Universität Augsburg
Institut für Geographie

University of Augsburg
Faculty of Applied Computer Science
Chair of Geoinformatics

Deutsches Zentrum
für Luft- und Raumfahrt

German Aerospace Center
Earth Observation Center
German Remote Sensing Data Center

Master Thesis

SAR Image Feature Analysis for Slum Detection in Megacities

Augsburg, February 13, 2017

Author
MATTHIAS WEIGAND
University of Augsburg
Matriculation No.: 1393747
Mail: matthias.j.weigand@gmail.com

Supervision
DR. MICHAEL WURM
German Aerospace Center
German Remote Sensing Data Center

PROF. DR. SABINE TIMPF
University of Augsburg
Institute of Geography

Acknowledgment

It is a pleasure to thank those who made this thesis possible.

I would like to express my special appreciation and thanks to my supervisor Dr. Michael Wurm who supported me during the thesis. In numerous discussions I experienced his continuous encouragement and constructive criticism which contributed to the success of this thesis. Especially, I would like to thank him for his perpetual support throughout the last year.

Furthermore, I am very grateful for the help of Dr. Andreas Schmitt, his information about multi-polarized SAR data and his help with the processing infrastructure. I also want to thank Dr. Hannes Taubenböck for his support during this thesis. I gratefully acknowledge the help and patience of Detlev Kosmann for his support with the processing infrastructure. Moreover, I would like to thank Martin Klotz for his interesting insights into accuracy assessment.

This study was compiled in cooperation with the *Deutsches Fernerkundungsdatenzentrum (DFD, German Remote Sensing Data Center)* at the *Deutsches Zentrum für Luft- und Raumfahrt (DLR, German Aerospace Center)* in Oberpfaffenhofen, Germany. The data used in this Master Thesis was provided by the TerraSAR-X science proposal LAN0860: *Spatio-Temporal Monitoring of Urbanization*.

I would also like to show my gratitude to Prof. Dr. Sabine Timpf who made this unique cooperation possible. I also would like to thank her for her constructive criticism and support.

I am grateful for the criticism and revisions of Marcel Stelter and Sebastian Größler.

Special thanks go to my family for their emotional support and during the last few years.

Finally, I want to thank Natalie Bienenstein for cheering me up and being there for me.

Contents

List of Figures

List of Tables

List of Abbreviations

AOI	Area of interest
CART	Classification and Regression Trees
CBA	Class Balanced Accuracy
CBR	Closing by reconstruction profile
CP	Closing profile
DMP	Differential Morphological Profile
ER	Error Rate
f	(Wave) Frequency
FS	Feature set
GLCM	Grey level co-occurrence matrix
GSO	Generic Slum Ontology
GUF	Global Urban Footprint
HH	Sent horizontally, received horizontally (polarized SAR signal)
$\hat{K}$	Kappa estimate
λ	Wave Length
LDA	Linear discriminant analysis
LU/LC	Land use / land cover
MDC	Mean standardized Discriminant Coefficient
MM	Mathematical morphology
MP	Morphological profile
OA	Overall Accuracy
OBR	Opening by reconstruction profile
O_m	Mutual Overlap
OOB	Out-of-bag estimate
OP	Opening profile
px	Pixel(s)
PPV	Positive Predictive Value
RF	Random Forest
SAR	Synthetic Aperture Radar
SE	Structuring Element
SEN	Sensitivity
SPE	Specificity
TDX	TanDEM-X
T_O	Threshold of the mutual overlap
TSX	TerraSAR-X
UN	United Nations
VH	Sent vertically, received horizontally (polarized SAR signal)
VI	Variable Importance
VV	Sent vertically, received vertically (polarized SAR signal)

Zusammenfassung

Die starken Veränderungen von urbanen Gebieten der letzten Jahrzehnte betreffen vorwiegend Entwicklungsländer. Besonders hier entwickelten sich Megastädte mit mehr als 10 Millionen Einwohnern, deren Anzahl sich Prognosen zufolge bis 2030 auf 41 erhöhen wird. Der zunehmende Zuzug in die Städte führt unweigerlich zur Entstehung von informellen Siedlungen mit schlechten Lebensbedingungen. Überwachung, Analyse und Kartierung von Slums sind nötig, um die wachsenden Herausforderungen zu meistern.

Daher werden in dieser Studie fernerkundliche SAR-Bilder zur großangelegten Kartierung von Slums in Mumbai, Indien, genutzt. In umfassenden Experimenten werden sowohl Textur- als auch morphologische Bildelemente hinsichtlich ihrer Eignung zur Slumabgrenzung im urbanen Raum analysiert. Hierbei werden zwei moderne, überwachte Klassifikationsalgorithmen miteinander verglichen. Die Klassifikationsgenauigkeit wird mithilfe eines flächendeckenden Referenzdatensatzes bestimmt.

Die Ergebnisse zeigen, dass es möglich ist, die urbane Landschaft unter Zuhilfenahme von Bildtexturmerkmalen und einem Random Forest Klassifikator mit einer Gesamtgenauigkeit von 88.58 % zu klassifizieren. Unter Berücksichtigung der klassenspezifischen Genauigkeitswerte zeigt sich jedoch, dass die Erkennung von Slums weiterhin eine Herausforderung darstellt. Mithilfe einer flächenbasierten Genauigkeitsabschätzung lässt sich erkennen, dass besonders kleinere Slums im urbanen Kontext schwer zu detektieren sind. Ein experimenteller Ansatz zur Merkmalsreduktion zeigt darüber hinaus, dass zur Erkennung von Slums mehr Bildmerkmale nötig sind als für die Abgrenzung von urbanen und nicht-urbanen Flächen.

Abstract

Urban areas have undergone major changes in the last decades. Especially in developing countries, megacities with more than 10 million inhabitants have developed. The number of such huge urban agglomerations is predicted to rise to 41 until 2030. The increasing influx of urban dwellers inevitably leads to the formation of informal settlements with poor living conditions. Monitoring, analysis and mapping of slums are necessary to provide information about such settlements and thus tackle these increasing challenges.

Therefore, SAR remote sensing images are used in this study for an extensive mapping of slums conducted in Mumbai, India. In a broad experimental setup textural and morphological features are analyzed for the slum discrimination in an urban landscape. Two state-of-the-art supervised classification algorithms are compared in the experiments. By utilizing an area-wide reference data set, detailed accuracy assessment techniques are applied to determine the classification quality.

The results indicate that it is possible to classify the urban landscape by using textural image features with an Overall Accuracy of 88.58% utilizing a Random Forest classifier. However, lower class specific accuracies of the slum areas show that slum mapping remains challenging. A patch based accuracy assessment proves that it is most difficult to detect small slum areas in the urban landscape. Furthermore, an experimental feature reduction experiment indicates that more image features are needed to detect slums than to discriminate urban and non-urban landscapes.

1 Introduction

1.1 Background

One of the most crucial challenges for mankind in the 21st century will be the increasing urbanization. It is the most important influencing factor to society's demographic alterations (United Nations 2010). The global urbanization trend of the last decades indicates that the urban living environment has undergone significant changes. Moreover, this trend has been predicted to continue and increase even further in the upcoming years (United Nations 2014).

The lion's share of the changes to urban population will take place in cities which are home to more than one million people (for comparison see figure 1.1). "Megacities", which are cities with more than ten million inhabitants, were home to seven percent of the world's urban population in 1990. By 2014 they increased their share to twelve percent of all urban dwellers around the world living in only 28 megacities. This trend is expected to continue until there will be more than 40 megacities in 2030, most of which being located in the Asian countries (United Nations 2014).

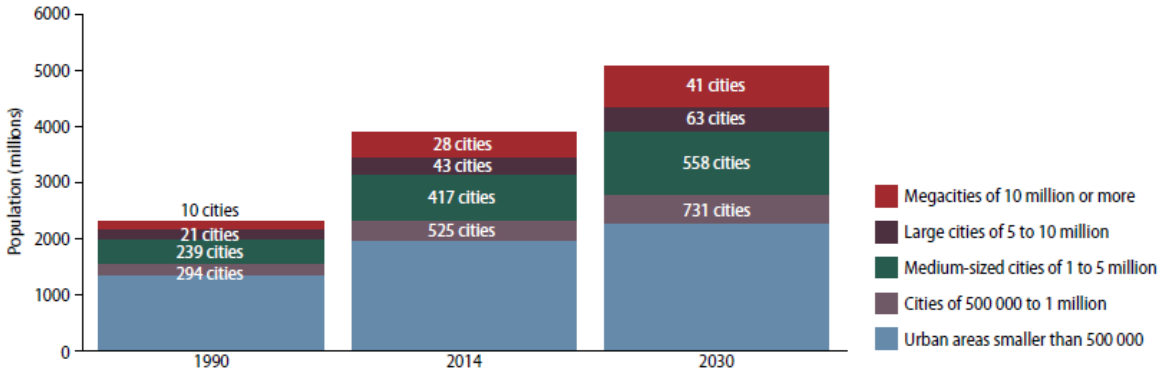


Figure 1.1: Distribution of the world's urban population by city size (United Nations 2014, p. 13)

The most significant changes to urban areas in the last decades took place in the developing countries around the world, where millions of people moved to the cities in the hope of finding labor. This leads to an increasing need of housing at affordable cost and inevitably to sub-standard living conditions and even illegal dwellings when capacities are exhausted (United Nations 2014). If accompanied by unbalanced economical growth, arising informal settlements, also called slums, will be a result of the concentration of poor urban dwellers. They are subsequently about to experience economical, political, cultural and social exclusion (United Nations 2010). On today's *planet of slums* (cf. Davis 2006) approximately more than one billion people are facing these problems and this number is expected to increase up to one and a half billion by 2020 (Arimah 2010). In consequence of the Millennium Development Goal 7D set by the United Nations to *"achieve, by 2020, a significant improvement in the lives of at least 100 million slum dwellers"* several countries, most prominently China, India and Indonesia, tried to face these problems of modern urbanization. Even though many slum dwellers globally experienced improvements in the living environment, for example by means of water sources, living space or sanitation facilities, further effort is needed to improve the living conditions of yet more people around the globe (United Nations 2013).

To be able to identify slums in the first place, it is necessary to define the terms informal

settlement and slum. The United Nations (2010) build their definition of slums upon the small scale of households:

> *"A slum household consists of one or a group of individuals living under the same roof in an urban area, lacking one or more of the following five amenities:*
>
> 1. *durable housing (a permanent structure providing protection from extreme climatic conditions);*
> 2. *sufficient living area (no more than three people sharing a room);*
> 3. *access to improved water (water that is sufficient, affordable and can be obtained without extreme effort);*
> 4. *access to improved sanitation facilities (a private toilet, or a public one shared with a reasonable number of people); and*
> 5. *secure tenure (*DE FACTO *or* DE JURE *secure tenure status and protection against forced eviction). [...]"*

Depending on the cultural region, the country, or their location within the city context, these sub-standard living environments are often named differently. For example Graybill et al. (2016, p. 31) state that *"[...] slums tend to be found in old, run-down areas of inner cities (sometimes, paradoxically on very valuable land)"*. Squatter settlements on the other hand are *"[...] usually located on the outskirts of cities in the developing world"* and are *"typically newer and comprised of makeshift dwellings erected without official permission on land not owned by squatters"*. These settlements lack essential services like electricity or sewerage and no plans of long-term development are conducted in order to improve living conditions. In India, slums have been defined by the Indian government's *Slum Areas (Improvement and Clearance) Act of 1956* as dwellings with buildings which

> *"[...] are by reason of dilapidation, overcrowding, faulty arrangement and design of such buildings, narrowness or faulty arrangement of streets, lack of ventilation, light or sanitation facilities, or any combination of these factors, are detrimental to safety, health or morals [...]".*

Yet, these poor living conditions and health-threatening circumstances do not only affect slum dwellers but can also have an impact on the health of the urban population when interaction is induced by mobility (Graybill et al. 2016). Therefore, it is even more necessary to develop ways to tackle these challenges to modern society. As one possible guideline, the United Nations (2013, p. 50) are building upon the lessons learned by well-functioning improvements of slum settlements and to tackle the growth of slums. They name monitoring and analysis as few of the important steps to improve the lives of many urban dwellers in the future.

In order to raise awareness of the location and spatial extent of such sub-standard living environments, it is necessary to develop mapping techniques which are capable of visualizing, outlining and quantifying slums in relevant regions. Yet, in the definition of slums by the United Nations (2010) several properties of the living environment are pointed out which in part cannot be derived from physical measurements. However, aside from these semantic properties of

slums, the physical environment can be utilized as the visual representation of these indicators (Taubenböck and Kraff 2014). An analysis of slum areas in different cities around the world showed that, despite individual local characteristics, the most important properties of slums are high building densities as well as low building heights, small building sizes and irregular geometric alignment (Taubenböck and Kraff 2015). These properties of the highly dynamic urban bodies require modern mapping techniques to keep track of differences in intra-urban structures such as slums. Therefore, in this study the term slum will be used for any kind of sub-standard or poor living environments which can be derived from physical measurements in urban areas. For reasons of readability, the term *slum* therefore comprises multiple terminologies such as *squatter settlement* or *informal settlement*.

1.2 State of Research

One method of slum monitoring and analysis is the interpretation of modern satellite images. These provide extensive inventories of physical measurements of the Earth's surface. However, in the case of slums, only limited official geographic information about the characteristics of these settlements is available. Data sources include for example census data but such data sources are often inconsistent or lack detailed information about the spatial extent (Kuffer et al. 2016).

In order to describe the physical properties of slums, Kohli et al. (2012) identified unique characteristics of slums from around the world in their Generic Slum Ontology (GSO). These characteristics refer to three spatial levels and are represented in six general indicators (see figure 1.2). While the level of "Environs" refers to the slums within their broader urban context and therefore mostly depends on auxiliary data such as socio-economic status, the levels of "Settlement" and "Object" relate to measurable physical properties of the urban environment at a high level of spatial detail. Today, with modern (very)-high resolution (HR / VHR) imaging satellites, these characteristics have become accessible for mapping and quantitative analysis.

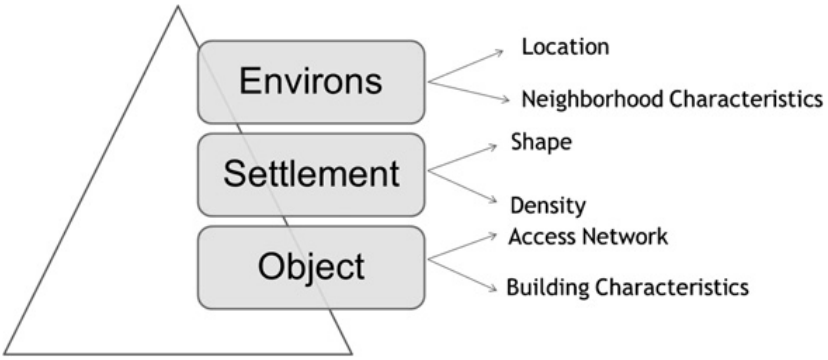


Figure 1.2: General indicators at three spatial levels to describe the Generic Slum Ontology (Kohli et al. 2012)

Following this ontology, the "Object" level indicators include for example building characteristics like the shape, size, materials or orientation. In recent studies it has been demonstrated that these properties can be extracted from VHR imagery. In fact, VHR optical data, topographic maps or high resolution digital elevation models (DEM) were used to delineate building footprints and heights or roofing materials (Niebergall et al. 2008; Tuia et al. 2009; Wurm et al. 2009; Baud et al. 2010; Wurm et al. 2011, 2014; Taubenböck and Kraff 2014). Also, linear geometries like

roads in informal settlements were extracted by Nobrega et al. (2008). However, only few studies concentrated on the extraction of information on Object level (Kuffer et al. 2016). Instead, research focused on the indicators on the "Settlement" level. These describe properties like the building density or shape of the slums on a city block or district level. Compared to formal settlements, slums show extremely high building densities. In order to measure this, in slum mapping textural measures derived from HR and VHR optical sensors have been proven to constitute useful information to describe the settlement density (Pacifici et al. 2009; Graesser et al. 2012; Engstrom et al. 2015; Duque et al. 2015).

While the aforementioned methods were mostly developed using optical sensors, clouds and other atmospheric influencing factors can impair the usability of such satellite images for interpretation. Especially in tropical regions, which most of the developing countries are located in, clouds are predominant throughout the year. Therefore it is difficult to find sufficient and coherent imagery for large areas of megacities.

Active imaging systems like Synthetic Aperture Radar (SAR) on the other hand can provide images even if the circumstances hinder optical sensors from getting unimpaired images. By actively "illuminating" the Earth's surface with long waved electromagnetic waves which penetrate by clouds and aerosols, these sensors are able to provide images under all weather conditions and also during the night. Moreover, SAR allows for acquisition of the shift in polarization of the electromagnetic waves and, by comparison of sent and received signal, draw conclusions about the surface properties.

Especially since the advent of high-resolution sensors like TerraSAR-X and TanDEM-X, SAR data can also be used for analyzing small scale features of the urban landscape. Hence, previous studies showed that individual characteristics of slums used to define the Generic Slum Ontology. For instance, the settlement density was described by textural features for urban land use classification (Du et al. 2015) and also slum discrimination (Dell'Acqua and Gamba 2003; Lisini et al. 2012; Wurm et al. 2017). Furthermore, the analysis of polarization content allows for the inclusion of information about the surface morphology – a promising improvement in urban area mapping (Chaabouni-Chouayakh and Datcu 2010).

