UNSUPERVISED DETECTION OF COMPOUND STRUCTURES USING IMAGE SEGMENTATION AND GRAPH-BASED TEXTURE ANALYSIS

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
Daniya Zamalieva

August, 2009

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.

Prof. Dr. Enis Çetin

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. Volkan Atalay

Approved for the Institute of Engineering and Science:

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

ABSTRACT

UNSUPERVISED DETECTION OF COMPOUND STRUCTURES USING IMAGE SEGMENTATION AND GRAPH-BASED TEXTURE ANALYSIS

Daniya Zamalieva
M.S. in Computer Engineering
Supervisor: Asst. Prof. Dr. Selim Aksoy
August, 2009

The common goal of object-based image analysis techniques in the literature is to partition the images into homogeneous regions and classify these regions. However, such homogeneous regions often correspond to very small details in very high spatial resolution images obtained from the new generation sensors. One interesting way of enabling the high-level understanding of the image content is to identify the image regions that are intrinsically heterogeneous. These image regions are comprised of primitive objects of many diverse types, and can also be referred to as compound structures. The detection of compound structures can be posed as a generalized segmentation or generalized texture detection problem, where the elements of interest are primitive objects instead of traditional case of pixels. Traditional segmentation methods extract regions with similar spectral content and texture models assume specific scale and orientation. Hence, they cannot handle the complexity of compound structures that consist of multiple regions with different spectral content and arbitrary scale and orientation.

In this thesis, we present an unsupervised method for discovering compound image structures that are comprised of simpler primitive objects. An initial segmentation step produces image regions with homogeneous spectral content. Then, the segmentation is translated into a relational graph structure whose nodes correspond to the regions and the edges represent the relationships between these regions. We assume that the region objects that appear together frequently can be considered as strongly related. This relation is modeled using the transition frequencies between neighboring regions, and the significant relations are found as the modes of a probability distribution estimated using the features of these transitions. Furthermore, we expect that subgraphs that consist of groups of strongly related regions correspond to compound structures. Therefore, we employ two different procedures to discover the subgraphs in the constructed graph. During the first procedure the graph is discretized and a graph-based knowledge discovery algorithm is applied to find the repeating subgraphs. Even though a single subgraph does not exclusively correspond to a particular compound structure, different subgraphs constitute parts of different compound structures. Hence, we discover compound structures by clustering the histograms of the subgraph instances with sliding image windows. The second procedure involves graph segmentation by using normalized cuts. Since the distribution of significant relations within resulting subgraphs gives an idea about the nature of corresponding compound structure, the subgraphs are further grouped by clustering the histograms of the most significant relations.

The proposed method was tested using an Ikonos image. Experiments show that the discovered image areas correspond to different high-level structures with heterogeneous content such as dense residential areas with high buildings, dense and sparse residential areas with low height buildings and fields.

Keywords: Image segmentation, object detection, texture analysis, graph-based analysis.

ÖZET

BİLEŞİK YAPILARIN GÖRÜNTÜ BÖLÜTLEME VE ÇİZGE TABANLI DOKU ANALİZİ İLE ÖĞRETİCİSİZ BULUNMASI

Daniya Zamalieva
Bilgisayar Mühendisliği, Yüksek Lisans
Tez Yöneticisi: Y. Doç. Dr. Selim Aksoy
August, 2009

Literatürdeki nesnesel görüntü analizi tekniklerinin ortak amacı görüntü türdeş bölgelere bölütlemek ve bunları sınıflandırmaktır. Fakat bu türdeş bölgeler, yeni nesil algılayıcılardan elde edilen yüksek uzamsal çözünürlüklü görüntülerde çok küçük detaylara karşılık gelmektedir. Görüntü içeriğini üst düzeyde anlamamızı sağlayan dikkate değer bir yöntem içsel olarak heterojen bölgelerin tanımlanmasıdır. Farklı tip temel nesnelerin birleşmesinden oluşan bu tür imge bölgeleri *bileşik yapılar* olarak da adlandırılır. Bileşik yapıların saptanması, pikseller yerine temel nesneler kullanan genellenmiş bölütleme veya doku analizi problemi olarak görülebilir. Geleneksel bölütleme yöntemleri benzer spektral içerikli bölgeleri bulurken, doku bulma teknikleri ise belirli bir ölçek ve yönelim gerektirir. Bundan dolayı bu iki teknik de değişik spektral içerik ve gelişigüzel ölçek ve yönelimli bileşik yapıların karmaşıklığıyla başa çıkamamaktadır.

Bu tez çalışmasında temel nesnelerden oluşan bileşik görüntü yapılarının bulunmasını sağlayan öğreticisiz bir yöntem önerilmektedir. İlk bölütleme adımı homojen spektral içerikli görüntü bölgeleri üretir. Sonrasında bölütleme sonuçları, düğümleri bölgeler ve kenarları bölgeler arasındaki ilişkiler olan bir ilişkisel çizgeye aktarılır. Birlikte sıkça görülen bölgeler çok ilgili olarak değerlendirilir. Bu ilişki komşu bölgelerdeki geçişlerin sıklığına bağlı olarak modellenir ve önemli ilişkiler, geçişlerin öznitelikleri kullanılarak oluşturulan olasılık dağılımındaki yerel enbüyük olarak bulunur. Ayrıca çok ilgili bölgeler içeren altçizgeler de bileşik yapılara karşılık gelmektedir. Bu yüzden kurulan çizgedeki altçizgeleri ortaya çıkarmak için iki farklı yöntem kullanılmaktadır. İlk yöntemde çizge ayrıklaştırılır ve tekrar eden altçizgeler çizge bazlı bilgi çıkarma algoritmasıyla bulunur. Tek başına bir altçizge belirli bir bileşik yapıya karşılık gelmese bile farklı altçizgeler bir bileşik yapının parçaları olabilir. Bundan dolayı bileşik yapılar, altçizgeler histogramlarının kayar imge pencereleri ile gruplandırılmaları sayesinde bulunur. İkinci yöntem düzgelenmiş kesitler algoritmasıyla çizge bölütlemesi içerir. Önemli ilişkilerin altçizgelerdeki dağılımı bize bileşik yapılar hakkında bir fikir vereceğinden, altçizgeler en önemli ilişkiler histogramı ile tekrar gruplandırılır.

Önerilen yöntem Ikonos görüntülerinde test edilmiştir. Deneyler sonucunda bulunan bölgelerin yüksel yoğunluklu yerleşim alanı, düşük yoğunluklu yerleşim alanı ve arazi gibi heterojen içerikli farklı üst düzey yapılara karşılık geldiği görülmüştür.

Anahtar sözcükler: Görüntü bölütleme, nesne sezimi, doku analizi, çizge tabanlı analiz.

Acknowledgement

I would like to express my deep thanks to my supervisor Asst. Prof. Dr. Selim Aksoy for his guidance and support throughout this work. It has been a valuable experience for me to work with him and benefit from his vision and knowledge in every step of my research.

I am also very thankful to Prof. Dr. Enis Çetin and Prof. Dr. Volkan Atalay for their suggestions on improving this work.

Certainly, I appreciate my family for their endless love, support and patience.

Besides, I would like to express my pleasure on being a part of RETINA team, and having such a nice friendship with the group members. Especially I would like to thank Onur, Sare, Aslı, Fırat and Bahadır for their support.

Finally I would like to thank Dr. James C. Tilton from NASA Goddard Space Flight Center for his valuable comments and provision of the RHSEG software.

This work was supported in part by the TUBITAK CAREER grant 104E074.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

1.1 Overview

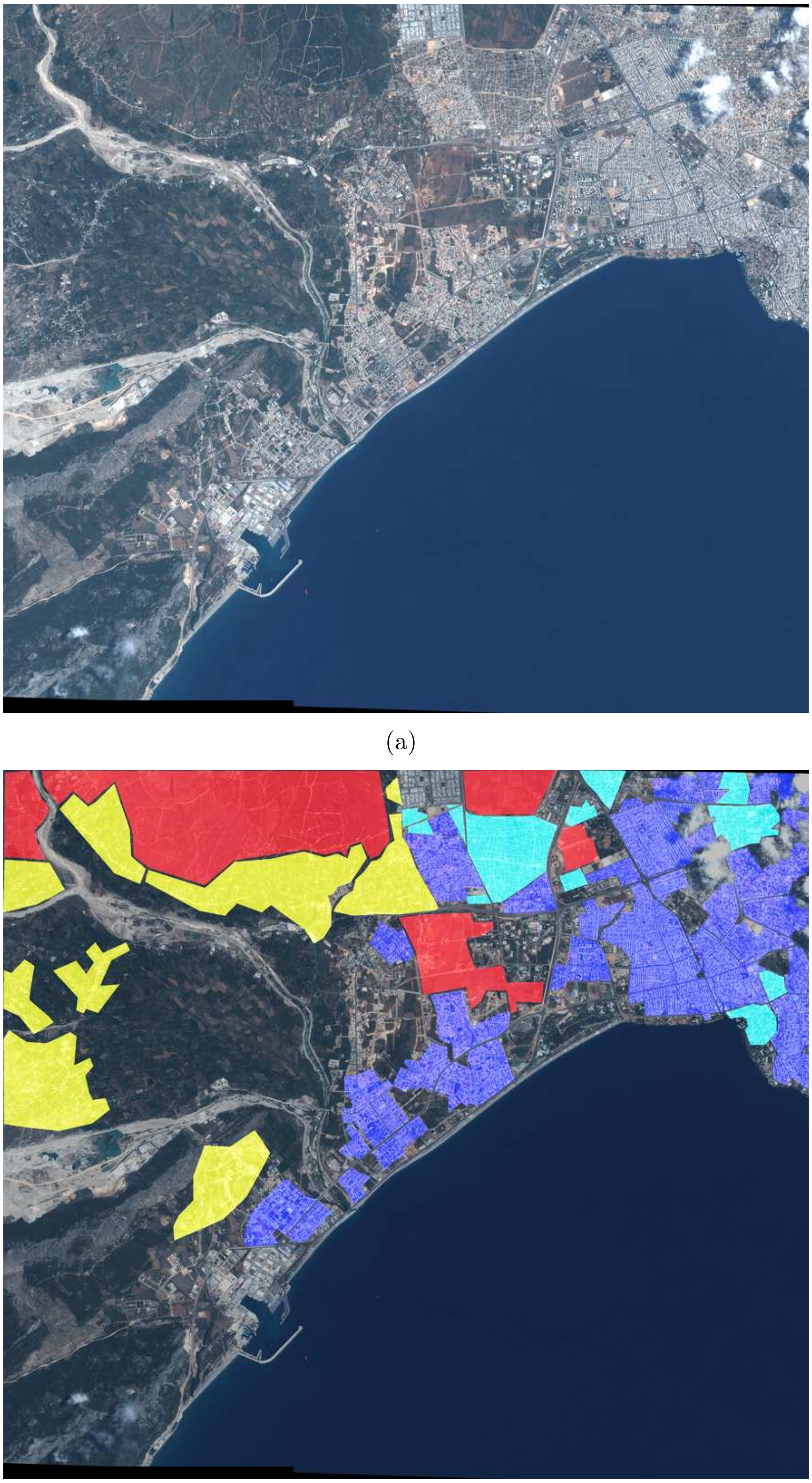
Constant increase in the amount of available high-resolution remotely sensed data is subsequently causing the demand for applications that aim automatic information extraction. A lot of effort has been spent on pixel-based analysis techniques [18]; however, several studies have shown that most of them are not competent enough to show high performance on this kind of data. To address this problem, the field of object-based image analysis has arisen in recent years [6].

The common goal of object-based image analysis techniques in the literature is to partition the images into homogeneous regions and classify these regions. However, such homogeneous regions often correspond to very small details in very high spatial resolution images obtained from the new generation sensors. One interesting way of enabling the high-level understanding of the image content is to identify the image regions that are intrinsically heterogeneous. These image regions are comprised of primitive objects of many diverse types, and can also be referred to as compound structures.

The compound structures generally correspond to high-level structures such as sparse and dense urban areas, forests, industrial and agricultural areas (see Figure 1.1). Thus, the identification of compound structures provides high level of abstraction beyond object-level analysis. In contrast to primitive objects (buildings, roads, etc.), the compound structures are able to capture more of the image content, and subsequently better summarize the scene. For high-level information extraction tasks, such as automated annotation of geospatial images, this is an inevitable and necessary step due to complexity and variability of object-level representation. Compound structures can also be used as contextual information for other detection or retrieval tasks.

However, the delineation of compound structures is a challenging task and most of the challenge originates from the nature of the compound structures. Since they are characterized by a mixture of primitives of several types, there is no limitation on the number or type of primitives within the compound structures and the amount of variation among the instances of the same type. While several segmentation algorithms have been proposed to partition images into homogeneous regions, the detection of meaningful regions that are internally heterogeneous is not a well-explored task. Hence, in order to obtain the compound structures further exploration must be performed.

A number of methods have been proposed for detection of compound structures of predefined types. These methods generally rely on a particular characteristics of a given compound structure type. For example, methods that aim detection [27] or classification [32, 11] of urban areas depend either on detection of buildings or their specific properties. For example, Stasolla and Gamba [27] proposed a procedure for extraction of human settlements from high-resolution synthetic aperture radar (SAR) images that uses the bright response production property of buildings. Dogrusoz and Aksoy performed classification of settlement areas as organized and unorganized by first detecting the buildings and then using them as primitives in both statistical [2] and structural [11] texture models. Unsalan and Boyer [32] suggested contructing graph where photometric straight line segments extracted from grayscale images are assigned to vertices and their spatial relationships are encoded by edges. They introduced a set of measures based on various properties of the graph and used these measures for classification



(a)

(b)

Figure 1.1: An Ikonos image of Antalya with 3551 × 3128 pixel size and 4 m spatial resolution, and some compound structures of interest: dense and sparse residential areas with different building size and fields.

of scenes as rural, residential and urban assuming that impact of human activity causes emergence of straight and smoothly curved contours and their spatial density and regularity increases with increasing development. One other example is the method proposed for the detection of harbors and golf courses [5], that relies on characteristic texture properties, namely on spatially recurrent patterns that are formed by boats and water in a harbor and trees and grass in golf courses.

Clearly, to provide the detection of compound structures regardless of their types, a generic unsupervised method that does not rely on particular properties of a certain compound structure type must be presented. This can be posed as a generalized segmentation problem because the goal is the delineation of regions of interest. However, traditional segmentation methods extract regions with uniform spectral content and cannot be used for detection of intrinsically heterogeneous regions. On the other hand, this is is also a generalized texture detection problem, because compound structures consist of spatial arrangements of image primitives. Traditional methods for texture detection that include co-occurrence matrix [22], Fourier transform [11, 5], and the autocorrelation functions [27] require the selection of specific scale and orientation which are not stable for compound structures. Standart texture models can perform well for detection or classification of compound structures when the image resolution is low, so that the level of detail is reduced. For example, the study presented in [17] performs the classification of built areas according to their density in low resolution (10 meter) SPOT panchromatic remote sensing images by employing algorithms based on occurrence frequency and co-occurrence matrices. However, when image resolution is very high, the complexity of compound structures cannot be handled by traditional texture models.

In this work, we focus on a general property of compound structures that is shared by all the compound structure types: the stong coupling between primitives. It is intuitive that the primitives that comprise compound structures are strongly related to each other. It can be assumed that the degree of this relationship is directly proportional to their transition frequency. For example, in case of forest, there is a high co-occurrence of tree crowns and their shadows. The similar assumption is used by [14] to provide a multiscale segmentation maps. However, their approach is dependent on preliminary clustering of primitives. Opposed to this, we aim to avoid the clustering of primitives or any label assignment, since it is a challenging problem and the errors at this step strongly affect the further analysis. To address this problem, we develop a procedure for transition frequency calculation without a preceeding transition or region type assignment.

In this thesis, we propose a generic unsupervised method for discovering interesting and significant compound objects regardless of their types. The method translates image segmentation into a relational graph, and applies two graph-based knowledge discovery algorithms to find the interesting and repeating substructures that may correspond to compound objects. The first step is image segmentation where the resulting regions correspond to primitive objects that have relatively uniform spectral content. The next step is the translation of this segmentation into a relational graph structure where the nodes represent the regions and the edges represent the relationships between these regions. We assume that the region objects that appear together frequently can be considered as strongly related. This relation is modeled using the transition frequencies between neighboring regions. Each transition is represented by a point in a multi-dimensional space. This space is modeled by a non-parametric probability distribution, and the local maxima found from the density function are assumed to correspond to the most frequently occurring and hence the most significant and important transitions. Finally, a graph whose edges encode this frequent spatial co-occurrence information is constructed, and subgraph analysis algorithms are used to discover substructures that often correspond to groups of region objects that occur together in high-level compound structures. The overview of the proposed framework is given in Figure 1.2.

1.2 Summary of Contributions

In this work, unlike the conventional object-based image analysis approach of finding homogeneous regions, we present an unsupervised method toward discovering compound image structures that are intrinsically heterogeneous. Opposed

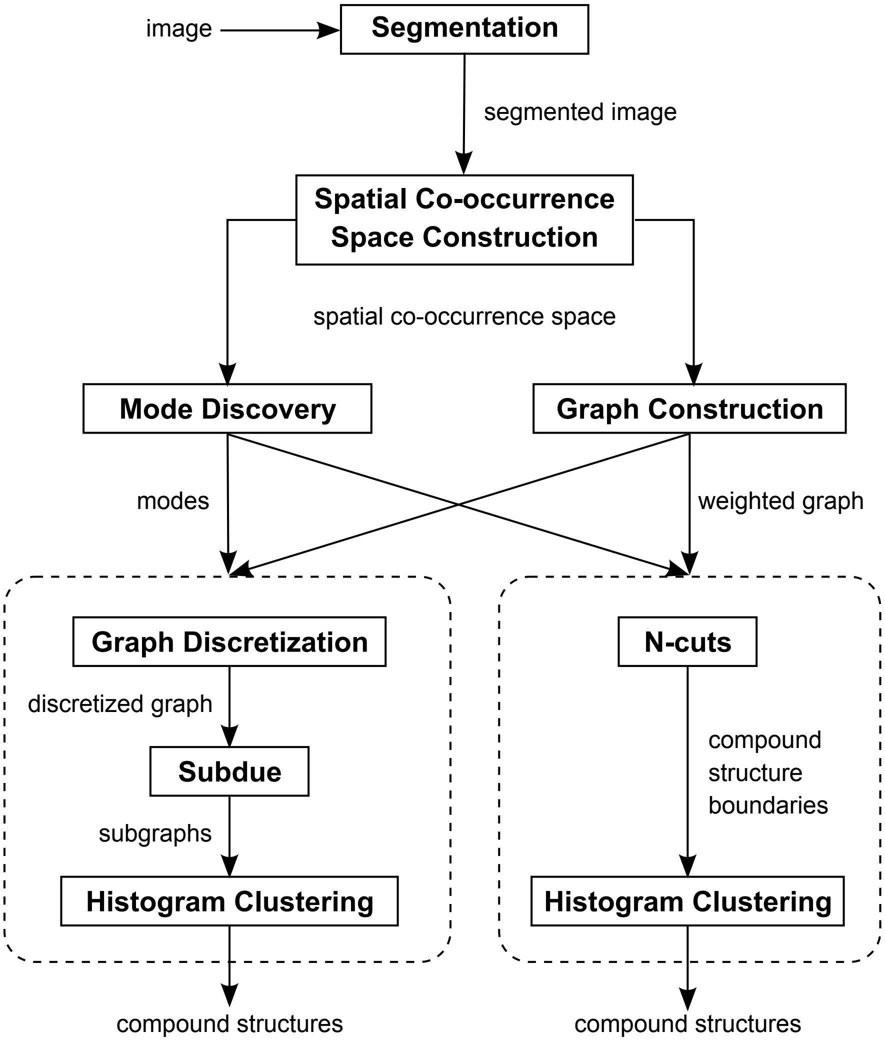


Figure 1.2: Overview of the proposed framework.

to the methods that aim to discover the compound structures of predefined types and rely on particular characteristics of a given compound structure type, we provide a generic method for discovering the compound structures regardless of their types.

