

CHAPTER 2

LITERATURE SURVEY

The classification of aerial and satellite remote sensing data has become a challenging problem due to the recent advances in remote sensor technology that led to higher spatial and spectral resolutions. The increased spectral and spatial resolution available has made detailed analysis of the data obtained possible and provided information that now plays an important role in a wide range of environmental disciplines.

Remote Sensing (RS) data is characterized by the dual nature of information it provides: the data can be viewed as a collection of spectra (spectral domain) where each pixel is a vector and the components are the reflectance values at a different wavelengths; or can be treated as a collection of images in the spatial domain at different wavelengths. Early attempts to analyze remote sensing data were limited to spectral domain and ignored the useful spatial information available. Consequently, joint spatial and spectral classifiers have been developed to analyze the remote sensing data better. The use of conventional methods on data with higher spatial and spectral resolution has resulted in the need for dealing with the problem of high dimensionality. Tackling the theoretical and practical problems that arise when dealing with a

vector space of high dimensionality is an active area of research. Also, the evolution in the properties of data itself has mandated the task of seeking methodologies that are based on sensor-invariant assumptions and ability to incorporate multi-source data in the classification process.

This chapter encompasses all the topics that are defined in the objectives of the study and provides an extensive review of the techniques that have been presented over the years. Section 1 of this chapter provides an introduction to the classification process, and reviews the past, present and proposed techniques available. Section 2 surveys the issues in HS data classification and elaborates on the problems of dimensionality reduction and need for incorporation of spatial information. Section 3 reviews the classification techniques available or proposed in the literature and Section 4 concludes the chapter.

2.1 Classification Techniques

Image classification is the process of creating a meaningful digital thematic map from an image data set. Generally, image classification techniques in remote sensing can be divided into supervised and unsupervised based on whether they include a priori knowledge during the decision process by using labeled training samples or not (Jain, 1986). Methods can be further sub-divided into parametric and non-parametric techniques based on whether the classifier assumes that data follows a specific distribution or not. The methodology of pattern classification applied to a particular problem depends on the data, the model of the data and the expected result of analysis (Bezdek, 1981).

2.1.1 Supervised Classification

Supervised classification techniques require training areas to be defined in order to determine the characteristics of each category. A discriminating function is modeled based on the information derived from the training data. Each pixel in the test image, is thus, assigned to one of the categories using the extracted discriminating information (Wilkinson, 2000). The main difference between unsupervised and supervised classification approaches is that the supervised classification requires training data.

2.1.2 *Parametric Classifier*

Parametric approaches to classification make use of a parameterized model of classes in the spectral feature space. These are generally more powerful than the non-parametric methods and lead to higher overall classification accuracy if the data used satisfies the requirements of the model.

2.1.2.1 *Bayesian Framework Based Classifiers*

Classifiers that rely on the Bayesian framework are one of the basic concepts in statistical pattern recognition. The Maximum likelihood classifier which is derived from the Baye's rule when classes have equal priors, is the most common supervised classification technique (Richards et al., 1999). The effectiveness of maximum likelihood classifiers depends on the estimation of the mean vector and covariance matrix for each spectral class which in turn is

dependent upon having sufficient number of training samples for each of the classes in consideration. Performance of these classifiers is affected, resulting in lower classification accuracy, if adequate multivariate statistical model is not available. The other drawback of this method is the computational cost required to classify each pixel. This is particularly important in circumstances where data to be classified are measured in a large number of spectral bands, or include many spectral classes to be discriminated. If the form of density function is unknown, it is possible to operate in a non-parametric mode.

2.1.3 Non-parametric Classifiers

These types of classifiers make no assumptions on the probability distribution and are often considered robust because they may work well for a wide variety of class distributions, as long as the class signatures are reasonably distinct. A wide variety of non-parametric spectral classifiers are available. These consist of statistical methods such as the parallelepiped classifier, minimum distance classifier and non-statistical methods such as neural networks, support vector machines, and decision tree classifiers.