However, despite all major progress in slum mapping, most studies focus on small regions for methodological development but do not yet create city-wide not to mention global slum inventories (Kuffer et al. 2016). Hence, the all-weather and all-day acquisition possibilities of (V)HR SAR imagery can significantly improve the data availability for large-area slum mapping.

1.3 Scope of this Study

The aim of this study therefore is to explore the capabilities of high-resolution partially polarized SAR imagery for large-area slum mapping. For this purpose, an extensive study area in the city of Mumbai, India, covering about 550 km² is chosen for investigation. Thus, more than 450 slums of various sizes, structures and within different urban contexts are included in this analysis. The area of investigation was covered by two partially polarized TerraSAR-X and TanDEM-X scenes from 2013.

The morphological differences between slums and other urban bodies influence the backscattering mechanisms of the electromagnetic waves of these SAR sensors. By utilizing the Ken-

naugh matrix these scattering mechanisms are resolved from the partially-polarized signal. Therefore, an existing framework is used to generate a set of Kennaugh elements (Schmitt et al. 2015) to include the polarization content of the actively acquired SAR images to improve slum mapping.

In a broad experimental setup, the discriminatory influence of different image characteristics for slum detection is examined. In addition to the polarization content of the images also textural and morphological elements were included to describe the slums on the Settlement level following the GSO. Therefore, textures based on the Gray Level Co-occurrence Matrix (GLCM) (Haralick et al. 1973) and differential morphological profiles (DMP) (Benediktsson et al. 2003) are derived. Afterwards the two modern machine learning algorithms Linear Discriminant Analysis (LDA) and Random Forest (RF) are used in a supervised classification scheme to derive labeled maps. Throughout the study these two classifiers are compared with respect to their classification quality in the context of slum mapping with SAR images.

Since slum dwellings often occupy only a small share of a city's total area (Taubenböck and Kraff 2015), a pixel based classification has to take this imbalanced class distribution into account. This very common problem in the field of data mining (Weiss 2004) also occurs in remote sensing classification approaches (e.g. Wright and Gallant 2007; Williams et al. 2009). Therefore, it is necessary to account for this imbalance both during the training data sampling for the creation of a supervised classification model as well as during the accuracy assessment of the results. Since accuracy measures aggregate myriads of correctly or incorrectly classified pixels into single numbers, it is important to be able to interpret the results correctly considering an imbalanced distribution. Thus, several sampling techniques and class specific accuracy assessment measures are discussed and applied to the case of slum detection.

This case study of slum mapping in the megacity Mumbai using SAR image features is guided by the following research questions:

1. Is it possible to delineate slums and formal settlements with partially polarized SAR imagery?

2. Which textural or morphological image features are best fitted for slum classification?

3. How can the classification accuracy be assessed sufficiently in an imbalanced setup?

At the beginning of this thesis the used image and reference data are introduced in detail and the creation of textural and morphological image features is described in section 2. Afterwards, in section 3, the methodological background of image classification is presented and training data sampling as well as accuracy assessment are approached on a theoretical level before elaborating the methods used in this study. Additionally, the experimental setup of the case study is described in detail in this section. Subsequently, The results are presented in section 4 which is followed by a detailed discussion in section 5 before the findings are concluded in section 6.

2 Image Data and Preparation

To analyze the suitability of SAR image features for slum mapping, the Indian city Mumbai was chosen as the study area. Several data sets were used to validate the classification results. In section 2.1 of this chapter the area of investigation will first be outlined with a short profile of the city followed by a description of the used reference data sets in section 2.2. Then, the used satellite data is described in detail in section 2.3. Afterwards, the calculated image features are introduced in section 2.4.

2.1 Study Area

Mumbai is located on the western coast of the Asian subcontinent India. Settled since the stone age the city of Mumbai, formerly called Bombay, was once of great importance for Portugal's and Britain's colonial empires. The city grew to become one of India's richest cities and the financial and commercial center of the country (Pacione 2006).

Mumbai's population increased from 3 million in 1950 to 12.5 million in 2011 within the city's administrative boundaries covering about 600km². The metropolitan area of Greater Mumbai counts about 21 million inhabitants (UNPD 2011). Since 1975, the constructed area has increased nearly eight-fold due to the enormous influx of people and restrictive urban policy (Taubenböck and Kraff 2015).

Despite the city's wealth, various slums are situated in Mumbai, including *Dharavi* which is said to be one of the largest slums in Asia. This slum alone is estimated to be home to about 300,000 to 1 million dwellers (Pacione 2006; Taubenböck and Kraff 2014). Depending on the reference the overall share of slum dwellers in Mumbai varies well between 39% (Pacione 2006, p. 235), 55% (Asha 2006) and up to nearly 60% (Dutt et al. 2016, p. 408). Even if they probably represent more than the half of the population of Mumbai, they only occupy about 15% of the urban area (Taubenböck and Kraff 2015). Therefore, the city of Mumbai is well-suited to develop and test slum mapping techniques.

2.2 Reference Data

Due to acquisition properties SAR, two areas of interest (AOIs) are selected which are presented as black boxes in figure 2.1. For this area of investigation, an area-wide set of reference polygons is used to identify the prevalent land cover types on the ground (Taubenböck and Wurm 2015). This data contains three semantic classes, namely *slums*, *urban* and *other*. While the first two classes include the urban land cover inventory such as informal and formal settlements, office buildings or industrial areas, the latter comprises all natural land cover types like vegetation, water and bare soil (see figure 2.1).

The reference for the land use / land cover (LU/LC) class *slum* was acquired from visual image interpretation of VHR optical satellite imagery by Taubenböck and Wurm (2015) and is manually adapted to the acquisition date of the used satellite images. The spatial level of detail is specified to city block level. Since the acquisition of the class *slum* is limited to the physical appearance of this specific settlement type, in this study the term slum is used for settlements which fit several morphological properties like

- densely built area (> 50% covered by houses / dwellings),
- low building heights (1 - 2 stories),
- low accessibility of major roads (only small paths between the dwellings),
- low amount of green space,
- mostly flat-roofed houses with low-quality building materials.

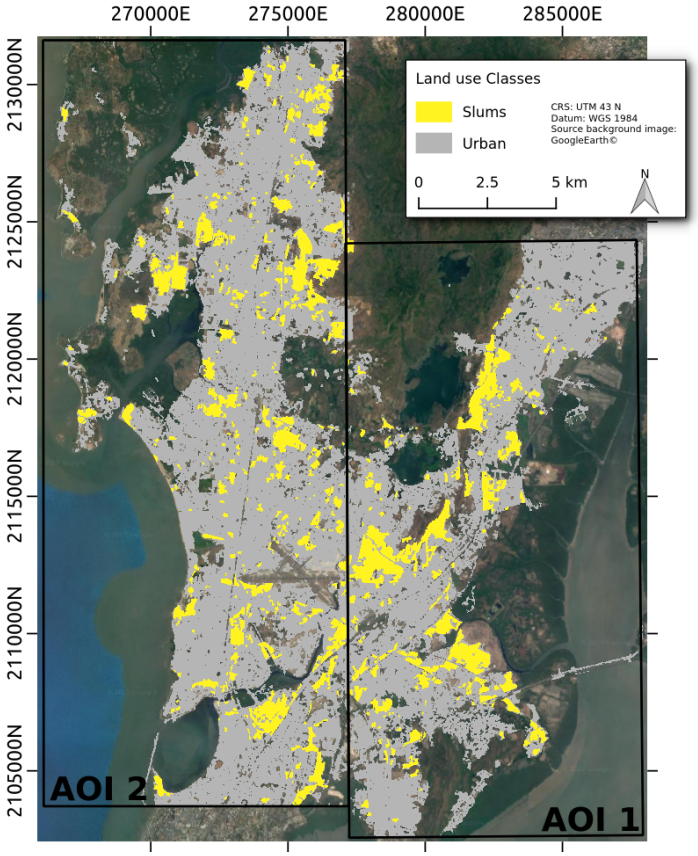


Figure 2.1: Reference data in Mumbai for the LU/LC classes *slum* and *urban*

An example of a slum in Mumbai within its urban context is presented in figure 2.2. The small shacks are built densely together with a homogeneous low building height. A chaotic system of small paths connects to only very few bigger access roads. Furthermore, the building materials, especially the rusted corrugated iron roofs, indicate lacking resources for high-quality housing. This example of the Dharavi slum also gives an impression of how slums are embedded next to non-slum environments houses.

The derived slum areas of Mumbai vary in size, shape and location in the city. Since it is the aim of this study to map slums based on textural and morphological image features, the slums' sizes are of major importance. The two AOIs hold a total of 454 slum patches (152 in AOI 1 and 302 in AOI 2) which cover about 1400ha and 2000ha, respectively. The distribution of the sizes of the individual slum patches for each AOI is depicted in figure 2.3a. The histograms indicate that most of the slums in this area are smaller than ten hectares and only few patches are larger

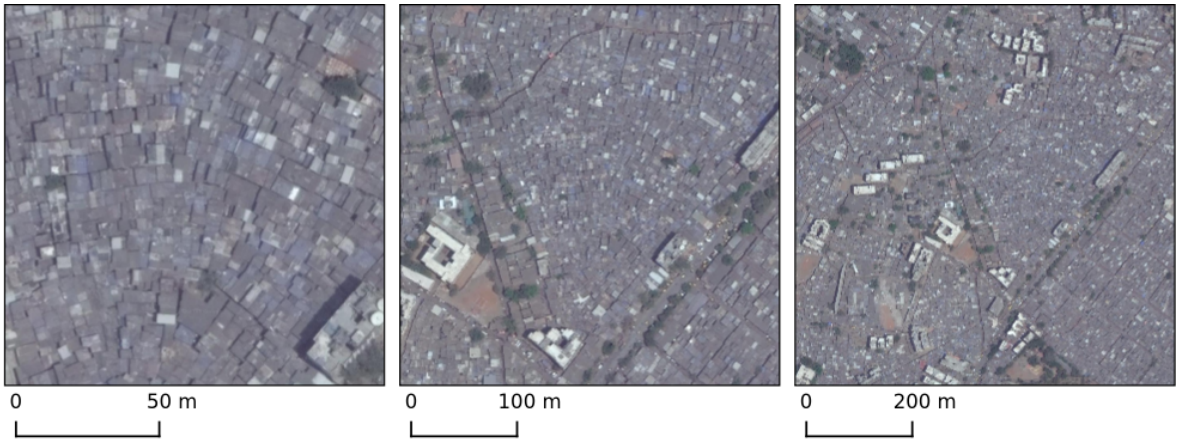
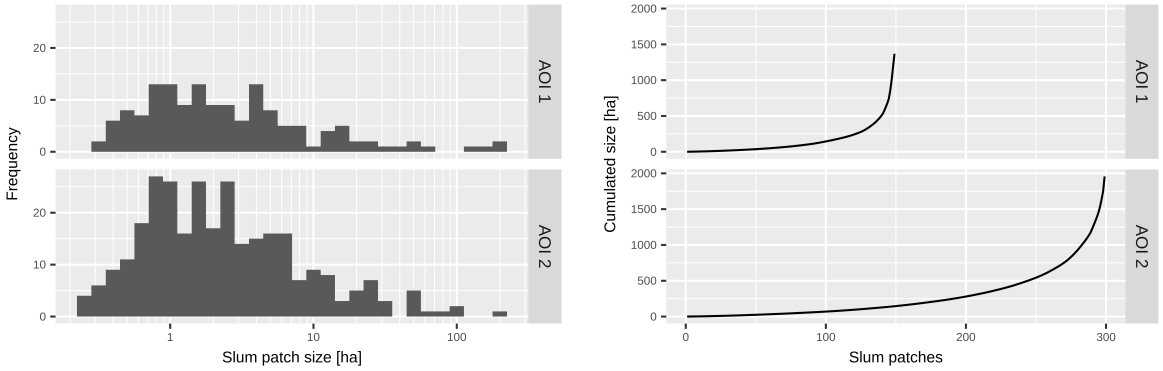


Figure 2.2: Dharavi slum in Mumbai at different scales (Image source: GoogleEarth©)



(a) Patch size distribution **(b) Cumulative size**

Figure 2.3: Slum patch size analysis

than that. This is also expressed by the cumulative size (figure 2.3b). The major share of the total slum area is held by only few huge patches. These differences in the sizes of slums have to be considered in the accuracy assessment, since especially the small slum patches (with a size of about 50m $\times$ 50m) are probably nested in other LU/LC classes and are therefore more difficult to delineate with the used image features (see figure 2.4).

The reference area for the land cover class *urban* is derived from the Global Urban Footprint (GUF) (Esch et al. 2013). It holds all man-made built-up urban landmass on the Earth's surface as binary image layer. With its 12m spatial resolution it provides fine-scale information about inventories of urban land cover with global coverage and can therefore also be used in regional studies (Klotz et al. 2016). However, the original GUF contains many artifacts of the *non-urban* class within urban areas holding very few pixels which are reduced by a 3 $\times$ 3 majority filter as shown in figure 2.4. By utilizing this unique data set the LU/LC class *urban* is available with an area-wide reference as well. From its original spatial resolution of 12m, the GUF is adapted to fit the resolution of the used satellite imagery, that is 2.5m.

The third LU/LC class *other* was assumed for all areas where neither the land use *urban* nor *slum* was prevalent. This class then included areas covered with water, forests, shrubs, bare soil

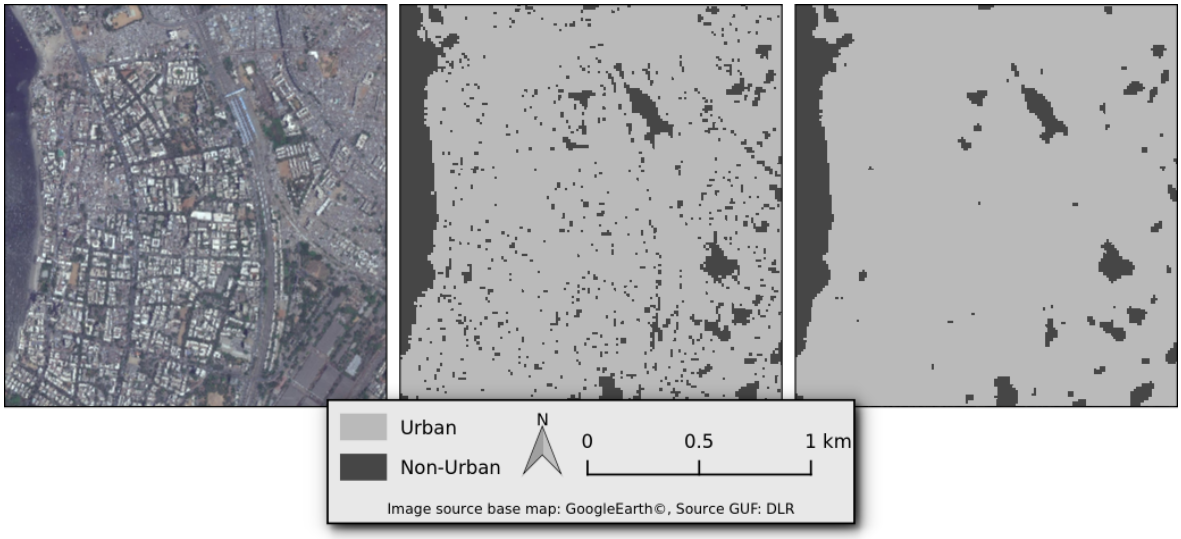


Figure 2.4: Global Urban Footprint before (middle) and after (right) artifact reduction

or other natural land cover. Moreover, since the GUF only includes built-up man-made structures, it has to be noted that also large, homogeneously flat surfaces like airport runways or parking lots were not included in the GUF and thus labeled as *other*.

2.3 Partially Polarized SAR Data

Originally developed based on a system to aid aircraft navigation, Radio Detection And Ranging (RADAR) imaging systems are used for remote sensing. In contrast to optical sensors, RADAR systems "illuminate" the Earth's surface, that is they actively send pulses of microwaves towards the ground and receive the backscattered signal (Moore et al. 1983, p. 430 f.). The wavelength λ of these microwaves is in a range of centimeters. Its length determines the penetration of surface materials as well as the minimum surface roughness needed for backscattering effects (Elachi 1987, p. 170). The way the microwaves are sent and received differs between *Real Aperture Radar* (RAR) and *Synthetic Aperture Radar* (SAR) systems. In RAR sensors the length of antenna sending out the microwaves determines along-track ground resolution, also called azimuth (Moore et al. 1983, p. 434). In order to generate high or very high ground resolutions, the antenna would have to be very long, which is not practicable in operational satellites (Elachi 1987, p. 203). In contrast, SAR systems synthesize a long antenna in order to achieve such high or very high resolutions (Moore et al. 1983, p. 445).

SAR sensors compensate the missing long antenna to generate high azimuth resolutions by using the Doppler effect. Multiple overlapping microwave beams illuminate the Earth's surface as illustrated utilizing a point target in figure 2.5. Thus, objects on the ground are illuminated multiple times. However, since one object should only be mapped once, the nearest position to the sensor has to be determined. This position can be identified by inferring it from the frequency f with which the wave has left the sensor. Waves that are backscattered by objects ahead of the sensor are compressed due to the Doppler effect, that is their frequency increases. In contrast, waves backscattered by objects lying behind the sensor are stretched, i.e. their frequency

decreases. Yet, the frequency is not changed when the object lies exactly beneath the sensor. Taking advantage of this physical phenomenon, it is possible to determine the wave which is sent out at the nearest point to an object (Elachi 1987, p. 207).