Our main contribution is the proposed spatial co-occurrence model that defines a feature space where each point corresponds to an inter-region transition so that features of the regions are encoded in the transition. The transitions that are similar in terms of their features are located close to each other in the spatial co-occurrence space. This enables the encoding of region features together with transition frequency. While similar transitions are pooled together to form dense clusters, seldom transitions are located sparsely. This model provides tolerance to small variations and noise in the region features. Furthermore, it can be easily extended with additional region features. Given this model, we propose that the significance of the particular transition can be found by using non-parametric probability density estimation. Note that our model does not depend on preliminary classification of regions or user-defined number of clusters. Complete description of spatial co-occurrence model is presented in Chapter 4.

One other contribution is the discovery of significant transitions in spatial co-occurrence space using non-parametric clustering and mode seeking. We state that points that corresponds to accumulations in the space can be considered as transitions of the same type. We suggest that local maxima (modes) of the probability density can be considered as centroids for these transition types and can be located by a mode seeking algorithm. This enables us to avoid assumptions about the cluster number and cluster shape and still obtain an implicit clustering of the space by assigning each transition to the closest mode. We also suggest algorithms for stabilizing the modes by mode merging and elimination based on symmetry. More information about mode discovery and the postprocessing steps is provided in Chapter 5.

Another contribution is the construction of a graph with vertices corresponding to primitive regions and the edges encoding the relationship degree between them. By analyzing the edge weights, we cluster the graph to find subgraphs, so that they are composed of vertices with corresponding edges that have high weights modeling frequent spatial co-occurrence. Furthermore, the subgraphs also contain neighborhood information among multiple region objects. Therefore, the subgraph nodes correspond to the region objects that occur together in a high-level compound structure. The details of graph construction and clustering are given in Section 6.1.

Finally, different from common approach of using histograms of primitives, we employ histograms of substucture instances (in Subdue case) and transitions (in normalized cuts case). Classic histograms that count the frequency of occurrence of objects/regions within a window ignore their spatial arrangements. In our case, the spatial arrangement is taken into account because it is encoded in subgraphs/transitions. Also encoding subgraphs/transitions in histograms results in more compact and more effective representations by significantly reducing the dimensionality of the histograms and consequently the computational cost of operations on them. More information on histogram construction and clustering is presented in Section 6.2 and Section 6.3.

1.3 Organization of the Thesis

The rest of the thesis is organized as follows. Chapter 2 summarizes the related work present in literature. In Chapter 3, the details of segmentation and feature extraction are given. Chapter 4 provides the description of the proposed spatial co-occurrence model. It also presents the details of probability density estimation. Next, Chapter 5 discusses the mode discovery in spatial co-occurrence space and postprosessing steps that aim elimination of redundant modes. In Chapter 6, we explain how we construct and cluster the graph to discover subgraphs that correspond to compound structures. We describe the used data set and provide experimental results in Chapter 7. Finally, Chapter 8 summarizes the work and discusses further research directions.

Chapter 2

Literature Review

In comparison to single object detection (such as buildings, roads, etc.), the studies that aim to detect compound objects are not encountered frequently in literature. Most of the state-of-the-art techniques aim the detection of compound structures of predefined types. The most common application is the detection and classification of built-up areas. The identification of precise location of built-up areas and assessment of settlement features is important for territorial planning and human security and safety decision process. Most of the methods proposed for detection or classification of built-up areas rely on particular characteristics of primitives that consitute them, namely buildings. For example, Stasolla and Gamba [27] proposed a procedure for extraction of human settlements from high-resolution synthetic aperture radar (SAR) images. They suggested that built-up areas can be considered as agglomerates of high intensity values since buildings usually produce bright responses in SAR images. They employed spatial indexes and mathematical morphology for detection of settlement's borders. Unsalan and Boyer [32] suggested constructing a graph where photometric straight line segments extracted from grayscale images are assigned to vertices and their spatial relationships are encoded by edges. They introduced a set of measures based on various properties of the graph and used these measures for classification of scenes as rural, residential and urban. This method relies on the fact that impact of human activity causes emergence of straight and smoothly curved contours and their spatial density and regularity increases with increasing development.

Dogrusoz and Aksoy performed classification of building groups as organized and unorganized by using both statistical [2] and structural [11] texture models by first detecting the buildings. In [2], they used buildings as textural primitives and employed co-occurrence-based spatial domain features and Fourier spectrum-based frequency domain features to model repetiveness and periodicity. In their later work [11], they constructed a graph whose nodes correspond to buildings and edges encode neighborhood information obtained through Voronoi tesselation. Then the graph was clustered by thresholding its minimum spanning tree and the resulting clusters were classified as regular or irregular according to the distributions of angles between neighboring nodes.

Apart from detection of built-up areas, several attemps have been made for detection of vineyards and orchards. Generally the proposed methods rely on the spatial arrangement of these structures. For example, the study presented in [34], employed Fourier transform based analysis for vineyard identification and characterization of previuosly delimited plots in 0.25 m spatial resolution images. Warner and Steinmaus [33] employed the spatial classification for identification of orchards and vineyards. Autocorrelation was calculated for the cardinal directions producing four one-dimensional autocorrelograms spaced 45° increments. The classification was performed by analyzing each of the four autocorrelograms for each pixel. One other example is the recent study by Delenne *et al.* [9] that compared two different approaches for vineyard detection. The first approach is based on directional variations of contrast feature calculated from co-occurrence matrices. The second approach is based on a local Fourier transform.

It is important to emphasize the frequent exploitation of texture models for the detection of compound structures [27, 11, 17, 22, 33, 34]. Similarly, the study illustrated in [5] performs the detection of harbors and golf courses by employing textural information. It learns the texture-motif model that corresponds to spatially recurrent patterns of image primitives for each compound object from a set of training examples and uses the learnt model for object detection. Gabor filters at different scales and orientations were used to extract features from the neighborhood of each pixel and Gaussian mixture-based clustering of pixels was employed to identify texture elements. Histograms of texture elements within a sub-window were used for detection of harbors and golf courses.

Multi-resolution analysis can change the amount of details in an image and may enable application of traditional texture models, for example, co-occurrence matrices with fixed displacement vectors and fixed window sizes. This can be useful for detection of compound structures of predefined types for which these displacement vectors can be defined a priori. However, the application of such methods is not straightforward for compound structures of different types because they contain different levels of detail that can emerge in different resolutions. As an example for the detection of specific compound structures in a particular resolution, the method introduced in [17] employs texture measurements to classify built areas according to their density into three categories: high, medium and sparse, in low resolution (10 meter) SPOT panchromatic remote sensing images. The authors developed three algorithms based on occurrence frequency and co-occurrence matrices. According to the output of the algorithms, built areas were classified by using supervised classification. Similarly, the method introduced in [22] performed the detection of built-up areas from satellite images with resolution approaching the size of buildings. It stated the assumption that the textural contrast is high in all directions within the built-up areas. The proposed procedure was based on fuzzy rule-based composition of anisotropic textural co-occurrence measures derived from satellite data by using gray-level co-occurrence matrix constructed for different distances and directions.

There is also a recent study [14] that uses the same assumption that the compound objects consist of strongly related primitives, as in this thesis. It aims to provide multiscale segmentation maps for remote sensing images by modeling transition frequency using Markov chains. Based on the initial segmentation, it finds the initial classes by first clustering primitives using color information and then using spatial information. These classes take on the role of states in the Markov chains. The image is scaned pixelwise along a given direction and the classes encountered along the path are encoded in Markov chain. During class merging procedure, the strongly interacting classes are merged first. Since this approach is strongly dependent on the length of boundaries between the regions, in their later work [23] the authors ehnance their model by considering the spatial distribution similarity of interacting regions besides the degree of their contact. Note that this method is dependent on preliminary clustering of primitives and the errors at this step strongly affect further analysis.

Chapter 3

Segmentation and Feature Extraction

First step in the proposed methodology is to perform segmentation to partition the image into regions and represent each region by its spectral and scale features. Details of image segmentation and feature extraction are discussed below.

3.1 Image Segmentation

Image segmentation is the first step in our study and it aims to partition the image into regions that have relatively uniform spectral content. The choice of the segmentation algorithm is important because the ensuing region-based analysis rely on the quality of the segmentation output. We selected the Recursive Hierarchical Segmentation (RHSEG) algorithm [29], because of high spatial fidelity of resulting segmentations and automatic production of hierarchical set of segmentations.

RHSEG is a computationally efficient recursive approximation of previously developed HSEG hierarchical image segmentation algorithm [28]. HSEG is a combination of spectral clustering and Hierarchical Step-Wise Optimization (HSWO).

HSWO is a form of region growing segmentation where each iteration aims to find best segmentation containing one region less than current segmentation [3, 31]. In contrast to HSWO, HSEG alternates region-growing iterations with spectral-clustering iterations. The logic behind this is that spatially adjacent regions merge during region growing iterations while non-spatially adjacent regions are merged by spectral clustering iterations. The addition of spectral clustering allows the produced segmentations to capture the spatial detail of images with greater fidelity and describe images in terms of region classes. Here, region classes are groups of spatially disjoint region objects and region objects are areas of spatially connected image pixels that correspond to image primitives.

Different priorities can be given to region growing (merges of spatially adjacent regions) and spectral clustering (merges of spatially non-adjacent regions). It can be controlled through the input parameter S_{wght}. This parameter varies from 0.0 to 1.0 and has the following effect according to its value:

- $S_{wght} = 0.0$, spatially non-adjacent region merges are not allowed,
- $0.0 < S_{wght} < 1.0$, spatially adjacent merges are given priority over spatially non-adjacent merges by a factor of $1.0/S_{wght}$,
- $S_{wght} = 1.0$, merges between spatially adjacent and spatially non-adjacent regions are given equal priority.

The advantage of combining region growing with spectral clustering can be demonstrated by comparing an image segmentation result from RHSEG with a result produced by HSWO. Figure 3.1 shows a 256 × 256 portion of an Ikonos image in true color, the region mean image from the RHSEG result using S_{wght} = 0.25, and the region mean image from the HSWO results. RHSEG and HSWO were both run until the region merging threshold of 10.0 was reached [30].

The output of RHSEG consists of the region class labels map at the finest level of segmentation detail (hierarchical level 0) and the region classes file that contains selected information about each region class at each hierarchical level. This file includes the region merges list feature that consists of the re-numberings

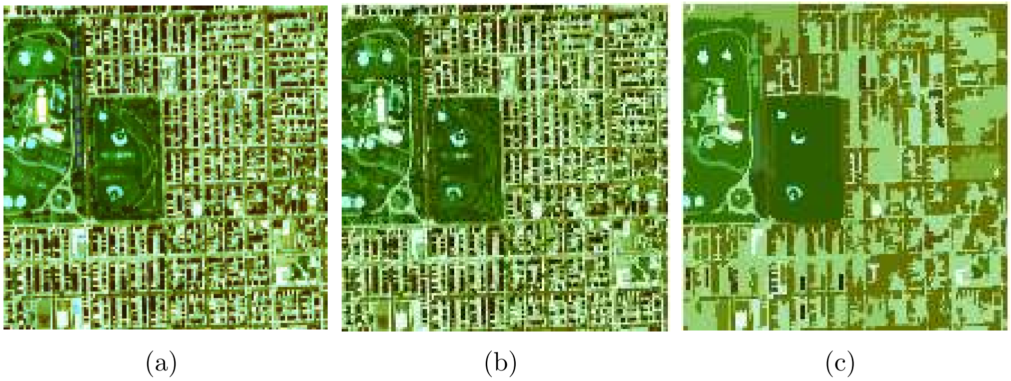


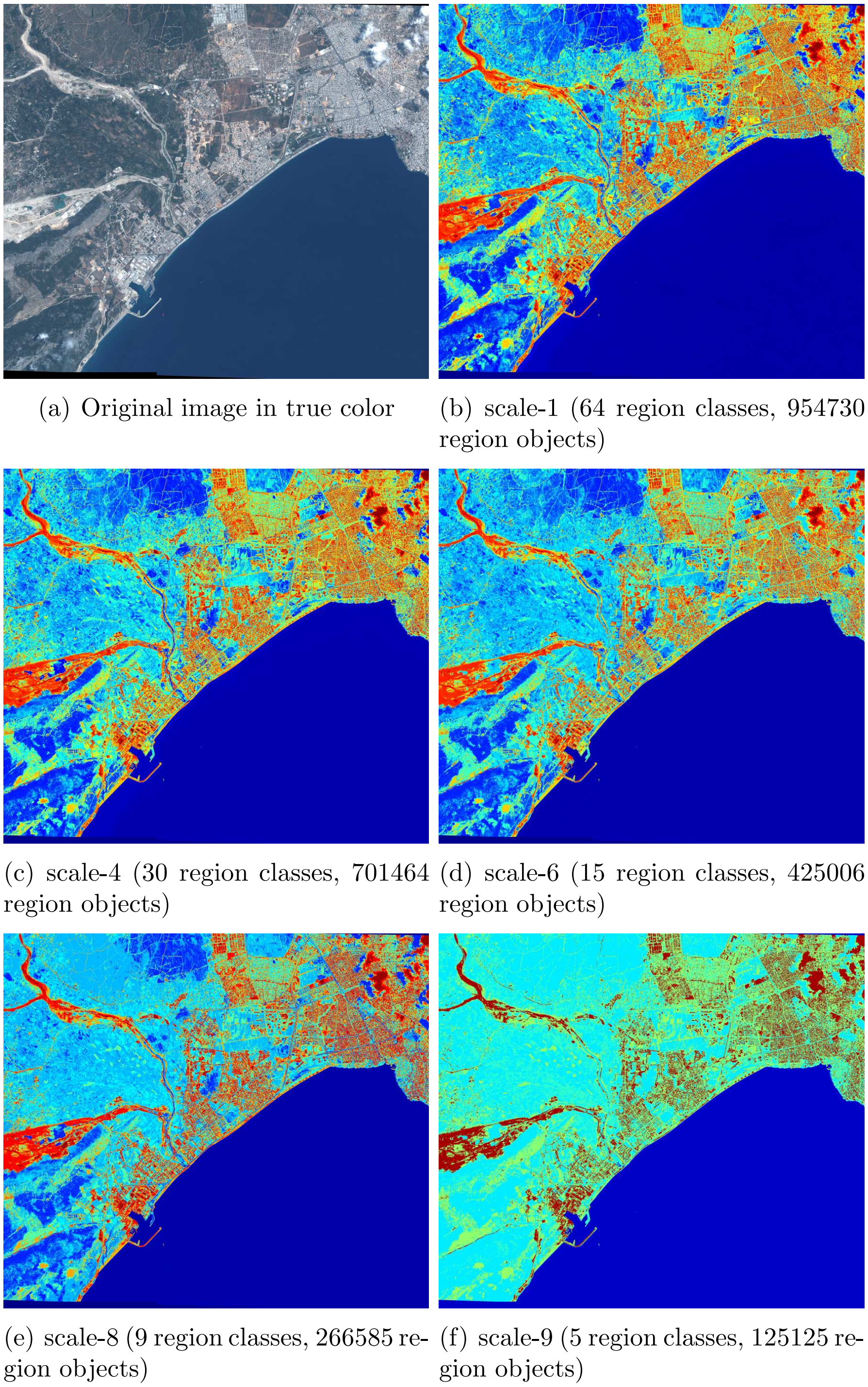
Figure 3.1: (a) A 256x256 portion of an Ikonos image in true color. (b) The region mean image from an RHSEG segmentation with $S_{wght} = 0.25$. (c) The region mean image from an HSWO segmentation.

of the region class labels map required to obtain the region class labels map for the second most detailed level (hierarchical level 1) through the coarsest (last) level of the segmentation hierarchy from the class label map. By examining this file, the segmentation at a desired hierarchy level can be obtained. Even though the whole hierarchy can be useful for object detection [1], it is possible to examine how the regions change at each level and choose the level of detail at which the particular regions are delineated. Figure 3.2 presents an example of segmentation detail varying with the levels in the hierarchy.

3.2 Feature Extraction

After the segmentation is performed, the image can be considered as a collection of regions. We want to represent each region in terms of a set of features that represent its content. These features must be able to adequately describe the region and capture the similarity between regions of the same type and dissimilarity between regions of different types. We choose to employ spectral and region size information for representing the regions.

In this case, spectral features are the red (r), green (g) and blue (b) channels of the image. Since a region generally comprises of a number of pixels, in order to



(a) Original image in true color

(b) scale-1 (64 region classes, 954730 region objects)

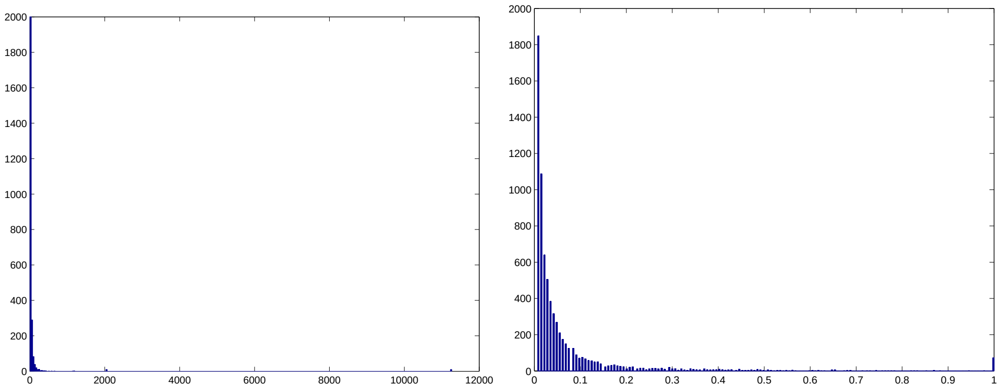
(c) scale-4 (30 region classes, 701464 region objects)

(d) scale-6 (15 region classes, 425006 region objects)

(e) scale-8 (9 region classes, 266585 region objects)

(f) scale-9 (5 region classes, 125125 region objects)

Figure 3.2: Visual bands of an Ikonos image of Antalya with 4 m spatial resolution, and the corresponding RHSEG results at different levels in the hierarchy. Default parameter values of RHSEG are used as explained in [21].



(a) Sample size value distribution

(b) Sample size value distribution after normalization

Figure 3.3: Region size normalization with elimination of extreme values by clipping at 1 percent.

extract spectral information we take the arithmetical average of all pixel values belonging to the given region for each channel.

However, color information alone is not always enough to discriminate between different region types. Hence, it is reasonable to use the region size information along with spectral features. Size of the region corresponds to the number of pixels associated with it.

Since spectral and size feature values have different ranges, feature normalization must be performed in order to equalize their ranges. Feature normalization is required to make feature components have similar effect during region comparison. To achieve this, each feature component is normalized to the [0,1] range by using linear scaling to unit range as

$$\tilde{x} = \frac{x - l}{u - l}, \tag{3.1}$$

where l and u are the lower and upper bound for a feature component x and $\tilde{x}$ is the normalized value.

