HIERARCHICAL SEGMENTATION, OBJECT DETECTION AND CLASSIFICATION IN REMOTELY SENSED IMAGES

A THESIS

SUBMITTED TO THE DEPARTMENT OF COMPUTER ENGINEERING

AND THE INSTITUTE OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

By

Hüseyin Gökhan Akçay

July, 2007

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Asst. Prof. Dr. Selim Aksoy(Advisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Asst. Prof. Dr. Pınar Duygulu Şahin

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Enis Çetin

Approved for the Institute of Engineering and Science:

Prof. Dr. Mehmet B. Baray
Director of the Institute

ABSTRACT

HIERARCHICAL SEGMENTATION, OBJECT DETECTION AND CLASSIFICATION IN REMOTELY SENSED IMAGES

Hüseyin Gökhan Akçay
M.S. in Computer Engineering
Supervisor: Asst. Prof. Dr. Selim Aksoy
July, 2007

Automatic content extraction and classification of remotely sensed images have become highly desired goals by the advances in satellite technology and computing power. The usual choice for the level of processing image data has been pixel-based analysis. However, spatial information is an important element to interpret the land cover because pixels alone do not give much information about image content.

Automatic segmentation of high-resolution remote sensing imagery is an important problem in remote sensing applications because the resulting segmentations can provide valuable spatial and structural information that are complementary to pixel-based spectral information in classification. In this thesis, we first present a method that combines structural information extracted by morphological processing with spectral information summarized using principal components analysis to produce precise segmentations that are also robust to noise. First, principal components are computed from hyper-spectral data to obtain representative bands. Then, candidate regions are extracted by applying connected components analysis to the pixels selected according to their morphological profiles computed using opening and closing by reconstruction with increasing structuring element sizes. Next, these regions are represented using a tree, and the most meaningful ones are selected by optimizing a measure that consists of two factors: spectral homogeneity, which is calculated in terms of variances of spectral features, and neighborhood connectivity, which is calculated using sizes of connected components. Experiments on three data sets show that the method is able to detect structures in the image which are more precise and more meaningful than the structures detected by another approach that does not make strong use of neighborhood and spectral information.

Then, we introduce an unsupervised method that combines both spectral and structural information for automatic object detection. First, a segmentation hierarchy is constructed and candidate segments for object detection are selected by the proposed segmentation method. Given the observation that different structures appear more clearly in different principal components, we present an algorithm that is based on probabilistic Latent Semantic Analysis (PLSA) for grouping the candidate segments belonging to multiple segmentations and multiple principal components. The segments are modeled using their spectral content and the PLSA algorithm builds object models by learning the object-conditional probability distributions. Labeling of a segment is done by computing the similarity of its spectral distribution to the distribution of object models using Kullback-Leibler divergence. Experiments on three data sets show that our method is able to automatically detect, group, and label segments belonging to the same object classes.

Finally, we present an approach for classification of remotely sensed imagery using spatial information extracted from multi-scale segmentations. Different structuring element size ranges are used to obtain multiple representations of an image at different scales to capture different details inherently found in different structures. Then, pixels at each scale are grouped into contiguous regions using the proposed segmentation method. The resulting regions are modeled using the statistical summaries of their spectral properties. These models are used to cluster the regions by the proposed grouping method, and the cluster memberships assigned to each region at multiple scales are used to classify the corresponding pixels into land cover/land use categories. Final classification is done using decision tree classifiers. Experiments with three ground truth data sets show the effectiveness of the proposed approach over traditional techniques that do not make strong use of region-based spatial information.

Keywords: Remote sensing images, hierarchical segmentation, unsupervised object detection, multi-scale classification, spatial information.

ÖZET

UYDU GÖRÜNTÜLERİNDE SIRADÜZENSEL BÖLÜTLEME, NESNE SEZİMİ VE SINIFLANDIRMA

Hüseyin Gökhan Akçay
Bilgisayar Mühendisliği, Yüksek Lisans
Tez Yöneticisi: Yard. Doç. Dr. Selim Aksoy
Temmuz, 2007

Yüksek çözünürlükteki uzaktan algılamalı uydu görüntülerinde bölütleme kent uygulamalarında önemli bir problemdir çünkü elde edilen bölütlemelerle sınıflandırma için piksel tabanlı spektral bilginin yanında uzamsal ve yapısal bilgiler elde edilebilir. Bu tezde, biçimbilimsel işlemlerle çıkarılan yapısal bilgi ve ana bileşenler analizi (ABA) ile özetlenen spektral bilgi kullanılarak gürültüden etkilenmeyen bölütler elde eden bir yöntem sunduk. Yapılan deneyler yöntemin görüntü üzerinde komşuluk bilgisini ve spektral bilgiyi beraber kullanmayan başka bir yönteme göre daha düzgün ve anlamlı yapılar bulduğunu göstermiştir.

Daha sonra, birden fazla ABA bandında ortaya çıkan sıradüzensel bölütlemeler arasından anlamlı yapılara denk gelen bağlı bileşenleri otomatik olarak seçmek için ise ögreticisiz bir yöntem sunulmuştur. Bu problem, verilen birden çok nesne/yapı için farklı sıradüzensel bölütlemelerden gelen çok sayıda aday bölgeden oluşan uzayda bir gruplama problemi olarak görülebilir. Bu amaçla, gruplama problemini çözmek için olasılıksal Gizli Degişken Analizi (OGDA) kullanmaktayız. Yapılan deneyler yöntemin aynı nesne sınıfına ait bölütleri otomatik olarak belirleyebildiğini göstermektedir.

Son olarak, birden fazla seviyede bölütleme sonucunda elde edilen bölgeleri kullanarak bir sınıflandırma yöntemi sunmaktayız. Farklı yapılardakı farklı ayrıntıları yakalamak için farklı yapısal öğe boyut aralıkları kullanılarak bir görüntünün birden fazla ölçekte temsil edilmesi amaçlanmaktadır. Her bir ölçekte bölütleme yapılmakta ve ortaya çıkan her bir bölüt içerisindeki piksellerin spektral özelliklerinin bir özeti ile temsil edilmektedir. Bu temsiller kullanılarak bölütler önerilen gruplama yöntemi ile gruplanmakta ve bölütlerin farklı ölçeklerdeki grup etiketleri piksellerin sınıflandırılmasında kullanılmaktadır. Son sınıflandırma karar ağacı sınıflandırıcısı ile yapılmaktadır. Yapılan deneyler

yöntemin uzamsal bilgiyi etkili bir şekilde kullanmayan klasik yönteme göre üstünlüğünü göstermektedir.

Anahtar sözcükler: Uydu görüntüleri, sıradüzensel bölütleme, öğreticisiz nesne sezimi, çok ölçekli sınıflandırma, uzamsal bilgi.

Acknowledgement

I would like to thank my advisor Dr. Selim Aksoy for his invaluable helps, expert guidance and motivation throughout my MS study. It has always been a pleasure to study with him.

I would like to thank Dr. Pınar Duygulu Şahin and Dr. Enis Çetin for reading my thesis and offering constructive comments.

I would like to thank Retina team members and my other friends for their moral support and nice friendship.

I would like to express my deepest gratitute to my family for their persistent support, understanding, love and for always being by my side. They make me feel stronger. I am incomplete without them.

Finally, I would like to thank Dr. David A. Landgrebe and Mr. Larry L. Biehl from Purdue University, Indiana, U.S.A., for the *DC Mall* data set, and Dr. Paolo Gamba from the University of Pavia, Italy, for the *Centre* and *University* data sets. This work was supported in part by the TUBITAK CAREER Grant 104E074 and European Commission Sixth Framework Programme Marie Curie International Reintegration Grant MIRG-CT-2005-017504.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Due to the constantly increasing public availability of high-resolution data sets, remote sensing image analysis has been an important research area for the last four decades. For example, nearly 3 terabytes of data are being sent to Earth by NASA's satellites every day [24]. There is also an extensive literature on classification of remotely sensed imagery using parametric or nonparametric statistical or structural techniques [47]. Advances in satellite technology and computing power have enabled the study of multi-modal, multi-spectral, multi-resolution and multi-temporal data sets for applications such as urban land use monitoring and management, GIS and mapping, environmental change, site suitability, agricultural and ecological studies.

The usual choice for the level of processing image data has been pixel-based analysis in both academic and commercial remote sensing image analysis systems. However, a recent study [74] that investigated classification accuracies reported in the last 15 years showed that there has not been any significant improvement in the performance of classification methodologies over this period. We believe that the reason behind this problem is the fact that there is a large semantic gap

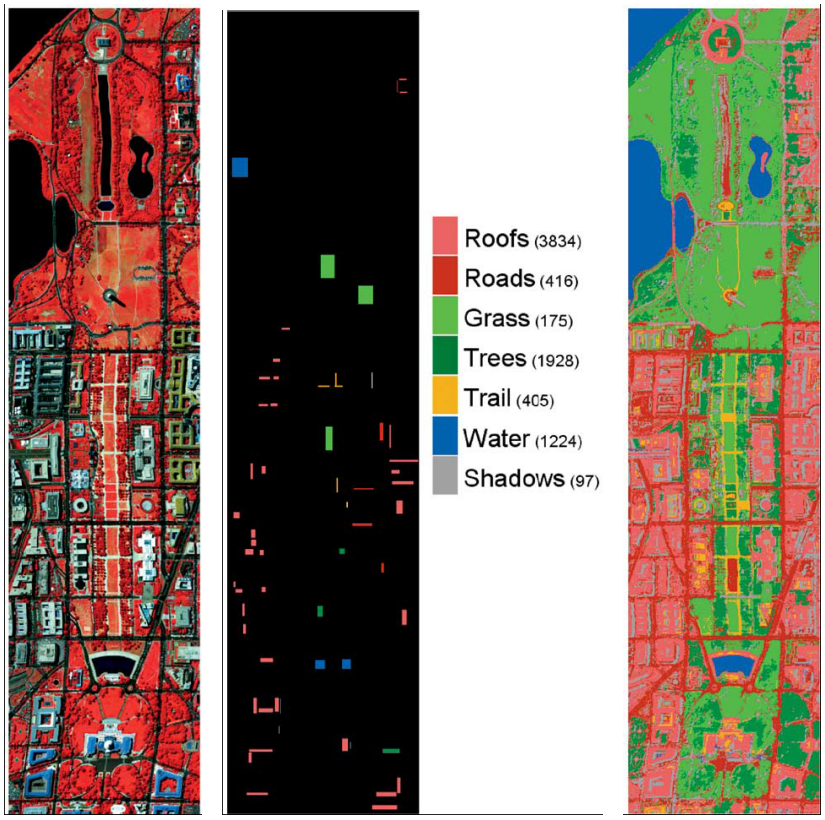
between the low-level features used for classification and the high-level expectations and scenarios required by the users. This semantic gap makes a human expert's involvement and interpretation in the final analysis inevitable and this makes processing of data in large remote sensing archives practically impossible.

The use of only pixel level features often does not meet the expectations as the resolution increases. Even though high success rates have been published in the literature using limited ground truth data, visual inspection of the results can show that most of the urban structures still cannot be delineated as accurately as expected. For example, if there is an area crowded by buildings in an image, classifying most of the pixels in that area as building while missing most of the roads and other small structures around them will still result in a high success rate if limited amount of pixel level ground truth is used for evaluation. In Figure 1.1, an example classification map with a pixel level quadratic Gaussian classifier is shown [3]. The classification accuracy is 93.9677% which is relatively high. However, most of the tiles on the left are merged, the boundaries are not explicit, and the thin asphalts and shadows between the tiles are also classified as tiles. These erroneous areas are not reflected in the numerical accuracy since the ground truth for testing such areas is not enough for a reliable evaluation. For example, no testing data for these thin roads and shadows are available. So, classifying these areas erroneously does not reduce the numerical classification accuracy. In Figure 1.2, an example classification map obtained in a recent study is shown [9]. The classification accuracy is 98.5% which is also relatively high. However, the classification map includes many wrongly classified pixels for all classes such as the roads and water on the upper part. The same reason also holds for this disagreement between the quantitative accuracy and the qualitative result such that testing data is insufficient for reliable evaluation.

The commonly used classifiers model image content using distributions of pixels in spectral or other feature domains by assuming that similar land cover structures will cluster together and behave similarly in these feature spaces. However, the assumptions for distribution models often do not hold for different kinds of data. Even when nonlinear tools such as neural networks or multi-classifier systems are used, the use of only pixel-based data often fails the expectations.

(a) False-color (b) Classification map

Figure 1.1: An example classification map with a pixel level quadratic Gaussian classifier for the *Centre* data set whose false color image is shown on the left. The corresponding ground truth maps for training and testing and class color codes are listed in Figure 3.2. (Images taken from [3].)



(a) False-color (b) Ground truth (c) Classification map

Figure 1.2: An example classification map (shown in (c)) obtained by a recent method [9] for the *DC Mall* data set whose false color image is shown in (a). The corresponding ground truth map and class color codes are listed in (b). 40 samples per class were used for training and the remaining were used for testing. (Images taken from [9].)

We believe that spatial and structural information should also be used for more intuitive and accurate classification. However, image segmentation is still an unsolved problem. Even though several approaches such as region growing, Markov random field models, and energy minimization have been shown to be useful in small data sets with limited detail, no generally applicable segmentation algorithm exists.

In this thesis, first, we describe a segmentation method for remote sensing images that uses the neighborhood and spectral information as well as the morphological information. First, principal components analysis (PCA) [23] is performed to extract the top principal components that represent the 99% variance of the whole data. Next, morphological opening/closing by reconstruction operations are performed on each PCA band separately using structuring elements in increasing sizes. These operations produce a set of connected components forming a hierarchy of regions for each PCA band. Then, the components at different levels of the hierarchy are evaluated as candidates for meaningful structures using a measure that consists of two factors: spectral homogeneity, which is calculated in terms of variances of multi-spectral features, and neighborhood connectivity, which is calculated using sizes of connected components. Finally, the components that optimize this measure are selected as meaningful structures in the image.

Then, we propose an unsupervised method for automatic selection of connected components corresponding to meaningful structures among a set of candidate regions from multiple PCA bands. Given multiple objects/structures of interest, the problem is seen as a grouping problem. To solve the grouping problem, we use the probabilistic Latent Semantic Analysis (PLSA) [37] technique which uses a graphical model for the joint probability of the regions and their features in terms of the probability of observing a feature given an object and the probability of an object given the region. The parameters of this graphical model are learned using the Expectation-Maximization algorithm. Then, for a particular region, the set of probabilities of objects/structures given this region can be used to assign an object label to this region.

Finally, we present an approach for classification of remote sensing images

using spatial information extracted from multi-scale segmentations. First, the proposed segmentation algorithm is applied using multiple disjoint structuring element size ranges corresponding to different scales. As the scale increases, the structuring elements gets larger in order first to capture the details and then the general image context. The resulting segments are modeled by their pixel properties, clustered by using the proposed object detection algorithm and classified according to their cluster labels. Final classification is done using decision tree classifiers.

1.2 Summary of Contributions

The goal of this thesis is to analyze remote sensing images and classify objects into land cover/use classes (e.g., buildings, trees, roads, etc.). The classification process is used as a crucial step to interpret the land in many different kinds of applications.

Today's trend in classification of remote sensing images is to do object-oriented classification rather than classifying single pixels. This requires segmentation of objects before assigning their class labels. Most of the previous segmentation work in the remote sensing literature are based on merging neighboring pixels according to user-defined thresholds on their spectral similarity. In this work, the approach we follow is to incorporate structural and shape information in the segmentation step. There are several approaches for extraction of these structures from the image data. However, most of the previous approaches try to solve the problem on specific images such as images of the same type of area and images where these structures are isolated. Our goal is to develop a generic model that can be applied to different types of images.

Seen from this aspect, [54] is related to our work in terms of using structural information in segmentation. In that work, Pesaresi and Benediktsson used morphological processing, which has recently become a popular approach for remote

sensing image analysis, in order to obtain structural information. They successfully applied opening and closing operations with increasing structuring element sizes to an image to generate morphological profiles for all pixels, and assigned a segment label to each pixel using the structuring element size corresponding to the largest derivative of these profiles. Even though morphological profiles are sensitive to different pixel neighborhoods, the segmentation decision is performed by evaluating pixels individually without considering the neighborhood information, and the assumption that all pixels in a structure have only one significant derivative maximum occurring at the same structuring element size may not always hold. However, different than that work, our method [2] uses morphological, neighborhood and spectral information at the same time to segment images. Morphological and neighborhood information are used to determine hierarchical candidate regions for the final segmentation. Then, the meaningful regions in the hierarchy are selected by testing the goodness of each candidate. In a previous work [66], the selection was done in a segmentation hierarchy manually. In this work, we do the selection process automatically by defining a measure for each candidate region and selecting the regions optimizing the measure.

Our object detection method [1] automatically selects meaningful structures among a set of candidate regions from multiple segmentations using probabilistic Latent Semantic Analysis (PLSA). PLSA was originally developed for statistical text analysis to discover topics in a collection of documents that are represented using the frequencies of words from a vocabulary. In our case, the documents correspond to image segments, the word frequencies correspond to histograms of pixel-level features computed as region-level features, and the topics to be discovered correspond to the set of objects/structures of interest in the image. Note that the whole learning and grouping algorithm proceeds in an unsupervised fashion. Russell *et al.* [58] used a different graphical model in a similar setting where multiple segmentations of natural images were obtained using the normalized cut algorithm by changing its parameters, and instances of regions corresponding to objects such as cars, bicycles, faces, sky, etc. were successfully grouped and retrieved from a large data set of images.

Finally, we describe a multi-scale region-based classification method [4] based

on our segmentation and object detection algorithms. Instead of classifying single pixels according to their spectral properties, we classify regions according to the statistical summary of their pixel properties. We model image content by obtaining segmentations in multiple scales. A recent similar approach [14] also did multi-scale region based classification. They obtained multiple hierarchical segmentations by applying region merging with different thresholds on spectral similarities. However, region merging-based methods work randomly and thresholding on spectral similarities does not effectively take into account the structural and shape information. Our method overcomes these disadvantages by the multi-scale nature of our segmentation algorithm. We obtain segmentations in multiple scales by applying the segmentation algorithm with successive disjoint structuring element size ranges.

1.3 Organization of the Thesis

The rest of the thesis is organized as follows. In Chapter 2, some of the previous work on remote sensing image classification is discussed. In Chapter 3, data and features used are introduced. In Chapter 4, a segmentation method for high-resolution remote sensing imagery is presented. For this purpose, a hierarchical segmentation tree is constructed by using both morphological and spectral information. In Chapter 5, a statistical method for unsupervised detection of objects in high-resolution remote sensing imagery is presented. The method uses the proposed segmentation method to find coherent groups of segments corresponding to objects from a set of hierarchical segmentations. In chapter 6, a multi-scale region-based approach for supervised classification of remotely sensed imagery is presented. The method constructs a scale-space by applying the proposed segmentation method in different scales. Then, a new set of features is formed by clustering the segments in multiple scales by the proposed object detection method. In Chapter 7, experiments are discussed. Finally, in Chapter 8, conclusions and future research directions are given.