2.1.3.1 Parallelepiped and Minimum Distance Classifier

The parallelepiped classifier (or box classifier) is perhaps the simplest of all non-parametric classification systems because this requires the least information from the user of supervised classification methods. In this method for each class specified, the user provides an estimate of minimum and

maximum values of each of the features used, from the training data, and an unknown pixel is classified if it lies inside any of the parallelepipeds. The problem with this technique is overlapping parallelepipeds. Several enhancements have been suggested to increase the utility of this classifier (Addington, 1975; Lillesand et al., 1994).

Another simple non-parametric classification method is the minimum distance classifier. This method uses the minimum distance between the pixel and centroid of the training class and a pixel is assigned to the class with the lower distance measure. Classification can be performed by using distance measures other than Euclidean (Wacker et al., 1972).

Methods of table-lookup classification, linear discriminant function can also be found in the literature (Richards et al., 1999). These set of classifiers though are simple and computationally efficient, have certain limitations. Sensitiveness to different degrees of variance in the spectral response data make these classifiers unusable in applications with narrow spectral classes and have high spectral variance.

2.1.3.2 *Artificial Neural Networks*

Neural Network (NN) classifiers have been in use for supervised classification of remote sensing data since the late 1980s. A NN is a form of artificial intelligence that imitates some function of the human brain and are general purpose computing tools that can solve complex non-linear problems (Fischer, 1996). The network comprises a large number of simple processing

elements linked to each other by weighted connections to a specified architecture. These networks learn from the training data by adjusting the connection weights (Bishop, 1995). They have been used in remote sensing and image analysis (Benediktsson et al., 1990; Foody et al., 1997). In remote sensing applications, the multi-layered feed forward network, and the Kohonen networks are generally used. These networks differ from each other in their approach to classifying remotely sensed data.

NNs have the advantage over traditional statistical methods in that they are distribution-free as they do not rely on any underlying statistical distribution of data and this makes them particularly valuable for multi-source applications. However, the NN approach can be computationally complex and requires a large number of training samples. Consequently the performance of a NN strongly depends on representative training samples, whereas statistical models require an appropriate model for each class (Benediktsson et al., 1990). Other limitations identified with back-propagation are slow convergence of the error back-propagation, the potential convergence to a local minimum and the incapacity to detect that an input sample has fallen in a region of the feature space without training data (Serpico et al., 1996; Bruzzone et al., 2004).

2.1.3.3 *Decision Tree Classifiers*

Decision Tree (DT) induction algorithms have been widely used in pattern recognition field for solving classification and related tasks (Quinlan, 1987). The use of DT for classification in remote sensing problems is described

in Swain et al., 1977. DT offers several advantages over classification algorithms traditionally used for land cover mapping. One advantage is the ability to effectively use both categorical and continuous predictor data with different measurement scales. Other advantages include the ability to handle non-parametric training and predictor data, good computational efficiency, and an intuitive hierarchical representation of discrimination rules (Pal et al., 2003). A major limitation of DT classifiers is the requirement of generation of rules that heavily depend on the knowledge from experts (Safavian et al., 1991).

2.1.3.4 Support Vector Machines

Support Vector Machines (SVM) are classification and regression methods which have been derived from statistical learning theory (Vapnik, 1995). The concept is based on optimal linear separating hyperplane that is fitted to the training patterns of two classes within a multi-dimensional feature space. The optimization problem that has to be solved relies on structural risk minimization and is aiming at a maximization of the margins between the hyperplane and closest training samples. If the two classes are non-separable, SVMs employ the kernel trick where a positive definite kernel function is used to map the input data into a high dimensional transformed feature space.