In case of spectral features, the lower and upper bounds are well defined since values for image channels have fixed ranges. However, there are no constraints for the size features. For example, there are some extremely high values in sample size value distribution illustrated in Figure 3.3(a). Obviously, the largest region size is not a good candidate for the upper bound, since the presence of a few regions that have very large sizes relative to other regions can drastically affect the normalization. In order to make the normalization more adequate, we eliminate extreme values by clipping the tail of the distribution. To define the clipping location, a certain percantage is set for the number of values to be excluded. The result of normalization after extreme value elimination is shown in Figure 3.3(b).

After the spectral and size features are extracted and normalized, each region R_i can be expressed by its feature vector $\mathbf{y}_{\boldsymbol{i}} = (r_i, g_i, b_i, s_i)$.

Chapter 4

Spatial Co-occurrence Model

The detection of compound structures can be posed as a generalized texture problem. Hence, one way for detection of compound structures is to employ traditional texture models [4, 10, 17]. Generally, texture models concern features that are related to periodicity, directionality or randomness. They include the co-occurrence matrix [15], Fourier transform, and the autocorrelation function [19]. For example, co-occurrence matrices computed at different inter-pixel distances and at particular orientation can be used to detect coarseness, directionality, and periodicity at a given orientation [36, 26, 10]. However, this model requires the selection of specific scale and orientation which are not stable for compound structures. Nonetheless, we can assume that compound structures consist of image primitives that are strongly related to each other.

In this work, we model the region relationships using the transition frequencies between neighboring regions in the image by assuming that the region objects that appear together frequently in the image can be considered as strongly related. One way to calculate the inter-region transition frequency is by determining the types of the regions in a transition and by counting the transitions involving the same types of region pairs. For example, RHSEG assigns a class label to each region and these labels can be used for further analysis as in [30], but they are only based on spectral properties of pixels which are generally noisy in high resolution. The determination of region type is a challenging classification problem, and errors at this step will result in misleading transition types. To avoid classification and its drawbacks, we propose a spatial co-occurrence model that enables transition frequency calculation without preceding transition or region type assignment. This model uses the multi-dimensional space where each point corresponds to an inter-region transition and enables the incorporation of region transition frequencies together with region features. This space is modeled by a non-parametric probability density distribution so that the probability value for each transition point corresponds to the frequency of its occurrence in the image. The details of spatial co-occurrence space construction and probability estimation are discussed below.

4.1 Spatial Co-occurrence Space Construction

Spatial co-occurrence space construction requires the definition for representation of inter-region transition. We assume that a transition can be fully described by a region pair between which it occurs. Each transition is defined by the features of the corresponding regions so that their contents can be incorporated in the model. In an image with N_R regions $R_i, i = 1, \dots, N_R$, the transition T_{ij} involving the regions R_i and R_j is represented by the concatenation of feature vectors of the two regions as $\mathbf{y}_{\boldsymbol{ij}} = (\mathbf{y}_{\boldsymbol{i}}, \mathbf{y}_{\boldsymbol{j}})$. Given the region feature vectors with 4 components, the feature vector for a transition corresponds to a point in the 8-dimensional spatial co-occurrence space. For simplicity, we refer to these points as $\mathbf{x}_{\boldsymbol{k}} \in \mathbb{R}^d, k = 1, \dots, N_T$ where $d = 8$ and N_T is the number of transitions.

To construct the spatial co-occurrence space, the transitions between each pair of neighboring regions are found and the corresponding feature vectors are extracted. Then, each transition is mapped to a point in the multi-dimensional space. Algorithm 1 describes the details of this procedure.

We assume that the transitions that involve two similar region pairs fall close to each other in the spatial co-occurrence space because regions with similar spectral content and sizes are expected to be similar in terms of their features.

Algorithm 1 Constructing the spatial co-occurrence space

```
$Regions = \{R_1, R_2, \ldots, R_{N_R}\}$
$Adjacent = \{\}$
$Transitions = \{\}$
for each $R$ in $Regions$ do
  $Adjacent = findAdjacentNeighbors(R)$
  for each $R_a$ in $Adjacent$ do
    $T = [R, R_a]$
    $\mathbf{x} = [\mathbf{y}, \mathbf{y_a}]$
    Add $T$ to $Transitions$
    Add point $\mathbf{y}$ to spatial co-occurrence space
  end for
end for
```

Consequently, the transitions that occur frequently cause the accumulation of points in the space. The significance of a given transition can be determined according to its position relative to these dense regions (see Section 4.2 for details).

While similar transitions are pooled together to form dense clusters, seldom transitions are located sparsely. This model provides tolerance to small variations and noise in the region features. Furthermore, it can be easily extended with additional region features.

To be able to provide visual example of spatial co-occurrence space, we use a simulated segmentation result shown in Figure 4.1, which was used by [30]. This simulated segmentation combines idealized segmentations of a residential area (most of the lower left quadrant), an apartment complex (most of the upper left quadrant), an industrial park (the upper right quadrant) and recreational parks (inserted in the apartment complex and residential quadrants) with a section of an actual segmentation of SAR data (lower right quadrant). This segmentation comprises 1439 regions and 3222 inter-region transitions, so the constructed spatial co-occurrence space contains 3222 points.

To get the general idea about the spatial co-occurrence space, we apply the Principal Component Analysis (PCA) [12] to reduce the space dimensionality. Then the space is visualized (see Figure 4.2(a)) by using the first two principal components. Although the illustrated space is an approximation to an actual

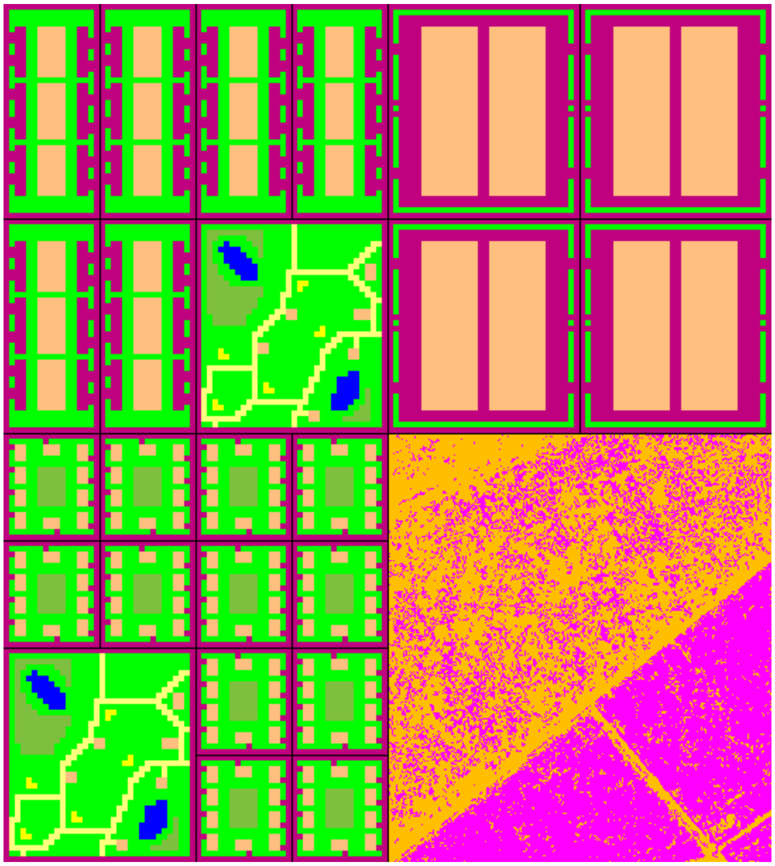
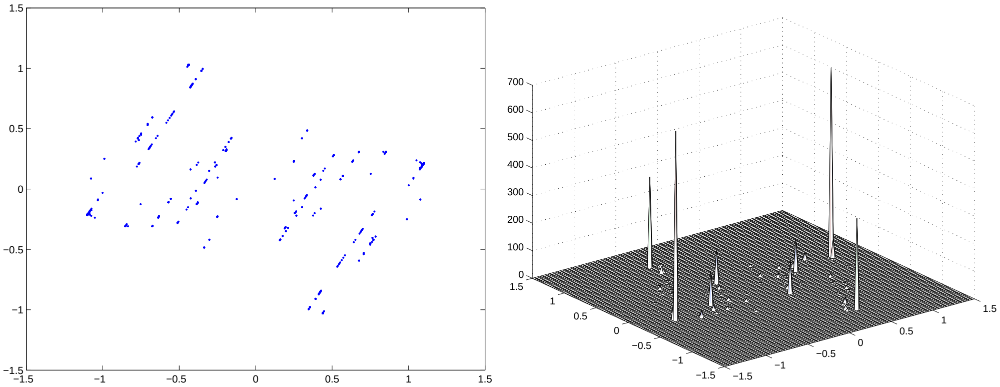


Figure 4.1: Simulated image segmentation (see text).

space, the accumulations of points can be observed. Used segmentation contains regions of exactly same color and size, therefore multiple transitions map to exactly same point in space. These types of accumulations can be better seen in Figure 4.2(b), where the 2-dimensional histogram of space points is illustrated. Note that the space is symmetrical due to the duality of transitions (transition from R_i to R_j implies transition from R_j to R_i).

4.2 Transition Probability Estimation

Once the spatial co-occurrence space is constructed, we aim to investigate the significance of each transition. Recall our assumption that region objects that appear together frequently in the image can be considered as strongly related,



(a) (b)

Figure 4.2: Visualization of spatial co-occurrence space constructed by using simulated image segmentation. (a) Plot of the space by projecting the data onto the first two principal components obtained by applying PCA. (b) 2-dimensional histogram of points in (a).

so the most recurrent transitions are the most important ones. Also recall that similar transitions that occur frequently cause the accumulation of points in the space. The significance of a particular transition can be determined according to its location relative to the dense areas in the spatial co-occurrence space. Namely, we can assign a particular weight to each transition by measuring the likelihood of the corresponding point in the space. We model the spatial co-occurrence space by a Parzen window-based non-parametric probability distribution, and the local maxima (modes) found from the probability density function correspond to the accumulations of points in the space. Given N_T data points $\mathbf{x}_{\boldsymbol{k}}$, $k = 1, \ldots, N_T$ in d-dimensional space, the density estimate at point $\mathbf{x}$ can be written as

$$p(\mathbf{x}) = \frac{1}{N_T} \sum_{k=1}^{N_T} K_H(\mathbf{x} - \mathbf{x}_{\boldsymbol{k}}), \tag{4.1}$$

where K($\mathbf{x}$) is a kernel window function and $\mathbf{H}$ is a symmetric positive definite $d \times d$ matrix representing the smoothing parameter (also called the bandwidth matrix). Assuming a Gaussian kernel with a smoothing parameter $\mathbf{H} = \sigma^2 \mathbf{I}$, the expression (4.1) yields

$$p(\mathbf{x}) = \frac{1}{N_T} \sum_{k=1}^{N_T} \frac{1}{(2\pi)^{d/2} |\mathbf{H}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{x}_{\boldsymbol{k}})^T \mathbf{H}^{-1} (\mathbf{x}-\mathbf{x}_{\boldsymbol{k}})}. \tag{4.2}$$

The complexity of this procedure can be decreased by using spatial data structures.

The points that fall within dense regions in the space would have more neighbours to contribute that results in higher probability values. Points that have high probabilities and constitute these dense regions stand for the most frequently occurring and hence the most important transitions.

The choice of the bandwidth matrix is critical because it strongly affects the smoothness of the resulting density. We want to optimize $\mathbf{H}$ so that it is a function of both N_T and the data itself. Different bandwidth selection algorithms were proposed; however, the ones that have practical use generally aim to estimate the smoothing parameter for univariate distributions. Therefore, we express the bandwith matrix as $\mathbf{H} = \sigma^2\mathbf{I}$, and reduce our problem to the estimation of σ.

To compute σ, we used a method based on leave-one out maximum likelihood estimation [13]. In this method, σ is computed as the value that optimized the product of the estimated densities at the sample points:

$$\arg\max_{\sigma} L(\sigma) = \prod_{j=1}^{N_T} \hat{F}_j(x) \tag{4.3}$$

in which

$$\hat{F}_j(x) = \sum_{i \neq j}^{N_T} \frac{1}{(\sigma\sqrt{2\pi})^m} \exp\left\{ - \frac{||x - x_i||^2}{2\sigma^2} \right\}. \tag{4.4}$$

Note that the contribution of the sample itself during the estimation of the density is omited. The optimization of (4.3) is always executed by finding the zero crossing(s) of its first derivative.

Chapter 5

Mode Discovery in Spatial Co-occurrence Space

At this step we want to delineate the clusters formed by the accumulations of the transition points. This will group the transitions and assign each transition a particular type. However, we do not want to obtain the exact clustering of the whole space. Instead, we aim to locate the dense regions and find the points that constitute these regions. We assume that the dense regions in this space correspond to the most frequently occurring and hence the most significant and important transitions. One way to discover these dense regions is to use a clustering algorithm such as EM-based mixture of Gaussians estimation, however, using this kind of clustering requires the assumption about cluster number and cluster shape that are not known a priori in our case. On the other hand, dense regions can be found by locating the modes (local maxima) of the estimated density. One possible method for locating these modes is the mean-shift algorithm [7]. This approach is non-parametric and it is very suitable for our method because it is also based on Parzen density estimation in a multi-dimensional space (similar to the spatial co-occurrence space proposed in Chapter 4).

We apply the mean-shift procedure to discover the modes in the previously constructed spatial co-occurrence space. Generally, the number of modes found

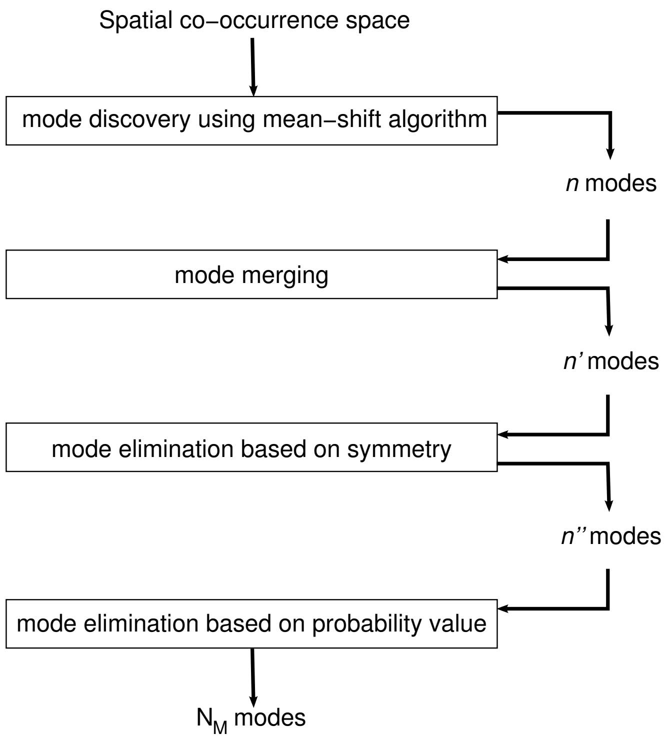

Spatial co−occurrence space
mode discovery using mean−shift algorithm
n modes
mode merging
n' modes
mode elimination based on symmetry
n'' modes
mode elimination based on probability value
N_M modes

Figure 5.1: Overview of mode discovery and postprocessing steps.

exceeds the actual number of modes due to the drawbacks of the mean-shift algorithm and the nature of spatial co-occurrence space. To overcome this problem, some of the modes are eliminated based on multiple criteria. Mode discovery and postprocessing steps are summarized in Figure 5.1 and explained in details in succeeding subsections.

5.1 Mode Discovery

Given N_T data points $\mathbf{x}_{\boldsymbol{k}} \in \mathbb{R}^d, k = 1, \ldots, N_T$, we want to find the location of the local maxima in the probability distribution fitted to the space. Starting from a randomly selected set of points, the algorithm computes the mean-shift vector at each point $\mathbf{x}$ as

$$\mathbf{m}(\mathbf{x}) = \frac{\sum_{k=1}^{N_T} \mathbf{x}_{\boldsymbol{k}} e^{-\frac{1}{2}D^2(\mathbf{x},\mathbf{x}_{\boldsymbol{k}},\mathbf{H})}}{\sum_{k=1}^{N_T} e^{-\frac{1}{2}D^2(\mathbf{x},\mathbf{x}_{\boldsymbol{k}},\mathbf{H})}} - \mathbf{x} \tag{5.1}$$

using the Parzen density gradient estimate at that point, and moves along this vector by iterating until the difference between two successive means is less than a threshold or the number of iterations reaches a maximum value. The points at which the algorithm converges are considered as the candidate modes. In (5.1),

$$D^2(\mathbf{x}, \mathbf{x}_{\boldsymbol{k}}, \mathbf{H}) = (\mathbf{x} - \mathbf{x}_{\boldsymbol{k}})^T \mathbf{H}^{-1} (\mathbf{x} - \mathbf{x}_{\boldsymbol{k}}) \tag{5.2}$$

is the Mahalanobis distance from $\mathbf{x}$ to $\mathbf{x}_{\boldsymbol{k}}$ and $\mathbf{H}$ is the symmetric positive definite dxd bandwidth matrix discussed in Section 4.2.

Ideally, the algorithm must be started from every point in the space to capture all modes. This can also provide implicit assignment to clusters if each point is assigned to a cluster corresponding to a mode it converged. However, running the algorithm for each point is computationally very expensive. For this reason it is more feasible to choose a sufficient number of points starting randomly so that the whole space is covered.

After the mean-shift algorithm is applied for sufficient number of observations, the points $\mathbf{m_1}, \mathbf{m_2}, \ldots, \mathbf{m}_{\boldsymbol{n}}$ of convergence correspond to the candidate modes. Running the mean-shift algorithm for the example presented in Figure 4.1 starting from 2000 different points results in $n = 375$ modes. The modes are shown in red in Figure 5.2. As expected, the modes are generally located at the peaks of the density. However, note that generally the number of candidate modes exceeds the actual number of modes due to the drawbacks of the algorithm and the nature of spatial co-occurrence space. Hence, postprocessing is required to eliminate some of the candidates.

