TECHNISCHE UNIVERSITÄT MÜNCHEN
Ingenieurfakultät Bau Geo Umwelt
Photogrammetrie und Fernerkundung

Context-based classification of urban structure types using spaceborne InSAR images

Tessio Novack

Dissertation

2016

TECHNISCHE UNIVERSITÄT MÜNCHEN
Ingenieurfakultät Bau Geo Umwelt
Photogrammetrie und Fernerkundung

Context-based classification of urban structure types using spaceborne InSAR images

Tessio Novack

Vollständiger Abdruck der von der Ingenieurfakultät Bau Geo Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzende:	Univ.-Prof. -Dr.-Ing. Liqiu Meng
Prüfer der Dissertation:	1. Univ.-Prof. Dr.-Ing. Uwe Stilla
	2. Univ.-Prof. Dr.-Ing. Uwe Sörgel Technische Universität Darmstadt

Die Dissertation wurde am 13.11.2015 bei der Technischen Universität München eingereicht und durch die Ingenieurfakultät Bau Geo Umwelt am 27.01.2016 angenommen.

Gedruckt bzw. veröffentlicht mit Unterstützung des
Deutschen Akademischen Austauschdienstes.

Abstract

The uncontrolled growth of urban areas has social and ecological consequences that may lead to various problems. Aiming to mitigate and avoid these problems, many planning tasks and geosciences studies are undertaken based on detailed geometric information of individual urban objects and their changes. However, several other of such tasks and studies rely not directly on the geometry of single objects, but rather on the semantic interpretation of larger aggregated areas, which are usually any of the city's administrative plots. This is the case with the mapping of so-called urban structure types (USTs). USTs classes are always assigned to spatial aggregates based on their social functionality and cultural value or, as is more often the case, on their environmental characteristics as well as the structure and spatial disposition of the objects that compose them.

This thesis proposes an approach to map USTs through context-based classification of high-resolution radar satellite images. The USTs are classified at the level of urban blocks, which are the areas enclosed by streets, railroad tracks and rivers. Two classification methods are applied to attain that goal, namely, Random Forest and a Minimum-Distance algorithm based on the relative entropy metric. As a way to increase the classification accuracy, contextual interactions between pairs of neighboring blocks are established using the framework of probabilistic graphical models. Different graph structures and parameterization of these models are described and compared.

The experiments were undertaken using four images from the TerraSAR-X satellite consisting of two interferometric image pairs obtained at the satellite's ascending and descending orbits. The images cover the central area of the city of Munich in Germany. The performance of the two classification methods as well as the influence of the different graph structures and model parameterizations are evaluated and compared based on different sets of training samples and considering five representative USTs classes.

It is shown that more accurate classifications are obtained when these are performed with attributes extracted from the images from both satellite orbits. The choice of the classification method, i.e. Random Forest or Minimum-Distance, has no significant influence on the accuracy of the results. By considering the contextual interaction between pairs of neighboring blocks, the classification accuracies could be increased to the extent of up to 10%, whereas the best context-based classification achieved the overall accuracy of 78%. In the final considerations, the limitations and advantages of automatically mapping USTs through the classification of remote sensing images are discussed.

Kurzfassung

Das unkontrollierte Wachstum der städtischen Bebauung hat verschiedene soziale und ökologische Folgen, die zu Problemen führen können. Die Erfassung urbaner Strukturen und ihre Veränderung sind zur Bereitstellung von Basisinformationen für Monitoring und Planung in den verschiedensten Bereichen, wie Stadtplanung oder Geo- und Umweltwissenschaften, von hohem Interesse. Für bestimmte Aufgabenstellungen steht jedoch nicht die detaillierte geometrische Beschreibung urbaner Objekte im Vordergrund, sondern es sind vielmehr semantische Interpretationen von aggregierten Bereichen gefragt, wie beispielsweise bei der Kartierung von Stadtstrukturtypen. Dabei erschließt sich häufig ein spezieller Stadtstrukturtyp eines Bereiches nicht nur aus den Oberflächeneigenschaften der Objekte selbst, sondern erst aus ihrem Kontext oder ihrer funktionellen Bedeutung.

Ziel der Arbeit ist es, aus hochaufgelösten Radarsatellitenbildern eine Kartierung von Stadtstrukturtypen durch kontextbasierte Klassifikation abzuleiten und deren Leistungsfähigkeit zu untersuchen. Dazu werden Stadtstrukturtypen auf Ebene von Blöcken klassifiziert, die durch Verkehrswege wie Straßen, Schienen und Flüsse eingegrenzt sind. Als Klassifikationsmethoden werden die Verfahren "Random Forest" und "Minimum-Distance" auf Basis von relativer Entropie eingesetzt. Zur Erhöhung der Leistungsfähigkeit der Klassifikation wird die Berücksichtigung von kontextuellen Zusammenhängen über die Nachbarschaft vorgeschlagen, welche durch probabilistische graphische Modelle beschrieben werden. Dazu werden unterschiedliche Graphenstrukturen gegenübergestellt und die Parametrisierung der Modelle beschrieben.

Für die Experimente werden vier TerraSAR-X Satellitenbilder herangezogen, die aus zwei interferometrischen Bildpaaren der aufsteigenden und absteigenden Bahn bestehen. Die Bilder decken den zentralen Teil der Stadt München in Deutschland ab. Die Untersuchung wird auf Basis der Klassifikation von fünf repräsentativen Klassen mit verschiedenen Trainingsdatensätzen durchgeführt. Es werden die Leistungsfähigkeit der Klassifikationsmethoden, sowie der Einfluss der Modellparameter und der Graphenstrukturen verglichen.

Es zeigt sich, dass die gemeinsame Verwendung der Bilder von beiden Bahnen bei der Extraktion von Attributen der Blöcke zu besseren Klassifikationsergebnissen führt. Die Wahl des Klassifikationsverfahrens ("Random Forest" oder "Minimum-Distance") hat keinen wesentlichen Einfluss auf die Ergebnisse. Durch die Berücksichtigung von kontextuellen Zusammenhängen zwischen benachbarten Blöcken wird die Genauigkeit der Klassifikation bis zu 10% erhöht, wobei die beste kontextbasierte Klassifikation eine Genauigkeit von 78% erreicht. In der Schlussbetrachtung werden die Grenzen der automatischen Ableitung von Stadtstrukturtypen aus Fernerkundungsdaten diskutiert.

Contents

List of Abbreviations

Abbreviation	Description	Page
BD	Building Density	33
CC	Connected Components	33
CPD	Conditional Probability Distribution	33
CRFs	Conditional Random Fields	36
DEM	Digital Elevation Model	26
GIS	Geographical Information System	21
GK	Gauss-Kruger	47
HR	High Resolution	16
HOGs	Histogram of Oriented Gradients	54
HSS	High-Resolution Staring Spotlight	54
InSAR	Interferometric Sythetic Aperture Radar	16
LBP	Loopy Belief Propagation	37
MAP	Maximum A Posteriori	37
MDP	Multidimensional Profile	53
MRFs	Markov Random Fields	35
PGM	Probabilistic Graphical Models	33
PPHT	Progressive Probabilistic Hough Transform	48
SAR	Sythetic Aperture Radar	16
SLC	Single-Look Complex	28
TS-X	TerraSAR-X	23
USTs	Urban Structure Types	16
VHR	Very High Resolution	16

List of Figures

List of Tables

1 Introduction

The percentage of people living in urban areas worldwide has recently surpassed the 50% mark and it is estimated that by 2050 it will reach up to 66%, corresponding to an addition of 2.5 billion people [Nations, 2014]. Although at slower rate, the sprawl and densification of urban areas has been taking place also in highly industrialized countries, whose urban population is presently of about 80% [Nations, 2014]. Only in Germany, urban areas have expanded in 20.3% from 1992 to 2013. In the period of 2010 to 2013, 73 hectares of land in average were converted each day to urban areas in this country [Bundesamt, 2015]. This process of urban sprawl and densification has obviously many consequences and is one of the main causes of several interrelated social and environmental impacts at the local, regional and global scales, such as:

- **Global scale:** deforestation and climate change, biodiversity and ecosystems endangerment, cultural standardization;
- **Regional scale:** increase in the emission of pollutants, much higher demand for energy, water and agricultural goods, increase in waste production, local climate change through the heat island effect, "forced" emmigration from rural to urban areas, urban mobility and traffic congestion issues, sanitation and public health problems;
- **Local scale:** landslides and flood risk, decrease in the quality of life through microclimate and environmental disconfort, air and noise pollution, threat of gentrification, shortage of housing and high rent prices.

Because of that, sustainable development and social equality have frequently been discussed and conceptualized in tight connection with urbanization. Since analysing the problem preceeds its solution, urban sprawl monitoring has become an intensively studied topic in recent years. In this context, remote sensing data and techniques offer many advantages. Several studies have reported the effectiveness of monitoring urban sprawl based on optical and microwave imagery [Pesaresi et al., 2013] [Esch et al., 2010] at both the regional and global scales [Patino & Duque, 2013] [Taubenboeck et al., 2012] [Weng, 2012] [Ban et al., 2015]. However, many of the abovementioned issues related to rapid and uncontrolled urbanization can be more effectively mitigated through planning actions undertaken locally at each city. These planning actions, with the main goals of reducing energy consumption and pollutants emission, are supposed to optimize the access to services and goods as well as the flux of energy and the outflow of disposals in a city. The physical basis on what these factors depend is the city's morphology. In fact, urban climate and radiant energy balance [Steemers, 2003] [Chun & Guldmann, 2014] [Sanaieian et al., 2014], traffic optimization [Weber et al., 2014], urban water and waste management [Chang & Wang, 1997] [Bach et al., 2015], energy consumption [Steemers, 2003] [Ratti et al., 2005] and carbon sequestration [Whitford et al., 2001] [Tratalos et al., 2007] have been studied intrinsically to the city's inner structure. Other factors that also significantly influence the well-being of citizens and the environmental impact of a city such as population density distribution [Li & Weng, 2005]

[Novack et al., 2011b], access to green areas [Barbosa et al., 2007] [Gill et al., 2008], suscetibility to noise pollution [Bolund & Hunhammar, 1999] [Hao et al., 2015] together with other quality of life indicators [Baud et al., 2010] [Hrehorowicz-Gaber, 2013] have also been analyzed in intrinsic relation to the distribution of a city's different types of settlements.

In Germany, these different types of urban settlements are categorized as urban morphology or urban structure types (USTs), which is the conventional term used as the translation of *Stadtstrukturtypen*. USTs serve as a common basis for the categorization of the urban surface which is supposed to guide interdisciplinary planning and monitoring tasks [Pauleit & Duhme, 2000]. The prefecture of Munich, for instance, uses its USTs map as a *"basis for developing urban and landscape plans and programs"*, more specifically *"to identify potential sites for certain projects and assess their impact on the surroundings"* and *"to find potential areas able to be densified"* [Munich, 2014]. Some USTs, based on their names, are defined only on the basis of functional and socioeconomic factors, but most of them are defined based on objective criteria, like the geometry, relative orientation and density of their buildings, as well as by environmental factors such as the presence of vegetation and water bodies. This enables to a great extent the distinction of these USTs by remote sensing methods which are, in comparison to ground surveys and visual interpretation, an automatic, objective, fast and probably lower-cost way to accomplish this task.

Over the last years, a large number of methods based on different remote sensing data have been proposed for the distinction of urban settlements [Patino & Duque, 2013] and the extraction of specific urban objects [Haala & Kada, 2010] [Rottensteiner et al., 2014]. Concerning airbone imaging systems, the detection and 3D reconstruction of single buildings has been performed extensively with airborne Synthetic Aperture Radar (SAR) [Soergel et al., 2009] [Michaelsen et al., 2010], Interferometric SAR (InSAR) [Stilla et al., 2003] [Thiele et al., 2007], optical [Jaynes et al., 2003] [Xiao et al., 2012] and laser scanning systems [Elberink & Vosselman, 2011] [Lu et al., 2014]. However, despite providing information at very high resolution (VHR) (< 1 m), airborne imaging systems usually cover a relatively limited area at each acquisition. This might make the distinction of USTs from large cities a costly task if flight campaings must take place specifically for that goal. Given that spaceborne HR data is by default already available, it might be prefereable to use this type of data depending on the resources available. Besides, optical spaceborne high-resolution (HR) (3 - 1 m) imaging systems provide detailed information over relatively large areas. Their potential for the detection and reconstruction of buildings in urban areas has also been reported by, for example, Tack et al. [2012], Ok [2013] and Turker & Koc-San [2015]. However, despite the high revisit rates guaranteed by the large amount of optical remote sensing satellites presently available, one must still rely on appropriate atmospheric and sunlight conditions at the time the satellite is passing over the study area. These constraints become a bit more critical if specific stereoscopy pairs are needed for 3D analysis. Another possibility is the discrimination of settlement types based on medium-resolution (30 - 10 m) SAR/InSAR datasets. Although covering large areas on the ground and providing 2D/3D image information independently from wheather and sunlight conditions, these datasets enable only the distinction of overgeneralized types of urban settlements, usually based on textural measures and other pixel-based attributes [Dell'Acqua et al., 2003]. Therefore, each of these types of remote sensing data offers advantages and constraints regarding the task of distinguishing USTs. It is then the conditions and requirements of each specific case what will determine the best trade-off between spatial resolution and coverage area as well as the most appropriate type of remote sensing imagery.

1.1 Motivation and Assumptions

The most modern HR spaceborne SAR sensors provide image data for large areas and at good spatial resolution, enabling the distinction of USTs from whole urban areas. Despite of that, efforts on the classification of USTs based on this type of remote sensing data are still very rare, as noticed by Chen et al. [2012], Patino & Duque [2013] and Walde et al. [2014]. This might be because of (1) the unintuitive visual interpretation of these images, (2) the challenge in extracting expressive features from them and (3) the distorted appearance of some objects due to the intrinsic acquisition geometry of these sensors. However, differently from optical data whose main information component is spectral, the backscattering energy from SAR sensors is dependent on the moisture content, the geometric structure of the objects on the earth's surface as well as on their orientation relative to the sensor. Furthermore, the possibility of deriving height information through interferometry makes this type of remote sensing data highly informative about the structure of urban areas. It is hence assumed that HR spaceborne SAR data offer good potential for distinguishing USTs. Furthermore, if a second dataset from a different looking direction is available, one has a comprehensive dataset and this potential is expected to increase significantly.

Regardless of the type of image data used, a key concept in the distinction of USTs is the consideration of *contextual* information. The concept of context in this case has two different meanings. The first one concerns the existence, structure and spatial distribution of different features inside an element of analysis. This is from the physical point of view what usually defines the UST class of this element. The second meaning of context concerns the spatial distribution of the USTs in the area under study. It is an assumption of this thesis that this distribution is not random, but rather that patterns frequently exist, for instance in the form of clusters of elements belonging to the same class. If this is the case, then knowing something about the probable UST of a location may be informative about the USTs of the neighboring locations. The assumption about the existence of patterns in the distribution of USTs is justified a priori by the first law of geography according to Tobler [1970], which states: "Everything is related to everything else, but near things are more related than distant things". This law can be interpreted as asserting that nature is almost never random and that cause-effect relations can be seen everywhere. Tobler emphasizes that this mutual cause-effect relations tend to be stronger between spatially close entities. One expression of this law in particular is of anyone's perception, namely, that similar things tend to be close to each other and close by things tend to be similar. The existence of patterns in the distribution of a city's UST is further supported by the existence in many cities of urban zoning laws, whose purpose is to organise the spatial distribution of the different types of settlements and their uses as a way to make the functioning and administration of the city more effective and easier.

The consideration of context in the distinction of USTs may be statistically modelled in a flexible way by means of the powerful framework of Probabilistic Graphical Models. By establishing a priori affinities or potentials for each possible joint class assignments of neighboring elements and considering also the a posteriori probabilities of each element belonging to each class, these models may be submitted to powerfull existing inference algorithms which are able to consider these element-wise class probabilities together with a possibly very complex network of contextual relationships between elements in oder to find the most probable classification of the entire area under study. Therefore, the uncertainty (or certainty) of element-wise classifications due to the fact that USTs are to some extent subjectively defined and have high inner-class variability is counterbalanced by the influence of neighboring elements. The intensity of this influence is defined, as will be shown, by the number of contextual relations and the potentials of the possible joint class assignments.

1.2 Goals, Research Questions and Contributions of this Thesis

The present work has two major purposes, namely, (1) to explore the full potential of state-of-the-art spaceborne InSAR imagery for distinguishing USTs through classification and (2) to investigate whether considering in their class assignments the mutual influences of contextually connected urban blocks causes, and to what extent, an improvement in the classification's overall accuracy [Congalton & Green, 1999].

More specifically, this thesis explicitly aims to answer the following questions:

- To what extent imagery from a state-of-the-art spaceborne HR InSAR system can be used for the automatic distinction of USTs?
- Which are the most expressive attributes for describing urban blocks in regard to their built-up structure and therefore to their UST classes?
- To what extent does the classification accuracy increases when attributes derived from the images acquired at the other looking direction (satellite orbit) are available?
- To what extent does the classification accuracy increases when the class assignment of each urban block is influenced by the possible class assignments of its neighboring blocks?
- Which of the experimented structures and parameterizations of the Probabilistic Graphical Models enable the highest increase in the overall accuracy of the USTs classification?

The main novelties and contributions of this thesis are:

- For the first time Probabilistic Graphical Models were applied for the classification of USTs. Different application-oriented ways of defining the models graphical structure and parameterization were proposed, evaluated and compared.
- HR spaceborne InSAR data acquired at ascending and descending looking directions was for the first time investigated regarding its potential for performing block-based distinction of USTs.
- As one of the ways of describing the built-up structure of urban blocks, multidimensional profiles were generated based on the application of a sequence of thresholding operations inside each block. A simple metric was used for the comparison of these profiles and hence for performing UST classification.

Other secondary contributions include:

- Strategies are proposed for effectively extracting meaningful lines and polygons from the InSAR images. These features are supposed to express parts of buildings and other man-made objects.
- Many attributes that describe the spatial disposition of these features inside the urban blocks were for the first time explored.
- For each block a network was generated connecting the polygons extracted in it. The spatial disposition of these polygons was then described by different network-based attributes.

- A polygon shape similarity measure was used to define the strength of contextual interaction between two neighboring blocks. The rationale behind it is that neighboring blocks of similar shape are more likely to belong to the same UST class.
- Contextual UST classifications were performed with Probabilistic Graphical Models containing 3rd-order interactions between neighboring blocks. Models with more than one a posteriori class probability term were also explored.

1.3 Structure of the Thesis

This thesis is structured as follows.

In Chapter 2, the most frequently used attributes for describing USTs are firstly presented. Following, it is shown and briefly discussed how the main man-made and natural features that characterize USTs typically appear on HR spaceborne InSAR imagery. Also in this chapter it is discussed the importance of having SAR images from different aspect angles for a more effective detection of features that characterize the USTs.

Chapter 3 addresses the theoretical aspects of Probabilistic Graphical Models. The main types of models and their differences are briefly presented as well as the steps performed during a widely used inference algorithm. Lastly, it is shown how these models are typically defined for performing contextual classification of remote sensing images.

Chapter 4 is the core of this thesis and describes the employed methods and how the methodological steps are interrelated as a means for attaining the goals presented above.

Chapter 5 presents the image and auxiliary data, as well as the study site and the ground-truth data of this thesis. It also describes how the considered UST classes were defined and which specific attributes were used to distinguish them. This chapter addresses as well the application strategies and parameterizations of the standard classification methods adopted in this thesis. The way in which different contextual classification models were evaluated and compared is also presented.

Chapter 6 presents and discusses the results from all performed experiments.

Based on the results obtained, Chapter 7 discusses in a more general perspective the potentials and drawbacks of approaching the application at hand with the proposed methods.

Chapter 8 closes the thesis by making the last commentaries and drawing the main conclusions. Suggestions for future works are made and tasks still to be accomplished are addressed.

I wish you a pleasent reading!

2 Urban Structure Types in Remote Sensing

Remote sensing methods have proven to be very efficient when the goal is to estimate and categorize certain physical aspects of the urban surface, such as degree of imperviousness [Weng, 2012] [Leinenkugel et al., 2011] [Im et al., 2012], vegetation cover [Yunhao et al., 2006] [Tigges et al., 2013] and building density [Chen et al., 2009] [Schmidt et al., 2010] [Yu et al., 2010]. Urban land cover classification is also a task that has been successfully performed with different types of methods and image data [Myint et al., 2011] [Novack et al., 2011a] [Okujeni et al., 2013]. An advantage from these applications is that the characterization of urban areas based on *objective* categories enables to a good extent the methodology to be more easely transferred to other cities with different morphologies and environmental conditions [Heinzel & Kemper, 2015] [Duque et al., 2015].

However, several important decision making and planning actions rely ultimately on the information about the city's land use, which is by its nature, even if the classes are formally defined, a *subjective* concept. (For all practical reasons, USTs can be considered without risk of misconception as a synonym for urban land use, although the earlier term refers mostly to german cities). In the cases where the goal is to distinguish USTs directly from remote sensing data, the expert is necessarily involved with the challenge that subjective and context-defined classes must be described by objective and quantitative image attributes. The assumption that this is to a certain extent a feasible task motivated this and many other works, all of which have in common the fact that context had to be in some way described and taken into consideration in the classification strategy. The extent to which context can be expressed is what most influences the accuracy of the end results. Therefore, special attention is dedicated in the next two sections to the state-of-the-art approaches in that regard. These approaches differ essentially in two aspects, namely, the elements of analysis (i.e. single pixels or image regions) and the descriptive attributes they rely on.

2.1 Pixel-Based Approaches

At pixel-based classification of urban settlements, the creation of meaningful image regions, either through segmentation or through the introduction of auxiliary Geographical Information Systems (GIS) data, is not a concern. Furthermore, such approaches frequently rely on medium to high spatial resolution (10 - 1 m) imagery, what has the consequence that the classification is provided for larger areas at relatively lower computational costs.

The pixel-based approaches for urban settlement distinction can be divided into three general groups. The first consists of works that performed standard pixel-based classification with large image datasets and powerfull classifiers [Dell'Acqua et al., 2003] [Hu & Weng, 2009]. They rely on the comprehensiveness of the image data, the expressiveness of the images pixel values and on

the discriminant power of classification methods such as fuzzy neuro-networks and self-organizing maps [Mather & Tso, 2009]. The second group is from works that have used spectral unmixing techniques to estimate the proportions of impervious, water and bare soil of each pixel and thus distinguish different environments inside the city [Phinn et al., 2002] [Cockx et al., 2014]. These approaches describe indirectly the inner-context of each pixel. The third and most common approach is to compute different texture metrics for each pixel based on a moving window and then perform standard per-pixel classification. The surroundings of each pixel is hence described according to the texture metric(s) and the size(s) of the kernel(s), i.e. the moving window, used. First order statistics [Schmidt et al., 2010], Haralick attributes [Dell'Acqua et al., 2003], wavelet transform [Myint, 2001], fractal indices [Thomas et al., 2008] and SAR-specific statistics [Reigber et al., 2008] [Esch et al., 2011] are some of the metrics frequently used to describe the textural context of each pixel. These kernel-based approaches have though the disadvantages that optimal kernel-sizes must be found (which may vary according to the data spatial resolution) and that the kernel is an artificial spatial entity most of the times not related to the morphology of the city.

Although they had their importance in the past, pixel-based approaches are limited if it is necessary to distinguish more than two or three general classes. Furthermore, they cannot be used if it is required that the classes instances coincide with the city's administrative parcels. Because of that, and due to the present availability of higher resolution sensors and user friendly object-based image analysis (OBIA) software that allow ease integration with GIS data [Benz et al., 2004] [PUC-Rio, 2015], region-based approaches have been adopted in most pertinent publications in the field.

2.2 Region-Based Approaches

Regarding region-based approaches for distinguishing USTs, the regions may be grid-cells [Novack & Stilla, 2014b] [Heinzel & Kemper, 2015] [Bach et al., 2015], user-defined analytical regions [Duque et al., 2006] [Duque et al., 2015] or some of the city's administrative parcels, like census tracts [Baud et al., 2010], urban districts [Yu et al., 2010] or, most frequently, the city's urban blocks [Banzhaf & Hoefer, 2008] [Novack & Kux, 2010] [Heiden et al., 2012] [Voltersen et al., 2014] [Novack et al., 2014]. Being the largest areas surrounded by streets, urban blocks are in most cities (and in all german large cities) the spatial units based on which urban zoning laws are defined and planning actions are conceived and put into practice. Therefore, region-based approaches have not only the advantage that many more descriptive attributes (textural, geometrical and contextual) may be explored, but also that regions like census tracts and urban blocks are frequently homogeneous regaring their built-up structures. Furthermore, classification results can be in these cases straightfowardly used for administrative purposes. Because of that, UST classifications performed for the german cities of Leipzig [Banzhaf & Hoefer, 2008], Jena [Walde et al., 2014], Berlin [Voltersen et al., 2014], Rostock [Walde et al., 2013], Erfurt [Walde et al., 2013], Cologne [Wurm et al., 2009], Dresden [Hecht et al., 2013] and Munich [Heiden et al., 2012] [Novack & Stilla, 2014a] have all considered the urban blocks as the analysis units.

As mentioned, an important advantage of region-based approaches is that the spectrum of descriptive features is much larger. In fact, all possible pixel-based attributes can be aggregated to the region level as first order statistics (i.e. mean, max., std. deviation etc.) [Schmidt et al., 2010] or by computing texture measures for the region as a whole. Van de Voorde et al. [2009] and Cockx et al. [2014], for example, computed the cumulative frequency distribution of the proportion of sealed surface from each of the blocks pixels and then fitted the cumulative frequency distributions to the exponential and sigmoid functions. The authors argument that the parameters of the fitted functions, the errors of fit and the difference between the fitted and observed cumulative frequency distributions are expressive metrics for estimating a block's UST. Recently, Duque et al. [2015]

estimated the degree of poverty from user-defined analytical regions of Medellin (Colombia) as a reflection of their settlement structures. They proposed a statistic regression model whose independent variables include several texture measures proposed by Haralick et al. [1973], as well as histogram and semivariogram measures from analytical regions. Schmidt et al. [2010] also proposed a regression model for estimating 2D building densities of urban blocks based on the texture measures by Haralick et al. [1973] and on first order statistics of TerraSAR-X (TS-X) intensity imagery. Novack & Stilla [2014b] have used TS-X data obtained in High-Resolution Spotlight Staring (HSS) mode and Histogram of Oriented Gradients (HOGs) [Dalal & Triggs, 2005] in order to classify urban blocks according to three general settlement types.

Texture metrics can be expressive and describe to a certain extent the structure of the settlements inside urban blocks. However, higher-level attributes that describe their *composition* have been more frequently resorted to when the goal is specically to classify USTs or the urban land use. Composition attributes are all created based on the same principle, namely, that firstly features are extracted from inside the blocks and then the existence, amount, area proportion and geometrical properties of these features are computed and used to describe the blocks USTs. Composition attributes have been adopted by many works aiming to classify USTs with multispectral [Lackner & Conway, 2008] [Voltersen et al., 2014], hyperspectral [Linden & Hostert, 2009] [Heiden et al., 2012], LiDAR [Heiden et al., 2012] [Walde et al., 2013] and SAR imagery [Novack & Stilla, 2014a] [Novack & Stilla, 2015]. In these and other related works, the features are patches from land cover classes extracted by a classification step performed at each block. The most common composition attributes are *number of*, *area*, *percentage area* and *maximum area*. If A_b denotes the area of urban block b, a_{ci} the area from patch i belonging to class c located inside b and N_c the number of patches from c inside b, then these four composition attributes can be calculated as:

$$area = \sum_{i=1}^{n} a_{ci}, \qquad max.\ area = \max_{i}(a_{ci}), \qquad per.\ area = \frac{\sum_{i=1}^{n} a_{ci}}{A_b} \times 100, \qquad number\ of = N_c. \tag{2.1}$$

These metrics can be combined for the calculation of other simple ones, like *patch density* (PD), *largest patch density* (LPD) and *mean area*:

$$PD = \frac{N_c}{A_b} \times 100, \qquad LPD = \frac{\max_{i}(a_{ci})}{A_b} \times 100, \qquad mean\ area = \frac{\sum_{i=1}^{n} a_{ci}}{N_c}. \tag{2.2}$$

Along with these five attributes, *aggregation index* (AI) has also been used for describing the structure of a land cover class inside a block [Baud et al., 2010] [Huck et al., 2011] [Novack et al., 2011b]. The AI metric is calculated by counting the number of neighboring pixels from the same class, called like-adjacencies, divided by the maximum possible number of such adjacencies. Figure 2.1(a) shows patches from a hypothetical class distributed in completely aggregated and disaggregated forms. In the first case, the AI value would be of exactly one and on the second case of exactly zero.

In addition to *area*, the first order statistics from other geometric properties of the land cover patches inside the blocks can be used as attributes as well. These geometric properties are usually *rectangular fit* (RF), *compactness*, *perimeter / area*, *major / minor axis length* etc. RF is the ratio between the area of a object and the area of its smallest enclosing rectangle and *compactness*, also refered as shape index, equals the perimeter of a feature divided by the minimum perimeter possible for a maximally compact feature. Figure 2.1(b) shows an example of two features with the same area but different compactness values.

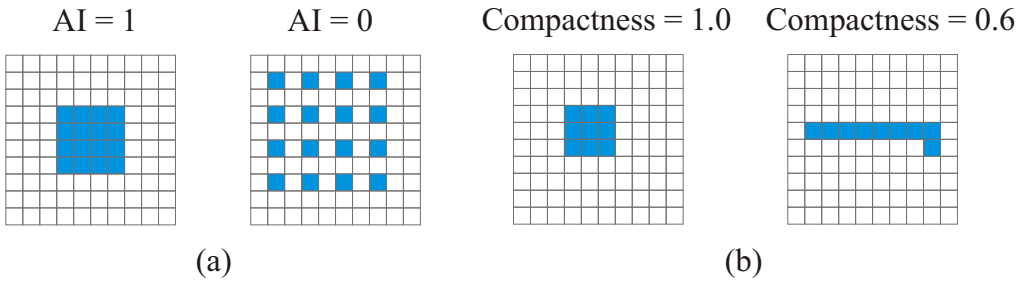

Figure 2.1: Two regions with different *aggregation indexes* for a certain class and two features with different compactness values. a) Two regions whose features from a certain class are distributed in completely aggregated (left) and disaggregated form (right). b) Two different features with same area but very different compactness values.

Composition attributes are usually calculated for general land cover classes, like vegetation, impervious surfaces, roofs and buildings. Numerous works have used them for the classification of urban land use and USTs [Pauleit & Duhme, 2000] [Herold et al., 2005] [Banzhaf & Hoefer, 2008] [Wurm et al., 2010] [Huck et al., 2011] [Coseo & Larsen, 2014] [Voltersen et al., 2014], all of which rely on the fact that the previous land cover classification is accurate enough so that these attributes are expressive for distinguishing the classes. This assumption is supported by the fact that the OBIA techniques resorted to in these works, such as multi-resolution segmentation, interface with GIS data and semantical and hierarchical structuring of classes, ensure the extraction of spatially consistent and semantically coherent land cover patches [Benz et al., 2004] [Blaschke, 2010].

Since buildings are the most important class of features for describing the built-up structure of an urban block, specific building metrics have been utilized as attributes whenever their footprints can be detected and their heights estimated. The most common of these metrics are *2D building density* (BD), *3D BD*, also known as normalized building volume, *average 3D BD* [Chen et al., 2009] and *average floor to area ratio (AFAR)* [Yu et al., 2010]. These metrics are computed as follows:

$$2D\ BD = \frac{\sum_{i=1}^{n} a_{buildings}}{A_b} \times 100, \qquad 3D\ BD = \frac{\sum_{i=1}^{n} V_i}{A_b}, \tag{2.3}$$

$$Av.\ 3D\ BD = \frac{\sum_{i=1}^{n} V_i}{A_b N_{buildings}}, \qquad AFAR = \frac{1}{N_b} \sum_{i}^{N_b} \sum_{j}^{t} A\_floor_j / A_b, \tag{2.4}$$

where V_i stands for the volume of building i, N_b is the number of buildings in block b with area A_b, j indices a floor from building i. Although *2D* and *3D BD* describe to a good extent the structure of any urban block, they do not take into account respectively the height and the number of buildings. In this sense, Average 3D BD is a more expressive metric. Other attributes, such as *average building height*, *average footprint size*, *average volume*, *maximum height*, *maximum footprint size*, *maximum volume*, *standard deviation of height*, *standard deviation of size*, *standard deviation of volume* and *number of buildings* have also been widely employed for region-based categorization of urban settlements [Yu et al., 2010] Heiden et al. [2012] [Hecht et al., 2013] [Voltersen et al., 2014] [Heinzel & Kemper, 2015].

Other works have put efforts in classifying individual buildings after extracting different geometric and contextual attributes from them [Steiniger et al., 2008] [Hecht et al., 2013]. The

context of the individual buildings are in these cases described by defining a buffer zone around each building and then counting the number of buildings, vegetation and road features inside the buffer, or by computing the mean, maximum and standard deviation of the buildings volume around the central building. Yu et al. [2010] and Lu et al. [2014] have used this approach to describe the settlements of urban blocks by generalizing to the whole block the attributes of the dominant building and those from secondary buildings. The drawback of these building metrics is that they require VHR 2D and 3D image data, what sometimes prohibitively increases processing time and data costs for larger areas applications.

The composition and building metrics presented above only describe partially the built-up structure of image regions. Another very important group of attributes for region-based classification of USTs can be called *configuration* attributes. Configuration attributes describe the spatial disposition and connectivity of the extracted features inside the regions. The motivation for exploring such attributes comes from the fact that certain UST classes like Perimeter Block Development, Row House Development and Regular Block Development can only be properly described and thus accurately distinguished if such attributes are considered. Despite of that, only few works on region-based classification of urban settlements have done this.

Simple examples of such configuration attributes are: *average nearest distance between buildings* [Yu et al., 2010], *minimal distance of the building contour to the closest block boundary* [Hecht et al., 2013], *number of buildings with orientation between...* (for different angle intervals) [Voltersen et al., 2014], etc. A type of more elaborated configuration attributes were proposed by Vanderhaegen & Canters [2010]. They defined for each block radial and contour profiles (Figure 2.2) and then calculated for each profile different indexes related to the building/non-building alternations, as well as to the lengths of the building and non-building profile segments. These indexes are: *normalized number of building/non-building alternations*, *average length of building/non-building segments* and *standard deviation of the normalized length of building/non-building profile segments.*

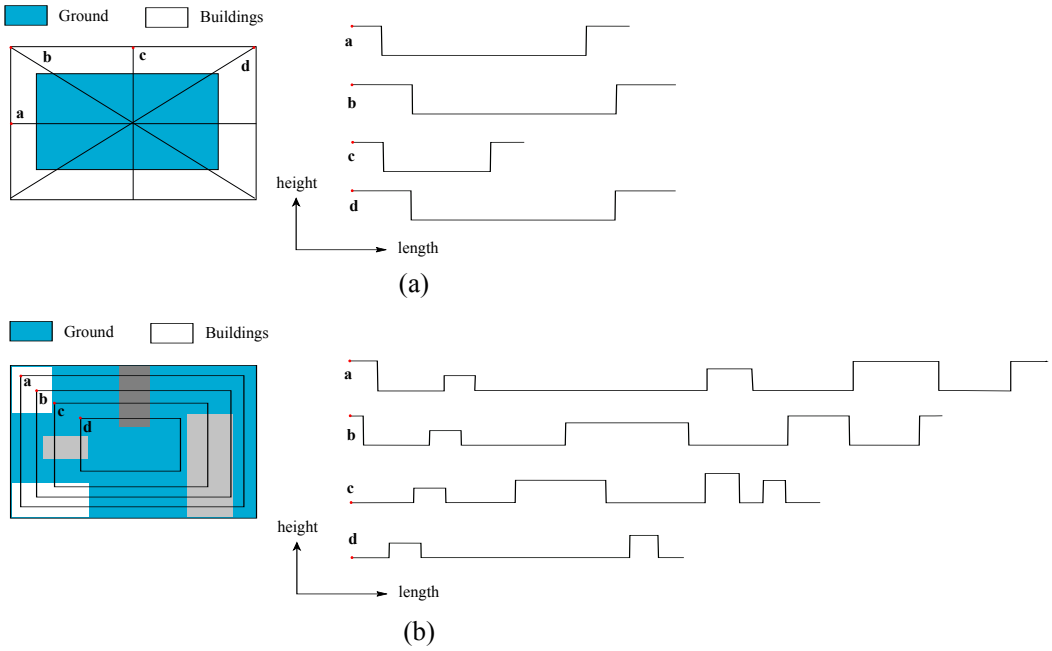

Figure 2.2: Radial and contour height profiles from two hypothetical urban blocks. a) Radial height profile from a hypothetical block. b) Contour height profile from a hypothetical block. Different indexes can be computed from the profiles. The indexes describe to some extent the disposition of the buildings inside the blocks. Adapted from Vanderhaegen & Canters [2010].

Another way of describing the configuration of features inside a block is by creating a *neighborhood graph* from the features inside it. A neighborhood graph is created usually by defining one node for each feature of one or more types of features (e.g. buildings, vegetation types, etc.) and then establishing edges between neighboring features. Two features can be considered neighbors based, for example, on the criteria of direct adjacency or distance thresholds. Let us assume that Figure 2.3(a) shows a synthetic image whose white and light grey objects have been extracted and shown in blue Figure 2.3(b). An example of a graph from the objects shown in blue is shown in Figure 2.3 (c). Several measures can then be derived from this graph as a way to express the disposition of the blue objects in the image region. Such measures include: number of edges connecting nodes from feature types c_i and c_j, mean orientation difference between connected nodes, mean distance between connected nodes, standard deviation of distance between connected nodes, nodes/edges proportion, proportion of edges connecting nodes from feature types c_i and c_j etc.

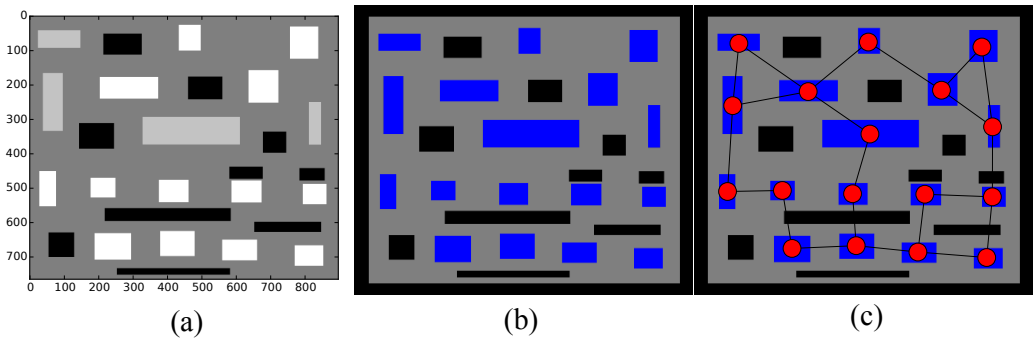

Figure 2.3: A neighborhood graph connecting features inside a hypothetical region. a) Synthetic image from a region. b) White and light grey features from the region extracted and shown in blue. c) An example of a neighborhood graph connecting the extracted features.

Although Hermosillaa et al. [2012], Walde et al. [2013] and Walde et al. [2014] have successfully used these graph metrics for classifying UST and urban land use, it seems that many possibilities of using graphs for describing the features disposition is yet to be explored. Assigning the nodes with different attributes (from the features they represent) and the edges with different weights enables the calculation of the above mentioned and many other measures, what may increase the expressiveness of the graph for describing the features spatial disposition in the image regions.

2.3 Appearance of Urban Features on InSAR Imagery

In this section it is briefly presented and discussed how buildings and vegetation, the two main features that characterize USTs, typically appear on HR InSAR imagery, that is to say on the coherence and intensity images, as well as on the interferometrically derived Digital Elevation Model (DEM) . The intention is to show to what extent and with what general strategy these two features can be detected and afterwards used for the computation of composition and configuration attributes from the urban blocks. It is also stressed in this section the importance of having intensity SAR images from the aspect angles of the ascending and descending satellite orbits for the detection of buildings and hence the block-based distinction of USTs.

The coherence image is an important type of data for urban applications. It has been exploited for settlement detection at the regional scale [Grey & Luckman, 2003], estimation of building heights [Fanelli et al., 2001] and the detection of human-induced scene changes [Milisavljevic et al., 2010]. Coherence is the complex cross-correlation coefficient from two co-registered SAR acquisitions obtained at the same looking direction. For many applications only the magnitude

of coherence is of interest, which ranges from 0 to 1. It can be computed for the whole scene by running a small window over the image pair and estimating it for each location. Besides being used to evaluate the reliability of height measures from the interferometric DEM and as a penalty function to optimise the co-registration of two SAR images [Woodhouse, 2006], the coherence measure is very relevant for land cover classification as well. For example, shadows (due to lack of signal), water bodies (due to specular reflection) and vegetated areas (due to the vertical spread of the scatters) have always very low coherence values, whereas man-made structures, due to their unchangeability between acquisitions, have high coherence values in cases the signal from these objects, due to their geometric structure, are scattered back to the sensor.