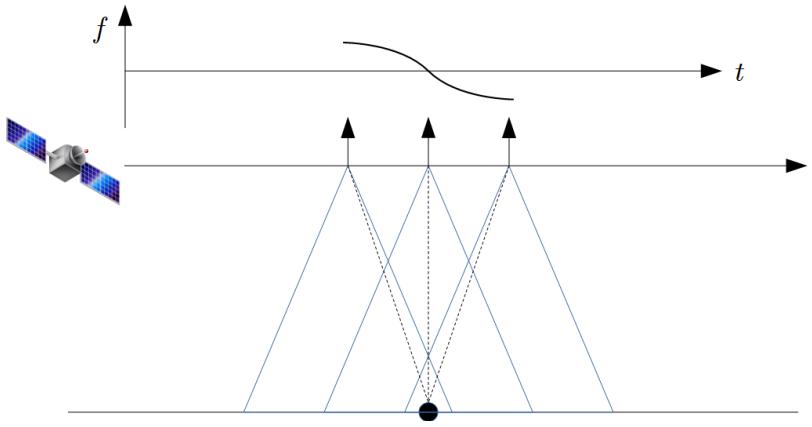


Figure 2.5: Doppler history of a point target as recorded by a SAR Sensor sending overlapping pulses of electromagnetic waves (adapted from Elachi (1987, p. 207))

In contrast to passive optical sensors, the backscattered signal of active SAR depends on the orientation of objects on the ground in relation to the sensor and its flight direction. Therefore, similar objects might have different signatures in the backscattered signal. Moreover, topography has a strong influence position of objects in the image since the distance to the sensor and therewith the travel time of the signal is changed.

The backscattered signal does not only contain information about the intensity but also about the polarization of the wave as a complex signal. Since it is also possible to determine the polarization incidence, which can be both horizontal (H) and vertical (V), four combinations of sent and received signal are possible: *HH*, *HV*, *VV* and *VH*. The polarization of the backscattered signal is influenced by the composition of the objects on the ground (Elachi 1987, p. 168 f.). The four polarizations (i.e. “quad-pol”) can be summarized in the Sinclair scatter matrix S:

$$S = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix} \tag{1}$$

By decomposing the polarizations, it is possible to differentiate the scattering mechanism which the signal has been exposed to. These mechanisms include odd bounce (e.g. on flat surfaces), double bounce (e.g. on house walls) and volume scattering (e.g. on vegetated land). Yet, since the acquisition of all four polarizations would decrease the spatial resolution, modern satellite systems do predominately acquire dual-pol SAR images (Schmitt et al. 2015). These polarizations are HH & VV as co-pol and HH & VH or VV & HV as cross-pol. However, common decomposition methods cannot be applied to these dual-pol images. To overcome this limitation, Kennaugh elements use the elements of the scatter matrix for decomposition (Schmitt and Brisco 2013). In this study the TSX and TDX images with polarization modes HH/VV and VV/VH are used. Their Kennaugh elements derived for co-pol data are

$$
\begin{aligned}
k_0 &= \frac{1}{2}\{|S_{HH}|^2 + |S_{VV}|^2\} \\
k_3 &= -Re\{S_{HH}S_{VV}^*\} \\
k_4 &= \frac{1}{2}\{|S_{HH}|^2 - |S_{VV}|^2\} \\
k_7 &= Im\{S_{HH}S_{VV}^*\}
\end{aligned} \tag{2}
$$

and for cross-pol

$$
\begin{aligned}
k_0 &= |S_{VV}|^2 + |S_{VH}|^2 \\
k_1 &= |S_{VV}|^2 - |S_{VH}|^2 \\
k_5 &= Re\{S_{VH}S_{VV}^*\} \\
k_8 &= -Im\{S_{VH}S_{VV}^*\}
\end{aligned} \tag{3}
$$

respectively. That is, k_0 measures the intensity of the backscattered signal. According to Schmitt and Brisco (2013), the Intensity k_0 and intensity differences k_3, k_4 etc. are normalized to achieve a uniform distribution.

$$k_0 = \frac{k_0 - 1}{k_0 + 1} \; and \; k_i = \frac{k_i}{k_0} \; for \; i \in [1, \ldots, 9] \tag{4}$$

These elements are grouped by the scattering events *Absorption* ($k_{1,2,3}$), *Diattenuation* ($k_{4,5,6}$) and *Retardance* ($k_{7,8,9}$) and the three coordinate directions parallel ($k_{1,4,7}$), diagonal ($k_{2,5,8}$) and circular ($k_{3,6,9}$) (cf. figure 2.6) (Kerwien 2007). Kennaugh elements are directional by definition which is why they range from -1 to 1. Additionally, the Kennaugh element framework (Schmitt et al. 2015) derives the absolute values of these scattering events. They are also used in this study to include both directional and non directional information about the scattering events. Furthermore, *Polarizance* reflects the l^2-norm of all Kennaugh elements k_{1-9} and an additional *Scale* layer derives information about non coherent scattering from multi-looking calculations.

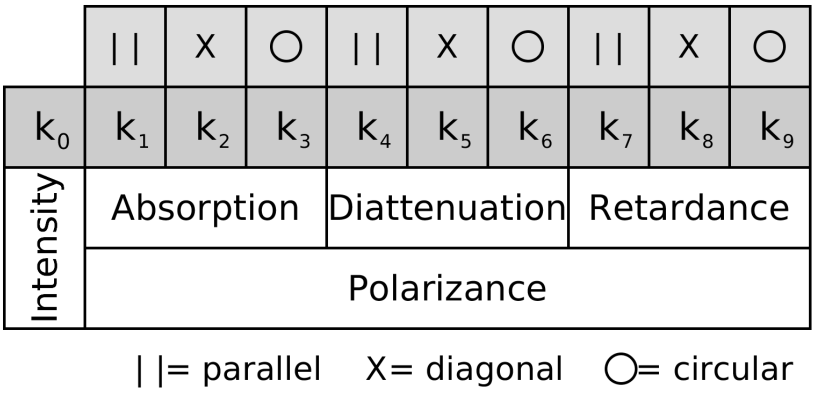


Figure 2.6: Kennaugh elements and derived scattering events

The data used in this study has been acquired by the sensors TerraSAR-X (TSX) and TanDEM-X (TDX) in X-band (that is with a wavelength of $\lambda = 3cm$). The individual acquisition properties and times are summarized in table 2. The Kennaugh elements were provided by

the automated Kennaugh-element framework (Schmitt et al. 2015). Figure 2.7 shows selected Kennaugh element images of the Dharavi slum in Mumbai. Water and vegetation (in the northern part of the image) are represented as more homogeneous and darker areas, whereas the urban and slum areas show more heterogeneous and brighter values.

Table 2: Acquisition properties of SAR image data

Polarization	AOI	Acquisition date	Flight direction	Sensor	Incidence angle
HH/VV	1	09/29/2013	Descending	TSX	33.7°
VV/VH	1	11/12/2013	Descending	TDX	33.7°
HH/VV	2	10/10/2013	Descending	TDX	34.7°
VV/VH	2	04/12/2013	Descending	TSX	34.7°

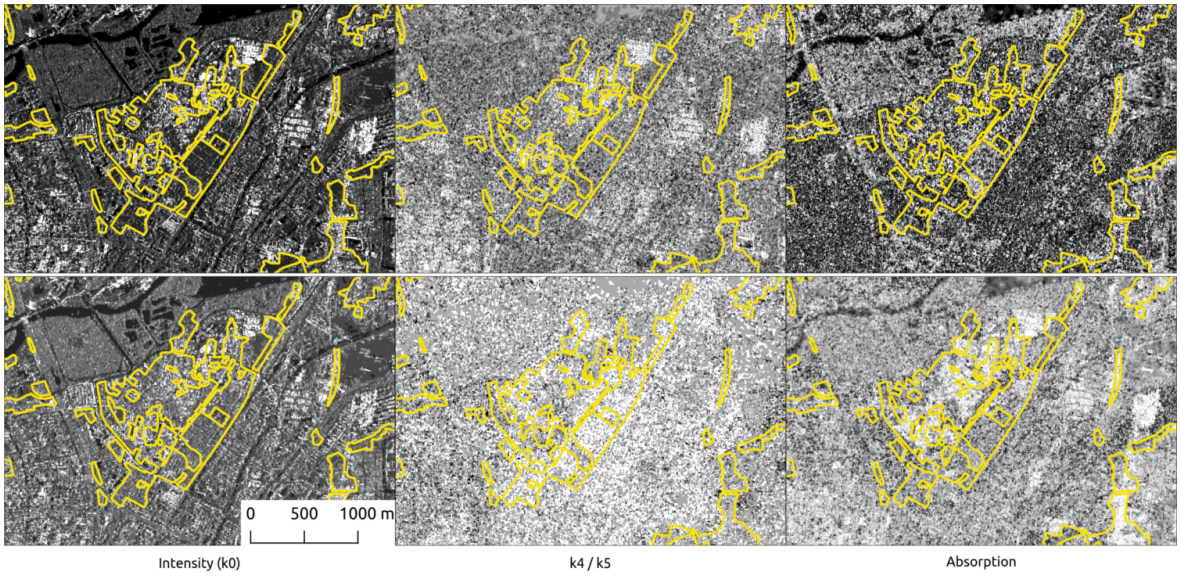


Figure 2.7: Kennaugh elements for HH/VV (top) and VV/VH (bottom) polarizations with slum outlines (yellow)

2.4 Image Features

In order to describe the image content, different image features are calculated for the used partially polarimetric information from Kennaugh-elements. The selection of these image features is based on the spatial domain of the reference data, that is city block level. This refers to the Settlement level in the GSO. Hence, textural parameters are investigated based on the *Grey Level Co-occurrence Matrix* and additionally *Differential Morphological Profiles* are applied. Both describe the images based on the spatial neighborhood of the pixels. Their derivations are explained in detail in the following sections.

2.4.1 Gray Level Co-Occurrence Matrix

The Grey Level Co-Occurrence Matrix (GLCM), introduced by Haralick et al. (1973), is an established way of describing texture in gray scale images based on gray tone spatial dependencies. Thus, they are well suited for object description in SAR images by means of neighborhood-aware enrichment (Shanmugan et al. 1981). GLCM have been proven to provide useful additional image features in various studies in the field of remote sensing (e.g. Soh and Tsatsoulis 1999; Habermeyer and Schmullius 1997; Pesaresi et al. 2008; Pacifici et al. 2009; Tuia et al. 2009; Graesser et al. 2012; Du et al. 2015; Masjedi et al. 2016).

The GLCM features describe the texture of an image I based on the neighboring conditions of individual gray levels within a specific area. This neighborhood, also called kernel K, is defined as a matrix with a size of $n \times n$ pixels. During the GLCM calculation the kernel is applied on the whole image I in a moving-window approach in a way so that at one point in the course of the calculation every pixel of the original image takes up the position of the kernel's centering pixel. Subsequently, the co-occurrence for each pixel can be calculated as described in the following and depicted in figure 2.8.

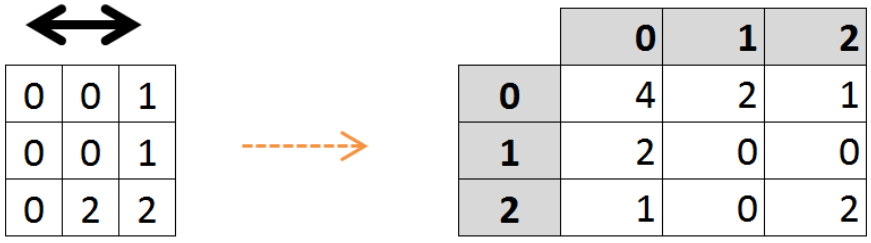


Figure 2.8: Example of a image kernel holding gray values (left) and the resulting secondary gray level co-occurrence matrix (right)

The pairs of neighboring pixels within the kernel K are transferred to a derivate matrix which then holds the number of co-occurrences of gray value pairs. This co-occurrence matrix is defined as an $m \times m$ matrix with m being the number of unique gray values in the image I. The distance between two compared pixels can be defined by the shift S, with $S = 1$ indicating a direct neighborhood of the respective pixels in a certain direction. Figure 2.8 presents an example of a simple 3×3 kernel (left) within an image. For this kernel the horizontal neighborhood with a shift of 1 is transferred to the derivate matrix (right). In this example the combination $0 \rightarrow 0$ or $0 \leftarrow 0$ occurs 4 times as it is transferred to the co-occurrence matrix. GLCM can be calculated in four directions (i.e. horizontally, vertically and (twice) diagonally) as shown in figure 2.9. By taking the average of all four directions (0°, 45°, 90° and 135°), a rotation invariant GLCM feature set is created.

On the basis of the co-occurrence matrix thusly created (cf. figure 2.8 right), several second-order metrics can be calculated using matrix algebra. Resulting values can then be returned into the centering pixel of K. In this study the *Mean*, *Variance*, *Homogeneity*, *Contrast*, *Dissimilarity*, *Entropy* and *Angular Second Moment* are calculated which are available in the *glcm* R package (Zvoleff 2016).

It is important to mention that such a moving-window approach leads to a smoothing of the

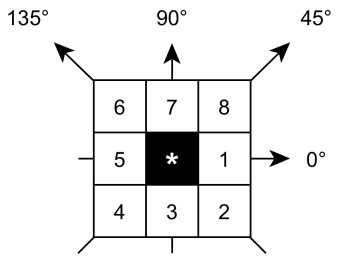


Figure 2.9: Possible shift directions for gray values in neighborhood of 9 connected pixels

image. This effect increases with larger kernel sizes as the percentage of change by moving of the kernel to the next pixel decreases with a higher amount of respected pixels. That is, by moving a 3×3 matrix one pixel forward 33% of the pixels are changed. Yet, when moving a 5×5 matrix one pixel further only 20% of the pixels are replaced. In other words, the ratio of changing pixels by moving a $n \times n$ kernel is therefore equal to $\frac{1}{n}$.

The dimensions of the kernels were chosen with regard to the size of the land use class of interest, namely *slums*. It was furthermore assumed that the size of kernel K would influence the quality of textural features for slum detection. Hence, the sizes of the kernel were set to 11×11, 21×21, 41×41, 81×81, 121×121 and 161×161 to test this hypothesis. Considering an image resolution of 2.5m, these kernels correspond to boxes with an edge length of 27.5m, 52.5m, 102.5m, 202.5m, 302.5m and 402.5m, respectively. Figure 2.10 shows the original HH/VVk_0 image and two derived GLCM textures of the Dharavi slum.

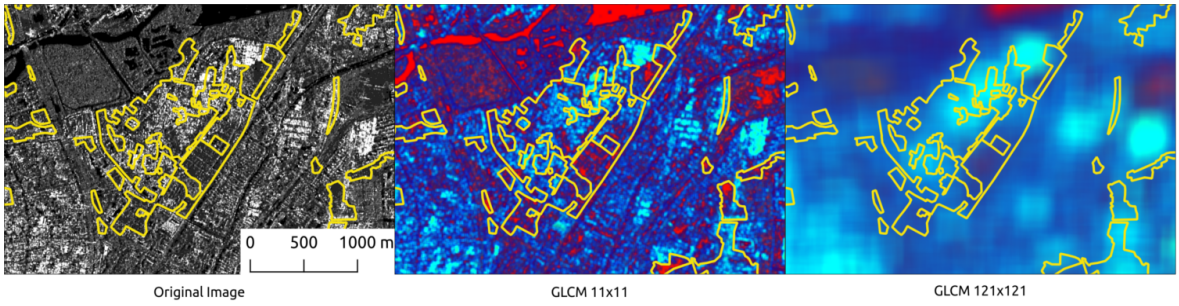


Figure 2.10: Example of a SAR image and its derived GLCM textures; Coloring of GLCM images: red: Angular second moment, green: Variance, blue: Entropy

It was found that the computational effort was heavily dependent on the number of unique gray levels in the image and with that the size of the co-occurrence matrix. Therefore, the original images were resampled to a value range from 0 to 10.

These 7 GLCM individual measures were calculated for the 9 Kennaugh element layers of 2 polarization modes, resulting in sets with a total of 126 layers. Furthermore, these sets were derived for 6 different kernel sizes and for both areas of interest. This highly dimensional data in combination with the reference data was later used as input to build supervised image classification models.

2.4.2 Mathematical Morphologies and Morphological Profiles

Additionally to GLCM textures, Differential Morphological Profiles (DMP) are used in this study to expand the image feature space. These profiles of mathematical morphology (MM), which was first introduced by Serra (1982), are based on set theory. Initially used to describe the shape of objects within a binary matrix environment, MM were adapted to gray scale images by Sternberg (1986) and Haralick et al. (1987). In this context, gray scale images are considered as topographical reliefs in which higher and lower gray values comply with higher and lower elevation, respectively (Sternberg 1986; Tuia et al. 2009). Mathematical morphology and morphological profiles have been applied in various settings for image analysis and remote sensing (e.g. Bellens et al. 2008; Fauvel et al. 2008; Tuia et al. 2009; Waske et al. 2009; Du et al. 2015; Geiß et al. 2016).