5.2 Mode Merging and Elimination

The convergence of the mean-shift algorithm is affected by the termination threshold and the number of maximum iterations allowed. Due to local details in the spatial co-occurrence space, starting at points that actually belong to the same mode may result in convergence at slightly different locations. One possible solution is to decrease the terminating threshold and increase the maximum number

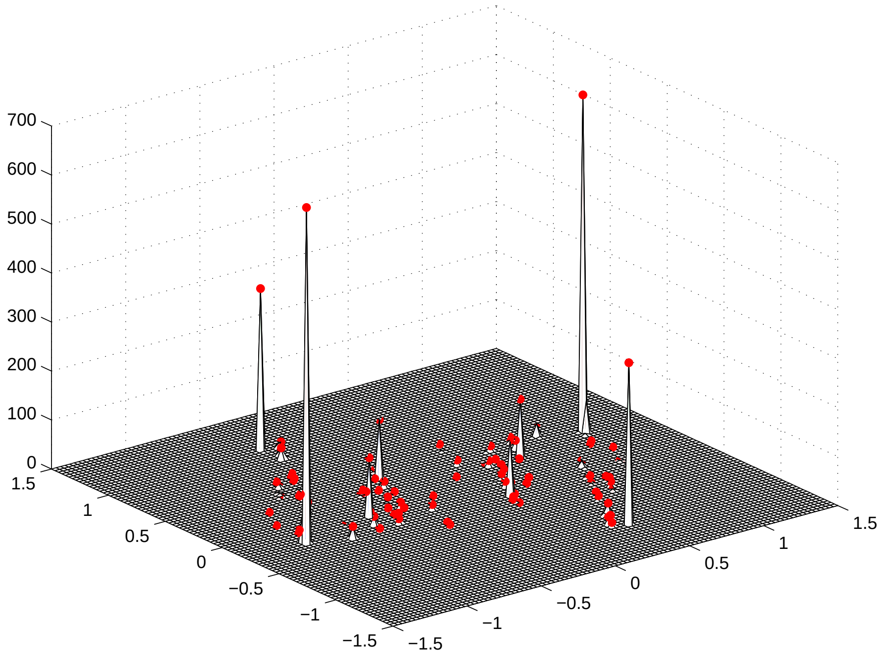


Figure 5.2: Original candidate modes discovered from the spatial co-occurrence space by using the mean-shift algorithm. The space is constructed by using the simulated image segmentation shown in Figure 4.1.

of iterations. However, while still this does not guarantee the convergence at exactly same point, it increases the computation time significantly. To eliminate such noisy convergence, we merge the candidate modes at a distance less than the bandwidth. We assume that these points correspond to the same mode. To merge the modes, hierarchical clustering is applied. We calculate the dissimilarity between the points by using (5.2), therefore the Mahalanobis distance between $\mathbf{m}_{\boldsymbol{i}}$ and $\mathbf{m}_{\boldsymbol{j}}$ that are closer than the bandwidth must not exceed 1. This can be derived by using (5.2). Let $\mathbf{m}_{\boldsymbol{i}}$ and $\mathbf{m}_{\boldsymbol{j}}$ be two candidate modes in the spatial co-occurrence space, so that $\mathbf{m}_{\boldsymbol{i}} = (m_{i1}, \ldots, m_{id})^T$ and $\mathbf{m}_{\boldsymbol{j}} = (m_{j1}, \ldots, m_{jd})^T$. The Mahalanobis distance between these two points can be expressed as

$$\begin{aligned} D^2(\mathbf{m}_{\boldsymbol{i}}, \mathbf{m}_{\boldsymbol{j}}, \mathbf{H}) &= (\mathbf{m}_{\boldsymbol{i}} - \mathbf{m}_{\boldsymbol{j}})^T \mathbf{H}^{-1} (\mathbf{m}_{\boldsymbol{i}} - \mathbf{m}_{\boldsymbol{j}}) \\ &= \frac{(m_{i1} - m_{j1})^2 + \ldots + (m_{id} - mjd)^2}{\sigma^2}. \end{aligned} \tag{5.3}$$

Since the employed bandwidth matrix is in the form $\mathbf{H} = \sigma^2 \mathbf{I}$, it can be said that all points within the hypersphere with radius σ centered at point $\mathbf{m}_{\boldsymbol{j}}$ can be merged with point $\mathbf{m}_{\boldsymbol{j}}$. The following inequality is true for every point $\mathbf{m}_{\boldsymbol{i}}$ within the hypersphere

$$(m_{i1} - m_{j1})^2 + \ldots + (m_{id} - m_{jd})^2 \leq \sigma^2, \tag{5.4}$$

which can be rewritten as

$$\frac{(m_{i1} - m_{j1})^2 + \ldots + (m_{id} - m_{jd})^2}{\sigma^2} \leq 1. \tag{5.5}$$

Combining the above equations yields

$$\frac{(m_{i1} - m_{j1})^2 + \ldots + (m_{id} - m_{jd})^2}{\sigma^2} = D^2(\mathbf{m}_{\boldsymbol{i}}, \mathbf{m}_{\boldsymbol{j}}, \mathbf{H}) \leq 1. \tag{5.6}$$

It can be observed that when the Mahalanobis distance between two points is less than or equal to 1, these points lie within the same bandwidth.

We use hierarchical clustering to find groups of points that are closer to each other than the bandwidth. When the hierarchical clustering tree is cut at the level corresponding to a Mahalanobis distance of 1, the points within the kernel bandwidth fall into the same cluster. To control cluster formation involving more than two points, we employ the complete linkage algorithm. This ensures that all points in a cluster lie within the bandwidth. Namely, for any cluster C, the following inequality holds:

$$\max\{D^2(\mathbf{m}_{\boldsymbol{i}}, \mathbf{m}_{\boldsymbol{j}}, \mathbf{H}) | \forall \mathbf{m}_{\boldsymbol{i}}, \mathbf{m}_{\boldsymbol{j}} \in C\} \leq 1. \tag{5.7}$$

After the clusters are obtained, one mode per cluster is selected by choosing the point that corresponds to the highest density calculated from (4.2). This results in n' modes ($n' < n$). Algorithm 2 describes the mode merging procedure.

The resulting set of modes provide an implicit clustering of the spatial co-occurrence space as any point in this space can be assigned to its closest mode. However, some clusters are redundant and some correspond to very sparse regions rather than accumulation of points. These clusters can be eliminated because we seek for the clusters that correspond to the most significant transitions. Note that we want to perform the elimination on cluster level rather than on mode level because applying clustering after eliminating the modes can result in improper

Algorithm 2 Mode merging

```
$CandidateModes = \{\mathbf{m_1}, \mathbf{m_2}, \ldots, \mathbf{m_n}\}$
$Distance = \{\}$
$ChosenModes = \{\}$
$Probs = \{\}$
for each $\mathbf{m_i}$ in $CandidateModes$ do
  for each $\mathbf{m_j}$ in $CandidateModes$ do
    Calculate $Distance[i][j]$ using (5.2)
  end for
end for
$hct$ = HierarchicalClustering($Distance$).
Cut $hct$ at level where distance is equal to 1 to obtain a clustering set $Clusters$ $= \{C_1, C_2, \ldots, C_{n'}\}$
for each $\mathbf{m_i}$ in $CandidateModes$ do
  Calculate $Probs(i)$ using (4.2)
end for
for each $C$ in $Clusters$ do
  Choose $\mathbf{m}$, $\mathbf{m} \in C$ with highest value in $Probs$
  Add $\mathbf{m}$ to $ChosenModes$ as representative for $C$
end for
```

Algorithm 3 Mode elimination based on symmetry

```
$CandidateModes = \{\mathbf{m_1}, \mathbf{m_2}, \ldots, \mathbf{m_{n'}}\}, k = 1, \ldots, n'$
Define $Labels$ as an array of zeros of size $n'$.
$ChosenModes = \{\}$
$l = 1;$
for each $m_i$ in $CandidateModes$ do
  if $Labels[i] == 0$ then
    $Labels[i] = l$
    $l = l + 1$
  end if
  for each $m_j$ in $CandidateModes$ do
    Calculate $dist1$ as a distance between $\mathbf{m_{i(1:d/2)}}$ and $\mathbf{m_{j(d/2+1:d)}}$ using (5.2)
    Calculate $dist2$ as a distance between $\mathbf{m_{i(d/2+1:d)}}$ and $\mathbf{m_{j(1:d/2)}}$ using (5.2)
    if $dist1 \leq 1$ and $dist2 \leq 1$ then
      $Labels[j] = Labels[i]$
    end if
  end for
end for
$ChosenInd$ = unique($Labels$)
$ChoosenModes = \{\mathbf{m_i}, \mathbf{m_i} \in CandidateModes, i \in ChosenInd\}$
```

cluster formation. Namely, transitions of different type can be assigned to the same cluster and noisy transitions can affect cluster integrity.

Redundant clusters are also present due to to the symmetric nature of the co-occurrence space. The symmetrical clusters correspond to the same transitions in terms of involved regions. The cluster symmetry information can be captured by examining the modes. Since T_{ij} is equivalent to transition T_{ji} and any mode $\mathbf{m}_{\boldsymbol{k}}$ can be represented as

$$\mathbf{m}_{\boldsymbol{k}} = (\mathbf{m}_{\boldsymbol{k(1:d/2)}}\mathbf{m}_{\boldsymbol{k(d/2+1:d)}}), \tag{5.8}$$

we compare the corresponding parts of the feature vectors of the candidate modes, and eliminate one of the modes corresponding to symmetric transitions. During comparison we follow the logic that is similar to that applied while mode merging. The corresponding parts of feature vectors are assumed to represent the same regions if the Mahalanobis distance between them is not greater than 1. This reduces the number of modes to n'', $n'' < n'$. Algorithm 3 describes the elimination procedure. Figure 5.3 presents the modes after elimination of the redundant symmetric modes.

Finally, the elimination of clusters that correspond to single points or sparse regions is important because these clusters generally correspond to noise. Similarly, these clusters can be discovered by examining the modes. The probability value of each mode is calculated by using the Parzen window-based estimator described by (4.1). Modes that have probability less than a predefined threshold and the corresponding clusters are eliminated. The modes $\mathbf{m_1}, \mathbf{m_2}, \ldots, \mathbf{m}_{\boldsymbol{N_M}}$ that are left at this step will be employed in further analysis (Figure 5.4). The resulting set of modes provide an implicit clustering of the spatial co-occurrence space as any point in this space can be assigned to its closest mode.

Selected N_M modes can be examined in terms of transitions assigned to them. Figure 5.5 presents 20 modes with the highest probability values discovered from the spatial co-occurrence space constructed by using the simulated segmentation result shown in Figure 4.1. It can be observed that mostly the discovered transitions are the most frequent and most important transitions that characterize particular compound structures. Notice that some transitions that involve

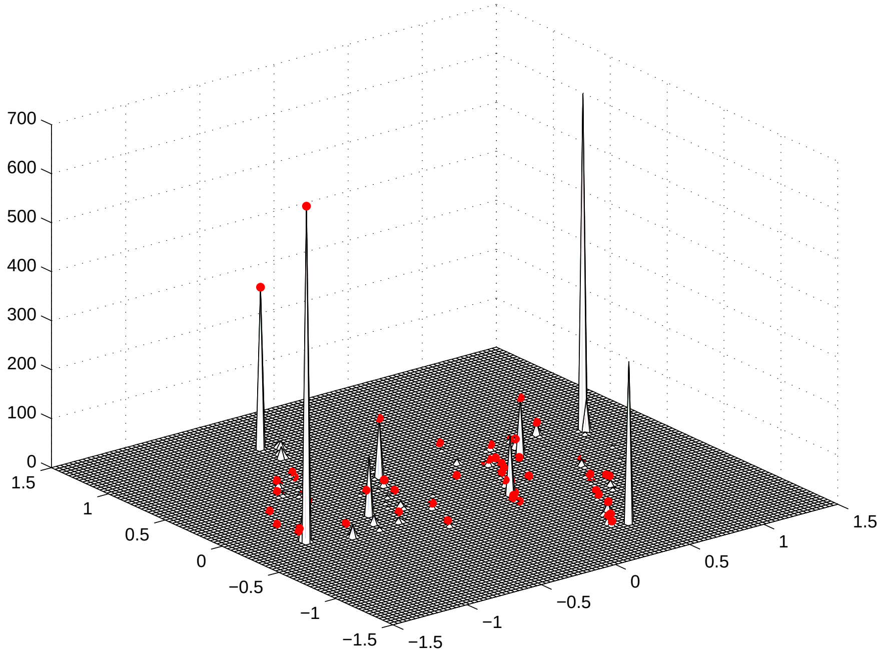


Figure 5.3: Candidate modes after the elimination based on mode symmetry. One of the modes from each pair of symmetric modes is eliminated.

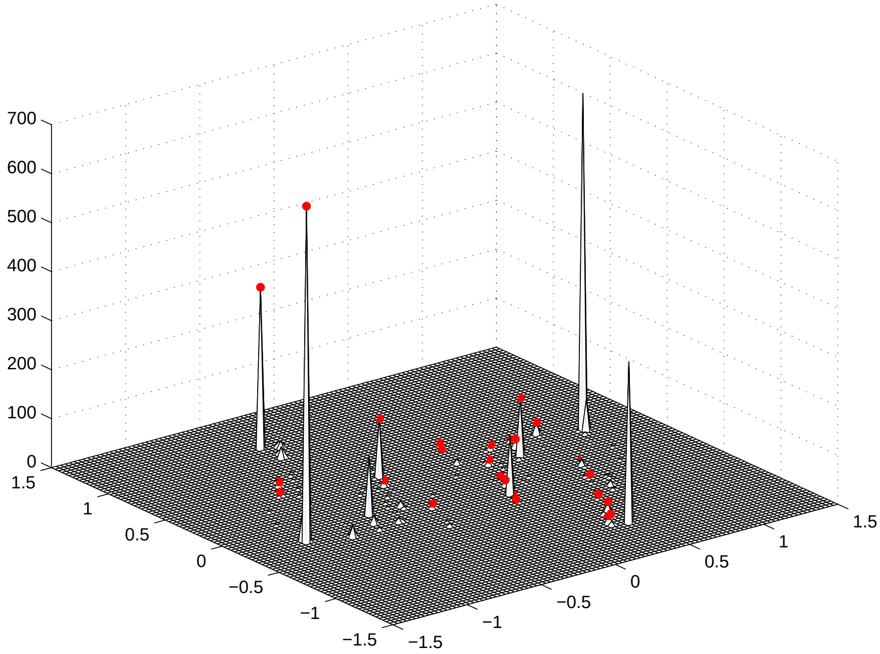


Figure 5.4: Finalized modes after all the postprocessing steps.

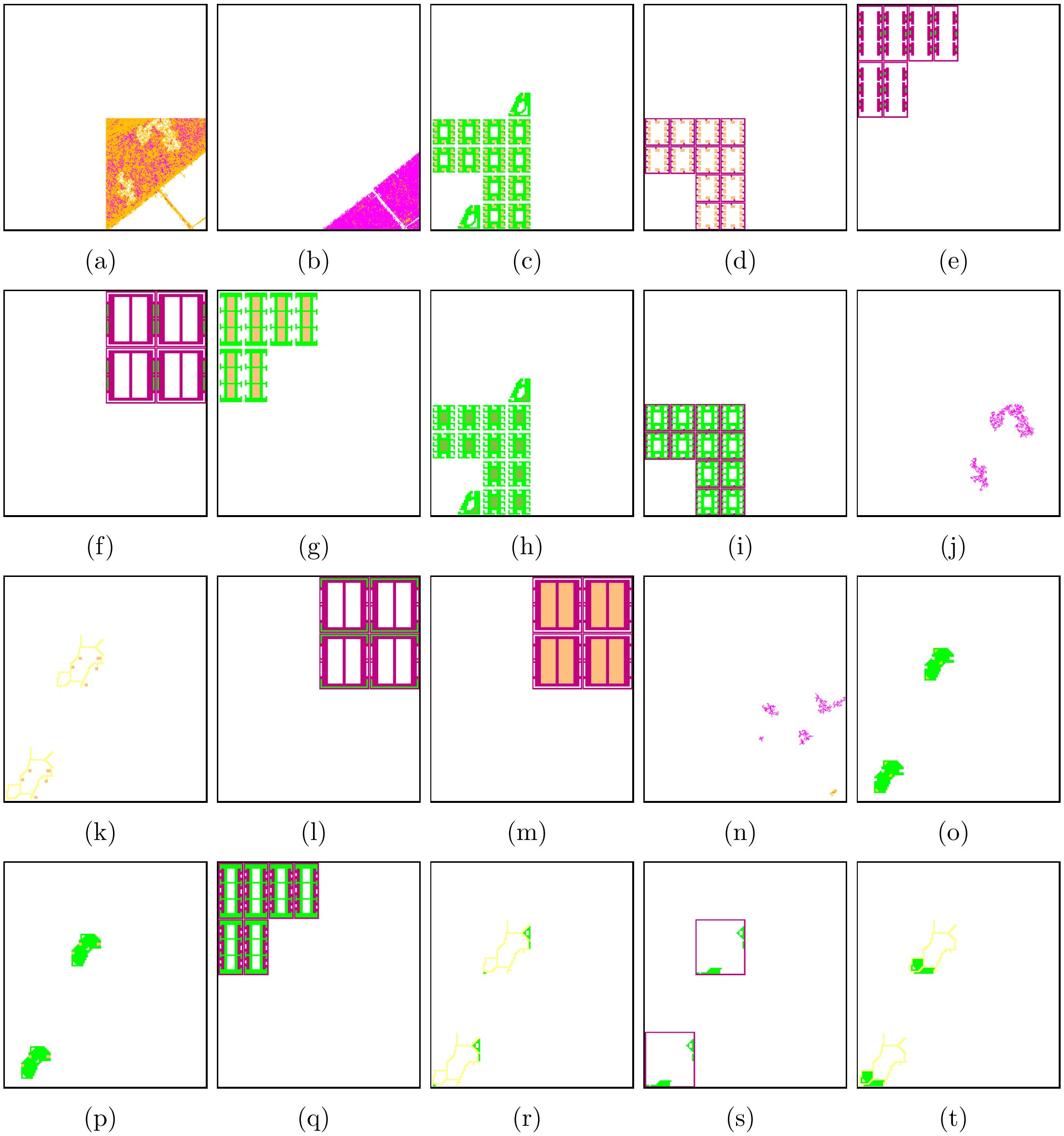


Figure 5.5: Most significant transitions discovered from the spatial co-occurrence space constructed by using the simulated segmentation result shown in Figure 4.1. Regions involved in transitions that were assigned to a given mode are shown in color.

regions with similar spectral content (for example, transitions on Figure 5.5(c) and Figure 5.5(g), Figure 5.5(d) and Figure 5.5(m)) are discriminated because of the addition of size features. There are also some transitions that correspond to noise, for example transitions on Figure 5.5(j) and Figure 5.5(n). They are selected as significant because they outnumber some of the important transitions. In real images, however, the frequency of noise transitions is very low relative to important transitions.

Chapter 6

Detection of Compound Structures

After the spatial co-occurrence space is constructed and the required information is extracted from it, we want to translate image segmentation into a relational graph by using this information. Details of graph construction and clustering are described below.

6.1 Graph Construction

At this step, we aim translation of segmentation into a relational graph structure. In the constructed graph, nodes represent the image regions and edges correspond to the relationship degree between these regions. It is common to use an unweighted graph and let the edges represent only the spatial adjacency [30]. However, by using this approach we may lose the detailed contextual information and the results may also suffer from the errors in segmentation (especially small details in urban areas in very high-resolution imagery such as Ikonos or Quickbird). An alternative is to set a fixed threshold for distance and connect the regions that are closer than the threshold with an edge. However, since this approach is scale dependent, it can often lead to the addition of unrelated neighbors in some cases while still losing some important neighbor information in some other cases. Moreover, the space proximity is not sufcient to thoroughly capture the relationship information; therefore, our objective is to concentrate on the proximity in the relationship as well.

The graph is constructed so that vertices represent regions and there is an edge between vertices that correspond to adjacent regions. Namely, for each region R_i there is a corresponding vertex R_i, and for each transition T_{ij} there is an edge connecting vertices R_i and R_j. To let the edges represent the relationship degree rather than only region adjacency, we assign a weight w_{ij} that is calculated as probability of transition corresponding to edge T_{ij} by using (4.2).

By analizing the edge weights, the graph is clustered to find the subgraphs, so that they are composed of vertices with corresponding edges that have high weights modeling frequent spatial co-occurrence. Furthermore, since the relational graph encodes the full spatial information in the image, the subgraphs also contain neighborhood information among multiple region objects. Therefore, the subgraph nodes correspond to the region objects that occur together in a high-level compound structure.