Figure 2.4 depicts a section of the urban area of Munich (Germany) as well as two of its urban blocks in optical and coherence images. The coherence image was computed from a pair of intensity images from the TS-X satellite with spatial resolution of approximately 1,1 m. As expected, wherever there is vegetation, water or shadow, the coherence values are low, expressed by darker pixels. Conversely, man-made structures like buildings and sealed areas appear in general as bright areas, corresponding to high coherence values. The yellow lines drawn in the images serve to highlight the borders between settlements and vegetated areas and river. The blue squares in Figure 2.4 (a) and (b) highlight a bridge surrounded by water and vegetation. The bridge appears as a strip of pixels with high values surrounded by dark pixels (Figure 2.4 (b)). This behaviour is also observable in Figure 2.4 (c) and (d). Nearly all man-made structures from the two blocks appear very bright, whereas the dark pixels between these structures are mostly from shadows and vegetated areas. The coherence image may therefore be used to estimate the proportion of vegetation and sealed areas, such as pavements and buildings, inside urban blocks. These proportions can be computed by choosing two optimal threshold values above and below which pixels are considered as belonging to man-made structures and vegetation/shadows respectively. This strategy has been applied for example by Heiden et al. [2012] and Voltersen et al. [2014], with the difference that they used the Normalized Difference Vegetation Index [Vina et al., 2011] image computed from optical data instead.

As discussed in Section 2.2, information about the height of the buildings is extremely valuable for the block-based distinction of USTs. From SAR imagery, height estimates can be derived through radargrammetry or interferometry [Woodhouse, 2006]. Each of these techniques offers pros and cons that should be considered in light of the application at hand and the availability of data and other resources. The main disadvantage of using radargrammetry to generate a DEM from urban areas lies in the fact that the complexity of these areas and thus the large dissimilarity of opposite-side images make their matching a hard task, specially for lower resolution spaceborne SAR data [Soergel, 2010]. Indeed, the majority of works on DEM generation from urban areas with SAR data have resorted to the technique of interferometry. However, the accuracy of the height measurements derived from InSAR data is limited by the side-looking acquisition geometry of SAR systems, which causes the undesired effects of layover and shadowing [Stilla et al., 2003]. Although these effects can be mitigated by means of the state-of-the-art techniques of Persistent Scatters [Gernhardt et al., 2010] and tomography [Schmitt et al., 2014], the large amount of data required to apply such techniques frequently becomes too expensive or is simply not available for some areas.

When the application's goals and constraints request that height information be derived based on spaceborne InSAR data, some factors must be considered in order to maximize the expressiveness of the generated DEM. The normal baseline between the acquisitions must configure a good trade-off between the height of ambiguity and the height sensitivity. Larger baselines will enable more accurate altitude measurements, increasing however the need for phase unwrapping, which is not trivial in urban areas [Stabel & Fischer, 2001]. Also, the acquisitions minimum and

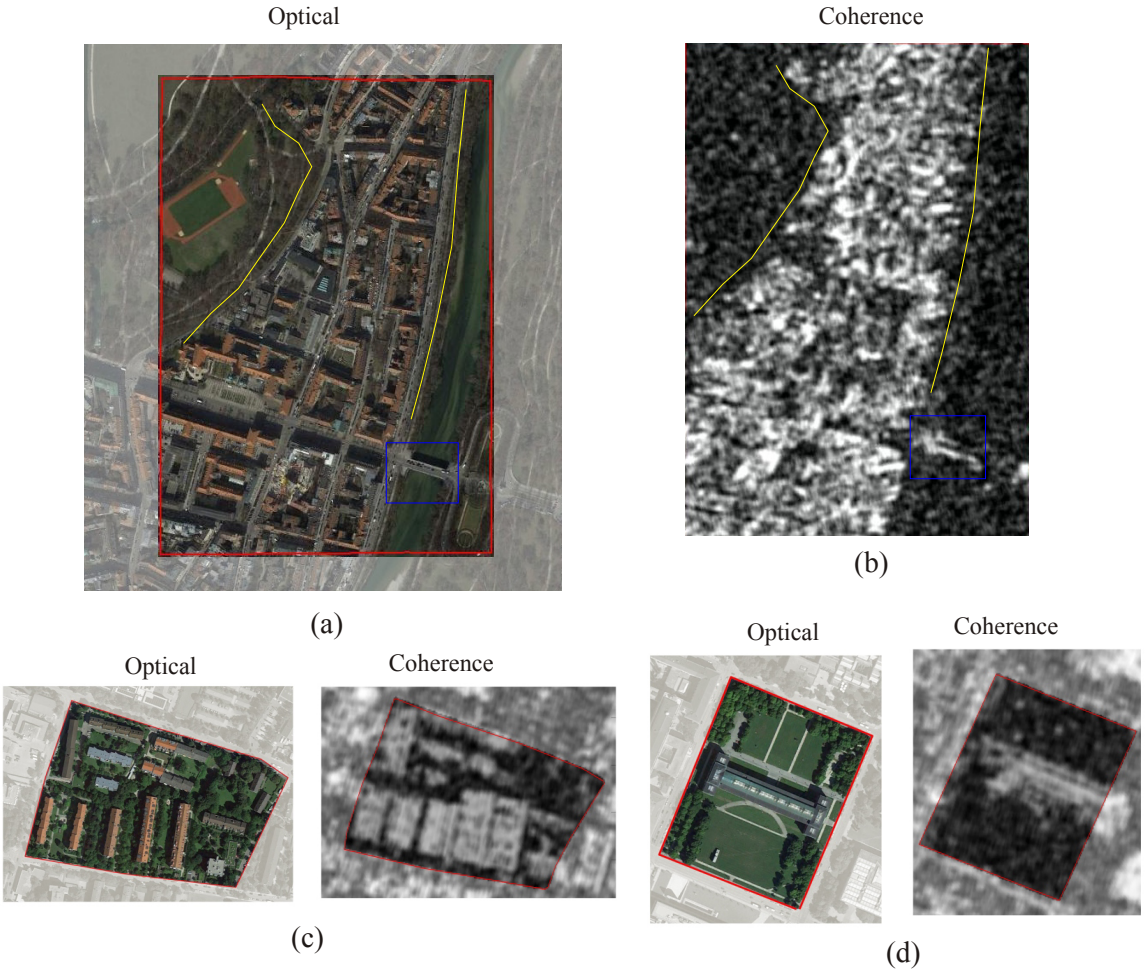

Figure 2.4: Appearance of man-made and natural features in optical and coherence images. a) Optical image from a section of Munich (Germany). b) The same section of (a) showed in the coherence image computed from HSS TS-X data with spatial resolution of approximately 1,1 m. c) and d) Two urban blocks with vegetation and man-made structures showed in optical and coherence images computed from the HSS TS-X data.

maximum incidence angles must not be such that the amount of shadows decreases at the expense of increasing too much the layover effect. Another aspect that should be considered is the possible change in the atmospheric conditions between the acquisitions. Since spaceborne InSAR data is frequently acquired at more than one orbit, one should care that the time gap is not too large and that the atmospheric conditions are therefore similar. Otherwise, the time decorrelation may increase the phase difference noise and thus damage the quality of the interferogram.

Even when all this is considered, two other aspects restrain the accuracy of building heights measurements made from spaceborne InSAR data. The first one is the still limited spatial resolution of spaceborne SAR systems. The second aspect regards the inherent complexity of most urban areas, what may hinder the distinction of the backscattered signals from, for example, two closeby buildings [Soergel, 2010]. Figure 2.5 shows a section of the city of Munich and two of its urban blocks in optical image and in a DEM generated from an interferometric pair of single-look complex (SLC) images from the TS-X satellite operating in HSS mode. The DEM has a spatial resolution of approximately 1,1 meter. It can be seen that in broader context and smaller scale, different building structures can be distinguished by an experienced interpreter (Figure 2.5 (b)), whereas when looking at single blocks individually it becomes difficult to visually interpret the

structure and disposition of the buildings. This has different reasons for the blocks shown in Figure 2.5 (c) and (d). In (c) the density and spatial complexity of the buildings distribution and their different shapes causes a complex mixture of signals together with the effects of foreshortening/layover and shadowing. In (d) the presence of vegetation in most of the block's area decrease the coherence of the phase measurements and thus the reliability of the height measurements. Any information on the built-up structure of single blocks can probably be better expressed as texture measures than as the attributes and distribution of extracted lines and flattened areas.

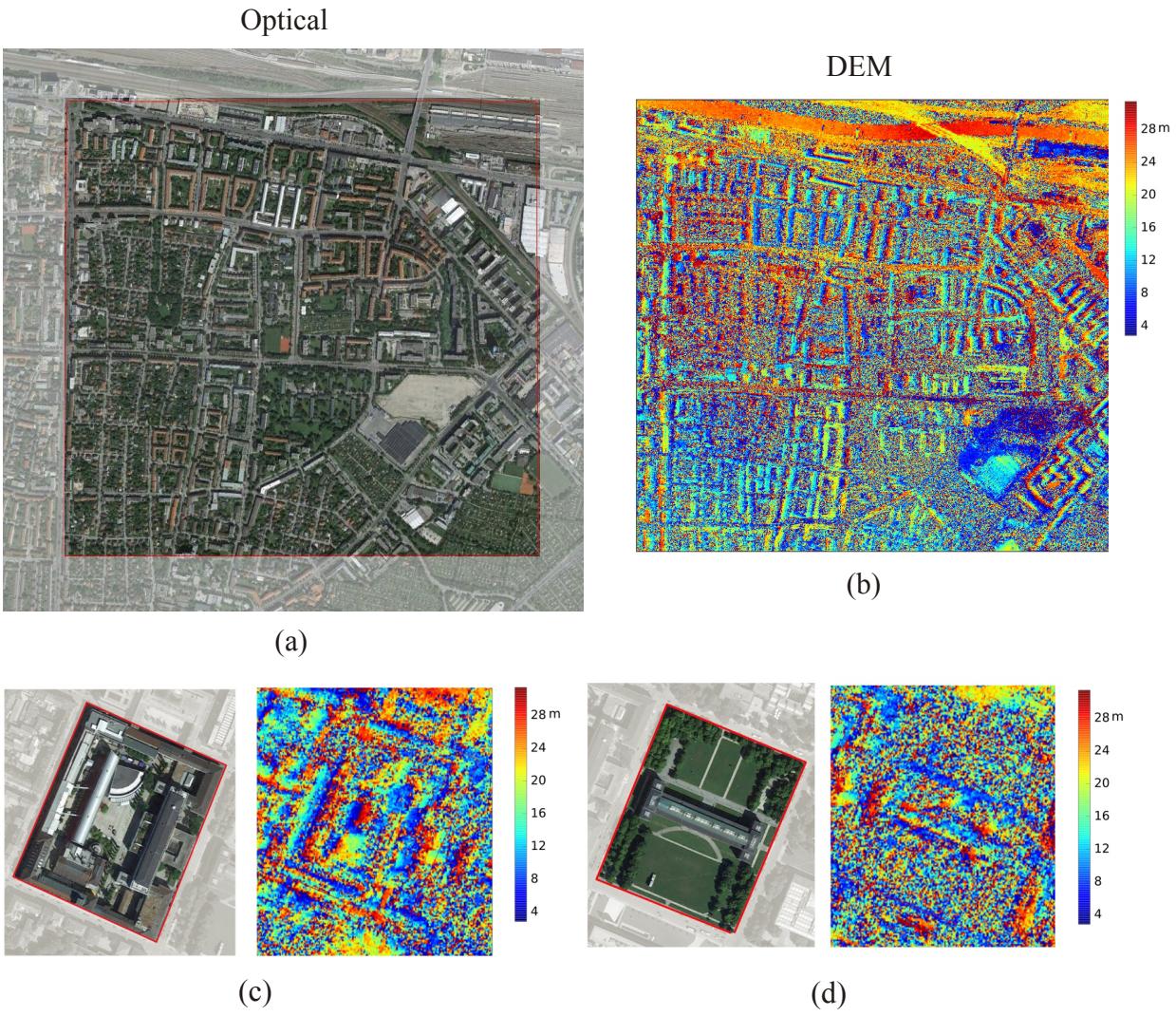

Figure 2.5: A section of Munich (Germany) and two of its urban blocks shown in optical image and as a DEM generated through interferometry with HSS TS-X data. a) Optical image from a section of Munich. b) DEM from the same area of (a) generated through interferometry. c) and d) Optical image and DEM from two individual urban blocks from Munich.

Probably the most important type of SAR data for urban area analysis is the intensity of the backscattered radar signal. It depends mainly on the shape and orientation of the scatters, as well as on their moisture content and the radar beam's incident angle. Because of volumetric absortion and diffused reflection, vegetated areas such as grass and trees usually appear dark and with moderately rough texture on intensity images. Calm water bodies will also appear dark because of specular reflection. Buildings and houses on the other hand can be identified both by very bright and very dark areas. Bright points, lines and areas are caused by the layover effect or by corner reflectors, whereas the dark areas are due to occluded areas, i.e. shadows. These effects

are inherent to the side-looking geometry of SAR systems and their intensity depends on the shape and orientation of the objects in relation to the SAR range direction [Stilla, 2007] [Soergel, 2010] [Brunner et al., 2010]. Based solely on the analysis of the corner reflection and shadow areas extracted from single HR SAR intensity images, some approaches have been proposed to detect and even reconstruct single buildings in urban areas [Tupin, 2003] [Brunner et al., 2010] [Lin et al., 2013]. Other works have extended this idea by resorting to airborne VHR SAR imagery obtained at multiple aspect angles [Michaelsen et al., 2005] [Ferro et al., 2009] [Maksymiuk & Stilla, 2012] [Schmitt et al., 2014].

Depending on the density and geometric complexity of the buildings, intensity images from spaceborne SARs may also be used to estimate to a certain extent the orientation and general geometrical characteristics of buildings [Brunner, 2009] [Ferro et al., 2011] [Cao et al., 2012a] [Cao et al., 2012b]. In this context, the chances of success increase if images of a same area are obtained at the satellite's two different looking directions. Figure 2.6 exemplifies this idea by showing two urban blocks from Munich in optical and intensity images acquired by the TS-X satellite at its ascending and descending orbits. It can be seen that buildings may appear very clearly in one image and yet be barely observable in the other. The intensity and position of the layover and shadow areas also vary a lot from one acquisition to the other. As mentioned, these effects depend on the SAR looking directions, the beam's incidence angles and the geometric properties and orientation of the buildings in relation to the SAR azimuth direction. It can be noticed, for example, that some smaller buildings from the block on the upper part of Figure 2.6 cannot be observed in any of the two SAR acquisitions. This is because their major axis are oriented more towards the range direction. On the lower part of Figure 2.6, a block is shown containing one single building. It exemplifies the fact that buildings may be detected by a bright line, caused by corner reflectors and the layover effect, or by a smooth shadow area whose shape corresponds to a certain extent the size and height from the building lying ahead of it.

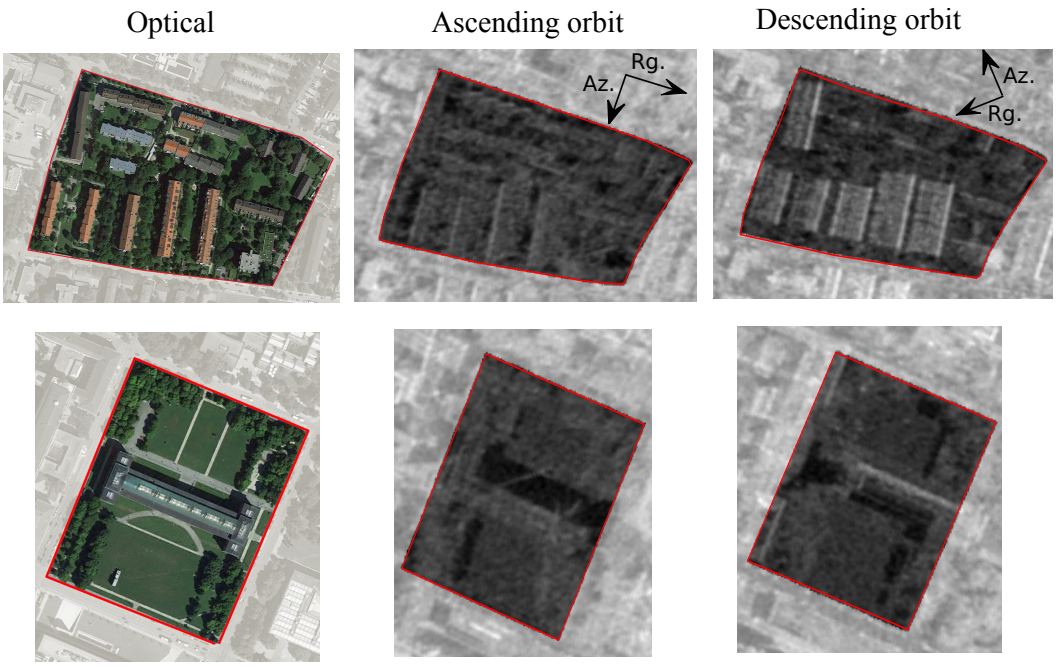

Figure 2.6: Two urban blocks imaged from the ascending and descending looking directions. Buildings may appear clearly in one intensity image and yet be barely observable in the other. The size and position of shadows and corner reflection areas also varies a lot according to the looking direction, the incidence angle, as well as the orientation and shape of the buildings.

Another important advantage of having spaceborne SAR imagery from the two looking directions is the possibility to mitigate the effect of layover in the context of detecting buildings inside urban blocks. It may happen that due to this effect a part of the backscattering signal from a building's facade gets assigned to pixels located outside the block to which it actually belongs. If this is the case in one of the acquisitions, it is likely that the layover effect from the other acquisition will cause the assignment of the backscattering signal from this part of the facade to pixels located inside the block. This is exemplified by the upper part of Figure 6.7. It shows a block from the city of Munich containing a high-rise building. Whereas part of the backscattering signal from the facade is assigned to pixels outside the block in the acquisition from the descending orbit, this effect is reversed in the acquisition from the ascending orbit. In the cases where only images from the same looking direction are available, the distinction, for example, of the commonly considered UST class High-Rise Buildings may therefore be compromised.

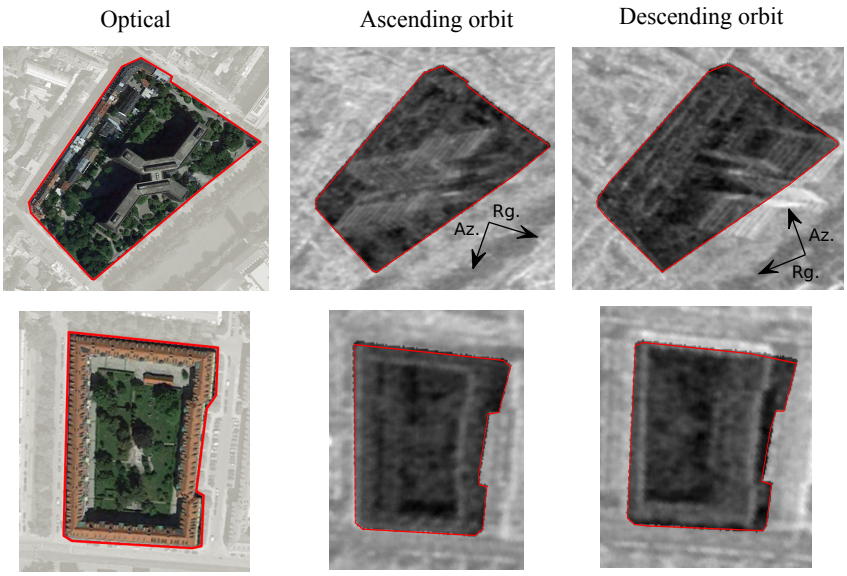

Figure 2.7: Layover effect on images from two urban blocks acquired at ascending and descending looking directions. This effect, depending on the looking direction, as well as the shape, size and position of the buildings, may cause the shifting of the signal from the facades to be assigned to pixels located outside the boundaries of the urban blocks. The availability of intensity images from both looking directions increases the chances of detecting these buildings.

In this context, layover may also hinder the detection of buildings when they are located too close to the boundaries of the blocks. This is frequently the case with UST classes Regular Block Development and specially Perimeter Block Development, as shown on the lower part of Figure 6.7. A possible strategy for identifying blocks belonging to this UST class is the detection of bright straight lines located close and parallel to the boundaries of the block. This strategy will fail though if such lines lay outside its boundaries. If, however, a second intensity image from the other looking direction is available, the risk of this happening decreases significantly. It is also interesting to notice that in neither intensity images alone the whole building is observable. Instead, corner reflection lines from the four facades can be detected, however in images from the two different acquisitions. This is then another evidence that having intensity images from both aspect angles increases the chances of detecting buildings inside urban blocks, what will presumably increase the USTs classification's accuracy in the end.

Lastly, it should be stressed that having spaceborne SAR acquisitions from both looking directions may also be advantageous for the correct distinction of USTs characterized by the presence of single-family houses. Figure 2.8 depicts two urban blocks from the city of Munich

that are oficially assigned to the UST class Single and Double House Development. As shown in the other figures of this section, the blocks were imaged from both looking directions of the TS-X satellite. It can be observed that at this specific alignement of the houses and incidence angles, the houses appear much clearer in the acquisition from the ascending orbit. Once again, the availability of only one acquisition from one of the looking directions could hamper the accurate classification of these and other similar blocks.

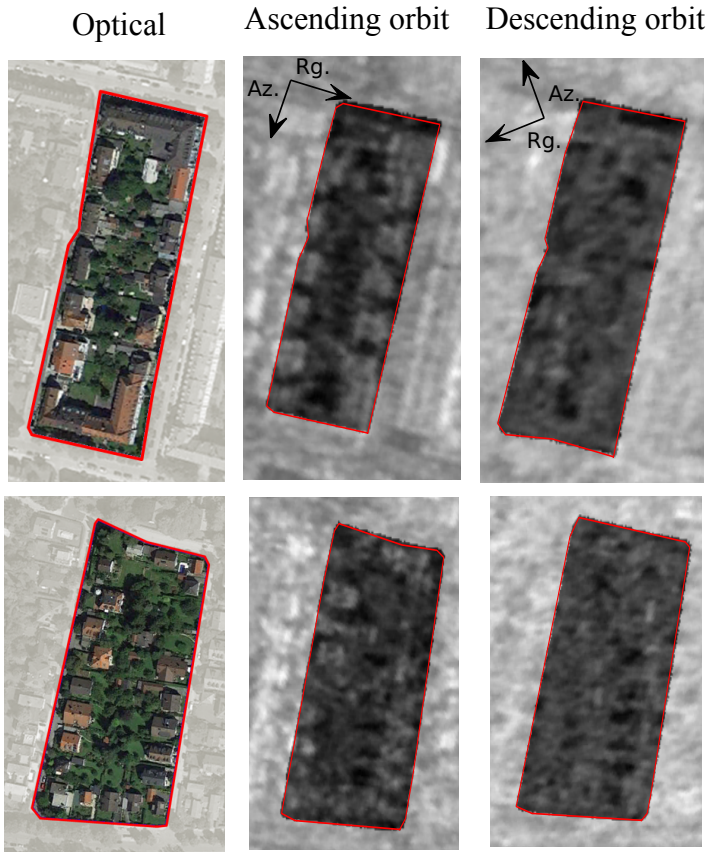

Figure 2.8: Two urban blocks with single-family houses shown in optical and intensity images obtained by the TS-X satellite at ascending and descending looking directions. The houses appear much clearer in the image from the ascending looking direction at this specific case. The availability of only one acquisition from one of the looking directions could hamper the accurate classification of these and other similar blocks.

3 Probabilistic Graphical Models

Probabilistic Graphical Models (PGMs) are a general framework for modeling systems that involve uncertainty and about which we want to reason [Koller & Friedman, 2009]. As it is frequently difficult to establish deterministic relations between variables, considering uncertainty is essential in many applications. In the context of PGMs, uncertainty is modeled using probability theory. PGMs can be of different types and be designed by a human expert or learned automatically from data. However, they all have in common the fact that they are a graph-based representation of a joint probability distribution. The graph consists of nodes and edges, where the nodes represent random variables (i.e. parts of the system) and the edges represent their probabilistic dependencies. The graph structure encondes the conditional independencies from the variables of the distribution and, therefore, how it factorizes. Figure 3.1 shows a simple PGM, the conditional independencies it encodes and how the joint distribution factorizes.

It is knowing how the joint distribution factorizes what makes reasoning on it computationally efficient. This is because usually considerably fewer parameters are needed to represent a factorized distribution than a joint one. For example, Figure 3.1 (a) shows on the left side a PGM containing five variables that may assume three values each. To represent a joint distribution of these five variables, 243 parameters would be needed (3 x 3 x 3 x 3 x 3 = 243). On the other hand, if it is known that this distribution factorizes according to the graph, only (3 x 3) + (3 x 3) + (3 x 3 x 3) + (3 x 3) = 54 parameters would be enough. The significant lower amount of parameters needed to represent a distribution with many variables allows the application of inference algorithms that work directly on the structure of the graph and are used for answering queries, like the probability of a variable's assignment given the observed value of other variables or the most probable joint assignment of all variables. PGMs also offer the advantage that the graph's structure is independent from the reasoning algorithm. Thus, editing one does not require altering the other. Furthermore, the graph usually reflects our understanding of the semantics of the problem. These advantages, combined with the fact that in many cases each variable is related to a limited set of other variables, make PGMs largely applicable.

PGMs can be categorized under two different types, namely, Bayesian networks, which are directed acyclic graphs, and Markov networks, which are undirected and possibly cyclic graphs. They differ regarding the conditional independencies they can encode and the type of interaction between two connected variables.

3.1 The Different Types of Models

Bayesian networks are acyclic graphs whose edges are directed and denote the direction of influence between two connected variables. Each variable in the graph is associated with a Conditional Probability Distribution (CPD) that defines a distribution over the values of the variable given each possible joint assignment of values from its parent-variables. Figure 3.2 shows a simple bayesian network and its CPDs.

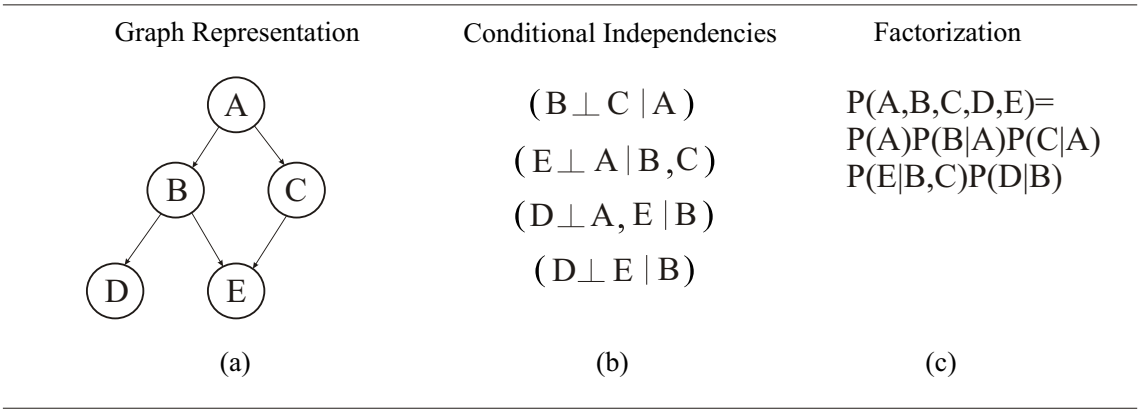

Figure 3.1: Example of a simple Probabilistic Graphical Model. a) The model's graph representation. b) The conditional independencies it encodes. c) The factorization of the joint probability distribution derived from knowing the variables conditional independencies.

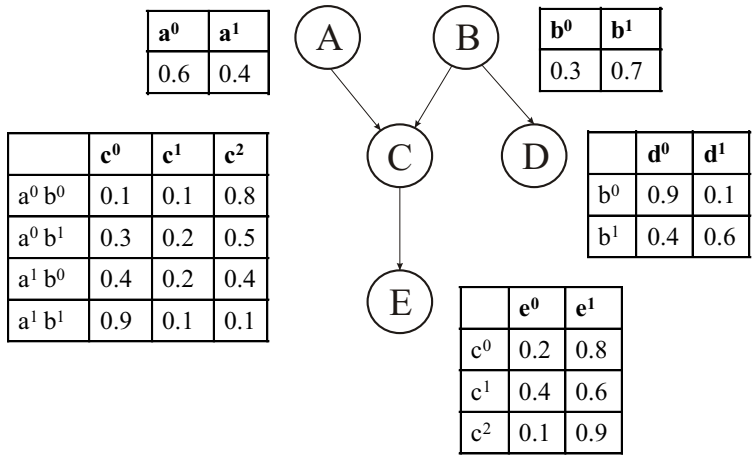

Figure 3.2: Example of a Bayesian network and its Conditional Probability Distributions.

Each factor of a Bayesian network is defined by a node X_i and its parent-nodes $Pa^G_{X_i}$. The set of conditional independencies a Bayesian network encodes can then be expressed by the following rule:

For each variable $X_i : (X_i \perp NonDescendants_{X_i} | Pa^G_{X_i})$. In words: X_i is independent from its non-descendants given its parents.

Like any PGM, the joint distribution it represents can be obtained by a product of factors. This is expressed formally by the chain rule for Bayesian networks:

$$P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i | Pa^G_{X_i}). \tag{3.1}$$

So, in the case of the network from Figure 3.2 the distribution factorizes as:

$$P(A, B, C, D, E) = P(A)P(B)P(D|B)P(C|A, B)P(E|C) \tag{3.2}$$

Bayesian networks have the advantage that when a new variable is added to the model, the CPD of the other variables do not change, what makes the contruction of such models easier.

They also can be in general more easely learned from data and admit the application of exact inference algorithms for performing queries on them [Koller & Friedman, 2009]. However, in many applications no directionalities in the influence between the variables can be defined. Furthermore, the relations of conditional independencies between the variables are frequently too complex and cannot be expressed with a directed model. In these cases, undirected models like Markov Random Fields (MRFs) can be used.

Because of undirectionality, the factors of a Markov network do not encode a CPD. Rather, they represent potentials (affinities) between the assignments of the variables in a factor. Figure 3.3 shows a simple MRF, its factors and their potentials, which can be any real positive number. As in the case of Bayesian models, the joint probability distribution of a MRF can be obtained through the product of its factors:

$$P(A, B, C, D) = \frac{1}{Z}\varphi_1(A, B)\varphi_2(B, C)\varphi_3(C, D)\varphi_4(D, A), \tag{3.3}$$

where $\varphi(X_i, X_j)$ represents the factor of variables X_i and X_j, as well as their potential function. Hence, for example, the probability of $P(a^0, b^0, c^1, d^1)$ would be calculated as:

$$P(a^0, b^0, c^1, d^1) = \frac{1}{Z}\varphi_1(a^0, b^0) * \varphi_2(b^0, c^1) * \varphi_3(c^1, d^1) * \varphi_4(a^0, d^1) = \frac{1}{Z}30 * 80 * 100 * 1, \tag{3.4}$$

where Z is the normalization term and is calculated by the summation of the product of each possible assignment of each factor of the network:

$$Z = \sum_{a,b,c,d} \varphi_1(a, b) * \varphi_1(b, c) * \varphi_1(c, d) * \varphi_1(d, a). \tag{3.5}$$

Because of prohibitive computational costs, this term cannot be calculated exactly in most real-world applications. Instead, some approximate method must frequently be used to perform any probabilty calculation on Markov networks.

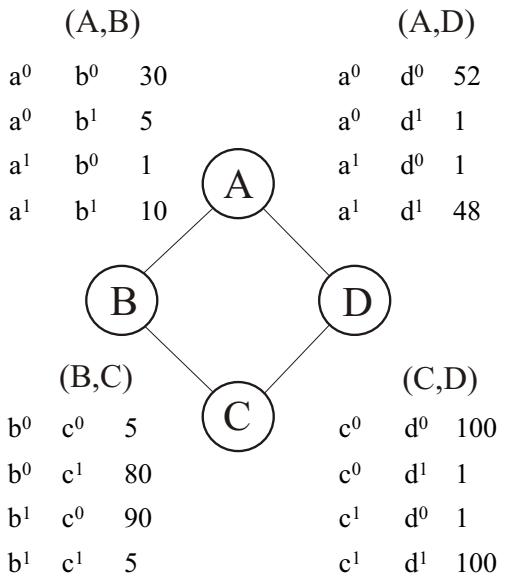

Figure 3.3: Example of a Markov network, its factors and their potentials.

MRFs are said to be more flexible than Bayesian models because they are able to capture more independency assumptions and because they enable the encoding of interdependencies between

variables. On the other hand, the parameterization of a Markov network, i.e. the definition of the factors potentials, is hardly done intuitively because they affect and are affected by the potentials from the other factors as well.

In many applications, the user is interested in performing only a specific prediction task, namely, one that concerns a specific set of variables in the network. Usually in these cases all other variables of the model are observed. Therefore, instead of $P(Y, X)$, it makes more sense to encode a conditional distribution $P(Y|X)$, where X and Y indicate the observed and unobserved variables, repectively. Since interest is only in predicting Y given X, it becomes unnecessary to model the dependencies and conditional independencies between the variables X. Hence, the model becomes discriminative regarding Y, where at the same time it avoids the modeling of probabilistic relations between X, which may be complex and not well understood. Conditional Random Fields (CRFs) are a type of undirected PGMs that encode such a conditional distribution as:

$$P(Y|X) = \frac{1}{Z(X)} \prod_{i=1}^{k} \varphi_i(D_i), \tag{3.6}$$

where D_i denotes the scope of factor φ_i, i.e. the variables it involves. In a CRF, the statement $D_i \subseteq X$ does not apply to any of the factors $\varphi_i(D_i)$. In other words, no factor will possess only observed variables. Another difference to MRFs regards the normalization term Z, which on CRFs is a function of X:

$$Z(X) = \sum_{Y} \overline{P}(X, Y), \tag{3.7}$$

where $\overline{P}$ denotes an unnormalized probability function. There will be then a different normalization term for each possible joint assignment x from X. Since X is always observed, the model is conditioned on x. This has the consequence that the model is reduced to a much simpler one. Figure 3.4 shows a simple CRF model before and after the conditioning on x. No matter how large the number of potentials from the factors $\varphi(x_i, y_i)$, what is defined by the cardinality of x_i and y_i, once conditioned on x_i these factors will be reduced to the number of potentials equal to the cardinality of y_i. The reduced factors $\varphi(x_i, y_i)$ and $\varphi(x_j, y_j)$ are then multiplied with factor $\varphi(y_i, y_j)$, so that their information is integrated in the reduced model. This multiplication has the effect of a reparameterization of the factors $\varphi(y_i, y_j)$. The operations of factor reduction and factor product are explained in the next section.

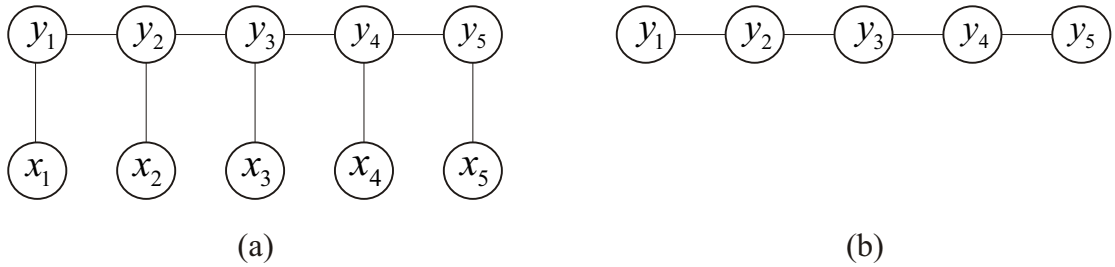

Figure 3.4: Conditioning of a CRF model to the observed data x. a) Structure of the model before conditioning. b) Structure of the model after conditioning.

The flexibility and powerfulness of CRFs come from the fact that it is possible to insert a large number of observed variables in the model according to domain knowledge without worrying about their probabilistic dependencies. Nonetheless, even after conditioning the model on x, the distribution over Y (i.e. over the factors $D_i \subseteq Y$) usually remains complex enough to make exact inference intractable. Because of that, a method for approximate inference must be used in order to perform probabilistic queries in the model.

3.2 Operations with Factors

As just mentioned, probability queries on MRFs and CRFs cannot be performed with exact inference methods. This would require the computation of the partition function Z, which is intractable at pratically any real-world application. Hence, some of the algorithms for approximate inference on PGMs [Andres et al., 2010] [Kappes et al., 2013] must be used. These algorithms are based on four operations performed on the model's factors, namely: product, marginalization, maximization and reduction.

Factor product, as any of the above operations, can only be performed between two factors whose scopes have at least one variable in common. So if $\varphi_i(S_i)$ and $\varphi_j(S_j)$ define two factors whose scopes are variables S_i and S_j respectively, then factor product is only possible if $S_i \cap S_j \neq \emptyset$. The scope of the resulting factor will be $S_i \cup S_j$. For example, considering the network shown at Figure 3.3, a possible factor product operation would be $\varphi_1(A, B) \times \varphi_2(B, C)$. This operation is shown in Figure 3.5 (a).

Factor marginalization and factor maximization are the main operations involved on inference on PGMs. Marginalization is performed when the potentials of a subset of variables from a factor must be computed. Considering the factor $\varphi(A, B, C)$ resulting from the product of factors $\varphi(A, B)$ and $\varphi(B, C)$, marginalizing out variable A can be expressed as $\sum_A \varphi(B, C)$. Figure 3.5(b) shows this operation. In order to marginalize A, the entries from the table on the left side of Figure 3.5(b) whose values of B and C match should be summed. Factor maximization, or max-marginalization, is a variant of marginalization in which instead of summing the values from the matching assignments of the variable to be max-marginalized, only the largest one is passed to the resulting factor. Therefore, the max-marginalization of variable A from factor $\varphi(A, B, C)$ is defined as $\max_A \varphi(B, C)$. Figure 3.5(b) shows this operation. The updated potentials from the resulting factor are shown between parenthesis.

The most simple operation on factors is reduction. It is applied anytime a variable in the model is observed. Then all the factors containing that variable are reduced. When a factor is reduced, all its entries that do not match a specific assignment from one or more variables in this factor are eliminated. Figure 3.5(c) shows the reduction of factor $\varphi(A, B, C)$ to the case where $C = c^1$.

3.3 MAP - Maximum A Posteriori Estimation

Image classification with PGMs consists in approximately estimating the *maximum a posteriori* (MAP) probability of Y given the observed x ($\arg\max_Y P(Y|x)$). That is, estimating the assignment of Y given x for which $P(Y|x)$ is the highest. One of the most commomnly used methods for inferring this MAP assignment is called Loopy Belief Propagation (LBP) [Kschischang et al., 2001]. The LBP algorithm is based on a so-called *cyclic cluster graph*. Firstly, it is described how a cyclic cluster graph is constructed and what properties it must have. Then, it is explained how the loopy message passing procedure enables the estimation of the MAP assignment.

If ξ is a PGM over variables V and with n factors φ, then a cluster graph Ψ is an undirected graph where each node is called a cluster C and $C_i \subseteq V$ for each C_i. Each edge between two clusters C_i and C_j is associated with the sepset $S_{i,j}$, such that $S_{i,j} \subseteq C_i \cap C_j$. A cluster graph must have two properties:

- **Family Preservation**: each factor φ_k from ξ is associated to one cluster C_i in Ψ such that $Scope[\varphi_k] \subseteq C_i$. More than one factor φ_k may be associated to C_i. In this case, these two

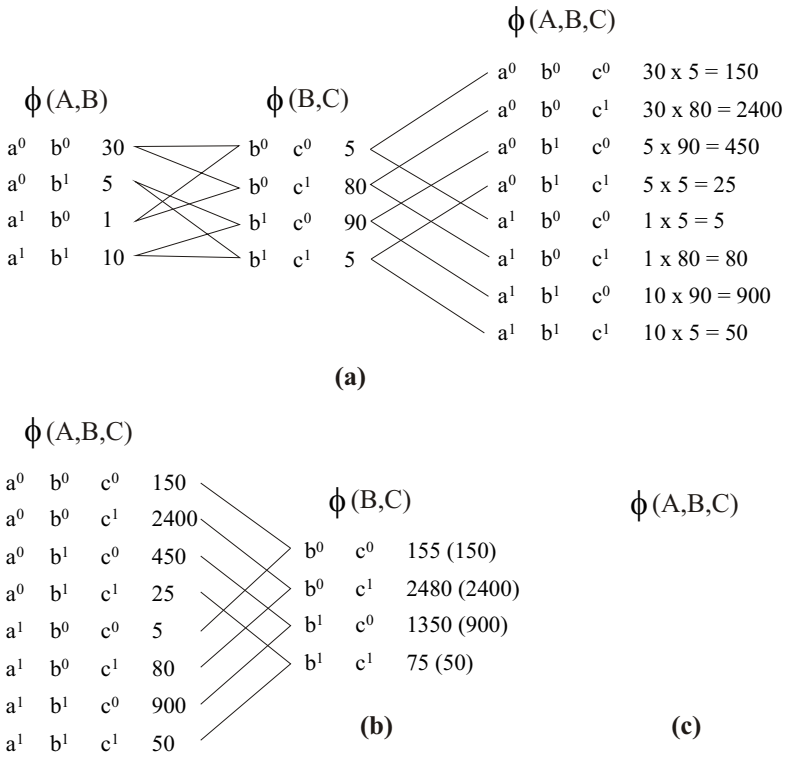

Figure 3.5: Examples of factor operations. a) Factor product. b) Factor marginalization and factor maximization (values showed in parenthesis). c) Factor reduction.

or more factors are multiplied, which implies that they must have at least one variable in common.