Chapter 2

Literature Review

As the amount of data to use increases day-by-day, content extraction and classification of remotely sensed imagery have become important research problems. Previous approaches modeled image content using only pixel level features whereas, today, techniques to incorporate spatial information into land cover/use interpretation have gained importance. In this chapter, we discuss some of the previous work on land cover/use classification. For better understandability, we divide the discussion into two categories: Pixel level techniques and spatial techniques using context.

2.1 Pixel Level Techniques

There is an extensive literature on pixel-level analysis of remotely sensed imagery [47]. In remote sensing images, a pixel may correspond to a large area covering different types of objects such as buildings, roads, shadows, grass, and trees [70]. For example, in a Landsat image, the characteristics of a 225 m^2 area are summarized in a pixel which may cause problems in pixel-level approaches. Pixel level techniques performs classification using distribution of pixels in spectral or other feature domains assuming that similar land structures will cluster

together and behave similarly in terms of summarized pixel level features. However, the assumptions for distribution models often do not hold for high-resolution data. In literature, traditional classifiers such as the maximum likelihood method [28, 11], n-dimensional probability density methods [15], artificial neural networks [29, 75, 13, 36, 7, 53, 61, 40, 38, 22], decision trees [31, 43], discriminant analysis [26, 27], genetic algorithms [69, 8] have been applied by using pixel-level information. However, these methods could not exceed a certain level in terms of accuracy rates.

A remote sensing image may have many feature bands corresponding to different wavelengths. For example, a multi-spectral image may have 4 to 7 bands, whereas a hyper-spectral image may have up to 240 bands [70]. The use of hyper-spectral data in classification has been studied in several works [20, 57]. However, most of the hyper-spectral bands are generally correlated and not useful for classification. Different methods, such as principal components analysis [34, 39, 45, 56, 44, 9], discriminant analysis feature extraction (DAFE) [47], the decision boundary feature extraction (DBFE) [47], feature selection [52, 68] and spectral unmixing [19, 17, 18, 63, 57, 16, 72], have been applied to reduce the dimensionality.

Even though pixel level techniques were popular when resolution was not high enough for applying other techniques [70], pixels alone do not give much information about image content as the resolution increases because they do not take into account contextual information during the labeling process. In general, the techniques that can be used for a better classification require the use of spatial and neighborhood information.

2.2 Spatial Techniques Using Context

We believe that, in addition to pixel-based spectral data, structural and spatial information should also be used to interpret land cover and land use. Traditionally, this is done by the use of textural, morphological, and region level features.

In literature, textural features, such as grey-level co-occurrence matrix (GLCM) [32, 64], normalized gray-level run lengths [73], or Markov random fields (MRF) [21], have been widely used to model spatial information. Li and Narayanan [48] used Gabor wavelet coefficients to characterize spatial information. In a recent approach, Bhagavathy and Manjunath [12] used Gabor texture filters with different scales and orientations, and performed Gaussian mixture-based clustering of pixels as texture elements. In another study, Shackelford and Davis [62] used texture measures extracted form normalized gray-level histogram, contextual and spectral information for classification of urban areas. In [76], Yu *et al.* extracted gradient based features for urban area detection. Unsalan and Boyer [71] modeled image windows by using edge information for urbanization detection. However, instead of modeling spatial context by pixel windows, an image segmentation approach may further improve the classification results.

Morphological processing has recently become a popular approach for incorporating structural information into classification. For example, Benediktsson *et al.* [10] applied morphological operators with different structuring element sizes to obtain a multi-scale representation of structural information, and used neural network classifiers to label pixels according to their morphological profiles. Each pixel feature was obtained as the difference between the multi-scale morphological profiles at successive scales.

Another method for incorporating spatial information into classification is through the use of regions obtained by segmentation. This is also referred to as object-oriented classification in the remote sensing literature. Image segmentation techniques [33] automatically group neighboring pixels into contiguous regions whose pixels are similar in terms of a criteria. Although image segmentation is heavily studied in computer vision and image processing fields, and despite the early efforts [42] that use spatial information for classification of remotely sensed imagery, segmentation algorithms have only recently started receiving emphasis in remote sensing image analysis. Examples of image segmentation in the remote sensing literature include region growing [25] and Markov random field models [60] for segmentation of natural scenes, hierarchical segmentation for image mining

[67], region growing for object level change detection [35], and boundary delineation of agricultural fields [59]. However, better segmentation of regions will definitely improve the object-oriented classification results. Most of the previous classification work use segmentation as a preprocess for incorporating spatial information. But, in general, simple and unreliable segmentation algorithms are applied. In [46], Kusaka and Kawata used an edge-based segmentation technique to find uniform regions, and classified these regions based on their spectral and spatial features. Bruzzone and Carlin [14] performed classification using the spatial context of each pixel according to a complete hierarchical multi-level representation of the scene. Their system was made up of a feature extraction module and a classifier. They applied a region merging-based hierarchical segmentation to the images to obtain segmentation results at different levels of resolution. Then, each pixel was characterized by a feature vector that included both the pixel level spectral information and the attributes of all the regions in which the pixel was included in the hierarchical segmentation. In a similar approach [4], we obtained a multi-resolution representation using wavelet decomposition [50, 49], segmented images at each resolution using clustering and mathematical morphology-based segmentation algorithms, and used region-based spectral, textural and shape features for classification. In [41], Katartzis *et al.* also modeled spatial information by segmenting images into regions and classifying these regions. Their system was based on a Markovian model, defined on the hierarchy of a multiscale region adjacency graph. In another study [65], Soh *et al.* presented a system for sea ice image classification which also segments the images, generate descriptors for the segments and then uses expert system rules to classify the images.

Chapter 3

Feature Extraction and Datasets

The algorithms presented in this thesis will be illustrated using three different data sets:

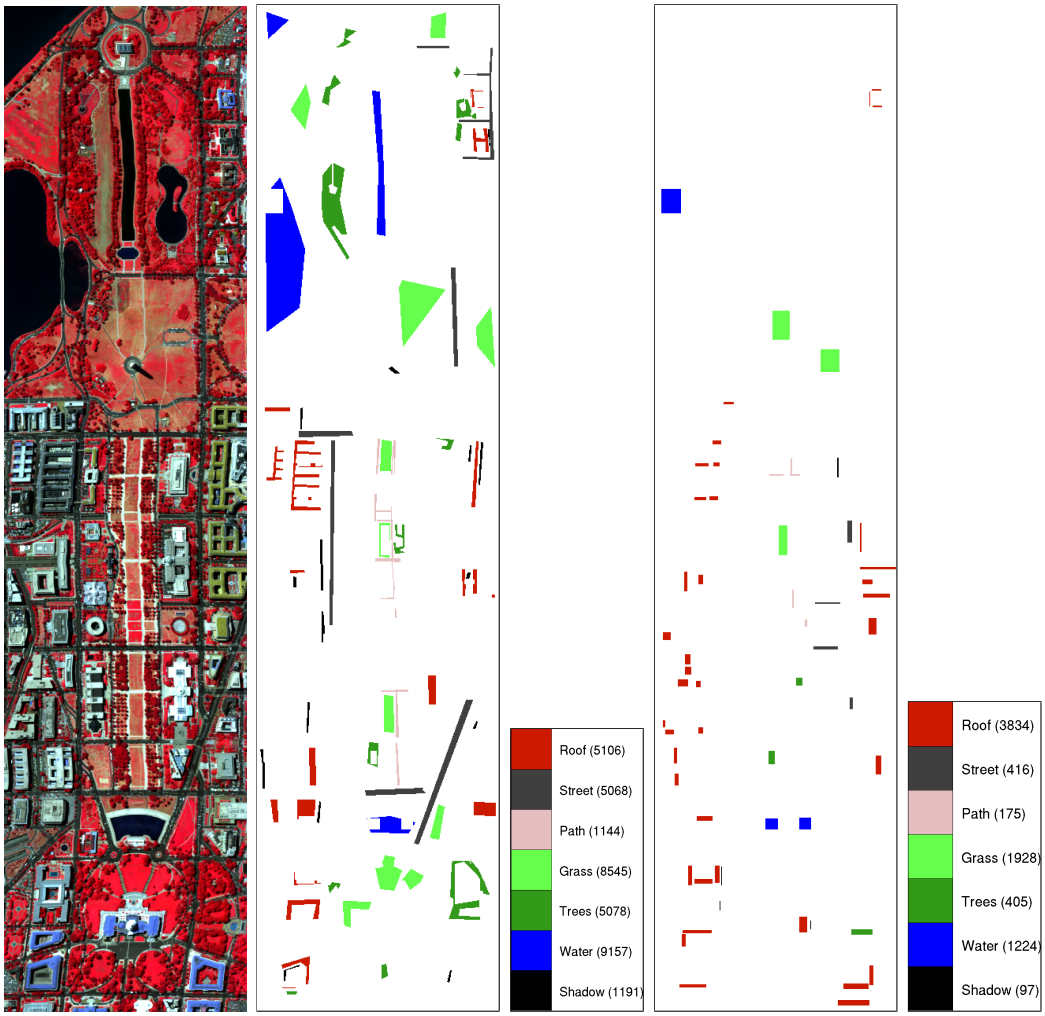
1. *DC Mall*: HYDICE (Hyperspectral Digital Image Collection Experiment) image with $1,280 \times 307$ pixels and 191 spectral bands corresponding to an airborne data flightline over the Washington DC Mall area.

   The *DC Mall* data set includes 7 land cover/use classes: roof, street, path, grass, trees, water, and shadow. A thematic map with ground truth labels for 8,079 pixels was supplied with the original data [47]. We used this ground truth for testing and separately labeled 35,289 pixels for training. Details are given in Figure 3.1.

2. *Centre*: DAIS (Digital Airborne Imaging Spectrometer) and ROSIS (Reflective Optics System Imaging Spectrometer) data with $1,096 \times 715$ pixels and 102 spectral bands corresponding to the city center in Pavia, Italy.

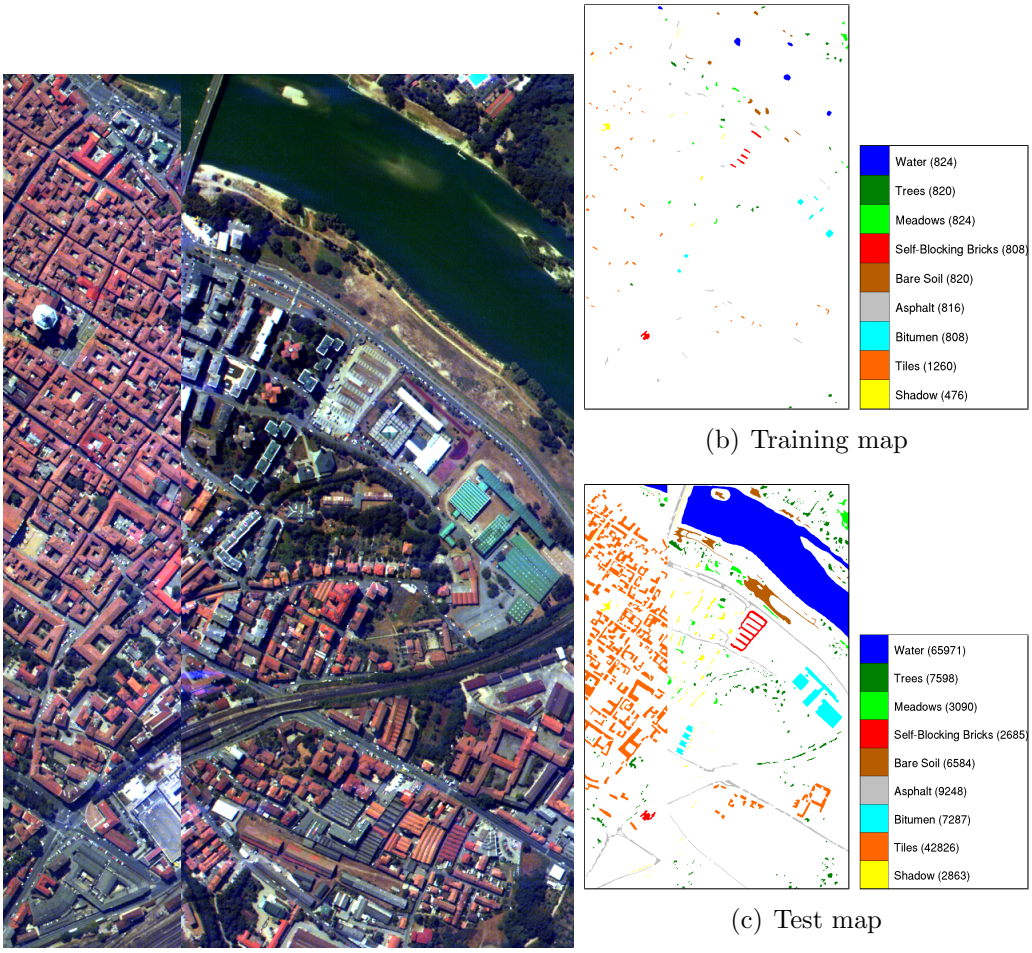
   The *Centre* data set includes 9 land cover/use classes: water, trees, meadows, self-blocking bricks, bare soil, asphalt, bitumen, tiles, and shadow. The thematic maps for ground truth contain 7,456 pixels for training and 148,152 pixels for testing. Details are given in Figure 3.2.

3. *University*: DAIS and ROSIS data with 610×340 pixels and 103 spectral



(a) *DC Mall* data (b) Training map (c) Test map

Figure 3.1: False color image of the *DC Mall* data set (generated using the bands 63, 52 and 36) and the corresponding ground truth maps for training and testing. The number of pixels for each class are shown in parenthesis in the legend. (Images taken from [3].)



(a) *Centre* data

(b) Training map

(c) Test map

Figure 3.2: False color image of the *Centre* data set (generated using the bands 68, 30 and 2) and the corresponding ground truth maps for training and testing. The number of pixels for each class are shown in parenthesis in the legend. (A missing vertical section in the middle was removed.) (Images taken from [3].)

bands corresponding to a scene over the University of Pavia, Italy.

The *University* data set also includes 9 land cover/use classes: asphalt, meadows, gravel, trees, (painted) metal sheets, bare soil, bitumen, self-blocking bricks, and shadow. The thematic maps for ground truth contain 3,921 pixels for training and 42,776 pixels for testing. Details are given in Figure 3.3.

In the rest of the chapter, pixel level characterization consists of spectral and textural properties of pixels that are extracted as described below.

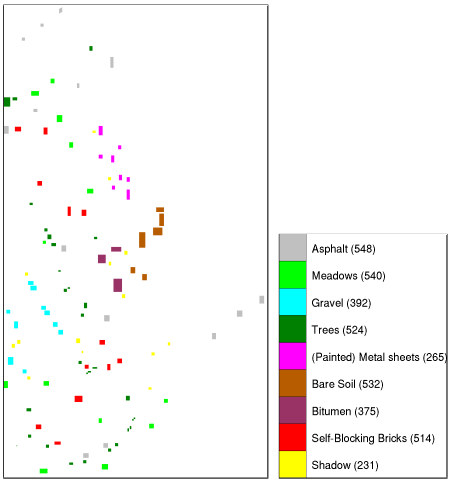
To simplify computations and to avoid the curse of dimensionality during the analysis of hyper-spectral data, we apply Fisher's linear discriminant analysis (LDA) [23] that finds a projection to a new set of bases that best separate the data in a least-squares sense. The resulting number of bands for each data set is one less than the number of classes in the ground truth. We also apply principal components analysis (PCA) [23] that finds a projection to a new set of bases that best represent the data in a least-squares sense. Then, we keep the top principal components representing the 99% variance of the whole data instead of the large number of hyper-spectral bands. The resulting number of bands for *DC Mall*, *Centre* and *University* data sets are 3, 3 and 4, respectively. In addition, we extract Gabor texture features [51] by filtering the first principal component image with Gabor kernels at different scales and orientations shown in Figure 3.4. We use kernels rotated by $n\pi/4, n = 0, \ldots, 3$, at 4 scales resulting in feature vectors of length 16.

Finally, each feature component is normalized by linear scaling to unit variance [5] as

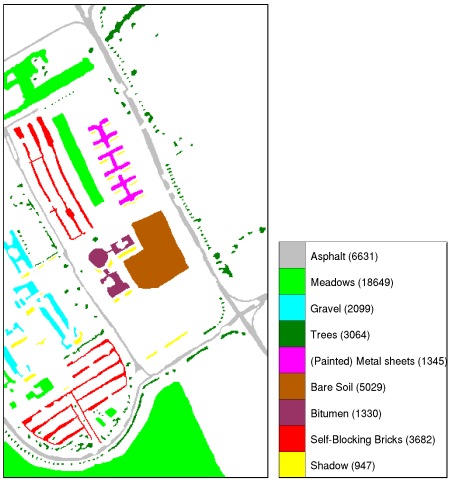
$$\tilde{x} = \frac{x - \mu}{\sigma} \tag{3.1}$$

where x is the original feature value, $\tilde{x}$ is the normalized value, μ is the sample mean, and σ is the sample standard deviation of that feature, so that the features with larger ranges do not bias the results. Examples for pixel level features are shown in Figures 3.5-3.7.

(a) *University* data



(b) Training map



(c) Test map

Figure 3.3: False color image of the *University* data set (generated using the bands 68, 30 and 2) and the corresponding ground truth maps for training and testing. The number of pixels for each class are shown in parenthesis in the legend. (Images taken from [3].)

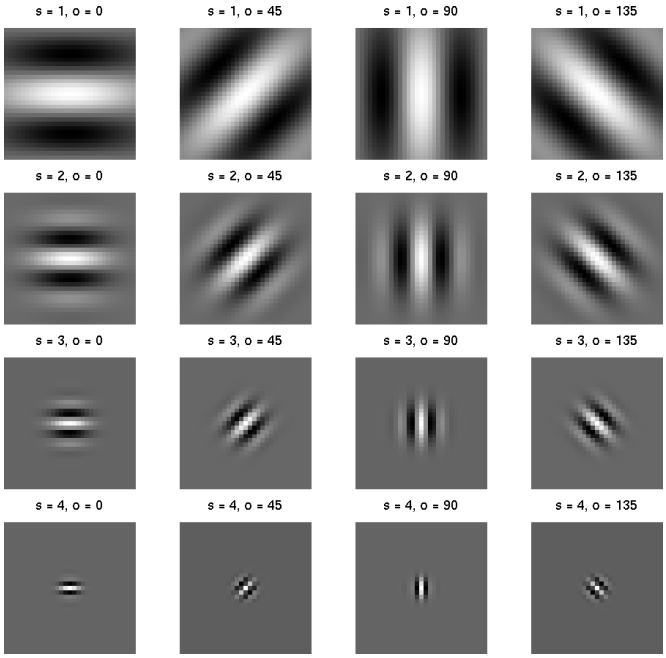


Figure 3.4: Gabor texture filters at different scales ($s = 1, \ldots, 4$) and orientations ($o \in \{0°, 45°, 90°, 135°\}$). Each filter is approximated using 31×31 pixels. (Image taken from [3].)