The final objective is to find compound structures that correspond to subgraphs of the complete scene graph. The subgraphs are discovered by using two different procedures. These procedures are discussed below in details.

6.2 Detection of Compound Structures using Subdue

In this work, we use a method that was introduced in [8] and was implemented in the Subdue system for graph-based knowledge discovery. The input and output of the system is a directed or an undirected graph with labeled vertices and edges, where input is the original graph and output is the discovered pattern or learned concept. The study presented in [30] applies Subdue to a graph constructed by using the information conveyed from the RHSEG segmentation output. Namely, each node is labeled with the region class label of the corresponding region object and the edges represent whether or not region objects are spatially adjacent. In our case, the input to the system is an undirected graph with labeled edges. To assign edge labels, we use N_M modes found by using the procedure described in Chapter 5. Given modes $m_1, m_2, \ldots, m_{N_M}$, the graph can be extended so that it reflects the transition type information. Transitions that were assigned to the same mode can be accepted as relations of the same type. Hence, transition type can be assigned to each edge according to the cluster label (between 1 and N_M). The edges that correspond to transitions that do not belong to any of the N_M modes are removed from the graph. Furthermore, in the constructed graph, the nodes are not labeled since we do no perform any classification of the regions after segmentation, so the relationship information is fully reflected by the edges and their labels.

Subdue searches for substructures (subgraphs) of the input graph that best compress this graph. The compression of the graph by a subgraph is defined as the replacement of this subgraph by a single node in the graph. The compression ability of a subgraph during the search is computed by the minimum description length heuristic [8]

$$Compression = \frac{DL(S) + DL(G|S)}{DL(G)} \tag{6.1}$$

where S is the subgraph being evaluated, $DL(S)$ is the description length of S, $DL(G|S)$ is the description length of the input graph G after it has been compressed using S, and $DL(G)$ is the description length of G. The description length of a graph is computed in terms of the number of bits required to encode that graph. The best subgraph is the one that minimizes (6.1).

The search is performed iteratively by compressing the graph with the best subgraph found in each iteration. The output is a list of subgraphs (in terms of nodes and edges they contain) that represent the discovered patterns together with all occurrences of each subgraph in the input graph. Figure 6.1 presents 3

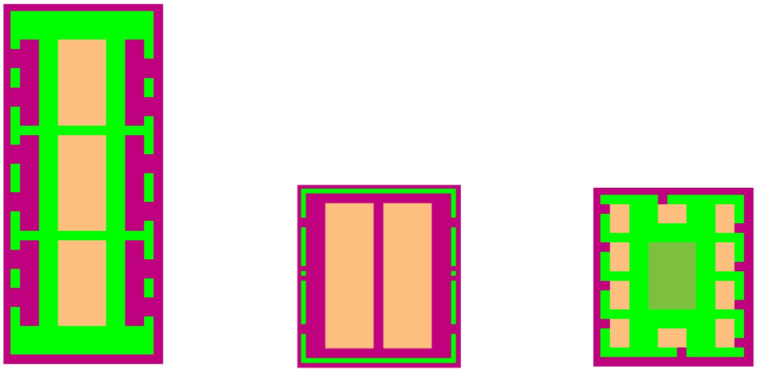


Figure 6.1: Most significant substructures discovered by Subdue.

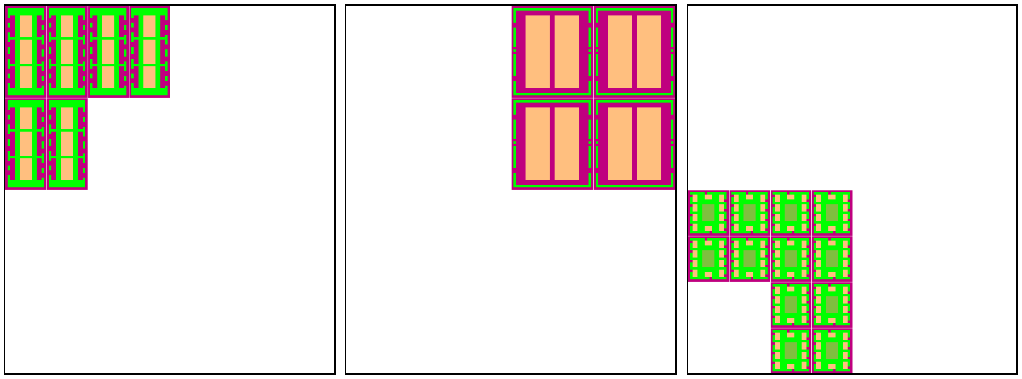


Figure 6.2: Most significant substructure instances.

most significant substructures discovered by Subdue for the simulated segmentation result shown in Figure 4.1. All occurrences of these substructures are shown in Figure 6.2.

It is important to emphasize that although in the simulated image the substructures discovered by Subdue correspond to compound structures, it can not always be so in real urban scenes. One reason for this is the fact that the compound structures are highly detailed and generally comprised of thousands of regions. Moreover, compound structures of the same type vary from one instance to another. On the other hand, Subdue searches for subgraphs that have exactly the same structure. This is possible only for small subgraphs that usually do not cover the whole compound structure. When we change the settings of Subdue to allow a slight deviation between subgraphs by enabling inexact subgraph matches, the run time of algorithm becomes extremely high and the algorithm does not converge. Therefore, we look for the exact matches and the discovered subgraph instances in complex urban scenes generally constitute parts of compound structures rather than whole compound structure.

However, generally the subgraph instances occur frequently in certain structures but rarely in others. Therefore, parts of the image with similar distribution of subgraph instances correspond to compound structures of the same type. To identify these image parts, we use the approach similar to that introduced in [16], where we used spatial relationship histograms to encode image content. In this case, we use histograms for describing the spatial distribution of subgraph instances. The square window centered at a pixel x can be represented by a histogram $h(x)$,

$$h(x) = [h_1(x), h_2(x), \ldots, h_{N_S}(x)], \tag{6.2}$$

where $h_s(x)$ is the number of recurrences of subgraph S in the window. The histogram is computed for each pixel of the image using the sliding windows, where N_S is the number of subgraphs. The constructed histograms are clustered by using the k-means algorithm. The number of desired clusters in this case is defined manually.

Note that classic histograms that count the frequency of occurrence of objects/regions within the window ignore their spatial arrangement. In our case, the spatial arrangement is taken into account because it is encoded in the subgraphs. Also encoding subgraph instances in histograms results in very compact and very effective representations by significantly reducing the dimensionality of the histograms and consequently the computational cost of manipulations on them.

6.3 Detection of Compound Structures using Normalized Cuts Algorithm

Recall that we construct a graph where vertices correspond to regions and edges correspond to transitions between them, or in other words, encode the region adjacency information. In addition, every edge T_{ij} is assigned a weight w_{ij} that is equivalent to transition probabilities and calculated by using (4.2). This extends the graph so that it encodes not only the spatial adjacency of regions but also their proximity in the relationship.

Also recall that our proposed idea for finding compound structures was based on the assumption that compound structures consist of image primitives that are strongly related to each other. Consequently, if we partition the graph into K disjoint sets, so that the sum of the weights of the edges within the set is maximized and the sum of the weights of the edges across the sets is minimized, these sets will correspond to compound structures.

Given a weighted undirected graph $G = \{V, E, W\}$, where V are vertices, E are edges and W is a symmetric nonnegative matrix representing the edge weights, the partition of G into two subgraphs A and B can be obtained by removing the edges connecting A and B, so that $A \cup B = V$ and $A \cap B = \emptyset$. The degree of dissimilarity between two subgraphs can be computed as the sum of weights of removed edges (also referred to as cut):

$$cut(A,B) = \sum_{u \in A, v \in B} w(u,v). \tag{6.3}$$

The normalized cut criterion [25] was introduced for the evaluation of the resulting partition:

$$Ncut(A,B) = \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(A,B)}{assoc(B,V)}, \tag{6.4}$$

where

$$assoc(A,V) = \sum_{u \in A, t \in V} w(u,t), \tag{6.5}$$

is the total connection from vertices of A to all vertices in the graph.

To generalize this bipartitioning-based normalized cuts criterion to multiclass problems, Yu and Shi introduced the multiclass normalized cuts [35]. It can be also referred to as simultanous K-way normalized cuts and denoted by $\Gamma_V^K = \{V_1, \ldots, V_k\}$. The goal is to maximize the K-way normalized associations function:

$$knassoc(\Gamma_V^K) = \frac{1}{K} \sum_{l=1}^{K} linkratio(V_l, V_l), \tag{6.6}$$

where

$$linkratio(A,A) = \frac{assoc(A,A)}{assoc(A,V)}. \tag{6.7}$$

The exact maximization of (6.6) is NP-complete, so Yu and Shi [35] developed an algorithm to find its discrete near-global optima. They first find the global optima in the relaxed continuous domain as the top K eigenvectors of $D^{-\frac{1}{2}}WD^{-\frac{1}{2}}$ subject to arbitrary orthogonal transforms. D is defined as a degree matrix

$$D = Diag(W1_N), \tag{6.8}$$

where Diag forms a diagonal matrix and 1_N stands for $1 \times N$ vector of all 1's. During the discretization step, they use singular value decomposition and non-maximum suppression in an iterative procedure to obtain the discrete solution closest to the continuous optima.

We use the simultanous K-way normalized cuts algorithm to obtain a partition $\{C_1, \ldots, C_K\}$ of graph G. Since the choice of K still remains an open problem, we assign the best K experimentally by examining the resulting partitions.

After the subgraphs are finalized, we can cluster them according to the distribution of transition types within each subgraph. Recall that the transition type can be assigned according to the closest mode among the selected N_M modes. Each subgraph C can be represented by a histogram $h(C)$

$$h(C) = [h_1(C), h_2(C), \ldots, h_{N_M}(C)], \tag{6.9}$$

where $h_i(C)$ is the number of transitions that were assigned to mode $\mathbf{m}_{\boldsymbol{i}}$ within the subgraph C.

After the corresponding histogram is computed for every subgraph, these histograms are clustered by using the k-means algorithm. The number of desired clusters is defined experimentally according to the number of different compound structure types present in the image.

Chapter 7

Experimental Results

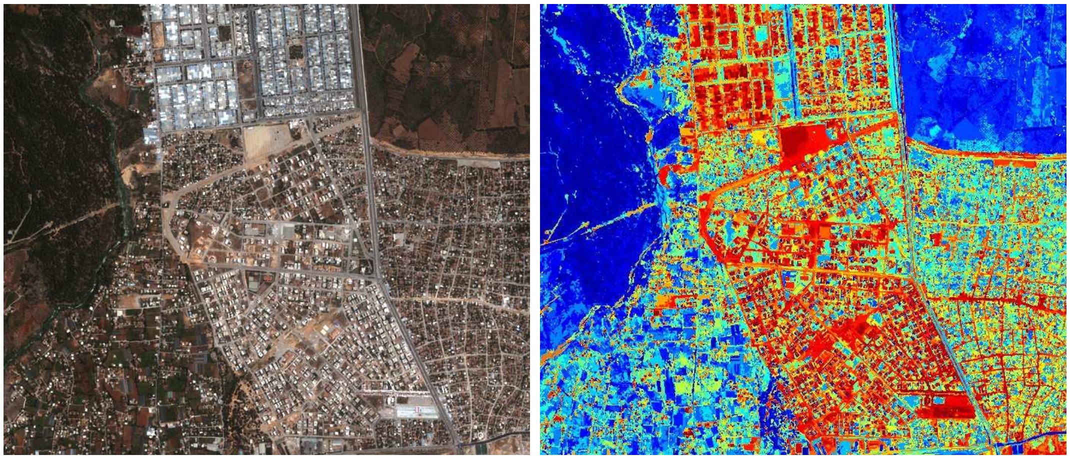
In this chapter, we present the results of the experiments conducted for the methods proposed in this thesis.

7.1 Dataset

The experiments were performed by using the Ikonos image of Antalya, Turkey, with 4 m spatial resolution and 3551 × 3128 pixel size. The image consists of 4 bands: red, green, blue and near-infrared. We use this image because its content is diverse, including several types of compound structures such as dense and sparse residenctial areas with large and small buildings and fields. The image is presented in Figure 7.4(a).

7.2 Experiments with Subdue

In this section, we present the experiments conducted for substructure discovery using the Subdue system. Due to the computational limitations of Subdue, we had to reduce the dataset by using part of the original image, namely the 700×600 pixel size section containing multiple compound structures (Figure 7.1(a)). The



(a) Original image in true color

(b) Segmentation in false color (51558 region objects)

Figure 7.1: Visual bands of an Ikonos image of Antalya with 4 m spatial resolution and 700×600 pixel size, and the selected RHSEG result. Default parameter values of RHSEG are used as explained in [21].

third segmentation scale (Figure 7.1(b)) was chosen by visual inspection among the 11 scales produced by RHSEG. According to this segmentation scale, there were 51,558 regions and 263,246 transitions present. This resulted in a spatial co-occurrence space containing 263,246 points. By using these points, the bandwidth parameter was estimated as $\sigma = 0.0188$.

The convergence threshold for the mean-shift algorithm was empirically set to 10^{-6} and the maximum number of iterations allowed was 4,000. We ran the algorithm 1,400 times starting at different sets of randomly selected points. This resulted in 1,197 unique candidate modes. After mode merging and the elimination of the symmetric modes, the number of modes was reduced to 271.

95 modes were chosen as significant ($N_M = 95$) by applying a threshold to the corresponding probability values. The Subdue algorithm was applied to the constructed graph, and the resulting substructures (subgraphs) were examined. Some example substructures and the corresponding region groups are shown in Figure 7.2. Even though a single substructure does not exclusively correspond to a particular compound structure, we can observe that different substructures constitute parts of different compound structures. For example, the substructure

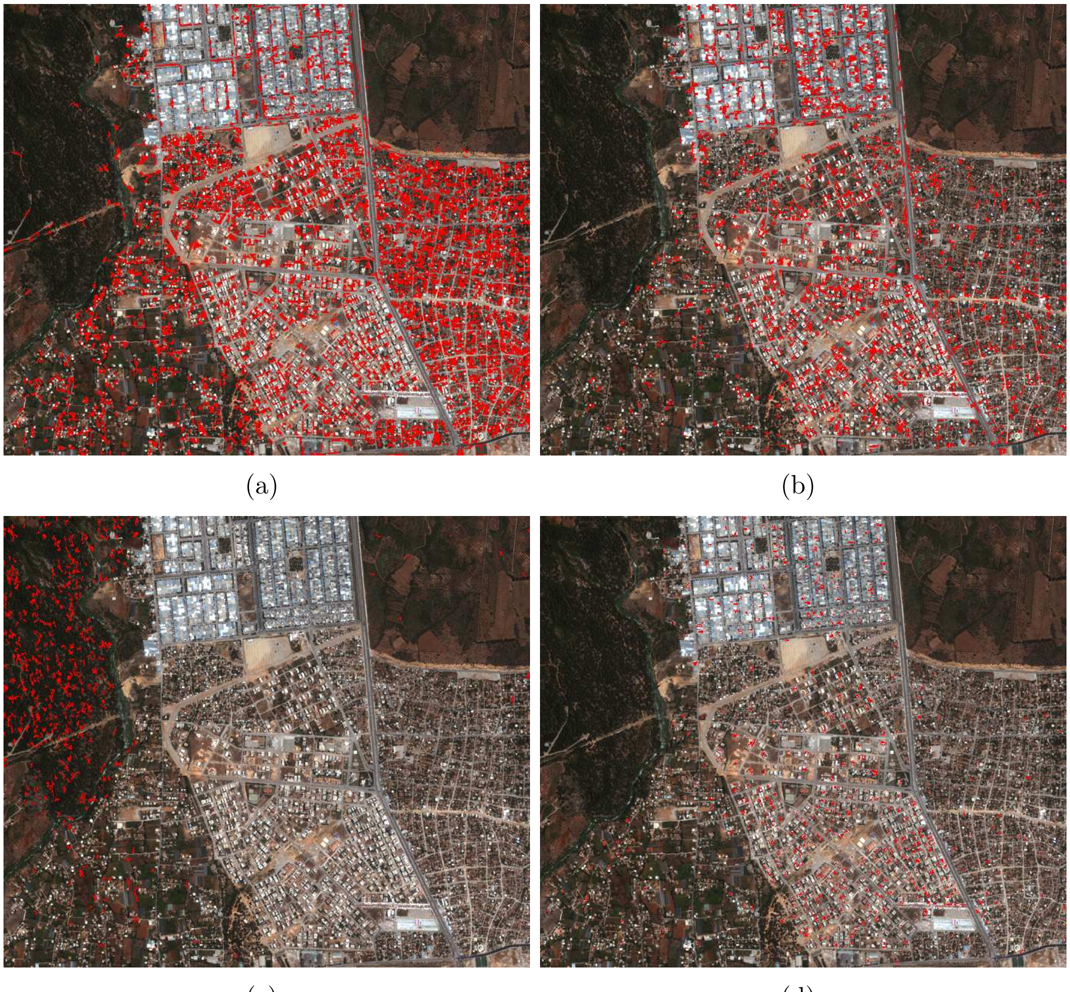
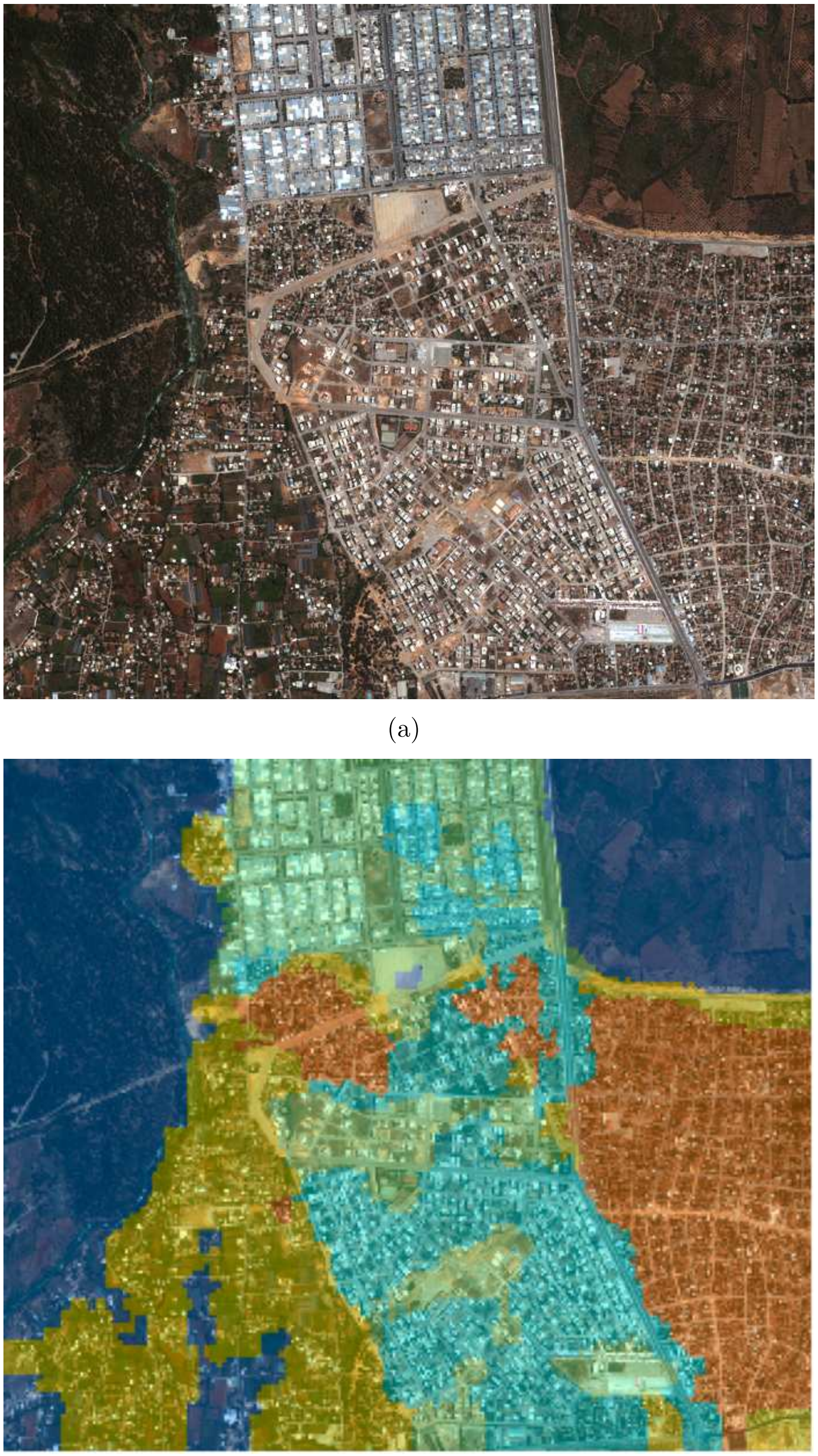


Figure 7.2: Example substructures obtained by graph analysis. The regions that are involved in different substructure instances are shown in red in different subfigures.



