote sensing data analysis and future challenges. *IEEE Geoscience and Remote Sensing Magazine*, 1(2):6–36, 2013. [Quoted on page 20]

[110] M. Fauvel, Y. Tarabalka, J. Benediktsson, et al. Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE*, 101(3):652–675, 2013. [Quoted on page 20]

[111] D. Landgrebe. *Signal theory methods in multispectral remote sensing*. John Wiley & Sons: New York, 2003. [Quoted on page 20]

[112] T. Bandos, L. Bruzzone, and G. Camps-Valls. Classification of hyperspectral images with regularized linear discriminant analysis. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(3):862–873, 2009. [Quoted on page 20]

[113] C. Bachmann, T. Ainsworth, and R. Fusina. Exploiting manifold geometry in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):441–454, 2005. [Quoted on page 20]

[114] C.-I. Chang. *Hyperspectral data processing: algorithm design and analysis*. John Wiley & Sons, 2013. [Quoted on pages 20 and 44]

[115] J. Bioucas-Dias and J. Nascimento. Hyperspectral subspace identification. *Geoscience and Remote Sensing, IEEE Transactions on*, 46(8):2435–2445, 2008. [Quoted on page 20]

[116] Q. Du and J. Fowler. Hyperspectral image compression using jpeg2000 and principal component analysis. *IEEE Geoscience and Remote Sensing Letters*, 4(2):201–205, 2007. [Quoted on page 20]

[117] G. Licciardi, F. Pacifici, D. Tuia, et al. 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. [Quoted on page 20]

[118] J. Li and J. Bioucas-Dias. Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data. In *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*, volume 3, pages III250–III253, 2008. [Quoted on page 20]

[119] P. Mitra, C. Murthy, and S. Pal. Unsupervised feature selection using feature similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3):301–312, 2002. [Quoted on page 20]

[120] A. Farahat, A. Ghodsi, and M. Kamel. An efficient greedy method for unsupervised feature selection. In *2011 IEEE 11th International Conference on Data Mining (ICDM)*, pages 161–170. IEEE, 2011. [Quoted on page 20]

[121] W. Liao, A. Pizurica, P. Scheunders, et al. Semisupervised local discriminant analysis for feature extraction in hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 51(1):184–198, Jan 2013. [Quoted on page 21]

[122] J. Sotoca and F. Pla. Supervised feature selection by clustering using conditional mutual information-based distances. *Pattern Recognition*, 43(6):2068–2081, 2010. [Quoted on page 21]

[123] Y. Liu, J. Li, A. Plaza, et al. Spectral partitioning for hyperspectral remote sensing image classification. *Geoscience and Remote Sensing Symposium, 2014. IGARSS'14. IEEE 2014 International*, pages 3434–3437, July 2014. [Quoted on pages 21 and 23]

[124] K. Wang, J. Zhang, D. Li, et al. Adaptive affinity propagation clustering. *Acta Automaatica Sinica*, 33(12):1242–1246, 2008. [Quoted on pages 21 and 23]

[125] R. Ammanouil, J. Abou Melhem, J. Farah, et al. Spectral partitioning and fusion techniques for hyperspectral data classification and unmixing. In *2014 6th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, pages 550–553, May 2014. [Quoted on page 21]

[126] N. Ly, Q. Du, and J. Fowler. Reconstruction from random projections of hyperspectral imagery with spectral and spatial partitioning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(2):466–472, 2013. [Quoted on page 21]

[127] S. Holzwarth, M. Habermeyer, R. Richter, et al. Hysens-dais 7915/rosis imaging spectrometers at dlr. In *3rd EARSeL workshop on imaging spectroscopy, Herrsching, 13-16 May 2003*, 2003. [Quoted on page 22]

[128] A. Müeller, R. Richter, M. Habermeyer, et al. Ares: a new reflective/emissive imaging spectrometer for terrestrial applications. In *International Symposium on Remote Sensing*, pages 159–166. International Society for Optics and Photonics, 2003. [Quoted on page 22]