The pioneering work of Gualtieri (Gualtieri et al., 2000) related to the use of SVM for classification of HS images has been followed by several researchers to analyze the theoretical properties and empirical performances of SVM applied to different kinds of classification problems. The SVM has become

popular as it has often been observed in many studies to classify data sets with an accuracy equivalent or higher than that derived from the application of range of alternative classifiers (Huang et al., 2002; Foody et al., 2004; Melgani et al., 2004). The success of SVMs is mainly due to the fact that SVMs do not require an estimation of statistical distributions of classes to carry out the classification task, but they define the classification model by exploiting the concept of margin maximization. Compared to NN approaches, SVMs exhibit higher generalization capability, robustness to Hughes phenomenon, lower effort required for model selection in the learning phase (Joachims, 1998) and optimality of solution obtained by the learning algorithm. Further details of SVM based classifiers are discussed in chapter 3.

2.1.4 Unsupervised Classification

In contrast to supervised approaches, which construct the decision boundaries from training data, unsupervised algorithms are based on set of unlabelled data. These unsupervised methods can be viewed as techniques of identifying natural groups or clusters in image data. The pixels within a cluster or group are more similar to each other than those pixels belonging to other clusters (Jain, 2000). Determination of clusters is performed by estimating the distances or comparison of variance within and between the clusters. The most popular cluster algorithms used in remote sensing image classification are ISODATA, k-means and SOM self organizing feature maps, an unsupervised neural classification method. Although the unsupervised procedures seem

flexible, the results of such methods are generally inferior to those achieved by supervised methods. This is partly because most real-world features exhibit complexity in their nature, and hence might not be easily separable in terms of their spectral signatures. In addition, the assumption behind the unsupervised approach that the pixels belonging to particular class will have similar spectral values in spectral space, and all classes are relatively distinct from each other in features space is difficult to satisfy in practice.

2.1.5 Advanced Classification Techniques

Recent studies on difficult classification problems using multi-source and HS data have led to composite (or hybrid) and ensemble classifiers (Crawford et al., 2003; Ham et al., 2005). Composite classifiers are based on combining multiple individual methods usually in a stacked topology (Wolpert, 1992) and make use of trained models whose expertise is combined to obtain an optimal classifier. Ensemble classification takes a slightly different approach where hundreds of classifiers are built and their decisions combined usually by a plurality vote or more sophisticated methods like consensus theory. In general, though these classifiers present viable alternatives, one drawback is to have to handle different learning algorithms resulting in an increase of processing complexity. Their effectiveness relies very much on the combining technique and ensemble does not always give a more accurate classification (Foody et al., 2007).

Numerous concepts for generating classifier ensembles have been introduced, boosting and bagging being the main approaches (Breiman, 1996; Freund et al., 1996). Though boosting can these techniques avoid over fitting with noiseless data and reduce the variance and the bias of the classification, sensitivity to noisy training data and computation time are some drawbacks.

2.2 Hyperspectral Data Classification

This section presents the classic problems encountered in HS data analysis and surveys the active research in this field.

Since the emergence of HS sensor technology in the recent decade, theoretical and experimental research revealed the issues in hyperspectral data analysis (Lee et al., 1993; Jimenez et al., 1998). Treatment of each pixel as a vector where each component contains specific information from a particular channel implies that size of vector is related to the number of bands that the sensor can collect. For HS data, 200 or more bands of the same scene are common and this translates to a pixel vector space of large dimensions. With the increasing dimensionality, many challenges arise as the space and geometric concepts self-evident in lower dimensions are necessarily not applicable. The problems introduced are sometimes referred to as curse of dimensionality. Donoho, 2000 gives a survey of high-dimensionality data analysis.