- **Running Intersection**: a cluster graph Ψ has the running intersection property if every variable $x \in C_i$ and $x \in C_j$ is present in every cluster and subset in the unique path between C_i and C_j. It also assumes that such path exists.

The family preservation property guarantees that each factor of ξ is present in exactly one cluster of Ψ. The running intersection property guarantees that there is no self-reinforcing loops regarding the beliefs from any variable in Ψ. Figure 3.6 shows a hypothetical PGM (a), an illegal cluster graph constructed out of it (b) and a legal one (c). In (b) there is a self-reinforcing loop between clusters 1, 2 and 4 concerning variable B. The beliefs regarding this variable are self-confirmed because they influence the beliefs about B from the same clusters it now receives information about it. This problem is solved in the cluster graph depicted in (c). It respects the family preservation for the graph in (a) and it has the running intersection property concerning all its variables.

Fortunately, there is a generic type of cluster graph that guarantees these two properties and that can be easily constructed independently from the complexity of ξ. These cluster graphs are called Bethe graphs [Koller & Friedman, 2009] and they have two layers of clusters. One layer contains exactly one cluster C_k for each factor φ_k of ξ such that $Scope[\varphi_k] = Scope[C_k]$. This guarantees family preservation. The second layer contains univariate clusters with each variable $x_i \in \xi$. An edge is then defined between the clusters C_k of the first layer and the univariate clusters of the second layer whose variables $x_i \in C_k$. Figure 3.7 shows the Bethe graph for the undirected PGM shown in Figure 3.6 (a).

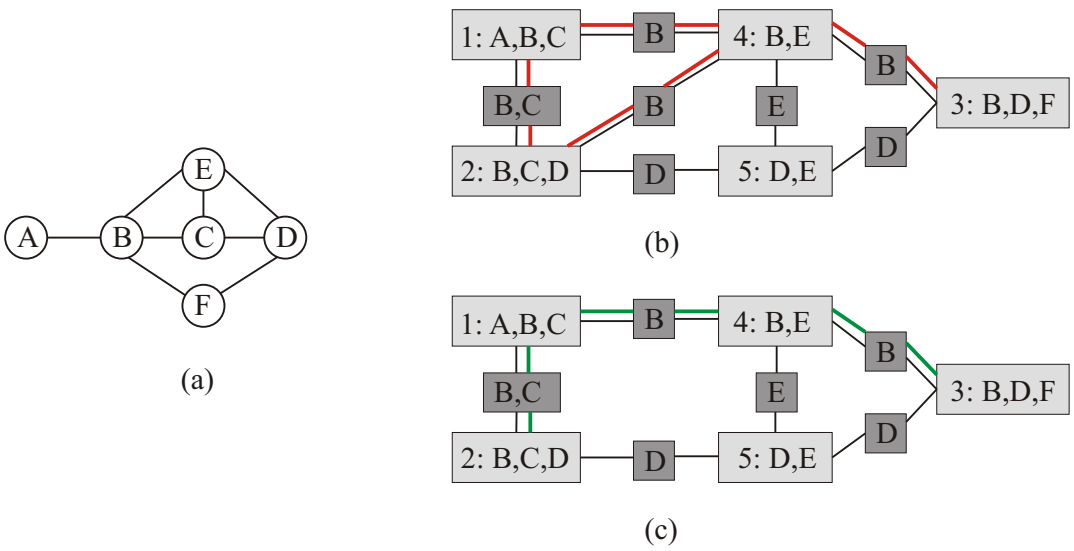

Figure 3.6: An undirected model and one example of an ilegal and of a legal cluster graph. a) Hypothetical Markov network. b) Cluster graph disrespecting the running intersection property for variable B. c) Example of a valid cluster graph for the respective Markov model.

The LBP algorithm transmits messages between every pair of connected clusters in the cluster graph. Messages have a direction. Hence a cluster graph with ten edges has twenty messages: one going from each cluster C_i to C_j and one going for each cluster C_j to C_i. The messages update the so-called *beliefs* about the variable(s) in the sepset from the two connected clusters C_i and C_j. The update of the beliefs is computed in the following way. Let ψ_i be the product of all factors from ξ associated to cluster C_i. For example, cluster C_1 from Figure 3.6 (c) is defined by the product of factors $\varphi(A, B)$ and $\varphi(B, C)$ from Figure 3.6 (a). The message from cluster C_k to C_i is defined as $\delta_{k \to i}$. The beliefs of subset $S_{i,j}$ at loop τ_n are defined as:

$$\tau_n(S_{i,j}) = \max_{C_i - S_{i,j}\psi(C_i)}, \; where \; \psi(C_i) = \psi_i . \prod_{k \in (Nb_i - \{j\})} \delta_{k \to i} \tag{3.8}$$

Each loop τ_n is completed when exactly one message is sent over each edge in each direction. This implies the updating of all beliefs. Since each cluster can only send a message after it has received the messages from all its neighbors except the one to which it is transmiting the message, the algorithm is initialized by setting an unit function for every sepset. This allows the algorithm to start with any message.

Strictly speaking, LBP will have converged when for each sepset $S_{i,j}$ the following condition holds:

$$\tau_{n-1}(S_{i,j}) = \tau_n(S_{i,j}) \tag{3.9}$$

This implies that every message in loop n is the same as in the previous loop. In reality, this may never occur, so the user is obligated to set a limited number of loops and/or apply a smoothing factor to the update of the messages:

$$\delta^n_{i \to j} = \lambda(\delta^n_{i \to j}) + (1 - \lambda)\delta^{n-1}_{i \to j} \tag{3.10}$$

When $\lambda = 1$, no smoothing takes place and when $0 < \lambda < 1$ the updated message gets averaged to its older one to the extent of $1 - \lambda$.

Once convergence is obtained, the approximated max-marginals of all variables have to be decoded in order to identify the most probable joint assignment. This is done by first multiplying

each cluster with all its incoming messages. Following, the most probable assignment at each cluster C_i must be consistent with the most probable assignment of all other clusters that contain any of the variables from $\psi_i(C_i)$. Figure 3.8 shows the pseudo-code of the Max-Product Belief Propagation inference algorithm, which is one variant of LBP. The other variant, called Max-Sum Belief Propagation, differs in the fact that product is substituted by summation and the factor potentials are previously transformed to logarithm form.

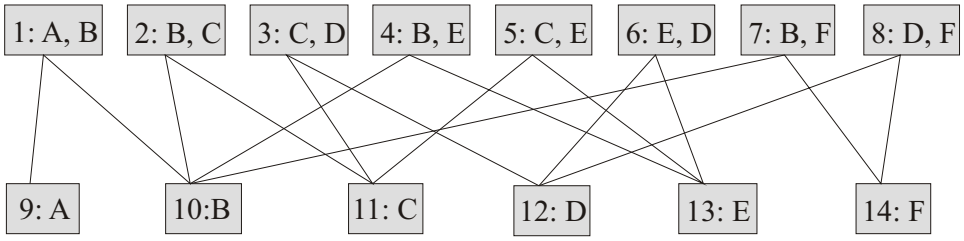

Figure 3.7: Bethe graph for the undirected graph of Figure 3.6 (a).

Procedure *Initialize Bethe Graph*;
Clusters are equal to the factors of the graph;
for *each cluster* C_i **do**
  $\beta_i \leftarrow \prod_{\varphi \,:\, a(\varphi)=i} \varphi$;
end
Initialize all values of all messages to 1;
for *each edge* $(i-j) \in \xi$ **do**
  $\delta_{i \to j} \leftarrow 1$;
  $\delta_{j \to i} \leftarrow 1$;
end

Procedure *Run Massage Passing Belief Propagation*;
while *convergence is not achieved* **do**
  Select $(i-j) \in \xi$;
  $\delta_{i-j}(S_{i,j}) \leftarrow$ **Procedure** *Max-Product Message*(i,j) ;
end

Procedure *Max-Product Message*;
i = sending clique;
j = receiving clique;
$\psi(C_i) \leftarrow \psi_i . \prod_{k \in (Nb_i - \{j\})} \delta_{k \to i}$;
$\tau(S_{i,j}) \leftarrow \max_{C_i - S_{i,j}} \psi(C_i)$;

Procedure *Compute Final Beliefs*;
for *each clique* i **do**
  $\beta_i \leftarrow \psi_i . \prod_{k \in Nb_i} \delta_{k \to i}$;
end

Figure 3.8: Pseudocode from the Max-Product Belief Propagation algorithm.

3.4 Remote Sensing Image Classification with Probabilistic Graphical Models

The apporaches for the classification of remote sensing images with PGMs can be categorized into two general types, namely, standard site-based and context-based classification. On site-based approaches, the class assignment of an image site, which is usually a single pixel or an image segment, is not influenced by the class assignments of its neighboring sites. Instead, only the variables of the site itself, which might eventually express some characteristics of its context, can influence its class assignment. The most frequently used type of PGM in these cases are Bayesian networks, which have been applied, for example, for land cover classification [Dlamini, 2011] [Mukashema et al., 2014], land use classification [Ropero et al., 2015] and the estimation of geo/biophysical parameters [Qu et al., 2008]. On the other hand, context-based classification with PGMs refer to the cases where there is mutual influence between neighboring image sites regarding their class assignments. In other words, there is probabilistic dependence between their unobserved variable *class*. In these cases, an undirected PGM must be used, for no directional causality can be established between the variable *class* of neighboring image sites.

To elaborate on how PGMs can be used for context-based image classification, some basic notation must first be defined. We denote i and j as two neighboring image sites and s_i and c_i respectively as the observed attributes and the unknown class label from i. The aim is to predict the set of labels $C = \{c_i, c_j, ..., c_n\}$ given the observed image data set $I = \{s_i, s_j, ..., s_n\}$ by means of a factorized probabilistic distribution. The factorization of such a distribution can be, though not always completely, visualized and represented as a graph $G = \{V, N\}$ defined by a set of nodes $N = \{n_i^s, n_i^c\}$ and vertices $V = \{v_{ii}^{sc}, v_{ij}^{cc}, v_{ij}^{cs}, v_{ij}^{ss}\}$. Although in theory G may have any possible structure (i.e. it may contain even all of the four types of vertices in it), the most commonly used structures for image filtering, segmentation and classification are the ones depicted in Figure 3.9 (a) and (b). In these graphs, each image site i has one node n_i^c (shown in pink) and at least one n_i^s (shown in blue). The latter represents an observed attribute from i, the former its unknown class label. Also, these graphs have only factors with two variables each connected by vertices of two types, namely v_{ii}^{sc} and v_{ij}^{cc}.

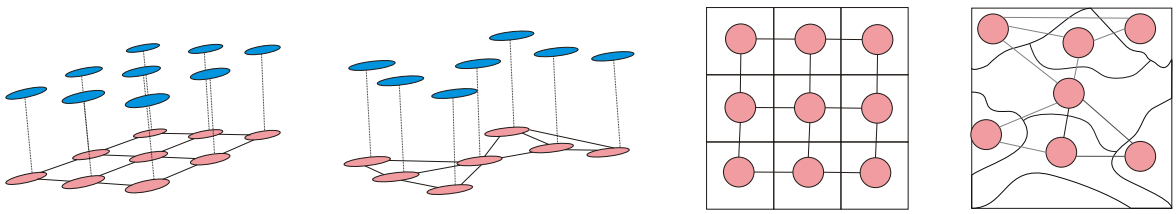

Figure 3.9: Commonly used structures of undirected graphs in image analysis applications. Blue and pink nodes represent observed and unobserved variables respectively. a) Regular grid structure used for the case when image sites are single pixels. b) Irregular grid, or region adjacency graph, used for the case when image sites are image segments. c) and d) Reduced Markov networks from (a) and (b) respectively after conditioning it on the observed variables. The nodes of the reduced graphs are depicted over the image sites they represent.

Formally speaking, the graphs in Figure 3.9 (a) and (b) represent MRFs models, i.e. they express the factorization of distributions of the type $P(C, I)$. These models can be formulated as the following product of factors:

$$P(C, I) = \frac{1}{Z} \prod_{c_i \in C} \varphi(c_i, s_i) \prod_{v_{i,j} \in V} \psi(c_i, c_j). \tag{3.11}$$

The terms $\varphi(c_i, s_i)$ are called *association* factors and define a function $f(s^i, c^i) \rightarrow p'$, where p' is the local class potential $p'(c_i|s_i)$. In other words, it defines the potentials of each possible k label c_i^k given the observed data at site i. The terms $\psi(c_i, c_j)$ are called *interaction* factors and define the potentials of each combination of values (c_i^k, c_j^k) from two neighboring sites i and j. Z is the normalization term. In the above models, the number of association factors is equal to the sum of neighboring sites j from each site i. Two sites can be considered neighbors based on different criteria, such as the euclidean distance or the adjacency between them. It should be noticed that a graph G will fail to represent the factorization of a distribution if the scope of at least one factor has more than two variables. Consider for instance the graph from Figure 3.9(b). It is impossible to know whether the distribution factorizes with seven association factors and ten pairwise interaction factors, or instead with seven association factors and four third-order factors, i.e. factors whose scopes have three random variables each. In this last case, the interaction factors would be of higher-order (generic name given when the model has factors with more than two variables) and equation 3.11 would be altered to be as follows:

$$P(C, I) = \frac{1}{Z} \prod_{c_i \in C} \varphi(c_i, s_i) \prod_{v_{i,j} \in V} \psi(c_i, c_j, c_m), \tag{3.12}$$

where sites i, j and m are considered to be mutual neighbors according to some criterium. Examples of image segmentation and labelling using higher-order PGMs can be found in Kohli et al. [2009] and Wegner et al. [2013].

Whether higher-order or not, the great flexibility of such models lies in the fact that there is complete freedom in defining the association and interaction potentials. In fact, they can be defined (1) by some function, (2) empirically by the user or (3) learned through supervised convex optimization [Lee et al., 2007] [Ganapathi et al., 2008] [Koller & Friedman, 2009]. Defining the potentials empirically is subjective and may become overly time-consuming. Learning them through optimization requires a large amount of training data [Koller & Friedman, 2009], specially when the cardinalities of the variables are high. Besides, experiments shown in Novack & Stilla [2014b] were not encouraging for the same application of this thesis. The option of defining the potentials by a function is on the other hand a convenient one. A standard procedure for defining the association potentials has been the use of powerfull classifiers to compute $p'(c_i|s_i)$, such as Support Vector Machines [Schnitzspan et al., 2008] [Zhong et al., 2014] or Random Forest [Niemeyer et al., 2013] [Novack & Stilla, 2015]. These powerfull classifiers can deal with many observed attributes from each i, in spite of those being continuos or sometimes inexpressive or correlated with each other. This makes the association potentials already very discriminative, hence increasing the chances of obtaining better results. The interaction potentials in their turn can be defined without much effort based on empirical knowledge or on collected samples. This is because the number of (c_i^k, c_j^k) combinations is only of $\frac{n(n+1)}{2}$, if we assume that the cardinality of c_i is of n for all sites.

Since the sites attributes I are always observed, the association factors are always reduced to n potentials, i.e. one for each n possible outcomes of c_i. After multiplying each of the reduced factors $\varphi(c_i, s_i)$ and $\varphi(c_j, s_i)$ to the respective $\psi(c_i, c_j)$, the graph becomes conditioned on I and the interaction factors get re-parameterized. Accordingly, the graphs of Figure 3.9 (a) and (b)

after being conditioned on I and reduced look like the ones depicted in Figure 3.9 (c) and (d). It can be argued that such models actually enconde a conditional distribution $P(C|I)$, since I is always observed and G always reduced to this observation. This would characterize these models as CRFs, given that Z becomes a function of I. In fact, a CRF is a special case of a MRF. Koller & Friedman [2009] defines a CRF as a undirected model which in all of its factors there is at least one unobserved variable. Other authors though prefere to define CRFs as when the interaction potential functions are a function of the observed variables as well [Kumar & Hebert, 2006] [Schnitzspan et al., 2008], as for example in the following model:

$$P(C, I) = \frac{1}{Z} \prod_{c_i \in C} \varphi(c_i, s_i) \prod_{v_{i,j} \in V} \psi(c_i, c_j, s_i, s_j). \tag{3.13}$$

Frequently, the image analysis problem at hand requires performing a specific prediction task with a model whose most variables are observed. In this context, one is not interested in estimating actual probabilites, but instead the outcome of C with the highest likelihood. Therefore, the partition function Z can be ignored. Also, because summations are better to deals with than multiplications, undirected models such as the MRF and the CRF shown in equations 3.11 and 3.13 can be reformulated to their corresponding energy functions:

$$\text{MRF: } E(C) = \sum_{i \in S} \varphi(s_i, c_i) + \sum_{i \in S} \sum_{j \in N_i} \psi(c_i, c_j), \tag{3.14}$$

$$\text{CRF: } E(C) = \sum_{i \in S} \varphi(s_i, c_i) + \sum_{i \in S} \sum_{j \in N_i} \psi(c_i, c_j, s_i, s_j). \tag{3.15}$$

Formulated as an energy function, the association and interaction potential functions from a model are transformed to their logarithm form. Consequently, the MAP configuration of C can be obtained by minimizing $E(C)$, since

$$E(C) \propto -\log P(C|I) \tag{3.16}$$

and therefore

$$C_{MAP} = \arg\max_C P(C|I) = \arg\min_C E(C). \tag{3.17}$$

This minimization can be performed without a problem, for example, with the LBP algorithm presented in Section 3.3. The only changes required are then the substitution of the multiplications by summations and the transforming of the association and interaction potentials to their negative logarithm form.

A convenient and common strategy for defining the potentials of the interaction factors of the MRF from equation 3.14 is through a Potts model:

$$\psi(c_i, c_j) = \begin{cases} 0 \text{ if } c_i = c_j \\ 1 \text{ if } c_i \neq c_j \end{cases}. \tag{3.18}$$

The Potts model forces any two neighboring sites indiscriminately to assume the same class label, creating therefore the effect of smoothing the spatial distribution of C by penalizing the energy everytime two neighbors should get different classes assigned to them. Obviuosly, this is compensated and regulated by the association potentials internally to the optimization process for inferring the C_MAP. In the context of CRFs, the potentials of the interaction factors may be defined on the other hand by a less brute and more flexible approach called generically *contrast-sensitive* Potts models [Hao et al., 2014]. These models have, for example, the form:

$$\psi(s_i, s_j, c_i, c_j) = \begin{cases} 0 \text{ if } c_i = c_j \\ -\ln\left(\|\mu_i - \mu_j\|_2\right) \text{ if } c_i \neq c_j \end{cases}, \tag{3.19}$$

where μ_i is the feature vector of site i with each feature m rescaled to the unit length as follows:

$$\mu_i^{m\prime} = \frac{\mu_i^m - \min(\mu^m)}{\max(\mu^m) - \min(\mu^m)}. \tag{3.20}$$

The potential function shown in equation 3.19 has the effect of penalizing the energy if the class of neighboring sites are different in proportion to the similarity of their feature vectors. In other words, the more similar their feature vectors are, the stronger the penalty will be. It is evident that many other criteria for penalizing the energy may be introduced as a contrast-sensitive Potts model. Whether as a MRF or a CRF, the interaction factors can be multiplied by a constant as a way to adjust the influence and consequently the degree of smoothing caused by the interaction potentials.

Another interesting property of undirected PGMs is that the model does not need to be limited to the association and interaction terms shown in equations 3.14 and 3.15. Other potential functions can be inserted as additional terms in the energy function. Each additional term is just another group of factors, each with at least one unobserved variable in it. A model for performing contextual and multitemporal classification may be defined, for example, by inserting a *temporal interaction* term to equation 3.14:

$$E(C,S) = \sum_{i\in S} \varphi(c_i, s_i) + \sum_{i\in S}\sum_{j\in N_i} \phi(c_i, c_j) + \sum_{i\in S}\sum_{k\in E_t}\sum_{l\in L_i^k} \psi(s_i^t, s_l^k, c_i^t, c_l^k), \tag{3.21}$$

where t and k are two neighboring epochs. $E_t = \{t-1, t+1\}$ is the set of temporal neighbors from epoch t. L_i^k is the set of image sites l that occupy the same spatial area as i but in epoch k instead of i. The temporal interaction potential models the dependency between the class labels and the observed data at consecutive epochs. Its potential function can be based on a transition matrix in which the conditional probabilities $P(c_i^t|c_l^k)$ for all $n * n$ label combinations are defined. Examples of such similar models can be found in Bruzzone et al. [2004], Feitosa et al. [2009] and Hoberg et al. [2015]. Besides multi-temporal classification, the tasks of multiscale [Su et al., 2011] and hierarchical [Alioscha-Perez & Sahli, 2014] classification of remote sensing imagery have also been performed with undirected PGMs with additional terms. Indeed, this possibility together with the freedom in defining the potential functions and the eventual consideration of higher-order factors is what makes PGMs highly attractive for the contextual classification of different remote sensing imagery in different applications.

4 Methodology

This chapter describes the methodological steps employed in this thesis. They are interconnected according to the flowchart shown in Figure 4.1. The paralelograms represent input data, the cubes represent processes and the rectangles represent the outcome from the processes. The concept *feature* refers to entities extracted from the images, as for example lines and polygons. The concept *attribute* refers to a variable that assumes a numerical value at each different instance, for example an urban block's area.

The following sections describe the main processes depicted in Figure 4.1 as well as other related steps. These main processes are: (1) extraction of urban blocks, (2) extraction of features inside the blocks, (3) computation of attributes from the blocks based on their features, (4) USTs classification and (5) USTs contextual classification.

4.1 Extraction of Urban Blocks

In this thesis, the image regions considered as analysis elements are urban blocks. These were extracted by firstly defining the study area as the intersection between the SLC image pairs. Following, the streets, rivers and railroads network was overlaid onto the study area. The blocks were then extracted as the closed regions whose borders coincide with the network. Blocks that were not entirely located inside the study area were excluded from the subsequent analysis.

4.2 Extraction of Features

One of the most important methodological steps of this thesis is the extraction of features inside the urban blocks. It is the geometrical properties and the spatial disposition of the features what gives the most expressive evidences about the built-up structure of the blocks. Therefore, a lot of efforts were put on the development of effective strategies for extracting meaningful features from inside the blocks. The next section describes the general feature extraction and analysis strategy. The subsequent ones describe the specific procedures employed for the extraction of the different types of features.

4.2.1 Extraction and Analysis Strategy

The streets, rivers and railroads network was converted from the WGS-84 coordinate system to the acquisition coordinate system, i.e. to the range and azimuth coordinates of the SAR image. In this system, all pixels in one of the image's columns correspond to the same SAR incidence angle, as all pixels in one line correspond to the same SAR azimuth instant. Figure 4.2 shows the intensity image and the boundaries of an urban block in WGS-84 and acquisition coordinates. From the perspective of feature extraction strategies, it is preferable to keep the images in acquisition geometry and know exactly the range and azimuth directions. In Figure 4.2 (a), for example, it can be seen that bright corner reflection lines will most probably appear

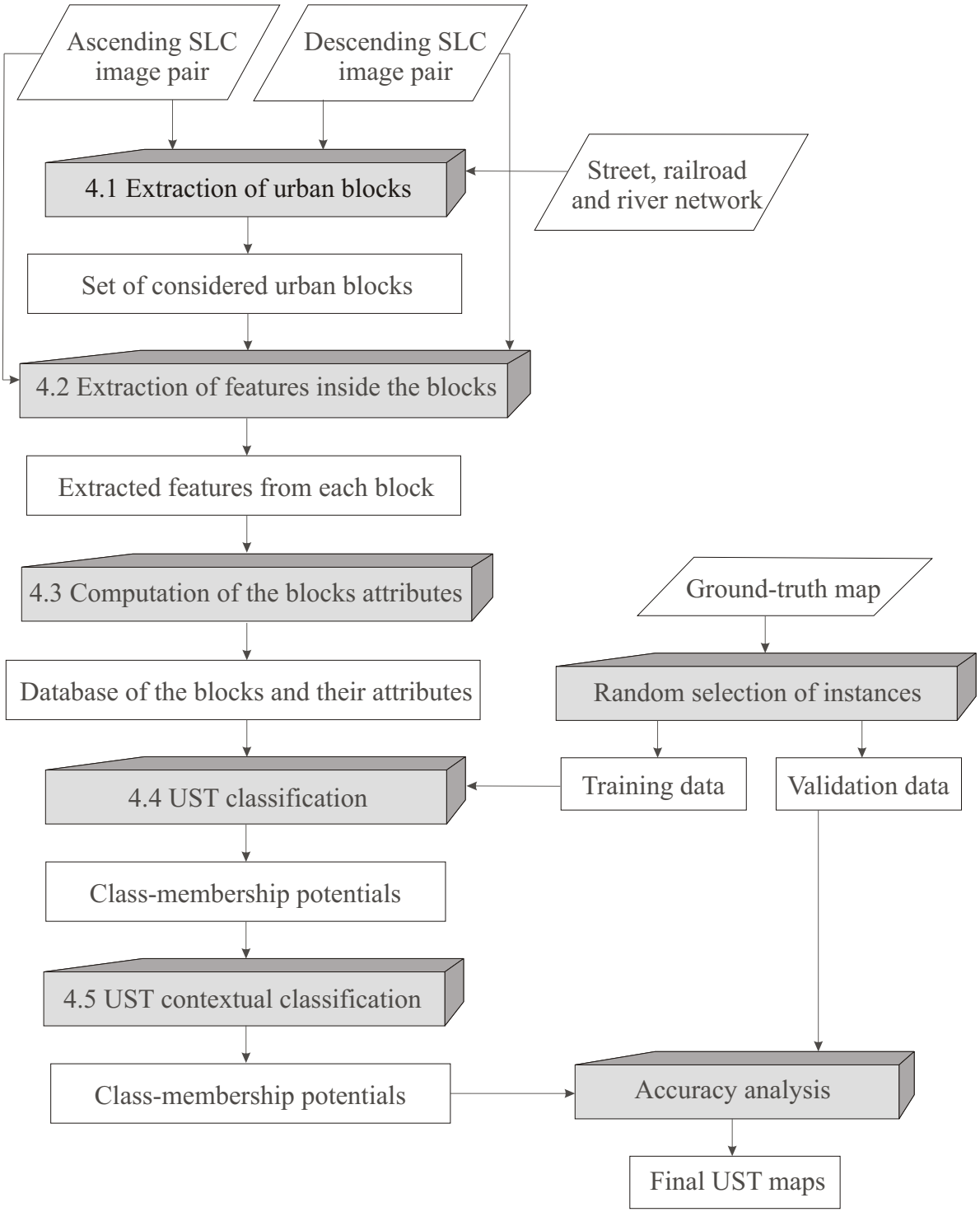

Figure 4.1: Methodological flowchart adopted in this thesis. The paralelograms represent input data, the cubes represent processes and the rectangles represent the outcome from the processes. The sections of this chapter in which the process are explained are indicated in the cubes.

perpendicular to the range direction. This facilitated the extraction of lines and increased the expressiveness of texture measures, what will be discussed in the respective sections.

In order to correctly calculate the geometric properties and position of the extracted features (e.g. length, area, orientation, distance to closest block boundary etc.), their coordinates, as well as those from their respective urban blocks, were reprojected from acquisition to WGS-84 coordinates and then from these to the local Gauss-Kruger (GK) - Zone 12 coordinates. The transformation from WGS-84 to GK-Zone 12 was performed by undertaking the steps and considering the parameters indicated by Wasmeier [2006].

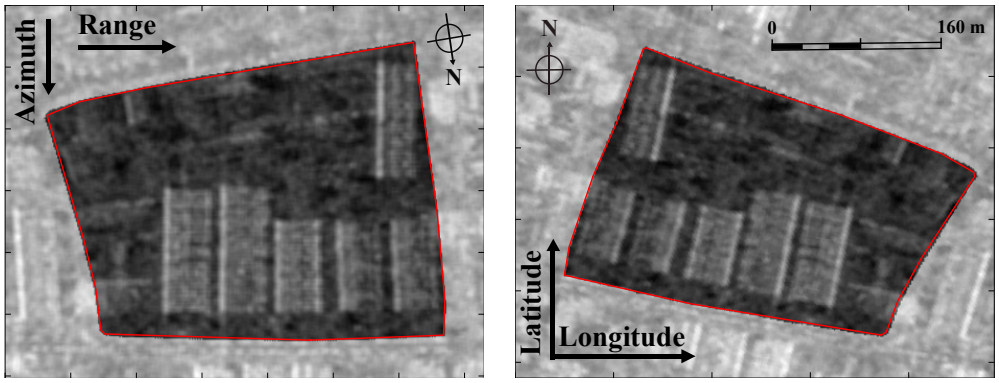

Figure 4.2: SAR intensity image and boundaries of an urban block shown in two different coordinate systems. a) Intensity image and boundaries of a block shown in acquisition geometry. b) Same intensity image and block boundaries projected to the WGS-84 coordinate system. Keeping the images in acquisition geometry facilitated the extraction of lines and increased the expressiveness of texture measures, as it will be discussed.

4.2.2 Extraction of Dark and Bright Areas

In SAR coherence images, as discussed in Section 2.3, bright and dark areas frequently correspond to sealed and vegetated areas respectively. In intensity images they are usually the consequences of the SAR-inherent corner reflection and shadowing effects, which in urban areas are dependent mainly on the buildings shape and orientation. The amount and geometrical properties of bright and dark areas are thus assumed to be closely related to the built-up structure of the urban blocks.

The extraction of these features was performed in the simplest way, namely, through thresholding of the input images. This operation creates a binary image based on a threshold value T and the following rule:

$$p(x,y) = \begin{cases} 0 \text{ if } g(x,y) < T \\ 1 \text{ if } g(x,y) \geqslant T \end{cases}, \tag{4.1}$$

where $p(x,y)$ and $g(x,y)$ are respectively a pixel position and its grey value. This exact rule was applied for extracting bright areas from the intensity and coherence images, whereas for the extraction of dark areas the inequality signs of equation 4.1 were obviously inverted.

After each thresholding operation, connected components (CCs) were created from the pixels with value 1 using the eight-pixel connectivity criterion [Gonzalez & Woods, 2001]. The area, orientation, and compactness of each CC were then computed and recorded.

4.2.3 Extraction of Lines

The number, length, position and orientation of lines extracted inside the urban blocks are very important evidences on their built-up structure. Therefore, a strategy was developed for extracting lines from the intensity images from each block. This strategy is summarized in the following workflow:

1. Edges Extraction → 2. Binarization → 3. Extraction of Lines → 4. Projection of Lines → 5. Filtering of Lines

The extraction of edges was performed by applying the Kirsch edge detector [Kirsch, 1971] over the intensity image of each block. The outcome of this step is an edge image E with the estimative of the edge gradient magnitude at each pixel. This is done by convolving the image I with different kernels $g^{(z)}$:

$$E_{n,m} = \max_{z\in[1,\ldots,8]} \sum_{i=-1}^{1} \sum_{j=-1}^{1} g_{ij}^{(z)} \times I_{n+i,m+j}, \tag{4.2}$$

such that $z \in [1, ..., 8]$ represents the compass directions ordered as N, NW, W, SW, S, SE, E and NE. The kernel used is always the same but rotated counterclockwise in 45° each time z changes from z to $z + 1$. Therefore:

$$g^{(1)} = \begin{bmatrix} +5 & +5 & +5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix},\ g^{(2)} = \begin{bmatrix} +5 & +5 & -3 \\ +5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix},\ g^{(3)} = \begin{bmatrix} +5 & -3 & -3 \\ +5 & 0 & -3 \\ +5 & -3 & -3 \end{bmatrix}, \text{... etc.}$$

After creating the edge image, it was binarized by the application of a global empirically defined threshold. Following, lines were extracted from this binary image using the Progressive Probabilistic Hough Transform (PPHT) method [Galamhos et al., 1999].

The standard version of this method, named Hough Transform [Duda & Hart, 1972], consists in representing lines with parameters r and θ. r is a vector that starts at the lower left origin of the image and ends at the closest point in the infinite line that goes through coordinate (i, j). θ is the angle between vector r and the x-axis of the image. Each possible line that goes through the image coordinate (i,j) will have an unique coordinate in Hough space ($r \in \mathbb{R}, \theta \in [0, \pi]$) and is defined as a function of θ : $r(\theta) = |i \cos\theta + j \sin\theta|$. One additional vote is then accumulated in the Hough space at each of these coordinates. This procedure is applied for every pixel (i, j) of the *binary* image whose value equals 1. Hough space coordinates with many votes define the parameters of pertinent lines in the image.

The problem however with the conventional Hough Transform is that it extract lines of infinite length parameterized only by its angle and distance from the origin. The PPHT method in the other hand is able to extract finite lines during the voting process by looking along the line's corridor and finding the longest segment of pixels either connected or separated by a gap not exceeding a given threshold [Galamhos et al., 1999]. The extraction of finite lines enables the calculation of their length. The following parameters of the PPHT method were defined empirically: (1) *voting threshold*: return only lines with number of votes larger than threshold; (2) *min. line length*: shorter lines get discarded and (3) *max. line gap*: merge two lines if their gap is shorter than parameter. Figure 4.3 depicts the results from the first three steps of the workflow for extracting lines inside the urban blocks.

It can be noticed from Figure 4.3 (e) that many extracted lines are redundant. The extraction of redundant lines is an inexorable effect of the PPHT method and the input binary image. In order to filter redundant ones, the lines were projected from acquisition to GK-Zone 12 coordinate

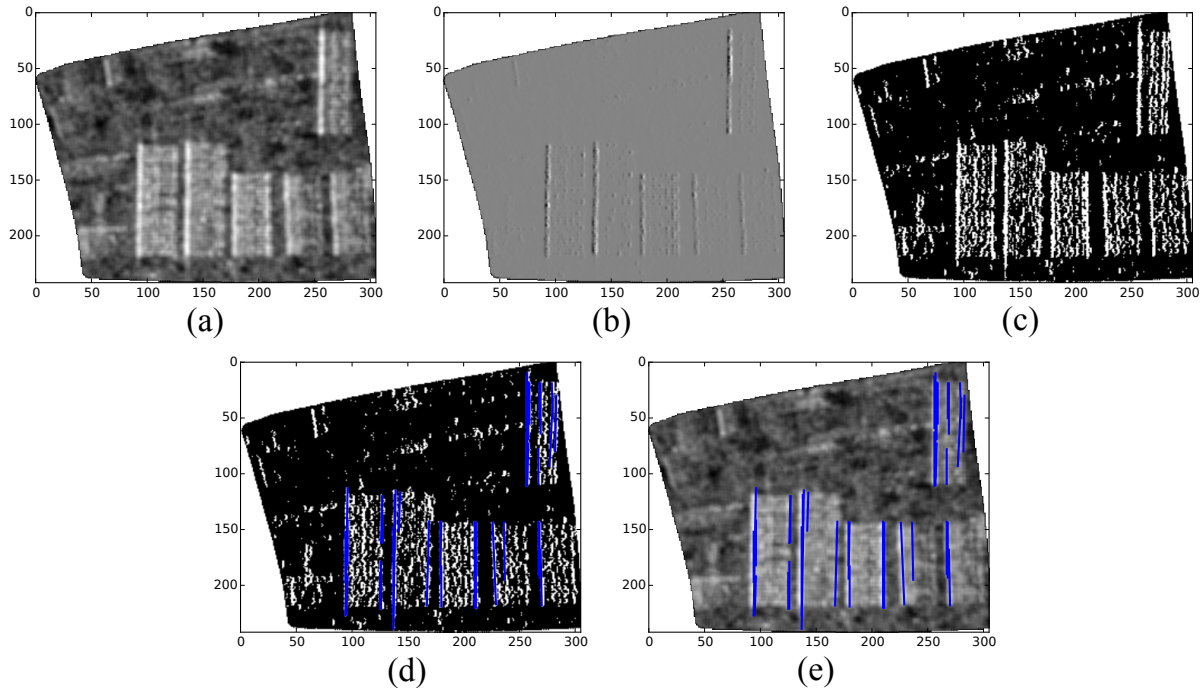

Figure 4.3: Partial results of the line extraction strategy. a) Input intensity image of an urban block. b) Edge image derived with the Kirsch [1971] filter. c) Binarization of the edge image through thresholding. d) Extraction of finite lines from the binary image using the PPHT method. e) Extracted lines overlaid onto the intensity image.

Line inherent:
Lenght
Orientation
Line to line related:
Distance to closest line
Orientation difference to closest line
Shortest distance between line and the most parallel line
Shortest distance between line and the most perpendicular line
Proportion between length of line and length of the closest line
Line to block boundaries related:
Distance to closest block boundary
Orientation difference to closest block boundary
Shortest distance between line and the most parallel block boundary
Shortest distance between line and the most perpendicular block boundary
Proportion between length of line and length of the closest block boundary

Table 4.1: Attributes computed for each line extracted inside each urban block.

system and their lengths and orientation were computed. Following, lines were kept, merged or eliminated according to basic criteria of proximity, joint length and relative orientation between every pair of close lines. Figure 4.4 shows in GK-Zone 12 coordinate system the boundaries of the urban block from Figure 4.3 and its extracted lines before and after the line filtering step. The green boxes show in detail that most redundant lines were successfully eliminated. For each of the remaining lines the attributes listed in Table 4.1 were computed.

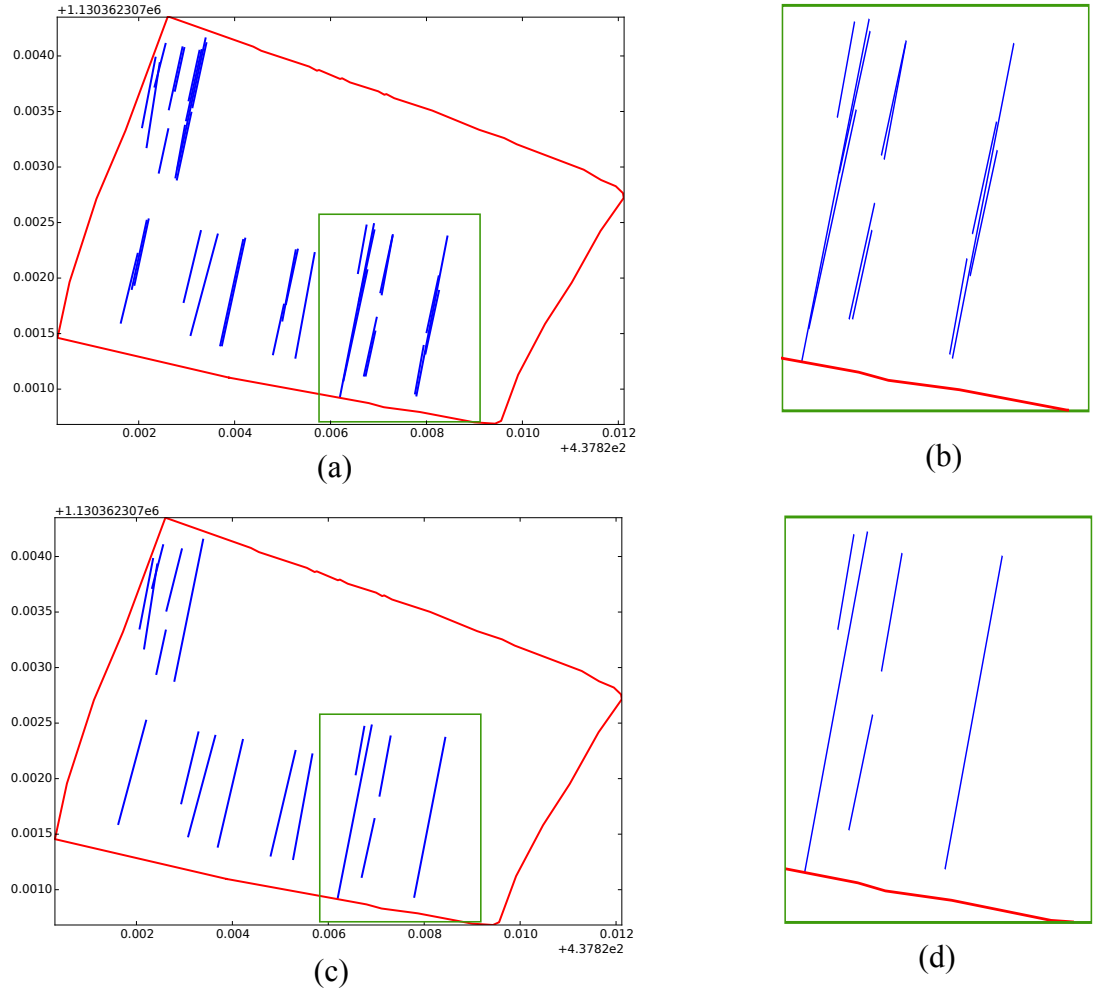

Figure 4.4: Example of the line filtering step. a) Boundaries and extracted lines from the block shown in Figure 4.3 projected to GK coordinate system. b) Lines before filtering shown in detail. c) Block lines remained after the filtering step. d) Remaining lines shown in detail.