(a)

(b)

Figure 7.3: (a) An Ikonos image of Antalya, Turkey and (b) segmentation obtained by clustering the substructure histograms of sliding image windows.

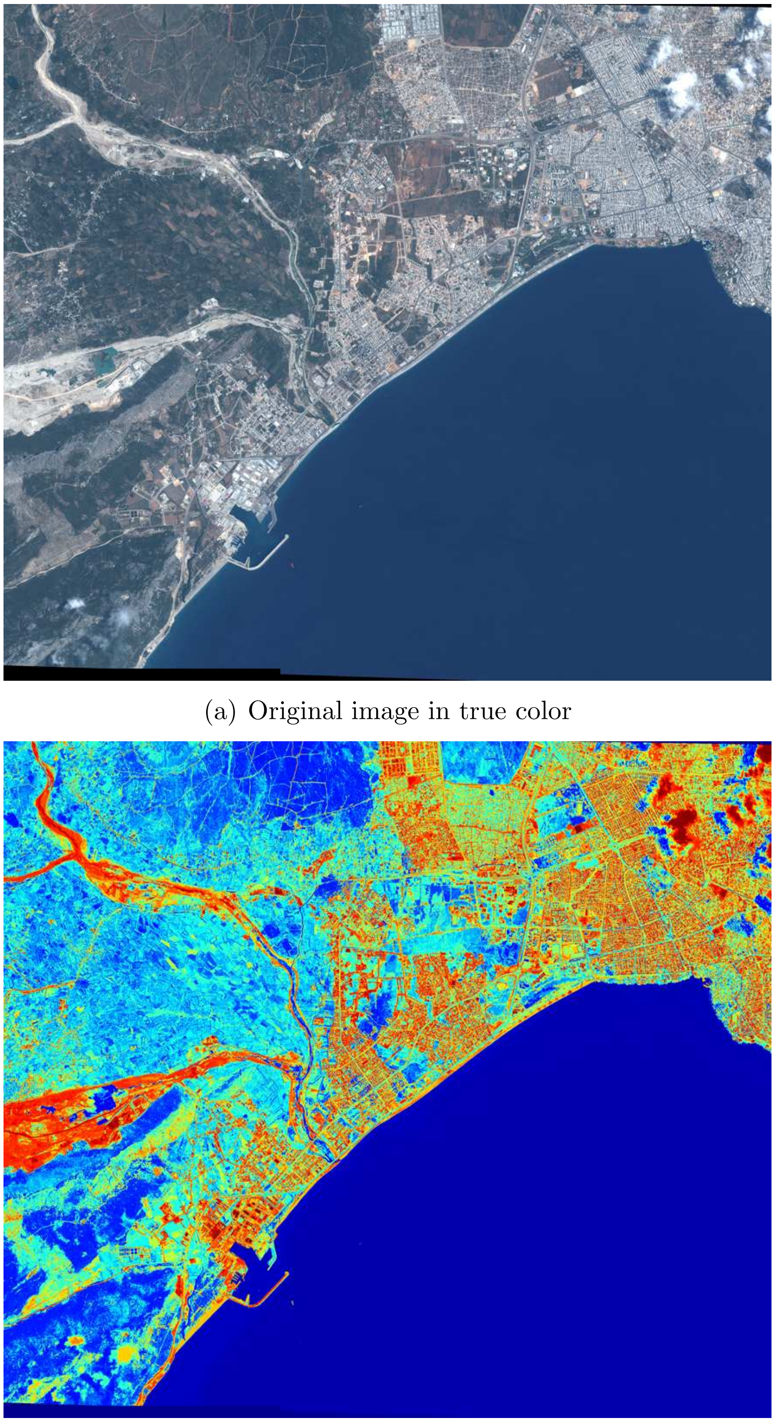
instances in Figure 7.2(a) mostly constitute the parts of residential areas with small buildings. Similarly, the instances in 7.2(b) mainly correspond to parts of an industrial area and a residential area with large buildings, and the instances in 7.2(c) are contained within a forest. Substructure instances in 7.2(d) correspond to roof tops of high buildings that were initially divided by segmentation.

To delineate the compound structures, we use the first 22 substructures discovered by Subdue, so the constructed histograms contain 22 bins. The histograms are calculated using 50×50 sliding windows with 5 pixels increments for computational efficiency. This resulted in 14,300 histograms. The result of clustering the histograms by using k-means with $k = 5$ is presented in Figure 7.3(b). Observe that compound structures of different types are assigned different labels and most of their boundaries are detected accurately.

7.3 Detection of Compound Structure using Normalized Cuts Algorithm

In this section, we present the experiments conducted for substructure discovery using the normalized cuts algorithm. The segmentation is performed by using the default parameters of RHSEG [21]. Among 11 levels in the produced hierarchy, the 4th level was chosen for further analysis as the most suitable according to the level of detail of regions. It is important to note that all 4 images bands were used for segmentation and 3 bands (red, blue and green) were used for further analysis. The original image and the corresponding segmentation is presented in Figure 7.4. This segmentation level contained 30 region classes and 701,464 different region objects.

Color and size features were extracted from each region object. Size values were normalized after clipping the 0.5 percent of extremely high values. After that, the region transitions were identified and the spatial co-occurrence space was constructed. Large number of regions (701,464) lead to a large number of transitions (3,459,910). This resulted in a spatial co-occurrence space containing



(a) Original image in true color

(b) RHSEG segmentation result in false color (30 region classes, 701464 region objects)

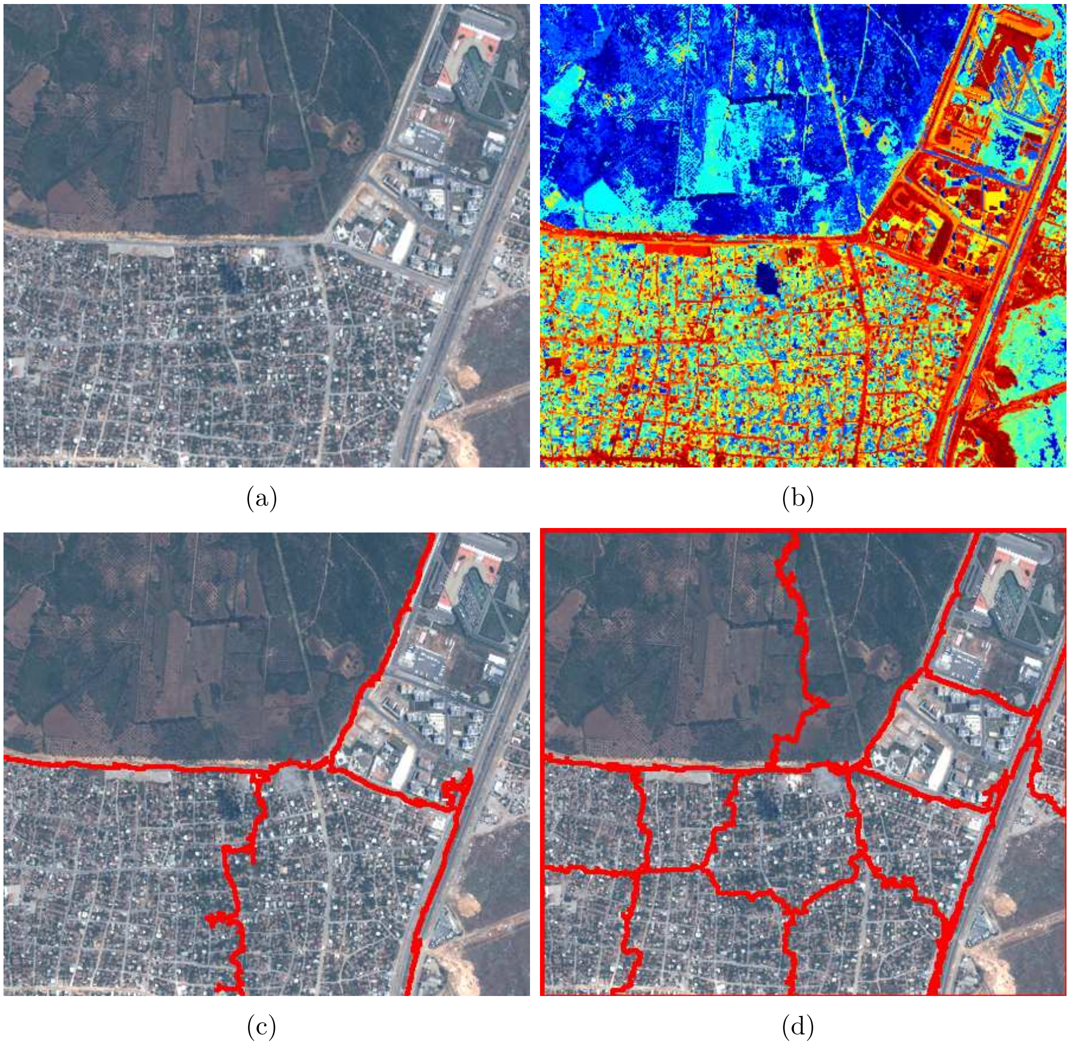
Figure 7.4: Visual bands of an Ikonos image of Antalya, Turkey, and the selected RHSEG result. Default parameter values of RHSEG are used as explained in [21].

3,459,910 points. To reduce the computational compexity, we used only half of this space choosing points by random sampling. Therefore, the employed spatial co-occurrence space consisted of 1,729,955 points. By using these data, the bandwidth parameter was estimated as $\sigma = 0.017$.

The convergence threshold for mean-shift algorithm was empirically set to 2.2204×10^{-16} and the maximum number of iterations allowed was 4,000. We ran the algorithm 1,100 times starting at different sets of randomly selected points. This resulted in 1,098 unique candidate modes. After mode merging and elimination based on symmetry, the number of modes was reduced to 109.

Ideally, our method requires that the graph is contructed by using the whole scene. However, the employed image contains very large number of regions, so manipulations on this graph have very high computational cost. Therefore, we divided the image into overlapping tiles of size 450×400. This resulted in 100 tiles. The graph is constructed for each tile and the normalized cuts algorithm is applied to each graph. We used the implementation of normalized cuts available online [24]. Since each tile has a different content, we had to determine different number of clusters (K) for each graph. Example segmentation can be shown in Figure 7.5. Note that with larger K, the compound structures of different types are fully separated. Even though sometimes the desired compound structures can be divided, they can be merged during further analysis.

After the normalized cuts clustering is applied for each image tile, we obtain the high-level segmentation of the whole image by merging the tiles. However, the segmentation of overlapping parts of two adjacent tiles may not always match. This can be better explained by an example shown in Figure 7.6. In this figure, the overlapping parts are bordered by white lines. Observe that the segmentation of these parts does not match exactly and there is no exact solution for merging the tiles. Hence, for simplicity, we decided to concatenate the tiles without merging their segmentations and let the substructures with similar content merge as a result of further analysis. The resulting segmentation of the whole image is presented in Figure 7.6. This image contains 616 substructures. These substructures can further be grouped and hence merged by clustering the histograms of



(a)

(b)

(c)

(d)

Figure 7.5: (a) Example image tile with (b) the corresponding RHSEG segmentation and segmentation results obtained by the normalized cuts algorithm with (c) K = 4 (d) K = 13.

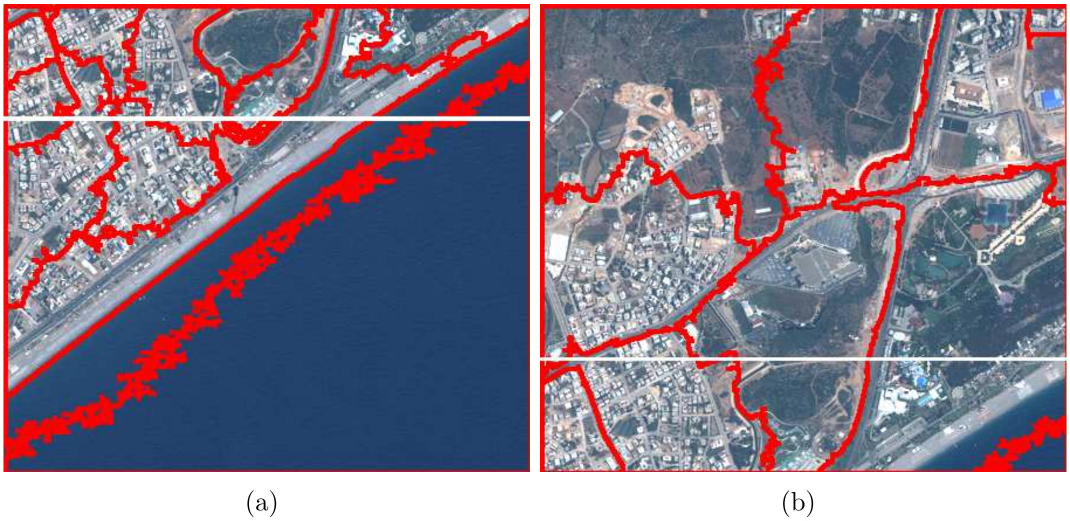


Figure 7.6: Images with overlapping parts with non-matching segmentation. The overlapping parts are shown by white lines.

modes.

Two parameters that can effect the final results are the number of modes (N_M) used in histogram construction and the number of clusters (k) given as input to k-means algorithm. The number of modes to be used and hence the dimensionality of the histograms can be chosen according to the probability of the modes calculated by using (4.2). The 109 finalized modes were sorted according their probability value. Figure 7.8 presents the assigned probability values. Observe that the probability values are very high for the first 5 modes.

As a next step, we want to examine the top 6 modes in terms of transitions assigned to them. Figure 7.9 presents the whole scene where regions involved in transitions that were assigned to a given node are shown in red. Observe that transitions in Figure 7.9(a) correspond to small vegetation primitives, and transitions assigned to second, fourth and sixth modes correspond to larger vegetation primitives. Transitions on Figure 7.9(e) captures the residential areas and transitions on Figure 7.9(c) covers the sea part. Note that transitions on Figure 7.9(b) and Figure 7.9(f) also contain some water transitions. We assume that this errors can emerge during the step there we assign transitions to the closest mode in the space in cases when the closest mode is not close enough to have the

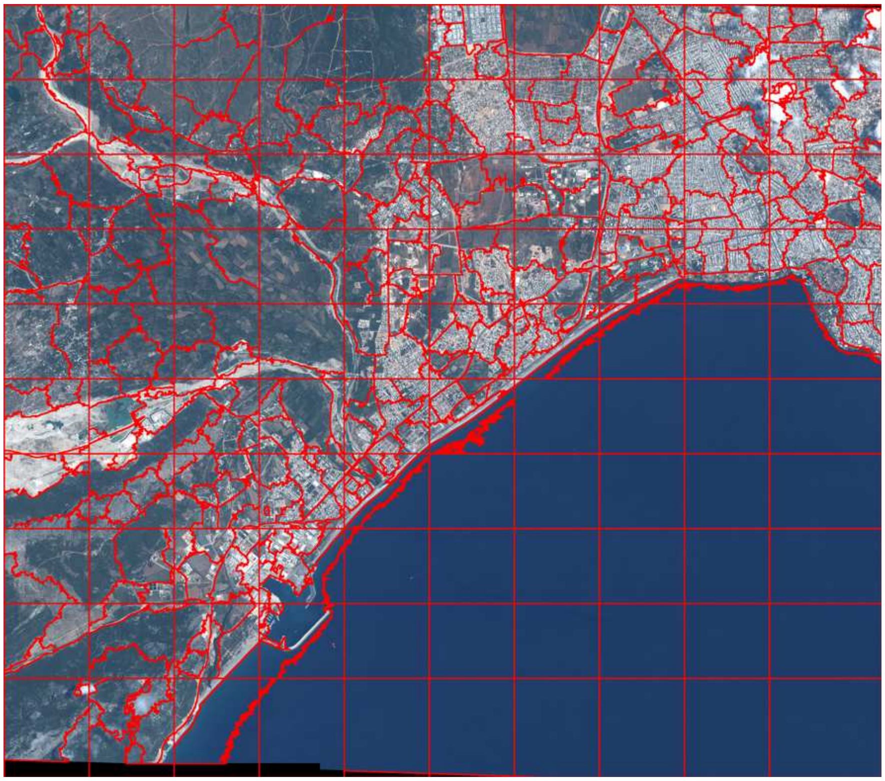


Figure 7.7: The partition of the whole scene obtained by merging the tiles.

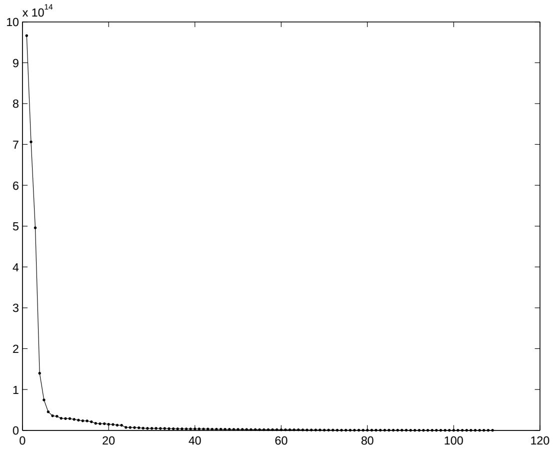


Figure 7.8: Probability values for each mode calculated by using (4.2).

same features.

To be able to evaluate the performance of the proposed method, we manually extracted masks that show the boundaries of dense residential areas with large buildings, dense residential areas with small buildings, sparse residential areas and fields. These masks are shown in Figure 7.10 in different color and details are given in Table 7.1.

Table 7.1: Number of labeled pixels for different area types.