The basic morphological operators are *Erosion* and *Dilation* which are depicted in figure 2.11. The structuring element (SE) – in this case a circular object – is applied to the binary image that holds an original object. The *Erosion* ϵ_N of an image $f(p)$ is the point-wise minimum of all gray values within the structuring element $N_G(p)$

$$\varepsilon_N f(p) = \{\wedge f(p')|p' \in N_G(p) \cup f(p)\}, \tag{5}$$

whereas the *Dilation* δ_N of the image is defined as the point-wise maximum within the structuring element (Pesaresi and Benediktsson 2001; Tuia et al. 2009)

$$\delta_N f(p) = \{\vee f(p')|p' \in N_G(p) \cup f(p)\}. \tag{6}$$

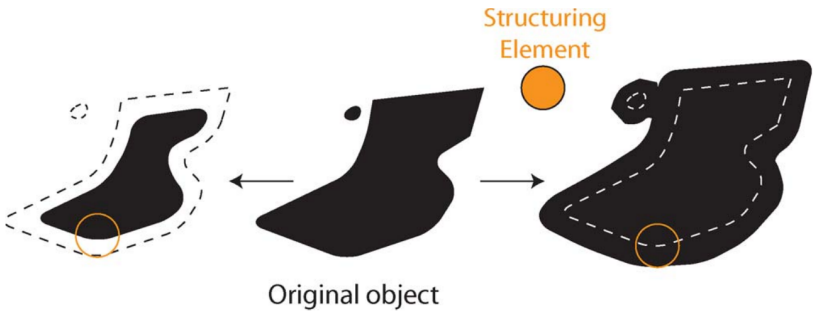


Figure 2.11: Example of Erosion (left) and Dilation (right) of a binary image using a circular structuring element (Tuia et al. 2009)

Consequently, in gray scale images erosion promotes dark features, dilation on the other hand promotes bright objects. Emerging from the combination of these two morphological operators, *Opening* $\gamma_N f(p)$ is the Dilation of an eroded image as

$$\gamma_N f(p) = \delta_N \varepsilon_N f(p), \tag{7}$$

whereas *Closing* $\varphi_N f(p)$ is the Erosion of a dilated image (Pesaresi and Benediktsson 2001) as

$$\varphi_N f(p) = \varepsilon_N \delta_N f(p). \tag{8}$$

Moreover, especially in gray scale images extending the morphological operators Opening and Closing with *Reconstruction* provide useful information for image analysis (Vincent 1993). Thus, two additional operators arise, *Opening By Reconstruction* (OBR) and *Closing By Reconstruction* (CBR). Basically, original gray values are opened by reconstruction when an eroded image I is reconstructed with the point-wise minimum of its iterative dilation J under I until stability is reached (Pesaresi and Benediktsson 2001).

$$\rho^I(J) = \bigvee_{n\geq 1} \delta^I_{(n)}(J) \mid \delta^I_{(n)} = \delta^I_{(n+1)} \tag{9}$$

In contrast, closing by reconstruction can be described as the dilated image I being reconstructed with the point-wise maximum of its iterative erosion J under I until stability is reached.

$$\rho^{*I}(J) = \bigwedge_{n\geq 1} \varepsilon^I_{(n)}(J) \mid \varepsilon^I_{(n)} = \varepsilon^I_{(n+1)} \tag{10}$$

It occurs, that the size of the structuring element influences the detectability of the individual object. Thus, small objects would be detected by a small SE and eliminated by a large one. Likewise, larger objects would be discovered by large SE while they would show high heterogeneity with small SE. Within the urban context one can assume a high variability of object sizes. Therefore, Pesaresi and Benediktsson (2001) added a multi-scale approach to mathematical morphologies: *Morphological Profiles* (MP). Morphological Profiles are created by the combination of Opening and Closing with differently sized structuring elements. The morphological opening profile $\Pi\gamma(x)$ is defined as

$$\Pi\gamma(x) = \{\Pi\gamma_\lambda : \Pi\gamma_\lambda = \gamma^*_\lambda(x),\ \forall\ \lambda\ \in [0,\ldots,n]\}, \tag{11}$$

whereas the morphological closing profile $\Pi\varphi(x)$ is defined as

$$\Pi\varphi(x) = \{\Pi\varphi_\lambda : \Pi\varphi_\lambda = \varphi^*_\lambda,\ \forall\ \lambda\ \in [0,\ldots,n]\}, \tag{12}$$

where $\gamma^*_0(x) = \varphi^*_0(x) = I(x)$ is the original value of the image I at point x. Consequently, these profiles hold the original image I and all morphological Openings and Closings (By Reconstruction) of n differently sized structuring elements. Therefore, it is possible to calculate the difference between each adjacent layer, resulting in a so called *Differential Morphological Profile* or *Derivate Morphological Profile* (DMP) (Pesaresi and Benediktsson 2001; Benediktsson et al. 2003). The differential opening profile is defined as

$$\Delta\gamma(x) = \{\Delta\gamma_\lambda : \Delta\gamma_\lambda = |\Pi\gamma_\lambda - \Pi\gamma_{\lambda-1}|,\ \forall\ \lambda\ \in [1,\ldots,n]\} \tag{13}$$

while the differential closing profile is defined as

$$\Delta\varphi(x) = \{\Delta\varphi_\lambda : \Delta\varphi_\lambda = |\Pi\varphi_\lambda - \Pi\varphi_{\lambda-1}|,\ \forall\ \lambda\ \in [1,\ldots,n]\}. \tag{14}$$

When combined to a full profile, DMPs hold information about which structures in the image can be detected by the different sized structuring elements. As figure 2.12 shows, small structures

like houses or huts are better detected by smaller $N_G(p)$, whereas larger structuring elements appear to incorporate structures on a larger city block level.

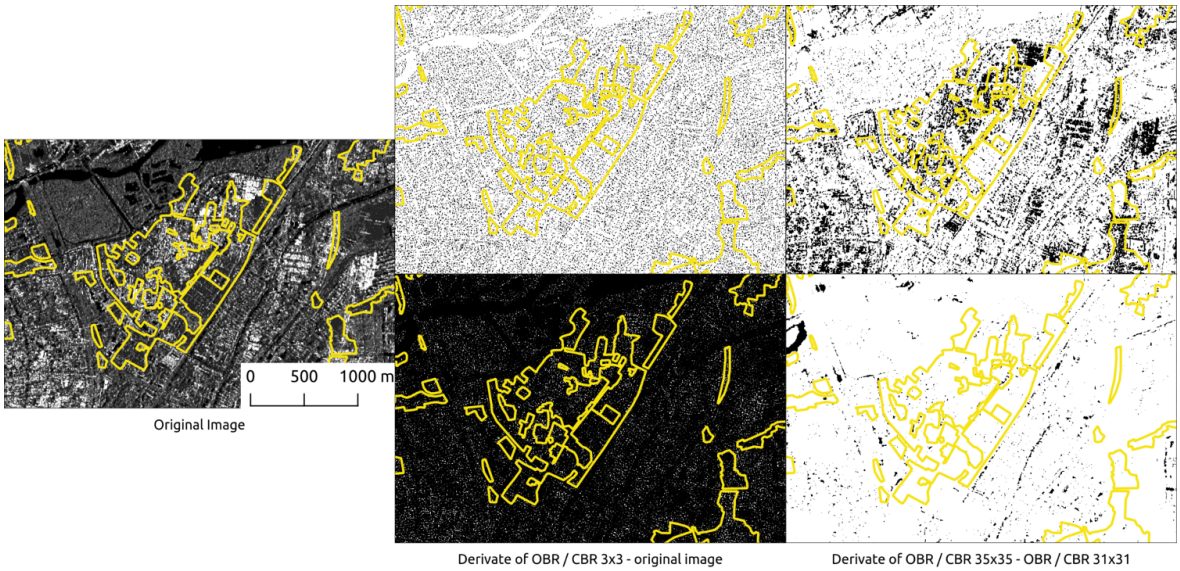


Figure 2.12: Derivate of a Morphological Profile of HH/VVk_0. Slums for orientation shown in yellow. Top: Opening By Reconstruction, bottom: Closing By Reconstruction

In this study, the basic morphological operators are calculated for the image data using the *EBImage* package in R (Pau et al. 2010). However, since reconstruction filtering is not available in the package, it is implemented as proposed by Vincent (1993, p. 18).

In this study, DMPs are calculated using a box shaped kernel as structuring element with sizes from 3 to 35 pixels with an increment of 4. This resulted in a profile depth of 9 layers for each opening and closing and is applied to the intensity Kennaugh elements k_0 layers of both polarization modes, thereby generating a feature space of 36 layers.

3 Image Classification

After introducing the used reference and image data, in this chapter the image classification process is described in detail. First, the classification approaches used in this study are outlined in section 3.1. Afterwards, the creation of a sufficient training data sample is exemplified on a theoretical basis in section 3.2 before introducing the overall and class specific accuracy assessment methods on an abstract level in section 3.3. At the end of this chapter in section 3.4, the experimental setup of this study is outlined.

3.1 Image Classification

The creation of thematic maps using the properties of satellite images is called classification. Thematic classes are assigned to each pixel by an algorithm based on the characteristics of the image features (Richards and Jia 2006, p. 193). As an example, a tiled roof reflects mostly red light, while a green forest can be easily distinguished by green light. By creating a two-dimensional feature space using red and green optical light, these two classes can usually be separated by a single line. Yet, in more complex classification problems with higher dimensional feature spaces, advanced classification algorithms are required to discriminate the labeled classes. Many of such algorithms are able to generate thematic classifications with high accuracies, however, they are not able to present a complete reflection of reality (Wurm 2013, p. 57).

In this study two classification methods – the *Linear Discriminant Analysis* (LDA) and the *Random Forest* (RF) algorithm – are compared with respect to their advantages and disadvantages for classification of partially polarized SAR images in an urban context. This comparison is performed both quantitatively and visually. This section outlines the basic functionalities of the two classification algorithms used.

3.1.1 Linear Discriminant Analysis

Introduced by Fisher (1936) for plant taxonomy, LDA has lately been used for classification of geographical data in various studies including hyperspectral image data, which is known for its high dimensional feature spaces (e.g. Gong et al. 1997; Herold et al. 2003; Puissant et al. 2005; Wurm et al. 2016). By projecting an n-dimensional feature space into a one dimensional coordinate system (Wurm et al. 2016), LDA is more efficient than various other modern classifiers. Subsequently, the detailed approach in forming a linear discriminant function is summarized according to Backhaus et al. (2011) and Wurm et al. (2016).

After defining the labeled classes or groups g (in this case $g = 3$: *Slums*, *urban* and *other*), LDA uses p variables to discriminate n observations (i.e. training points). Therefore, $k = min(g-1, p)$ discriminant functions Y are calculated using the p variables

$$Y_K = b_0 + b_1X_1 + b_2X_2 + \cdots + b_kX_k, \tag{15}$$

where Y is the discriminant variable, b_0 is a constant, b_i are the discriminant coefficients and X_i are the feature variables. In a case with more than two groups the first discriminant function is

estimated by maximizing the discriminatory criterion

$$\Gamma = \frac{inter - group\ scattering}{intra - group\ scattering} = \frac{\sum_{K=1}^{g} N_k(\bar{Y}_K - \bar{Y})^2}{\sum_{K=1}^{g}\sum_{J=1}^{N_K}(\bar{Y}_{JK} - \bar{Y}_K)} \tag{16}$$

with the training points N ($N = 1, \ldots, n_G$) of the group G ($G = 1, \ldots, g$). Its maximum value

$$\gamma = max(\Gamma) \tag{17}$$

is also called *Eigenvalue* and defines the first discriminant function. All further functions are arranged to describe the remaining scattering. After defining all k discriminant functions, an unlabeled feature space can be assigned to the g classes.

As one useful tool, especially with huge feature spaces, the importance of the individual variables p to the class discrimination can be assessed. The *Mean standardized Discriminant Coefficient* (MDC) $\bar{b}_P$ describes the discriminatory influence of the variable P in all discriminant functions by

$$\bar{b}_P = \sum_{K=1}^{k} |b^*_{PK} \cdot PV_K \tag{18}$$

with the standardized discriminant coefficient

$$b^*_{PK} = b_{PK} \cdot s_P \tag{19}$$

and the the percent of variance

$$PV_K = \frac{\gamma_K}{\sum_{i=1}^{k} \gamma_i} \tag{20}$$

of the discriminant function K. s_P describes the standard deviation of the feature variable P. Thus, it is possible to identify the most important descriptive variables. Especially for applications with high dimensional spaces it is often necessary to reduce the number of variables to achieve a reduction of computational cost. At the end of the study an overview of the MDC values is presented and discussed. The Linear Discriminant Functions and classification models are calculated using the R package *MASS* (Venables and Ripley 2002).

3.1.2 Random Forest

Another way of classifying unlabeled observations from a feature space is based on decision trees. In recent years a classification algorithm using large ensembles of such trees has gained interest: The *Random Forest* (RF) classifier. Introduced by Breiman (2001), the Random Forest creates a collection of individual decision trees based on randomly picked samples from all training observations. This classification algorithm is is capable of competing with other state-of-the-art classification methods like Support Vector Machines, Neural Networks, Discriminant analysis, Generalized Linear Models and others within a variety of classification setups (Fernández-Delgado et al. 2014). Also, in the fields of image processing and remote sensing, results of

higher accuracies were obtained using RF than with other classifiers (Rodriguez-Galiano et al. 2012; Kulkarni and Lowe 2016) and it was therefore used in several studies (e.g. Zhu et al. 2012; Beijma et al. 2014; Du et al. 2015; Engstrom et al. 2015; Peerbhay et al. 2015; Belgiu and Drăguţ 2016; Hariharan et al. 2016). Even in environments with a highly imbalanced class distribution (see Japkowicz and Stephen (2002)) Random Forests proved to perform well in comparison to other state-of-the-art classification algorithms, like Support Vector Machines, Bagging or Boosting (Khalilia et al. 2011; Stumpf and Kerle 2011). Moreover, Pal (2005) clarified that the RF algorithm is easier to use than Support Vector Machines, since there are only few user defined parameters in RF. However, the RF algorithm was called a "black box" (e.g. Prasad et al. 2006), as it was not possible to examine how the individual trees are grown. Much like the LDA, RF provides the possibility to assess the importance of the input variables for the classification outcome. With large feature spaces, this is deemed useful for feature selection in order to reduce computational effort (Belgiu and Drăguţ 2016).

The creation of the Random Forest of trees is based on the bootstrap aggregating, also called "bagging", (Breiman 1996) of the input variables. While bagging the input data, different subsets of the data are randomly permuted with replacement. This approach has been demonstrated to improve classification accuracies and to reduce the instability of the basic learning mechanism (Dietterich 2000). Random Forests extend bagging by using a random set of variables at each split during the creation of each individual tree. Thus, accuracy is improved while strength is maintained and furthermore the approach is faster than bagging (Breiman 2001).

The processes of the classification and prediction of Random Forest models are depicted in figure 3.1. In the training phase (A) a user defined number $ntree$ of Classification and Regreession Trees (CART) (Breiman et al. 1984) is built. Here, two examples are shown in which different sets of variables i and observations j are chosen to build individual classification trees. In the classification phase (B), an unlabeled observation (d) is classified with all trees. The class which is ultimately assigned to the observation is determined by the most frequent result of all trees (i.e. majority vote).

During the creation of a Random Forest numerous trees are built. Consequently, one major advantage of RF is that only limited generalization error is possible, which is the error resulting from the total population in contrast to the error of the training data. That is, Random Forests are not prone to overfitting (Prasad et al. 2006). Another advantage of RF is that only two parameters have to be defined by the user: the number of trees $ntree$ built and the number of variables $mtry$ used for each split during the creation of the trees. Even though it was found that the classification accuracy is more sensitive to $mtry$ than to $ntree$ (Ghosh et al. 2014; Kulkarni and Sinha 2012), increasing the former has a big influence on the computational costs (Belgiu and Drăguţ 2016).

In this study, the Random Forests were calculated with the *randomForest* package for *R* (Liaw and Wiener 2002). The two user defined parameters were set to $ntree = 750$ and $mtry$ was set to the square root of the total number of input variables (which is the default value within the package). To reduce computational time, the model creation was adapted for parallel processing.