4.2.4 Extraction of Polygons

Polygons were extracted with the aim of detecting as accurately as possible compact features that indicate the presence of buildings in the urban blocks. For that, a strategy was developed consisting of a top-down followed by a bottom-up step.

The top-down step consists in the creation of a Max-Tree from the intensity image of each block, whereas the bottom-up step consists in the filtering of meaningful nodes from the tree. A Max-Tree is a hierarchical represention of the structure of an image concerning its pixel values $g(x, y)$ [Salembier et al., 1998]. It is created by a sequence of image binarization steps - each of which consisting in the application of equation 4.1 - for n different thresholds T, such that $T_0 < T_1 < T_2 < ... < T_n$. At each binarization step, CCs are created which are represented as nodes in the tree. A tree has the amount of levels equal to the chosen number of binarization steps. Two CCs extracted respectively at binarization steps performed with T_i and T_{i+1} are parent and son if $CC_{(g)}^{T_i} \cap CC_{(g)}^{T_{i+1}} \neq \emptyset$, then surely $CC_{(g)}^{T_{i+1}} \subseteq CC_{(g)}^{T_i}$. In this case, a directed edge is set in the tree from $CC_{(g)}^{T_i}$ to $CC_{(g)}^{T_{i+1}}$. Figure 4.5 illustrates the creation of a Max-Tree from a synthetic image with four binarization levels. Min-Trees are exactly the same as Max-Trees,

only that the signs of equation 4.1 are inverted and $T_0 > T_1 > T_2 > ... > T_n$. Max-Trees were created for the extraction of bright polygons and Min-Trees for the extraction of dark polygons. In both cases, each CC was approximated to a smoother polygon by the Douglas-Peucker algorithm [Douglas & Peucker, 1973] and then projected to GK coordinate system. After that, the CCs *area*, *compactness* and *length-to-width ratio* were computed and stored in their respective nodes in the tree.

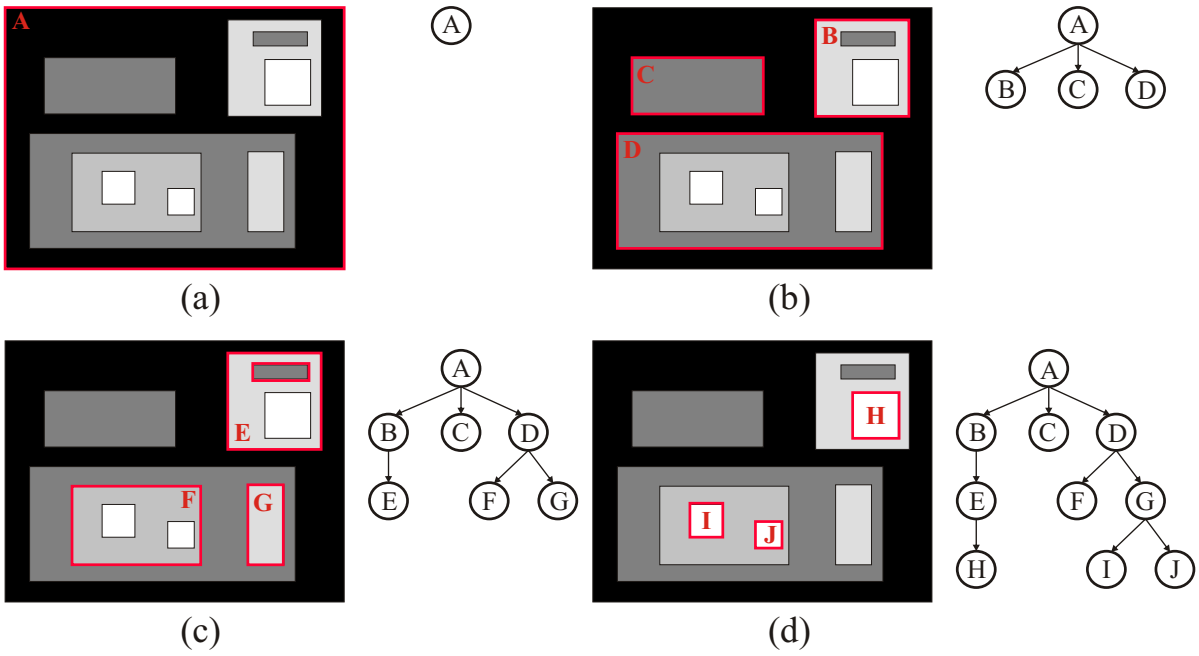

Figure 4.5: Example of a Max-Tree with four binarization levels created from a synthetic image. The tree nodes IDs are coded with capital letters. a) to d) CCs extracted with thresholds T_0 to T_3 and the respective nodes in the tree. Son-nodes are always entirely comprised in their parent-nodes.

Max- and Min-Trees are very effective for extracting meaningful polygons from images. However, the best of these polygons, approximated from CCs and represented by nodes in the tree, cannot be selected simply based on their attributes. This is because there might be *spatial conflicts* between them. This is exemplified by nodes G and D from the Max-Tree from Figure 4.5. It would make no sense, for example, to consider both of these nodes as representing features from two different buildings. Rather, one has to choose between keeping D and discarding all of its successors, or instead to keep G, and maybe F also, and discard all its ancestors. The selection/discard of polygons and hence the solving of spatial conflicts was performed at the bottom-up step of the polygon extraction strategy.

The bottom-up step, as the name implies, starts at the leaf nodes and goes up the tree comparing each node with its parent. At each comparison one of the nodes is discarded, namely, the one that matches the least the expected type of polygons. This matching is measured by pertinence functions which express in the interval from 0 to 1 how similar a node is, regarding a specific attribute, from an ideal polygon. As mentioned, three attributes were considered: *area*, *compactness* and *length-to-width ratio*. The pertinence values from each of these three attributes have to be somehow aggregated into a total pertinence value. The way in which this was done is shown in Figure 4.6. Firstly, the nodes attribute values are normalized according to certain boundary values and submitted to their associated pertinence functions. The pertinence values are then aggregated to a total one according to a simple hierarchy of aggregation operators. Finally, the node with highest total pertinence is kept. In case it is the son of the node against

with it is competing, it gets reconnected to its grand-father, which is the node it will be compared with in the next step of the bottom-up process.

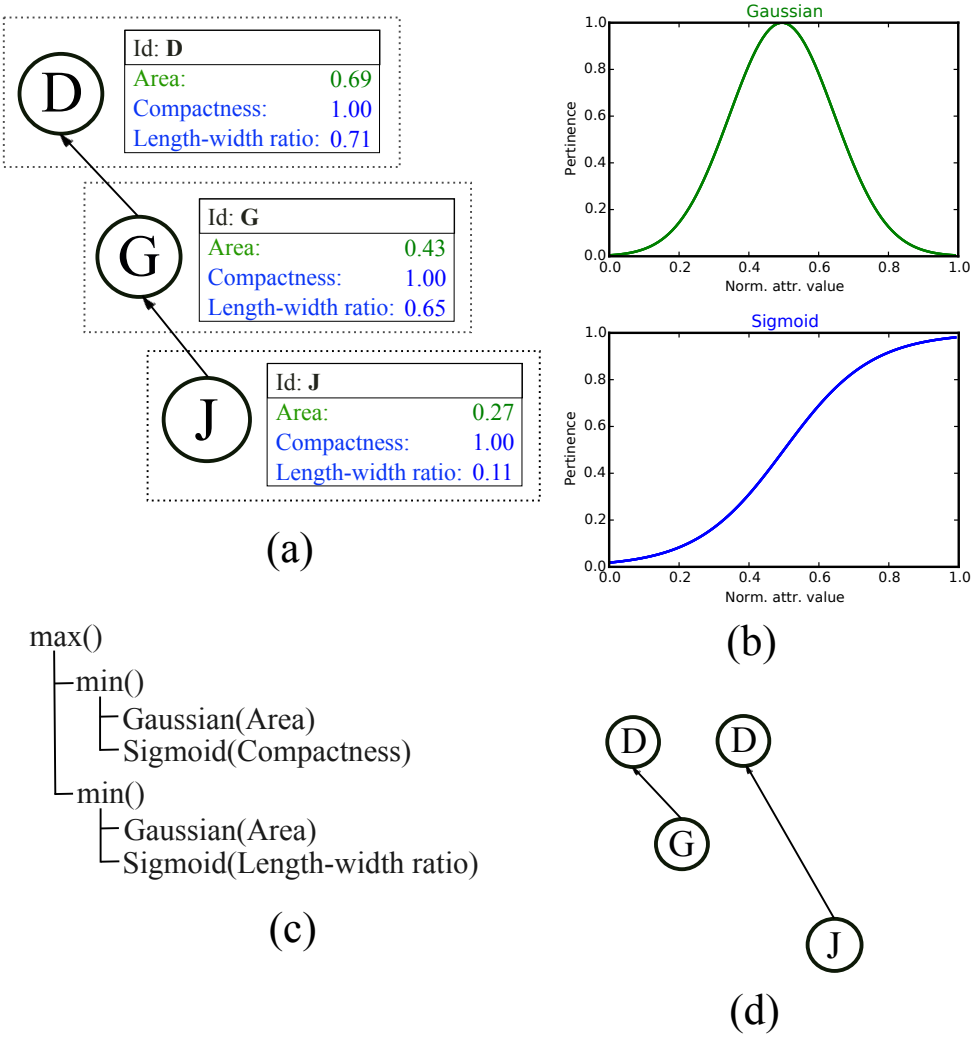

Figure 4.6: Bottom-up process for selecting meaningful polygons from the Max- and Min-Trees. a) In each of the tree's node the attributes of the polygon it represents are stored in normalized form. b) The pertinence functions to which the node's attribute values are submitted (according to the color association shown here). c) Hierarchy of aggregation rules responsible for computing a single pertinence value for each node. d) The two possible restructuring of the tree after the discard of the node with lowest pertinence.

The final result is a block's set of polygons with no spatial overlap. Polygons with pertinence value under a certain threshold were discarded. The remaining ones are expected to match to a good extent the sorts of polygons that ideally represent building features in the SAR-intensity image. The advantage of this strategy is that no segmentation is performed, what spares the sometimes unintuitive tunning of segmentation parameters. The only parameters that have to be set are the thresholds T_i and the pertinence functions, which can be standard ones like the gaussian, the sigmoid or the linear function.

4.3 Urban Block Attributes

In this section, the attributes from the urban blocks that were considered for USTs classification are described. These attributes are of three types, namely, (1) attributes computed from the CCs

of dark and bright areas, (2) attributes that describe the urban blocks according to the composition and configuration of lines and polygons inside them and (3) texture attributes extracted from the interferometrically derived DEM.

4.3.1 From Dark and Bright Areas

After extracting bright and dark areas from the coherence and intensity images and computing the CCs area, orientation and compactness, several *composition* attributes expressing the number, proportion and general geometric properties of each CC type were computed for each urban block. Some of these attributes were presented in equations 2.1 and 2.2. The complete list of these attributes, as well as the threshold values applied for extracting each type of CC are presented in Chapter 5.

In addition to these composition attributes, Multidimensional Profiles (MDP) were also generated from the extraction of bright and dark areas. A MDP is a sequence S of multivariate observations $x_i(t) \in [i = 1, ..., n; t = 1, ..., m]$, where x_i indices a variable from the observation t. It is assumed that the observations are logically ordered, as in a time-series analysis or in a systematic altering of parameters. MDPs can be represented as a $observations \times variables$ $(n \times m)$ matrix or as a profile in feature space, where one axis of the space represents the observations t and the other axes the variables x_i computed at each t. In the present context, t is a threshold value and the x_i variables are the composition attributes computed from the extracted CCs.

4.3.2 From the Extracted Lines and Polygons

As another way to describe the built-up structure of the urban blocks, several attributes were computed which regard the geometrical properties and the spatial configuration of lines and polygons inside the blocks. Such attributes were conceived with the aim to objectively express the signs considered by a human interpreter when visually distinguishing USTs on remote sensing images. Such attributes can be assigned to one of the following types:

- Composition attributes. Examples: Number of lines, Number of polygons larger than 500 m^2.
- Proportion attributes. Examples: Proportion between the number of lines and the number of block boundaries, Proportion between the joint area of all polygons and the block's area.
- Descriptive attributes. Examples: Mean length of lines, Area of largest polygon.
- Feature-Feature configuration attributes. Examples: Number of pairs of perpendicular lines, Standard deviation of the distance between parallel polygons.
- Feature-Block Boundary configuration attributes. Examples: Distance between the largest line and the closest block boundary, Orientation difference between the largest polygon and the block boundary most parallel to it.

As explained in Section 4.2.1, these attributes were computed with the features and the blocks boundaries coordinates projected to the local GK-Zone 12 coordinate system. The distance between polygons was measured between their centers of mass. The distance between lines was computed as the distance between their closest points. The complete list of these five types of attributes is presented in Chapter 5.

The spatial configuration of polygon features inside the blocks was also described based on *network-based* attributes. Thus, for each block a network was created in which the nodes represent

the polygons inside it and the edges define contextual relations between them. The networks were created in the following way. At each polygon's center of mass a node was set. Then, edges between each node and its two nearest-neighbors were created. Finally, each node was associated with the attributes of the polygon it represents. These attributes are: *area*, *compactness*, *length to width ratio*, *orientation*, *distance to closest block boundary*, *orientation difference to closest block boundary* and *pertinence*. Also, the edges connecting every two nodes were associated with the attributes *distance between the nodes*, *orientation difference between the polygons* and *mean pertinence of the polygons*. Figure 4.7 shows the network created from the polygons extracted inside an urban block. Several properties of the networks structures were extracted and considered as attributes from the respective blocks.

Another set of network-based attributes considered in this thesis concerns the spatial autocorrelation of the polygons (i.e. the nodes) attributes. A network's spatial autocorrelation is a measure of the extent to which connected nodes tend to have similar values regarding a given attribute. A common way to measure the spatial autocorrelation of a network's attribute is through the Moran's Index I [Moran, 1950], which is computed as:

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{X})(x_j - \bar{X})}{\sum_i (x_i - \bar{X})^2}, \tag{4.3}$$

where x_i is the value of node for variable x, $\bar{X}$ is the mean value of x and w_{ij} is an element of a matrix of spatial weights. The weights are defined as zero when two nodes are not connected and as their normalized distance otherwise.

The justification for considering spatial autocorrelation measures as a block's attributes is that some USTs are characterized by the presence of buildings with similar geometrical properties and constant relative orientation. In these cases, the Moran's Index is expected to be higher. This is also expected to happen when two similar polygons resulting from the extraction of a bright and a dark area are found very close to each other, what indicates the presence of a well-detected building.

4.3.3 From the Digital Elevation Model

As discussed in Section 2.3, DEMs generated from spaceborne InSAR data are not very expressive regarding the built-up structures from urban blocks. Nevertheless, it was assumed based on visual analysis that texture measures could be to some extent informative for the block-based distinction of USTs. Hence, Histograms of Oriented Gradients (HOGs) [Dalal & Triggs, 2005] were explored as texture descriptors.

The first step for the generation of HOGs is the computation of the gradient images in x and y directions by performing convolution filtering repectively with kernels $[-1, 0, 1]$ and $[-1, 0, 1]^T$. The second step is the creation of a histogram whose n bins represent different orientations from $0°$ to $180°$ discretized into n intervals. This histogram is accumulated by letting each pixel in the block to cast a weighted vote to each bin. The magnitude $|\nabla|$ and orientation θ of the gradient from each pixel define respectively the weight of the vote and to which orientation bin the pixel votes. These two measures are computed as follows:

$$\theta = \operatorname{atan2}(dx, dy), \quad |\nabla| = \sqrt{dy^2 + dx^2}. \tag{4.4}$$

A single unnormalized histogram with eight bins was accumulated for each urban block and each of the two DEMs. These eight bins values were used as attributes in the UST classification.

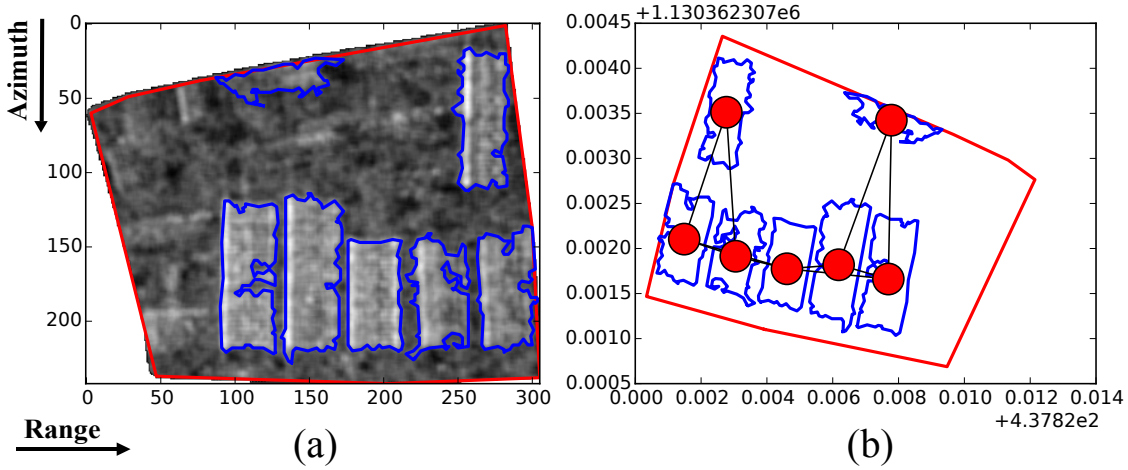

Figure 4.7: A network connecting the extracted polygons from an urban block. a) The extracted polygons and boundaries from an urban block shown in acquisition coordinates. b) The boundaries and polygons from the same block projected to GK coordinates. From the projected polygons in (b) a network was defined where nodes are set on the center of mass from each polygon and edges are defined between each node and its two-nearest neighbors. Block attributes are then computed as the structural properties of the network as well as the spatial autocorrelation measurements concerning the nodes attributes.

4.4 Urban Structure Types Classification

This section presents the two approaches with which the classification of USTs was performed. The classification results served as a basis for measuring the increase of accuracy achieved when context is considered in the classification and, as it will be explained, as a integrated part of the contextual classification approaches that are proposed in this thesis. Before that, however, the process of class generalization is addressed.

4.4.1 Class Generalization

Class generalization is the process of grouping specific classes into more general ones based on certain criteria. The criteria applied for grouping the specific UST classes were their physical and principally semantic similarities. The grouping of classes is inevitably subjective and dependent on the specific study area and other boundary conditions, such as the type of utilized remote sensing data and the ultimate use of the final classification. The official UST classes from the study area and their generalization are presented in Chapter 5.

4.4.2 Comparison of Multidimensional Profiles

The comparison of MDPs has several applications in the fields of computer vision [Kadous, 2002], biomedicine [Zhang et al., 1995] and data mining [Agrawal et al., 1993]. Although, to the best of our knowledge, it has not yet been explored for remote sensing imagery classification, it is assumed that MDPs can effectively express the structure of image regions regarding their pixel values. Classification of these image regions can be performed by simply collecting sample MDPs from each class and assigning the other regions to the classes according to a similarity metric.

Different types of metrics and measures have been proposed for comparing MDPs [Lee et al., 2000] [Yang & Shahabi, 2007] [Shao et al., 2011]. They all have in common the fact that it is possible to evaluate the uncertainty of the class assignments of a region by analysing the similarity

of its MDPs to the reference MDPs from all considered classes. These similarities can hence be understood as values of membership to the classes.

In this thesis, three types of MDPs were generated for each urban block, namely, two resulting from the extraction of bright areas from the intensity and coherence images and one resulting from the extraction of dark areas from the intensity images. As explained in Sections 4.2.2, each MDP was generated by applying a ordered sequence of threshold values to the images and by computing after each thresholding operation several block-based attributes regarding the extracted CCs.

The comparison of MDPs was performed using the *relative entropy* metric, also known as the Kullback-Leibler distance [Kullback & Leibler, 1951]. The Kullback-Leibler distance D between two blocks A and B over a random variable x at different observations t is given by

$$D(\,A(x)|\,|B(x)\,) = \sum_{t \in T} A(x(t)) * \log \frac{A(x(t))}{B(x(t))}, \tag{4.5}$$

where t is in this case a threshold value. Thus, $A(x(t))$ is the value of the attribute x computed from the CCs extracted by applying threshold t over the image of block A. As several attributes are computed from each block at the application of each threshold value $t \in T$, and as these attributes are considered independent from each other, the Kullback-Leibler distance D between two blocks A and B is computed simply as the sum of the univariate distance from A and B for each of the n attributes as

$$\begin{aligned} D(\,A(x_1,...,x_n)|\,|B\,(x_1,...,x_n)) = D(\,A(x_1)|\,|B\,(x_1)) + D(\,A(x_2)|\,|B\,(x_2)) + ... \\ ... + D(\,A(x_n)|\,|B\,(x_n)). \end{aligned} \tag{4.6}$$

Before it is possible to compare two MDPs using the relative entropy metric, the values of each x_i need to be normalized, since they probably have very different domains.

4.4.3 The Random Forest Algorithm

Random Forest is a classification algorithm developed by Breiman [2001] that provides comparable results to, and frequently better than, other powerful classifiers such as Suport Vector Machines, Logistic Regression and Decision Trees [Walton, 2008], [Novack et al., 2011a] [Rodriguez-Galiano et al., 2012]. Random Forest is said to be an ensemble and bagging method because it comprises several single classifiers $\{h(x, \Theta_k), k = 1, ...\}$, each of which contributes with one vote of same weight to the class assignment of an input element. Each k classifier is fitted based on a different independently generated subset of variables x and training samples $\{\Theta_k\}$ with the same class probability distribution.

In the Random Forest algorithm, decision trees are considered as the single classifiers. A decision tree is a top-down hierarchical sequence of binary splits of the data which generates a tree-structured set of rules. Figure 4.8 shows an example of a hypothetical decision tree. At the root of a tree, the dataset is in its entirety and at its leaves only instances that belong to the same class are to be found. All other nodes are thresholding rules that split the data in two. The optimal variable x_i and threshold value t for performing the split can be found by minimizing the Gini index of impurity:

$$S(x_i, t_j)^* = \underset{x_i \in [1,...,n], t_j \in \{D\}}{\arg\min} \; 1 - \sum_c p_c^2(x_i, t_j), \tag{4.7}$$

where $S(x_i, t_j)^*$ is the optimal split of a node and p_c is the proportion of class c at the data subset created by applying threshold t_j on variable x_i. D are all values assumed by variable x_i on the

training set. The minimization of equation 4.9 is performed at each node of a decision tree with the exception of its leaves.

The fact that each decision tree in Random Forest is derived with a random subset of variables and training samples reduces the correlation between the decision trees. It reduces also the computational cost of the algorithm and the risk of over-fitting, unlike methods based on boosting. In fact, the computing time is in the order of $T\sqrt{MN\log(N)}$, where T is the number of trees, M is the number of variables that each split evaluates and N is the number of training samples used to create each decision tree. All these parameters are user-defined.

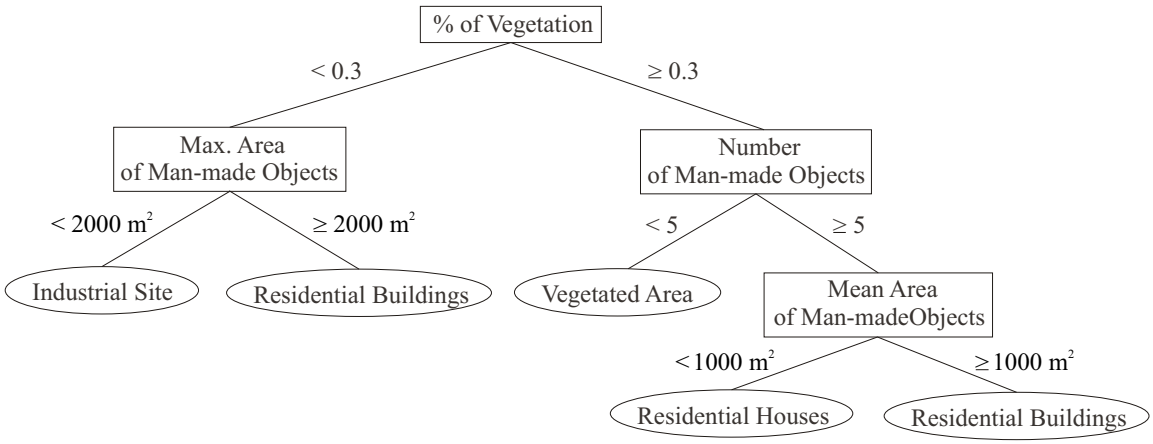

Figure 4.8: Example of a hypothetical decision tree. At its root the whole dataset is in its entirety and at its leaves only instances from a same class. All other nodes are thresholding rules that split the data in two according to a thershold that minimizes the Gini impurity index.

Random Forest has also the advantages that it is robust against noise and outliers, it can handle a very large number of variables without deletion and it can estimate the importance of variables during classification. This last aspect of the algorithm is important both for dimensionality reduction and for knowledge acquisition on which variables most influence the classification's accuracy. Variable importance estimation was performed with the mean accuracy decrease strategy [Hastie et al., 2001] in the following way: each tree of the forest is fitted and its accuracy is computed based on the samples of the sample set that were not used to fit the trees (called out-of-bag subset), then the values of variable m^{th} from the out-of-bag samples are randomly permuted and the accuracy of each tree is re-computed. For each tree and each variable m, the misclassification rate of the out-of-bag samples before and after m is permuted is compared. The rate difference for each m is averaged by the number of trees fitted with a subset of input variables that include m.

4.5 Contextual Analysis and Classification of Urban Structure Types

This section is comprised of three subsections. The first one describes the measures used to evaluate the spatial autocorrelation of the UST classes. The second describes how the shape similarity of two neighboring blocks was evaluated. The last and most important subsection presents the models proposed for performing contextual classification of USTs with undirected PGMs.

4.5.1 Spatial Autocorrelation of the Classes

Spatial autocorrelation analysis of the USTs was performed in order to evaluate the extent to which the assumption that neighboring blocks tend to pertain to the same UST class can be justified. The stronger the spatial autocorrelation, the stronger is this assumption. Two urban blocks are considered to be contextually related if they are mutual neighbors. The discussion on the criteria for establishing neighborhood relationships between blocks is left for Section 4.5.3.

The spatial autocorrelation can be measured globally or for each class separetely. To the best of our knowledge, the only ways of analysing global and per-class spatial autocorrelation of *categorical* variables is respectively with the Assortativity Index [Newman, 2003] and with Join Counts statistics [Cliff & Ord, 1981]. The assortativity index is computed in the following way:

$$Assortativity = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i}, \tag{4.8}$$

where $\mathbf{e}$ is a matrix whose elements e_{ij} express the proportion of pairwise neighboring relationships in which a block belongs to class i and its neighbor to class j. Thus,

$$\sum_{ij} e_{ij} = 1, \ \sum_i e_{ij} = a_i \text{ and } \sum_j e_{ij} = b_j. \tag{4.9}$$

As these neighboring relationships are undirected, the matrix e has the properties that $e_{ij} = e_{ji}$ and $a_i = b_i$. If the network is perfectly assortative, i.e. when every block is only a neighbor of blocks from the same class, then the assorativity index is 1 and $\sum_i e_{ii} = 1$. If the network is perfectly disassortative, i.e. when each block is only connected to neighbors of different classes, then $\sum_i e_{ii} = 0$ and the index will be positive and smaller than 1. Figure 4.9(a) shows sixteen elements belonging to four classes. Each element is a neighbor from the elements directly beside it as well as with those directly above and below it, as Figure 4.9(a) indicates. Table 4.2(a) shows the number of each class combination from these pairwise neighboring relationships, whereas Table 4.2(b) shows the respective proportion of each of these class combinations among all existing neighboring relationships. According to equation 4.8, the assortative index computed from the matrix shown in Table 4.2(b) is of approximately 0.44. For the purpose of comparison, the assortative index of the example from Figure 4.9(b) is of 0.55.

The Join Counts statistics follow the principle of counting seperatly for each class k the three possible class configurations among all pairs of neighboring blocks. These three configurations are WW, BB and WB, where W stands for white or 0, i.e. when a block i belongs to any class other than k, and B stands for blue or 1, i.e. when the block belongs to class k. These three configurations are counted as:

$$BB = (1/2) \sum_i \sum_j y_i y_j w_{ij}, \tag{4.10}$$

$$WW = (1/2) \sum_i \sum_j (1 - y_i)(1 - y_j) w_{ij}, \tag{4.11}$$

$$BW = (1/2) \sum_i \sum_j (y_i - y_j)^2 w_{ij}, \tag{4.12}$$

where y is the vector of observations, in this case of blocks, and w_{ij} is a spatial weights matrix in binary form. Hence, w_{ij} is 1 when block i and j are neighbors and 0 otherwise. Figure 4.9(c) and (d) show respectively the binary maps from classes 'Purple' and 'Yellow' according to Figure 4.9(a). Based on these binary maps and the neighborhood criteria presented above, the Join

Counts statistics from these classes are computed. It can be observed that the statistics BB, WW and BW for class 'Purple' are respectively of 4, 14 and 6. For the class 'Yellow' these statistics are of 2, 16 and 6.

Inference for the Join Counts statistics should, according to Cliff & Ord [1981], be performed computationally based on n random permutations, where at each permutation the blue and white observations are randomly moved spatially. Based on the permutations, a reference probability distribution is obtained and pseudo p-values can be computed as:

$$p = (m + 1)/(n + 1), \tag{4.13}$$

where m is the number of values from the reference distribution that are equal to or greater than the observed Join Counts. Since this is a one-sided test, if the p-value for BW is very small, that does not mean that there is negative autocorrelation for the given class, but instead, it indicates the presence of WW and BB autocorrelation in the data.

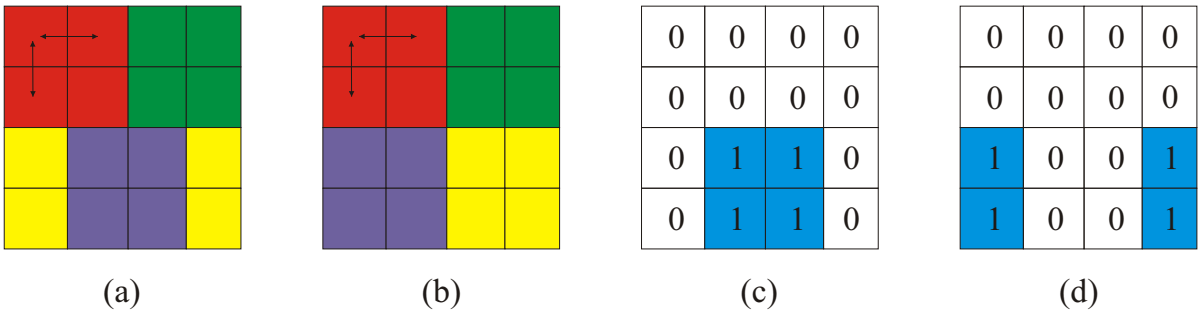

Figure 4.9: Examples for the explanation of the spatial autocorrelation measures for categorical data. a) A map with sixteen elements, four classes and twenty-four pairwise neighboring relationships. b) A variation of the map shown in (a). c) Binary map from the class 'Purple' from the map shown in (a). d) Binary map from the class 'Yellow' from the map shown in (a).

4.5.2 Urban Blocks Shape Comparison

As it will be discussed, the strength of the interaction between two neighboring blocks regarding their class assignments was also defined in proportion to the similarity of their shapes. Based on empirical observation, it was assumed that the probability of two neighboring blocks belonging to the same class is higher if their shape is similar. This can be observed mainly in older cities whose street network has not been entirely planned.

In this thesis, the L_2-based Turning Function metric proposed by Arkin et al. [1991] was used for measuring the shape similarity between two neighboring blocks. This metric represents a polygon A as its Turning Function $\Theta_A(s)$ in which s is the position in the normalized perimeter of A in relation to an intial point O on A's boundary (i.e. where $s = 0$). $\Theta_A(s = 0)$ is equal to the counterclockwise tangent between the segment of A in which O is located and the line parallel to the x-axis. This first angle is called v. As s increases by going along the perimeter of A in counterclockwise direction, $\Theta_A(s)$ will not change until s reaches the first vertice of A. At that point a "jump" occurs in the function relative to the counterclockwise tangent between the segments directly before and after s. If the turn s makes at that point is counterclockwise, this angle is positive and $\Theta_A(s)$ increases, otheriwse the angle is negative and $\Theta_A(s)$ decreases. Figure 4.10 shows an example of a hypothetic polygon A and what its Turning Function $\Theta_A(s)$ would be.

	Red	Green	Yellow	Purple	Total
Red	8	2	1	1	12
Green	2	8	1	1	12
Yellow	1	1	4	4	10
Purple	1	1	4	8	14
Total	12	12	10	14	

(a)

	Red	Green	Yellow	Purple	a_i
Red	0.16	0.04	0.02	0.02	0.25
Green	0.04	0.16	0.02	0.02	0.25
Yellow	0.02	0.02	0.08	0.08	0.21
Purple	0.02	0.02	0.08	0.16	0.29
b_j	0.25	0.25	0.21	0.29	

(b)

Table 4.2: Example used to explain the computation of the assortative index. a) Number of each class combination of the neighboring relationships from the elements in Figure 4.9(a). b) Proportion of each class combination among all existing neighboring relationships from the example in Figure 4.9(a).

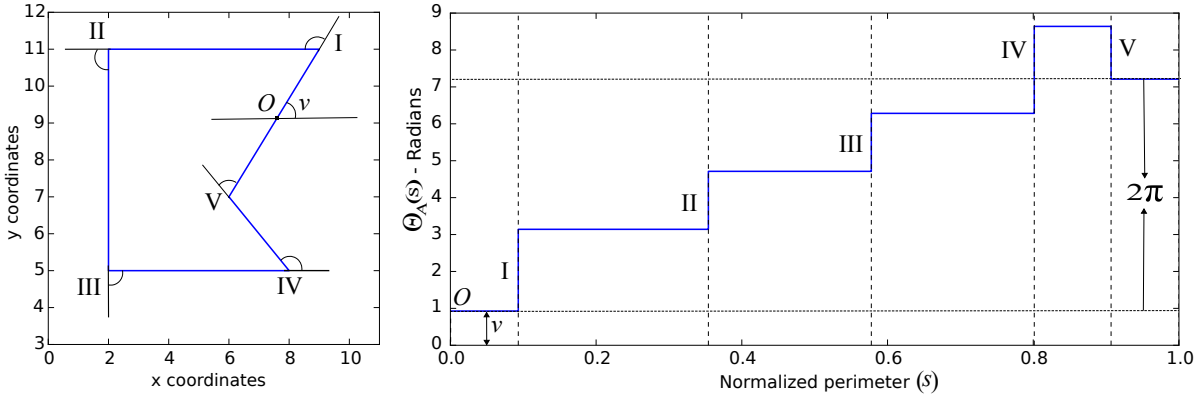

Figure 4.10: A polygon represented in the cartesian plane and its Turning Function. a) A hypothetic polygon named A represented in the cartesian plane. At the initial point O in A, where $s = 0$, the Turning Function $\Theta_A(s)$ is equal to the counterclockwise tangent of angle v measured in relation to the segment in which O is located and the line parallel to the x-axis. b) The Turning Function $\Theta_A(s)$ of the polygon A shown in (a). $\Theta_A(s)$ increases abruptly every time s encounters a vertice of A and s turns counterclockwise. If a clockwise turn takes place, $\Theta_A(s)$ decreases abruptly. These descontinuities are indicated with dashed lines in (b). Upper-case roman numbers indicate in (a) and (b) the vertices of A and the corresponding discontinuities in the Turning Function $\Theta_A(s)$.

The function $\Theta_A(s)$ can be extended to the whole real line by continuily accumulatint it as s goes around the normalized perimeter of A. Therefore, $\Theta_A(1) = v + 2\pi$, $\Theta_A(s+1) = \Theta_A(s) + 2\pi$ and $\Theta_A(s-1) = \Theta_A(s) - 2\pi$. Rotating the polygon A by θ corresponds to a vertical shift of $\Theta_A(s)$ to $\Theta_A(s) + \theta$, whereas altering the initial point O by an amount $t \in [0, 1]$ causes the horizontal shift of $\Theta_A(s)$ to $\Theta_A(s + t)$. The problem of measuring the similarity between polygons A and B is the one of finding t^* and θ^* for which the difference between the integrals of the Turning Functions of A and B is minimized. In other words, it is the problem of finding the optimal reference point and orientation of A for which the shapes of A and B are most similar. This similarity can be measured by the L_2-based Turning Function metric as follows:

$$S_2^{A,B} = \left\{ \min_{t \in [0,1]} \left[\int_0^1 [\Theta_A(s+t) - \Theta_B(s)]^2 ds - [\theta^*(t)]^2 \right] \right\}^{\frac{1}{2}}. \tag{4.14}$$

Fortunately, $S_2^{A,B}(t, \theta)$ is a convex function of θ for any fixed value of t and the optimal rotation θ^* for any given t is simply $a - 2\pi t$ [Arkin et al., 1991]. a is the difference between the areas under $\Theta_B(s)$ and the initial function $\Theta_A(s)$ given by:

$$a = \int_0^1 \Theta_B(s) - \int_0^1 \Theta_A(s). \tag{4.15}$$

Since the functions $\Theta_A(s + t)$ and $\Theta_B(s)$ are both piece-wise constant, the integral in equation 4.14 is very easy to compute. The values of s for which $\Theta_A(s)$ changes are named *discontinuities* and they occur, as mentioned, at each vertice of A. In Figure 4.10 (b) these discontinuities are indicated as dashed lines. If A and B have respectively m and n vertices, then plotting $\Theta_A(s+t)$ and $\Theta_B(s)$ in the interval $t = [0, 1]$ is expected to produce $m+n$ discontinuities. The rectangles in this plot bounded on the sides by two discontinuities and on the top and bottom by $\Theta_A(s+t)$ and $\Theta_B(s)$ are called *strips*. The area within each *strip* can be easely computed as its width (ds) times the square of its height, defined by $|\Theta_A(s + t) - \Theta_B|$. Adding up the areas of all strips gives the value of the integral in equation 4.14. Given that $\theta^*(t)$ is obtained as $a - 2\pi t$, the only problem left is to find t^*. Arkin et al. [1991] prove that only $m * n$ values of t must be evaluated to find it, namely, the values t at which a discontinuity in $\Theta_A(s+t)$ is aligned with a discontinuity in $\Theta_B(s)$. This makes it very simple to minimize equation 4.14 and thus compute the similarity between two urban blocks based on their Turning Functions. Figure 4.11 shows the Turning Functions from two hypothetical urban blocks named Red and Blue before and after $\Theta_{Blue}(s)$ gets shiftted by an amount t which makes a discontinuity in $\Theta_{Red}(s)$ coincide with a discontinuity in $\Theta_{Blue}(s + t)$.

4.5.3 Contextual Classification Models

The contextual classification of USTs was performed using different undirected PGMs. The models differ according to the following aspects: (1) strategy for defining the graph's structure, (2) the type of function for defining the interaction potentials (3) whether the model has one or two groups of association potentials and (4) the order of the model, i.e. the number of unobserved variables present in the interaction factors. For the rest of this section, the notation from Section 3.4 will be adopted.