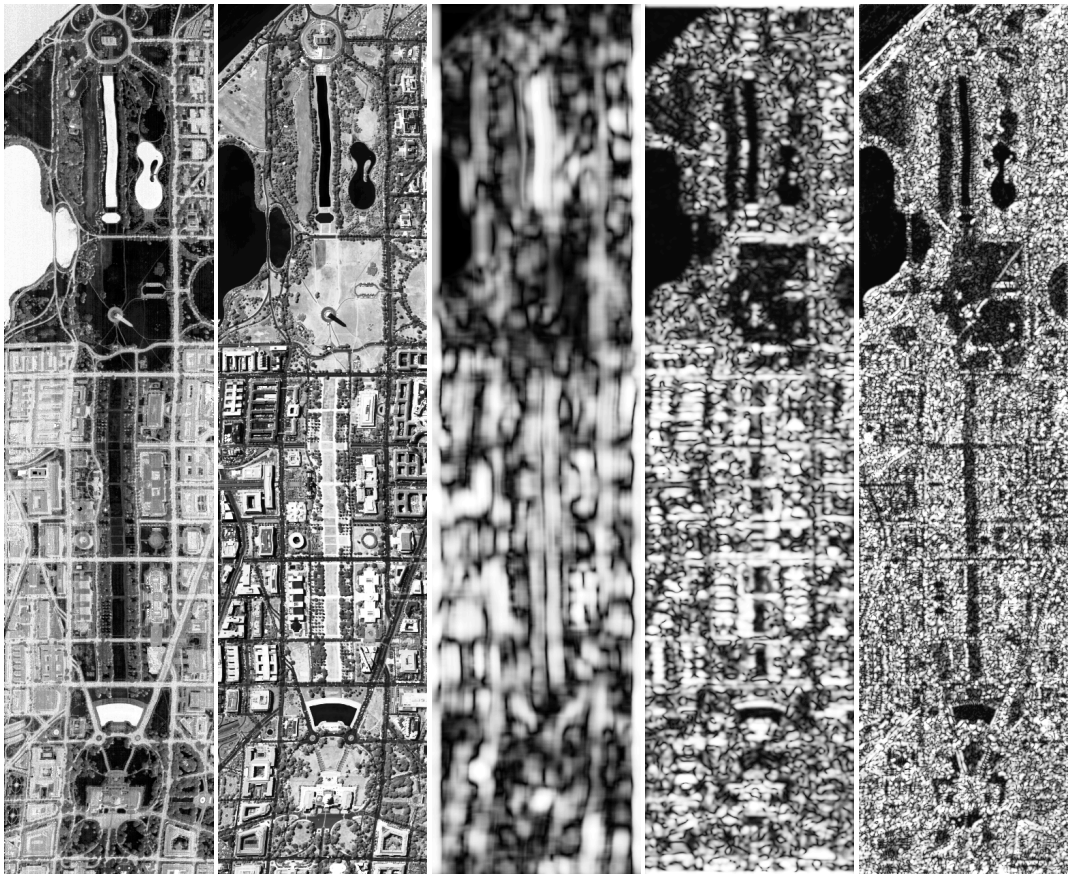


Figure 3.5: Pixel feature examples for the *DC Mall* data set. From left to right: the first LDA band, the first PCA band, Gabor features for 90 degree orientation at the first scale, Gabor features for 0 degree orientation at the third scale, and Gabor features for 45 degree orientation at the fourth scale. Histogram equalization was applied to all images for better visualization. (Images taken from [3].)

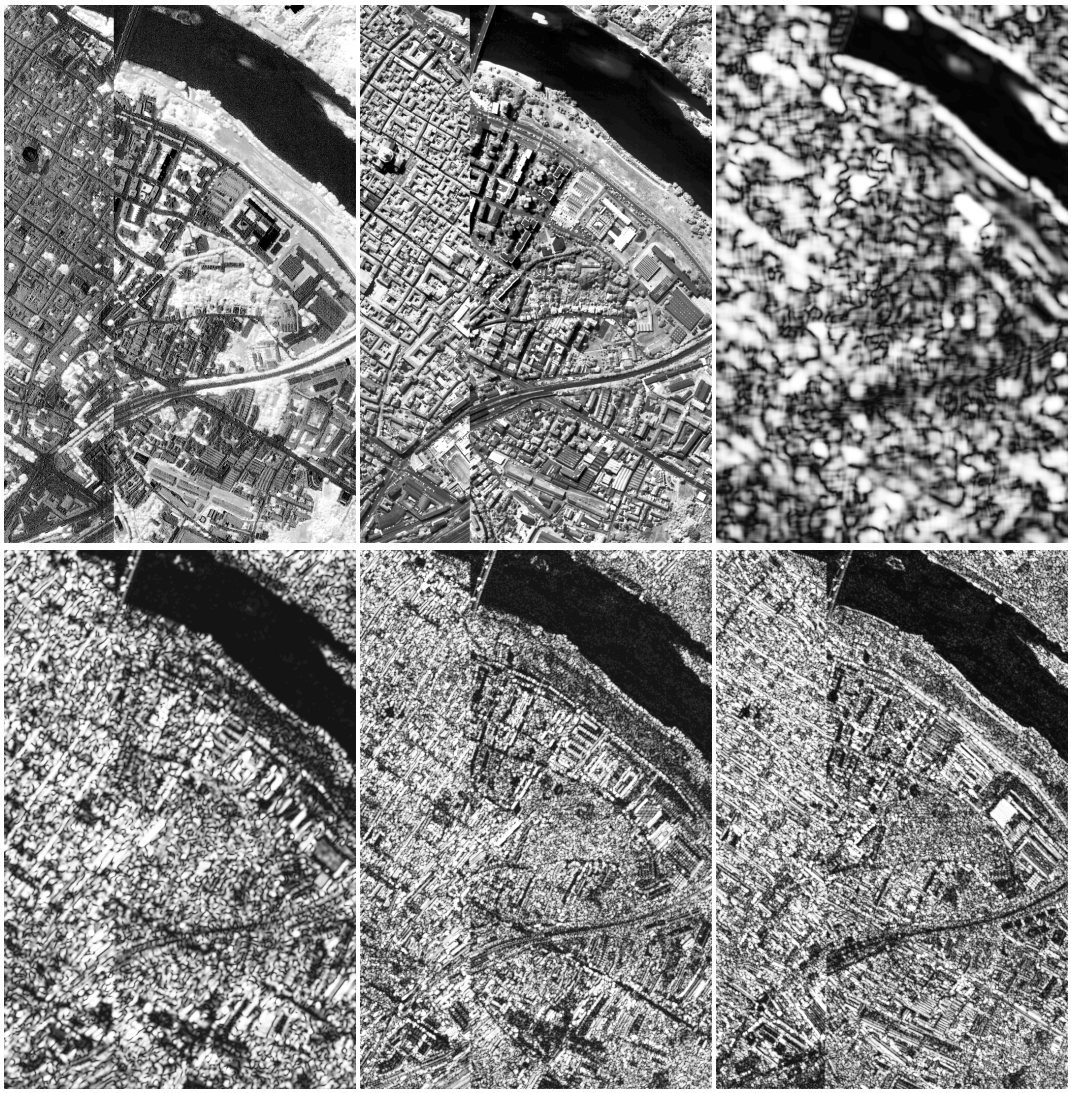


Figure 3.6: Pixel feature examples for the *Centre* data set. From left to right, first row: the first LDA band, the first PCA band, Gabor features for 135 degree orientation at the first scale; second row: Gabor features for 45 degree orientation at the third scale, Gabor features for 45 degree orientation at the fourth scale, and Gabor features for 135 degree orientation at the fourth scale. Histogram equalization was applied to all images for better visualization. (Images taken from [3].)

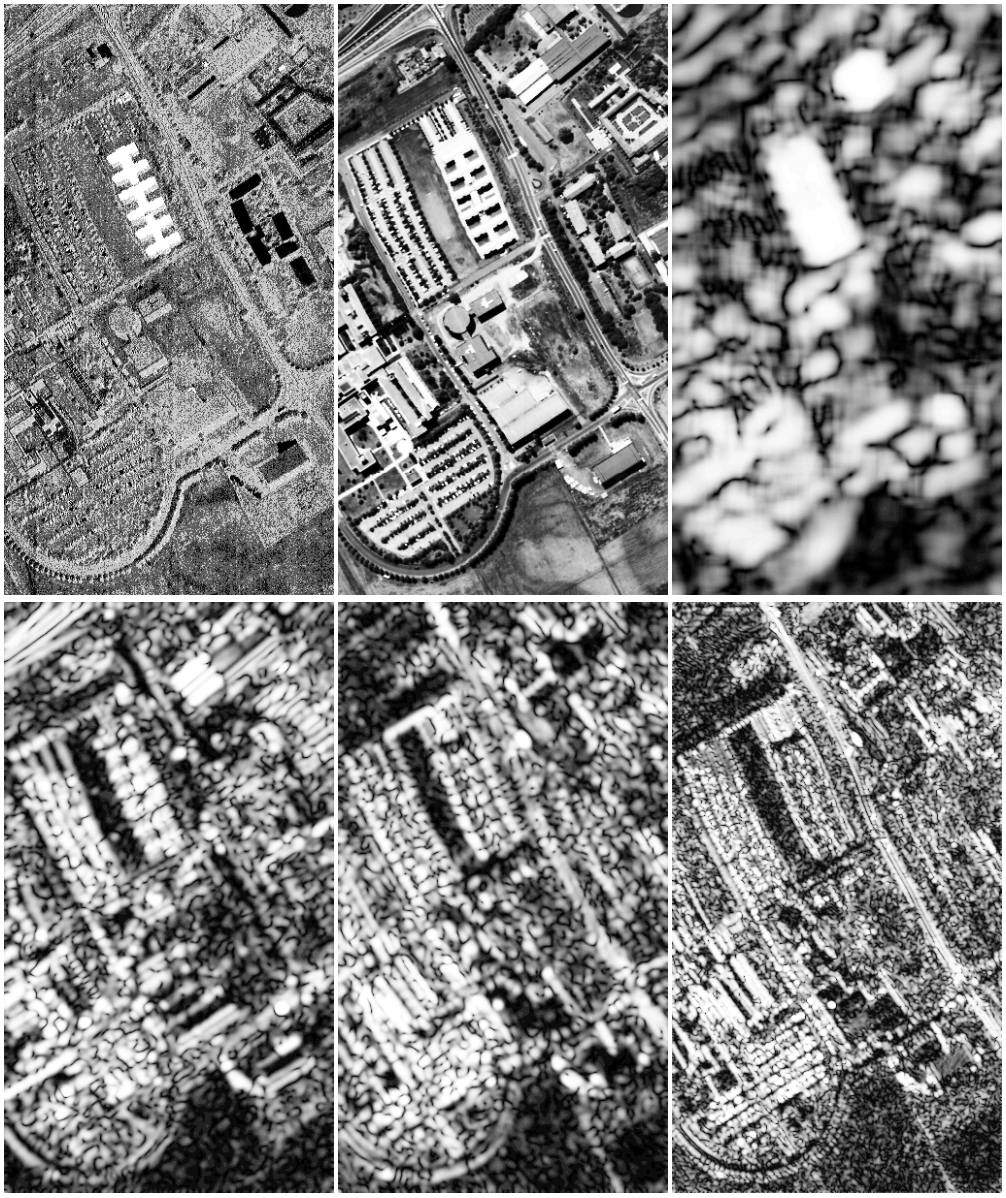


Figure 3.7: Pixel feature examples for the *University* data set. From left to right, first row: the first LDA band, the first PCA band, Gabor features for 45 degree orientation at the first scale; second row: Gabor features for 45 degree orientation at the third scale, Gabor features for 135 degree orientation at the third scale, and Gabor features for 135 degree orientation at the fourth scale. Histogram equalization was applied to all images for better visualization. (Images taken from [3].)

Chapter 4

Image Segmentation

In this work, our goal is to develop a segmentation algorithm for partitioning images into spatially contiguous regions so that the structural information can be modeled using the properties of these regions in classification. Many popular image segmentation algorithms in the computer vision literature that are based on clustering assume that images have a moderate number of objects with relatively homogeneous features, and cannot be directly applied to high-resolution remote sensing images that contain a large number of complex structures. Furthermore, another popular approach of edge-based segmentation becomes hard for such images because of the large amount of details. Moreover, watershed-based techniques are also not very useful because they often produce oversegmented results mostly because of irrelevant local extrema in images. A common approach is to apply smoothing filters to suppress these extrema but lots of details in high-resolution images may be lost because spatial support of these details are usually small. Therefore, most of the segmentation work in the remote sensing literature are based on merging neighboring pixels according to user-defined thresholds on their spectral similarity. As an alternative, proximity filtering and morphological operations can also be used as post-processing techniques to pixel-based classification results for segmenting regions [6].

In a related work, Pesaresi and Benediktsson [54] performed segmentation using morphological characteristic of pixels in the image. In their approach, opening

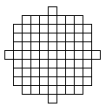
and closing operations with increasing structuring element (SE) sizes were successively applied to an image to generate morphological profiles for all pixels, and the segment label of each pixel was assigned as the SE size corresponding to the largest derivative of these profiles. A problem with that approach is that it assumes all the pixels in a particular structure have only one significant derivative maximum occurring at the same SE size. However, our experiments have shown that many pixels in most structures often have more than one significant derivative maximum. Furthermore, even though morphological profiles are sensitive to different pixel neighborhoods, the segmentation decision is performed by evaluating pixels individually without considering the neighborhood information.

In this chapter, we present a method that uses the neighborhood and spectral information as well as the morphological information. We first apply principal components analysis to hyper-spectral data to obtain representative bands. Then, we extract candidate regions on each principal component by applying opening and closing by reconstruction operations. For each principal component, we represent the extracted regions by a hierarchical tree, and select the most meaningful regions in that tree by optimizing a measure that consists of two factors: spectral homogeneity, which is calculated in terms of variances of multi-spectral features, and neighborhood connectivity, which is calculated using sizes of connected components.

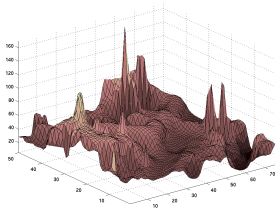
4.1 Morphological Profiles

Morphological opening and closing operations are used to model structural characteristics of pixel neighborhoods. These operations are known to isolate structures that are brighter and darker than their surroundings, respectively. Contrary to opening (respectively, closing), opening by reconstruction (respectively, closing by reconstruction) preserves the shape of the structures that are not removed by erosion (respectively, dilation). In other words, image structures that the SE cannot be contained are removed while others remain (see Figure 4.1 for an illustration).

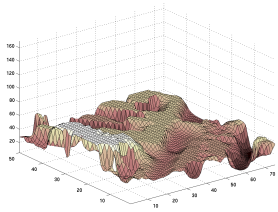
(a) Gray-scale image



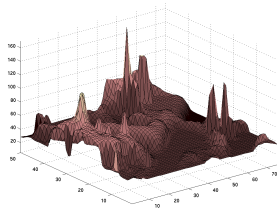
(b) SE



(c) 3-d representation of (a)

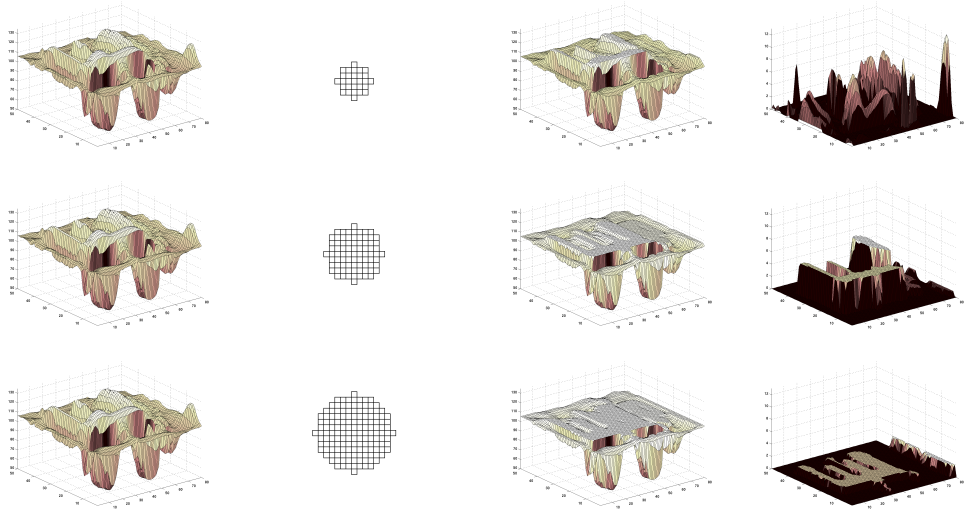


(d) After opening by reconstruction



(e) After closing by reconstruction

Figure 4.1: Opening and closing by reconstruction example. (d) and (e), respectively, show 3-d representations of the results of opening and closing by reconstruction applied on (a) with a disk SE of size 5. In (d) and (e), respectively, brighter and darker image structures that the SE cannot be contained are removed by flattening the pixels on the structure into same level.



(a) Gray-scale image (b) SE (c) Opening profile (d) Derivative of opening profile

Figure 4.2: The morphological opening profile and the derivative of the morphological opening profile example. The original image in (a) is applied opening by reconstruction with increasing SE sizes (disk SEs of sizes 3, 5 and 7 are used as shown in (b)) to obtain the opening profile in (c). The difference between consecutive steps of the opening profile is stored as the derivative of the opening profile as shown in (d).

These operations are applied using increasing SE sizes to generate multi-scale characteristics called morphological profiles For completeness, we review the concepts of the morphological profile and the derivative of the morphological profile as defined by Pesaresi and Benediktsson [54] below (see Figure 4.2 for an illustration).