Samples which are not selected by the bagging of the training data set can be used for a performance evaluation of each tree as an out-of-bag (OOB) estimate. Thus, about one third of the training data set can be used for variable importance (VI) estimation. This importance is

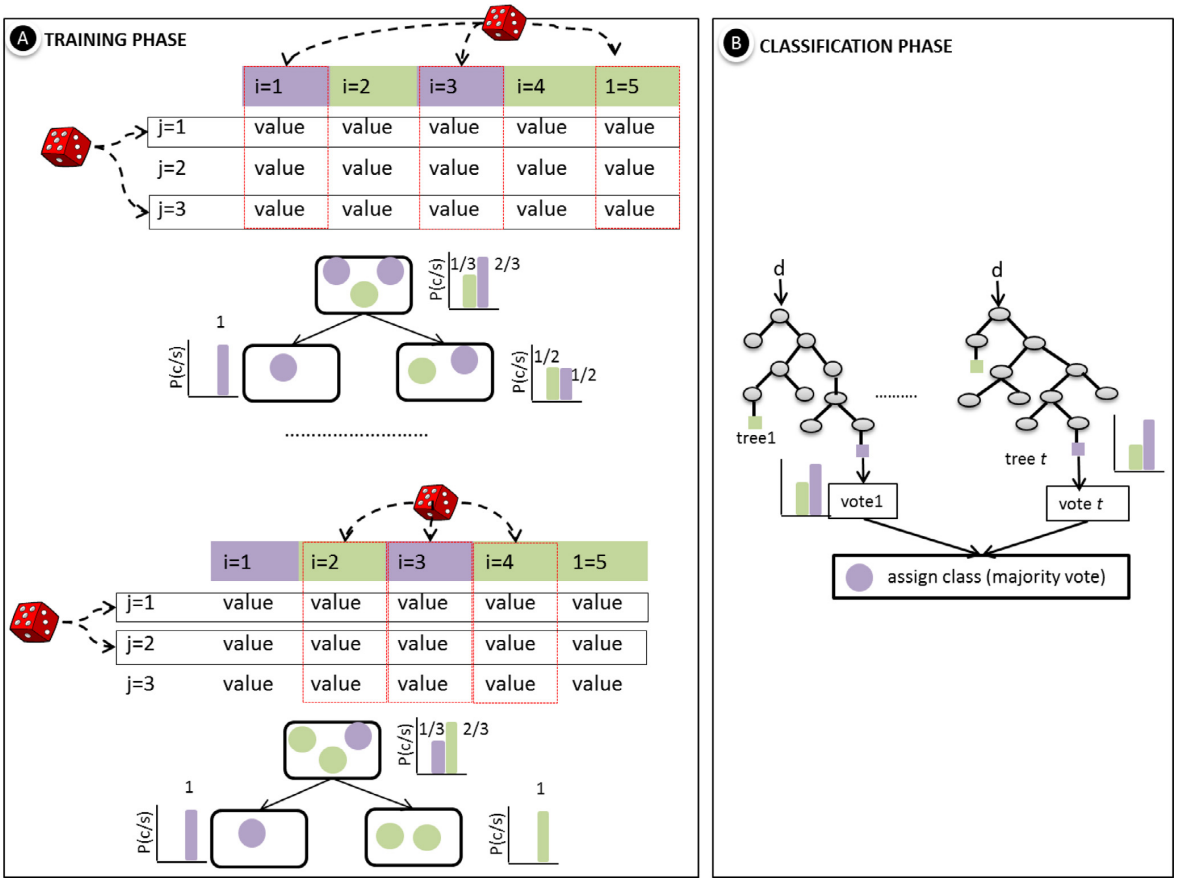


Figure 3.1: Training and classification phase of RF classifier (image source: Belgiu and Drăguţ (2016))

evaluated based on the decrease of the OOB accuracy when excluding certain input variables for the creation of the trees. VI is therefore a measure of relative importance among all predictor variables (Prasad et al. 2006), similar to the MDC value of LDA.

VI calculated from RF based on CART (also used in the *randomForest* package) is often used for feature importance assessment. However, Strobl et al. (2007) demonstrated empirically that VI exhibits a bias in favor of variables that have more breaking points as well as correlated predictor variables. To overcome this bias Strobl et al. (2008) developed conditional inference trees (Hothorn et al. 2006), as a new approach to form random trees. Implemented in the R package *party*, these forests were used for VI-ranking in order to show the influence of feature reduction exemplarily in this case study.

3.2 Training Data Sampling

The collection of training data samples is often an expensive task, since *in situ* measurements take up a lot of time and oftentimes financial effort (Congalton and Green 2008, p. 63). In this study the rare case of having an area-wide reference data set, however, introduced the issue of how to create a sufficient training data set. Basically, a training sample represents points which are connected to both LU/LC classes from the reference data and the image features on a spatial level. This sample is used to create the feature space needed for model training. By having an area-wide reference layer, it is possible to imagine several spatial patterns in which these points could be aligned. Subsequently, the advantages and disadvantages of the individual samples are

approached and their influence on the classification outcome is discussed on a theoretical level.

A few of such spatial alignments of the training samples will be outlined while presenting the considerations taken into account for their creation in detail. In the following, four different sampling distributions are outlined, each building upon its prior approaches. The aim of this chapter is *1.) to demonstrate the importance of training data sampling* and *2.) to present the training data set used in the final methodological work flow of this study.* Thereby the focus rests on the demands on the training data set in order to create an optimal classification model.

Beforehand, however, it is important to summarize the special characteristics of the data used in this study:

- Texture features calculated using a moving-window approach lead to smoothing of the image content.
- The reference shows an extremely imbalanced class distribution.
- In SAR imagery the orientation of objects on the Earth's surface in relation to the sensor has significant influence on the backscattering characteristics.
- The size of the sample has a huge impact both on the representation of the actual feature space and on computational resources needed for the creation of the model.
- Due to the large dimensions of the feature space (holding 120+ image layers), it is assumed to be impossible for the user to conceive its characteristics.

Manual Selection of Training Areas

One approach of collecting training samples is to delimit individual training areas manually for each LU/LC class, which might seem to hold characteristic structures. This is one of the most common ways of collecting samples in remote sensing, since it comes with very low financial and time effort. Within the selected areas all pixels are used for training purposes as it is shown in figure 3.2. During the validation process, these manually selected areas are excluded from the accuracy assessment to avoid unwanted spatial cross-validation.

However, this again induces a certain kind of subjectivity to the training process. In fact, it is impossible to depict a sufficient feature space by manual selection, since the variation of intra-urban classes is too broad. This is the case, especially when considering the composition of spatial neighborhood environments (which is also described by kernel based textural image features) and the orientation of structures on the ground in dependency to the sensor's position. Furthermore, this in remote sensing commonly used scheme of training data collection is irrelevant in this study, since area-wide reference data was available. Thus, cost-intensive *in situ* data acquisition is not driven by minimizing financial or temporal cost.

Random Spatial Sample

Another very common way of establishing a sample is by creating randomly distributed points in space. Thereby, the user's subjective assumptions about representation of the labeled classes within the feature space are negligible, which might not represent significant characteristics of

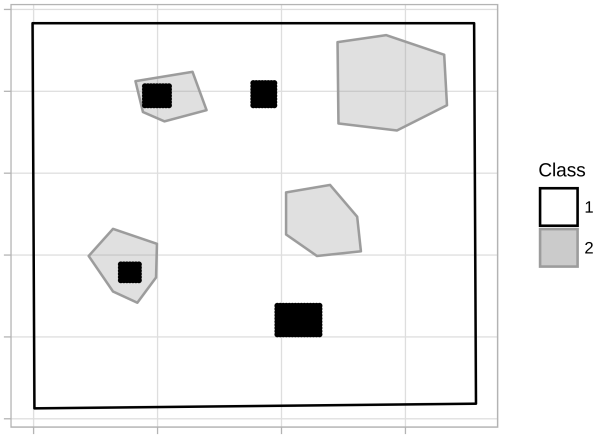


Figure 3.2: Manually selected areas training for training point generation

the numerical feature space. However, as it can be seen in figure 3.3, especially in a setup with extremely imbalanced class distributions, the total number of points within the small class might be relatively small. Furthermore, the points might lie in very close spatial proximity to each other. Thus, it cannot be ruled out that some lie within the same pixel. Moreover, the smoothing effect of a moving kernel approach might lead to only negligible differences between adjacent pixels.

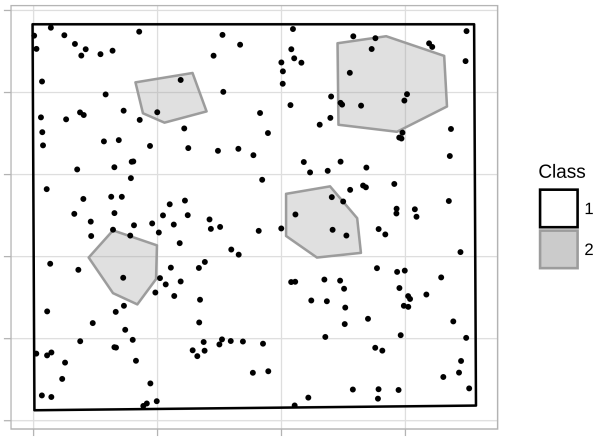


Figure 3.3: Randomly distributed spatial sample points

Spatially Equally Distributed Sampling

To tackle the problems of randomly selected training points, a spatially equally distributed training data set could be established (see figure 3.4). This is the only way to ensure that no pixel is used twice for the model training. For the creation and testing of the classification model such a data set could be divided randomly into two parts. Yet, in this study it was necessary to provide an extensive training data set in order to describe all the imaginable neighborhood relationships, especially for the discrimination between the two intra-urban LU/LC classes *slum* and *urban*. Therefore, a grid width of 37.5 m ($\widehat{=}$ 15 pixels) was chosen as a trade-off between generating

enough training points – especially for the small slum class – and the largest possible distance between the individual points. However, with increasing kernel size (and thus a higher degree of smoothing) the training points and their surroundings were labeled as the class of the training point itself. This lead to spatially biased results in favor of larger kernels, as at one point the surrounding pixels included validation points. Visually the classification results were also very smoothed and incapable of depicting the real borders of the individual land use classes.

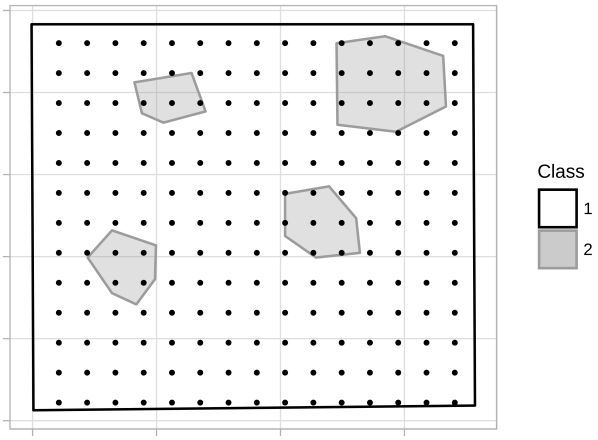


Figure 3.4: Regular grid of spatial points

Spatially Incoherent Equally Distributed Sampling

All sampling methods above show disadvantages because of a spatial interference of training and validation data leading to a spatial bias, small amounts of training points in a case of imbalanced class distribution or subjectivity. Another approach of sampling uses the advantages of an area-wide reference data set, as present in this study, by dividing the image in half. One part of the image is then used for the training of a classification model. Afterwards, this model is applied to classify the other part of the image and vice versa (see figure 3.5).

The points are sampled equally distributed in space to include all possible characteristics of the urban landscape. Furthermore, an unbiased spatially incoherent accuracy assessment can be applied. Therefore, the influence of smoothing introduced by moving kernel textural features can be excluded.

However, one small drawback of this setup is that the training and prediction process has to be separated into two subtasks. In order to apply an accuracy assessment that includes all pixels, the two results have to be combined into one error matrix.

On the basis of the unique conditions in this study, training points were selected by *spatially incoherent equally distributed sampling*. For that the images were split in half horizontally in a northern and southern part. The land use classes *urban* and *other* were sampled with every 100^{th} pixel (i.e. 10×10 pixels). Due to the under-representation of *slums*, this LU/LC class was sampled with every 16^{th} pixel (i.e. 4×4 pixels). Table 3 shows the number of training points used per class for each area of interest as well as the total number.

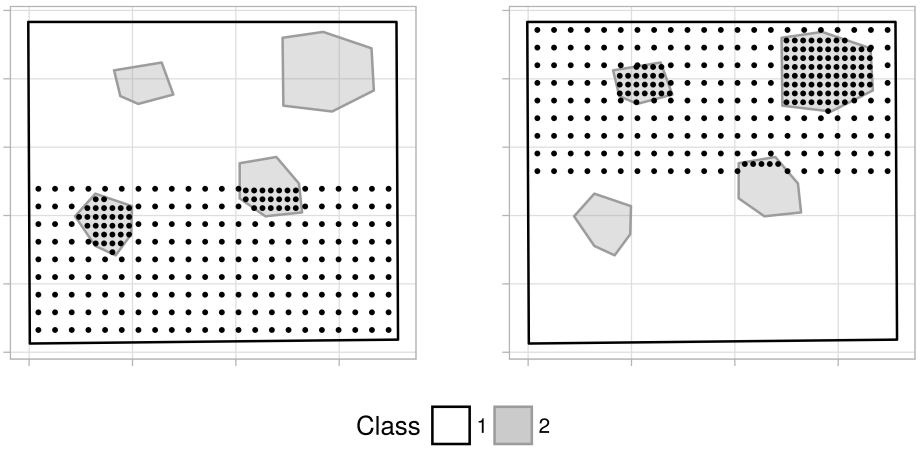


Figure 3.5: Two sets of spatially incoherent equally distributed training points

Table 3: Number of training points collected per class for each area of interest

LU/LC class	AOI 1	AOI 2
Slum	142,430	196,661
Urban	104,662	182,541
Other	250,588	282,780
Total	**497,680**	**663,478**

3.3 Accuracy Assessment of an Imbalanced Class Distribution

In the field of supervised image classification, a machine learning model is applied to classify whole satellite image scenes or areas of interest. This model is trained with only a small representative subset of all the points to label (cf. section 3.1) and can therefore not represent the immense number of real properties but is only able to approximate these. For this reason, it is of major importance to assess the quality of this prediction process. Only thusly is it possible to receive a quantitative impression of the classification quality.

The accuracy of a remote sensing image classification can be assessed using different measures, both for the overall impression as well as for each individual class label. However, such measures compress the classification result into few numbers and can consequently be accompanied by information loss (Mosley 2013). Hence, it is important to select suitable accuracy measures for assessing the quality of the classification in this study.

		Reference	
		1	0
Prediction	1	*TP*	*FP*
	0	*FN*	*TN*

Figure 3.6: Error matrix for a binary classification case

Facing an imbalance of the land use classes in this study, some accuracy measures do not reflect the visual impression of the classification quality. Therefore, in this section the advantages and disadvantages of different overall and class specific accuracy measures are discussed by means of their usability within an imbalanced data set. At first, theoretical considerations are applied to detect the assets and drawbacks of the individual accuracy measures. Afterwards, best suited accuracy measures are selected for the quality assessment during the study at hand.

In remote sensing accuracy assessment is performed using the error matrix (Congalton and Green 2008, p. 57 f.), which is an $n \times n$ matrix describing the correctly and incorrectly classified observations of an n-class problem. An exemplary error matrix for a two class case (*event* (1) or *no event* (0)) is depicted in figure 3.6. and holds the labels of all observations for reference and prediction. Thus, it is possible to quantify the observations which are labeled *a.)* correctly as event, meaning true positives TP, *b.)* falsely as event, that are false positives FP, *c.)* correctly as no event, namely true negatives TN, and *d.)* falsely as no event, that are false negatives FN.

Scaling up from the binary case, the matrix can also be built for n classes which would result in an $n \times n$ matrix. However, for this theoretical approach to class specific accuracy measures the binary problem is used for reasons of clarity and comprehensibility.

3.3.1 Overall Accuracy Measures

The most common measure of determining the accuracy of a classified image is the *Overall Accuracy* OA. It determines the percentage of correctly classified pixels (Congalton and Green 2008, p. 16 f.).

$$OA = \frac{TP + TN}{total\ population} = \frac{TP + TN}{TP + FP + FN + TN} \tag{21}$$

Another measure widely used for quality assessment in remote sensing is Kappa (Cohen 1960). It is used to estimate the difference between an achieved classification result and a statistical fluke. $\hat{K}$ is an estimate of the Kappa coefficient (Congalton and Green 2008) and is calculated with

$$\hat{K} = \frac{n \sum_{i=1}^{k} n_{ii} - \sum_{i=1}^{k} n_{i+} n_{+i}}{n^2 - \sum_{i=1}^{k} n_{i+} n_{+i}} \tag{22}$$

assuming k as the number of rows of the error matrix, n_{ii} as the value in the i-th row and i-th column, n_{i+} as the sum of the i-th row and n_{+i} the column sum, respectively. The values ranging from -1 to 1 were classified according to their strength of agreement of the classification result by Landis and Koch (1977) as it can be seen in table 4.

Table 4: Classification of Kappa values

Kappa value	Strength of agreement
< 0.00	poor
0.00 - 0.20	slight
0.21 - 0.40	fair
0.41 - 0.60	moderate
0.61 - 0.80	substantial
0.81 - 1.00	almost perfect

Even if OA and $\hat{K}$ provide a good estimate of the classification as a whole, small classes in case of an imbalanced class distribution are under-represented in this measure (Mosley 2013). To address this shortcoming, class specific accuracy measures can be calculated. The *Class Balanced Accuracy* CBA weights the classification based on the absolute representation of the classes in the reference (Mosley 2013).

$$CBA = \frac{\sum_{i}^{k} \frac{n_{ii}}{max(n_{i+}, n_{+i})}}{k} \tag{23}$$

with k being the number of classes (and therefore the number of rows and columns of the error matrix) and n_{ii}, n_{i+} and n_{+i} defined as previously. CBA normalizes the overall accuracy by assuming an equal distribution of all classes.

3.3.2 Class Based Accuracy Measures

In addition to overall accuracy measures it is possible to assess the classification quality with regard to the individual labeled classes. The aim is to get an impression about the individual over and under classification as well as the reliability of the result for each class. In a binary case these classes correspond to positives P and negatives N.

The *Sensitivity* (Altman and Bland 1994a), sometimes also referred to as Producer's Accuracy, determines the share of correctly captured pixels of one class in relation to their complete reference.

$$Sensitivity = \frac{TP}{P} = \frac{TP}{TP + FN} \tag{24}$$

Sensitivity can therefore be seen as a measure which indicates how much of the area of a certain class is classified correctly.

On the contrary, *Specificity* measures the number of pixels which are labeled with any other

class (i.e. negatives N) in relation to the total number of negatives (Altman and Bland 1994a).

$$Specificity = \frac{TN}{N} = \frac{TN}{TN + FP} \tag{25}$$

Thus, Specificity is a good indicator of the amount of over-classification by a certain class.