[129] J. Pearlman, S. Carman, C. Segal, et al. Overview of the hyperion imaging spectrometer for the nasa eo-1 mission. *Geoscience and Remote Sensing Symposium, 2001. IGARSS'01. IEEE 2001 International*, 7:3036–3038, July 2001. [Quoted on page 22]

[130] T. Cocks, R. Jenssen, A. Stewart, et al. The hymaptm airborne hyperspectral sensor: the system, calibration and performance. In *EARSEL Workshop on Imaging Spectroscopy, Zurich*, 1998. [Quoted on page 22]

[131] 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. [Quoted on pages 23, 27, 28 and 31]

[132] J. Hartigan and M. Wong. Algorithm as 136: A k-means clustering algorithm. *Applied statistics*, pages 100–108, 1979. [Quoted on page 23]

[133] A. Jain and R. Dubes. *Algorithms for clustering data*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988. [Quoted on page 23]

[134] S.-Z. Li. *Markov random field modeling in computer vision*. Springer-Verlag New York, Inc., 1995. [Quoted on pages 28, 70, 74 and 96]

[135] J. Li, J. 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, 2013. [Quoted on pages 28, 70 and 71]

[136] Q. McNemar. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2):153–157, 1947. [Quoted on page 33]

[137] A. Plaza, P. Martinez, J. Plaza, et al. Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):466–479, 2005. [Quoted on page 40]

[138] S. Serpico, M. D'Inca, F. Melgani, et al. Comparison of feature reduction techniques for classification of hyperspectral remote sensing data. In *International Symposium on Remote Sensing*, pages 347–358. International Society for Optics and Photonics, 2003. [Quoted on page 40]

[139] 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, 2004. [Quoted on page 40]

[140] G. Camps-Valls and L. Bruzzone. Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 43(6):1351–1362, 2005. [Quoted on page 40]

[141] B. Waske, S. van-der Linden, J. Benediktsson, et al. Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 48(7):2880–2889, 2010. [Quoted on pages 40 and 41]

[142] M. Fauvel, C. Dechesne, A. Zullo, et al. Fast forward feature selection for the nonlinear classification of hyperspectral images. *arXiv preprint arXiv:1501.00857*, 2015. [Quoted on page 40]

[143] G. Byrne, P. Crapper, and K. Mayo. Monitoring land-cover change by principal component analysis of multitemporal landsat data. *Remote Sensing of Environment*, 10(3):175–184, 1980. [Quoted on page 40]

[144] L. Ma, M. Crawford, and J. Tian. Local manifold learning-based-nearest-neighbor for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4099–4109, 2010. [Quoted on page 40]

[145] W. Chen, J. Huang, J. Zou, et al. Wavelet-face based subspace lda method to solve small sample size problem in face recognition. *Int. J. Wavelets Multiresolut Inf. Process.*, 07(02):199–214, mar 2009. [Quoted on page 40]

[146] Z. Zhu, S. Jia, S. He, et al. Three-dimensional Gabor feature extraction for hyperspectral imagery classification using a memetic framework. *Information Sciences*, 298:274–287, mar 2015. [Quoted on page 40]

[147] P. Ceccato, N. Gobron, S. Flasse, et al. Designing a spectral index to estimate vegetation water content from remote sensing data: Part 1: Theoretical approach. *Remote Sensing of Environment*, 82(2):188–197, 2002. [Quoted on page 40]

[148] D. Odermatt, A. Gitelson, V. Brando Ernesto, et al. Review of constituent retrieval in optically deep and complex waters from satellite imagery. *Remote sensing of environment*, 118:116–126, 2012. [Quoted on page 40]

[149] Y. Qian, F. Yao, and S. Jia. Band selection for hyperspectral imagery using affinity propagation. *IET Computer Vision*, 3(4):213–222, 2009. [Quoted on page 40]