Let γ_λ be the morphological opening by reconstruction operator using SE with size λ (in our case λ represents the radius of a disk shaped SE) and $\Pi_\gamma(x)$ denote the opening profile at pixel x of image I. $\Pi_\gamma(x)$ is defined as the vector

$$\Pi_\gamma(x) = \{\Pi_{\gamma_\lambda} : \Pi_{\gamma_\lambda} = \gamma_\lambda(x), \forall\lambda \in [0, \ldots, m]\}. \tag{4.1}$$

Also, let φ_λ be the morphological closing by reconstruction operator using SE with size λ and $\Pi_\varphi(x)$ denote the closing profile at pixel x of image I. Similarly, $\Pi_\varphi(x)$ is defined as the vector

$$\Pi_\varphi(x) = \{\Pi_{\varphi_\lambda} : \Pi_{\varphi_\lambda} = \varphi_\lambda(x), \forall\lambda \in [0, \ldots, m]\}. \tag{4.2}$$

The definition of opening and closing by reconstruction operations imply that $\Pi_{\gamma_0}(x) = \Pi_{\varphi_0}(x) = I(x)$. The derivative of the morphological profile (DMP) is defined as a vector where the measure of the slope of the opening-closing profile is stored for every step of an increasing SE series [54]. The derivative of the opening profile $\Delta_\gamma(x)$ is defined as the vector

$$\Delta_\gamma(x) = \{\Delta_{\gamma_\lambda} : \Delta_{\gamma_\lambda} = |\Pi_{\gamma_\lambda} - \Pi_{\gamma_{\lambda-1}}|, \forall\lambda \in [1, \ldots, m]\}. \tag{4.3}$$

Similarly, the derivative of the closing profile $\Delta_\varphi(x)$ is defined as the vector

$$\Delta_\varphi(x) = \{\Delta_{\varphi_\lambda} : \Delta_{\varphi_\lambda} = |\Pi_{\varphi_\lambda} - \Pi_{\varphi_{\lambda-1}}|, \forall\lambda \in [1, \ldots, m]\}. \tag{4.4}$$

Then, the DMP $\Delta(x)$ can be written as the vector

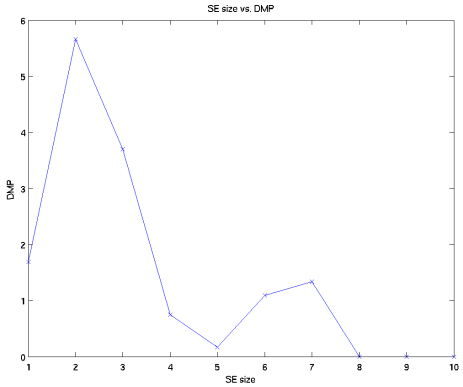
$$\Delta(x) = \left\{ \begin{array}{c} \Delta_{c+\lambda} : \Delta_{\gamma_\lambda}, \forall\lambda \in [1, \ldots, m] \\ \Delta_{c-\lambda+1} : \Delta_{\varphi_\lambda}, \forall\lambda \in [1, \ldots, m] \end{array} \right\} \tag{4.5}$$

for an arbitrary integer c with m equal to the total number of iterations [54].

In their segmentation scheme, Pesaresi and Benediktsson [54] define an image segment as a set of connected pixels showing the greatest value of the DMP for the same SE size. That is, the segment label of each pixel is assigned according to the SE size corresponding to the largest derivative of its profiles. Their scheme works well in images where the structures in the image are mostly flat so that all pixels in a structure have only one derivative maximum. A drawback of this scheme is that neighborhood information is not used while assigning segment labels to pixels. This results in lots of small noisy segments in images with non-flat structures where the scale with the largest value of the DMP may not correspond to the true structure (see Figure 4.3 for an illustration). In our approach, we do not consider pixels alone while assigning segment labels. Instead, we also take into account the behavior of the neighbors of the pixels.

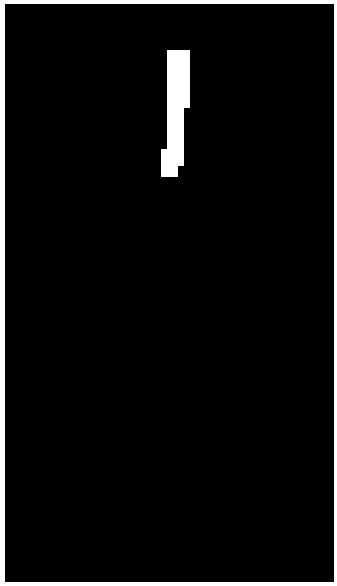
4.2 Hierarchical Region Extraction

In our segmentation approach, our aim is to determine the regions by applying opening and closing by reconstruction operations. We assume that pixels with a

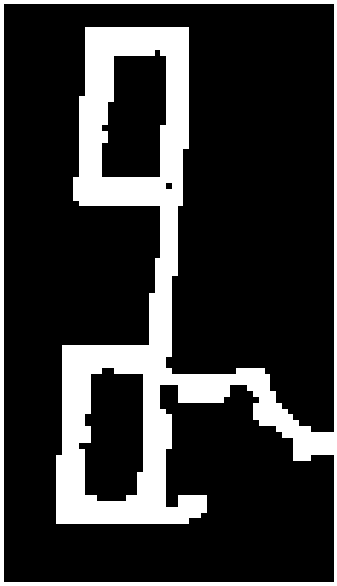


(a) DMP of the pixel marked in (b)

(b) Sample pixel marked on the image



(c) Region for SE size 2



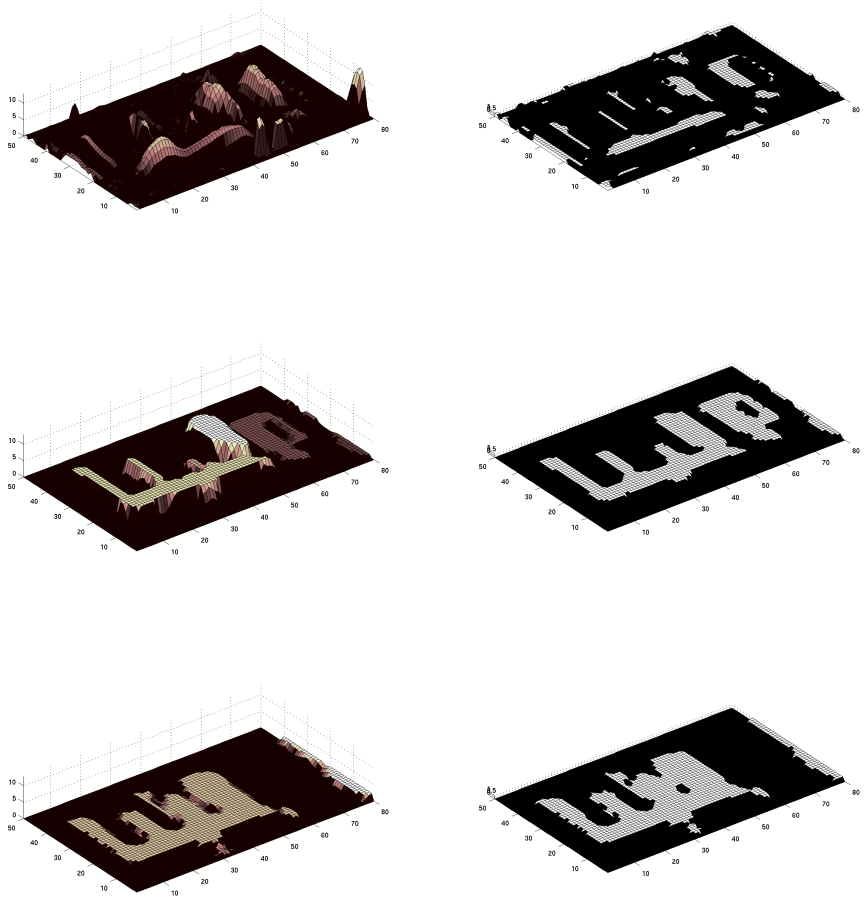
(d) Region for SE size 3

Figure 4.3: The greatest value in the DMP of the pixel marked with a blue + in (b) is obtained for SE size 2 (derivative of the opening profile of the 3rd PCA band is shown in (a)). (c) shows the region that we would obtain if we label the pixels with the SE size corresponding to the greatest DMP. The region in (d) that occurs with SE size 3 is more preferable as a complete structure but it does not correspond to the scale of the greatest DMP for all pixels inside the region.

positive DMP value at a particular SE size face a change with respect to their neighborhoods at that scale. The main idea is that a neighboring group of pixels that have a similar change for a particular SE size is a candidate region for the final segmentation. These groups can be found by applying connected components analysis to the DMP at each scale (see Figure 4.4 for an illustration). Then, we use the spectral angle mapper (SAM) as a rough measure to check the spectral homogeneity within each group. SAM between two vectors s_i and s_j is calculated as: $SAM(s_i, s_j) = \cos^{-1}(\frac{s_i \cdot s_j}{\|s_i\|\|s_j\|})$. Using SAM, spectral similarity of pixel vectors $\{p\}_i^K$ in the region R_k relative to the vector s_j is computed as: $S(R_k, s_j) = (1/K) \sum_{i=1}^{K} SAM(s_j, p_i)$ where K is the number of pixels in the region R_k. Then, Plaza and Tilton [55] define a measure of homogeneity in the region R_k as: $S(R_k, c_k)$ where $c_k = (1/K) \sum_{i=1}^{K} p_i$ is the centroid of R_k. S indicates how similar are spectral information of the pixels in a region. The less the S, the more the region is homogeneous. The connected components whose average DMP values are greater than 0.2 and average SAM values are less than 0.095 are considered in the rest of the analysis. These thresholds are chosen empirically.

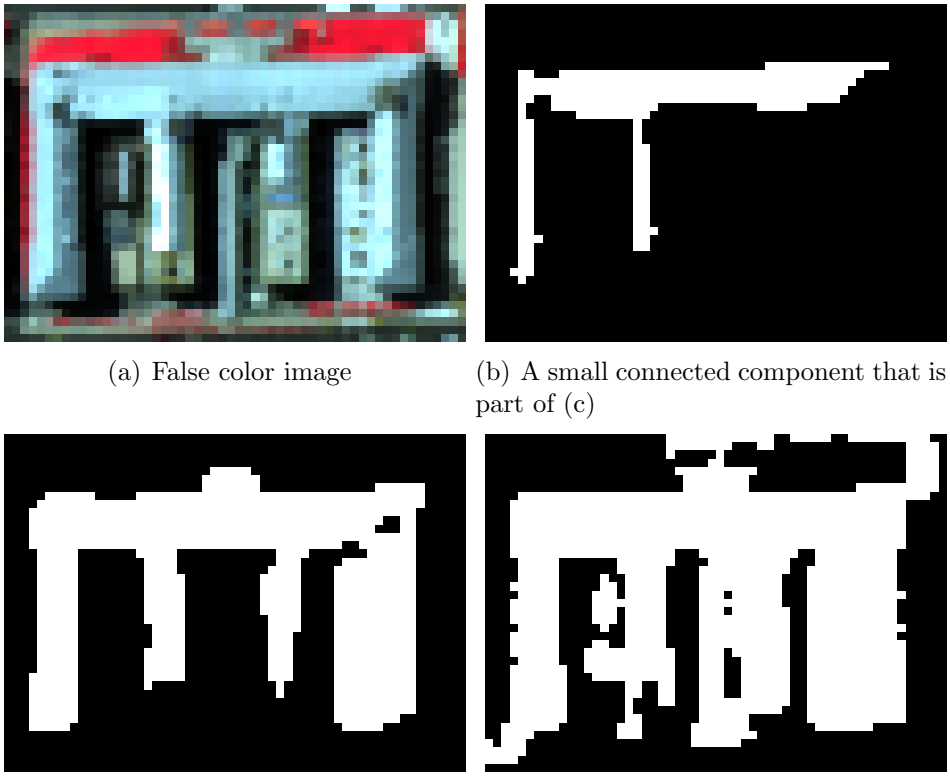
Considering the fact that different structures have different sizes, we apply opening and closing by reconstruction using SEs in increasing sizes from 1 to m. However, a connected component appearing for a small SE size may be appearing because heterogeneity and geometrical complexity of the scenes as well as other external effects such as shadows produce texture effects in images and result in structures that can be one to two pixels wide [54]. In this case, there is most probably a larger connected component appearing at the scale of a larger SE and to which the pixels of those noise components belong. On the other hand, a connected component that corresponds to a true structure in the final segmentation may also appear as part of another component at larger SE sizes. The reason is that a meaningful connected component may start merging with its surroundings and other connected components after the SE size in which it appears is reached. Figure 4.5 illustrates these cases.

For each opening and closing profile, through increasing SE sizes from 1 to m, each morphological operation reveals connected components that are contained within each other in a hierarchical manner where a pixel may be assigned to



(a) Opening DMP (b) $DMP > 0$

Figure 4.4: In (b), the pixels whose DMP (DMP of the 2nd PCA band is shown in (a)) are greater than 0 are shown. Each connected component at each scale is a candidate region for the final segmentation.



(a) False color image

(b) A small connected component that is part of (c)

(c) The preferred connected component

(d) A large connected component where (c) started merging with others

Figure 4.5: Example connected components for a building structure. These components appear for SE sizes 3, 5 and 6, respectively, in the derivative of the opening profile of the 2nd PCA band.

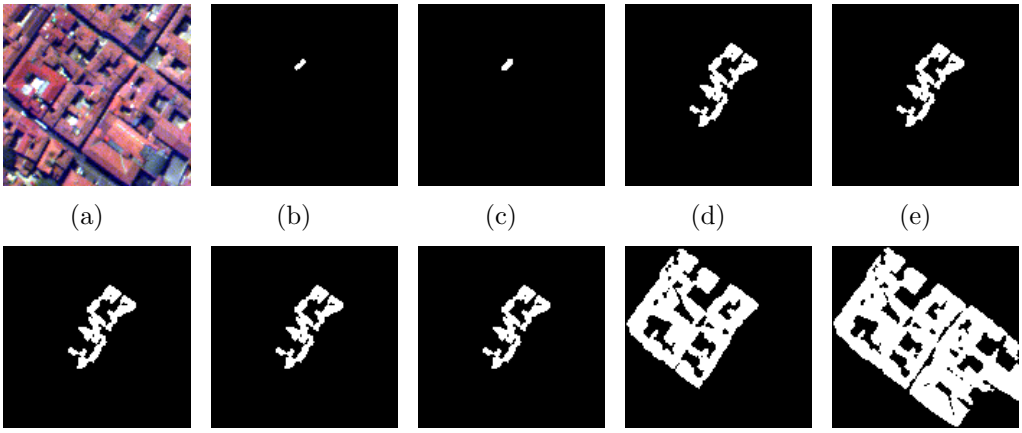


Figure 4.6: Example connected components appearing for SE sizes from 2 to 10 in the derivative of the opening profile of the 3rd PCA band. These regions are contained within each other in a hierarchical manner. Note that the components do not change in some of the scales.

more than one connected component appearing at different SE sizes (see Figure 4.6). We treat each component as a candidate meaningful region. Using these candidate regions, a tree is constructed where each connected component is a node and there is an edge between two nodes corresponding to two consecutive scales (SE sizes differ by 1) if one node is contained within the other. Leaf nodes represent the components that appear for SE size 1. Root nodes represent the components that exist for SE size m. Since we use a finite number of SE sizes, there may be more than one root node. In this case, there will be more than one tree and the algorithms described in the next section are run on each tree separately.

Figure 4.7 shows an example tree where the nodes are labeled as i_j with i denoting the node's level and j denotes the number of the node from left to right in level i. For example, node 3_3 has two children nodes 2_4 and 2_5, and its parent is node 4_1. The reason of node 2_3 having only one child may be that either no new connected component appears in level 2 or node 2_3 is formed by merging of node 1_4 with its surrounding pixels that are not included in any connected component in level 1. The same reasons also hold for node 3_2. Figure 4.8 shows a part of an example tree constructed by candidate meaningful regions

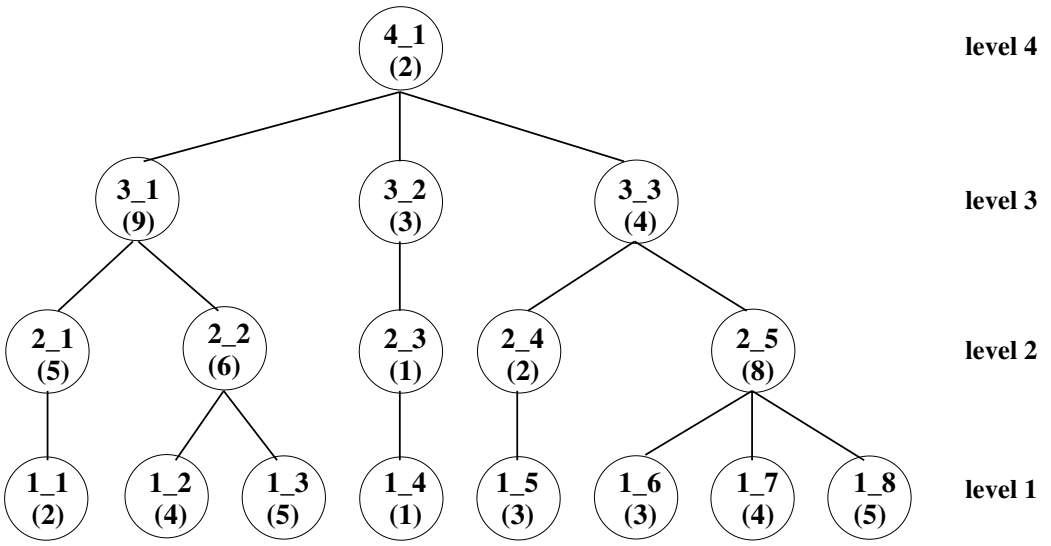


Figure 4.7: An example tree. Node i_j is a connected component that exists for SE size i. j denotes the sequence number of the node from left to right in level i.

appearing in five levels.

4.3 Region Selection

After forming a tree for each opening and closing profile, our aim is to search for the most meaningful connected components among those appearing at different SE sizes in the segmentation hierarchy. With a similar motivation in [66], Tilton analyzed hierarchical image segmentations and selected the meaningful regions manually. Then, Plaza and Tilton [55] investigated how different spectral, spatial and joint spectral/spatial features of regions change from one level to another in a segmentation hierarchy with the goal of automating the selection process in the future. In this thesis, each node in the tree is treated as a candidate region in the final segmentation, and selection is done automatically as described below.