The *Positive Predictive Value* (PPV), also known as User's Accuracy, (Altman and Bland 1994b) can be utilized to quantify the relation between correctly classified pixels TP and the total number of thusly predicted pixels.

$$PPV = \frac{TP}{TP + FP} \tag{26}$$

Therefore, the PPV is an indicator of how many pixels labeled with a certain class actually match this exact class in the reference.

3.3.3 Interpretation of Accuracy Measures

To approach the description of overall and class specific accuracy, respectively, six different theoretical scenarios of a two class problem are depicted in figure 3.7. These examples are used to characterize both overall and class specific accuracy measures in an imbalanced scenario which holds 10 observations (e.g. pixels) of positives and 100 observations of negatives, respectively. They highlight scenarios which might occur during a supervised classification process, with the individual error matrices quantifying the classification results.

In detail, example 1 (see fig. 3.7a) shows only correctly classified positives and negatives. On the other hand, in example 2 (see fig. 3.7b) positives are only classified outside of their corresponding reference area resulting in false negatives (FN) and false positives (FP). Example 3 (see fig. 3.7c) represents a classification scenario, which is neither covering the positive reference area completely nor creating any over-classification, whereas example 4 (see fig. 3.7d) holds very little TP classifications but yet six times more over-classified pixels. Example 5 and 6 (see fig. 3.7e and 3.7f) show the same ratio between correctly classified and over-classified observations, yet varying their absolute number.

To identify the advantages and disadvantages of these overall and class specific accuracy measures, they are calculated for the examples in figure 3.7 and summarized in table 5. The class specific accuracy measures refer to the *positives* in the examples which, based on their total count, represent the smaller group.

Table 5: Accuracy measures for theoretical classification examples

Examples	1	2	3	4	5	6
OA	1.00	0.81	0.95	0.86	0.90	0.90
$\hat{K}$	1.00	-0.1	0.65	0.05	0.39	0.24
CBA	1.00	0.45	0.73	0.51	0.67	0.56
Sensitivity	1.00	0.00	0.50	0.10	0.40	0.20
Specificity	1.00	0.90	1.00	0.94	0.96	0.98
PPV	1.00	0.00	1.00	0.14	0.50	0.50

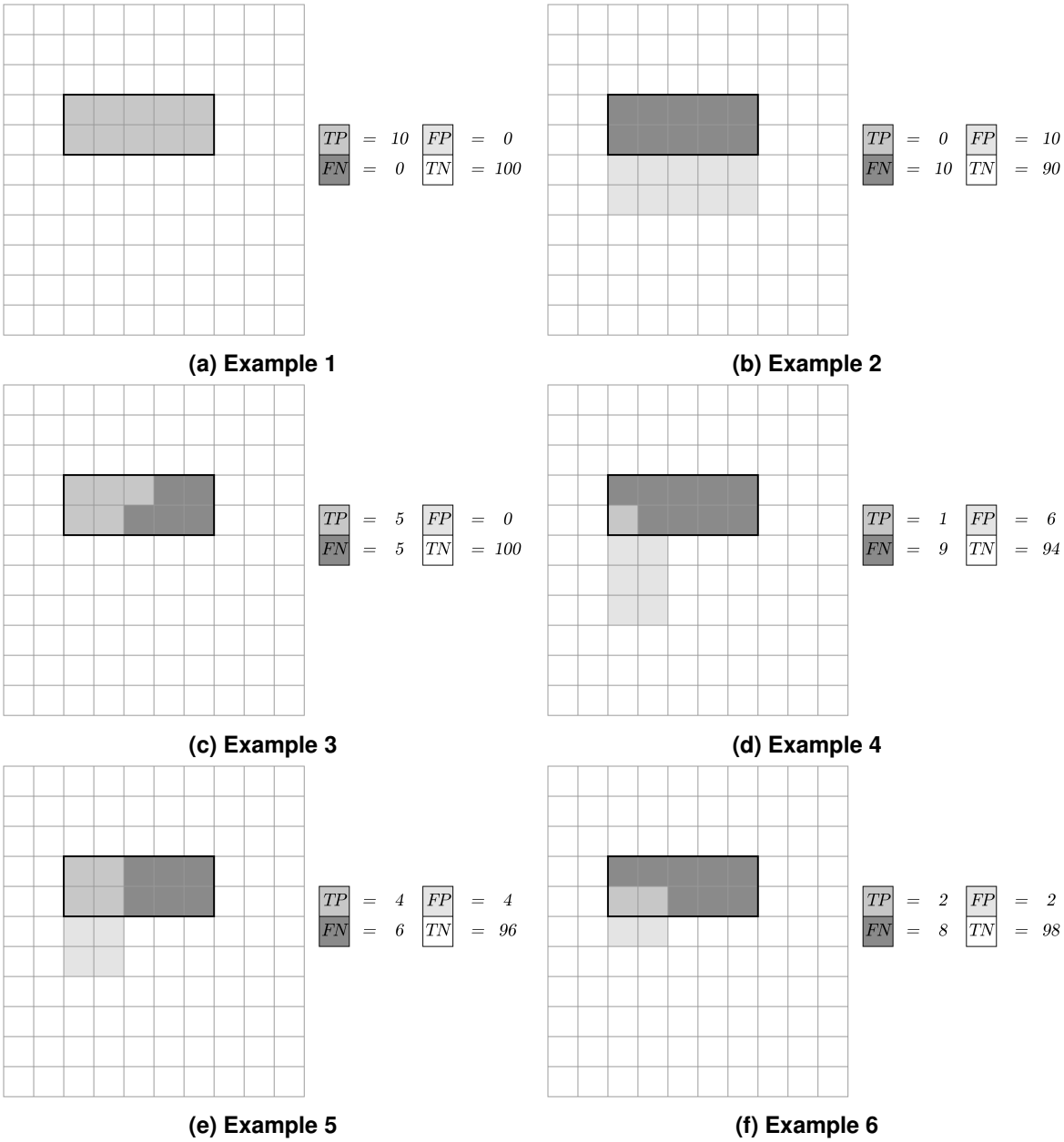


Figure 3.7: Theoretical examples of a classification result in a two class (positive / negative) example with error matrices

Example 1 shows a perfect classification with no misclassification. This is also indicated by all accuracy measures taking the value 1.0. In contrary, examples 2 through 6 contain misclassification which are also represented by the overall and class specific accuracy measures. In example 2, no positives are classified correctly, thus creating a Sensitivity of 0. While the overall accuracy decreases only slightly, $\hat{K}$ indicates a very low strength agreement. The over-classification of 10 cells leads to a decrease of 0.1 from the optimal Specificity. Since not a single pixel is classified correctly, the PPV is also 0. In example 3 no over-classification is present but only 50% of the positives are classified correctly. Thus, the Sensitivity decreases to 0.5 while Specificity and PPV both indicate no over-classification.

In example 4 only a small share of the positives is classified correctly combined with a strong over-classification. Therefore, both sensitivity and PPV are low. Also, the Kappa estimate and CBA show a high degree of randomness in the classification result. Examples 5 and 6 show the same relation between correctly and incorrectly classified positives with differing absolute numbers. While the Sensitivity and the Specificity are able to outline this difference, the PPV remains constant and therefore allows only implicit conclusions about the total over-classification.

In conclusion it can be stated that while the OA only changes slightly in this case of an imbalanced class distribution, the balanced measure CBA is more sensitive in the small class of positives. It is important to recognize that even slight changes of the Specificity of small classes could represent strong over-classification in relation to the TP values. Especially in cases with imbalanced class distribution, it is therefore necessary to interpret class specific accuracies based on the three parameters Sensitivity, Specificity and PPV. Only thus is it possible to provide a detailed statement on the misclassification.

3.4 Experimental Setup

The experimental setup of this study is designed to evaluate the discriminatory influence of textural and morphological image features for slum detection in SAR images. Furthermore, the effect of differently sized kernels and the behavior of two classification methods is compared in this domain. Therefore, several classification experiments are conducted using different image feature sets. The structure of the experiments is outlined in detail in this section.

For the experiments, two in remote sensing applications frequently used classifiers (see section 3.1) are selected to compare the quality of a classification of highly dimensional feature spaces: Linear Discriminant Analysis and Random Forest. Using a training data sample of spatially incoherent equally distributed training samples (see section 3.2), these two classifiers are trained using the different sets of image features. After the prediction process, both overall and class specific accuracy measures are applied (see section 3.3) to describe the discriminatory quality of each individual result.

The setup is established based on prior assumptions concerning the spatial domain of the objects to be detected, in this case city blocks. GLCM textures describe the homogeneity or heterogeneity of a surface based on the image gray levels. Following the Generic Slum Ontology for describing the slums, this can be seen as a proxy to the settlement density. In contrast, mathematical morphologies merge structures with similar gray values. Therefore, they are capable of detecting rough cuts and contrasts in the spatial neighborhood.

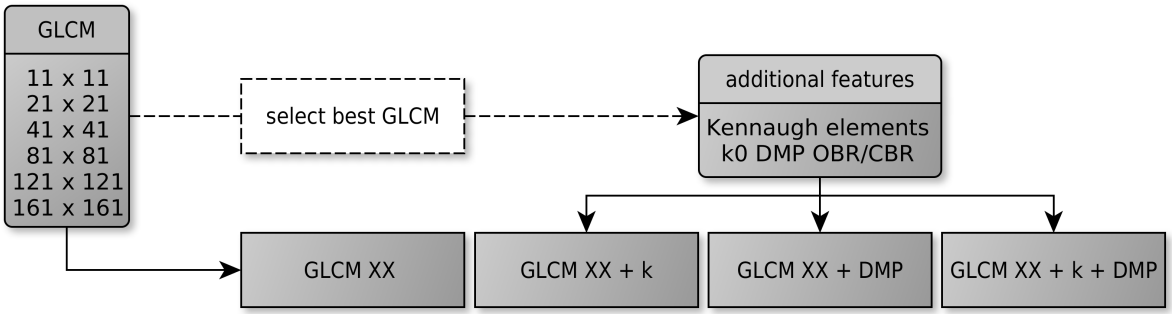


Figure 3.8: Schematic illustration of the experimental setup which is applied for both classifiers (LDA and RF)

Taking these considerations into account, the image features are separated into different sets in order to test their capabilities for the discrimination of intra-urban classes. Therefore, these sets were assembled in an extensive test setup (see figure 3.8). First, the textural image features sets are analyzed separately focusing on the question which kernel size holds the highest accuracies. This is determined for both classification methods (LDA and RF). Furthermore, the individual properties of the classification results are discussed in detail. Afterwards, this best discriminating GLCM kernel sizes of each classifier are combined with the following additional features:

a) Kennaugh elements (k_0, k_3, k_4 and k_7 for HH/VV polarized data, k_0, k_1, k_5 and k_8 for VV/VH polarized data)

b) Differential Morphological Profile (DMP) created from the Opening and Closing By Reconstruction of the backscattering intensity k_0 for both polarization modes

To identify the various feature sets throughout this study, a systematic nomenclature is established. It follows the following structure:

$$FS\,^{GLCM\ kernel\ size}_{additional\ features}$$

Thus, the feature set including the GLCM texture derived with a 81 $\times$ 81 kernel, the Kennaugh elements and the differential morphological profile of k_0 is termed $FS\,^{81}_{k+DMP}$.

Since this study mainly focuses on the mapping of slum areas, additionally to the pixel based accuracy assessment an object based accuracy assessment method is applied. By dividing the slums into classes according to their size, they are analyzed separately. Hereby, it is possible to determine the influence of object properties on the ability of the classifier to map slum areas.

The two classifiers used in this study enable an assessment of the image feature importances on the classification result. Hence, these variable importance measures are used to perform an exemplary feature reduction. Thus, it is possible to receive an impression of the dependency of the classification result on the number of image features used for mapping.

4 Results

Following the experimental setup, at first all textural image features with varying kernel sizes were classified. The quantitative results of the accuracy assessment are presented in this section. Based on both overall (section 4.1.1) and class specific accuracy measures (section 4.1.2) the best discriminating textural feature set is selected from FS^{11}, FS^{21}, FS^{41}, FS^{81}, FS^{121} and FS^{161}. Building upon this feature set, extended feature sets are created using the polarized data (HH/VV k_0, k_3, k_4, k_7 and VV/VH k_0, k_1, k_5, k_8) as well as the DMP of the backscattering intensity (k_0). Subsequently they are analyzed in terms of classification quality in section 4.2. Afterwards, the classified slum areas are investigated on the level of individual objects in section 4.3. In section 4.4 the results of the experimental feature reduction approach are presented.

4.1 SAR Texture

In this section the results of the supervised classification of the textural image features are presented. Hence, first the overall classification results are described followed by the class specific accuracy of the slum areas.

4.1.1 Overall Classification Results

For the assessment of the overall classification quality, the measures Overall Accuracy (OA), Kappa estimate ($\hat{K}$) and Class Balanced Accuracy (CBA) are calculated. They are presented in table 6 where they are differentiated by classifier and feature set. Furthermore, these measures are sketched for visual interpretation in figure 4.1.

Table 6: Results of overall accuracy measures for GLCM texture feature sets

	LDA						RF					
	FS^{11}	FS^{21}	FS^{41}	FS^{81}	FS^{121}	FS^{161}	FS^{11}	FS^{21}	FS^{41}	FS^{81}	FS^{121}	FS^{161}
OA	76.97	83.23	87.43	**87.90**	86.28	85.28	79.11	84.95	88.33	**88.58**	87.52	86.53
$\hat{K}$	58.78	69.19	76.59	**77.45**	74.70	72.99	61.71	71.80	77.83	**78.09**	75.95	73.98
CBA	58.74	67.50	74.20	**75.60**	72.99	70.98	60.13	68.97	76.06	**77.56**	72.16	67.24

Most strikingly, it can be seen that all three measures initially indicate a gain in accuracy from FS^{11} through FS^{81}. The feature set FS^{81} holds the highest values for OA, $\hat{K}$ and CBA with both classifiers. By a further increase of the spatial neighborhood, however, all accuracy measures start decreasing. The differences between the medium and large sized kernels (41 – 161) vary only little, whereas the small kernels show higher differences.

The comparison of the two classifiers reveals that with RF higher results can be attained (with only one exception: CBA of FS^{161}). By means of Overall Accuracy the highest results for LDA and RF are 87.90 and 88.58, respectively. The Kappa estimate of 0.7745 and 0.7809 indicate a substantial strength of agreement for both classifiers with FS^{81}. Additionally, CBA is highest for this feature set with 75.60 and 77.56, respectively. Yet, lower values of CBA in comparison to OA in all experiments indicate significant misclassification of the small group (i.e. slums). To identify the deficits of the classification, it is of the utmost importance to inspect this small group.

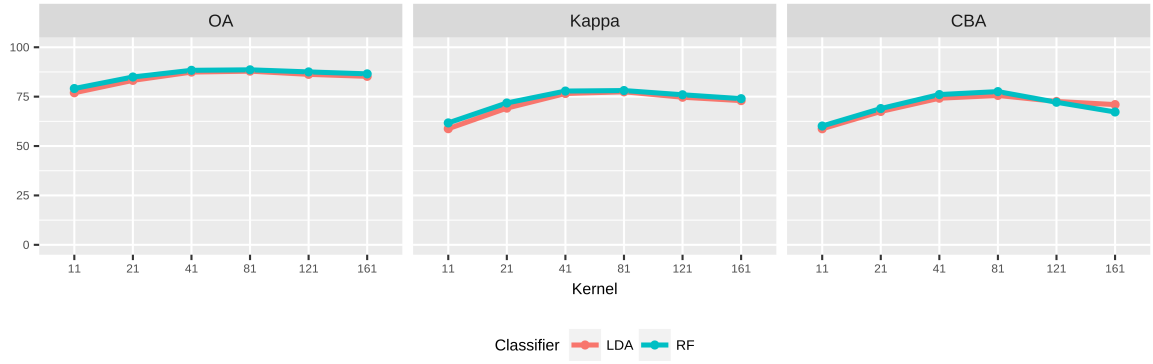


Figure 4.1: Overall Accuracy measures

4.1.2 Classification Results of Slums

With the land use class of interest in this study (i.e. slums) being under-represented in terms of spatial extent, the Overall Accuracy is more influenced by the prevalence of non-slum pixels. To focus the evaluation on this very small group, the quality of the classified slum areas is assessed by using four different measures. These are Sensitivity (SEN), Specificity (SPE), Positive Predictive Value (PPV) and Error Rate (ER). They provide an impression of over- and under-classification of this class within the whole classification result. The results of the class specific accuracy analysis for the pixels labeled as slums are shown in table 7 and illustrated in figures 4.2 and 4.3.