The main issues in analyzing HS data can be summarized as:

1. The second-order statistic plays an important role in classification. Lee et al., 1993 show that using the variance of multivariate data led to significantly better classification results than considering only the mean as the dimensionality of data increases.
2. In high-dimensional space, normally distributed tend to concentrate in the tails, while uniformly distributed data tend to the corners. This is also commonly referred to as "concentration of measure" phenomenon (Donoho, 2000).
3. Hughes effect: with a limited number of training samples, there is a classification accuracy penalty as the number of features increase beyond a certain point (Hughes, 1968).
4. As shown in Fukunga, 1990, the number of training samples required for good estimation of parameters is linearly proportioned to the dimensionality and quadratic ally related to the order of the classifier.
5. Large spatial variability of each land-cover class and a likelihood of the presence of a high proportion of mixed pixels.

Issue 1 is difficult to satisfy in most practical scenarios, from 2 the local neighborhood is very likely to be empty and makes statistical estimation inadequate. The requirement that available training samples proportionally increase with dimensionality is not always satisfiable thus bringing up the need to placate Hughes effect. Dimensionality reduction, discussed in the following section, thus appears to be a major factor in HS classification. The internal

spatial variability and increase in the possibility of recording multiple thematic classes in each fundamental spatial unit, the pixel, is inherent side effect of increased spatial resolution of sensors. Tackling this problem requires incorporating spatial signatures in the classification process. However, in literature several HS data classification methods are presented that exclude dimensionality reduction and rely on non-parametric techniques or clever stacking of multiple classification schemes.

2.2.1 Dimensionality Reduction

Dimensionality reduction is a technique aimed at reducing the dimensionality of data by mapping them onto another space of a lower dimension, without discarding any meaningful information. Feature selection is the technique of selecting a subset of relevant features, while feature extraction is a method of combining features - both in order to obtain relevant representation in a lower-dimensional space. Feature-selection techniques perform a reduction of spectral channels by selecting a representative subset of original feature by employing a selection criterion and a search strategy. The former aims at assessing the discrimination capabilities of a chosen subset according to statistical distance measures and the latter defines an optimization approach to identify the best subset of features according to the selection criterion. Since the identification of optimal solution is computationally unfeasible, techniques that lead to sub-optimal solutions are normally used. Among the search strategies proposed in literature, Sequential Forward Search

(Kittler, 1978), the more effective sequential forward selection (Pudil et al., 1994) and Steepest Ascent (Serpico et al., 2002) are notable.

Feature extraction approaches address the problem of feature reduction by transforming the original feature space into a feature space of lower dimensionality which contains most of the original information. Transformations based on statistical analysis have proved be useful for classification (Chiang, 2001; Keshava, 2004). Transformation can be supervised, where training set data are available and the transformation is performed according to the properties of training set or unsupervised where the algorithm works directly on the data without any ground truth. The effectiveness of supervised method is correlated with how well the training set represents the whole data set, can be extremely time consuming and hence is not widely used. Unsupervised case does not focus on class discrimination, but looks for another representation of data in a lower-dimensional space. Principal Component Analysis (PCA), where data is projected into a subspace that minimizes the reconstruction error in the mean square sense was shown to be competitive for the purpose of classification regardless of the their theoretical limitations (Landgrebe, 1980; Lennon et al., 2001). The advantages of PCA relatively lower complexity and absences of parameters. Methods such as Decision Boundary Feature Extraction method (DBFE) (Lee et al., 1993) and Discriminate Analysis Feature Extraction (DAFE) (Lee et al., 1993; Jimenez et al., 1998) proved to be very effective when the training set is sufficiently large, capable of providing a minimum number of transformed features. Spectral unmixing for feature extraction prior to

supervised classification of hyperspectral data using SVMs was explored in Dopido et al., 2011.

The above mentioned dimension reduction methods transform the high-dimensional features space to a low-dimensional space by linear or nonlinear transformations, and will result in the loss of original physical interpretation of the individual bands in the HS image. Unfortunately, information content in HS images does not always match such projections (Kaewpijit, 2003).