Ideally, we expect a meaningful region to be as homogeneous as possible. However, in the extreme case, a single pixel is the most homogeneous. Hence, we also want a region to be as large as possible. In general, a region stays almost the same (both in homogeneity and size) for some number of SEs, and then faces a

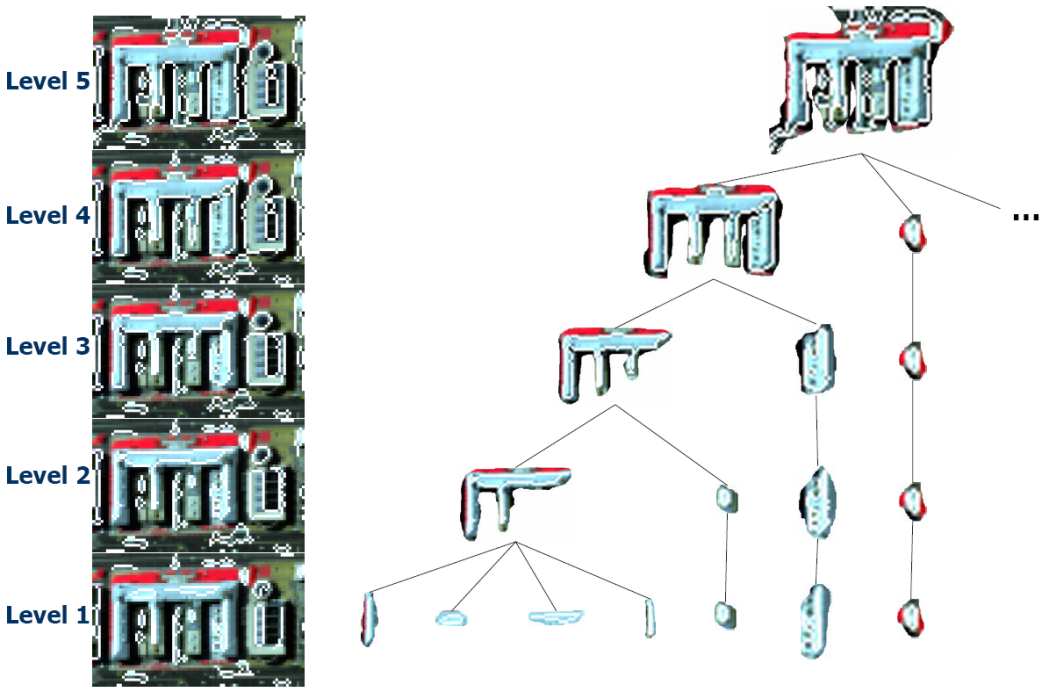


Figure 4.8: An example tree where each candidate region is a node.

large change at a particular scale either because it merges with its surroundings to make a new structure or because it is completely lost. Consequently, the size we are interested in corresponds to the scale right before this change. In other words, if the nodes on a path in the tree stay homogeneous until some node n, and then the homogeneity is lost in the next level, we say that n corresponds to a meaningful region in the hierarchy.

With this motivation, to check the meaningfulness of a node, we define a measure consisting of two factors: spectral homogeneity, which is calculated in terms of variances of spectral features, and neighborhood connectivity, which is calculated using sizes of connected components. Then, starting from the leaf nodes (level 1) up to the root node (level m), we compute this measure at each node and select a node as a meaningful region if it is the most homogeneous and large enough node on its path in the hierarchy (a path corresponds to the set of nodes from a leaf to the root).

In order to calculate the homogeneity factor in a node, we use the fact that pixels in a correct structure should have not only similar morphological profiles,

but also similar spectral features. Thus, we calculate the homogeneity of a node as the standard deviation of the spectral information of the pixels in the corresponding region where the spectral information of a pixel consists of the PCA components representing the 99% variance of the whole data. However, while examining a node from the leaf up to the root in terms of homogeneity, we do not use the standard deviation of the node directly. Instead, we consider the difference of the standard deviation of that node and its parent. What we expect is a sudden increase in the standard deviation. When the standard deviation does not change much, it usually means that small sets of pixels are added to the region or some noise pixels are cleaned. When there is a large change, it means that the structure merged with a larger structure or it merged with other irrelevant pixels disturbing the homogeneity in the node. Hence, the difference of the standard deviation in the node's parent and the standard deviation in the node should be maximized while selecting the most meaningful nodes.

To be able to calculate a single value for the difference of the standard deviations, we must find a single value for the standard deviation of the PCA components of the pixels in a node and its parent. Let the number of PCA components be d. We find a single standard deviation value for the d-dimensional PCA components X of the pixels in node n by projecting X into a 1-dimensional representation where we find the standard deviation [30]. X is projected onto the vector connecting the average of the PCA components of n and the average of the PCA components of the parent of n. This is the vector where it is believed to be important for separating the data. Let c_1 be the average of the PCA components of n, c_2 be the average of the PCA components of the parent of n, and $v = c_1 - c_2$ be the d-dimensional vector connecting c_1 and c_2. The projection $x_i\prime \in X\prime$ of PCA components $x_i \in X$ of each pixel i is done by: $x_i\prime = \frac{<x_i,v>}{||v||^2}$. $X\prime$ is a 1-dimensional representation of X projected onto v. After projecting the PCA components of a node and its parent into a 1-dimensional representation, the standard deviation of the projected data for each node is found.

However, using only the homogeneity factor will favor small structures because in the extreme case a single pixel is the most homogeneous. To overcome this problem, the number of pixels in the region corresponding to the node is

introduced as another factor to create a trade-off. As a result, the goodness measure M for a node n is defined as

$$M(n) = D(n, parent(n)) \times C(n) \tag{4.6}$$

where the first term is the standard deviation difference between the node's parent and itself, and the second term is the number of pixels in the node. The node that is relatively homogeneous and large enough will maximize this measure and will be selected as a meaningful region.

Given the value of the goodness measure for each node, we find the most meaningful regions as follows. Suppose $T = (N, E)$ is the tree with N as the set of nodes and E as the set of edges. The leaf nodes are in level 1 and the root node is at level m. Let P denote the set of all paths from the leaves to the root, and $M(n)$ denote the measure at node n. $descendant(n)$ denotes descendant nodes of node n. We select $N^* \subseteq N$ as the final segmentation such that

1. $\forall a \in N^*, \forall b \in descendant(a)$,
   $M(a) \geq M(b)$,

2. $\forall a \in N \setminus N^*$,
   $\exists b \in descendant(a) : M(a) < M(b)$.

3. $\forall a, b \in N^*$,
   $\forall p \in P : a \in p \rightarrow b \notin p$,
   $\forall p \in P : b \in p \rightarrow a \notin p$,

4. $\forall p \in P$,
   $\exists a \in p : a \in N^*$,

The first condition requires that any node in N^* must have a measure greater than all of its descendants. The second condition requires that no node in $N \setminus N^*$ has a measure greater than all of its descendants. The third condition requires that any two nodes in N^* cannot be on the same path (i.e., the corresponding regions cannot overlap). The fourth condition requires that every path must include a node that is in N^*.

We use a two-pass algorithm for selecting the most meaningful nodes (N^*) in the tree. The bottom-up (first) pass aims to find the nodes whose measure is greater than all of its descendants (condition 1). The algorithm first marks all nodes in level 1. Then, starting from level 2 up to the root level, it checks whether each node in each level has a measure greater than or equal to those of all of its children. The greatest measure, seen so far in each path, is propagated to upper levels so that it is enough to check only the children, rather than all descendants, in order to find whether a node's measure is greater than or equal to all of its descendants'.

After the bottom-up pass marks all such nodes, the top-down (second) pass seeks to select the nodes satisfying, as well, the remaining conditions (2, 3, 4). It starts by marking all nodes as *selected* in the root level if they are marked by the bottom-up pass. Then, in each level until the leaf level, the algorithm checks for each node whether it is marked in the bottom-up pass while none of its ancestors is marked. If this condition is satisfied, it marks the node as *selected*. Finally, the algorithm selects the nodes that are marked as *selected* in each level as meaningful regions.

Below, we give the algorithm for selecting the most meaningful nodes in the hierarchical tree. In the algorithm, $children(n)$ denotes children nodes of node n.

Algorithm 1 Region Selection

```
Run Bottom-Up algorithm
Run Top-Down algorithm
for each level l = 1 to m do
  for each node n in level l do
    select n as a meaningful region if it is marked as selected
  end for
end for
```

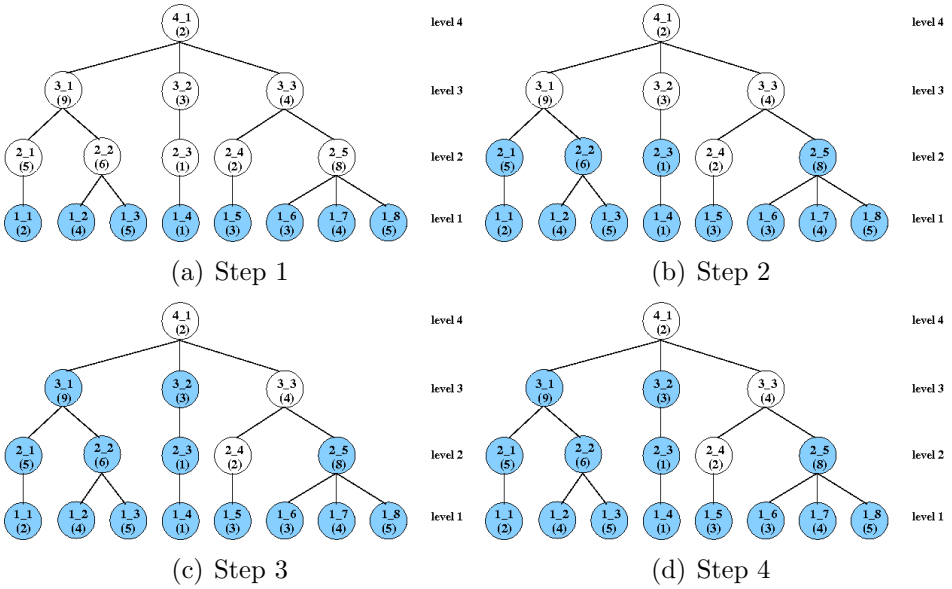
In order to illustrate an example run of these algorithms, a measure is given, in parenthesis, in each node i_j in the example tree of Figure 4.7. After we run the Bottom-Up algorithm, each node $1\_j (1 \leq j \leq 8)$ is marked in the beginning of the algorithm. Then, as we move upwards nodes $2\_1, 2\_2, 2\_3, 2\_5$ in level 2,

Algorithm 2 Bottom-Up algorithm

Mark all nodes in level 1
for each level $l = 2$ to m **do**
  for each node n in level l **do**
    if $M(n) \geq max\{M(a)|a \in children(n)\}$ **then**
      mark n
    else
      $M(n) = max\{M(a)|a \in children(n)\}$
      leave n unmarked
    end if
  end for
end for

Algorithm 3 Top-Down algorithm

Mark all nodes in level m as *selected* if they're already marked in Bottom-Up
for each level $l = m - 1$ to 1 **do**
  for each node n in level l **do**
    if $parent(n)$ is marked as *selected* or *parent-selected* **then**
      mark n as *parent-selected*
    else
      if $parent(n)$ is not marked in Top-Down and n is not marked in Bottom-Up **then**
        leave n unmarked
      else
        mark n as *selected*
      end if
    end if
  end for
end for



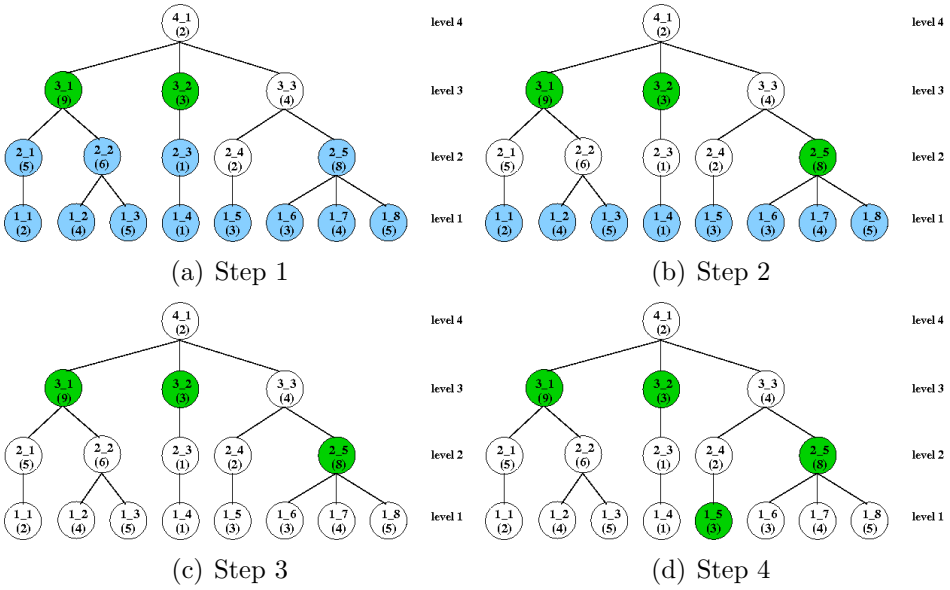
(a) Step 1 (b) Step 2

(c) Step 3 (d) Step 4

Figure 4.9: An example run of Bottom-Up algorithm on the tree in Figure 4.7. Beginning from the leaves until the root, the nodes whose measure are greater than all of its descendants (satisfying condition 1) are colored with blue in each step.

nodes 3_1 and 3_2 are marked in level 3 since each of them is greater than or equal to all of its descendants. Then, we run the Top-Down algorithm and mark nodes 3_1, 3_2, 2_5 and 1_5, satisfying the four conditions defined above, as *selected*. Figures 4.9 and 4.10 show the marked nodes in each step of the Bottom-Up and the Top-Down algorithms. In the Bottom-Up algorithm, the marked nodes are colored with blue and in the Top-Down algorithm, the marked nodes are colored with green.

After selecting the most meaningful connected components in each opening and closing tree separately, the next step is to merge the resulting connected components. The problem occurs when two connected components, where one is selected from the opening tree and the other is selected from the closing tree, intersect. In this case, the intersecting pixels are assigned to the connected component whose goodness measure is greater.



(a) Step 1 (b) Step 2

(c) Step 3 (d) Step 4

Figure 4.10: An example run of Top-Down algorithm on the tree in Figure 4.9(d). Beginning from the root until the leaves, the nodes marked in Bottom-Up algorithm which satisfy, as well, the remaining conditions (2, 3, 4) are marked with green in each step. When the algorithm ends, the green nodes are selected as the most meaningful nodes in the tree.

Chapter 5

Object Detection

In Chapter 4, we described a method that used the neighborhood and spectral information as well as the morphological information for segmentation. After principal components analysis (PCA), morphological profiles were generated for each PCA band separately. These operations produced a set of connected components forming a hierarchy of segments for each PCA band. Then, a measure that combined spectral homogeneity and neighborhood connectivity was designed to select meaningful segments at different levels of the hierarchy.

The experiments in Section 7.1 show that the combined measure is able to detect structures in the image that are more precise and more meaningful than the structures detected by the approach in [54]. An important observation is that different structures appear more clearly in different principal components. For example, buildings can be detected accurately in one component but roads, trees, fields and paths can be detected accurately in other components (see Figures 5.1 and 5.2 for examples). Information from multiple PCA components must be combined for better overall detection.

In this chapter, we present an unsupervised algorithm for automatic selection of segments from multiple segmentations and PCA bands. The input to the algorithm is a set of hierarchical segmentations corresponding to different PCA bands. The goal is to find coherent groups of segments that correspond to meaningful

Figure 5.1: Example segmentation results (overlaid as white on false color and zoomed) for the *DC Mall* data set. The left, middle and right images show the extracted segments in the first, second and third PCA bands, respectively.

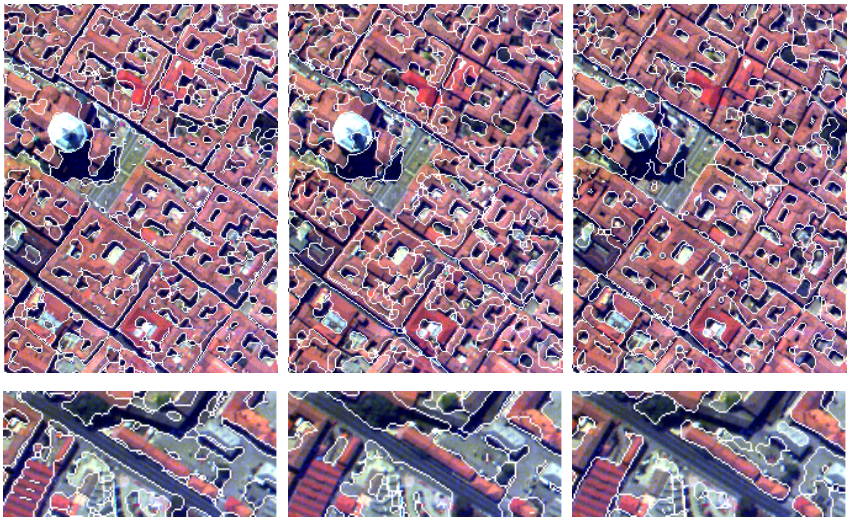


Figure 5.2: Example segmentation results (overlaid as white on false color and zoomed) for the *Centre* data set. The left, middle and right images show the extracted segments in the first, second and third PCA bands, respectively.

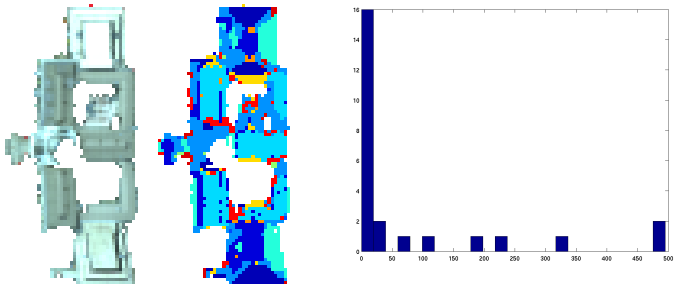
structures. The assumption here is that, for a particular structure (e.g., building), the "good" segments (i.e., the ones containing a building) will all have similar features whereas the "bad" segments (i.e., the ones containing multiple objects or corresponding to overlapping partial object boundaries) will be described by a random mixture of features. Therefore, given multiple objects/structures of interest, this selection process can also be seen as a grouping problem within the space of a large number of candidate segments obtained from multiple segmentations. We use the probabilistic Latent Semantic Analysis (PLSA) algorithm [37] to solve the problem. The resulting groups correspond to different types of objects in the image.

5.1 Modeling Segments

The grouping algorithm consists of three steps: extracting segment features, grouping segments, detecting objects. In the first step, each segment is modeled using the statistical summary of its pixel content. First, all pixels in the image are clustered by applying the k-means algorithm [23] in the spectral (PCA, LDA) and textural (Gabor) feature domains. This corresponds to quantization of the feature values. Then, a histogram is constructed for each segment to approximate the distribution of these quantized values belonging to the pixels in that segment. This histogram is used to represent the segment in the rest of the algorithm. (Note that any discrete model of the segment's content can also be used by the grouping algorithm in the next section.)

5.2 Grouping Segments

In this work, we use the probabilistic Latent Semantic Analysis (PLSA) algorithm [37] to solve the grouping problem. PLSA was originally developed for statistical text analysis to discover topics in a collection of documents that are represented using the frequencies of words from a vocabulary. In our case, the

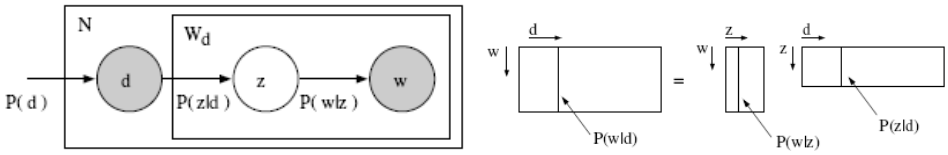


(a) Segment (b) K-means quantization (c) Histogram

Figure 5.3: A segment modeling example. First, all pixels in the image are clustered by the k-means algorithm. The resulting pixel labels of the segment in (a) is shown in (b). Then, a histogram (shown in (c)) is constructed to represent the segment.

documents correspond to image segments, the word frequencies correspond to histograms of pixel-level features, and the topics to be discovered correspond to the set of objects/structures of interest in the image. Russell *et al.* [58] used a different graphical model in a similar setting where multiple segmentations of natural images were obtained using the normalized cut algorithm by changing its parameters, and instances of segments corresponding to objects such as cars, bicycles, faces, sky, etc., were successfully grouped and retrieved from a large data set of images.