Table 7: Results of class specific accuracy measures for slums using GLCM texture feature sets

	LDA						RF					
	FS^{11}	FS^{21}	FS^{41}	FS^{81}	FS^{121}	FS^{161}	FS^{11}	FS^{21}	FS^{41}	FS^{81}	FS^{121}	FS^{161}
SEN	0.683	0.750	**0.790**	0.785	0.763	0.732	0.681	**0.698**	0.679	0.569	0.441	0.324
SPE	0.878	0.922	0.946	**0.952**	0.940	0.935	0.888	0.931	0.958	0.974	0.980	**0.984**
PPV	0.272	0.392	0.495	**0.524**	0.465	0.434	0.290	0.405	0.523	0.596	**0.597**	0.580
ER	1.045	0.858	0.715	**0.690**	0.773	0.834	1.030	0.897	**0.799**	0.835	0.962	1.096

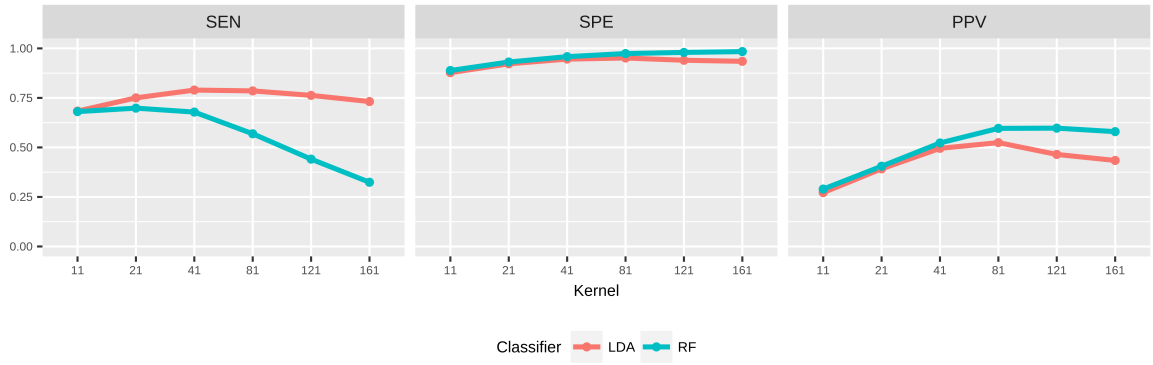


Figure 4.2: Class specific accuracy measures of slums

The Sensitivity quantifies the share of referenced slum pixels (i.e. positives) that are recognized correctly by the classifier. The results show that both classifiers gain Sensitivity with increasing GLCM kernel size until they reach a maximum and then start to decrease again. LDA reaches the highest Sensitivity of 0.79 with FS^{41}, whereas RF reaches 0.698 with FS^{21}. For all GLCM feature sets LDA is able to achieve higher sensitivity values than RF. Furthermore, the slope of the decrease with increasing kernel size is steeper for RF than for LDA.

Quantifying the share of over-classified pixels in comparison to all pixels outside the slum reference (i.e. negatives), the Specificity is not as sensitive to small numbers of misclassified observations as Sensitivity in a case with an imbalanced class distribution. Therefore, the Specificity values show very high values in all experiments, but still indicate a gradual change with varying kernel size. That is, in contrast to Sensitivity the Random Forest classifier outperforms LDA by means of Specificity for every GLCM feature set. Furthermore, RF gains Specificity by increasing kernel size up until 161 $\times$ 161 (SPE = 0.984), whereas LDA reaches a peak at FS^{81} (SPE = 0.952) and decreases with larger kernel sizes. Therefore, RF is able to minimize the number of over-classified slum pixels even with very large GLCM kernels. LDA on the other hand shows higher over-classification of slum areas with strongly smoothed textures.

The Positive Predictive Value describes the ratio of pixels correctly classified as slum in relation to over-classified pixels. Similar to Sensitivity and Specificity, the classification accuracy benefits from increasing kernel size by means of PPV. LDA reaches a peak at FS^{81} (PPV = 0.524) and RF at FS^{121} (PPV = 0.597), respectively. Similar to the Specificity, RF outperforms LDA in terms of PPV in all GLCM feature sets.

The Error Rate (ER) as a combined measure for over- and under-classification also shows a gradual change over the varying GLCM kernels. The Error Rates for the experiments with the textural image features are shown in figure 4.3. The lowest error is reached for LDA with FS^{81} of 0.690 and for RF with FS^{41} of 0.799. From these best results the classifications show increasing Error Rates both with smaller and larger kernels.

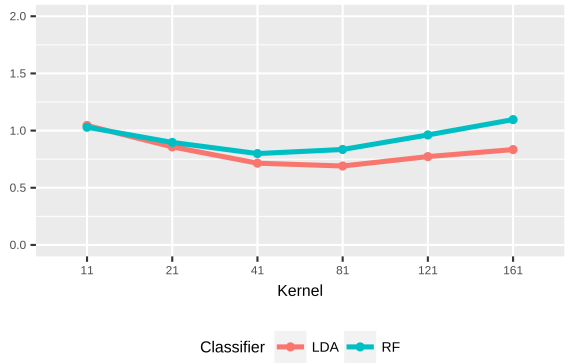


Figure 4.3: Error Rate of pixels labeled as slums

4.2 Extended Feature Sets

Based on the overall accuracy measures of the classification results the GLCM features with a kernel size of 81 $\times$ 81 were extended by additional image features, since they showed the best

results. These experiments were also performed using the two classification methods LDA and RF. Subsequently, the labeled images were quantitatively analyzed according to their classification accuracy. Table 8 summarizes the overall accuracies of FS^{81} (for comparison), FS^{81}_{k}, FS^{81}_{DMP} and FS^{81}_{k+DMP}.

Table 8: Results of overall accuracy measures for extended feature sets

	LDA				RF			
	FS^{81}	FS^{81}_{k}	FS^{81}_{DMP}	FS^{81}_{k+DMP}	FS^{81}	FS^{81}_{k}	FS^{81}_{DMP}	FS^{81}_{k+DMP}
OA	87.90	87.96	87.96	87.99	88.58	88.58	88.60	88.60
$\hat{K}$	0.774	0.776	0.776	0.776	0.781	0.781	0.781	0.781
CBA	75.60	75.72	75.73	75.79	77.56	77.74	77.74	77.79

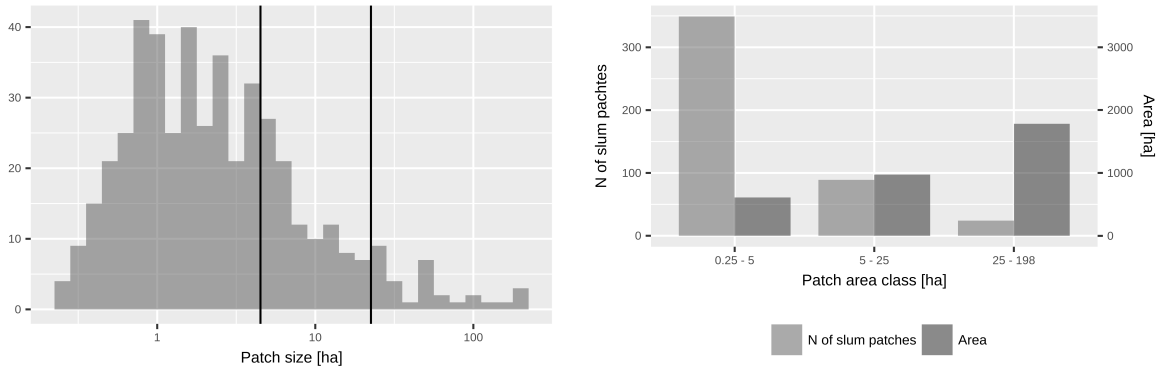
The comparison of FS^{81} and all extended feature sets reveals that the extension brings only very small improvements by means of classification accuracy. This applies both to the Kennaugh elements and the differential morphological profile. However, there is no decrease in accuracy by extending the feature set. For both classifiers, the best result is achieved with FS^{81}_{k+DMP} and an Overall Accuracy of 87.99 using LDA and 88.60 using RF, respectively. Yet, the most significant variations appear in the Class Balanced Accuracy (CBA). That indicates that the changes are mostly affecting the under-represented classes. Similar tendencies are also represented in the class specific accuracy measures (see table 9). Extending the feature sets leads to slight increases of the Sensitivity of the classified slums. Furthermore, the PPV increases slightly for the experiments with LDA and decreases for RF experiments.

Table 9: Results of class specific accuracy measures for slums using extended feature sets

	LDA				RF			
	FS^{81}	FS^{81}_{k}	FS^{81}_{DMP}	FS^{81}_{k+DMP}	FS^{81}	FS^{81}_{k}	FS^{81}_{DMP}	FS^{81}_{k+DMP}
SEN	0.785	0.787	0.788	0.789	0.569	0.574	0.573	0.575
SPE	0.952	0.952	0.952	0.952	0.974	0.974	0.974	0.973
PPV	0.524	0.527	0.527	0.528	0.596	0.594	0.595	0.594
ER	0.690	0.686	0.685	0.683	0.835	0.832	0.832	0.831

4.3 Dependence of Classification Accuracy on Slum Patch Size

The pixel based accuracy assessment assumes all slum areas as one spatial entity. However, when considering individual patches of slums, it is possible to create a broader impression about the capabilities of partially polarized SAR data for slum mapping. Since the image processing methods are based on focal operations of different kinds, the size of the individual patches is crucial for its detectability. To investigate the influence of varying kernel sizes on the different patches, an object based accuracy assessment is applied following the approach of Klotz et al. (2016). Therefore, the reference slums in this study are divided into three classes by their area as indicated in figure 4.4a. The referenced slum patches are separated into classes holding slums with sizes of 0.25 – 5ha, 5 – 25ha and 25 – 198ha. Figure 4.4b depicts the number of slum patches as well as the cumulative area of the slums within the individual classes.



(a) Histogram of all slums in both AOIs separated into three size classes

(b) Number of slums and cumulative area per size class

Figure 4.4: Arrangement of three classes of slums by patch size

By dividing all patches into three groups it is attempted to maintain a balance between the total size of each group and keeping enough patches per group to provide a statistically representative sample. However, the small group holds only 18% of the total area (N = 349), the second group holds about 29% (N = 89), whereas the third group holds 53% of the total area but contains only 24 reference patches.

To analyze the classification results on patch level, adjacent pixels labeled as slums are grouped to individual patches. Following the concept of Klotz et al. (2016), these patches are subsequently compared to the reference slum patches on spatial level by means of their mutual overlap O_m with the reference. Once crossing the overlap threshold $T_O = 50\%$ of the reference patch's size, a classified object can be called true positive. This can be transferred to calculate the Sensitivity. Since it is not possible to assign the negatives to any specific class, the PPV was assumed static for all three classes of each individual feature set. Thus, an Error Rate was calculated for the three size classes for all experiments with textural image features. The results are shown in figure 4.5.

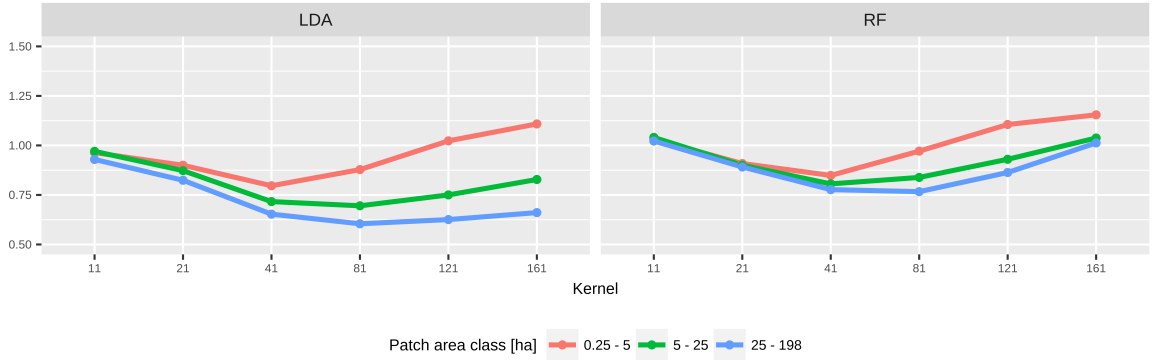


Figure 4.5: Error Rate of patch based accuracy assessment by patch size classes and GLCM kernels

The Error Rates represent a similar curved trend of the ER by varying kernel size which is also pictured by the pixel based accuracy measures. Yet, these results show that in all experiments the largest patches were classified with lower ER than the smaller ones. The lowest ER was

achieved using LDA with a kernel size of FS^{41} for the smallest patches (0.25 – 5 ha) and with a kernel of FS^{81} for the two larger groups. RF, on the other hand, shows the lowest ER with FS^{41} for the small and medium-sized patches, whereas the largest patches are classified with the least error using FS^{81}.

4.4 Experimental Feature Reduction

Both classifiers used in this study provide the possibility of estimating the importance of the 126 individual features on the classification result. In an experimental approach for feature reduction the influences of reduced feature sets are examined. These estimates are derived for all textural features with a kernel size of 81 $\times$ 81 as this feature set shows the highest accuracy overall for both classifiers. The variable importance for the experiments was calculated using the *Mean standardized Discriminant Coefficient* in case of LDA and the *Mean Decrease in Accuracy* in case of RF. The rankings of the most important features and their relative influence on the classification result are depicted in figure 4.6. Since the Random Forest algorithm uses random samples of input variables and observations for its creation which might influence the variable importance measure, a series of 100 model runs was performed. Figure 4.6 (r.) shows the average variable importances of all models combined for the individual features with the antennas indicating the standard deviations.

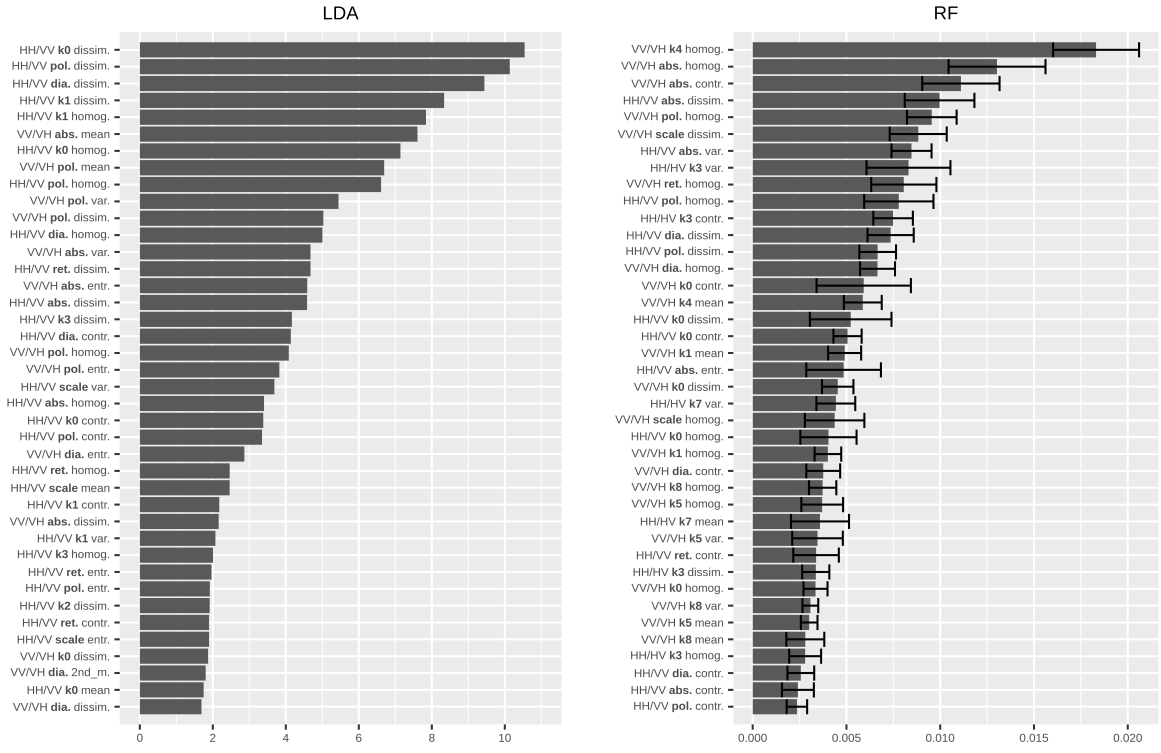


Figure 4.6: Relative ranking of the 40 most important variables for classification with LDA and RF

When examining the features appearing in these top 40 of the most important variables, it can be stated that the two classifiers rank the importance of the image layers differently. On the one hand, the polarization mode of the acquired images and the polarized content (i.e. the Kennaugh

element) influences the variable importance differently. That is, 26 out of the 40 most important features for LDA are of the HH/VV polarization, RF ranks keep a higher balance between the two polarization modes (21$\times$ VV/VH vs. 19$\times$ HH/VV). On the other hand, the seven GLCM texture measures are also unequally distributed. The most important features for LDA often are the *Dissimilarity* texture, whereas RF favors the *Homogeneity* texture.

In order to assess the reduced feature set's influence on the classification result a series of tests is applied. In these tests the n (where $n \in \{10, 20, \ldots, 120, 126\}$) most important textural features are used to train and classify the image content. The results are compared by means of overall accuracy measures (see figure 4.7) as well as class specific accuracy for the slums (see figure 4.8). In this experiment the accuracies are assessed using a pixel based approach.

Looking at the Overall Accuracy (fig. 4.7, l.) of the experiments, it can be seen that for both classifiers the accuracy of the results are reduced when diminishing the feature set to the 10 most important features. By increasing the number of features to 20 for RF and 40 for LDA, respectively, OA reaches a plateau. From there on the increment in the number of image features induces only slight changes to the OA. The trend of the Kappa estimate $\hat{K}$ follows alike. In contrast, the Class Balanced Accuracy changes more when increasing the number of features. At all times RF performs slightly better than LDA in terms of OA, $\hat{K}$ and CBA.