Area Type	**Number of Labeled Pixels**
dense residential areas with large buildings	1108564
dense residential areas with small buildings	257927
sparse residential areas	720376
fields	766243

For a given number of modes N_M and a given number of clusters k, we want to evaluate the resulting clustering set according to the ground truth. To choose the cluster that is the best candidate for the particular groundtruth we employ one-to-one matching. We construct a bipartite graph, where one set of nodes corresponds to obtained cluster labels, and the other node set corresponds to ground truth labels. The match between each two nodes is weighted by the F1 score that can be defined as

$$F1 = \frac{2 \times precision \times recall}{precision + recall}, \tag{7.1}$$

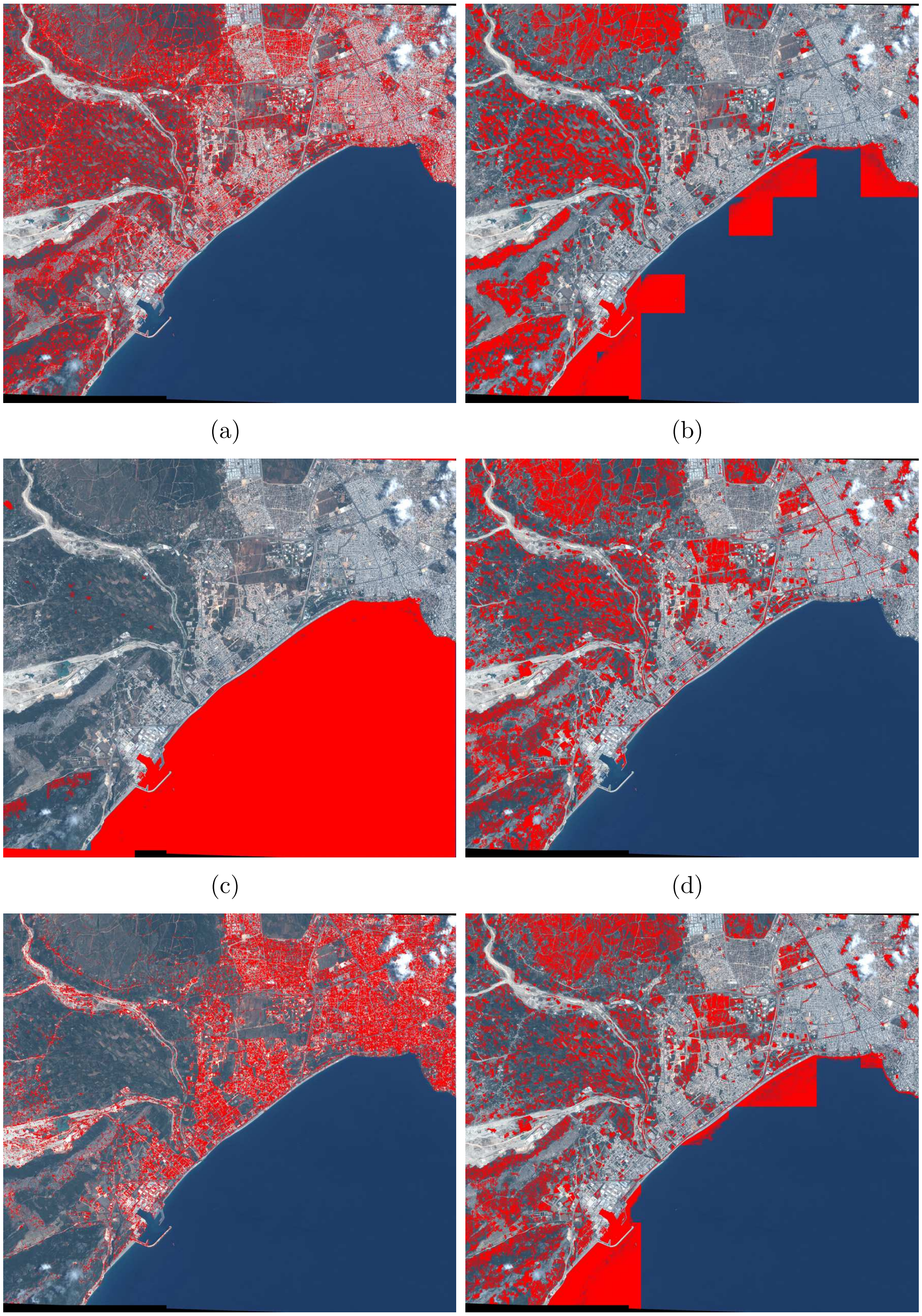
where

$$precision = \frac{\text{\# of correctly detected pixels}}{\text{\# of all detected pixels}}, \tag{7.2}$$

and

$$recall = \frac{\text{\# of correctly detected pixels}}{\text{\# of all pixels in the groundtruth}}. \tag{7.3}$$

The best one-to-one matching configuration between nodes of two sets is found by using Munkres Assignment Algorithm (also known as the Hungarian Algorithm) [20].



(a)

(b)

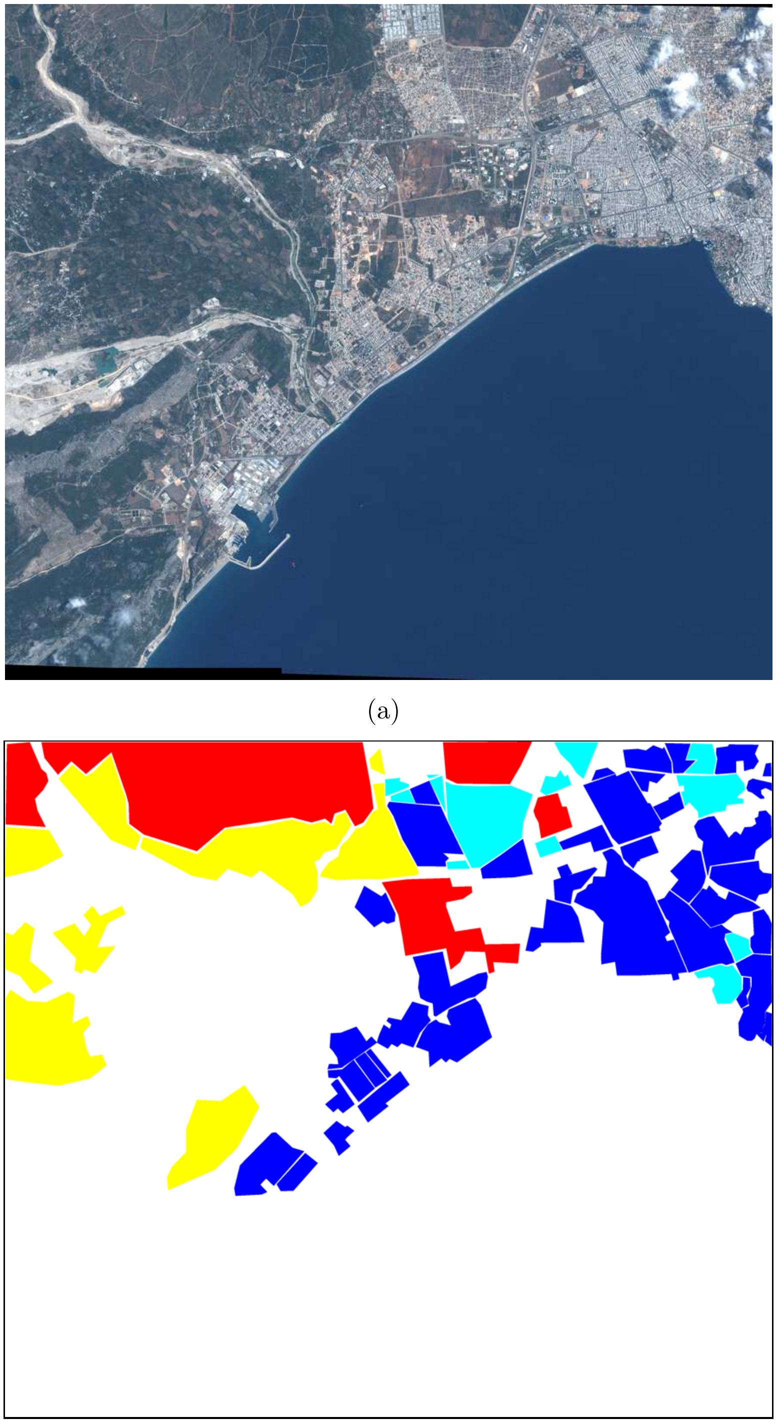
(c)

(d)

(e)

(f)

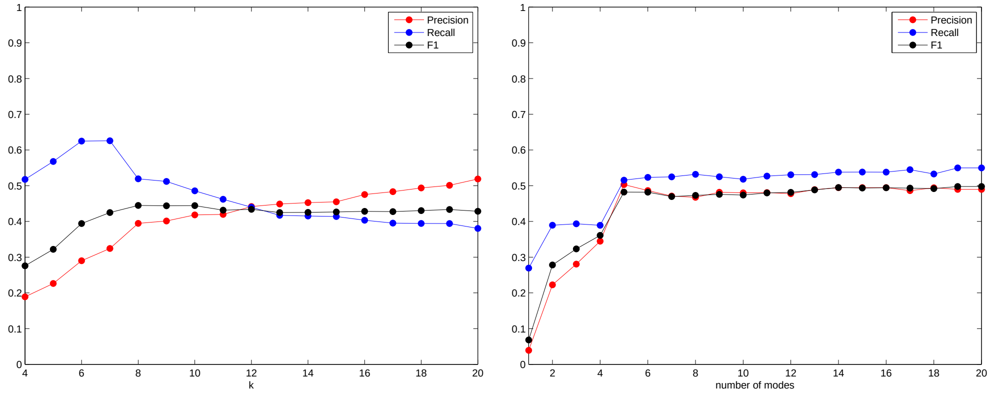
Figure 7.9: Transition assignments for top 6 modes. The regions that are involved in transitions assigned to different modes are shown in red in different subfigures.



(a)

(b)

Figure 7.10: (a) Visual bands of an Ikonos image of Antalya, Turkey and (b) the ground truth extracted from this image. The dense residential areas with large buildings are shown in dark blue, dense residential areas with small buildings are shown in light blue, sparse residential areas are shown in yellow and fields in red.



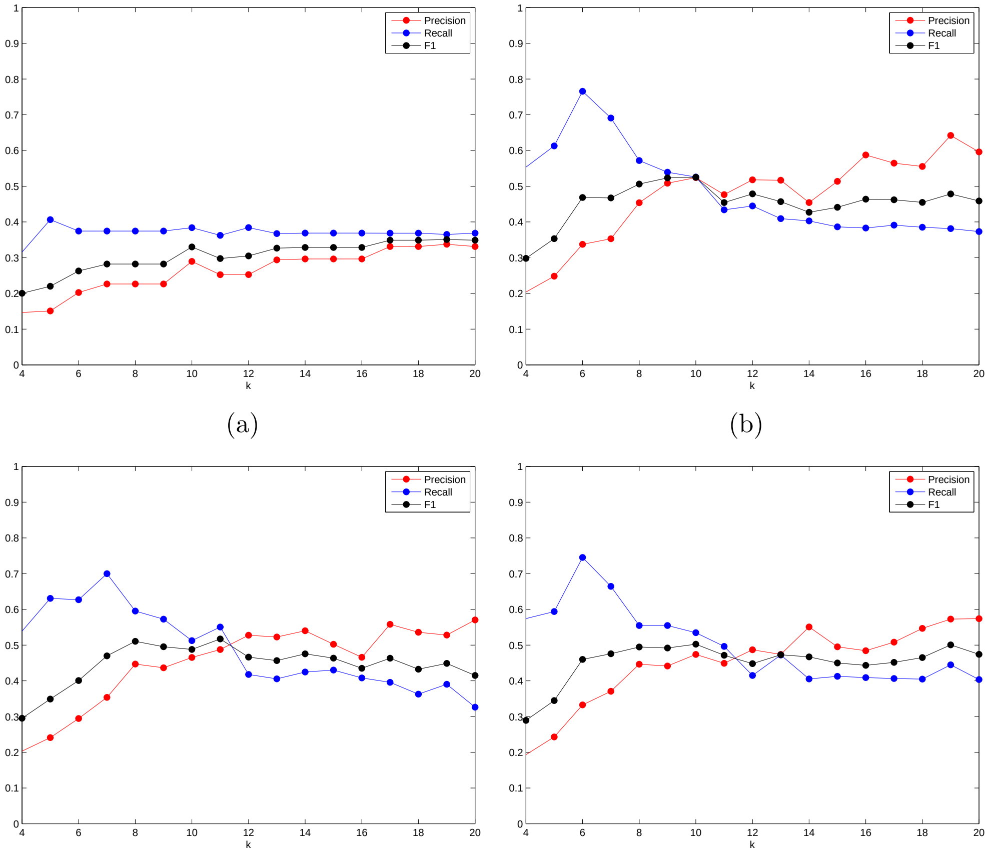
(a) (b)

Figure 7.11: (a) Plot of k versus average F1 scores, precision and recall and (b) N_M versus average F1 scores, precision and recall.

We built 20 sets of histogram for different mode numbers, namely from 1 mode to 20 modes and obtain partition of each set by using different k values (from 4 to 20). Since the cluster set created by k-means can change with every run of the algorithm, we perform clustering of each histogram set 10 times and choose the one with the highest average F1 score. Figure 7.11(a) shows how average precision, recall and F1 score changes with respect to k. Average measures are computed by using the clustering results of all 20 histogram sets for a given k. Observe that as the value of k increases, the average precision increases while avarage recall decreases. This behavior is caused by evaluation using one-to-one matching. Here, k can be chosen as 12 for optimal precision and recall.

Next, we want to examine how the performance changes with the change of number of modes. The plot of number of the modes used in histogram construction versus average precision, recall and F1 scores is presented in Figure 7.11(b). Average measures are computed by using the clustering results for all used numbers of clusters for a given N_M. It can be concluded that the performance does not change significantly after $N_M \geq 5$.

The plots in Figure 7.12 show how precision, recall and F1 score changes with different k according to the particular N_M. Observe that the lowest performance is for $N_M = 3$, and in this case both precision and recall do not exceed 0.4 for all



(a)

(b)

(c)

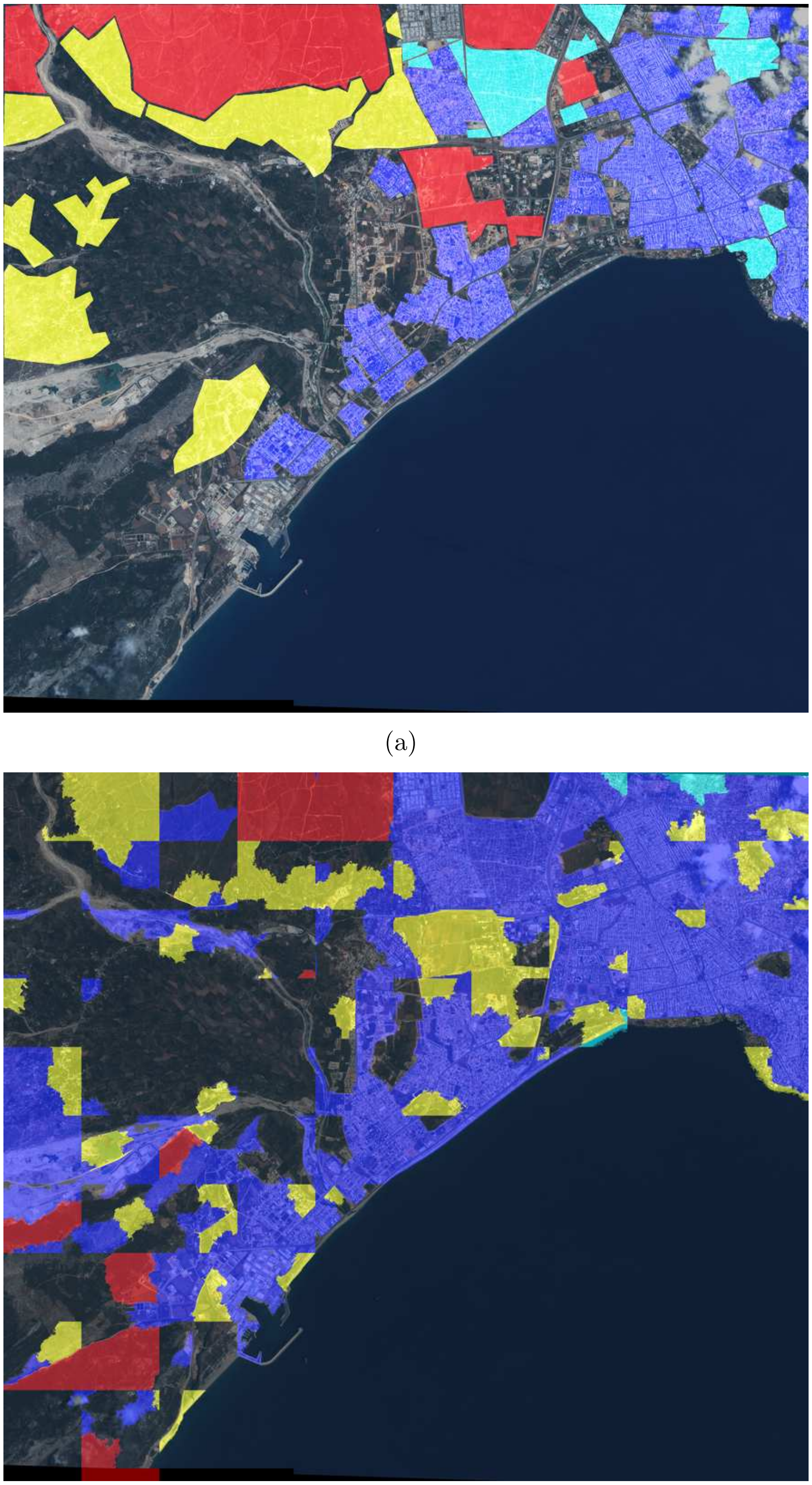
(d)

Figure 7.12: Plot of k versus F1 scores, precision and recall for (a) $N_M = 3$, (b) $N_M = 5$, (c) $N_M = 10$, (d) $N_M = 12$.

ks.

Finally, we present some visual results for detection of compound structures. Figures 7.13 - 7.16 illustrate the detection results for the plots given in Figure 7.12. In each case k that causes the highest performance is chosen.