The PLSA technique uses a graphical model for the joint probability of the segments and their features in terms of the probability of observing a feature given an object and the probability of an object given the segment. Suppose there are N segments (documents) having content coming from a distribution (vocabulary) with M pixel feature values (words). The collection of segments is summarized in an N-by-M co-occurrence table n where $n(d_i, w_j)$ stores the number of occurrences of feature value w_j in segment d_i. In addition, there is a latent object type (topic) variable z_k associated with each observation, an observation being the occurrence of a feature in a particular segment.



(a) (b)

Figure 5.4: PLSA graphical model. In (a), nodes inside a rectangular box are replicated the number of times in the top left corner. Filled nodes are observed random variables whereas unfilled are unobserved. In PLSA, the object specific spectral probability, $P(w_j|z_k)$, and the segment specific object probability, $P(z_k|d_i)$, are used to compute the segment specific spectral probability, $P(w_j|d_i)$, as shown in (b). (Images taken from [58].)

Let $P(w_j|z_k)$ denote the object-conditional probability of feature w_j occurring in object z_k, and $P(z_k|d_i)$ denote the probability of object z_k observed in segment d_i. PLSA uses the graphical model shown in Figure 5.4(a) for the joint probability $P(w_j, d_i, z_k)$ and the generative model $P(d_i, w_j) = P(d_i)P(w_j|d_i)$ for feature content of segments can be computed using the conditional probability (see Figure 5.4(b) for illustration)

$$P(w_j|d_i) = \sum_{k=1}^{K} P(w_j|z_k)P(z_k|d_i). \tag{5.1}$$

Then, the object specific feature distribution $P(w_j|z_k)$ and the segment specific feature distribution $P(w_j|d_i)$ can be used to determine similarities between object types and segments (explained in the next section).

In PLSA, the goal is to identify the probabilities $P(w_j|z_k)$ and $P(z_k|d_i)$. These probabilities are learned using the Expectation-Maximization (EM) algorithm [37]. In the E-step, the posterior probability of the latent variables are computed based on the current estimates of the parameters as

$$P(z_k|d_i, w_j) = \frac{P(w_j|z_k)P(z_k|d_i)}{\sum_{l=1}^{K} P(w_j|z_l)P(z_l|d_i)}. \tag{5.2}$$

In the M-step, the parameters are updated to maximize the expected complete

data log-likelihood as

$$P(w_j|z_k) = \frac{\sum_{i=1}^{N} n(d_i, w_j)P(z_k|d_i, w_j)}{\sum_{m=1}^{M}\sum_{i=1}^{N} n(d_i, w_m)P(z_k|d_i, w_m)}, \tag{5.3}$$

$$P(z_k|d_i) = \frac{\sum_{j=1}^{M} n(d_i, w_j)P(z_k|d_i, w_j)}{n(d_i)}. \tag{5.4}$$

The E-step and the M-step are iterated until the difference between consecutive expected complete data log-likelihoods is less than a threshold or the number of iterations exceeds a threshold.

5.3 Detecting Objects

After learning the parameters of the model, we want to find good segments belonging to each object type. This is done by comparing the feature distribution within each segment, $p(w|d)$, and the feature distribution for a given object type, $p(w|z)$. The similarity between two distributions can be measured using the Kullback-Leibler (KL) divergence $D(p(w|d)\|p(w|z))$. Then, for each object type, the segments in an image can be sorted according to their KL divergence scores, and the most representative segments for that object type can be selected. However, if there are two segments within an object type extracted from different principal components and at least one of them overlaps with the other by a predetermined percent of its whole area, the less representative (the one with a larger KL divergence score) structure is removed from the topic to avoid showing multiple segments of the same object.

Chapter 6

Region-based Classification

In this chapter, we present an approach for classification of remotely sensed imagery using spatial information extracted from multi-scale segmentations. We model spatial information by segmenting images into spatially contiguous segments and classifying these segments according to the statistical summary of their pixel properties. First, the original image is segmented using clustering-based and mathematical morphology-based algorithms. The parameters of these methods are adjusted so that oversegmented regions are also produced to capture the details of small structures. Then, we perform segmentation on each PCA band using the the algorithm in Chapter 4 with disjoint SE size ranges to model image content in different levels. These levels are used to capture different details inherently found in different structures.

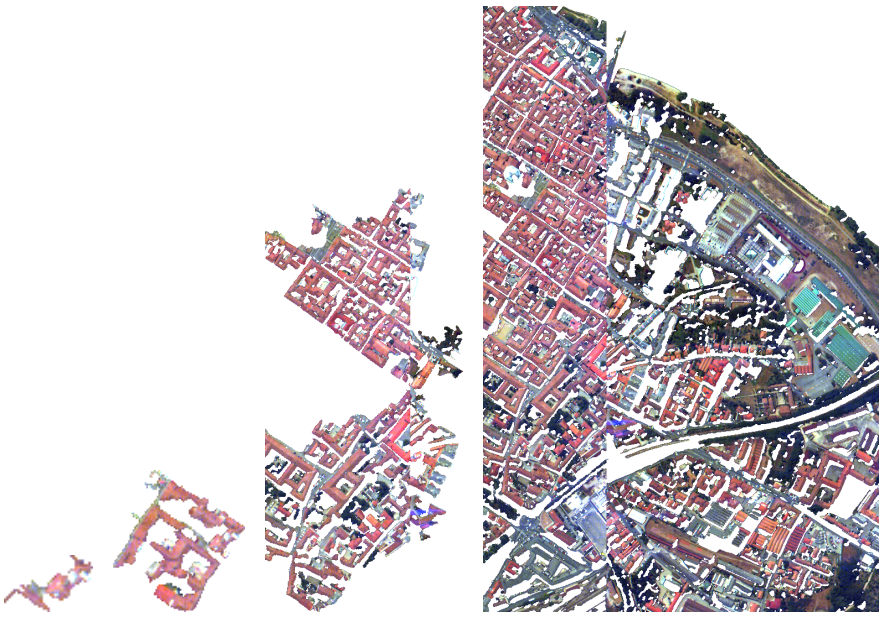
Resulting segments are modeled using the statistical summaries of their spectral properties. Then, these attributes are used as features to cluster the segments. Finally, the cluster memberships assigned to each segment in multiple levels of the scale hierarchy are used to classify the corresponding pixels into land cover/land use categories defined by the user. Final classification is done using decision tree classifiers.

6.1 Multi-scale Segmentation

First, the k-means algorithm [23] is used to cluster the spectral data represented by PCA and LDA, and Gabor texture features. After this unsupervised clustering step, each pixel is assigned the label of the cluster that it belongs in the pixel-level feature space. Since the k-means algorithm uses only pixel-level information and ignores spatial correlations, the resulting segmentation may contain isolated pixels with labels different from those of their neighbors. We use an iterative split-and-merge algorithm [6] to convert this intermediate step to contiguous segments.

This procedure corresponds to a spatial smoothing of the clustering results. The parameters for the algorithms were empirically chosen to produce oversegmented segments to capture the details of small structures.

After an oversegmentation is obtained, in order to model the multi-scale context of each pixel, we apply the algorithm in Chapter 4 on each PCA band with different disjoint SE size range sets instead of applying the algorithm with a fixed SE size range. Then, the initial oversegmentation and each segmentation obtained within each range correspond to different scales on each PCA band. The reason for using different SE size ranges is that different ranges capture different details inherently found in different structures. As the SE size gets coarser, larger structures that provide the general image context can be represented without being convoluted with the details (see Figure 6.2 for an illustration). For example, a part of a building can be identified in the first scale, the whole building can be identified in the second scale, a neighboring group of buildings can be identified in the third scale and then up to the largest scale, the context can be modeled with spatially increasing structures (see Figure 6.1 for an illustration). If the number of scales is increased large enough, at the largest scale the image can be identified as two objects: Land and water. It is therefore natural to analyze first the image content with coarse structures and then gradually analyze small structures [50]. This process is also similar to the strategy used by the human vision system [49].



(a) A building part (b) A single building (c) A building neighborhood (d) A large land area

Figure 6.1: An example hierarchical framework for analyzing the image. A part of a building segment is recognized by considering its context at different scales.

6.2 Classifying Segments

Image classification is usually done by using pixel features as input to classifiers such as minimum distance, maximum likelihood, neural networks or decision trees. However, large within-class variations and small between-class variations of these features at the pixel level and the lack of spatial information limit the accuracy of these classifiers.

In this work, we perform classification using segment level information. First, the segments at all ranges on each PCA band and in the initial oversegmentation are clustered using the algorithm in Chapter 5. The grouping process assigns a cluster label to each segment. These segment level labels can be converted to pixel level features by collecting the labels of the segments at multiple scales corresponding to each pixel. Note that a pixel may not be included in any segment in a scale because the DMP of that pixel does not change within the SE size range of that scale. In this case, the segment label of the pixel is assigned the label of

(a) Oversegmentation (First scale) (b) Second scale (c) Third scale

(d) Fourth scale (e) Fifth scale (f) Sixth scale

Figure 6.2: Multi-scale segmentation example (The first PCA band). We use six successive scales as the multi-scale representation for all data sets. The SE size ranges corresponding to different scales were chosen as: $[3, 8], [9, 13], [14, 23], [24, 43], [44, 73]$. Segment boundaries are marked as white.

the segment in which the pixel is included at the previous scale of the same PCA band. Then, each pixel is assigned a new feature vector of length $c \times a + 1$ where c is the number of PCA bands, and a is the number of scales per PCA band. We add one for the segment label in the initial oversegmentation.

In the next chapter, we evaluate the performance of the new features for classifying pixels into land cover/land use categories defined by the user. Classification is done using a binary decision tree classifier with the gini impurity criterion [23], and its performance is compared to that of a traditional maximum likelihood classifier with the multivariate Gaussian with full covariance matrix assumption for each class.

Chapter 7

Experiments and Results

We applied the proposed segmentation, object detection and classification algorithms to *DC Mall*, *Centre* and *University* data sets.

7.1 Evaluation of Segmentation

First, the tree structure was constructed for each PCA band separately and the regions were selected from each tree individually. Figures 7.1- 7.2 show example segmentation results for *DC Mall*, *Centre* and *University* data sets, respectively. Structuring element sizes from 3 to 15 were used for both opening and closing profiles for both data sets. We present the zoomed versions of the results for several example areas to better illustrate the details for high-resolution imagery and for clarity of the presentation on thesis. The results obtained by the algorithm in [54] are also given for the same areas.

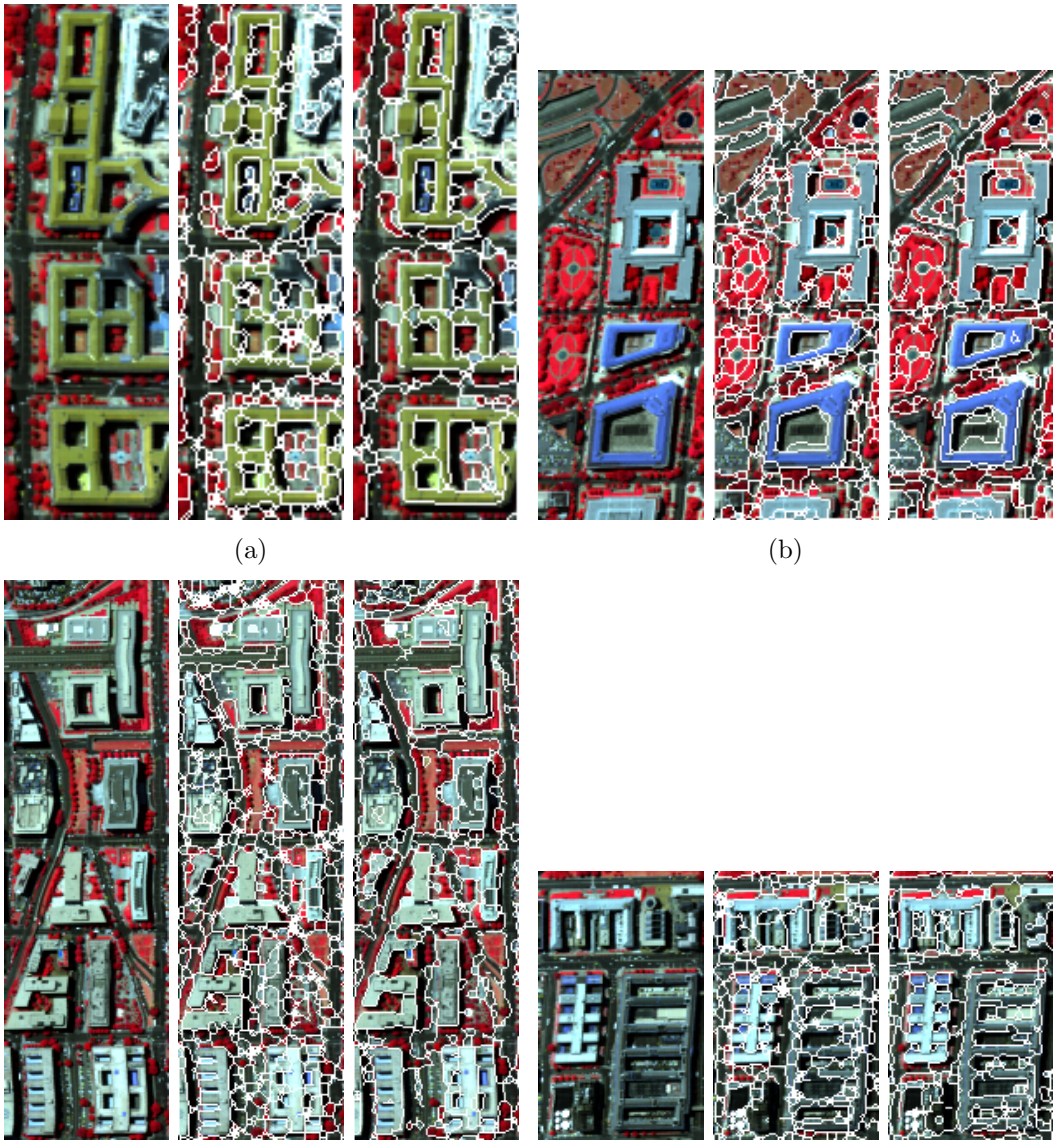
The results show that our segmentation algorithm usually finds structures as a whole but the method of [54] often oversegments them and produces small regions. These small regions occur because the segment label assignment is done for each pixel individually by only considering the greatest value in its DMP. Thus, noisy pixels that are different from their neighborhoods may produce small

regions because they may have large values occurring at scales corresponding to small SE sizes. However, our algorithm considers both the morphological characteristics encoded in the DMP and the spectral information measured in terms of the standard deviation within contiguous groups of pixels. It also considers the consistency of these values within neighboring pixels forming large connected components. As a result, the combined measure that uses both spectral and neighborhood information is both robust to noise and consistent within detailed structures in high-resolution images. In all of the examples, our algorithm is able to extract many meaningful regions as whole segments.

Another important observation is that different structures are extracted more clearly in different principal components. For example, the structures in Figures 7.1(a)-7.1(b) are found in the second PCA band of the *DC Mall* data set like many other buildings. The structures in both Figures 7.2(a), 7.2(c) and 7.2(d) are found in the third PCA band of the *Centre* data set but the structures in Figure 7.2(b) are found in the first PCA band. The structures in Figure 7.3(a) are found in the first PCA band of the *University* data set but the structures in Figure 7.3(b) are found in the third PCA band. The reason that a particular structure being extracted better in a particular PCA band is that the pixels belonging to that structure are found lighter or darker than their surroundings on that PCA band. This motivates the next step on merging the results from individual PCA bands as a final segmentation for an image. As a final note, we also observed that the texture effects produced by vegetation in some of the PCA bands result in small regions in those areas. We will investigate additional multi-spectral features (e.g., NDVI) to improve the segmentation for such regions.

7.2 Evaluation of Object Detection

The first step was hierarchical segment extraction. Disk structuring elements with radii from 3 to 15 were used for both opening and closing profiles for all data sets. The tree structure described in Section 4.2 was constructed for each PCA band separately, and the segments were selected from each tree independently. For the



(a) (b)

(c) (d)

Figure 7.1: Example segmentation results for the *DC Mall* data set. The left image shows the false color representation, the middle one shows the result of the algorithm in [54], and the right one shows the result of the proposed approach.