[150] K. Zhao, D. Valle, S. Popescu, et al. Hyperspectral remote sensing of plant biochemistry using bayesian model averaging with variable and band selection. *Remote Sensing of Environment*, 132:102–119, 2013. [Quoted on page 40]

[151] V. Lombardo, M. Musacchio, and M. Buongiorno. Error analysis of subpixel lava temperature measurements using infrared remotely sensed data. *International Geophysical Journal*, 191(1):112–125, 2012. [Quoted on page 40]

[152] D. Tuia, G. Camps-Valls, G. Matasci, et al. Learning relevant image features with multiple-kernel classification. *IEEE Transactions on Geoscience and Remote Sensing*, 48(10):3780–3791, 2010. [Quoted on pages 40 and 41]

[153] M. Gong, M. Zhang, and Y. Yuan. Unsupervised band selection based on evolutionary multiobjective optimization for hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 54(1):544–557, 2016. [Quoted on page 40]

[154] J. Han, D. Zhang, G. Cheng, et al. Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning. *IEEE Transactions on Geoscience and Remote Sensing*, 53(6):3325–3337, 2015. [Quoted on page 40]

[155] L. Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996. [Quoted on page 40]

[156] R. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. *Machine learning*, 37(3):297–336, 1999. [Quoted on pages 40 and 41]

[157] L. Rokach. Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1-2):1–39, 2010. [Quoted on page 40]

[158] J. Benediktsson, J. Chanussot, and M. Fauvel. Multiple classifier systems in remote sensing: from basics to recent developments. In *Multiple Classifier Systems*, pages 501–512. Springer, 2007. [Quoted on pages 40 and 41]

[159] W. Di and M. Crawford. Active learning via multi-view and local proximity co-regularization for hyperspectral image classification. *IEEE Journal of Selected Topics in Signal Processing*, 5(3):618–628, 2011. [Quoted on pages 40, 41 and 50]

[160] M. Woźniak, M. Graña, and E. Corchado. A survey of multiple classifier systems as hybrid systems. *Information Fusion*, 16:3–17, 2014. [Quoted on pages 40 and 41]

[161] T. Dietterich. Ensemble methods in machine learning. In *Multiple classifier systems*, pages 1–15. Springer, 2000. [Quoted on pages 41 and 43]

[162] S. Kawaguchi and R. Nishii. Hyperspectral image classification by bootstrap adaboost with random decision stumps. *IEEE Transactions on Geoscience and Remote Sensing*, 45(11):3845–3851, 2007. [Quoted on pages 41 and 43]

[163] T. Ho, J. Hull, and S. Srihari. Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(1):66–75, 1994. [Quoted on page 41]

[164] H. Wang, F. Nie, and H. Huang. Multi-view clustering and feature learning via structured sparsity. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 352–360, 2013. [Quoted on pages 41 and 50]

[165] P. Ghamisi, J. Benediktsson, and S. Phinn. Fusion of hyperspectral and lidar data in classification of urban areas. In *2014 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 181–184. IEEE, 2014. [Quoted on page 41]

[166] G. Cheng, J. Han, and L. Guo others. Effective and efficient midlevel visual elements-oriented land-use classification using VHR remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 53(8):4238–4249, Aug 2015. [Quoted on page 41]

[167] J. Han, P. Zhou, D. Zhang, et al. Efficient, simultaneous detection of multi-class geospatial targets based on visual saliency modeling and discriminative learning of sparse coding. *ISPRS Journal of Photogrammetry and Remote Sensing*, 89:37–48, 2014. [Quoted on page 41]

[168] X. Huang, Q. Lu, and L. Zhang. A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 90:36–48, 2014. [Quoted on page 41]

[169] G. Zhang and X. Jia. Feature selection using kernel based local fisher discriminant analysis for hyperspectral image classification. In *2011 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1728–1731. IEEE, 2011. [Quoted on page 41]