Another set of methods view the problem of dimensionality reduction from the perspective of redundancy among multiple narrowly separated bands and noise reduction. Redundancy can cause convergence instability of models and variations due to noise in redundant data propagate through a classification or discrimination model. The solution that these of methods try to arrive at is to reduce the number of bands by eliminating redundancy and selecting only those bands that might provide the highest discriminating factors while excluding the formation of a new features and are collectively referred to as band selection or band reduction methods.

Band selection can be conducted based on the availability of class information. When class information is known, supervised band selection can be done. Whereas, unsupervised selection methods have to be implemented to find the most informative and distinctive bands in the absence of priori class information.

A large group of supervised band-selection algorithms calculate class separability where a subset of bands is selected using some distance measure as a

metric. Gill et al., 1991 and Han et al., 2001 used the regression method for selecting a subset of bands. In Wittten et al., 2000 instance based method uses inverse Euclidean distance weighting of the k-nearest neighbors to predict any number of continuous variables which was then used to pick the optimal subset of bands.

Several band reduction techniques are based on spectral similarity measures that analyzed a scalar measure of overlap between multiple vector signals. Bands that are deemed similar are coalesced or replaced by the best representative sample. The critical component of these algorithms is the operator known as distance metric that mathematically quantifies the similarity between two spectra. Numerous distance metrics have been proposed and in use. Euclidean distance is the most popular metric as it has an intuitive appeal to evaluate the proximity of object in two- or three-dimensional space and works well when a data set has compact or isolated clusters. City-block distance (also called Hamming distance, Manhattan distance or Taxicab metric) between two vectors is the sum of differences of their corresponding components. This metric is not as popular as the Euclidean one, but has the big advantage of being computationally cheap. Spectral Angle Mapper (SAM), another popular measure, determines the spectral similarity between two spectra by computing the angular difference of the two pixel vectors (Kruse et al., 1993). Other well known measures include Jaccards coefficient (Salton et al., 1983), Jensen-Shannon divergence (Rao, 1982), Kullback-Liebler divergence measure (Kullback, 1969).

Similarity measurements based on Information Theory were also used in literature. Mutual information (MI) that measures the joint entropy of two random variables was introduced as a similarity measure between images simultaneously by Maes et al., 1997 and Viola et al., 1997. Guo et al., 2008 found that the selection based on simple criterion of only retaining bands with high mutual information helps reduce the data set size. Shinjin et al., 2009 presented a band grouping method based on conditional mutual information between adjacent bands in the pre-classification phase and reported some computational cost savings. Sotoca et al., 2006 proposed a band reduction method using correlation among bands based on MI as well. MI does not perform well when the features were highly correlated and the direct way of selecting features by jointly maximizing the MI would suffer from combinatorial explosion.

Several other band reduction techniques were reported as well. Serpico et al., 2007 put forward a procedure to extract spectral channels of variable bandwidths and spectral positions. Huang et al., 2005 presented a feature weighting method for a band selection, which was based on pair-wise separability criterion and matrix coefficient analysis through principal component analysis. Huang et al., 2006 proposed a hybrid genetic algorithm for feature selection, which consisted of the local and global searches. Yang et al., 2011 proposed a supervised band selection algorithm that uses known class signatures only without examining the original bands.

The band reduction methods though most likely help improve classification accuracy; rigorous quantitative analysis might be required to choose the appropriate method to obtain the best tradeoff between accuracy versus computational requirements.

2.2.2 Incorporating Spatial Domain

Over the years the spectral as well as the spatial resolution of remote sensing sensors has increased greatly. This increase of spectral and spatial resolution added to the challenges of handling HS data. The fact that each spectral signature measures the response of multiple underlying elements at each spatial unit adds to the spatial variability and increases the occurrence of mixed or heterogeneous pixels. Classification of the heterogeneous areas cannot be accomplished by relying on spectral properties of the data alone and will need to incorporate the information related to spatial arrangement of the pixels in the neighborhood.