The first aspect that should be considered when developing a PGM for performing contextual classification of UST is the criterion according to which two urban blocks i and j are considered to be neighbors. Obviously, if they are neighbors, there is contextual interaction between them and

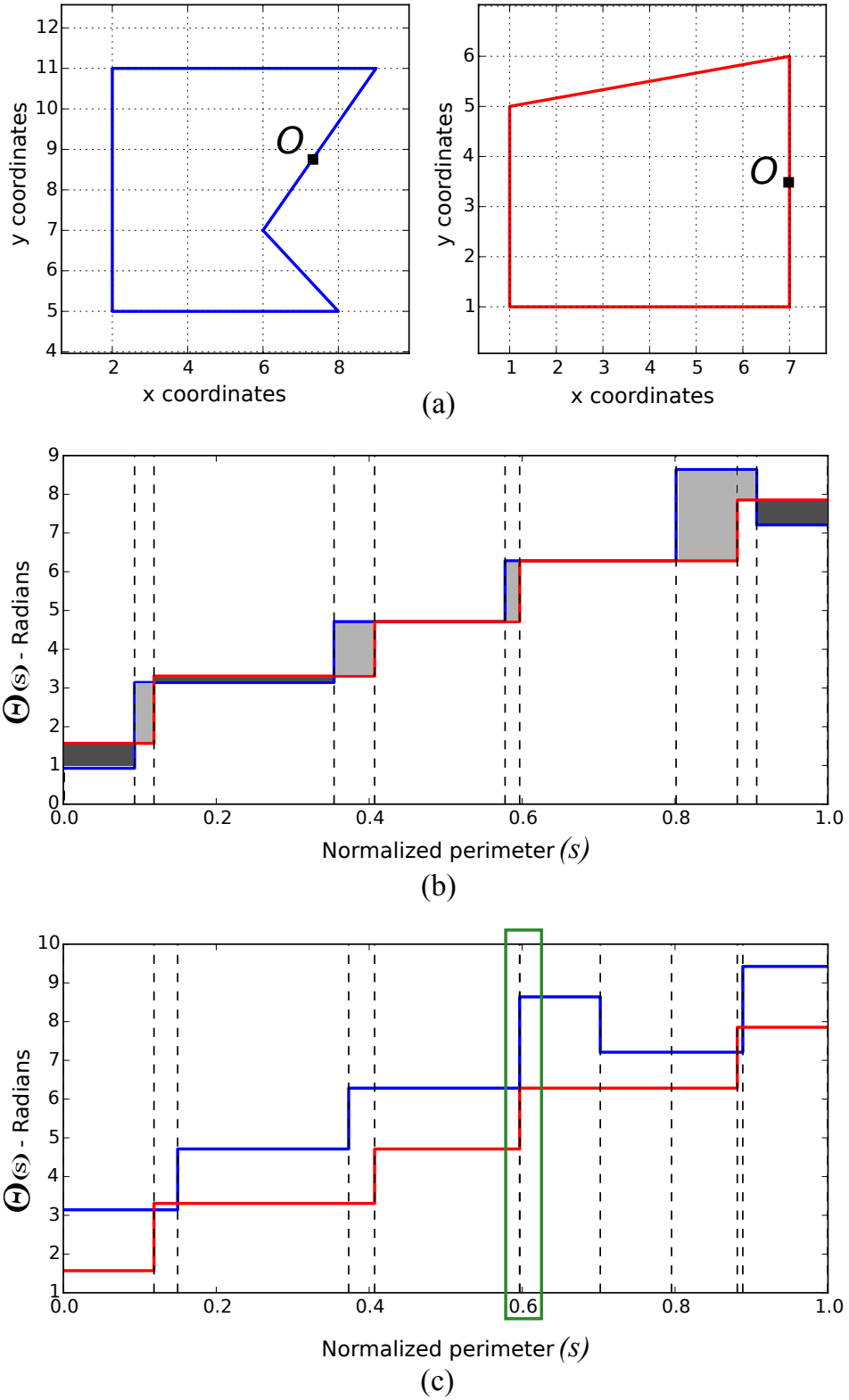

Figure 4.11: Horizontal shifting of the Turning Function of a hypothetical urban block. a) Two hypothetical urban blocks represented in cartesian space and the origin points O in their perimeter where $t = 0$. b) The Turning Functions $\Theta_{Blue}(s)$ and $\Theta_{Red}(s)$ of the two blocks. The areas in light grey indicate the intervals for which a (equation 4.15) is positive. The areas in darker grey indicate the intervals for which a is negative. c) Horizontal shift of $\Theta_{Blue}(s)$ to the amount of t so that a discrepancy from $\Theta_{Blue}(s+t)$ overlies horizontally a discrepancy from $\Theta_{Red}(s)$ (highlighted by the green rectangle).

they are thus probabilistically dependent. In this thesis, four criteria for establishing neighboring relationships between blocks were considered. The first three of them are defined as follows:

$$j \in N_i \text{ if } \begin{cases} \text{Criterion 1: } d(i,j) < th \\ \text{Criterion 2: } d(i,j) < th_i(length_i * 1,5) \wedge dist.(i,j) < th' \\ \text{Criterion 3: } j \in k\text{-nn}(i) \wedge d(i,j) < th' \end{cases}, \tag{4.16}$$

where N_i is the set of neighboring blocks from i, $d(i,j)$ is the euclidean distance between block i and j, th denotes a fixed distance threshold and th' denotes an upper distance threshold above which two blocks are not considered neighbors. $d(i,j)$ was always measured between the centers of mass of the urban blocks. The first criterion considers two blocks as neighbors if they are not further apart as a fixed th, the second criterion is just like the first with the exception that th is adaptively defined as the length of block i times 1,5. The factor of 1,5 was chosen because, having the center of mass of block i as its origin, the radius of $0,5 * length(i)$ will usually include mostly the area of the block itself. Thus, the adaptive radius of $1,5 * length(i)$ will include as neighbors of i the blocks j that are apart from i to the extent of its length. The third criterion considers as neighbors from i its k nearest-neighboring blocks. The fourth criterion considers blocks i and j as mutual neighbors if they are adjacent to each other. In this case, the streets are considered the borders between the blocks. The adjacencies between the blocks was extracted by firstly creating a binary image where the blocks are the foreground and the streets, rivers and railroads network is the background. Following, the euclidean distance transform of this binary image was created. Finally, a watershed segmentation was performed over the distance transform image in order to extract the exact borders between the blocks. If two blocks share a border, they are considered neighbors. This process is depicted in Figure 4.12.

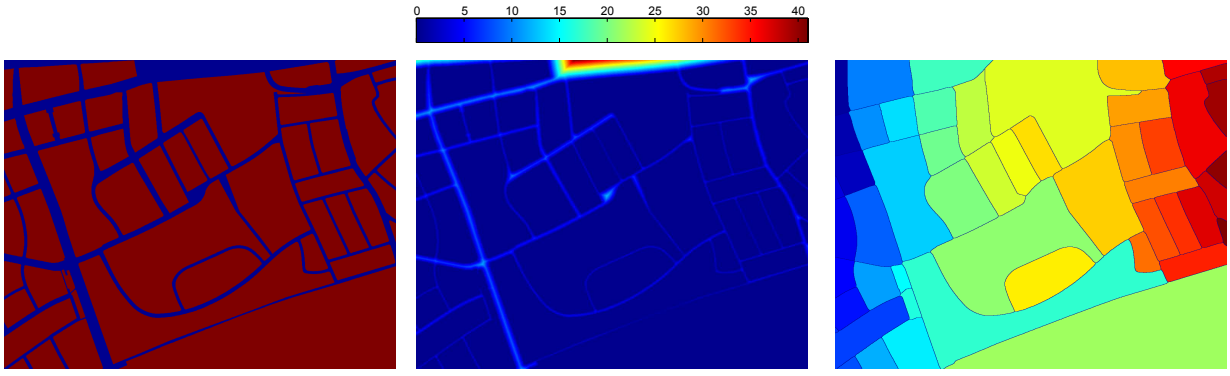

Figure 4.12: Steps performed for extracting the adjacency between urban blocks. a) Input binary image. b) Euclidean distance transform of the binary image. c) Outcome of the watershed segmentation applied on the distance transform image. Two blocks are considered to be neighbors if they share a border on the watershed image.

It is possible to depict with Figure 4.13 the construction of a simple undirected graph for contextual UST classification. In this graph, blocks are considered neighbors if the euclidean distance between their centers of mass is not larger than a threshold th. Each block's unobserved variable c is represented in the graph by a node and an edge is created between each pair of neighbors. If it is defined that, besides its neighbors unobserved variable c_j, c_i depends also on its observed s_i, then the resulting graph will have the number of association factors equal to the number of blocks in the study site and the number of pairwise interaction factors equal to the number of pairs of neighboring blocks. The process depicted in Figure 4.13 for one block

i is performed for each block of the study site. The repeated interaction factors are afterwards discarded.

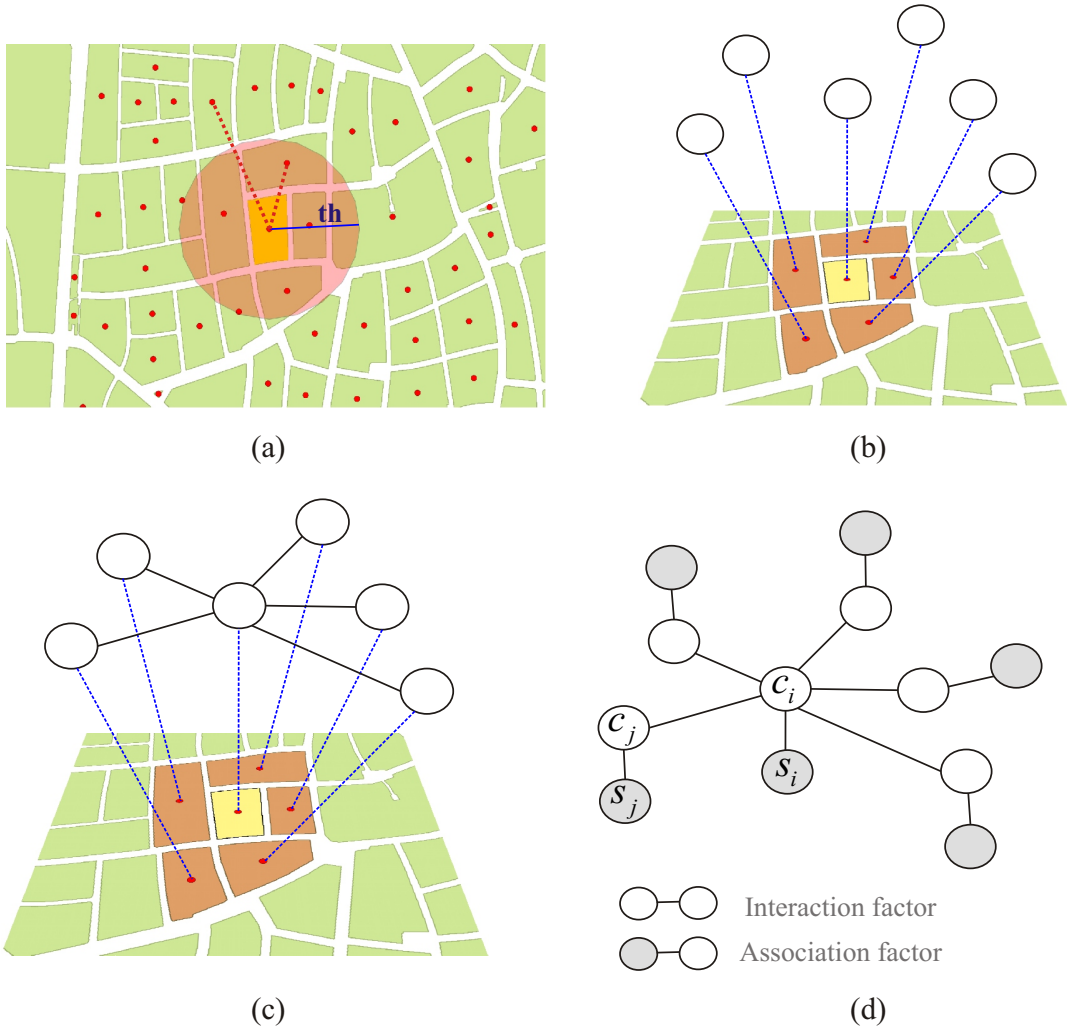

Figure 4.13: Construction of a simple undirected Probabilistic Graphical Model for Urban Structure Types classification. a) Blocks are considered neighbors if the euclidean distance between their centers of mass is not larger than a threshold th. b) Each block's unobserved variable c is represented in the graph by a node. c) An edge is created between each pair of neighbors. d) If it is defined that, besides its neighbors unobserved variable c, c_i depends also on its observed s_i, then the resulting graph will have the number of association factors equal to the number of blocks in the study site and the number of pairwise interaction factors equal to the number of pairs of neighboring blocks.

Independently of which criterion is used to define the pairwise neighboring relations between blocks, the graph shown in Figure 4.13 is a MRF. It is the simpliest type of PGM with which contextual UST classification was performed in this thesis and it can be formulated as an energy function as follows:

$$E(C) = \sum_{i \in S} \varphi(c_i, s_i) + \lambda \sum_{i \in S} \sum_{j \in N_i} \psi(c_i, c_j). \tag{4.17}$$

Its association functions were defined as

$$\varphi(c_i, s_i) = -\ln(P(c_i = l_k | s_i)), \tag{4.18}$$

where $P(c_i = l_k|s_i)$ is the probability of class label l_k given the data s_i. This probability is given by the Random Forest classifier as the number of votes from the decision trees to class label l_k divided by the total number of decision trees. The association functions $\psi(s_i, s_j)$ were defined as Potts models [Hao et al., 2014]:

$$\psi(c_i, c_j) = \begin{cases} 0 \text{ if } c_i = c_j \\ 1 \text{ if } c_i \neq c_j \end{cases}. \tag{4.19}$$

The parameter λ in the energy function above has the purpose of balancing the influence of the interaction factors in relation to the association factors, as the former, depending on th, may be much more numerous than the latter. As it will be shown, the parameter λ was tunned to the best value between 0 and 1.

A simple CRF version of the model from equation 4.19 was formulated as follows:

$$E(C) = \sum_{i \in S} \varphi(c_i, s_i) + \lambda \sum_{i \in S} \sum_{j \in N_i} \psi(s_i, s_j, c_i, c_j). \tag{4.20}$$

The association potentials in this model are defined just as before. The interaction potentials on the other hand are defined by a so-called contrast-sensitive Potts model, which considers the l_2-norm of the difference between the normalized attribute vectors of the neighboring urban blocks:

$$\psi(s_i, s_j, c_i, c_j) = \begin{cases} 0 \text{ if } c_i = c_j \\ -\ln\left(\|s_i - s_j\|_2\right) \text{ if } c_i \neq c_j \end{cases}. \tag{4.21}$$

In this contrast-sensitive Potts function, the more similar the attribute vectors of the neighboring blocks, the higher is the cost for assigning them to different classes. It can be seen from Figure 4.14 that defining the cost as $-\ln\left(\left\|s_i - s_j\right\|_2\right)$ instead of $1 - \left(\left\|s_i - s_j\right\|_2\right)$ makes the model more strict against assigning different classes to neighboring blocks.

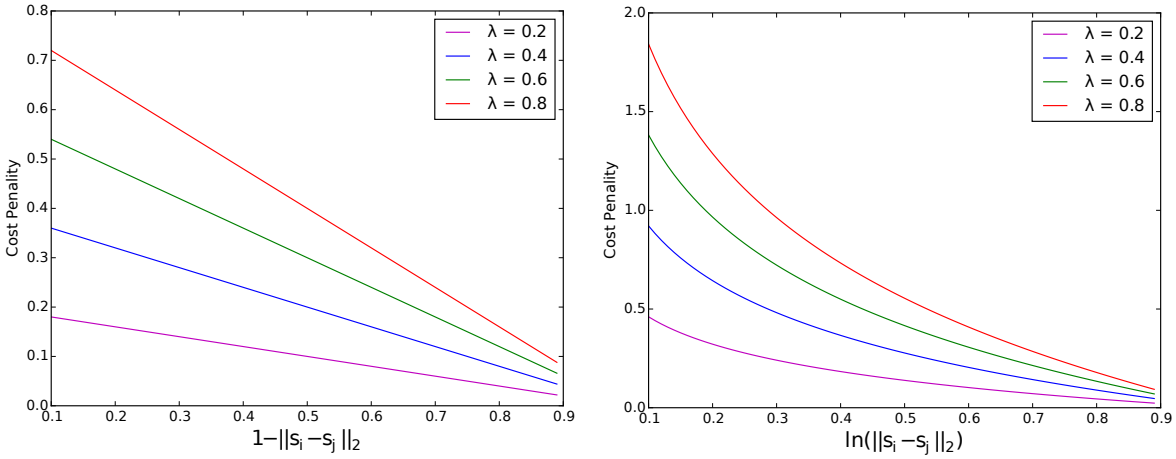

Figure 4.14: Different cost functions for different λ values. a) The costs of assigning neighboring blocks to different classes is weaker in this case. b) This function makes the costs higher, what causes that models with similar functions are more strict against assigning different classes to neighboring blocks.

An alternative way of defining the interaction potentials was also experimented with the following contrast-sensitive Potts function:

$$\psi(s_i, s_j, c_i, c_j) = \begin{cases} 0 \text{ if } c_i = c_j \\ -\ln\left(\left\|s_i^w - s_j^w\right\|_1\right) \text{ if } c_i \neq c_j \end{cases} \tag{4.22}$$

where s_i^w and s_j^w are the attribute vectors from neighboring sites i and j weighted by the attribute importance vector computed by the Random Forest classifier. In other words, it is the negative natural logarithm of the weighted mean of the difference.

A similar CRF model was tested which considers the shape similarity between neighboring blocks i and j for defining the cost in case they are assigned to different classes. This model is formulated in the following way:

$$E(C) = \sum_{i \in S} \varphi(c_i, s_i) + \lambda \sum_{i \in S} \sum_{j \in N_i} \psi(S_2^{i,j}, c_i, c_j) \tag{4.23}$$

and their interaction potential functions are given by:

$$\psi(S_2^{i,j}, c_i, c_j) = \begin{cases} 0 \text{ if } c_i = c_j \\ -\ln\left(S_2^{i,j}\right) \text{ if } c_i \neq c_j \end{cases}, \tag{4.24}$$

where $S_2^{i,j}$ is the shape similarity between blocks i and j computed with the L_2-based Turning Function measure presented in Section 4.5.2.

Given that standard UST classification was performed not only with the Random Forest algorithm, but also by means of MDP comparison, the MRF and CRF models from equations 4.19 and 4.22 were extented to accomodate a second type of association potential functions. These functions are defined by the strategy for computing the membership from blocks to UST classes based on MDPs comparison. The MDP creation and comparison strategies are explained respectively in Sections 4.3.1 and 4.4.2, whereas the way in which the final membership values from each block to each class was computed is addressed in Chapter 5 at Section 5.2.3. The MRF model with this additional association term is formulated as:

$$E(C) = \sum_{i \in S} \varphi(c_i, s_i) + \sum_{i \in S} \delta(c_i, s_i) + \lambda \sum_{i \in S} \sum_{j \in N_i} \psi(c_i, c_j), \tag{4.25}$$

whereas its CRF version is formulated as

$$E(C) = \sum_{i \in S} \varphi(c_i, s_i) + \sum_{i \in S} \delta(c_i, s_i) + \lambda \sum_{i \in S} \sum_{j \in N_i} \psi(s_i, s_j, c_i, c_j). \tag{4.26}$$

The interaction potential functions from these two models were, just as before, defined by equations 4.21 and 4.23. The association terms are two: $\varphi(c_i, s_i)$, defined as the outcome from the Random Forest algorithm for each block i and $\delta(c_i, s_i)$, which is defined by the MDPs comparison strategy.

All the models presented above have in common the fact that all their interaction factors have no more than two unobserved c variables. As mentioned in Section 3.4, models with interaction factors with more than two of such variables are called higher-order models. In this thesis, classification experiments were performed also with third-order MRFs defined as:

$$E(C) = \sum_{i \in S} \varphi(c_i, s_i) + \lambda \sum_{i \in S} \sum_{j \in N_i} \psi(c_i, c_j, c_m). \tag{4.27}$$

The interactions potentials from these models were defined by the simple third-order Potts function:

$$\psi(c_i, c_j, c_m) = \begin{cases} 0 \text{ if } c_i = c_j = c_m \\ 1 \text{ otherwise} \end{cases}. \tag{4.28}$$

The construction of third-order factors was done in the following way. First, pairwise neighboring relationships were defined based on the criterion of 3-nearest neighbors. Then, for each neighboring pair of blocks i and j, the closest block m to j which is also a neighbor from i was searched. In this way, a factor $\psi(c_i, c_j, c_m)$ was inserted in the model. This process, shown graphically in Figure 4.15, was repeated for every block i and afterwards all repeated factors were discarded.

The models presented above express the types of UST contextual classifications that were performed. Minimizing their energy functions is equivalent to finding the most-probable classification of the study area, as explained in Section 3.3. As mentioned, the most-probable classification can only be found approximately when dealing with undirected PGMs. In all cases, the algorithm used for that task was the Max-Sum LBP, whose equivalent Max-Product LBP is presented in Figure 3.8. The actual experiments performed and their results will be presented respectively in Chapters 5 and 6.

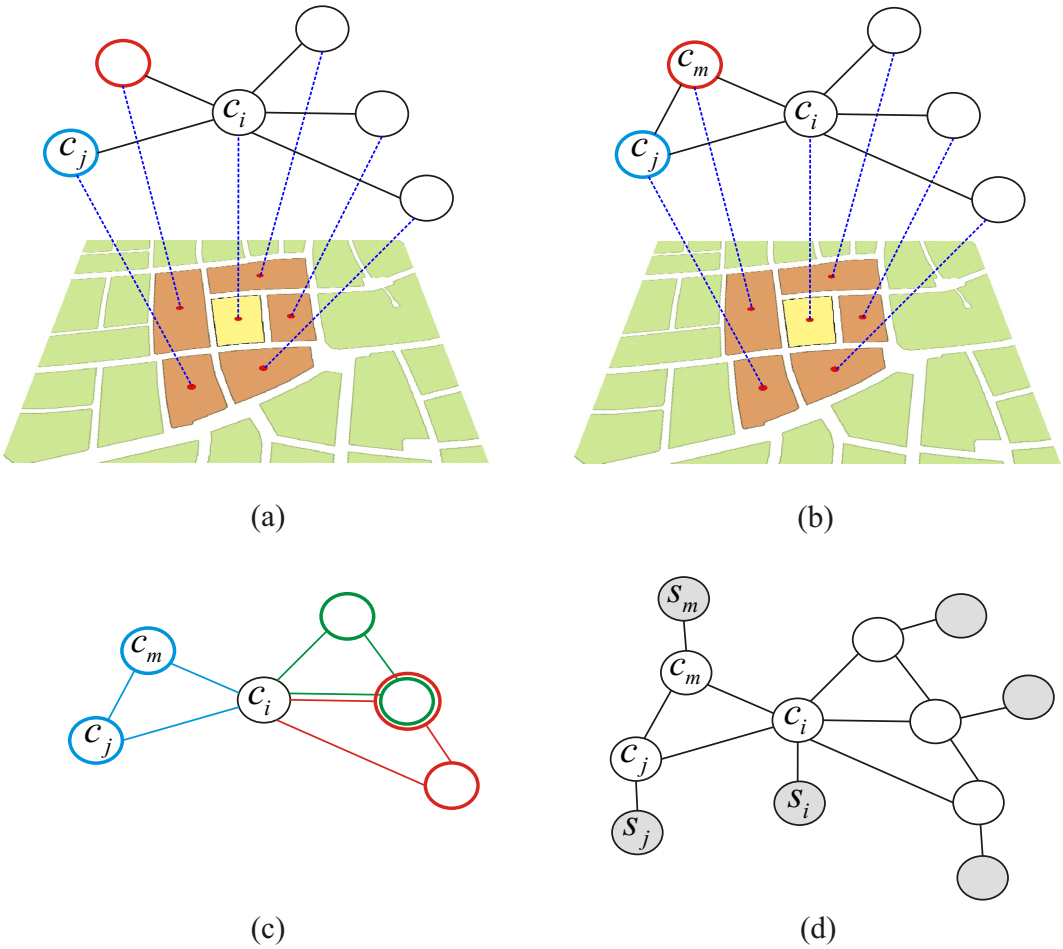

Figure 4.15: Strategy for building third-order interaction factors from a third-order MRF. Firstly, pairwise neighboring relationships were defined based on the criterion of 3-nearest neighbors. a) For each neighboring pair of blocks i and j, the closest block m (shown in red) to j (shown in blue) which is also a neighbor from i is searched. b) An interaction factor containing variables c_i, c_j and c_m is thus inserted in the model. c) This process is repeated for the other neighbors of i and for all other blocks in the study area. The different colors represent different thrid-order interaction factors. d) The final local graph structure for these six blocks.

5 Data and Experiments

This chapter presents in its first part the image, auxiliary and ground-truth data used in this thesis as well as its study site. The second part describes the experiments and how some of the methods presented in the previous chapter were applied.

5.1 Used Data and Study Site

5.1.1 Image Data

The extraction of features from the urban blocks was performed based on SLC images acquired by the TS-X satellite operating at HSS mode. The images have a nominal spatial resolution of approximately 1,1 meter and were kindly provided by the German Aerospace Center (DLR). Table 5.1 presents the main acquisition parameters. The acquisition dates were chosen after confirming close to clear-sky conditions on the respective days. The minimum and maximum incidence angles of the chosen acquisitions are such that the layover and shadowing effects are, among the imagery available at that time, most balanced. The four images presented in Table 5.1 configure two interferometric pairs, in which one was obtained at ascending and the other at descending looking direction. Table 5.2 presents the main parameters of these two InSAR pairs. Because the area covered by the images is relatively flat and its surface height differences almost never exceeds 80 meters, the heights of ambiguity of both InSAR datasets were considered to be appropriate for the application at hand.

Date	Range Spacing	Azimuth Spacing	Ran. x Azi. Looks	Min. Inc. Angle	Max. Inc. Angle	Size
18.05.2011	0.45	0.86	1 x 1	22.33°	23.63°	10364 x 6054
09.06.2011	0.45	0.86	1 x 1	22.23°	23.63°	10364 x 6082
09.05.2011	0.45	0.87	1 x 1	24.73°	25.94°	11166 x 5342
20.05.2011	0.45	0.87	1 x 1	24.72°	25.94°	11168 x 5982

Table 5.1: Main acquisition parameters of the TS-X images used in this thesis.

Looking Direction	Date	Polarization	Perp. Baseline	Height of Ambiguity
Ascending	18.05.2011 09.06.2011	VV VV	51.16 m	70.86 m
Descending	09.05.2011 20.05.2011	VV VV	120.16 m	27.32 m

Table 5.2: Main parameters of the InSAR image pairs used in this thesis.

From both InSAR image pairs shown in Table 5.2, two DEMs were generated using the public software *Nest* developed by the European Space Agency (https://earth.esa.int/web/nest/home). The sequence of steps performed and the respective parameters applied in each of them are shown in Table 5.3.

Steps:	**Parameters:**
Co-registration	900 Ground Control Points
Generation of the flattened interferogram	-
Estimation of the coherence image	Window size: 10x2 (az. x rg.)
Filtering of the interferometric phase	Method: Goldstein, Alpha: 0.5, Block size: 32
Multi-looking of the filtered phase	Rg x Az. looks: 2x3
Phase unwrapping	Method from Costantini [1998]
Geocoding	-

Table 5.3: Steps and parameters applied for the generation of InSAR DEMs.

5.1.2 Study Site

The TS-X images presented in the previous section cover a significant area of the city of Munich (Germany). Figure 5.1 shows the exact footprints of the ascending and descending acquisitions. Each footprint covers an area of approximately 10 x 5 km. The study site was defined as the intersection area of these two footprints. Its central point coincides with the historical center of the city.

Munich was chosen as this thesis's study site due to the following conveniences: (1) there is ample availability of TS-X imagery from this city, (2) reliable auxiliary and ground-truth data are available and (3) its built-up structures are representative of other medium to large European cities.

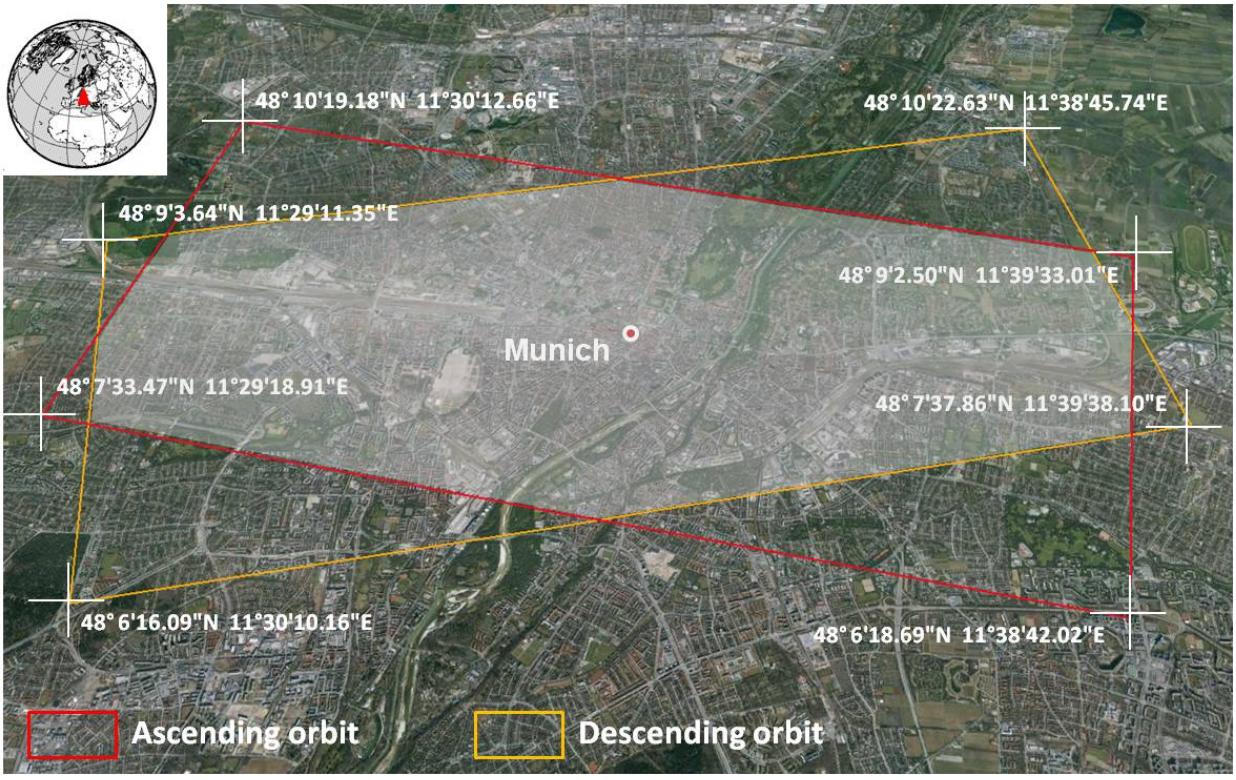

Figure 5.1: The image data footprints and the study site of this thesis. Each footprint covers an area of approximately 10 x 5 km. Their intersection area was defined as study site. Its central point coincides with the historical center of Munich.

5.1.3 Auxiliary and Ground-Truth Data

The only auxiliary data used in all experiments of this thesis are the streets, rivers and railroads network as well as the official UST map from the city of Munich. Both data were kindly provided in shapefile (.shp) format by the city hall's Department of Health and Environment (Referat fuer Gesundheit und Umwelt). As explained in Section 4.1, the streets, rivers and railroads network was used to extract the urban blocks from the study site, whereas the official UST map from Munich was considered as ground truth data and used for collecting training samples and assessing the classifications accuracies. Fortunately, the latest version of this map is from 2011, which is also the year in which the SAR images were acquired. Figure 5.2 shows the official UST map from the study site. It contains 1696 urban blocks assigned to 27 different classes.

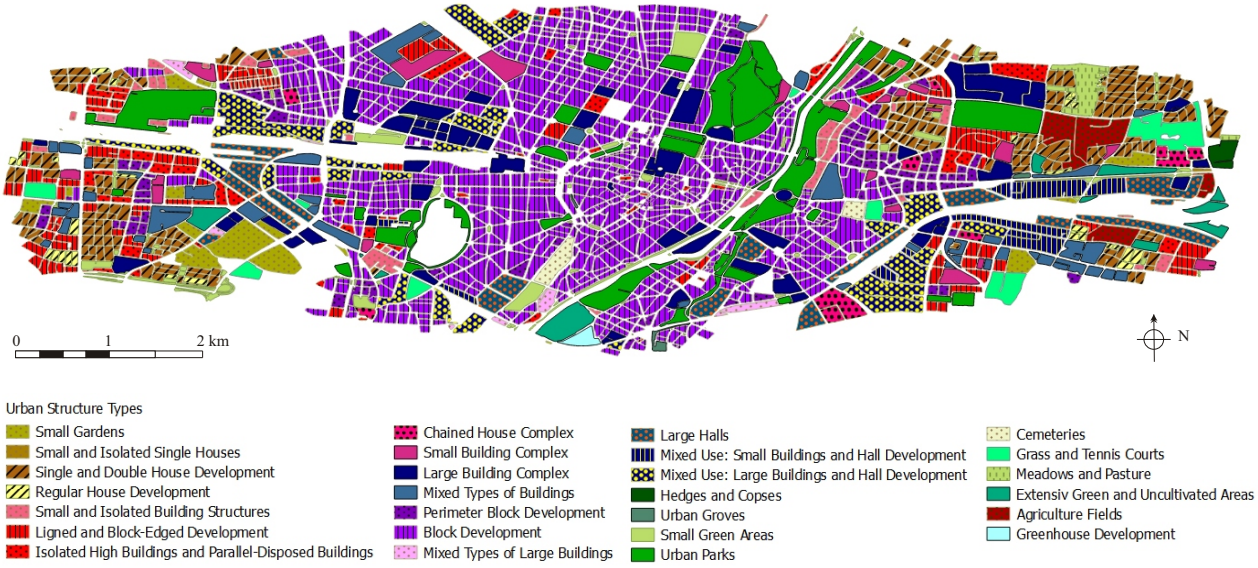

Figure 5.2: Official Urban Structure Types map from the study site of this thesis. It contains 1696 urban blocks assigned to 27 different classes. The data was kindly provided by the city hall of Munich.

5.2 Experiments

5.2.1 Class Generalization and Sample Selection

As it can be observed in Figure 5.2, many of the official UST classes from the study site have a low number of instances. Moreover, some of them are defined exclusively by functional factors, i.e. they do not have physical characteristics that make them distinguishiable from other similar classes. Because of that, a generalization of the classes was performed. It comprised two steps. The first one involved assigning mixed blocks to one single UST class. Among the 1695 blocks of the study site, 21.89% of them are assigned to more than one UST class. These blocks were assigned to a single UST class, namely, to the one with largest relative area. The second step concerned the semantic grouping of the original 27 UST classes into five general ones. Table 5.4 shows to which general classes the original ones were assigned. The number of instances from each class is shown in parenthesis after their names. As mentioned in Section 4.4.1, the criteria for merging classes were their structural and functional similarities. The outcomes of the class generalization process are a group of five UST classes, each with a reasonable number of instances, and a ground-truth map in which each block is assigned to a single UST class.

The selection of samples for training the classification models was performed in two ways, namely, randomly and in supervised way. In both cases, the same amount of samples from each of the five classes was selected. This amount was defined as half the number of instances from the UST class with least instances. It can be seen from Table 5.4 that this class is Large Buildings and Industrial Areas and that it has 126 blocks officially assigned to it. Therefore, the number of samples from each of the five classes collected at all classification experiments was of 63. In total, twenty-six different groups of training samples were collected: one in supervised way and the other twenty-five randomly.

Dense Block Development (DBD) (691):
Block Development (691)
Detached and Semi-detached Housing (DSDH) (282):
Small and Isolated Building Structures (47)
Small and Isolated Single Houses (1)
Single and Double House Development (201)
Regular House Development (33)
Large Buildings and Industrial Areas (LBIA) (126):
Greenhouse Development (1)
Large Building Complex (56)
Large Halls (20)
Mixed Use: Large Buildings and Hall Development (28)
Mixed Use: Small Buildings and Hall Development (12)
Regular Block Development (RBD) (319):
Chained House Complex (10)
Isolated High Buildings, Parallel-Disposed Buildings (54)
Lined and Block-Edged Development (90)
Perimeter Block Development (10)
Mixed Types of Buildings (46)
Small Buildings Complex (22)
Parks, Squares and Vegetated Areas (PSVA) (278):
Agriculture Fields (4)
Cemeteries (2)
Extensiv Green and Uncultivated Areas (11)
Hedges and Copses (2)
Meadows and Pasture (2)
Small Gardens (16)
Small Green Areas (179)
Squares and Other Green Areas (50)
Urban Groves (4)

Table 5.4: Generalization of the official UST classes. The number of instances and the abbreviation of each class is shown in parenthesis after their names. The five classes written in blank letters are the ones considered in all classification experiments.

5.2.2 Considered Attributes

In this section, the attributes considered in the standard classification of USTs are presented. Together with the attributes computed from the features extracted inside the urban blocks, the following block shape attributes were also considered in the classification: Area, Compactness, Major-minor axis ratio, Orientation, Perimeter, Perimeter-area-ratio.

Table 5.5 presents the composition and configuration attributes of the blocks concerning the lines extracted in them. These attributes were computed twice, i.e. once considering the lines extracted from the intensity image from the ascending orbit and once considering the lines extracted from the intensity image from the descending orbit. Lines were considered to be parallel to each other if their relative orientation is below 15°. They are considered to be perpendicular if their relative orientation is between 75° and 90°.

As explained in Section 4.3.2, networks were created for each urban block connecting the polygon features extracted in it. For each block two networks were created: one connecting the bright and dark polygons extracted from the intensity image from the ascending orbit and another connecting the bright and dark polygons extracted from the intensity image from the descending orbit. Some of these attributes are related to the structure of the network and others to the Moran's I of the nodes attributes. These statistics of the Moran's I were estimated both computationally, i.e. based on 9999 permutations, and analitically based on the assumption of normality. The attributes computed from these networks and considered for classification are the ones listed in Table 5.7. The density (d) of the network is computed as:

$$d = \frac{2m}{n(n-1)}, \tag{5.1}$$

where n is the number of nodes and m is the number of edges in the network.

The attributes computed from the bright and dark areas extracted as CCs through the binarization by thresholding operations are the following: Number of CCs, Number of CCs larger than 200 pixels, Number of CCs larger than 500 pixels, Number of CCs larger than 1000 pixels, Joint CCs area, Mean CC area, Maximum CC area, Standard deviation of the CCs area, Proportion of the CCs area, Maximum compactness of the CCs, Compactness of the largest CC, Maximum length-to-width ratio of the CCs, Minimum perimeter-area ratio of the CCs, Number of CCs with length-to-width ratio higher than 5 and Number of CCs with length-to-width ratio higher than 10.

5.2.3 Attribute Relevance Analysis

In order to gain insight on the expressiveness of the considered block attributes, an attribute relevance analysis was performed using the Random Forest algorithm. As described in Section 4.4.3, the attributes relevance were measured through the mean accuracy decrease strategy. Based on each of the twenty-six training sample sets, a Random Forest model was fitted and intrinsically to it the attributes relevance was computed. The final relevance of each attribute was considered as the average relevance from these twenty-six Random Forest models.

5.2.4 Standard Classifications

The standard classifications served as references for evaluating the accuracy improvement achieved by considering context in the classifications. They are also comprised in the contextual classification models as the association potential functions. The standard classifications were performed, as addressed above, with the Random Forest classifier and by means of comparing the blocks MDPs.

For the creation of each MDP from each block, the input image's 200-quantiles were firstly computed. Then, the quantiles were set as the 200 threshold values T, with each of which CCs were extracted through binarization by thresholding operations, as explained in section 4.2.2. After each thresholding operation, the CC attributes described in the previous section were computed. For each block, six MDPs were created. The types of MDPs are defined by the input image and

Line attributes:
Maximum line length
Mean angle difference between a line and its closest line
Mean angle diff. between a line and the block boundary closest to it
Mean angle diff. between a line and the block boundary most parallel to it
Mean angle diff. between a line and the line most parallel to it
Mean angle diff. between a line and the line most perpendicular to it
Mean distance between a line and its closest line
Mean distance between a line and the line most parallel to it
Mean distance between a line and the line most pependicular to it
Mean distance between a line and the block boundary closest to it
Mean distance between a line and the block boundary most parallel to it
Mean distance between a line and the block boundary most perpendicular to it
Mean orientation of the lines
Mean length of the lines
Min. angle diff. between a line and the block boundary closest to it
Min. distance between a line and the line most parallel to it
Min. distance between a line and the line most pependicular to it
Min. distance between a line and the block boundary most parallel to it
Min. distance between a line and the block boundary most perpendicular to it
Number of lines
Number of lines longer than 50 m
Number of lines longer than 100 m
Std. dev. of angle difference between a line and its closest line
Std. dev. of distance between a line and its closest block boundary
Std. dev. of distance between a line and its closest line
Std. dev. of distance between a line and the line most parallel to it
Std. dev. of distance between a line and the line most perpendicular to it
Std. dev. of the lines length
Std. dev. of the orientation of all lines

Table 5.5: The composition and configuration attributes computed from the lines extracted from each urban block. These attributes were computed once considering the lines extracted from the intensity image from the ascending orbit and once considering the lines extracted from the intensity image from the descending orbit.

the type of CCs extracted. These six types of MDPs are the following: (1) Bright areas from the intensity image from the ascending orbit, (2) Bright areas from the intensity image from the descending orbit, (3) Dark areas from the intensity image from the ascending orbit, (4) Dark areas from the intensity image from the descending orbit, (5) Bright areas from the coherence image from the ascending orbit and (6) Bright areas from the coherence image from the descending orbit. Each MDP was compared with the MDPs of the same type from the classes reference MDPs and the mean similarity values for each of the classes was computed. These values express the similarities of the block to the UST classes in regard to that type of MDP. Using the criterion of highest similarity, six classifications were produced, i.e. one for each MDP type. The overall accuracies of these classifications were then considered as weights of the weighted mean of the similarity values from each block to each class. Finally, a final classification was produced by the highest weighted mean similarity criterion. The weighted mean similarity to each class is also what was used as the association potential in some of the contextual classification models.