[170] X. Jia, B. Kuo, and M. Crawford. Feature mining for hyperspectral image classification. *Proceedings of the IEEE*, 101(3):676–697, 2013. [Quoted on pages 41 and 44]

[171] J. Li, P. Marpu, A. Plaza, et al. Generalized composite kernel framework for hyperspectral image classification. *Geoscience and Remote Sensing, IEEE Transactions on*, 51(9):4816–4829, 2013. [Quoted on page 41]

[172] W. Di and M. Crawford. View generation for multiview maximum disagreement based active learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 50(5):1942–1954, 2012. [Quoted on pages 41 and 56]

[173] J. Ham, Y. Chen, M. Crawford, et al. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):492–501, 2005. [Quoted on page 41]

[174] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995. [Quoted on page 42]

[175] J. Bioucas-Dias and M. Figueeredo. Logistic regression via variable splitting and augmented lagrangian tools. *Tech. Rep.*, 2009. [Quoted on pages 42, 50, 56, 83, 84 and 97]

[176] F. Condessa, J. Bioucas-Dias, C. Castro, et al. Classification with reject option using contextual information. In *2013 IEEE 10th International Symposium on Biomedical Imaging (ISBI)*, pages 1340–1343. IEEE, 2013. [Quoted on pages 42, 46 and 47]

[177] F. Condessa, J. Bioucas-Dias, and J. Kovacevic. Supervised hyperspectral image classification with rejection. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 2600–2603. IEEE, 2015. [Quoted on pages 42, 46 and 47]

[178] P. Bajcsy and P. Groves. Methodology for hyperspectral band selection. *Photogrammetric Engineering & Remote Sensing*, 70(7):793–802, 2004. [Quoted on page 44]

[179] E. Sarhrouni, A. Hammouch, and D. Aboutajdine. Band selection and classification of hyperspectral images using mutual information: An algorithm based on minimizing the error probability using the inequality of fano. In *2012 International Conference on Multimedia Computing and Systems (ICMCS)*, pages 155–159. IEEE, 2012. [Quoted on page 44]

[180] K. Sun, X. Geng, L. Ji, et al. A new band selection method for hyperspectral image based on data quality. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):2697–2703, 2014. [Quoted on page 44]

[181] C.-I. Chang and S. Wang. Constrained band selection for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 44(6):1575–1585, 2006. [Quoted on pages 44 and 45]

[182] C. Chang and C. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. [Quoted on pages 46 and 56]

[183] M. Fauvel, J. Chanussot, and J. Benediktsson. A spatial–spectral kernel-based approach for the classification of remote-sensing images. *Pattern Recognition*, 45(1):381–392, 2012. [Quoted on pages 48 and 70]

[184] G. Mountrakis, J. Im, and C. Ogole. Support vector machines in remote sensing: a review. *Journal of Photogrammetry and Remote Sensing*, in press, 2013. [Quoted on page 50]

[185] T. Kavzoglu and I. Colkesen. A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation*, 11(5):352–359, 2009. [Quoted on page 50]

[186] Y. Tarabalka, J. Benediktsson, J. Chanussot, et al. Multiple spectral–spatial classification approach for hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4122–4132, 2010. [Quoted on pages 70 and 71]

[187] G. Zhang, X. Jia, and J. Hu. Superpixel-based graphical model for remote sensing image mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 53(11):5861–5871, 2015. [Quoted on pages 70, 71 and 94]

[188] L. Fang, S. Li, X. Kang, et al. Spectral-spatial classification of hyperspectral images with a superpixel-based discriminative sparse model. *IEEE Transactions on Geoscience and Remote Sensing*, 53(8):4186–4201, 2015. [Quoted on pages 70 and 71]

[189] J. Li, M. Khodadadzadeh, A. Plaza, et al. A discontinuity preserving relaxation scheme for spectral–spatial hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(2):625–639, 2016. [Quoted on pages 70, 71, 84 and 97]