Spatial information, such as texture and context has been widely used. Texture depends on the neighborhood of the pixel and the contextual information represents external or non-remotely-sense information such as elevation values or data derived from soil or geology maps. Textural features contain information about spatial distribution of tonal variation within a band. Contextual features contain the information derived from blocks of image data surrounding the area being analyzed.

A number of textural measures have been proposed in literature, including the grey-level co-occurrence matrix (Harlick et al., 1973), autoregressive models (Frankot et al., 1987), Fourier transform and Fractal based texture (Keller et al., 1989). Recently, wavelet-based texture features (Fukuda et al., 1999) have been used in classification of remotely sensed data. A considerable amount of research has been done to investigate the effectiveness of texture features and a number of studies have shown the classification accuracy can be improved by using texture features in combination with the spectral features (Barber et al., 1991; Mather, 1999).

Contextual information is generally calculated by using a moving window or image segmentation. With moving windows, contextual information is extracted from the pixel's immediate neighborhood by imposing a search window, computing a contextual parameter within that window and assigning that value to the original pixel. Contextual data can be incorporated either during the classification routine or post-classification (Kartikeyan et al., 1994; Barnseley et al., 1996). Post-classification context integration generally involves filtering with a majority filter, whereby each pixel is recorded to the majority class of a neighborhood (Groom et al., 1996). Not only does this operation reduce the salt-and-pepper effect typical for per-pixel classifier, it also results in larger classification units that might adhere more to the human perception of land cover (Stuckens et al., 2000) which depends on the neighborhood of the pixel has been widely used. This increase of spectral and spatial information enables us to apply more deterministic approaches to the classification process

rather than the using only basic statistical assumptions. One of the challenges of HS processing is the large spatial variability within each pixel, resulting in mixed pixels, due to the fact that each spectral signature generally measures the response of multiple underlying components at each site.

Segmentation approach on the other hand, divides the image into contiguous clumps of pixels called segments or regions. All pixels of a segment or region are then assumed to belong to the same information class. Object-based classification builds on this by grouping the spatially adjacent pixels into spectrally homogenous objects first and then conducts classification on objects as the minimum processing units. Kettig et al., 1976 proposed and built a spectral-spatial classifier called extraction and classification of homogenous objects (ECHO) based on this idea.

Another possible approach in order to deal with the high-dimensional nature of HS data is to consider the geometrical properties rather than the statistical properties of the classes. Mathematical morphology provides a well-established theory for analyzing spatial relationship between sets of pixels. In Pesaresi et al., 2001 the concept of morphological profiles was proposed and used for segmentation of high-resolution remotely sensed images. Derivatives of morphological profiles are used for automatic detection of structures in the images (Pesaresi et al., 2007) and as class signatures for the purpose of classification (Plaza et al., 2005).

2.3 Hyperspectral Classification Methods in Literature

The rest of this chapter reviews the recent research done to tackle issues of HS data mentioned in section 2.2.2. Non-SVM literature is presented followed by SVM related literature.

In Camp-Valls et al., 2007 semi-supervised graph based method to handle the high number of unlabeled samples present in the image. Though the method is shown to provide relatively good performance and have good robustness and stability, choosing the appropriate composite kernel is not intuitive and computational complexity gets too high when the sample size is large. Chang et al., 2003 used a combination of greedy modular Eigen space and positive Boolean function algorithms to come up with a new supervised classification technique. Anthony et al., 2006 investigated the use fuzzy unsupervised learning vector quantization NN in the context of HS AVIRIS image classification and have shown that unlike conventional HS data processing methods, end members for a given scene which can be difficult to determine with confidence are not required for neural analysis while achieving good classification accuracies. In Jonathan et al., 2008, Adaboost and Random Forest tree-based ensemble classifiers were evaluated against NN classifier and demonstrated that the ensemble classifiers were effective in dealing with classification of airborne HS imagery. A novel pair wise decision tree framework is proposed for HS classification, where no partitions and clustering are needed in Chen et al., 2007. Benediktsson et al., 2003 used a Conjugate-Gradient NN for classification of aircraft scanned HS remote sensing data. In

Mianji et al., 2011, relevance vector machine was used to develop a new supervised classification method for HS images. HS image classification was achieved by the using features reconstructed as the sum of lower order intrinsic mode functions of each band in Demir et al., 2011. A semi-supervised learning framework based on the tri-training scheme and decision fusion is proposed for classification in Huang et al, 2012. A genetic fuzzy-rule based hyperspectral classification system is developed in Stavrakoudis et al., 2012.