Polygon attributes:
Maximum pertinence
Mean area
Mean angle difference between a polygon and its closest block boundary
Mean distance between a polygon and its closest block boundary
Mean distance between a polygon and the polygon most parallel to it
Mean distance between a polygon and the polygon most perpendicular to it
Mean, max. and std. dev. of the polygons area
Mean, max. and std. dev. of the polygons length-to-width
Mean, max. and std. dev. of the polygons compactness
Mean, max. and std. dev. of the polygons orientation
Number of pairs of polygons parallel to each other
Number of pairs of polygons perpendicular to each other
Number of polygons
Number of polygons with area $> 3^{rd}$, 4^{th} and 5^{th} 5-quantiles of the area values
Number of polygons with pertinence > [0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.5]
Number of pol. with pertinence > [3^{rd}, 4^{th}, 5^{th}] of the 5-quantiles of the pertinence values
Std. dev. of the angle diff. between a polygon and its closest block boundary
Std. dev. of the distance between a polygon and its closest block boundary
Std. dev. of the distance between a polygon and the polygon most parallel to it
Std. dev. of the distance between a polygon and the polygon most perpendicular to it
Std. dev. of the polygons orientation
3^{rd}, 4^{th} and 5^{th} 5-quantiles of the polygons area
3^{rd}, 4^{th} and 5^{th} 5-quantiles of the pertinence values

Table 5.6: The composition and configuration attributes computed from the polygons extracted from each urban block. These attributes were computed four times for each block according to the type of polygon (from dark or bright areas) and the satellite orbit (ascending or descending).

The classifications performed with the Random Forest algorithm considered the 129 attributes presented in Tables 5.5 and 5.6, as well as the 16 HOGs attributes. To this vector, the CC attributes presented in the previous section were also inserted. However, instead of extracting the CCs with two-hundred thresholds, as in the case of the generation of MDPs, the CCs were extracted with only three different thresholds. These are the input image's 5^{rd}, 7^{th} and 9^{th} 10-quantiles. This adds up to 45 computed attributes for each of the 6 types of CC presented above. Thus, in total 415 attributes were submitted to the Random Forest classifier, i.e. 129 + 16 + (45 * 6). The number of decision trees from the Random Forest was set to 100. (Although this might seem a low number, exploratory experiments revealed that the performance was never improved when the number of decision trees was set to 1000 or more). The number of attributes evaluated when looking for the best split at each tree bifurcation was set to the root square of the total number of attributes. Attribute evaluation was performed internally to the fitting of the model and the attributes with relevance below the mean relevance value were discarded. Then the model was re-fitted with only the most pertinent attributes.

5.2.5 Contextual Classification Models

In order to facilitate the report of the results on the contextual classification of USTs, the different models presented in Section 4.5.3 were coded according to Table 5.7. These models were evaluated and compared based on twenty-six different training sample sets. As explained in Section 5.2.1, one of these sample sets was collected in supervised way, whereas the other twenty-five were collected randomly. The actual classifications, as mentioned in Section 4.5.3, were in all

Network structure attributes:
Density of the graph
Highest mean membership of two connected nodes
Mean node pertinence
Membership value of the node with highest membership
Number of nodes
Number of edges
Number of edges / number of nodes
Number of edges connecting parallel nodes
Number of edges connecting parallel nodes / number of edges
Number of edges connecting parallel nodes / Number of edges (...)
(...) connecting perpendicular nodes
Number of edges connecting perpendicular nodes
Number of edges connecting perpendicular nodes / number of edges
Std. dev. of the membership from the two connected nodes (...)
(...) with highest membership mean
Network Moran's *I* of attributes:
Area
Distance to closest block boundary
Length to width ratio
Orientation
Orientation difference to closest block boundary
Rectangular Fit
For the attributes above:
Expected I based on random permutation of the values
Expected I under normality assumption
Difference between I and expected I based on random permutations
Difference between I and expected I under normality assumption
P-value of I based on random permutation of the values
P-value of I under normality assumption (one-sided)

Table 5.7: Attributes from the networks connecting the polygon features inside the blocks. For each block two networks were created: one connecting the bright and dark polygons extracted from the intensity image from the ascending orbit and another connecting the bright and dark polygons extracted from the intensity image from the descending orbit.

cases performed by minimizing the models energy functions with the Max-Sum LBP inference algorithm.

Each of the models from Table 5.7 were tested on graph structures defined by the four criteria presented in Section 4.5.3. Regarding criterion 1, two fixed distance thresholds th were tested, namely, of 140 and 240 m. 142.4 is the 0.9 quantile of the blocks attribute 'length of longest line', whereas 244.7 is the 0.8 quantile of the blocks length. Hence, these two numbers were rounded down to 140 and 240 and then considered as the fixes distance threshold. Regarding criteria 2 and 3, a fixed upper distance threshold th' of 300 m was defined. The number of k nearest-neighbors considered when applying criterion 3 was of 3.

Code	Energy Function	Interaction Potentials	Type
M1	Eq. 4.19	Eq. 4.21	MRF
M2	Eq. 4.27	Eq. 4.21	MRF
M3	Eq. 4.29	Eq. 4.30	MRF
M4	Eq. 4.22	Eq. 4.23	CRF
M5	Eq. 4.22	Eq. 4.24	CRF
M6	Eq. 4.25	Eq. 4.26	CRF
M7	Eq. 4.28	Eq. 4.23	CRF

Table 5.8: Coding of the contextual classification models proposed and tested in this thesis.

6 Results

In this chapter, the results obtained in this thesis are presented. It has three major sections. The first one briefly analyses the relevance of the block attributes considered in this thesis. It also compares the overal accuracies of USTs classifications performed with the two different methods, i.e. Random Forest and MDP comparison, and with sets of attributes derived from images obtained at one and both looking directions. The second section analyses the performance of the different contextual classification models proposed in this thesis and takes a look at the misclassifications of specific representative classifications. The third and last section analyses the spatial autocorrelation of the official UST map of the study area according to neighbourhood relations defined by different criteria.

6.1 Attribute Relevance and Standard Classifications Analysis

6.1.1 Attribute Relevance

This section presents the most relevant attributes considered in this work. The computation of the attributes relevance was performed as described in Section 5.2.3. Table 6.1 shows from each type of attributes the eight best-ranked ones. Before their names, the abbreviations indicate from which image it was computed, i.e. whether from the coherence (Coh.) or intensity (Int.) image from the ascending (Asc.) or the descending (Des.) looking direction. It can be observed that the most relevant groups of attributes are the ones derived from the extracted bright and dark areas as well as from the extracted polygons. The single most relevant attributes however are surprisingly those that concern the blocks shape (area, perimeter and perimeter-area ratio). This is presumably because the values of all other attributes are to some extent dependent on the blocks size and shape. The least relevant group of attributes turned out to be the network attributes, followed by some of the HOGs and lines ones.

From the third-six attributes shown in Table 6.1, exactly eighteen of them were derived from the ascending and eighteen from the descending images, what supports the assumption that it is important to consider imagery obtained at both orbits. Among the HOG attributes, however, the best ranked ones were computed from the DEM generated from the ascending SLC images. In fact, this interferometric pair has a higher height of ambiguity (Table 5.2), what makes the problem of phase unwrapping less critical and thus increases the expressiveness of this DEM in comparison to the one derived from the descending image pair. Regarding the generic types of the attributes, i.e. composition or configuration, it can be seen from Table 6.1 that the majority of the lines attributes are configuration ones (6 out of 8), whereas among the polygon attributes only the two most relevant ones are configuration attributes.

6.1.2 Standard Classifications

In this section, the overal accuracies from the standard USTs classifications are reported. Based on each of the twenty-five different sets of training samples collected randomly, six classifications

Name	Relevance
Polygon Attributes:	
Asc.: Mean orientation angle of bright polygons	12,0 %
Asc.: Mean orientation angle of dark polygons	11,2 %
Asc.: Mean compactness of dark polygons	10,6 %
Des.: Maximum length-to-width ratio of bright polygons	10,3 %
Asc.: Mean length-to-width ratio of bright polygons	10,3 %
Des.: Mean area of bright polygons	10,1 %
Des.: Number of bright polygons with pertinence $> 3^{rd}$ 5-quantile	9,6 %
Des.: Mean length-to-width ratio of bright polygons	9,3 %
Network Attributes:	
Des.: Highest mean membership of two connected nodes	5,5 %
Des.: Number of edges	2,8 %
Des.: Mean node pertinence	2,7 %
Des.: Diff. between obs. I and exp. I under norm. of the nodes orientation	1,7 %
Asc.: Mean node pertinence	1,6 %
Asc.: I p-value under norm. for orient. diff. between pol. and clos. blk. bnd.	1,5 %
Asc.: Diff. between obs. and exp. I under norm. for the nodes orient.	1,4 %
Des.: Diff. between obs. and exp. I based on permutations for the nodes area	1,4 %
Lines Attributes:	
Asc.: Min. angle diff. between a line and the block boundary closest to it	8,7 %
Des.: Min. distance between a line and the block boundary most parallel to it	7,7 %
Asc.: Std. dev. of distance between a line and the line most perpendicular to it	7,6 %
Asc.: Mean length of lines	7,6 %
Des.: Std. dev. of distance between a line and its closest block boundary	7,4 %
Asc.: Maximum line length	7,3 %
Asc.: Std. dev. of distance between a line and its closest block boundary	7,2 %
Des.: Mean distance between a line and the block boundary closest to it	6,8 %
HOGs and Block Shape Attributes:	
Block perimeter-area ratio	19,8 %
Block area	15,1 %
Block perimeter	14,4%
Asc.: HOG 2 (22° - 44°)	8,9 %
Asc.: HOG 1 (0° - 22°)	7,3 %
Block orientation	7,2 %
Asc.: HOG 3 (44° - 66°)	7,1 %
Asc.: HOG 4 (66° - 88°)	6,8 %
Dark and Bright Areas Attributes:	
Asc., Int.: Number of CCs $>$ 1000 pxl. - 9^{th} 10-quantiles	12,8 %
Des., Coh.: Number of CCs $>$ 500 pxl. - 7^{th} 10-quantiles	11,5 %
Des., Int.: Area proportion of bright CCs - 7^{th} 10-quantiles	10,9 %
Des., Coh.: Area proportion of dark CCs - 9^{th} 10-quantiles	10,5 %
Asc., Int.: Number of CCs $>$ 1000 pxl. - 5^{th} 10-quantiles	10,3 %
Des., Int.: Number of CCs $>$ 1000 pxl. - 9^{th} 10-quantiles	10,0 %
Des., Int.: Number of CCs $>$ 1000 pxl. - 7^{th} 10-quantiles	9,7 %
Des., Int.: Area proportion of bright CCs - 9^{th} 10-quantiles	9,6 %

Table 6.1: The most relevant attributes from each type. Before their names, the abbreviations indicate from which image it was computed, i.e. whether from the coherence (Coh.) or intensity (Int.) image from the ascending (Asc.) or the descending (Des.) looking direction.

were performed combining the two different methods with the three different attribute sets. The sets of attributes are composed of the ones derived from the ascending, the descending imagery as well as from attributes derived from imagery obtained at both looking directions. Figure 6.1 shows the plot graphs from the overal accuracies of these classifications performed as explained in Section 5.2.4. The experiments Ids identify the twenty-five different sample sets used to train the models. It can be seen from Figure 6.1 that considering attributes from both image datasets always causes an increase in the overall accuracy. The highest overal accuracies achieved by both methods are very similar (of approximately 69%), as highlighted by the dashed lines in Figure 6.1. It can be seen also that classifications performed with attributes and MDPs derived from the ascending images achieved most of the twenty-five times higher overal accuracies as those performed with attributes and MDPs from the descending images.

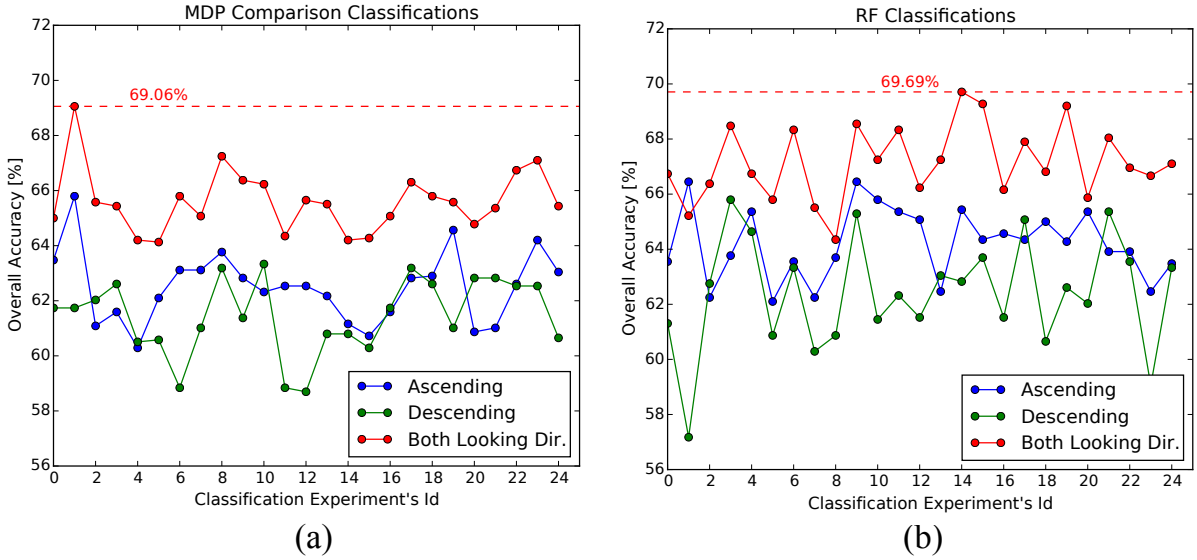

Figure 6.1: Comparison of classifications performed with different methods and attribute sets. Twenty-five different training samples sets were randomly selected. With each sample set, six classifications were performed combining the two different methods with the three different attribute sets. The attribute sets are defined based on whether the attributes were derived from the ascending, the descending images or from the images acquired at both looking directions. a) Plot of the accuracies of classifications performed with the MDP comparison strategy. b) Plot of the accuracies of classifications performed with the Random Forest algorithm.

Table 6.2 presents the first-order statistics of the overal accuracies from Figure 6.1. Regarding the Random Forest method, in comparison to classifications performed considering only attributes from the ascending images, an average accuracy improvement of 3% was achieved when attributes from both image datasets were considered. The accuracy improvement is in average of 5% in comparison to classifications performed only with attributes from the descending images. Regarding the classification strategy of comparing the blocks MDPs, if the MDPs from both image datasets are considered, an average accuracy improvement of respectively 3% and 4% is obtained in comparison to classifications performed with MDPs from the ascending and the descending datasets alone. It is also noticeable from Table 6.2 that in general the Random Forest method performed slightly better than the strategy of comparing the blocks MDPs.

Looking Direction	Method	Min. (%)	Mean (%)	Max. (%)	Std. Dev.
Ascending	Random Forest	61.30	64.09	66.44	1.35
	MDPs Comp.	60.02	62.48	65.79	1.27
Descending	Random Forest	57.17	62.41	65.79	1.96
	MDPs Comp.	58.69	61.45	63.33	1.34
Both Looking Dir.	Random Forest	64.34	67.15	69.69	1.30
	MDPs Comp.	64.13	65.61	69.06	1.11

Table 6.2: First-order statistics of the overall accuracies from the classifications shown in Figure 6.1. An average accuracy increase of up to 5% is obtained if attributes from both image datasets are considered.

6.2 Contextual Classifications

In this section, the results from the performed contextual classifications are presented. The first subsection shows the extent to which the overal accuracies increase when the context of each block is considered. The performance of each of the considered models, which differ according to their parameterizations and the graph structures on which they are applied, is reported and compared. The second subsection approaches the misclassifications of the standard and contextual classifications performed based on samples collected in supervised way.

6.2.1 Parameterizations and Structures Comparison

Regarding their parameterizations, seven different types of models, coded according to Table 5.8, were tested each on five different neighbourhood structures. Therefore, thirty-five different contextual classifications were performed with each training sample set. This section first evaluates the classifications performed based on the samples collected in supervised way. Following, the comparisons are extended to the classifications performed based on twenty-five different sample sets collected randomly.

As discussed in Section 4.5.3, all the models considered in this thesis contain a parameter λ which weights the influence of the interaction factors and thus establishes the strength of the influence of context in the classification. This parameter can be defined empirically based on the ratio between the number of blocks, which is the number of association factors, and the number of pairwise neighbouring relationships between blocks, which is the number of pairwise interaction factors. Table 6.3 shows this ratio and the number of interaction factors for the five types of neighbourhood structures considered in this thesis and presented in Section 5.2.5. Instead of defining λ empirically however, an exhaustive search of the optimal λ was performed by analysing the overal accuracies obtained with a hundred different λ values from 0 to 1. The overal accuracies refer to the classifications performed with training samples collected in supervised way. Figure 6.2 shows for the four different CRF models considered and the five different neighbourhood structures the overal accuracies achieved with each of this *lambda* values. A dashed line indicates the highest accuracies achieved by each CRF model. The continuous black line indicates the overall accuracy of the standard classification, which was performed with the Random Forest classifier and based on samples collected in supervised way. The class membership values output by the Random Forest algorithm for each block were considered as the association potentials, as explained in Section 4.5.3. It can be noticed that M7 is the model that performed the best with all graph structures, followed, with the exception of Figure 6.2(a), by M4.

Figure 6.3 shows for each neighbourhood structure and for model M7 and each of the three MRF models considered the profiles of λ values and respective overal accuracies achieved. Just as in Figure 6.2, the dashed lines indicate the highest overall accuracy achieved by each model

	Int. Factors	Ass./Int. Factors
Fixed Radius of 140 m	2233	0.75
3-Nearest Neighbours	1487	1.13
Adaptive Radius	2360	0.71
Fixed Radius of 240 m	7178	0.23
Block-adjacency	4280	0.39

Table 6.3: Number of interaction factors from the five different neighbourhood structures and the ratio between association and interation factors. The number of association factors, and hence of blocks from the study site, is of 1695.

and the continuous black line indicates the overall accuracy of the standard classification. It can be observed that none of the MRF models performed better than the best CRF model (M7). Among the MRF models, with the exception of Figure 6.3(b), M1 and specially M2 are the ones that performed best.

The maximum overal accuracies achieved by each model and neighbourhood structure combination are reported in Table 6.4. The highest overall accuracy achieved with each different graph structure is highlighted in blue. All models except M3 achieved the highest overal accuracies when applied on the graph structure defined by a fixed radius of 240 m. M3 performed better when the graph structure is defined by the criterion of block adjacency. Among all MRF models (M1, M2 and M3) tested with the structure defined by the criterion of 3-nearest neighbours, M3 is the one that performed best. When considering this type of structure, M3 also achieved higher overal accuracies than the other CRF models (M4, M5, M6, M7) with the exception of M7. In fact, M7 is the model that achieved the highest overal accuracies when considering any of the five graph structures. It is then logical that the overall highest accuracy among all models and structures combination was achieved by M7 with a graph/neighbourhood structure defined by a fixed radius of 240 m. Given that M7 performed better than all other CRF models, it can be concluded that apparently it worths the effort of inserting the additional association factors defined by the outcomes of the MDPs-based classification. This is partially confirmed by the overal accuracies achieved by the MRF models. M2, which also includes these extra association factors, performed better than the other two MRF models when three out of five graph structures are considered. Furthermore, among the MRF models, M2 obtained the highest mean and maximum overall accuracy over all five structures. It can be observed from Table 6.4 that, except when considering the 3-nearest neighbours structure, third-order interaction factors do not increase the accuracy. (It does however increases processing time!). It should be stressed that in general the CRF models performed better than the MRF ones. Also, despite being the simplest MRF and CRF models, M1 and M4 achieved decent overal accuracies, specially when applied on structures defined by the criteria of block adjacency and of a fixed radius of 240 m. In this former case, M1 and M4 actually achieved the highest overal accuracies among the MRF and CRF models respectively. Surprisingly though, M5 and M6 only performed better than M1 and M3 considering the mean and maximum statistics.

In order to evaluate the consistency of the models, their performances on twenty-five different classification experiments were analysed. Each experiment differs only on the set of training samples, which, as mentioned, was in all cases collected randomly. At each of the classification experiments, the best λ parameter for each model and graph structure was searched. This search consisted in firstly identifying the best λ value for the respective model and graph structure according to the classifications based on the training samples collected in supervised way. Following, each model was tested with twenty λ values centered on this value of reference and differing on 0.01. For example, it can be seen from Figure 6.2(a) that model M4 achieved the highest accuracy

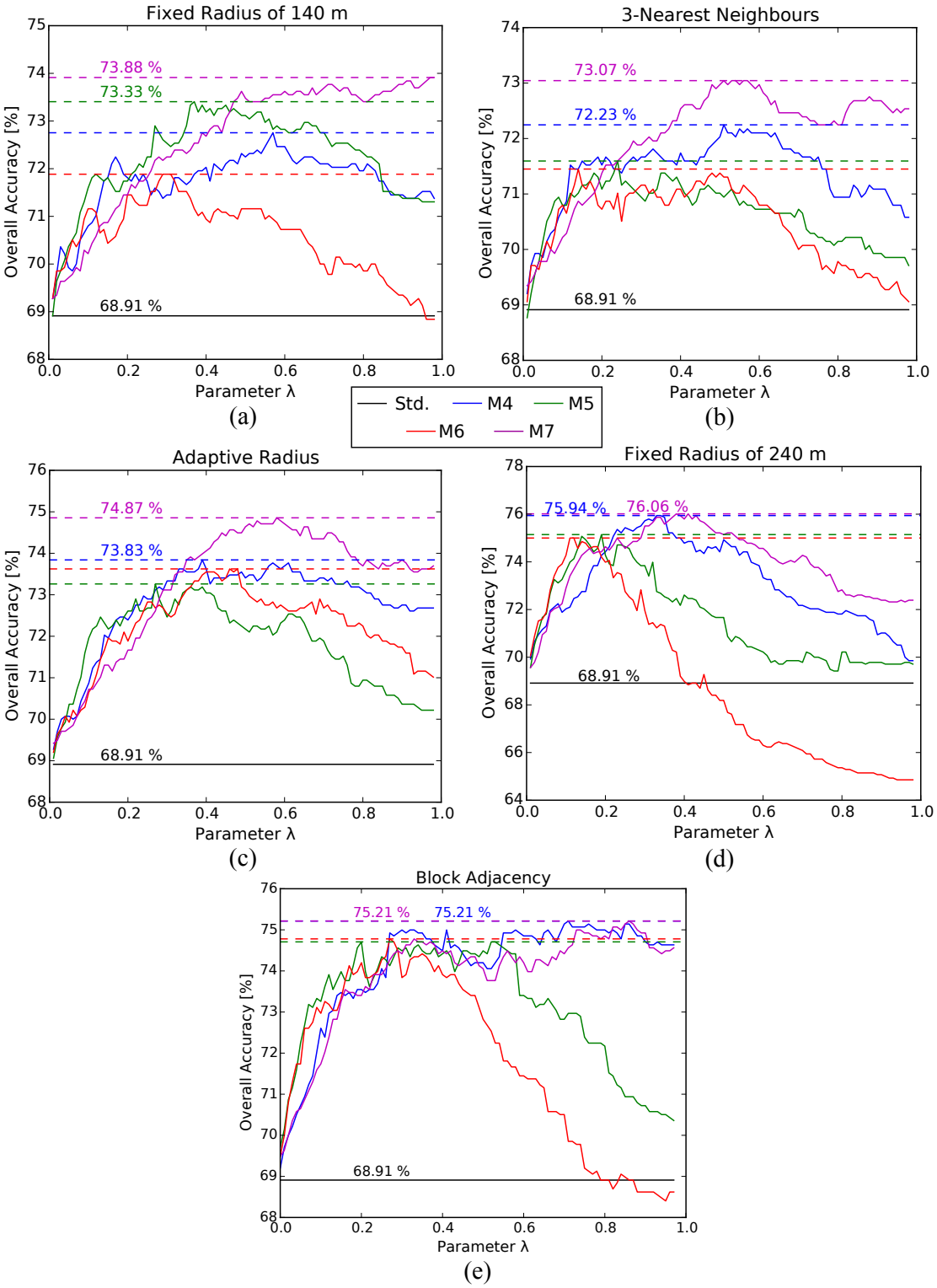

Figure 6.2: Overall accuracies achieved by setting the *lambda* parameter to a hundred values between 0 to 1 for the CRF models considered. a) to e) Results for the graph structures defined by the five different criteria considered in this thesis. A dashed line indicates the highest accuracies achieved by each CRF model. The continuous black line indicates the overall accuracy of the standard classification, which was performed with the Random Forest classifier and based on samples collected in supervised way.

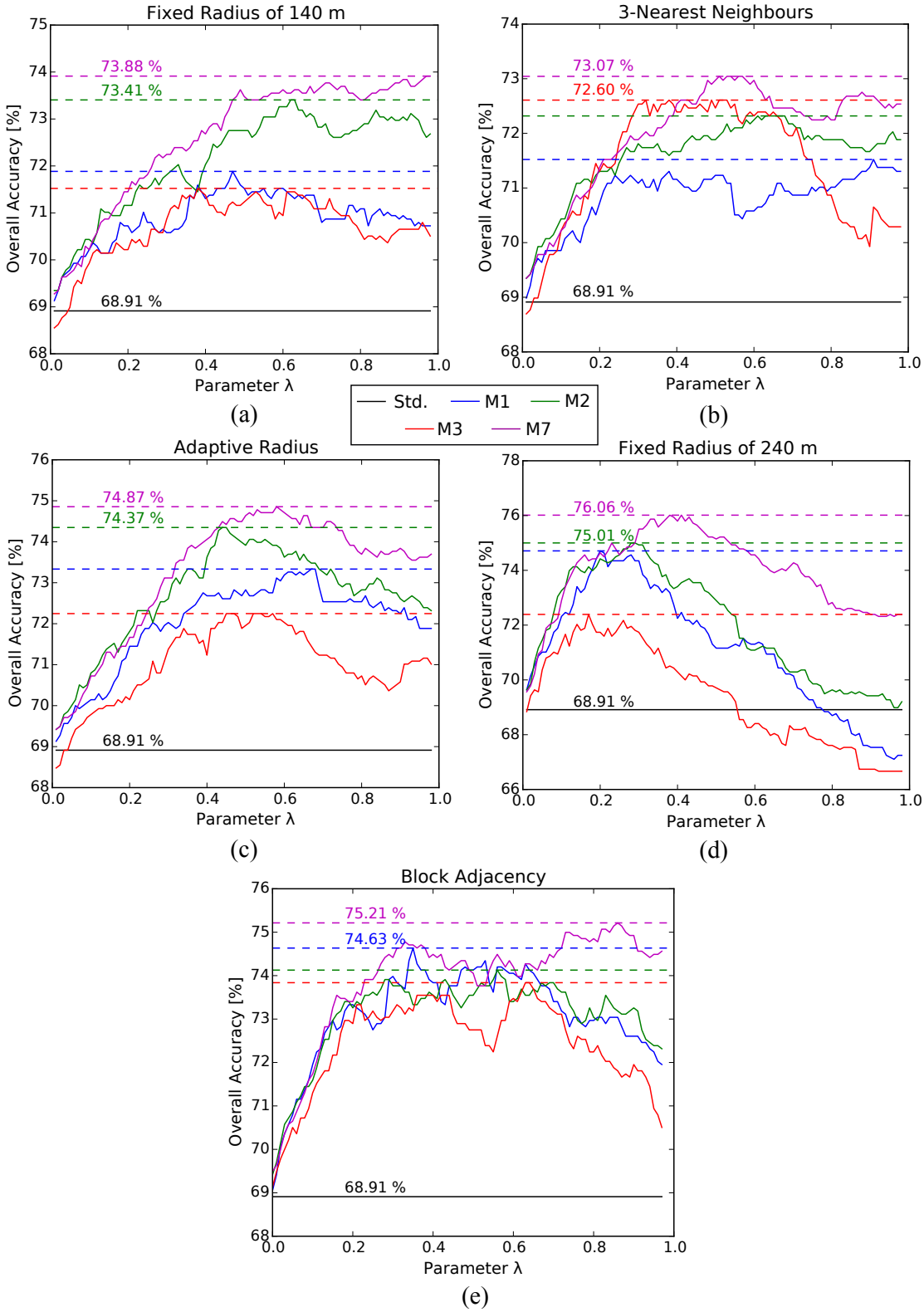

Figure 6.3: Overall accuracies achieved by setting the *lambda* parameter to a hundred values between 0 and 1 for M7 and the MRF models considered. a) to e) Results for the graph structures defined by the five different criteria considered in this thesis. A dashed line indicates the highest accuracies achieved by each model. The continuous black line indicates the overall accuracy of the standard classification, which was performed with the Random Forest classifier and based on samples collected in supervised way.

	M1	M2	M3	M4	M5	M6	M7	Mean	Max.
Fix. rad. of 140 m	71.88	73.40	71.52	72.75	73.40	71.88	73.91	72.67	73.91
3-nearest neighb.	71.52	72.31	72.60	72.24	71.59	71.44	73.04	72.10	73.04
Adaptive rad.	73.33	74.34	72.24	73.84	73.26	73.62	74.85	73.64	74.85
Block Adj.	74.63	74.12	73.83	75.21	74.70	74.77	75.21	74.63	75.21
Fixed rad. of 240 m	74.71	75.00	72.39	75.94	75.14	75.00	76.01	74.88	76.01
Mean	73.21	73.83	72.51	73.99	73.61	73.34	74.60		
Max.	74.71	75.00	73.83	75.94	75.14	75.00	76.01		

Table 6.4: Overall accuracies achieved by the different contextual classification models applied at different graph/neighbourhood structures. The contextual classifications were performed based on training samples collected in supervised way. The highest overall accuracy achieved with each structure is highlighted in blue.

with a λ of about 0.6. Hence, at each of the twenty-five experiments conducted with M4 and the structure defined by the fixed radius of 140 m, the best λ was found by testing the twenty values between 0.5 and 0.7.

Figure 6.4 depicts the plot graphs with the overal accuracies achieved by each model and graph structure at each of these twenty-five classification experiments. It can be seen that the standard classifications overal accuracies vary significantly according to the experiment. This is reasonable since the number of samples from each class is relatively low (63) and the inner-class variability is high. Surprisingly, however, the contextual classification model that performs best at each experiment varies a lot for each of the considered graph structures. This leads to the conclusions that there is no overall best model and that, instead, the models performances depend strongly on the set of training samples, based on which the standard classification and thus the association potentials are defined.

The performance and consistency of the models can also be evaluated based on Table 6.5. It presents for each graph structure the first-order statistics of the models overal accuracies achieved throughout the twenty-five experiments. It can be observed that indeed the graph structure with which the highest overal accuracies are achieved is the one defined by a fixed radius of 240 m, followed by the ones defined by the blocks adjacencies and by the blocks adaptive radius. In Table 6.5, the highest mean and maximum overal accuracies achieved with each structure is highlighted. When contextual classification is performed with the structure defined by the blocks 3-nearest neighbours, just as when the samples were collected in supervised-way, the highest overall accuracy is achieved with model M3, which contains third-order interaction factors. When the structure was defined by the blocks adjacencies, the highest mean and maximum overal accuracies were achieved by M6, which takes into consideration the blocks shape similarity for defining their potential of belonging to the same class. Differently from when the samples were collected in supervised way, the overall highest mean and maximum overal accuracies were achieved by model M4 instead of by model M7. However, it can be observed that M7 is after M4 the model that most of the times performed best. Regarding the MRF models, M1 overcome most of the times the performances from M2 and M3, what also does not agree with the results shown in Table 6.4, where it can be seen that M2 is the model which performed best for three out of five graph structures.

Table 6.5 also presents the standard deviations from the overal accuracies achieved by each model when applied over each of the five graph structures. The overal accuracies standard deviations are never higher than 1.89 and do not vary a lot, which means that the mean overall accuracy improvement achieved by considering context in the classification is representative. The mean overall accuracy improvement is also shown in Table 6.5 for each model and structure com-

Mean Overall Accuracy (%):

	M1	M2	M3	M4	M5	M6	M7
Fix. rad. of 140 m	72.77	72.81	71.68	73.25	72.91	72.67	73.25
3-nearest neighb.	71.71	71.51	72.42	71.84	71.80	70.91	71.92
Adaptive rad.	73.64	73.30	73.00	74.03	73.75	73.47	73.90
Block Adj.	74.82	74.44	73.48	74.24	73.86	74.80	74.22
Fixed rad. of 240 m	75.17	75.16	73.61	75.93	75.51	75.35	75.48

Maximum Overall Accuracy (%):

	M1	M2	M3	M4	M5	M6	M7
Fix. rad. of 140 m	75.07	75.43	74.56	75.86	75.65	75.28	75.94
3-nearest neighb.	73.91	74.20	75.43	73.69	74.05	73.62	73.55
Adaptive rad.	76.08	75.72	75.36	77.39	76.44	76.88	76.30
Block Adj.	76.81	76.30	75.72	76.66	76.37	76.95	76.23
Fixed rad. of 240 m	77.46	77.02	75.79	78.26	77.39	76.95	77.68

Standard Deviation:

	M1	M2	M3	M4	M5	M6	M7
Fix. rad. of 140 m	1.22	1.38	1.31	1.44	1.39	1.41	1.30
3-nearest neighb.	1.18	1.17	1.25	1.26	1.04	1.22	1.01
Adaptive rad.	1.27	1.33	1.26	1.53	1.40	1.51	1.39
Block Adj.	1.35	1.26	1.89	1.38	1.34	1.29	1.32
Fixed rad. of 240 m	1.40	1.27	1.20	1.22	1.22	1.08	1.27

Mean Accuracy Improvement (%):

	M1	M2	M3	M4	M5	M6	M7
Fix. rad. of 140 m	5.47	5.56	4.36	5.98	5.63	5.35	5.98
3-nearest neighb.	4.39	4.22	5.10	4.50	4.48	3.56	4.61
Adaptive rad.	6.37	6.05	5.69	6.76	6.48	6.18	6.59
Block Adj.	7.50	7.12	6.16	6.92	6.54	7.48	6.90
Fixed rad. of 240 m	7.91	7.91	6.30	8.63	8.23	8.06	8.19

Table 6.5: First-order statistics of the overall accuracies from the twenty-five contextual classification experiments. The highest mean and maximum overal accuracies achieved with each structure is highlighted.

bination. It can be seen from this part of the table that applying any of the CRF models (M4, M5, M6 and M7) over the structure defined by the fixed radius of 240 m is expected to deliver the highest overall accuracy improvements.

6.2.2 Misclassification Analysis

This section briefly discusses the missclassifications from the standard and best contextual classifications obtained based on the samples collected in supervised way. As shown by Table 6.4 and Figures 6.2 and 6.3, the best contextual classification was achieved with model M7 applied on the neighbourhood structure defined by a fixed radius of 240 m. Figure 6.5 shows this and the standard classification, as well as the official UST map of the study site generalized to the five classes according to Table 5.4. An overall accuracy improvement of approximately 7,2% (from 68,9% to 76,06%) was achieved by considering contextual relationships in the classification. It can be seen from Figure 6.5 that a smoothing effect is obtained by considering context in the classification. This was expected, since the energy functions of which the models are consisted are penalized

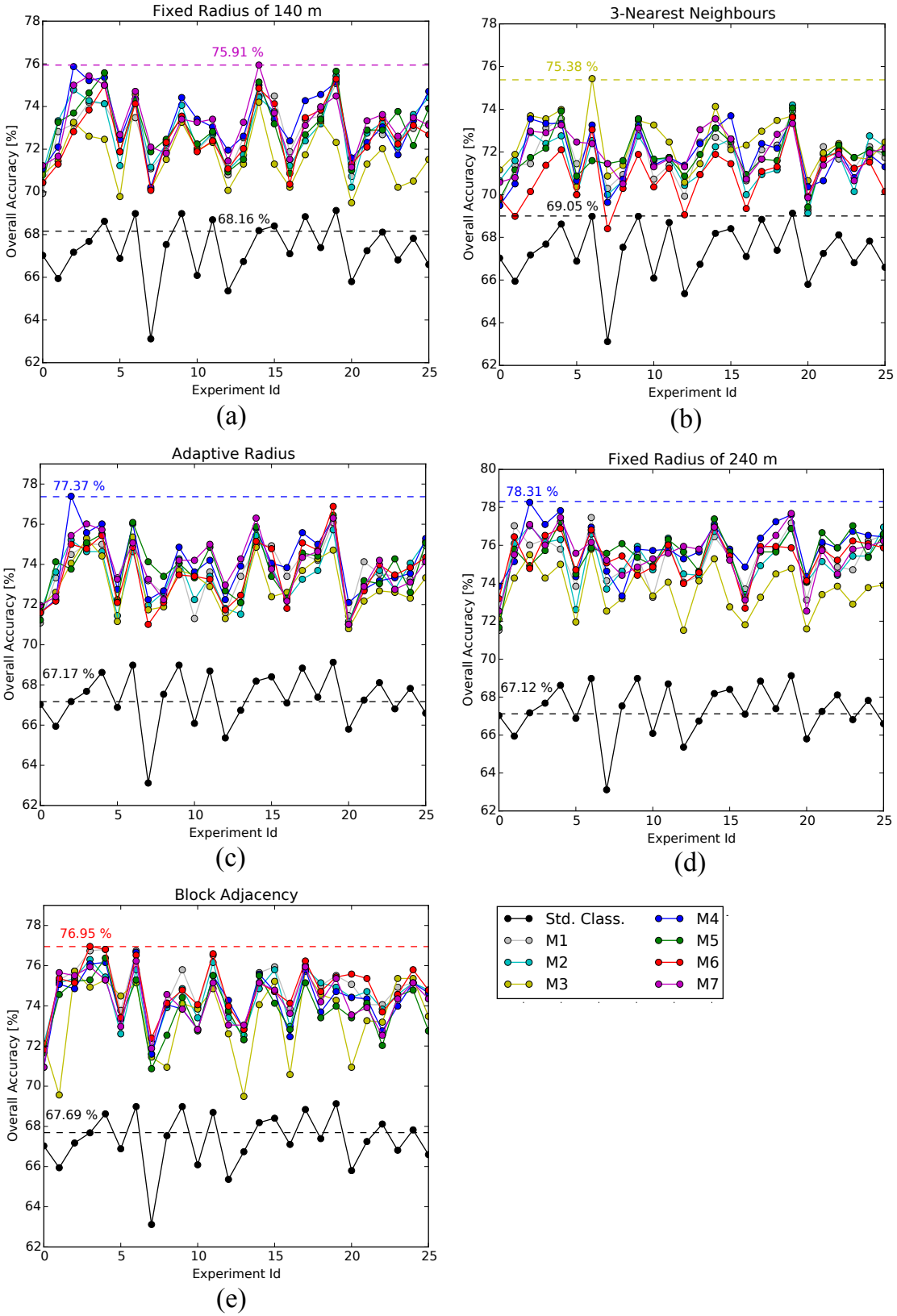

Figure 6.4: Plot graphs with the overall accuracies achieved by each model and graph structure at each of the twenty-five classification experiments. The experiments differ only on the training sample sets. a) to e) Plot graphs from the different graph/neighbourhood considered.

when two neighbouring blocks assume different classes. This smoothing effect causes most of the times the correction of misclassifications from the standard classification. However, at times it causes also incorrect class assignment changes. Examples of these two effects are shown in Figure 6.6.

In order to further analyse the percentage of blocks that had successfull and unsuccessfull class changes, Table 6.6 was created. It shows for model M7 applied over each of the five considered graph structures the percentange of the four possible before-and-after classification status, which are the following: correct-false (CF), false-false (FF), false-correct (FC) and correct-correct (CC). These percentages are shown only for when model M7 is applied, because it is when the highest overall accuracy improvement is achieved for each of the five graph structures, as it can be seen in Table 6.4. Table 6.6 shows that, for all graph structures, the percentages of CF is always lower than FC, indicating that there was more successfull class changes than unsuccessfull ones. The difference between FC and CF is largest for the structure defined by a fixed radius of 240 m. The structure which yields the most class assignment changes is the one defined by the blocks adjacencies.