[190] R. Jain, R. Kasturi, and B. Schunck. *Machine vision*, volume 5. McGraw-Hill New York, 1995. [Quoted on page 70]

[191] S. Warfield, K. Zou, and W. Wells. Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23(7):903–921, 2004. [Quoted on page 70]

[192] J. Cheng, J. Liu, Y. Xu, et al. Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. *IEEE Transactions on Medical Imaging*, 32(6):1019–1032, 2013. [Quoted on page 70]

[193] S. Thayammal and D. Selvathi. A review on segmentation based image compression techniques. *Journal of Engineering Science and Technology Review*, 6(3):134–140, 2013. [Quoted on page 70]

[194] Y. Zhang. A review of recent evaluation methods for image segmentation. In *Sixth International, Symposium on Signal Processing and its Applications, 2001*, volume 1, pages 148–151. IEEE, 2001. [Quoted on page 70]

[195] M. Preetha, L. Suresh, and M. Bosco. Image segmentation using seeded region growing. In *2012 International Conference onComputing, Electronics and Electrical Technologies (ICCEET)*, pages 576–583. IEEE, 2012. [Quoted on page 70]

[196] Y. Boykov, V. Lee, H. Rusinek, et al. Segmentation of dynamic nd data sets via graph cuts using markov models. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2001*, pages 1058–1066. Springer, 2001. [Quoted on page 70]

[197] T. Chan and L. Vese. Active contours without edges. *IEEE transactions on Image processing*, 10(2):266–277, 2001. [Quoted on page 70]

[198] C. Li, C-Y. Kao, J. Gore, et al. Minimization of region-scalable fitting energy for image segmentation. *IEEE Transactions on Image Processing*, 17(10):1940–1949, 2008. [Quoted on page 70]

[199] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741, 1984. [Quoted on page 70]

[200] C. Nieuwenhuis, E. Töppe, and D. Cremers. A survey and comparison of discrete and continuous multi-label optimization approaches for the potts model. *International journal of computer vision*, 104(3):223–240, 2013. [Quoted on pages 70, 74, 94 and 96]

[201] M. Klodt, T. Schoenemann, K. Kolev, et al. An experimental comparison of discrete and continuous shape optimization methods. In *Computer Vision–ECCV 2008*, pages 332–345. Springer, 2008. [Quoted on page 70]

[202] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, 2004. [Quoted on page 70]

[203] Y. Boykov and G. Funka-Lea. Graph cuts and efficient nd image segmentation. *International journal of computer vision*, 70(2):109–131, 2006. [Quoted on page 70]

[204] V. Jeleň, J. Janáček, and Z. Tomori. Mobility tracking by interactive graph-cut segmentation with bi-elliptical shape prior. In *2010 IEEE 8th International Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pages 225–230. IEEE, 2010. [Quoted on page 70]

[205] H. Wang and D. Koller. Subproblem-tree calibration: A unified approach to max-product message passing. In *ICML (2)*, pages 190–198, 2013. [Quoted on pages 70 and 71]

[206] A. Globerson and T.S. Jaakkola. Fixing max-product: Convergent message passing algorithms for map lp-relaxations. In *Advances in neural information processing systems*, pages 553–560, 2008. [Quoted on pages 70 and 71]

[207] T. Werner. A linear programming approach to max-sum problem: A review. *IEEE transactions on pattern analysis and machine intelligence*, 29(7):1165–1179, 2007. [Quoted on pages 70 and 71]

[208] V. Kolmogorov and C. Rother. Minimizing nonsubmodular functions with graph cuts-a review. *IEEE transactions on pattern analysis and machine intelligence*, 29(7):1274–1279, 2007. [Quoted on page 70]

[209] C. Rother, V. Kolmogorov, V. Lempitsky, et al. Optimizing binary mrfs via extended roof duality. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007. [Quoted on pages 70 and 71]