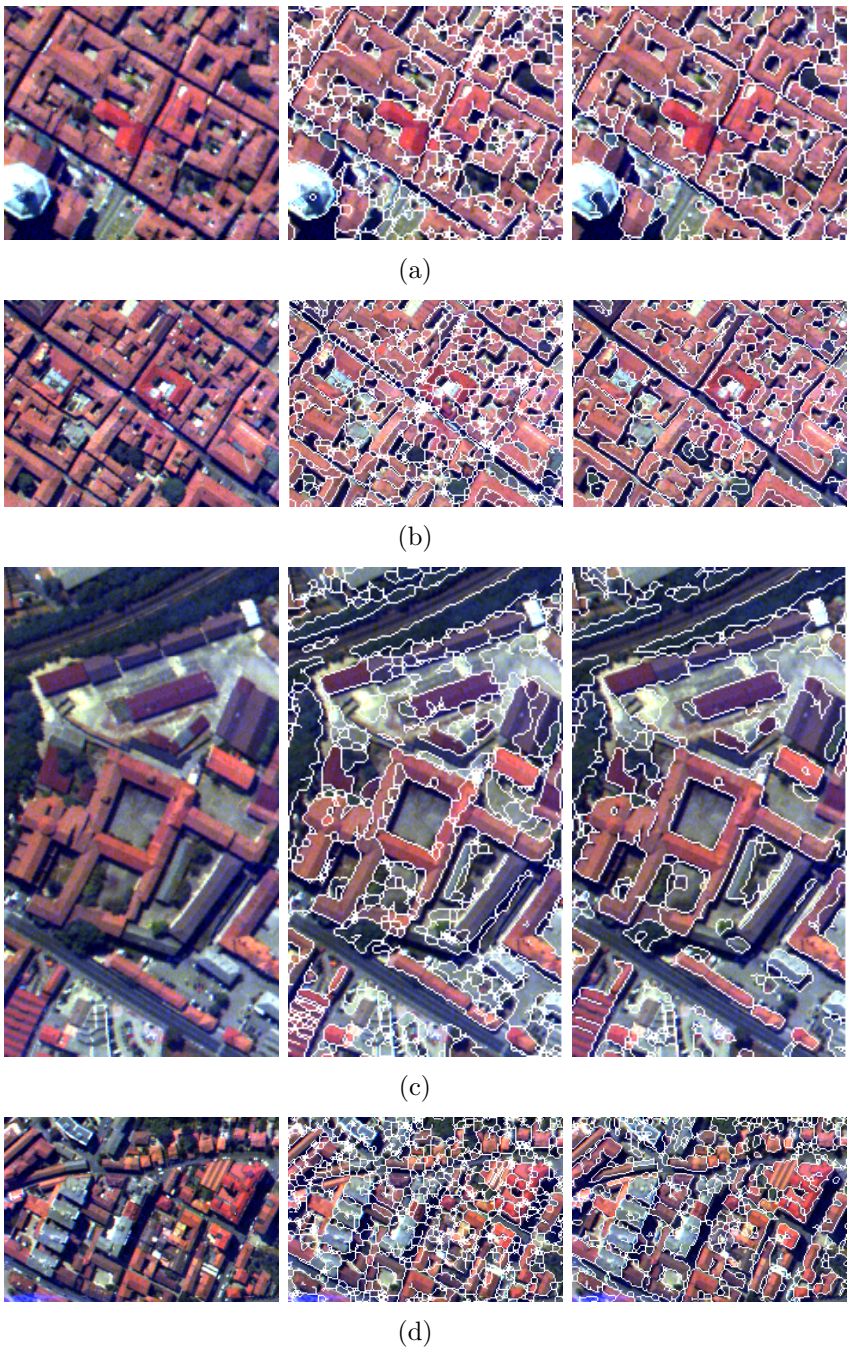
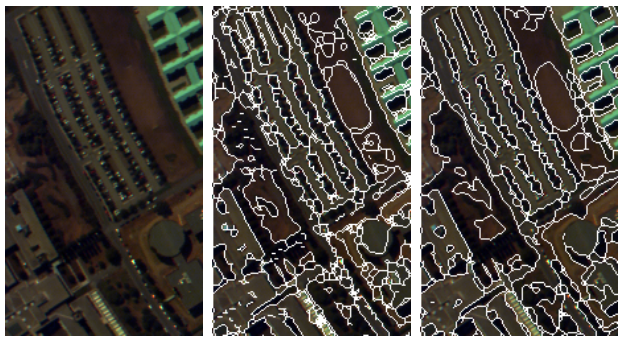
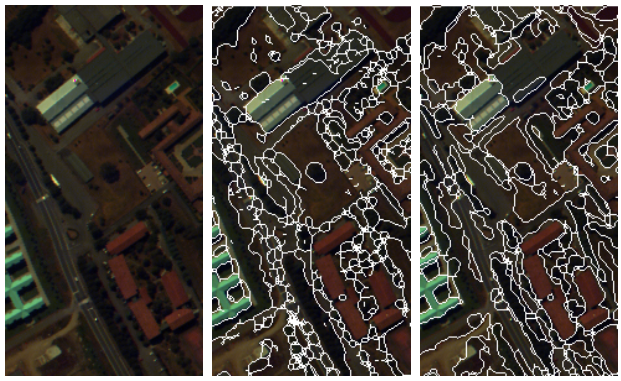


Figure 7.2: Example segmentation results for the *Centre* data set. The left image shows the false color representation, the middle one shows the result of the algorithm in [54], and the right one shows the result of the proposed approach.



(a)



(b)

Figure 7.3: Example segmentation results for the *University* data set. The left image shows the false color representation, the middle one shows the result of the algorithm in [54], and the right one shows the result of the proposed approach.

DC Mall data set, 818, 843 and 796 segments were found in the first, second, and third principal components, respectively. For the *Centre* data set, 1782, 1805 and 1510 segments were found in the first, second, and third principal components, respectively. For the *University* data set, 568, 563, 558 and 378 segments were found in the first, second, third, and fourth principal components, respectively

The next step was to find coherent groups of segments that corresponded to different objects. First, all pixels in the image were clustered using their their PCA values corresponding to the 99% variance, LDA and Gabor bands. The k-means algorithm was used with k empirically selected as 25 for clustering. Then, for each segment, a histogram with 25 bins was constructed by counting the number of pixels belonging to each spectral cluster within that segment. Next, the PLSA algorithm was used to learn the spectral data distributions for the segments and the object types. The number (K) of latent object type variables (z_k) was set to 50 in the experiments. The parameters of the distribution models were learned using the EM algorithm.

In the final step, the KL divergence score between each segment and each object type was computed, and the segments were grouped as belonging to the object type where the KL score was the smallest. Segments within each group were further sorted according to these scores, and the most representative segments for each object type were selected. Since the segments were extracted from different PCA bands, some of the segments could overlap. When the overlap between two segments belonging to the same group was more than 30% of the area of one of the segments, the one with a larger KL divergence score was removed.

Figures 7.4-7.6 show example results for *DC Mall*, *Centre*, and *University* data sets, respectively. The sub-figures b, c, and d of 7.4-7.5 present the segments belonging to the groups that mostly contain buildings, roads, and vegetation respectively. Examination of individual groups showed that segments corresponding to objects (i.e., "good" segments) were mostly placed into coherent groups. For example, man-made structures such as buildings placed in the same group also had very similar spectral characteristics (e.g., roofs with similar colors) and buildings in different groups had different spectral attributes. Similarly, most of the streets

Table 7.1: Precision values on three object types from the *DC Mall* data set.

Building	Road	Vegetation
70.4787	41.0359	73.0469

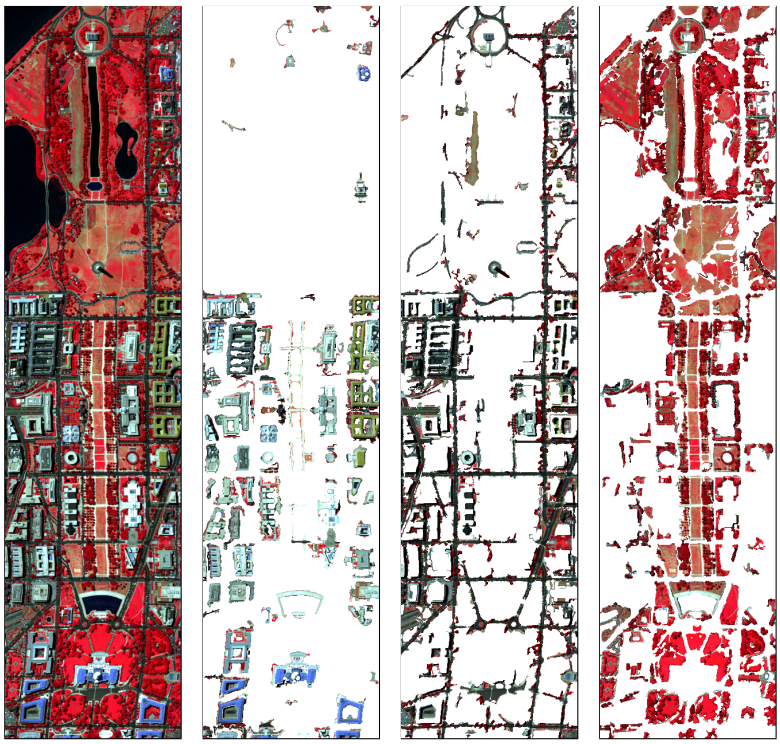
and paths (roads) were grouped correctly. However, there were also some minor confusion caused by shadows and small errors in the initial segmentations. We believe that including new features, in addition to the spectral bands, in the clustering of pixels for modeling the segments will eliminate most of these problems.

Retrieval performance is also evaluated quantitatively on the *DC Mall* data set where each region is labeled with its object type. Table 7.1 shows the precision values for each object type. Average precision is 61.5205%. The proposed method performs better for buildings and vegetation than roads. This is because some segments corresponding to roads are occluded with parts of buildings and vegetation.

Overall, the results show that the proposed algorithm is able to merge the segmentation results from multiple PCA bands by grouping the segments and performing object detection by selecting the most representative segments corresponding to object classes in an unsupervised mode. Future work will include designing automatic methods for selecting the number of object types (topics) in the PLSA algorithm. We will also create object level ground truth for quantitative performance evaluation of *Centre* and *University* data sets.

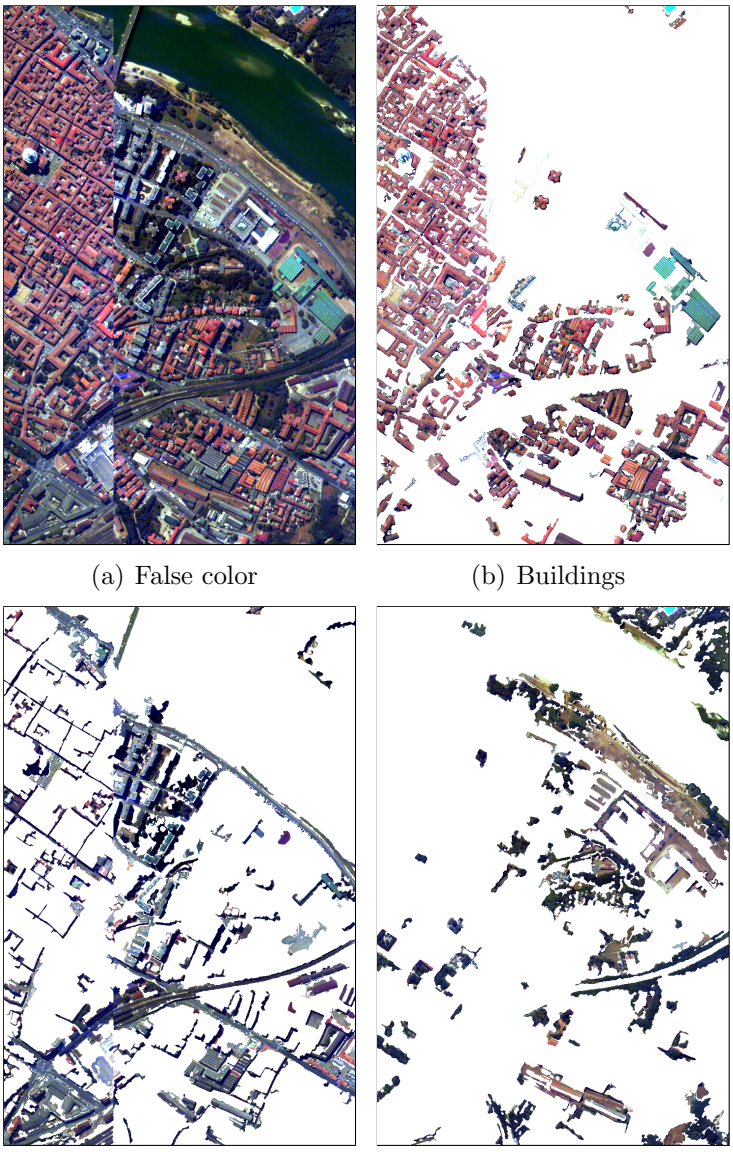
7.3 Evaluation of Classification

We evaluate the proposed feature extraction and classification algorithms both quantitatively and qualitatively. First, the original image is segmented using clustering-based and mathematical morphology-based algorithms as described in Section 6.1 to produce oversegmented segments. The value of k for clustering was set to 20. The minimum area threshold for merging was set to 10 pixels. The



(a) False color (b) Buildings (c) Roads (d) Vegetation

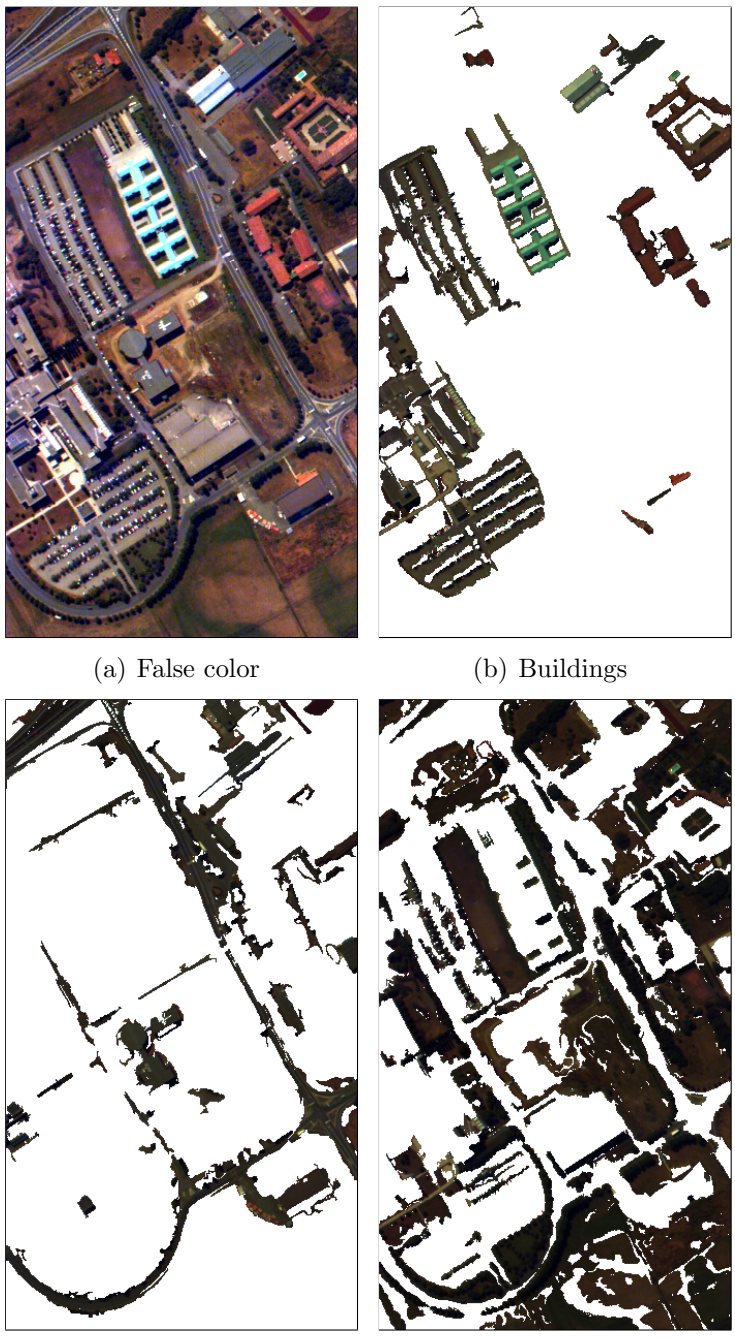
Figure 7.4: Examples of object detection for the *DC Mall* data set.



(a) False color (b) Buildings

(c) Roads (d) Vegetation

Figure 7.5: Examples of object detection for the *Centre* data set.



(a) False color (b) Buildings

(c) Roads (d) Vegetation

Figure 7.6: Examples of object detection for the *University* data set.

neighborhood size for growing was fixed as 3×3. Then, we perform segmentation on each PCA band using the algorithm in Chapter 4 with disjoint SE size ranges to model image content in different levels. We use five successive scales as the multi-scale representation for all data sets. The SE size ranges corresponding to different scales were chosen as: $[3, 8], [9, 13], [14, 23], [24, 43], [44, 73]$. For each range, respectively, the connected components whose number of pixels are greater than 25, 50, 100, 100 and 100 are considered in the segmentation process. Resulting segments are modeled using the statistical summaries of their spectral properties represented by PCA, LDA and Gabor bands. Then, these attributes are used as features to cluster the segments using the algorithm in Chapter 5 where the EM algorithm was used with k empirically selected as 25 for clustering for quantization. The number (K) of latent object type variables (z_k) was set to 20 in the experiments. Finally, the cluster memberships assigned to each segment in multiple levels of the scale hierarchy are used to classify the corresponding pixels into land cover/land use categories defined by the user. In particular, for the *DC Mall*, *Centre* data sets, the length of a pixel's feature vector is 16 and for the *University* data set the length is 21. Final classification is done using decision tree classifiers. The training and test ground truth data are shown in Figures 3.1-3.3.

Confusion matrices for the cases where region-based features were used with the decision tree classifier are shown in Tables 7.2-7.4. The classification performances of both classifiers (region level, quadratic Gaussian) are summarized in Table 7.5. For qualitative comparison, the classification maps for both classifiers for all data sets were computed as shown in Figures 7.7-7.9.

The results show that the proposed approach performed better than the traditional maximum likelihood classifier with Gaussian density assumption for all data sets. Using texture features in addition to the spectral ones improved the performance of both approaches. In addition, using multi-scale spatial information with region features improved the results for the proposed approach further but the maximum likelihood classifier could not avoid producing groups of misclassified pixels due to the lack of spatial information. The most significant improvement was obtained applying region level classification for the *University*

Table 7.2: Confusion matrix when region features were used with the decision tree classifier for the *DC Mall* data set (testing subset). Classes were listed in Fig. 3.1.

		Assigned							Total	% Agree
		roof	street	path	grass	trees	water	shadow		
True	roof	3824	0	0	2	8	0	0	3834	99.7392
	street	0	414	0	0	0	0	2	416	99.5192
	path	0	0	175	0	0	0	0	175	100.0000
	grass	1	0	0	1927	0	0	0	1928	99.9481
	trees	0	0	0	0	405	0	0	405	100.0000
	water	1	0	0	0	0	1223	0	1224	99.9183
	shadow	1	2	0	0	0	0	94	97	96.9072
Total		3815	428	180	1928	405	1224	99	8079	99.7896

Table 7.3: Confusion matrix when region features were used with the decision tree classifier for the *Centre* data set (testing subset). Classes were listed in Fig. 3.2.

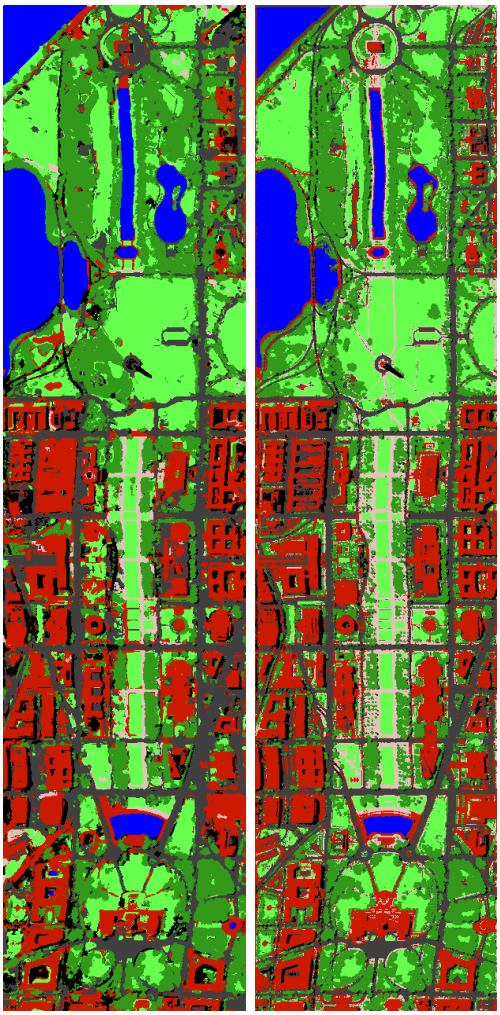
		Assigned									Total	% Agree
		water	trees	meadows	bricks	bare soil	asphalt	bitumen	tiles	shadow		
True	water	65971	0	0	0	0	0	0	0	168	65971	100.0000
	trees	0	6142	1127	113	63	55	45	18	35	7598	80.8371
	meadows	0	155	2842	0	49	3	0	41	0	3090	91.9741
	bricks	0	1	76	2551	39	5	0	0	13	2685	95.0093
	bare soil	1	4	56	944	5318	5	73	42	41	6584	80.7716
	asphalt	0	84	77	23	32	8828	13	81	110	9248	95.4585
	bitumen	0	0	52	14	279	296	6297	336	13	7287	86.4142
	tiles	0	690	56	14	20	35	11	40590	1410	42826	94.7789
	shadow	100	18	79	5	0	170	28	50	2413	2863	84.2822
Total		65842	6352	4288	2634	6364	10009	7023	42380	3260	148152	95.1401

Table 7.4: Confusion matrix when region features were used with the decision tree classifier for the *University* data set (testing subset). Classes were listed in Fig. 3.3.