Misclassifications are most commonly analysed based on a confusion matrix. Table 6.7 shows the confusion matrices of the standard classification as well as of the contextual classification shown in Figure 6.5. As mentioned, the contextual classification shown in Figure 6.5 is the one produced with model M7 and the graph structure defined by a fixed radius of 240 m. It yielded an increase in the Kappa index of 0.09 (from 0.57 to 0.66). Aiming to spare space, the names of the classes in Table 6.7 and in the following ones are written in abbreviated form. These abbreviations are first indicated in Table 5.4 and are adopted for the rest of this section.

Regarding the confusion matrix of the standard classification in Table 6.7, it can be seen that classes PVA, RBD and DSDH get mutually confused. This is probably because these are classes where vegetation occupies frequently a good proportion of the blocks area. Classes DBD, RBD and LBIA get also mutually confused presumably because they share the trait of having buildings which sometimes may appear with similar shapes and densities. Regarding the confusion matrix of the contextual classification, Table 6.7 shows that the user's accuracy has declined only for classes PVA and LBIA. This might be because the distribution of the blocks belonging to these classes are not as clustered as the others, what makes them more vulnerable to the inherent smoothing effect of the contextual classification. If the association potential of the blocks correctly classified as PVA or LBIA is not strong enough, it is likely that they will suffer incorrect class changes. Morevover, class LBIA is the most heterogeneous one concerning its blocks shape and appearance in the SAR images. These might be reasons for this class reaching the lowest user's accuracy. The producer's accuracy on the other hand has increased for all the five classes.

Another analysis presented in this section refers to relationship between misclassifications and mixed blocks, i.e. blocks originally assigned to more than one UST class. As mentioned, according to the official UST map of this thesis study site, 21.89% of the 1695 urban blocks are mixed. In the standard classification performed with the training samples collected in supervised way, the percentage of mixed blocks among the misclassified ones is of 28.67%, whereas the percentage of misclassified blocks among the mixed ones is of 33.15%. These percentages from the contextual classifications performed with model M7 and the five considered graph structures can be seen in Table 6.8. According to it, these percentages do not change considerably depending on the graph structure. It also leads to the conclusion that the misclassifications are only partially caused by mixed blocks and that approximately two thirds of them are correctly classified, what is not so different than the approximately 66% to 75% correctness of these classifications.

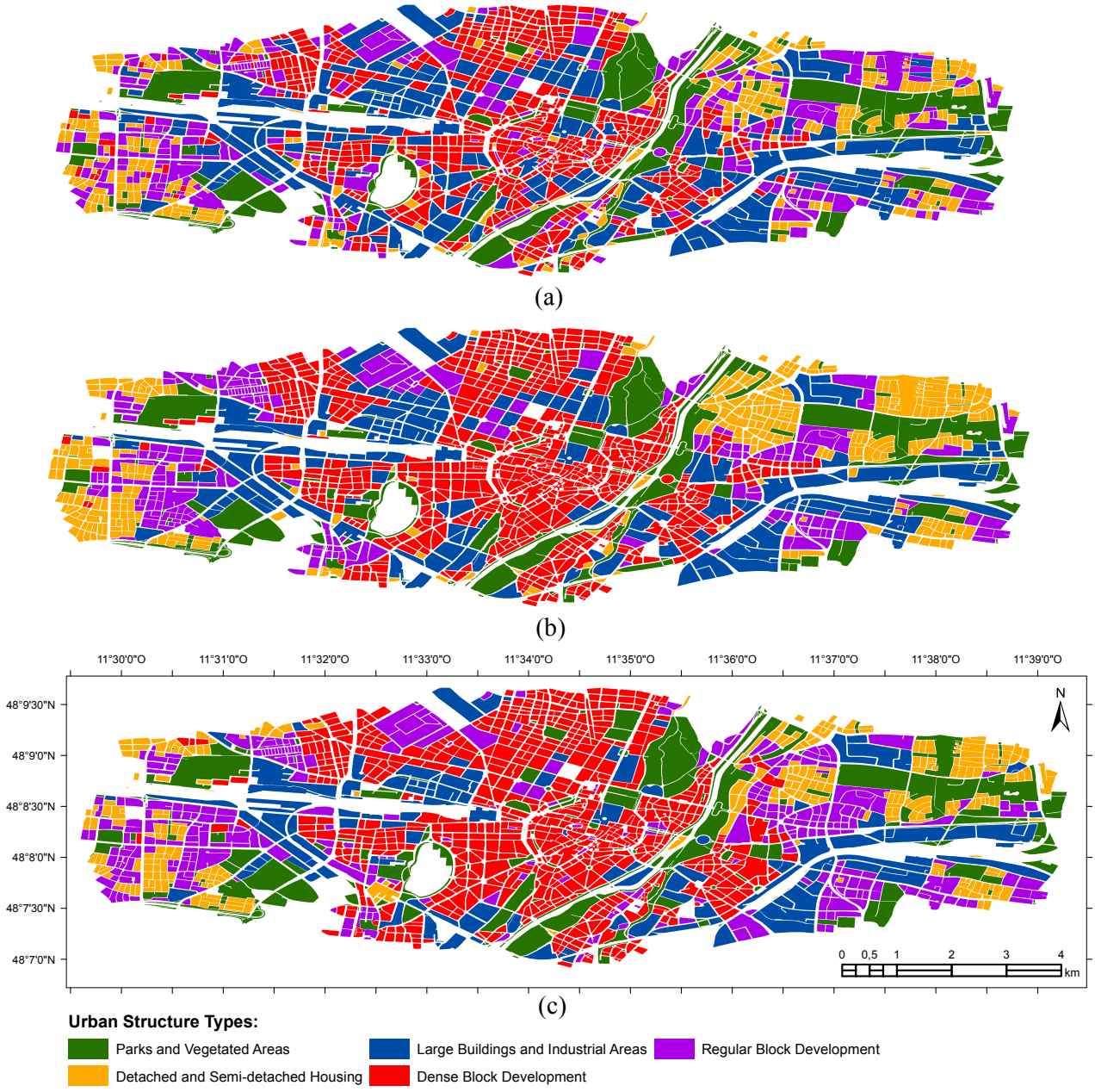

Figure 6.5: Standard and contextual classification and ground-truth map from the study site. a) Standard classification performed by the Random Forest algorithm based on the training samples collected in supervised way. b) Contextual classification performed with model M7 and the graph structure defined by a fixed radius of 240 m. c) the ground-truth UST map from the study site for the five general UST classes considered. The overal accuracies of the standard and contextual classifications are of 68,9% and 76,06% respectively.

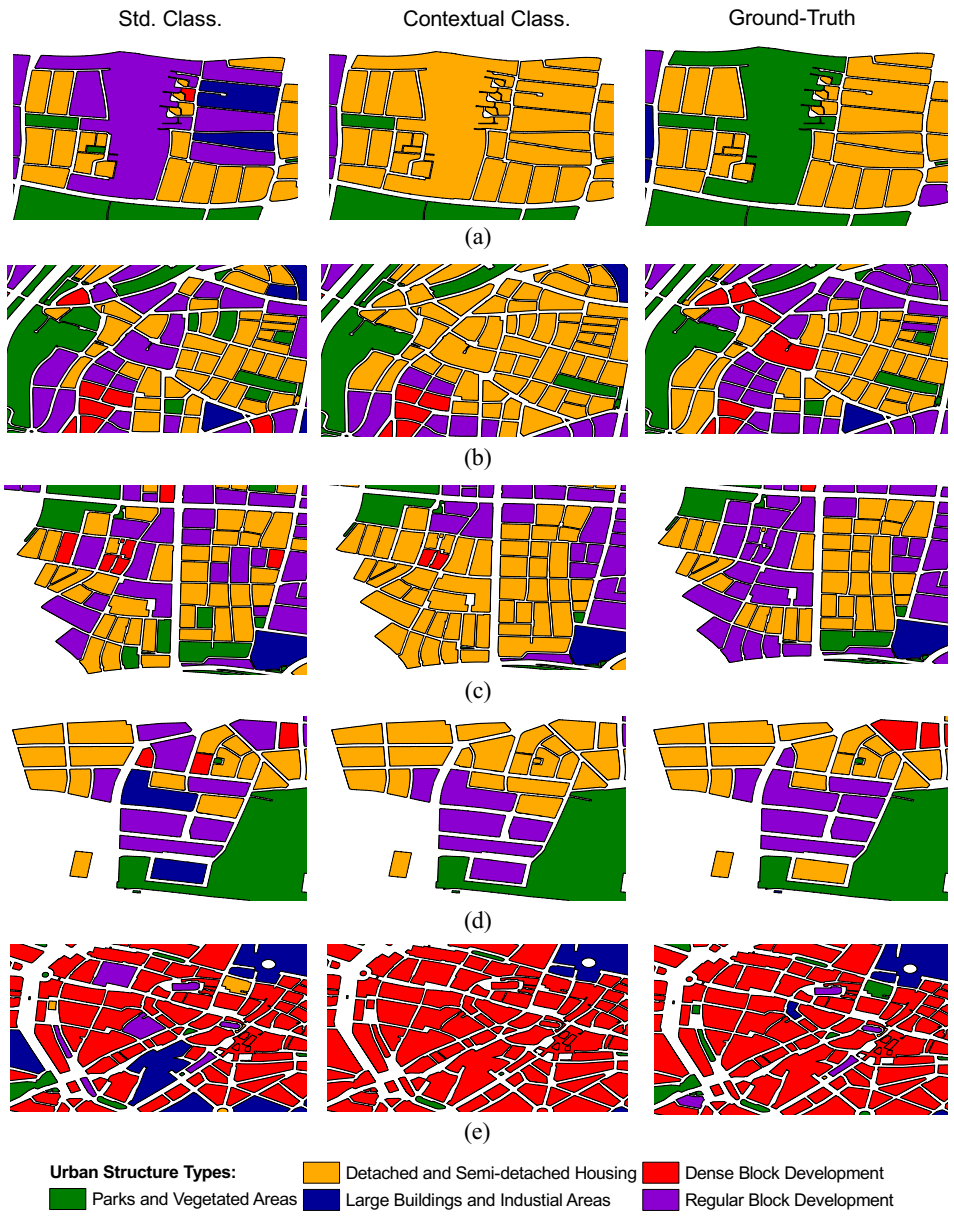

Figure 6.6: Details of the standard and contextual classifications as well as of the ground-truth map. a) to e) Examples of successfull and unsuccessfull class changes caused by the smoothing effect from the contextual classification.

	CF	FF	FC	CC
Fixed radius of 140 m	4.8%	17.4%	7.9%	69.8%
3-neareast neighbours	1.9%	20.2%	5.0%	72.7%
Adpative radius	2.7%	18.1%	7.1%	71.9%
Block adjacency	9.8%	13.09%	12.21%	64.83%
Fixed radius of 240 m	5.2%	15.4%	9.8%	69.4%

Table 6.6: Percentages of the before-and-after class assignment changes regarding the contextual classifications performed with model M7 and the five graph structures considered in this thesis. Note: CF: correct-false, FF: false-false, FC: false-correct, CC: correct-correct. For all five graph structures the percetages of CF is always lower than FC, indicating that there was always more successfull class changes than unsuccessfull ones.

	Mapped Class							
	PVA	DSDH	LBIA	DBD	RBD	Total	User's Acc. (%)	Prod.'s Acc. (%)
Ground Truth								
Standard classification:								
PVA	184	10	3	3	15	215	85.58	84.40
DSDH	22	144	6	12	35	219	65.75	57.60
LBIA	2	5	42	6	7	62	67.77	30.00
DBD	3	25	61	491	48	628	78.18	85.68
RBD	7	66	28	64	91	256	35.54	46.42
Total	218	250	140	573	196			
Contextual classification:								
PVA	165	17	3	18	12	215	76.74	89.18
DSDH	9	177	2	4	27	219	80.82	62.54
LBIA	2	7	41	9	3	62	66.12	43.61
DBD	4	13	26	566	19	628	90.12	86.14
RBD	5	69	22	60	100	256	39.06	62.11
Total	185	283	94	657	161			

Table 6.7: Confusion matrix of the standard and contextual classifications performed with training samples collected in supervised way. The contextual classification was performed with model M7 and a graph structure defined by the criterion of fixed radius of 240 m. The overall accuracies achieved by the standard and the contextual classifications are of 68.98 % and 76.01 % respectively. The Kappa index improved from 0.57 to 0.65.

	Mix. blks. among misclass.	Misclass. among mix. blks.
Fixed radius of 140 m	31.30%	31.81%
3-nearest neighbors	30.05%	30.46%
Adaptive radius	32.77%	31.27%
Block adjacency	33.43%	35.04%
Fixed radius of 240 m	30.77%	29.11%

Table 6.8: Percentages of mixed blocks among the misclassified ones and the percentages of misclassified blocks among the mixed ones for the five considered graph structures.

6.3 Spatial Autocorrelation Analysis

In this section, the assumption that the spatial distribution of the USTs in the study site is not random is evaluated. The extent to which the USTs distribution is spatially autocorrelated is considered to be the extent to which this assumption holds. In the ground-truth map shown in Figure 6.5(c) it can be seen that blocks are frequently surrounded by blocks of the same class. In other words, the UST classes are to a good extent spatially clustered and not randomly distributed. Besides this qualitative visual analysis, two measures for evaluating the spatial autocorrelation of the USTs were used. These measures, presented in Section 4.5.1, are the assortativity index and the Join Counts statistics. These analysis were performed based on the ground-truth map from Figure 6.5(c) which is the generalization of the official UST map from Munich shown in Figure 5.2 according to Table 5.4. Another purpose of the autocorrelation analysis is to investigate whether there is correlation between spatial autocorrelation values and the contextual classifications overal accuracies achieved with the different graph/neighborhood structures.

6.3.1 Based on the Assortativity Index

Table 6.9 shows for the different neighborhood structures the assortativity index, the number of pairs of neighbors as well as the overal accuracies achieved by different contextual classification models. The assortativity index is 1 if all blocks are always neighbors of blocks from the same class and it is 0 if no pair of neighboring blocks belongs to the same class. It can be seen that all neighborhood structures yield good assortative mix, indicating global spatial autocorrelation of the USTs. From observing Table 6.9, one notices that as the number of interaction factors increases, the assortativity index tends to decrease. The only exception to that trend is the neighborhood criterion defined by an adaptive radius. A possible reason might be that, according to this neighborhood defining criterion, small blocks will only have its nearby blocks as neighbors, whereas very large blocks might not be associated to any neighboring pair due to the maximum distance threshold th' (see equation 4.18). Therefore, if it is assumed that closeby blocks of similar size have higher chance of belonging to the same class, this structure is likely to yield higher assortativity values than the one defined by a fixed radius of 140 m, for example.

According to Table 6.9, there is correlation between the number of interaction factors (i.e. pairs of neighboring blocks) and the obtained overal accuracies. When the former increases the overal accuracies obtained with models M1, M2, M4, M6 and M7 also increase. Thus, with the exception of the neighborhood structure defined by the criterion of adaptive radius, there is inverse correlation between the assortativity index and the obtained overal accuracies and the number of interaction factors.

			Overall Accuracy (%)				
	Assort. Idx.	Pairwise Neigh.	M1	M2	M4	M6	M7
Fixed radius of 240 m	0.36	7178	74.71	75.00	75.94	75.00	76.06
Block adjacency	0.41	4280	74.63	74.12	75.21	74.77	75.21
Fixed radius of 140 m	0.43	2233	71.88	73.40	72.75	71.88	73.91
3-nearest neighbors	0.45	1487	71.52	72.31	72.24	71.44	73.04
Adaptive radius	0.46	2360	73.33	74.34	73.84	73.62	74.85

Table 6.9: Assortativity index computed for the different graph/neighborhood structures. The number of interaction factors and the overall accuracy obtained by different models is also shown for the five graph structures considered.

6.3.2 Based on the Join Count Statistics

The analysis of the USTs spatial autocorrelation based on the JC statistics has focused on three neighborhood structures only, namely, the ones defined by (i) the blocks 3-nearest neighbors, (ii) the blocks adjacencies and (iii) by a fixed radius of 240 m. The first of these structures was chosen because it has the least pairs of neighboring blocks and yet the second largest assortativity index. The second was chosen because it is arguably the most intuitive and generic way of establishing neighboring relationships between blocks and it does not require the definition of any parameter to do so. The third of these structures was chosen because it is the one that yielded the highest overal accuracies, as already discussed. Table 6.10 shows for each of the five considered UST classes the number of BB, BW and WW occurences from all pairs of neighboring blocks defined by each of these three graph structures. B represents the occurence of class k and W the occurence of any other class. Hence, BB stands for the pairs of neighbors in which both blocks belong to k, as BW represents the case in which one block belongs to k and the other to a class different than k. Table 6.10 also shows the mean and maximum values of BB, BW and WW computed from 10.000 random permutations of the B and W values from each class.

The existence of spatial autocorrelation for a certain class can be attested when the number of observed BB pairs is higher than the mean number of BB pairs from the n permutations, as well as when the number of observed BWs pairs is lower than the mean BW pairs randomly produced in the permutations. With a single exception concerning class Large Buildings and Industrial Areas and the structure defined by 3-nearest neighbors, this occurs with all classes and structures. In fact, the number of observed BB occurences is always fairly larger than the mean number of BB occurences computed from the permutations. The opposite holds for the BW occurences, i.e. the number of observed BW pairs is always significantly lower than the mean one computed from the permutations. It can also be noticed from Table 6.10 that very frequently the number of BB pairs is larger than the maximum number of BB pairs produced by the random permutations. Accordingly, the number of observed BW pairs is also very frequently lower than the minimum number of BW occurences produced in the random permutations. The few exceptions concern classes Large Buildings and Industrial Areas and Parks and Vegetated Areas, which are the ones with fewer instances. Therefore, there is higher probability that in at least one of the permutations the autocorrelation for these classes will be higher than the actually observed one.

Based on the permutations, pseudo p-values for the BB and BW counts can be computed using equation 4.15. The pseudo p-values for the BB counts express the percentage of times that the observed BBs were counted to be lower than the BBs from one of the n permutations. Pseudo p-values for the BW counts express the number of times that the number of observed BW were counted to be higher than that of one of the n permutations. Table 6.11 depicts these pseudo p-values for all five classes and the three graph/neighborhood structures considered in this section. As it can be seen, the null hypothesis of spatial randomness can be rejected most of the times. The results for the BW counts show for all classes a probability of 1.0 under the null hypothesis of negative spatial autocorrelation. Again, this strongly supports the assumption that there is positive spatial autocorrelation on the USTs distribution.

6.3.3 Based on the Classes neighboring Relationship Proportions

The spatial autocorrelation of the USTs can also be analysed by creating for each class k histograms of the proportions of neighboring relationships between k and all UST classes. As with the JC statistics, such histograms were produced only for the graph/neighborhood structures defined by the criteria of 3-nearest neighbor, blocks adjacencies and fixed radius of 240 m. These histograms are shown in Figure 6.7. In general terms, they do not change considerably according to the neighborhood structure. The most significant common trait is that blocks from class Dense Block Development, along with those from classes Regular Block Development and Detached and Semi-Detached Housing appear clustered concerning the three different neighborhood structures. In other words, blocks from these classes are predominantly neighbors from blocks of the same class. Therefore, the assortativity values computed for these three structures are mostly influenced by these three classes, specially by class Dense Block Development, which, as it can be seen from Table 5.4, is the one with most instances.

Regarding the neighborhood structure defined by the blocks adjacencies, it can be noticed that with the exception of class Parks and Vegetated Areas, all other classes mantain neighboring relationships mostly with blocks from the same class. This also happens when the structure is defined by the 3-nearest neighbor criterion. However, in this case the exception is with class Large Buildings and Industrial Areas. These two classes are the only ones at which a clustered spatial distribution does not hold when longer range contextual relations are considered by defining the neighborhood structure based on the criterion of a fixed radius of 240 m. This effect is surely

3-Nearest neighbors:

	Observed			Randomly Permuted			
Classes:	BB	BW	WW	Mean BB	Mean BW	Max. BB	Min. BW
PVA	118	333	1036	38.87	403.80	67	315
DSDH	179	159	1149	39.99	408.32	72	329
LBIA	4	44	1439	7.81	201.11	25	144
DBD	475	318	694	240.50	715.73	308	645
RBD	147	226	1114	51.12	450.35	82	367

Block adjacency:

	Observed			Randomly Permuted			
Classes:	BB	BW	WW	Mean BB	Mean BW	Max. BB	Min. BW
PVA	144	911	3218	111.69	1160.92	158	1053
DSDH	349	535	3389	114.80	1173.53	160	1054
LBIA	116	466	3691	22.52	576.98	42	482
DBD	1369	945	1959	691.44	2057.27	784	1911
RBD	419	802	3052	146.78	1293.15	199	1156

Fixed Radius of 240 m:

	Observed			Randomly Permuted			
Classes:	BB	BW	WW	Mean BB	Mean BW	Max. BB	Min. BW
PVA	303	1605	5274	187.93	1950.68	252	1774
DSDH	689	917	5576	193.11	1972.39	257	1755
LBIA	54	395	6733	37.85	970.62	73	812
DBD	2406	1659	3117	1161.70	3456.24	1318	3274
RBD	677	1320	5185	247.25	2174.04	336	1997

Table 6.10: Join Count statistics of all five classes according to three different neighborhood structures. The number of permutations was set to 10.000. The numbers support the assumption that there is spatial autocorrelation of the USTs for the three graph/neighborhood structures considered.

Class	pseudo p-value BB			pseudo p-value BW		
	Fx. Rd. of 240 m	Blk. Adj.	3-NN	Fx. Rd. of 240 m	Blk. Adj.	3-NN
PVA	0.0	0.003	0.0	1.0	1.0	1.0
DSDH	0.0	0.0	0.0	1.0	1.0	1.0
LBIA	0.016	0.0	0.94	1.0	1.0	1.0
DBD	0.0	0.0	0.0	1.0	1.0	1.0
RBD	0.0	0.0	0.0	1.0	1.0	1.0

Table 6.11: Pseudo p-values of the Join Counts statistics for the five classes and the three different neighborhood structures. The pseudo p-values are inferred for each class based on the random permutations of the B and W values and using equation 4.15.

influenced by the facts that these are the classes with fewer instances and usually the largest block sizes.

6.3.4 Considering the Blocks Shape Similarity

This section intends to analyse whether the probability of two neighboring blocks belonging to the same class increases as the similarity of these blocks shapes also increases. The assumption that this does happen was the rationale for creating the contextual classification model M6. Figure 6.8 shows the cumulative frequency distributions of the shape similarity measure introduced in Section 4.5.2 (equation 4.16) for three different neighborhood structures. It also shows together with these cumulative frequency distributions the percentage of pairs of neighboring blocks belonging to the same and to different classes and whose shape similarity values lie below the value on the abscissa. It can be observed that the lower the values on the abscissa, i.e. the higher the shape similarities, the more probable it is that a pair of neighboring blocks with equal or higher similarity belong to the same class. This trend persists through the accumulation of more than 80% of the pairs of neighboring blocks. It is not that intense in the case when neighbors are defined by the blocks adjacencies. However, it is strong enough in the cases of the other two structures so that it can be stated that the assumption holds. In the cases of all three structures, the overall probability that two neighboring blocks belong to the same class is of about 56%, which means that, regardless of the criterion for establishing neighboring relationships between blocks, context can provide valuable information in estimating the UST class from each block.

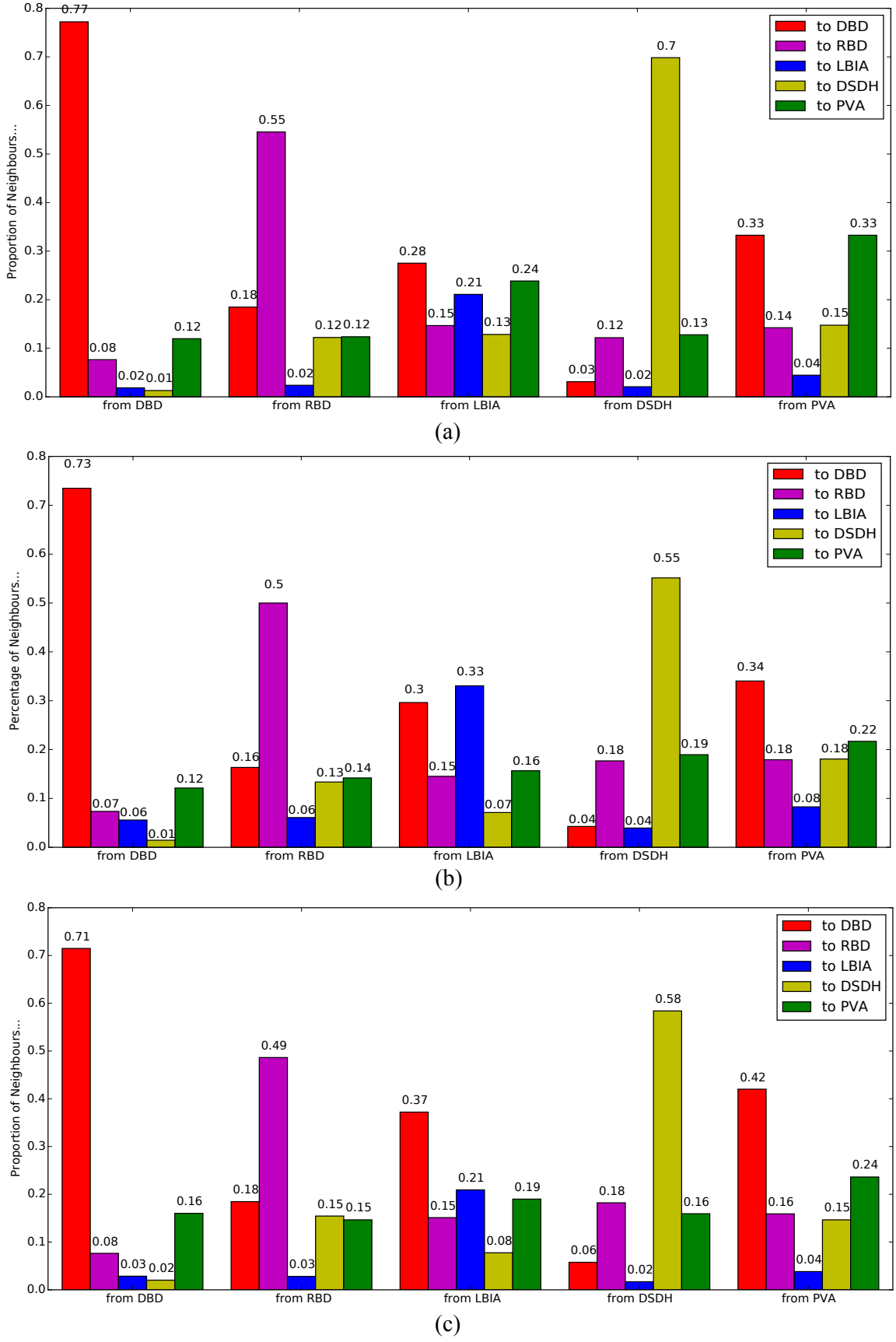

Figure 6.7: Histograms of the proportions of class-wise relationships between neighboring blocks. a) Histograms for the case when neighboring relationships are defined by the 3-nearest neighbors criterion. b) Histograms for the case when neighboring relationships are defined by the criterion of block adjacency. c) Histograms for the case when neighboring relationships are defined by a fixed radius of 240 m.

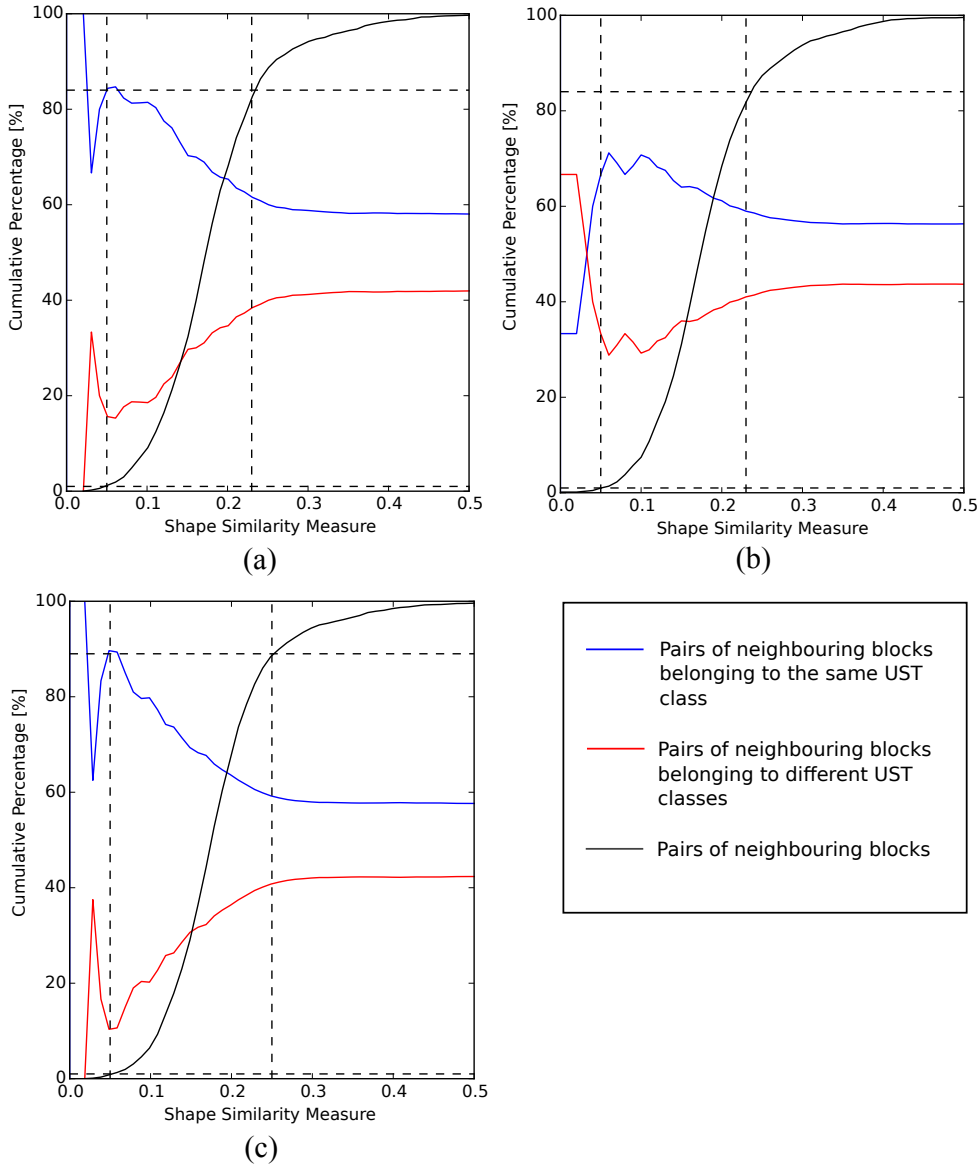

Figure 6.8: Cumulative frequency distribution of the shape similarity measure for three different graph definition criteria. a) Cumulative frequency distribution for the case when the neighboring relationships between blocks is defined by the criterion of 3-nearest neighbors. b) Cumulative frequency distribution for the case when the neighboring relationships between blocks is defined by the criterion of block adjacency. c) Cumulative frequency distribution for the case when the neighboring relationships between blocks is defined by the criterion of a fixed radius of 240 m.

7 Discussion

7.1 On the Application and Dataset

The distinction of USTs using remote sensing imagery is for a number of reasons a difficult task. USTs are defined based on physical, environmental and functional aspects which are considered at the level of some spatial aggregate, like user-defined grids, census tracts or, more commonly, urban blocks. By using his empirical knowledge and subjectiveness, it might be intuitive for a human expert to distinguish the USTs of urban blocks. However, it is not easy to translate the evidences and criteria a human considers into quantitative image-derived attributes. The challenge lies in computing attributes that expressively describe the composition and structure of different urban features, e.g. buildings, sealed areas and vegetated areas, found inside these spatial aggregates. The fact though that USTs are at times partially (at times even entirely) defined by functional aspects sets limits to their distinction based on quantitative image-derived attributes. It seems thus reasonable to assume that the upper limit of the extent to which USTs can be distinguished based on RS imagery depends on the extent to which the classes are defined by physical and environmental aspects. It should be mentioned that fortunately functional aspects have frequently a physical expression, as in the cases of residential and industrial uses which give rise to very distinguishable types of buildings and urban environments. A way to get around this dependency on physical aspects is by considering more general classes resulting from the grouping of the original and more specific ones based on their physical similarities. Whether and how this can be made depends on the specific study site and on the purposes for distinguishing USTs.

If on one hand it is hard to emulate human reasoning through image-derived attributes and classification strategies, the automatic distinction of USTs from remote sensing imagery has the advantage that it is performed based on objective criteria and thus human subjectiveness is eliminated. This is preferable when the UST map is used as an input to systems dedicated to model phenomena like radiant energy balance and carbon sequestration, or when it is used as a variable for estimating climate confort and other quality of life indicators. In these types of analyses, the input data referring to the city's built-up structures have to concern the physical and environmental aspects of the classes, such as the presence and structure of vegetated areas and the size and density of the buildings. Remote sensing imagery and methods offer great potential for categorizing urban settlements according to such criteria.

Independently from the methods applied, the distinction of USTs with remote sensing data requires the use of imagery with spatial resolution high enough so that the features that characterize the USTs can be effectively detected. Spaceborne sensors, both SAR and optical ones, fulfill this requirement well, what is important regarding the distinction of USTs in parts of the world where airborne remote sensing data may be lacking. In comparison to optical imagery though, the distinction of USTs with SAR/InSAR imagery is a bit more difficult. As discussed in Chapter 2, the most seminal works on block-based distinction of USTs rely on a previously performed, and sometimes detailed, land cover classification. However, few experts would disagree that urban land cover classification is harder to perform with SAR/InSAR data. While on optical sensors

the backscattered energy is strongly influenced by the physical-chemical properties of the objects on the surface, on SAR systems it is dependent on the moisture content, geometrical structure, roughness, condictivity and orientation of the objects in relation to the sensor's azymuth direction. This requires, or at least encourages, the development of approaches specific for the case when the objects must be extracted from the urban blocks using SAR/InSAR images. This thesis advogates that block attributes computed from InSAR imagery should rely on the expressiveness of the coherence image for the extraction of vegetated and sealed areas, as well as on the intensity image for the extraction of shadow and corner reflection features. As shown in this thesis though, 3D information extracted from interferometrically derived DEMs are not expressive enough for accurately extracting buildings inside urban blocks. This is mainly because of the complexity of urban areas regarding the SAR-inherent effects, what strongly influences the performance of phase unwrapping methods and hence the accuracy of InSAR-derived DEMs.

Another idiosyncracy of USTs distinction and urban land use classification is that, differently from land cover classes, the built-up structure pattern of each UST class may vary significantly from city to city. Although in a region, country and even a continent cities may be structurally similar, environmental, cultural and economic conditions are frequently the causes for city-specific class definitions and class appearances. This makes the transferability of any methodological approach more difficult, not to mention specific model parameterizations. Moreover, UST classes are internally heterogenous regarding their defining traits and regarding also the shape and size of the blocks assigned to a class. That makes each block unique and only to a certain extent similar to other blocks from the same UST class. These factors motivate the adoption of approaches which take *uncertainty* into account at assigning classes to blocks.

7.2 On the Proposed Approach

This thesis proposes a block-based approach for distinguishing USTs. Urban blocks were chosen as elements of analysis because they are usually the spatial entities at the level of which USTs and urban land use classes are defined by official planning and city administration agencies. This is because blocks are bounded by streets, which in turn define the structure of a city's circulation dynamics as well as the largest areas available for construction in a city. Thus, urban zoning laws and other intervention and monitoring regulations are defined based on this type of spatial entity. Classifying USTs at the level of urban blocks require their extraction, what can be performed straightforwardly if a street network is available. Fortunately, such a data is frequently available and can be provided by the city administration and planning agencies.

The main factors that enable an UST class to be effectively distinguished from remote sensing images are two, namely, the defining traits and the number of instances of this class. The former concerns the aspects and criteria that defines a class as opposed to all others. If an UST class is defined essentialy by a functional or cultural aspect (e.g. sport facilities, historical areas) most likely such class will be frequently misclassified. The second aspect refers to the necessity of training the classification algorithm with a reasonable number of samples from each class. Hence, specific and rare classes should also be merged with others. In order to increase the number of distinguishable classes, the study site would thus have to be enlarged so that enough training samples from the additional classes could be collected. This thesis has satisfactorily distinguished five general UST classes. The five classes cover the whole spectrum of the most commom built-up and permeable surface structures.

Instead of distinguishing USTs based on attributes that regard the structure of the land cover classification from inside each block, the strategy adopted in this work was of extracting lines and polygons features from inside each block and then computing attributes that decribe the

presence, structure and spatial disposition of these features. Although in this way a previous land cover classification could be spared, the expressiveness of such attributes depends on favourable relative orientation between the buildings and the SAR range axis. This is because in order for buildings to appear in a SAR intensity image in the form of shadow and double-bounce features, this appropriate relative orientation must exist. This dependency can partially but surely be mitigated by considering images acquired at the satellite's both looking directions. Ideally, having SAR acquisitions from multiple aspects would guarantee the appearance of such shadow and double-bounce features, whereas in optical imagery the appearance and shape of shadows depend on the illumnination conditions and the lines may not correspond to roof edges. Although several attributes were explored which describe the spatial disposition of the lines and polygons features inside the blocks, many of these attributes were found to be less expressive than more simple and commonly used composition attributes, like the features-to-block area proportion, the mean length of lines, etc. This is probably because such attributes are more generic and hence describe more expressively the variability of appearances from the UST classes.

Nowadays, many powerful classification methods can be applied on remote sensing images. Each method poses advantages and disadvantages in light of the conditions and constraints of each application. In this thesis, the classification of USTs based on the comparison of MDPs generated for each block has achieved comparable results as the ones obtained with Random Forest, which is one of the algorithms of choice for several remote sensing imagery classification applications. The proposed strategy for extracting CCs and creating MDPs with them can easely be transposed, since it does not require the tunning of any parameter except the number of percentiles of the images digital numbers, which are to be considered as the thresholds values based on which the CCs are extracted through image thresholding. The use of the relative entropy metric for comparing the blocks MDPs does not require the setting of any parameter either.

Regarding the motivation for considering context in the classification of USTs, spatial autocorrelation analyses performed in this thesis has shown that for the city of Munich, a typical large european city, the five considered classes are not spatially distributed in random manner. They are in fact disposed in spatial clusters. This can be expected to occur in cities with similar growth and planning history and more generally on cities where the distribution of USTs is regulated and controlled by urban zoning laws, a common fact in many important cities at the present time.

Regaring the method through which contextual relations between blocks are considered in the classification, it is safe to say that PGMs are presently the most statistically sound and flexible framework for that. All constituent parts of a PGM (i.e. its structure, potential functions and inference algorithm) are independent from each other, what enables easy adaption to other conditions, such as study site, input data and computational resources. One particularly important advantage is the fact that any classifier can be used as the association potential functions, as long as it outputs for each element a relevance value to each of the considered classes. Likewise, the interaction potential functions can also be altered or adapted without any influence on the other model parts. In this thesis, I investigated and compared what seems to be the most general neighborhood structures and interaction potential functions for the block-based distinction of USTs.

7.3 On the Assumptions and Results

The results obtained in this thesis support the assumption that considering the context of each urban block in the UST classification increases the overall accuracy. At times, the accuracy improvement has reached up to 10%. In the high majority of times though it was of about 7 to 8%. The obtaining of standard and contextual classifications with accuracies of approximately

70% and 80% respectively supports the assertion that HR spaceborne InSAR images can be used to distinguish general USTs.

Section 2.3 shows examples of urban blocks where certain features would not be possible to be extracted if only images from one looking direction were available. Since most attributes considered in the classifications are computed based on such extracted features, it was assumed that classifications performed with attributes derived from images acquired at both looking directions would always achieve higher accuracy indexes. The results from classification experiments conducted with one training sample set collected in supervised way and with other twenty-five sample sets collected randomly have shown that this assumption could be confirmed. In all of these twenty-six classification experiments, the accuracy has increased in average 5% when considering also attributes computed from images obtained at the other looking direction.

8 Conclusion and Outlook

This chapter closes the thesis by making the last commentaries and drawing the main conclusions. Suggestions for future works are made and tasks still to be accomplished are addressed.

8.1 Summary and Conclusions

The main goal of this thesis is to distinguish USTs based on remote sensing imagery. The proposed approach considers the urban blocks as the elements of analysis, what is in agreement with the official UST map from the study site, i.e. Munich (Germany), and from many other cities in the world as well. Nothing hinders though the application of the methods proposed in this work on other types of spatial entities like census tracts or user-defined grids and regions. UST classes are defined mainly by the presence, geometrical structure and spatial disposition of urban features, e.g. buildings and vegetated areas, located inside these spatial entities. This fact, together with the number of shapes and sizes that urban blocks may assume, gives rise to great variability in the appearance of each UST class. The conception and computation of attributes that expressively describe the blocks according to these features is an art, for one is not dealing with classes that have a more or less fixed spectral or structural signature. On the contrary, the classes are complex and heterogenous as the ways in which human constructions may be.