In the last decade, since the introduction of SVMs for classification of remote sensing data, many studies have been published in remote sensing literature on the application of SVMs for HS data analysis and relative effectiveness of SVMs over other classification algorithms. In Gualitieri et al., 2000 the authors introduced SVM for classification of AVIRIS 224 spectral band data and discussed the robustness of SVM to the curse of dimensionality. Melgani et al., 2004 proposed a theoretical discussion and experimental analysis for understanding the potential of SVM as a spectral classifier. Camp-Valls et al., 2005 presented a kernel based framework and assessed the performance of the SVM against regularized radial basis function NN, kernel Fischer discriminant analysis, and regularized Ada-boost. Dixon et al., 2008 presented a study on the comparison between SVM and NN for classification of RS data. Foody et al., 2004 evaluated a single SVM against a series of classifiers that are widely used in RS, with particular regard to the effect of training set size on classification accuracy and noted that most accurate classifications are obtained with the SVM approach.

Several advanced SVM-based classifiers have been developed for facing complex problems related to the properties of HS images. Semi-supervised SVM classifiers have been proposed for dealing with strongly ill-posed problems, where very few training samples are available. Bruzzone et al., 2005 presented a semi-supervised SVM based on transductive inference that exploits a specific iterative algorithm that gradually searches a reliable separating hyperplane in the kernel space that incorporates both labeled and unlabeled samples in the training phase. Laplacian SVM technique that adopts an additional regularization term on the geometry of both labeled and unlabeled samples by using Laplacian graph was introduced by Belkin et al., 2006. Marconcini et al., 2009 presented a novel composite semi-supervised SVM that exploits the use of unlabeled data for increasing the reliability of the training phase and applied composite kernel functions for utilizing spectral and spatial information simultaneously.

Studies that include spatial-context information of single pixel in the classification process are abundant as well. Farag et al., 2005 proposed a framework for applying maximum a posteriori estimation principle in remote sensing image segmentation, which incorporates contextual and geometrical information in the SVM classification process by means of Markov Random Field (MRF). Bruzzone et al., 2009 described a context-sensitive semi-supervised SVM that includes the usage of contextual information during the training phase in order to improve the robustness to possible mislabeled training patterns. Fauvel et al., 2008 proposed a method based on fusion of morphological information and the original HS data for urban area classification

using SVMs. In Tarabalka et al., 2010 spatial contextual information by means of MRF regularization was used in refining the classification results obtained by a probabilistic pixel-wise SVM HS image classifier. Another interesting approach is the application of active learning concepts with SVMs as discussed in Mitra et al., 2004.

2.4 Summary

In this chapter, an introduction to classification methods and review of issues and techniques in classification of hyper-dimensional remote sensing data was presented along with literature survey of adoption of SVMs. In addition, the need for dimensionality reduction and incorporation of spatial information in the classification process was also discussed. It can be concluded that unless the dimensionality of data sets can be reduced and spatial correlation between adjacent pixels is included, classification accuracy will suffer. Many of the proposed methods, though are proven to be effective, the computation complexity is too cumbersome to be widely adopted.

To overcome the unacceptable computational cost while not trading off the classification accuracy, this thesis proposes an integrated sensor invariant unsupervised dimensionality reduction approach and joint spectral-spatial classifier based on the use of morphological profiles for classification of high resolution images. Basics of SVM and extended morphological tools will be discussed in Chapter 3.