		Assigned									Total	% Agree
		asphalt	meadows	gravel	trees	m. sheets	bare soil	bitumen	bricks	shadow		
True	asphalt	6078	144	68	22	118	0	0	198	3	6631	91.6604
	meadows	444	15579	0	2014	0	90	302	98	122	18649	83.5380
	gravel	8	0	835	0	2	0	0	1254	0	2099	39.7808
	trees	6	69	1	2956	3	3	19	3	4	3064	96.4752
	m. sheets	0	1	2	0	1342	0	0	0	0	1345	99.7770
	bare soil	0	577	0	0	102	4151	78	0	121	5029	82.5413
	bitumen	0	38	0	0	0	4	1285	0	3	1330	96.6165
	bricks	71	8	353	3	0	1	0	3246	0	3682	88.1586
	shadow	88	0	6	0	1	0	0	23	829	947	87.5396
Total		5102	18299	2237	4227	1350	4110	1521	4706	1224	42776	84.8630

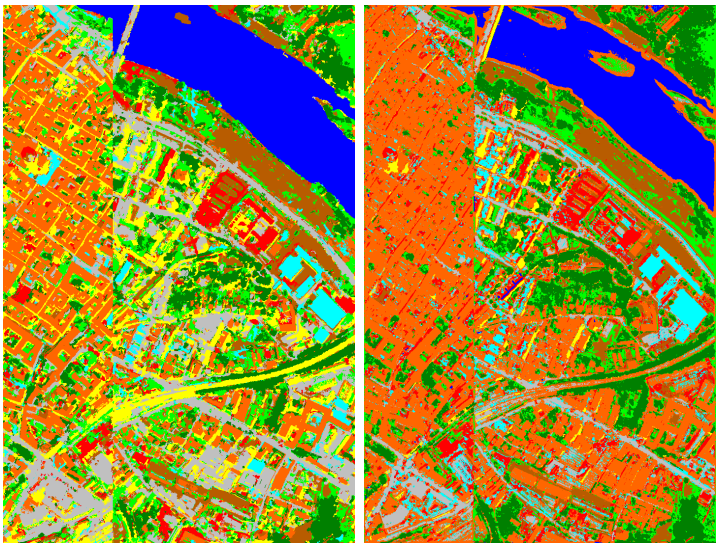
Table 7.5: Summary of classification accuracies using the region level classifier and the quadratic Gaussian classifier.

	DC Mall	*Centre*	*University*
Region level	99.7896	95.1401	84.8630
Quadratic Gaussian	99.3811	93.9677	81.2792



(a) Region level (b) Quadratic Gaussian

Figure 7.7: Final classification maps with the region level classifier and the quadratic Gaussian classifier for the *DC Mall* data set. Class color codes were listed in Figure 3.1. ((b) taken from [3].)



(a) Region level (b) Quadratic Gaussian

Figure 7.8: Final classification maps with the region level classifier and the quadratic Gaussian classifier for the *Centre* data set. Class color codes were listed in Figure 3.2. ((b) taken from [3].)

(a) Region level (b) Quadratic Gaussian

Figure 7.9: Final classification maps with the region level classifier and the quadratic Gaussian classifier for the *University* data set. Class color codes were listed in Figure 3.3. ((b) taken from [3].)

data set. The performances of the pixel level classifier for *DC Mall* and *Centre* data sets were already quite high. In all cases, region level classification performed better than the traditional pixel level classifier.

An important observation is that even though high classification accuracies are seen in the numerical results, the classification maps of especially the pixel level classifier include wrongly classified pixels for all images due to the lack of spatial information [3]. The reason is that those wrongly classified pixels are not included in test ground truth maps. Some examples are the upper part of the *DC Mall* data, tiles on the left of the *Centre* data and many areas in the *University* data. For a more reliable evaluation of the methods, the ground truths must include many more pixels.

Chapter 8

Conclusions and Future Work

In this thesis, we firstly described a method for segmentation of urban structures in high-resolution images. The first step was to extract structural information using morphological opening and closing by reconstruction operators. Principal components analysis bands were used to summarize hyper-spectral data and the morphological operators were applied to each band separately. Then, candidate segments were extracted by applying connected components analysis to the pixels selected according to their morphological profiles obtained using increasing structuring element sizes. Next, these segments were represented using a tree, and the most meaningful ones were selected by optimizing a measure that consisted of two factors: spectral homogeneity, which was calculated in terms of variances of spectral features, and neighborhood connectivity, which was calculated using sizes of connected components. We evaluated the proposed approach on three data sets. The experiments showed that our method that considers morphological characteristics, spectral information, and their consistency within neighboring pixels is able to detect structures in the image which are more precise and more meaningful than the structures detected by another approach that does not make strong use of neighborhood and spectral information.

Then, we described an unsupervised method for automatic selection of segments corresponding to meaningful structures among a set of candidate segments from multiple hierarchical segmentations. Segmentation was done by combining

structural information extracted by morphological processing with spectral information summarized using principal components analysis. Segments that maximized a measure consisting of spectral homogeneity and neighborhood connectivity were selected as candidate structures for object detection. The segments coming from multiple PCA bands were grouped using the probabilistic Latent Semantic Analysis algorithm where the resulting groups of coherent segments corresponded to different object types. We evaluated the proposed approach on three data sets. The experiments showed that our method is able to automatically detect and group structures belonging to the same object classes.

Finally, we presented an approach for classification of remotely sensed imagery using multi-scale and spatial techniques. We model image content at different scales by applying the proposed segmentation algorithm with disjoint SE size ranges to obtain contiguous segments at each scale. The resulting segments were modeled using the statistical summaries of their pixel-level content. Then, these models were used to cluster the regions by the proposed clustering algorithm, and the cluster labels assigned to each segment in multiple scales were used to classify the corresponding pixels with a decision tree classifier. We investigated the performance of multi-scale analysis and region features in classification. Experiments with two data sets showed the effectiveness of the proposed approach over the traditional maximum likelihood classifier because of the use of spatial information extracted from multi-scale segmentations.

Even though the results look satisfactory both numerically and visually, better evaluation of the classification technique is needed. This motivates an important future work on gathering ground truth data with large coverage. We will also prepare ground truth data for quantitative evaluation of segmentation and object detection techniques.

Bibliography

[1] H. G. Akcay and S. Aksoy. Automated detection of objects using multiple hierarchical segmentations. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, Barcelona, Spain, July 2007.

[2] H. G. Akcay and S. Aksoy. Morphological segmentation of urban structures. In *Proceedings of 4th IEEE GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, Paris, France, April 11–13, 2007.

[3] S. Aksoy. Spatial techniques for image classification. In C. Chen, editor, *Signal and Image Processing for Remote Sensing,*. Taylor & Francis, 2006.

[4] S. Aksoy and H. G. Akcay. Multi-resolution segmentation and shape analysis for remote sensing image classification. In *Proceedings of 2nd International Conference on Recent Advances in Space Technologies*, Istanbul, Turkey, June 9-11 2005.

[5] S. Aksoy and R. M. Haralick. Feature normalization and likelihood-based similarity measures for image retrieval. *Pattern Recognition Letters*, 22(5):563–582, May 2001.

[6] S. Aksoy, K. Koperski, C. Tusk, G. Marchisio, and J. C. Tilton. Learning Bayesian classifiers for scene classification with a visual grammar. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):581–589, March 2005.

[7] P. M. Atkinson and A. R. L. Tatnall. Neural networks in remote sensing. *International Journal of Remote Sensing*, 18(4):699–709, November 1997.

[8] S. Bandyopadhyay and S. K. Pal. Pixel classification using variable string genetic algorithms with chromosome differentiation. *IEEE Transactions on Geoscience and Remote Sensing*, 39(2):303–308, February 2001.

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

[10] J. A. Benediktsson, M. Pesaresi, and K. Arnason. Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. *IEEE Transactions on Geoscience and Remote Sensing*, 41(9):1940–1949, September 2003.

[11] J. A. Benediktsson, P. H. Swain, and O. K. Ersoy. Neural network approaches versus statistical methods in classification of multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):540–552, July 1990.

[12] S. Bhagavathy and B. Manjunath. Modeling and detection of geospatial objects using texture motifs. *IEEE Transactions on Geoscience and Remote Sensing*, 44(12):3706–3715, December 2006.

[13] H. Bischof, W. Schenider, and A. J. Pinz. Multispectral classification of landsat images using neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 30(3):482–490, May 1992.

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

[15] H. Cetin, T. A. Warner, and D. W. Levandowski. Data classification, visualization, and enhancement using n-dimensional probability density functions (npdf): Aviris, tims, tm, and geophysical applications. *Photogramm. Eng. Remote Sens.*, 59(12):1003–1109, 1993.

[16] C. I. Chang. Target signature-constrained mixed pixel classification for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 40(5):1065–1081, May 2002.

[17] C. I. Chang and C. M. Brumbley. A kalman filtering approach to multispectral image classification and detection of changes in signature abundance. *IEEE Transactions on Geoscience and Remote Sensing*, 37(1):257–268, January 1999.

[18] C. I. Chang and H. Ren. An experiment-based quantitative and comparative analysis of target detection and image classification algorithms for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 38(2):1044–1063, March 2000.

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

[20] F. DellAcqua, P. Gamba, A. Ferrari, J. A. Palmason, J. A. Benediktsson, and K.Arnason. Exploiting spectral and spatial information in hyperspectral urban data with high resolution. *IEEE Geoscience and Remote Sensing Letters*, 1(4):322–326, October 2004.

[21] X. Descombes, M. Sigelle, and F. Preteux. Estimating Gaussian Markov random field parameters in a nonstationary framework: application to remote sensing imaging. *IEEE Transactions on Image Processing*, 8(4):490–503, April 1999.

[22] P. Dreyer. Classification of land cover using optimized neural nets on spot data. *Photogramm. Eng. Remote Sens.*, 59(5):617–621, 1993.

[23] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, Inc., 2000.

[24] S. S. Durbha and R. L. King. Knowledge mining in earth observation data archives: a domain ontology perspective. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, volume 1, September 2004.

[25] C. Evans, R. Jones, I. Svalbe, and M. Berman. Segmenting multispectral Landsat TM images into field units. *IEEE Transactions on Geoscience and Remote Sensing*, 40(5):1054–1064, May 2002.

[26] S. E. Franklin. Discrimination of subalpine forest species and canopy density using digital casi, spot pla, and landsat tm data. *Photogrammetric Engineering and Remote Sensing*, 60(10):1233–1241, 1994.

[27] S. E. Franklin. Parametric and nearest-neighbor methods for hybrid classification: A comparison of pixel assignment accuracy. *Photogramm. Eng. Remote Sens.*, 60(12):1439–1448, 1994.

[28] G. Frizzelle and A. Moody. Mapping continuous distributions of land cover: A comparison of maximum-likelihood estimation and artificial neural networks. *Photogramm. Eng. Remote Sens.*, 67(6):693–705, 2001.

[29] P. Gamba and B. Houshmand. An efficient neural classification chain of sar and optical urban images. *International Journal of Remote Sensing*, 22(8):1535–1553, 2001.

[30] G. Hamerly and C. Elkan. Learning the k in kmeans. In *Advances in Neural Information Processing Systems (NIPS)*, volume 17, 2003.

[31] M. Hansen, R. Dubayah, and R. Defries. Classification trees: An alternative to traditional land cover classifiers. *International Journal of Remote Sensing*, 17(5):1075–1081, 1996.

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

[33] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision*. Addison-Wesley, 1992.

[34] J. C. Harsanyi and C. I. Chang. Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach. *IEEE Transactions on Geoscience and Remote Sensing*, 32(4):779–785, July 1994.

[35] G. G. Hazel. Object-level change detection in spectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(3):553–561, March 2001.

[36] P. D. Heerman and N. Khazenie. Classification of multispectral remote sensing data using a backpropagation neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 30(1):81–88, January 1992.

[37] T. Hofmann. Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, 42(1–2):177–196, January–February 2001.

[38] C. Y. Ji. Land-use classification of remotely sensed data using kohonen self-organizing feature map neural networks. *Photogramm. Eng. Remote Sens.*, 66(12):1451–1460, 2000.

[39] 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, January 1999.

[40] I. Kanellopoulos and G. G. Wilkonson. Strategies and best practice for neural network image classification. *International Journal of Remote Sensing*, 18(4):711–725, November 1997.

[41] A. Katartzis, I. Vanhamel, and H. Sahli. A hierarchical markovian model for multiscale region-based classification of vector-valued images. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):548–558, March 2005.

[42] R. L. Kettig and D. A. Landgrebe. Classification of multispectral image data by extraction and classification of homogeneous objects. *IEEE Transactions on Geoscience Electronics*, GE-14(1):19–26, January 1976.

[43] A. S. Kumar and K. L. Majumder. Information fusion in tree classifiers. *International Journal of Remote Sensing*, 22(5):861–869, 2001.

[44] S. Kumar, J. Ghosh, and M. M. Crawford. Best bases feature extraction algorithms for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7):1368–1379, July 2001.

[45] T. M. Kuplich, C. C. Freitas, and J. V. Soares. The study of ers-1 sar and landsat tm synergism for land use classification. *International Journal of Remote Sensing*, 21(10):2101–2111, 2000.

[46] H. E. T. Kusaka and Y. Kawata. Classification of the spot image using spectral and spatial features of primitive regions that have nearly uniform color. *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):749–752, July 1990.

[47] D. A. Landgrebe. *Signal Theory Methods in Multispectral Remote Sensing*. John Wiley & Sons, Inc., 2003.

[48] J. Li and R. M. Narayanan. Integrated spectral and spatial information mining in remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 42(3):673–685, March 2004.

[49] S. Mallat. Wavelets for a vision. *P-IEEE*, 84:604–614, 1996.

[50] S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(2):674–693, July 1989.

[51] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):837–842, August 1996.

[52] P. W. Mausel, W. J. Kramber, and J. K. Lee. Optimum band selection for supervised classification of multispectral data. *Photogrammetric Engineering and Remote Sensing*, 56:55–60, 1990.

[53] J. D. Paola and R. A. Schowengerdt. A review and analysis of backpropagation neural networks for classification of remotely sensed multispectral imagery. *International Journal of Remote Sensing*, 16(16):3033–3058, 1995.

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

[55] A. J. Plaza and J. C. Tilton. Automated selection of results in hierarchical segmentations of remotely sensed hyperspectral images. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, volume 7, pages 4946–4949, Seoul, Korea, July 25–29, 2005.

[56] H. Ren and C. I. Chang. A generalized orthogonal subspace projection approach to unsupervised multispectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 38(6):2515–2528, November 2000.

[57] S. Roessner, U. Heiden, and H. Kaufmann. Automated differentiation of urban surface based on airborne hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7):1525–1532, July 2001.

[58] B. C. Russell, A. A. Efros, J. Sivic, W. T. Freeman, and A. Zisserman. Using multiple segmentations to discover objects and their extent in image collections. In *Proceedings of IEEE Transactions on Computer Vision and Pattern Recognition*, volume 2, pages 1605–1614, New York, NY, June 17–22, 2006.

[59] A. Rydberg and G. Borgefors. Integrated method for boundary delineation of agricultural fields in multispectral satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(11):2514–2520, November 2001.

[60] A. Sarkar, M. K. Biswas, B. Kartikeyan, V. Kumar, K. L. Majumder, and D. K. Pal. A MRF model-based segmentation approach to classification for multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 40(5):1102–1113, May 2002.

[61] S. B. Serpico and F. Roli. Classification of multisensor remote sensing images by structured neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 33(3):562–578, May 1995.

[62] A. K. Shackelford and C. H. Davis. A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas. *IEEE Transactions on Geoscience and Remote Sensing*, 41(9):1920–1932, September 2003.

[63] C. Small. Estimation of urban vegetation abundance by spectral mixture analysis. *International Journal of Remote Sensing*, 22(7):1305–1334, 2001.

[64] P. C. Smits and A. Annoni. Updating land cover maps by using texture information form very high-resolution space-born imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3):1244–1254, May 1999.

[65] L. K. Soh, C. Tsatsoulis, D. Gineris, and C. Bertoia. Arktos: an intelligent system for sar sea ice image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 2(1):229–248, January 2004.

[66] J. C. Tilton. Analysis of hierarchically related image segmentations. In *Proceedings of IEEE GRSS Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, Washington, DC, October 27–28, 2003.

[67] J. C. Tilton, G. Marchisio, K. Koperski, and M. Datcu. Image information mining utilizing hierarchical segmentation. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, volume 2, pages 1029–1031, Toronto, Canada, June 2002.

[68] F. Tsai and W. D. Philpot. A derivative aided hyperspectral image analysis system for land cover classification. *IEEE Transactions on Geoscience and Remote Sensing*, 40(2):416–425, February 2002.

[69] B. C. K. Tso and P. M. Mather. Classification ofmultisource remote sensing imagery using a genetic algorithm and markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3):1255–1260, May 1999.

[70] C. Unsalan. *Multispectral Satellite Image Understanding*. PhD thesis, The Ohio State University, 2003.

[71] C. Unsalan and K. L. Boyer. A theoretical and experimental investigation of graph theoretical measures for land development in satellite imagery. *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, 27(4):575–589, April 2005.

[72] J. Verhoeye and R. D. Wulf. Land cover mapping at sub-pixel using linear optimization techniques. *Remote Sensing of Environment*, 79:96–104, 2002.

[73] J. S. Weszka, C. R. Dyer, and A. Rosenfeld. A comparative study of texture measures for terrain classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-6:269–285, 1976.

[74] G. G. Wilkinson. Results and implications of a study of fifteen years of satellite image classification experiments. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):433–440, March 2005.

[75] T. Yoshida and S. Omatu. Neural network approach to land cover mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 32(5):1103–1109, September 1994.

[76] S. Yu, M. Berthod, and G. Giraudon. Toward robust analysis of satellite images using map informationapplication to urban area detection. *IEEE Transactions on Geoscience and Remote Sensing*, 37(4):1925–1939, May 1999.