USTs official maps are frequently produced based on aspects that cannot be quantitatively or objectively formulated. Frequently, it is not even produced and updated by the same expert, that is to say, with the same subjectivity. To evaluate automatic USTs classifications taking into consideration the official UST maps as ground-truth is surely the most appropriate thing to do. However, the performance of a classification approach should not be evaluated solely based on the crisp analysis of whether a class assignment corresponds exactly to the one from the official map. Perhaps, automatic USTs classifications should also be evaluated based on the consistency of the classes, i.e. based on whether the instancies from each class are similar regarding their built-up structure. In other words, the question that should be asked is: is the UST classification meaningful and therefore useful regarding the purposes for which the UST map is used? In this sense, the main point concerns not so much the achieved overall accuracy, but rather if the attributes that are to be used to distinguish the classes are reasonable and if the features based on which the attributes are computed could be effectively extracted.

This thesis main argument is that the most important key concept in the distinction of USTs based on remote sensing imagery is the description, modeling and consideration of *context*. This was approached in two different levels, namely, internally to each urban block and externally at the level of the mutual influences between neighboring blocks regarding their class assignments.

The first level refers to the spatial disposition of the image features which presumably indicate the presence and geometrical properties of certain urban features, such as buildings and vegetated areas. In this thesis, several attributes were proposed and used which describe the position and orientation of the features in relation to each other and to the boundaries of the blocks. The

features spatial disposition was also described by means of creating for each block a network connecting the features extracted in it and then by extracting the network's structural properties and autocorrelation measures referring to the features different geometrical properties. Although this is certainly pertinent from a theoretical perspective, the practice has shown that more simple attributes are most of the times more relevant. The relevance of the attributes, however, should be considered in light of the classification method used to evaluate them. Perhaps, knowledge-based approaches, which are flexible and powerful in terms of emulating human reasoning, could consider attributes that become relevant according to their class-membership functions and in the context of the user-defined class despriction rules, as done by Novack & Kux [2010] and Novack et al. [2014]. The difference in the potential of optical and SAR imagery for distinguishing USTs lies therefore in the effectiveness with which certain image features can be extracted. In other words, if the features are effectively extracted, the class-describing attributes will be expressive. In this sense, this thesis contributed in showing that having imagery from the satellite's ascending and descending looking directions considerably increases the chances of effectively extracting these features. It was also demonstrated by experiments and results that the classification accuracy would have been lower if imagery from only one of the satellite's looking directions were available.

The second level at which context was considered in the distinction of USTs refers to the mutual influence of neighboring blocks regarding their class assignment. These influences were modelled by means of different PGMs. It is secure to say that in the state-of-the-art PGMs offer the most flexible and statistically sound approach to model contextual relationships between spatial entities in geographical analysis and image processing applications. Based on the assumption that the spatial distribution of the USTs in a city is not random, this thesis is the first to explore contextual information and probability theory as a way to model uncertainty and explore the information from the context of each block as a way to improve the accuracy of standard block-based classification. Presumably, the accuracy improvement observed in all contextual classifications performed in this thesis is partially caused by the effect that blocks whose shape and appearance differ considerably from the ones from the class to which it belongs become less vulnerable to misclassifications if they are surrounded by blocks with strong association potentials to the correct and same class as the concerning block. Commonly adopted and proposed ways of modeling the contextual interaction between urban blocks were considered by means of different parameterizations of the PGMs as well as by defining the structure of the models based on different criteria. The overal accuracy improvements obtained in all contextual classification experiments showed that this is a worthwhile effort in the UST and urban land use classification.

8.2 Outlook

The results presented in this thesis and on correlated papers published during its realization are expected to encourage the consideration of contextual relations between neighboring blocks regarding their class assignments. PGMs are such a flexible and powerfull framework that it is hard to imagine another approach through which contextual relations can be modelled and considered in block-based USTs and land-use classifications. As mentioned, one important feature of PGMs is that they allow the inclusion of other terms in the model that are not the necessarly existent association and interaction terms. In this thesis a third term was included in two of the models, but they were only a second association term. As shown in Section 3.4, multi-temporal and multi-scale classification has been performed with models containing such additional terms. Therefore, multi-temporal, hierarchical and contextual classification of USTs are expected to be performed with PGMs in the near future.

In this thesis, the parameterization of the interaction factors were defined by standard and contrast-sensitive Potts models. These functions penalize the energy when two neighboring blocks

are assigned to different classes. This might smooth the classification at occasions on which it should not, that is to say when two different classes are related in such a way that causes them to be spatially close to one another. A simple example concerns the UST class 'Parks and Vegetated Areas', which is frequently observed to be surrounded by blocks belonging to other related classes, like 'Detached and Semi-detached Housing' or 'Dense Block Development'. Inversely, it might be less probable that two blocks belonging to class 'Parks and Vegetated Areas' are found to be direct neighbors. These type of contextual relationships can easely be modelled by the parameterization of the interaction factors.

As shown in Section 6.2.2, approximately 30% of the blocks in the study site are assigned to more than one UST class. However, the issue of mixed blocks was not addressed in this work. To the best of our knowledge, this is a gap still to be filled in the field of block-based UST distinction. The detection, internal segmentation and classification of mixed blocks brings the question on how to evaluate the accuracy of these tasks when they are performed automatically. At least in the case of Munich, it seems very hard to obtain by automatic methods the same internal segmentation of mixed blocks as the ones from the official map. Figure 8.1 shows two examples of mixed blocks concerning their segmentation and classification before and after the class generalization process explained in Section 5.2.1. It can be seen that their exact internal segmentation could hardly be obtained by an automatic method.

An important requirement for improving the results obtained with the proposed approach is the improvement of the spatial resolution of the (In)SAR images. This would enable a better effectiveness of the proposed line and polygon extractors as well as the development of more powerfull ones, making the attributes based on these features more expressive. Furthermore, the interferometrically derived DEMs would also be expressive enough to enable much more accurate distinction of certain classes. The basic condition for enhancing the spatial resolution of the SAR imagery is the acquisition of imagery by airborne sensors. Airborne SARs are able to acquire images at different, also orthogonal, aspect angles. Having SAR/InSAR acquisitions from four orthogonal aspects angles increases tremendously the chances of effectively and accurately extracting image features which indicate the presence and geometrical properties of the buildings inside the blocks, as demonstrated by Schmitt [2014]. However, it should be stressed, none of this improvements on the side of the input image data reduces the inherent complexity of the UST classes. Hence, the difficulty of distinguishing them from remote sensing imagery would fundamentally persist.

Another aspect of the field of USTs distinction with remote sensing imagery that has not been explored yet concerns the extraction of streets from the imagery as a way to abdicate from auxiliary data provided by third-parties. Automatically extracting lines corresponding to streets and using those as boundaries from closed segments which hopefully correspond to the urban blocks of the city is an effort that would be fairly appreciated by the scientific community in this field. Such an approach would be in particular pertinent in cities where auxiliary data concerning the streets network is not available or is not reliable.

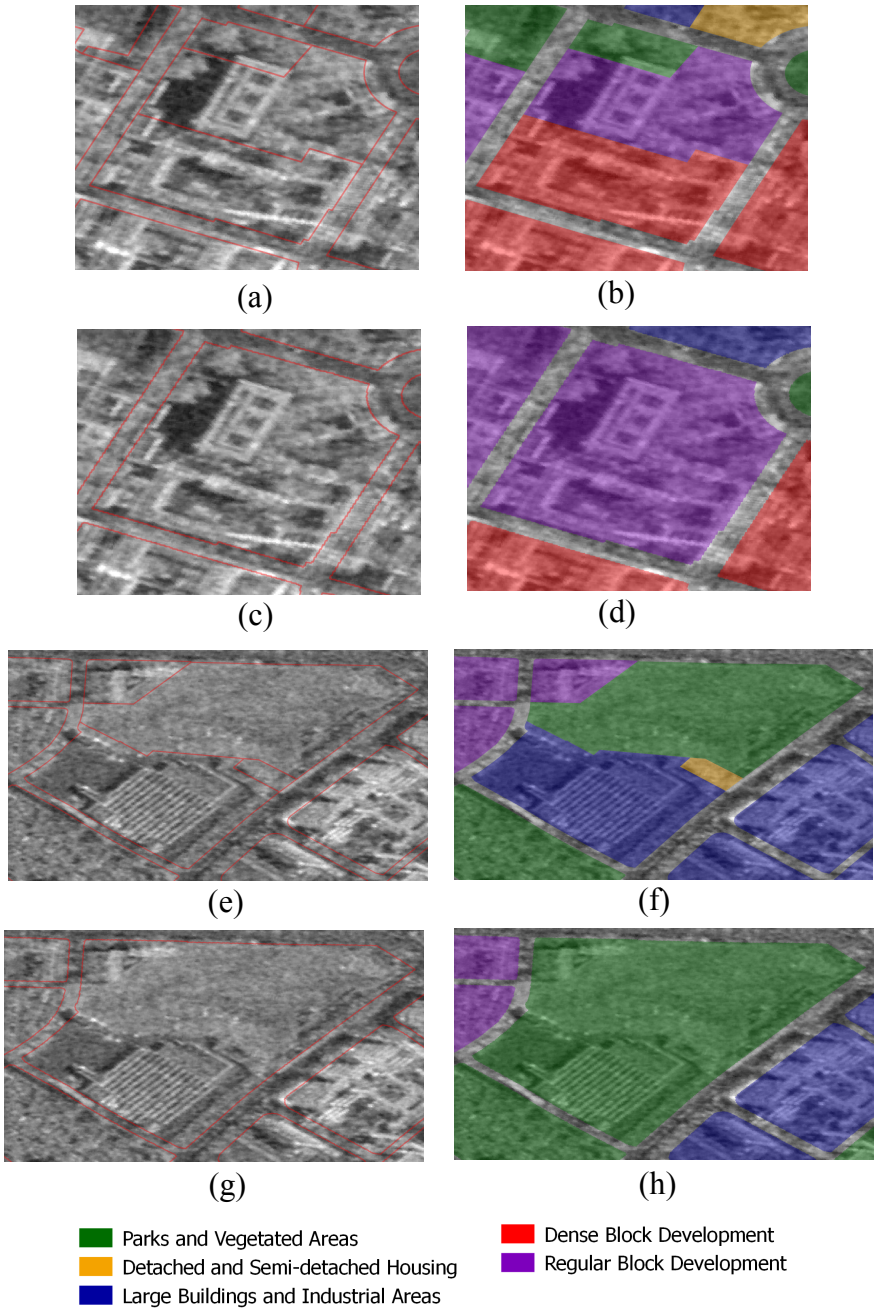

Figure 8.1: Two examples of mixed blocks concerning their segmentation and classification before and after the class generalization process explained in Section 5.2.1. a) and e) Ground-truth internal segmentation of two mixed blocks. b) and f) Ground-truth original classification of two mixed block. c) and g) Merge of the internal segments of the two mixed block. d) and h) Result of the class generalization process which produced the final ground-truth classification of the two mixed blocks.

Bibliography

Agrawal R, Imielinski T, Swami A (1993) Database mining: a performance perspective. IEEE Transactions on Knowledge and Data Engineering, 5: 914–925.

Alioscha-Perez M, Sahli H (2014) Efficient learning of spatial patterns with multi-scale conditional random fields for region-based classification. Remote Sensing, 6: 6727–6764.

Andres B, Kappes JH, Köthe U, Schnörr C, Hamprecht FA (2010) An empirical comparison of inference algorithms for graphical models with higher order factors using OpenGM. In: Goesele M, Roth S, Kuijper A, Schiele B, Schindler K (eds) Pattern Recognition, volume 6376 of Lecture Notes in Computer Science (pp. 353–362). Berlin: Springer.

Arkin E, Chew L, Huttenlocher D, Kedem K, Mitchell J (1991) An efficiently computable metric for comparing polygonal shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13 (3): 209–216.

Bach P, Staalesen S, McCarthy D, Deletic A (2015) Revisiting land use classification and spatial aggregation for modelling integrated urban water systems. Landscape and Urban Planning, 143: 43–55.

Ban Y, Jacob A, Gamba P (2015) Spaceborne SAR data for global urban mapping at 30 m resolution using a robust urban extractor. ISPRS Journal of Photogrammetry and Remote Sensing, 103: 28–37.

Banzhaf E, Hoefer R (2008) Monitoring urban structure types as spatial indicators with CIR aerial photographs for a more effective urban environmental management. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 1 (2): 129–138.

Barbosa O, Tratalos J, Armsworth P, Davies R, Fuller R, Johnson P, Gaston K (2007) Who benefits from access to green space? A case study from Sheffield (UK). Landscape and Urban Planning, 83 (23): 187–195.

Baud I, Kuffer M, Pfeffer K, Sliuzas R, Karuppannan S (2010) Understanding heterogeneity in metropolitan India: the added value of remote sensing data for analyzing sub-standard residential areas. International Journal of Applied Earth Observation and Geoinformation, 12 (5): 359–374.

Benz U, Hofmann P, Willhauck G, Lingenfelder I, Heynen M (2004) Multi resolution object oriented fuzzy analysis of remote sensing data for GIS ready information. ISPRS Journal of Photogrammetry and Remote Sensing, 58 (34): 239–258.

Blaschke T (2010) Object based image analysis for remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing, 65 (1): 2–16.

Bolund P, Hunhammar S (1999) Ecosystem services in urban areas. Ecological Economics, 29 (2): 293–301.

Breiman L (2001) Random forests. Machine Learning, 1: 5–32.

Brunner D (2009) Advanced methods for building information extraction from very high resolution SAR data to support emergency response. PhD thesis, University of Trento.

Brunner D, Lemoine G, Bruzzone L, Greidanus H (2010) Building height retrieval from VHR SAR imagery based on an iterative simulation and matching technique. IEEE Transactions on Geoscience and Remote Sensing, 48 (3): 1487–1504.

Bruzzone L, Cossu R, Vernazza G (2004) Detection of land-cover transitions by combining multidate classifiers. Pattern Recognition Letters, 25 (13): 1491–1500.

Bundesamt U (2015) Siedlungs- und Verkehrsflaeche. online. accessed in 30.10.2015.

Cao Y, Su C, Liang J (2012a) Building unit density detection from high resolution TerraSAR-X image based on mathematical morphological operators. In: IEEE International Geoscience and Remote Sensing Symposium, 1: 5978–5981.

Cao Y, Su C, Liang J (2012b) High resolution SAR building detection with scene context priming. In: IEEE 11th International Conference on Signal Processing, 3: 1791–1794.

Chang N, Wang S (1997) A fuzzy goal programming approach for the optimal planning of metropolitan solid waste management systems. European Journal of Operational Research, 99 (2): 303–321.

Chen X, Yu Y, Zhu P (2009) Study from building density to building 3D density. In: International Conference on Management and Service Science, 1: 1–5.

Chen Z, Y. Z, Guindon B, Esch T, Roth A, Shang J (2012) Urban land use mapping using high resolution SAR data based on density analysis and contextual information. Canadian Journal of Remote Sensing, 38 (06): 738–749.

Chun B, Guldmann J (2014) Spatial statistical analysis and simulation of the urban heat island in high-density central cities. Landscape and Urban Planning, 125: 76–88.

Cliff A, Ord J (1981) Spatial Processes: Models and Applications. London: Pion Limited.

Cockx K, Van de Voorde T, Canters F (2014) Quantifying uncertainty in remote sensing-based urban land-use mapping. International Journal of Applied Earth Observation and Geoinformation, 31: 154–166.

Congalton RG, Green K (1999) Assessing the accuracy of remotely sensed data : principles and practices. Mapping science series. Boca Raton, London, New York: Lewis Publishers.

Coseo P, Larsen L (2014) How factors of land use/land cover, building configuration, and adjacent heat sources and sinks explain urban heat islands in Chicago. Landscape and Urban Planning, 125: 117–129.

Costantini M (1998) A novel phase unwrapping method based on network programming. IEEE Transactions on Geoscience and Remote Sensing, 36: 813–821.

Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1: 886–893.

Dell'Acqua F, Gamba P, Lisini G (2003) Improvements to urban area characterization using multitemporal and multiangle SAR images. IEEE Transactions Geoscience and Remote Sensing, 41 (9): 1996–2004.

Dlamini W (2011) Application of a bayesian network for land-cover classification from a Landsat 7 ETM+ Image. International Journal of Remote Sensing, 32 (21): 6569–6586.

Douglas DH, Peucker TK (1973) Classics in Cartography: Reflections on Influential Articles from Cartographica, volume 10, chapter Algorithms for the reduction of the number of point required to represent a digitized line or its caricature, (pp. 112–122). Chichester: John Wiley and Sons.

Duda RO, Hart PE (1972) Use of the Hough transformation to detect lines and curves in pictures. Communications of the ACM, 15: 11–15.

Duque J, Ramos R, Surinach J (2006) Supervised regionalization methods: a survey. International Regional Science Review, 8: 1–31.

Duque JC, Patino JE, Ruiz LA, Pardo-Pascual JE (2015) Measuring intra-urban poverty using land cover and texture metrics derived from remote sensing data. Landscape and Urban Planning, 135: 11–21.

Elberink S, Vosselman G (2011) Quality analysis on 3D building models reconstructed from airborne laser scanning data. ISPRS Journal of Photogrammetry and Remote Sensing, 66 (2): 157–165.

Esch T, Schmidt M, Breunig M, Felbier A, Taubenboeck, H.ck H, Heldens W, Riegler C, Roth A, Dech S (2011) Identification and characterization of urban structures using VHR SAR data. In: IEEE International Geoscience and Remote Sensing Symposium, 1: 1413–1416.

Esch T, Thiel M, Schenk A, Roth A, Mueller A, Dech S (2010) Delineation of urban footprints from TerraSAR-X data by analyzing speckle characteristics and intensity information. IEEE Transactions on Geoscience and Remote Sensing, 48 (2): 905–916.

Fanelli A, Ferri M, Santoro M, Vitale A (2001) Analysis of coherence images over urban areas in the extraction of buildings heights. In: IEEE/ISPRS Joint Workshop Remote Sensing and Data Fusion over Urban Areas, 1: 69–73.

Feitosa R, Costa G, Mota G, Pakzad K, Costa M (2009) Cascade multitemporal classification based on fuzzy Markov chains. ISPRS Journal of Photogrammetry and Remote Sensing, 64 (2): 159–170.

Ferro A, Brunner D, Bruzzone L (2009) An advanced technique for building detection in VHR SAR images. In: Earth Resources and Environmental Remote Sensing/GIS Applications, 7477: 1–12.

Ferro A, Brunner D, Bruzzone L, Lemoine G (2011) On the relationship between double bounce and the orientation of buildings in VHR SAR images. IEEE Geoscience and Remote Sensing Letters, 8 (4): 612–616.

Galamhos C, Matas J, Kittler J (1999) Progressive probabilistic Hough transform for line detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1: 554–560.

Ganapathi V, Vickrey D, Duchi J, KolZhu P (2008) Constrained approximate maximum entropy learning of markov random fields. In: Proceedings of the Twenty-Fourth Conference Conference on Uncertainty in Artificial Intelligence, 1: 196–203.

Gernhardt S, Adam N, Eineder M, Bamler R (2010) Potential of very high resolution SAR for persistent scatterer interferometry in urban areas. Annals of GIS, 16 (2): 103–111.

Gill S, Handley J, Ennos A, Pauleit S, Theuray N, Lindley S (2008) Characterising the urban environment of UK cities and towns: A template for landscape planning. Landscape and Urban Planning, 87 (3): 210–222.

Gonzalez RC, Woods RE (2001) Digital Image Processing. Boston, MA, USA: Boston: Addison-Wesley Longman Publishing, 2nd edition.

Grey W, Luckman A (2003) Mapping urban extent using satellite radar interferometry. Photogrammetric Engineering & Remote Sensing, 69: 957–961.

Haala N, Kada M (2010) An update on automatic 3D building reconstruction. ISPRS Journal of Photogrammetry and Remote Sensing, 65 (6): 570–580.

Hao M, Shi W, Deng K, Zhang H (2014) A contrast-sensitive Potts model custom-designed for change detection. European Journal of Remote Sensing, 47: 643–654.

Hao Y, Kang J, Krijnders J (2015) Integrated effects of urban morphology on birdsong loudness and visibility of green areas. Landscape and Urban Planning, 137: 149–162.

Haralick R, Shanmugam K, Dinstein I (1973) Textural Features for Image Classification. IEEE Transactions on In Systems, Man and Cybernetics, 3: 610–621.

Hastie T, Tibshirani R, Friedman J (2001) The Elements of Statistical Learning. Springer Series in Statistics. New York: Springer.

Hecht R, Herold H, Meinel G, Buchroithner M (2013) Automatic derivation of urban structure types from topographic maps by means of image analysis and machine learning. In: 26th International Cartographic Conference.

Heiden U, Heldens W, Roessner S, Segl K, Esch T, Mueller A (2012) Urban structure type characterization using hyperspectral remote sensing and height information. Landscape and Urban Planning, 105 (4): 361–375.

Heinzel J, Kemper T (2015) Automated metric characterization of urban structure using building decomposition from very high resolution imagery. International Journal of Applied Earth Observation and Geoinformation, 35: 151–160.

Hermosillaa T, Ruiza L, Recioa J, Cambra-Lopezb M (2012) Assessing contextual descriptive features for plot-based classification of urban areas. Landscape and Urban Planning, 106: 124–137.

Herold M, Couclelis H, Clarke K (2005) The role of spatial metrics in the analysis and modeling of urban land use change. Computers, Environment and Urban Systems, 29 (4): 369–399.

Hoberg T, Rottensteiner F, Feitosa R, Heipke C (2015) Conditional random fields for multitemporal and multiscale classification of optical satellite imagery. IEEE Transactions on Geoscience and Remote Sensing, 53 (2): 659–673.

Hrehorowicz-Gaber H (2013) Effects of transformations in the urban structure on the quality of life of city residents in the context of recreation. Bulletin of Geography. Socio-Economic Series, 21 (21): 61–68.

Hu X, Weng Q (2009) Estimating impervious surfaces from medium spatial resolution imagery using the self-organizing map and multilayer perceptron neural networks. Remote Sensing of Environment, 113 (10): 2089–2102.

Huck A, Hese S, Banzhaf E (2011) Delineating parameters for object-based urban structure mapping in Santiago de Chile using QuickBird data. In: ISPRS Hannover 2011 Workshop, 38(4/W19): 1–6.

Im J, Lu Z, Rhee J, Quackenbush L (2012) Impervious surface quantification using a synthesis of artificial immune networks and decision/regression trees from multi-sensor data. Remote Sensing of Environment, 117: 102–113.

Jaynes C, Riseman E, Hanson A (2003) Recognition and reconstruction of buildings from multiple aerial images. Computer Vision and Image Understanding, 90 (1): 68–98.

Kadous MW (2002) Temporal classification: extending the classification paradigm to multivariate time series. PhD thesis, University of New South Wales.

Kappes J, Andres B, Hamprecht F, Schnorr C, Nowozin S, Batra D, Kim S, Kausler B, Lellmann J, Komodakis N, Rother C (2013) A comparative study of modern inference techniques for discrete energy minimization problems. In: IEEE Conference on Computer Vision and Pattern Recognition, 1: 1328–1335.

Kirsch R (1971) Computer determination of the constituent structure of biological images. Computers and Biomedical Research, 4: 315–328.

Kohli P, Ladicky L, Torr P (2009) Robust higher order potentials for enforcing label consistency. International Journal of Computer Vision, 82: 302–324.

Koller D, Friedman N (2009) Probabilistic Graphical Models: Principles and Techniques. London: MIT Press.

Kschischang F, Frey B, Loeliger H (2001) Factor graphs and the sum-product algorithm. IEEE Transactions on Information Theory, 47: 498–519.

Kullback S, Leibler R (1951) On information and sufficiency. Annals of Mathematical Statistics, 22: 79–86.

Kumar S, Hebert M (2006) Discriminative random fields. International Journal of Computer Vision, 68 (2): 179–201.

Lackner M, Conway T (2008) Determining land-use information from land cover through an object-oriented classification of IKONOS imagery. Canadian Journal of Remote Sensing, 34 (2): 77–92.

Lee S, Chun S, Kim D, Lee J, Chung C (2000) Similarity search for multidimensional data sequences. In: 16th International Conference on Data Engineering, 1: 599–608.

Lee S, Ganapathi V, Koller D (2007) Advances in Neural Information, volume 19, chapter Efficient structure learning of markov networks using L1-regularization, (pp. 817–824). Cambridge: MIT Press.

Leinenkugel P, Esch T, Kuenzer C (2011) Settlement detection and impervious surface estimation in the Mekong delta using optical and SAR remote sensing data. Remote Sensing of Environment, 115 (12): 3007–3019.

Li G, Weng Q (2005) Using Landsat ETM+ imagery to measure population density in indianapolis, Indiana, USA. Photogrammetric Engineering and Remote Sensing, 71: 947–958.

Lin H, Zhou X, Kuang G (2013) Building detection from urban SAR image using building characteristics and contextual information. EURASIP Journal on Advances in Signal Processing, 56 (1): 1–16.

Linden S, Hostert P (2009) The influence of urban structures on impervious surface maps from airborne hyperspectral data. Remote Sensing of Environment, 113 (11): 2298–2305.

Lu Z, Im J, Rhee J, Hodgson M (2014) Building type classification using spatial and landscape attributes derived from LiDAR remote sensing data. Landscape and Urban Planning, 130: 134–148.

Maksymiuk O, Stilla U (2012) A concept for building reconstruction from airborne multi-aspect SAR data. In: IEEE International Geoscience and Remote Sensing Symposium., 1: 7405–7408.

Mather P, Tso B (2009) Classification Methods for Remotely Sensed Data. Environmental engineering. Boca Raton: Taylor and Francis.

Michaelsen E, Soergel U, Thoennessen U (2005) Potential of building extraction from multi-aspect high-resolution amplitude SAR-data. In: Joint workshop of ISPRS/DAGM Object Extraction for 3D City Models Road Databases and Traffic Monitoring., XXXVI-3/W24: 149–154.

Michaelsen E, Stilla U, Soergel U, Leo D (2010) Extraction of building polygons from SAR images: grouping and decision-level in the GESTALT system. Pattern Recognition Letters, 31 (10): 1071–1076.

Milisavljevic N, Closson D, Bloch I (2010) Detecting human-induced scene changes using coherent change detection in SAR images. In: W. W, Szekely B (eds) ISPRS TC VII Symposium, XXXVIII, Part 7B: 389–394.

Moran P (1950) Notes on continuos stochastic phenomena. Biometrika, 37: 17–23.

Mukashema A, Veldkamp A, Vrieling A (2014) Automated high resolution mapping of coffee in Rwanda using an expert bayesian network. International Journal of Applied Earth Observation and Geoinformation, 33: 331–340.

Munich (2014) Referat fuer Gesundheit und Umwelt - Landeshauptstadt Muenchen.

Myint S (2001) A robust texture analysis and classification approach for urban land-use and land-cover feature discrimination. Geocarto International, 16 (4): 29–40.

Myint S, Gober P, Brazel A, Grossman-Clarke S, Weng Q (2011) Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. Remote Sensing of Environment, 115: 1145–1161.

Nations U (2014) World urbanization prospects: the 2014 revision. United Nations - Department of Economic and Social Affairs, Population Division, Technical report.

Newman M (2003) Mixing patterns in networks. Physical Review, 67: 1–14.

Niemeyer J, Rottensteiner F, Soergel U (2013) Classification of urban LiDAR data using conditional random field and random forests. In: Joint Urban Remote Sensing Event: 139–142.

Novack T, Esch T, Kux H, Stilla U (2011a) Machine learning comparison between WorldView-2 and QuickBird-2-simulated imagery regarding object-based urban land cover classification. Remote Sensing, 3: 2263–2282.

Novack T, Kux H (2010) Urban land cover and land use classification of an informal settlement area using the open-source knowledge-based system InterIMAGE. Journal of Spatial Science, 55: 23–41.

Novack T, Kux H, Feitosa R, Costa G (2014) A knowledge-based, transferable approach for block-based urban land-use classification. International Journal of Remote Sensing, 35: 4739–4757.

Novack T, Kux H, Freitas C (2011b) Geocomputation, Sustainability and Environmental Planning, volume 348, chapter Estimation of population density of census sectors using remote sensing data and spatial regression, (pp. 111–122). Berlin:.

Novack T, Stilla U (2014a) Classification of urban settlements types based on space-borne SAR datasets. In: ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, II-7: 55–66.

Novack T, Stilla U (2014b) Discriminative learning of conditional random fields applied to the classification of urban settlement types. In: Gemeinsame Tagung 2014 der DGfK, der DGPF, der GfGI und des GiN, 23: 8p.

Novack T, Stilla U (2015) Discrimination of urban settlement types based on spaceborne SAR datasets and a conditional random fields model. In: ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, II-3/W4: 143–148.

Ok A (2013) Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts. ISPRS Journal of Photogrammetry and Remote Sensing, 86: 21–40.

Okujeni A, Linden S, Tits L, Somers B, Hostert P (2013) Support vector regression and synthetically mixed training data for quantifying urban land cover. Remote Sensing of Environment, 137: 184–197.

Patino J, Duque J (2013) A review of regional science applications of satellite remote sensing in urban settings. Computers, Environment and Urban Systems, 37: 1–17.

Pauleit S, Duhme F (2000) Assessing the environmental performance of land cover types for urban planning. Landscape and Urban Planning, 52 (1): 1–20.

Pesaresi M, Huadong G, Blaes X, Ehrlich D, Ferri S, Gueguen L, Halkia M, Kauffmann M, Kemper T, Lu L, Marin-Herrera M, Ouzounis G, Scavazzon, M. Soille P, Syrris V, Zanchett L (2013) A global human settlement layer from optical HR/VHR RS data: concept and first results. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 6: 2102–2131.

Phinn S, Stanford M, Scarth P, Murray A, Shyy P (2002) Monitoring the composition of urban environments based on the vegetation-impervious surface-soil (VIS) model by subpixel analysis techniques. International Journal of Remote Sensing, 23 (20): 4131–4153.

PUC-Rio (2015) InterIMAGE - interpreting images freely.

Qu Y, Wang J, Wan H, Li X, Zhou G (2008) A Bayesian network algorithm for retrieving the characterization of land surface vegetation. Remote Sensing of Environment, 112 (3): 613–622.

Ratti C, Baker N, Steemers K (2005) Energy consumption and urban texture. Energy and Buildings, 37 (7): 762–776.

Reigber A, Jager M, He W, Ferro-Famil L, Hellwich O (2008) Detection and classification of urban structures based on high-resolution SAR imagery. In: 3rd TerraSAR-X Science Team Meeting.

Rodriguez-Galiano V, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez J (2012) An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing, 67: 93–104.

Ropero R, Aguilera P, Rumí R (2015) Analysis of the socioecological structure and dynamics of the territory using a hybrid Bayesian network classifier. Ecological Modelling, 311: 73–87.

Rottensteiner F, Sohn G, Gerke M, Wegner J, Breitkopf U, Jung J (2014) Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. ISPRS Journal of Photogrammetry and Remote Sensing, 93: 256–271.

Salembier P, Oliveras A, Garrido L (1998) Antiextensive connected operators for image and sequence processing. IEEE Transactions on Image Processing, 7: 555–570.

Sanaieian H, Tenpierik M, Linden K, Seraj F, Shemrani S (2014) Review of the impact of urban block form on thermal performance, solar access and ventilation. Renewable and Sustainable Energy Reviews, 38: 551–560.

Schmidt M, Esch T, Klein D, Thiel M, Dech S (2010) Estimation of building density using Terrasar-X-Data. In: IEEE International Geoscience and Remote Sensing Symposium: 1936–1939.

Schmitt M (2014) Reconstruction of urban surface models from multi-aspect and multi-baseline interferometric SAR. PhD thesis, Technische Universitaet Muenchen.

Schmitt M, Schonberger J, Stilla U (2014) Benefit of using multiple baselines and multiple aspects for SAR interferometry of urban areas. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7 (10): 4107–4118.

Schnitzspan P, Fritz M, Schiele B (2008) Hierarchical support vector random fields: joint training to combine local and global features. In: Forsyth D, Torr P, Zisserman A (eds) European Conference on Computer Vision, volume Part II (pp. 527–540). Berlin: Springer.

Shao J, Tao Shen H, Huang Z, Zhou X (2011) Exploring distributional discrepancy for multidimensional point set retrieval. IEEE Transactions on Multimedia, 13: 71–81.

Soergel U (2010) Radar Remote Sensing of Urban Areas, chapter Review of radar remote sensing on urban areas, (pp. 1–47). Dordrecht: Springer.

Soergel U, Michaelsen E, Thiele A, Cadario E, Thoennessen U (2009) Stereo analysis of high-resolution SAR images for building height estimation in cases of orthogonal aspect directions. ISPRS Journal of Photogrammetry and Remote Sensing, 64 (5): 490–500.

Stabel E, Fischer P (2001) Satellite radar interferometric products for the urban application domain. Advances in Environmental Research, 5 (4): 425–433.

Steemers K (2003) Energy and the city: density, buildings and transport. Energy and Buildings, 35 (1): 3–14.

Steiniger S, Lange T, Burghardt D, Weibel R (2008) An approach for the classification of urban building structures based on discriminant analysis techniques. Transactions in GIS, 12: 31–59.

Stilla U (2007) High resolution radar imaging of urban areas. In: Fritsch D (ed) Photogrammetric Week, 07. Wichmann: Heidelberg: 149–158.

Stilla U, Soergel U, Thoennessen U (2003) Potential and limits of InSAR data for building reconstruction in built up areas. ISPRS Journal of Photogrammetry and Remote Sensing, 58 (1-2): 113–123.

Su X, He C, Feng Q, Deng X, Sun H (2011) A supervised classification method based on conditional random fields with multiscale region connection calculus model for SAR image. IEEE Geoscience and Remote Sensing Letters, 8: 497–501.

Tack F, Buyuksalih G, Goossens R (2012) 3D building reconstruction based on given ground plan information and surface models extracted from spaceborne imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 67: 52–64.

Taubenboeck H, Esch T, Felbier A, Wiesner M, Roth A, Dech S (2012) Monitoring urbanization in mega cities from space. Remote Sensing of Environment, 117: 162–176.

Thiele A, Cadario E, Schulz K, Thönnessen U, Soergel U (2007) Building recognition from multi-aspect high-resolution InSAR data in urban areas. IEEE Transactions on Geoscience and Remote Sensing, 45: 3583–3593.

Thomas I, Frankhauser P, Biernacki C (2008) The morphology of built-up landscapes in Wallonia (Belgium): A classification using fractal indices. Landscape and Urban Planning, 84: 99–115.

Tigges J, Lakes T, Hostert P (2013) Urban vegetation classification: benefits of multitemporal RapidEye satellite data. Remote Sensing of Environment, 136: 66–75.

Tobler W (1970) A computer movie simulating urban growth in the Detroit region. Economic Geography, 46: 234–240.

Tratalos J, Fuller R, Warren P, Davies R, Gaston K (2007) Urban form, biodiversity potential and ecosystem services. Landscape and Urban Planning, 83 (4): 308–317.

Tupin F (2003) Extraction of 3D information using overlay detection on SAR images. In: Joint Workshop on Remote Sensing and Data Fusion over Urban Areas: 72–76.

Turker M, Koc-San D (2015) Building extraction from high-resolution optical spaceborne images using the integration of support vector machine (SVM) classification, Hough transformation and perceptual grouping. International Journal of Applied Earth Observation and Geoinformation, 34: 58–69.

Van de Voorde T, Canters F, van der Kwast J, Engelen G, Binard M, Cornet Y (2009) Quantifying intra-urban morphology of the greater Dublin area with spatial metrics derived from medium resolution remote sensing data. In: Joint Urban Remote Sensing Event: 1–7.

Vanderhaegen S, Canters F (2010) Developing urban metrics to describe the morphology or urban areas at block level. In: GEOBIA 2010: Geographic Object-Based Image Analysis, XXXVIII-4: 1–6.

Vina A, Gitelson AA, Nguy-Robertson AL, Peng Y (2011) Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. Remote Sensing of Environment, 115: 3468–3478.

Voltersen M, Berger C, Hese S, Schmullius C (2014) Object-based land cover mapping and comprehensive feature calculation for an automated derivation of urban structure types at block level. Remote Sensing of Environment, 154: 192–201.

Walde I, Hese S, Berger C, Schmullius C (2013) Graph-based mapping of urban structure types from high-resolution satellite image objects - case study of the german cities Rostock and Erfurt. IEEE Geoscience and Remote Sensing Letters, 10 (4): 932–936.

Walde I, Hese S, Berger C, Schmullius C (2014) From land cover-graphs to urban structure types. International Journal of Geographical Information Science, 28: 584–609.

Walton J (2008) Subpixel urban land cover estimation: comparing cubist, random forests, and support vector regression. Photogrammetric Engineering and Remote Sensing, 74: 1213–1222.

Wasmeier P (2006) Transformations between WGS84 and local coordinate systems. Technische Universitaet Muenchen, Technical report.

Weber N, Haase D, Franck U (2014) Assessing modelled outdoor traffic-induced noise and air pollution around urban structures using the concept of landscape metrics. Landscape and Urban Planning, 125: 105–116.

Wegner J, Montoya-Zegarra J, Schindler K (2013) A higher-order CRF model for road network extraction. In: IEEE Conference on Computer Vision and Pattern Recognition.: 1698–1705.

Weng Q (2012) Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. Remote Sensing of Environment, 117: 34–49.

Whitford V, Ennos A, Handley J (2001) City form and natural processindicators for the ecological performance of urban areas and their application to Merseyside (UK). Landscape and Urban Planning, 57 (2): 91–103.

Woodhouse I (2006) Introduction to Microwave Remote Sensing. Boca Raton: Taylor and Francis.

Wurm M, Taubenboeck H, Dech S (2010) Quantification of urban structure on building block level utilizing multisensoral remote sensing data. In: Earth Resources and Environmental Remote Sensing/GIS Applications, 7831: 78310H–78310H–12.

Wurm M, Taubenboeck H, Roth A, Dech S (2009) Urban structuring using multisensoral remote sensing data by the examples of german cities Cologne and Dresden. In: Urban Remote Sensing Joint Event: 1–8.

Xiao J, Gerke M, Vosselman G (2012) Building extraction from oblique airborne imagery based on robust facade detection. ISPRS Journal of Photogrammetry and Remote Sensing, 68: 56–68.

Yang K, Shahabi C (2007) An efficient k-nearest neighbor search for multivariate time series. Information and Computation, 205 (1): 65–98.

Yu B, Liu H, Wu J, Hu Y, Zhang L (2010) Automated derivation of urban building density information using airborne LiDAR data and object-based method. Landscape and Urban Planning, 98 (4): 210–219.

Yunhao C, Peijun S, Xiaobing L, Jin C, Jing L (2006) A combined approach for estimating vegetation cover in urban/suburban environments from remotely sensed data. Computers and Geosciences, 32: 1299–1309.

Zhang X, Begleiter H, Porjesz T, Wangt T, Litket A (1995) Event related potentials during object recognition tasks. Brain Research Bulletin, 38: 531–538.

Zhong Y, Zhao J, Zhang L (2014) A hybrid object-oriented conditional random field classification framework for high spatial resolution remote sensing imagery. IEEE Transactions on Geoscience and Remote Sensing, 52: 7023–7037.

Lebenslauf

Name	Tessio Novack
Geburtstag und -ort	30.10.1982 in Brazil
Wohnort	Ricarda-Huch-Strasse 1 80807 Muenchen

Ausbildung/Tätigkeit

2002 – 2006	Bachelor of Science in Geography. University of Sao Paulo (USP)
2005 – 2006	GIS expert at the Brazilian Institute for Geography and Statistics, Brazil-Sao Paulo
2006 – 2007	GIS expert at the Forestry Institute of Sao Paulo State, Brazil-Sao Paulo
2007 – 2009	Master of Science in Remote Sensing. Brazilian Institute for Space Research (INPE)
2010 – 2010	Associated researcher at INPE
2010 – 2015	PhD scholarship holder from the Deutscher Akademischer Austausch Dienst

Acknowledgment

I hereby express my gratitude for the financial support given by the Deutscher Akademischer Austauschdienst (DAAD) and for the providing of the image and cadastral data by the Deutsches Zentrum für Luft- und Raumfahrt (DLR) and the Referat für Gesundheit und Umwelt der Stadt München respectively.

I would like to deeply and sincerly thank Prof. Dr.-Ing Uwe Stilla for accepting to be my advisor, for the interesting talks and for the constructive suggestions given along the making of this thesis. His character and scientifc perceptiveness will always be inspiring for me.

I would also like to thank all my colleagues at the Department of Photogrammetry and Remote Sensing of the Technical University of Munich for all the help given when I faced apparently unsolvable problems, for the sharing of thoughts and insights and specially for making the atmosphere at the office a very pleasant one.

With all my love, I thank my father, mother and brother for supporting me in all sorts of ways during this very special, but at times difficult, period of my life.

Most and above all, I thank God.

"Reason has always existed but not always in a reasonable form." - Karl Marx