[210] Q. Xu, Q. Chen, S. Yang, et al. Superpixel-based classification using k distribution and spatial context for polarimetric sar images. *Remote Sensing*, 8(8):619, 2016. [Quoted on page 71]

[211] J. Bioucas-Dias, F. Condessa, and J. Kovačević. Alternating direction optimization for image segmentation using hidden markov measure field models. In *IS&T/SPIE Electronic Imaging*, pages 90190P–90190P. International Society for Optics and Photonics, 2014. [Quoted on pages 71, 72, 74, 76, 78, 94 and 96]

[212] V. Kolmogorov and M. Wainwright. On the optimality of tree-reweighted max-product message-passing. *arXiv preprint arXiv:1207.1395*, 2012. [Quoted on page 71]

[213] F. Condessa, J. Bioucas-Dias, and J. Kovačević. Supervised hyperspectral image segmentation: a convex formulation using hidden fields. In *IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS'14)*, Lausanne, Switzerland, June 2014. [Quoted on pages 71, 72, 74, 75, 76 and 94]

[214] R. Achanta, A. Shaji, K. Smith, et al. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, 2012. [Quoted on pages 72, 76, 83, 95 and 97]

[215] M. Afonso, J. Bioucas-Dias, and M. Figueiredo. An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Transactions on Image Processing*, 20(3):681–695, 2011. [Quoted on pages 72, 76 and 77]

[216] J. Marroquin, E. Santana, and S. Botello. Hidden markov measure field models for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(11):1380–1387, 2003. [Quoted on page 72]

[217] A. Figueiredo. Bayesian image segmentation using wavelet-based priors. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005*, volume 1, pages 437–443. IEEE, 2005. [Quoted on page 72]

[218] S. Lefkimmiatis, A. Roussos, M. Unser, and P. Maragos. *Convex generalizations of total variation based on the structure tensor with applications to inverse problems.* Springer, 2013. [Quoted on page 72]

[219] X. Bresson and T. Chan. Fast dual minimization of the vectorial total variation norm and applications to color image processing. *Inverse problems and imaging*, 2(4):455–484, 2008. [Quoted on page 72]

[220] B. Goldluecke, E. Strekalovskiy, and D. Cremers. The natural vectorial total variation which arises from geometric measure theory. *SIAM Journal on Imaging Sciences*, 5(2):537–563, 2012. [Quoted on pages 72 and 75]

[221] C. Zach, D. Gallup, J. Frahm, et al. Fast global labeling for real-time stereo using multiple plane sweeps. In *VMV*, pages 243–252, 2008. [Quoted on pages 72 and 74]

[222] F. Condessa, J. Bioucas-Dias, and J. Kovacevic. Segsalsa-str: A convex formulation to supervised hyperspectral image segmentation using hidden fields and structure tensor regularization. *arXiv preprint arXiv:1504.07028*, 2015. [Quoted on page 72]

[223] D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 271–279, 1989. [Quoted on pages 74, 94 and 96]

[224] P. Combettes and J-Ch. Pesquet. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer, 2011. [Quoted on pages 75, 78 and 79]

[225] J. Kovačević, V. Goyal, and M. Vetterli. *Fourier and Wavelet Signal Processing.* Cambridge University Press, 2015. [Quoted on page 79]

[226] J. Li, J. Bioucas-Dias, and A. Plaza. Semi-supervised hyperspectral image classification based on a markov random field and sparse multinomial logistic regression. In *2009 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 3, pages III–817. IEEE, 2009. [Quoted on pages 83 and 84]

[227] 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, 1990. [Quoted on page 83]

[228] M. Volpi and V. Ferrari. Semantic segmentation of urban scenes by learning local class interactions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–9, 2015. [Quoted on page 86]

[229] M. Fauvel, J. Benediktsson, J. Chanussot, et al. Spectral and spatial classification of hyperspectral data using svms and morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 46(11):3804–3814, 2008. [Quoted on page 94]