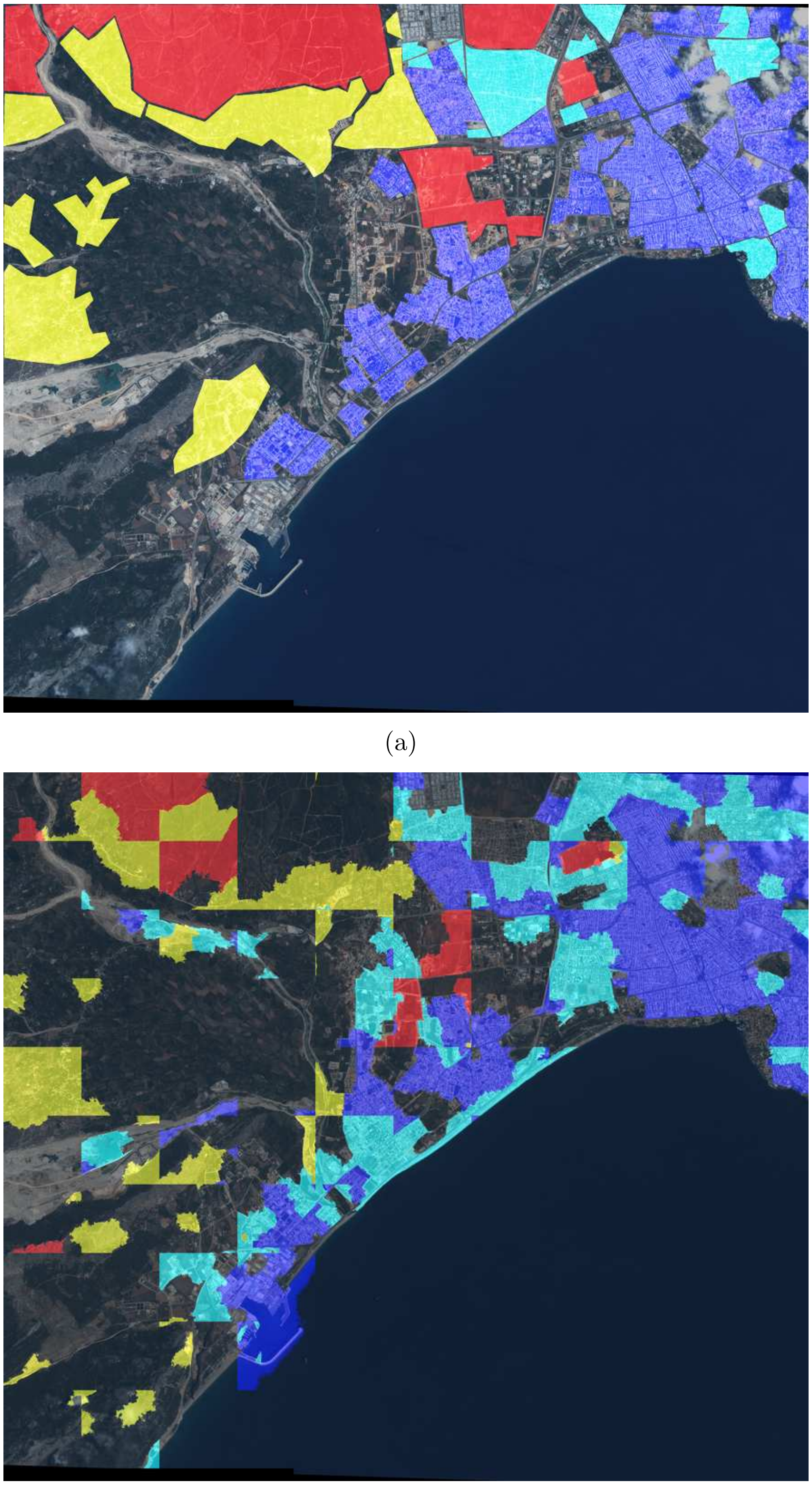
Observe that in Figure 7.13 the delineation of compound structure is poor. Also most of the areas are found to be dense residential areas with large buildings. The results in Figure 7.14 and 7.15 are similar, as expected. Also note that most of the dense residential areas with large buildings are delineated accurately. By comparing these result with the result presented in Figure 7.16, we can observe how precision increases while recall decreases with larger k.



(a)

(b)

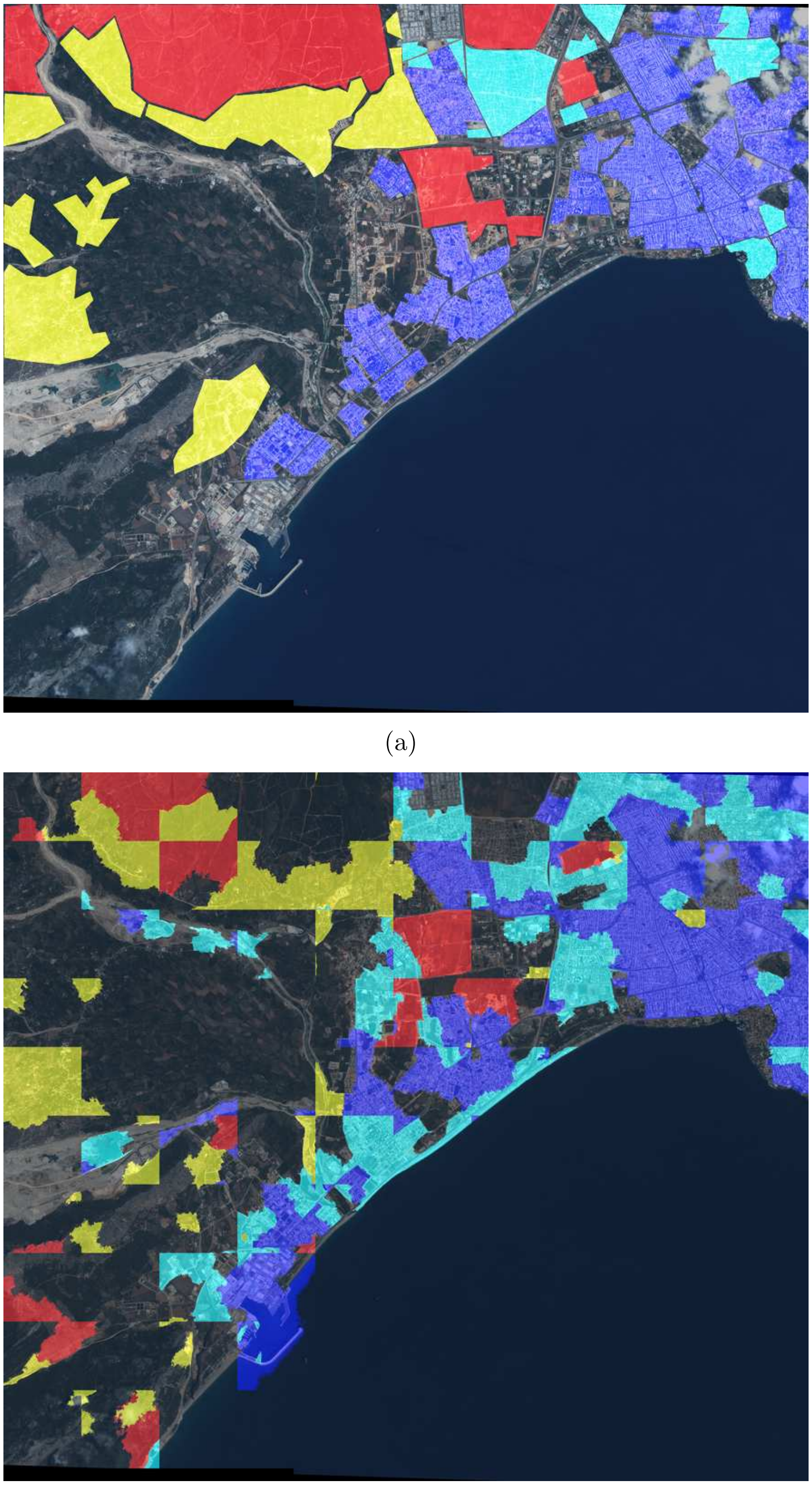
Figure 7.13: (a) The ground truth and (b) the result of compound structure detection with $N_M = 3$ and $k = 10$.



(a)

(b)

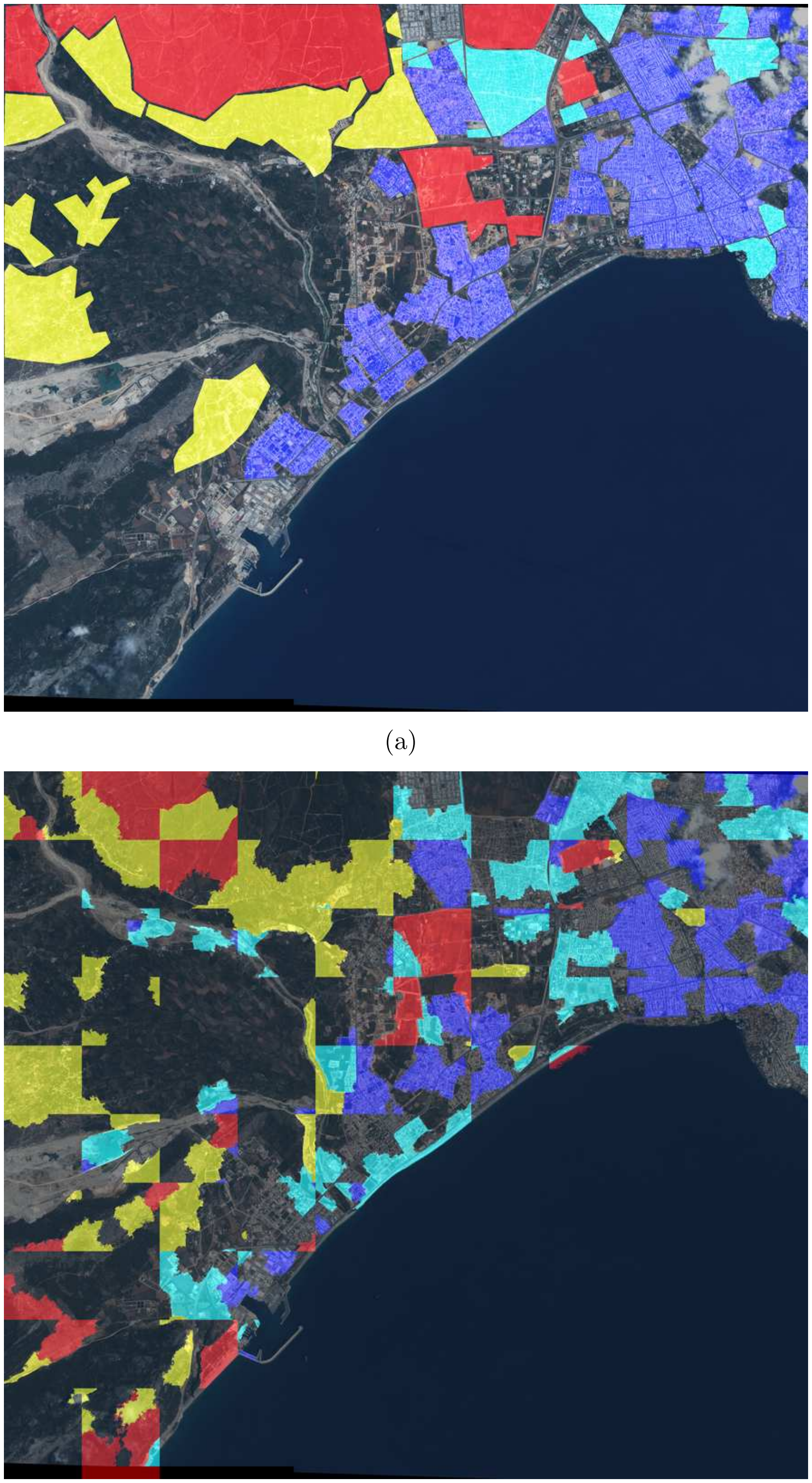
Figure 7.14: (a) The ground truth and (b) the result of compound structure detection with $N_M = 5$ and $k = 10$.



(a)

(b)

Figure 7.15: (a) The ground truth and (b) the result of compound structure detection with $N_M = 10$ and $k = 11$.



(a)

(b)

Figure 7.16: (a) The ground truth and (b) the result of compound structure detection with $N_M = 12$ and $k = 13$.

Throughout the experiments, we observed that the quality of the initial segmentation strongly influences the effectiveness of the following graph analysis. We also observed how the accurracy changes with different number of clusters and different number of modes and noted that the differentiation between various compound structures can be performed when number of used modes is greater than 4.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

In this thesis, we presented an unsupervised method toward discovering compound image structures that were comprised of complex groups of simpler primitive objects. We mentioned the importance of compound structures and discussed that in contrast to primitive objects (buildings, roads, etc.), the compound structures were able to capture more of the image content and subsequently better summarize the scene. We discussed the challenges of detection of compond structures and stated that the traditional segmentation and texture detection methods were not able to handle the complexity of compound structures, so there was a need for generic unsupervised method that can perform detection of compound structures regardless of their types and without preceeding classification of primitives. We suggested to focus on a general property of compound structures that is shared by all the compound structure types: the stong coupling between primitives. We assumed that the primitives that comprised compound structures were strongly related to each other and the degree of this relationship was directly proportional to their transition frequency. As a result, we developed a procedure for transition frequency calculation without a preceeding transition or region type assignment.

The proposed algorithm consists of four main steps. The initial step is image segmentation that produces image regions with homogeneous spectral content. The next step is the construction of spatial co-occurrence space, where each point corresponds to an inter-region transition. The importance of each transition was calculated as likelihood of the corresponding point in the space. The forth step is the identification of most significant relations which were discovered by using the local maxima (modes) in the probability distribution of points in spatial co-occurrence space. The last step is the translation of image segmentation into a relational graph where vertices correspond to regions and edges correspond to inter-region transitions with weights calculated as probability of the matching transition, and discovery of compound structures as the subgraphs of this graph. The substructures were discovered by using two different approaches. One of them is the graph-based knowledge discovery system Subdue that searches for repeating subgraphs within the graph. The other approach is clustering of the graph by using normalized cuts algorithm to obtain subgraphs that consist of regions that are strongly related to each other.

In experimental work, we evaluated the performances of compound structure detection using Subdue and normalized cuts algorithm. Visual result provided for evaluation of the approach using Subdue showed that when the exact match for subgraphs is employed, the discovered substructures correspond to parts of compound structures. We discussed the trade-off between exact and inexact matches and used the histograms of subgraph instances to delineate the compound structures. We also evaluated the performance of compound structure detection using normalized cuts algorithm. We observed how the accurracy changes with different number of clusters and different number of modes and noted that the differentiation between various compound structures can be performed when number of used modes is greater than 4. We compared our results with the ground truth classes and obtained high recall values. We concluded that our method is capable of discriminating compound structures of different types successfully.

8.2 Future Work

Throughout the experiments, we observed that the quality of the initial segmentation strongly influences the effectiveness of the following graph analysis. Therefore, we aim to improve the segmentation result so that the primitives are detected in the most accurate way. Besides, we plan to employ additional features such as shape of primitives and consequenlty extend the spatial co-occurrence space so that it encodes more information. We believe that this will improve the performance of detection of compound structures that generally consist of primitives of particular shape, for example, residential areas and forests. We also plan to try either different implementations of normalized cuts algorithm or other graph clustering algorithms to avoid the division of image into tiles.

Bibliography

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

[2] S. Aksoy and E. Dogrusoz. Modeling urbanization using spatial building patterns. In *Proceedings of 4th IAPR International Workshop on Pattern Recognition in Remote Sensing, Hong Kong*, 2006.

[3] J. M. Beaulieu and M. Goldberg. Hierarchy in picture segmentation: A stepwise optimization approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(2):150–163, 1989.

[4] S. Bhagavathy. Modeling and detection of geospatial objects using texture motifs. Master's thesis, University of California, Santa Barbara, CA, December 2005.

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

[6] T. Blaschke, S. Lang, and G. Hay. *Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications*. Springer, 2008.

[7] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[8] D. J. Cook and L. B. Holder. Graph-based data mining. *IEEE Intelligent Systems*, 15(2):32–41, 2000.

[9] C. Delenne, S. Durrieu, G. Rabatel, M. Deshayes, J. S. Bailly, C. Lelong, and P. Couteron. Textural approaches for vineyard detection and characterization using very high spatial resolution remote sensing data. *International Journal of Remote Sensing*, 29(4):1153–1167, 2008.

[10] E. Dogrusoz. Generalized texture models for detecting high-level structures in remotely sensed images. Master's thesis, Bilkent University, Ankara, Turkey, June 2007.

[11] E. Dogrusoz and S. Aksoy. Modeling urban structures using graph-based spatial patterns. In *IEEE International Geoscience and Remote Sensing Symposium*, 2007.

[12] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification.* 2nd edition, John Wiley and Sons, Inc., 2000.

[13] R. P. W. Duin. On the choice of smoothing parameters for parzen estimators of probability density functions. *IEEE Transactions on Computers*, C-25(11):1175–1179, 1976.

[14] R. Gaetano, G. Scarpa, and G. Poggi. Hierarchical texture-based segmentation of multiresolution remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 47(7):2129–2141, 2009.

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

[16] F. Kalaycılar, A. Kale, D. Zamalieva, and S. Aksoy. Mining of remote sensing image archives using spatial relationship histograms. In *IEEE International Geoscience and Remote Sensing Symposium*, volume 3, July 2008.

[17] V. Karathanassi, C. H. Iossifidis, and D. Rokos. A texture-based classification method for classifying built areas according to their density. *International Journal of Remote Sensing*, 21(9):1807–1823, 2000.

[18] D. Landgrebe. *Signal theory methods in multispectral remote sensing*. Wiley-Interscience, 2005.

[19] H. Lin, L. Wang, and S. Yang. Extracting periodicity of a regular texture based on autocorrelation functions. *Pattern Recognition Letters*, 18(5):433–443, 1997.

[20] J. Munkres. Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, 5(1):32–38, 1957.

[21] NASA Goddard Space Flight Center. *RHSEG and HSEGViewer User's Manual*. 2005.

[22] M. Pesaresi, A. Gerhardinger, and F. Kayitakire. A robust built-up area presence index by anisotropic rotation-invariant textural measure. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 1(3):180–192, Sept. 2008.

[23] G. Scarpa, R. Gaetano, M. Haindl, and J. Zerubia. Hierarchical multiple markov chain model for unsupervised texture segmentation. *IEEE Transactions on Image Processing*, 18(8):1830–1843, August 2009.

[24] J. Shi. MATLAB Normalized Cuts Segmentation Code, available at http://www.seas.upenn.edu/ jshi/software/, 2009.

[25] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[26] V. Starovoitov, S.-Y. Jeong, and R.-H. Park. Texture periodicity detection: features, properties, and comparisons. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 28(6):839–849, November 1998.

[27] M. Stasolla and P. Gamba. Spatial indexes for the extraction of formal and informal human settlements from high-resolution SAR images. *IEEE Journal of Selected Topics in Aapplied Earth Observations and Remote Sensing*, 1(2):98–106, 2008.

[28] J. C. Tilton. Image segmentation by region growing and spectral clustering with a natural convergence criterion. In *IEEE International Geoscience and Remote Sensing Symposium Proceedings*, volume 4, pages 1766–1768 vol.4, Jul 1998.

[29] J. C. Tilton. Parallel implementation of the recursive approximation of an unsupervised hierarchical segmentation algorithm. In A. Plaza and C.-I. Chang, editors, *High-performance Computing in Remote Sensing*. 2007.

[30] J. C. Tilton, J. C. Cook, and N. Ketkar. The integraton of graph-based knowledge discovery with image segmentation hierarchies for data analysis, data mining and knowledge discovery. In *IEEE International Geoscience and Remote Sensing Symposium*, volume 3, July 2008.

[31] J. C. Tilton and S. C. Cox. Segmentation of remotely sensed data using parallel region growing. In *International Geoscience and Remote Sensing Symposium Digest*, volume 1, pages Section WP–4, paper 9, San Francisco,CA, 1983.

[32] 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, 2005.

[33] T. A. Warner and K. Steinmaus. Spatial classification of orchards and vineyards with high spatial resolution panchromatic imagery. *Photogrammetric Engineering and Remote Sensing*, 71:179–187, 2005.

[34] T. Wassenaar, J. Robbez-Masson, P. Andrieux, and F. Baret. Vineyard identification and description of spatial crop structure by per-field frequency analysis. *International Journal of Remote Sensing*, 23(17):3311–3325, 2002.

[35] S. X. Yu and J. Shi. Multiclass spectral clustering. In *IEEE International Conference on Computer Vision*, pages 313–319 vol.1, 2003.

[36] S. Zucker and D. Terzopoulos. Finding structure in co-occurrence matrices for texture analysis. *Computer Graphics and Image Processing*, 12(3):286–308, 1980.