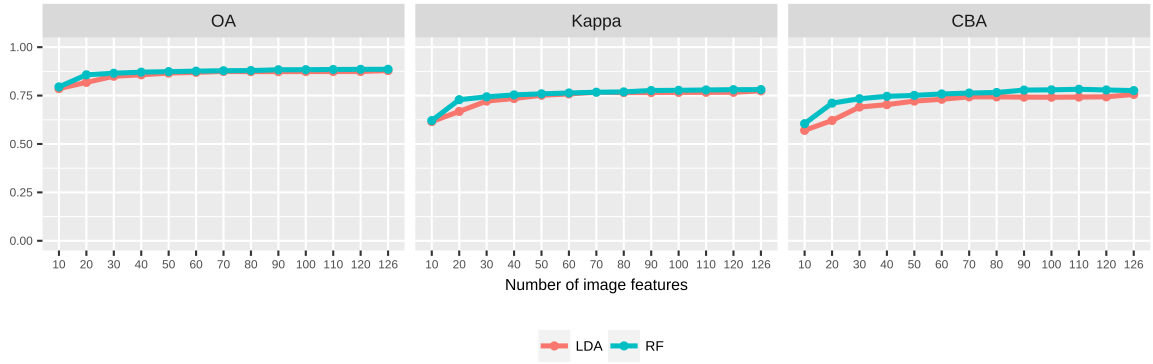


Figure 4.7: Overall accuracy measures by the n most important features

The graph of Sensitivity (see figure 4.8, l.) shows a continuous increase with a rising number of features. LDA shows a higher total increase in SEN (0.657 – 0.785) than RF (0.507 – 0.569). The Specificity shows similar trends similar to the overall accuracy statistics. After an initial steep increase, SPE does only change very little with 20 (40 for LDA) and more features. The PPV shows the highest increase of all class specific accuracy measures with more features, that is 0.300 for LDA and 0.337 for RF, respectively.

Figure 4.8: Class specific accuracy measures by the n most important features

5 Discussion

In this section the results of this study are discussed. First, the suitability of texture derived from partially polarized SAR imagery for automated slum mapping is summarized. Subsequently, an interpretation of the benefit of extending textural features with differential morphological profiles is made. Afterwards, the importance of the spatial domain of slums within the urban context is outlined before examining the results of the experimental feature reduction. Lastly, the two classification algorithms used in this study are compared in consideration of advantages and disadvantages for slum detection with high-dimensional SAR image feature spaces.

5.1 Slum Mapping Using SAR Texture

The analyses of the classification accuracy of textural image features derived from several kernels reveal that the data is in principle suited for discriminating an urban landscape with accuracies of over 87 percent. Yet, the success of this approach for slum mapping depends on the size of the spatial neighborhood taken into account. All overall accuracy measures (OA, $\hat{K}$ and CBA) show that the classification quality first increases with larger kernels and reaches a peak at a kernel size of 81 $\times$ 81. By further expanding the GLCM kernel size the classification accuracy decreases. Thus, this 81 $\times$ 81 kernel can be identified as the most suitable GLCM kernel in terms of overall classification quality in this setup. Yet, the balanced perspective achieved by the CBA highlights that the smaller class, i.e. the slums, are classified with a lower precision than the larger classes.

This is confirmed by the class specific accuracy assessment. It shows that one best discriminating kernel cannot be easily defined for the slum areas by only using the Overall Accuracy and Kappa estimate. This difficulty is also confirmed by the visual impression of the classification result. Figures 5.1 and 5.2 show two exemplary urban neighborhoods comprising differently sized slum patches.

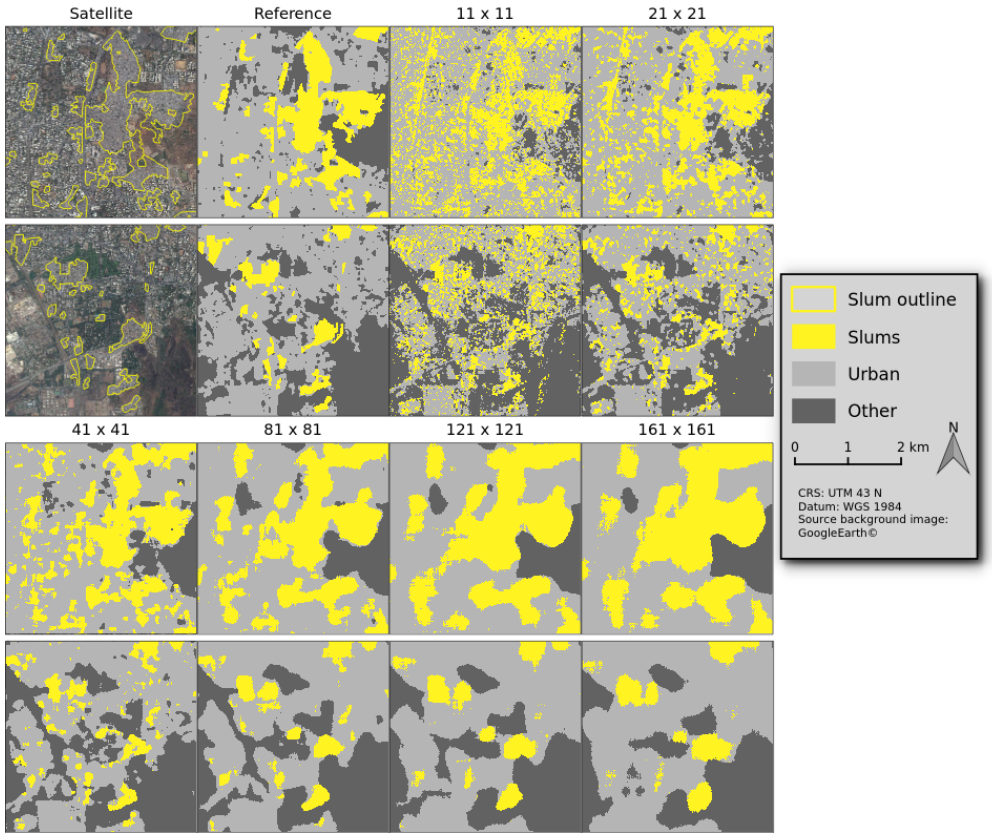


Figure 5.1: Comparative alignment of all GLCM kernels classified by Linear Discriminant Analysis

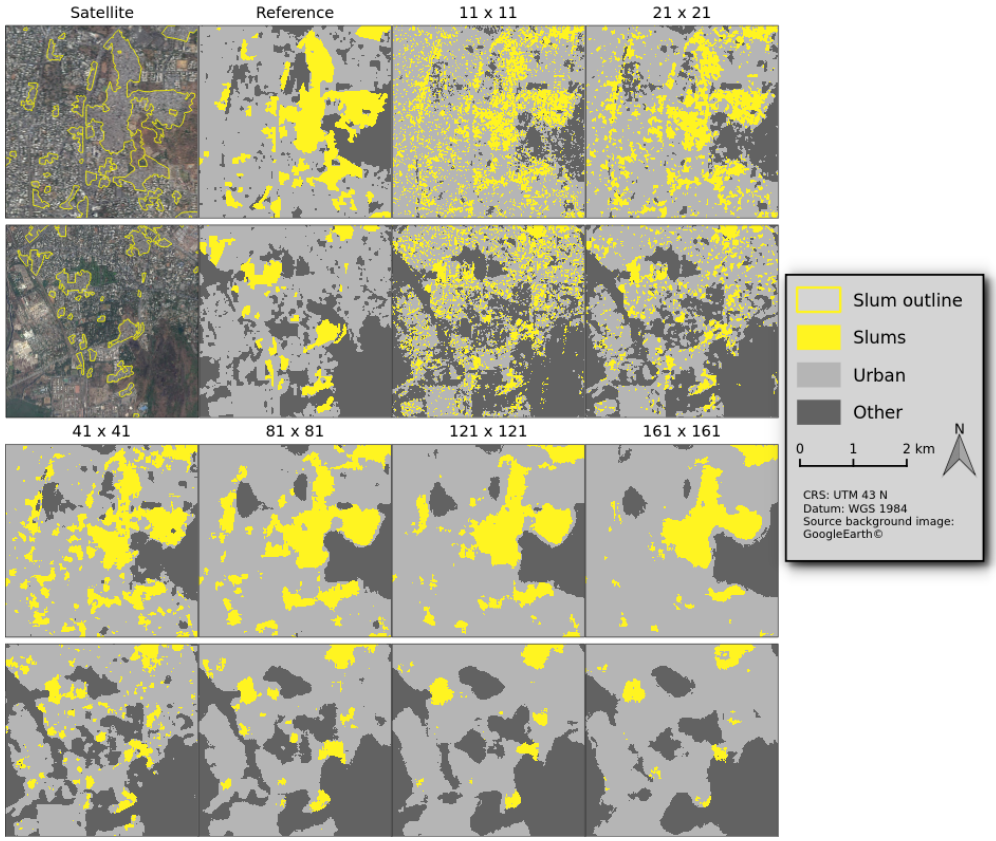


Figure 5.2: Comparative alignment of all GLCM kernels classified by Random Forest

These examples depict the gradual change of smoothing introduced by the varying kernel size. While the classifications results with small kernels (11 $\times$ 11 and 21 $\times$ 21) show a very noisy scattering of slum pixels, the larger kernels incrementally lead to smoother and more homogeneous patch arrangements. Obviously, the small kernels are not able to delineate the feature space in a distinct way. Thus, the strong over-classification induced is also represented in the very low PPV values for these feature sets. That is, only one quarter of all pixels classified as slums for FS^{11} actually lie within the slum reference. In contrast, the highest PPV is reached with medium-sized kernels where more than 50 percent of mapped slum pixels match the reference (FS^{81}). If the GLCM kernel size is too large, more over-classification is caused around the referenced slum areas since the classifiers, especially LDA, tend to extend the slum areas. Therefore, the classification only profits from larger kernel sizes up to a certain point.

These two examples also explain the differences between the two classifiers in Sensitivity and PPV. Whereas LDA classifies slums more generously, RF delineates them more conservatively. Therefore, the slum patches are visibly bigger in the LDA classification and fill the reference areas more than RF, referring to Sensitivity. Yet, this also leads to more over-classification which is indicated in a lower PPV for LDA in comparison to RF. However, the PPV of the slums is relatively low for both classifiers. That is, a maximum of 59% of the slum areas mapped are also referring to slums.

The Specificity represents the amount of over-classification for each experiment. Both classifiers first benefit from an increasing kernel size, since the over-classification is visibly reduced with medium-sized kernels. RF continues to decrease in over-classification for all experiments, whereas LDA introduces over-classification with kernel sizes larger that 81 pixels. It can be seen that this is caused by the aforementioned extension of the slum areas by LDA.

The Error Rates of the individual mapped results seem to reflect the visual experience of a possible “best” result. While the small kernel sizes (11 and 21) show very heterogeneous patterns, the large kernels (121 and 161) tend to eliminate most of the small slum patches. The medium-sized kernels (41 and 81) are able to maintain balance between accurate spatial delineation and enough smoothing in order to create homogeneous patches.

The need for inspection of both overall and class specific accuracy measures is highlighted by the strong differences between their variations. In case of an imbalanced class distribution a separate discussion of the accuracy measures is necessary, since overall accuracy measures do not reflect the trends of small classes sufficiently.

Although an enormous amount of training samples is used for the creation of the classification models (cf. table 3), some slums cannot be detected sufficiently. In fact, the slum in the center of figure 5.3 features very regular structures which are quite unique in the whole area of investigation. It is detected quite accurately by LDA during the classification, however, only small parts are correctly classified by the RF classifier. Obviously, this is an extreme phenomenon of the already observed more conservative slum classification of RF. This leads to the assumption that the random forest model is not able to extrapolate the feature spaces like LDA. Thus, RF seems to require a higher number of training points to capture all imaginable spatial contexts and settings.

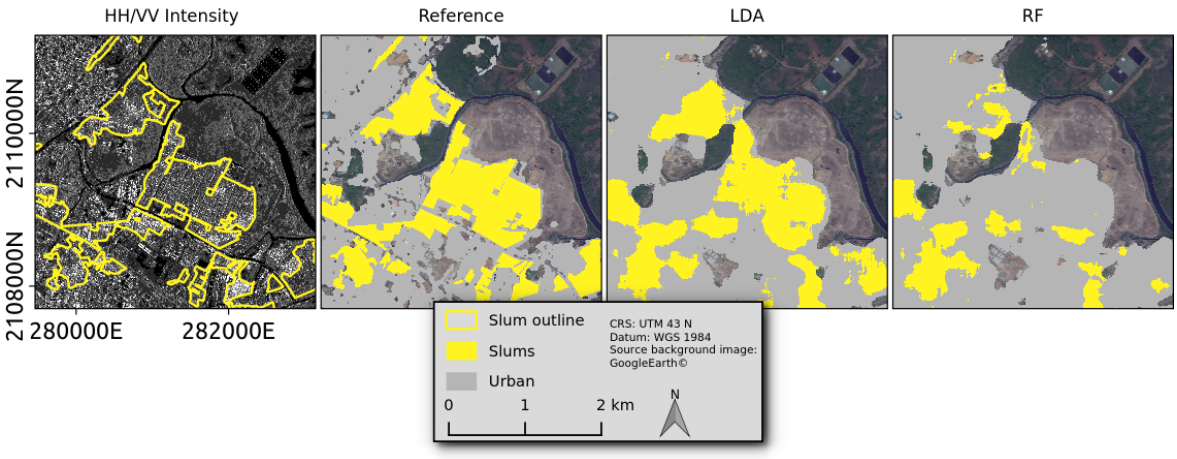


Figure 5.3: Example of a strongly organized slum showing significant differences in the classification results of LDA and RF

The classification results of both classifiers with FS^{81} are mapped in figure 5.4. They show that the slums basically are classified spatially along the referenced slum areas. That allows the conclusion that GLCM textures are able to describe unique properties of slums within an urban environment. Following the results from the overall and class specific accuracies, it can be seen that LDA classifies larger patches of slums than the RF classifier. Furthermore, LDA shows some linear features, especially in the north-western part of the classified area. These refer to linear rail tracks and cannot be detected correctly as urban. This can be seen as the drawback of the feature space extrapolation mentioned above. RF on the other hand does not misclassify these areas.

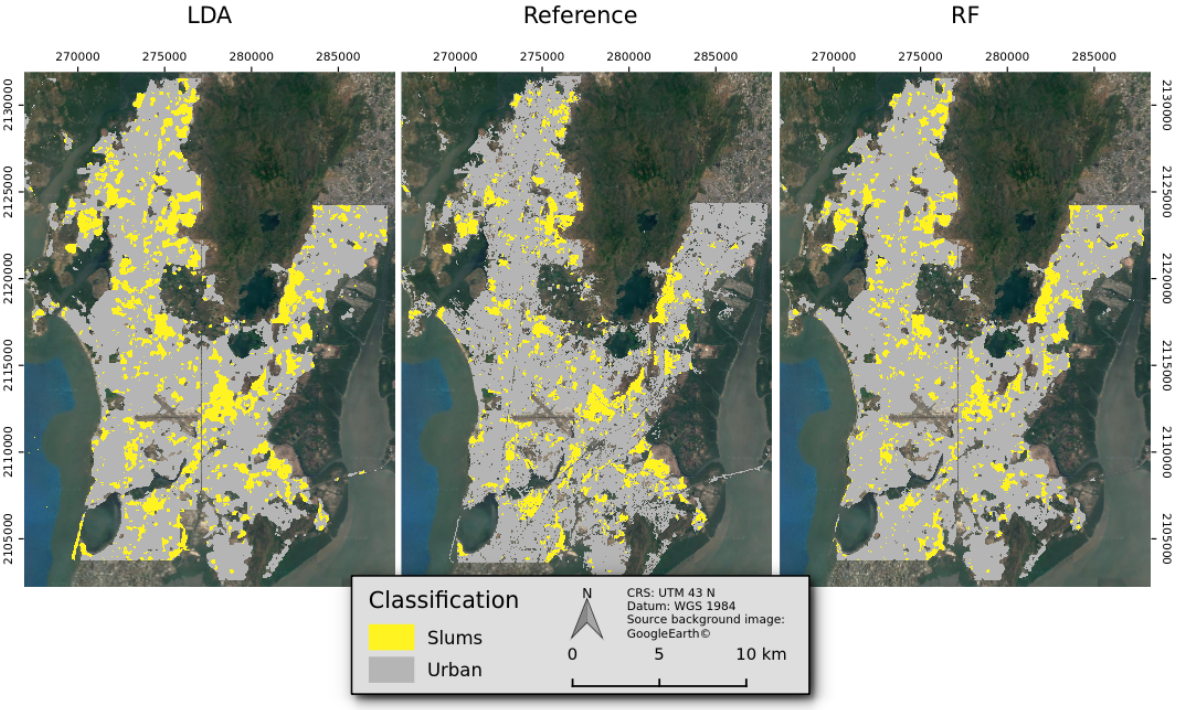


Figure 5.4: Reference and classification results with FS^{81} for both classifiers

5.2 Benefit of Extending Texture with Morphology

The enlargement of the GLCM feature space is tested with FS^{81} since these textures provide the best overall accuracies for both classifiers. This is achieved by adding the basic partially polarized Kennaugh element information and the Derivate of Morphological Profiles derived from the backscattering intensity (k_0). The accuracies, both overall and class specific measures, showed that an extension of GLCM texture features with these features is able to increase the classification accuracies in all cases. However, these changes were relatively little and come with an additional computational effort that is necessary for creating the DMP layers.

The visual impression of the classification result in figure 5.5 outlines these little differences. Both classifiers do not take significant visible advantages of the extension of the feature sets with DMP or Kennaugh elements. In contrast to the prior assumptions, the addition of DMP features does not boost the detection of the exact slum edges.

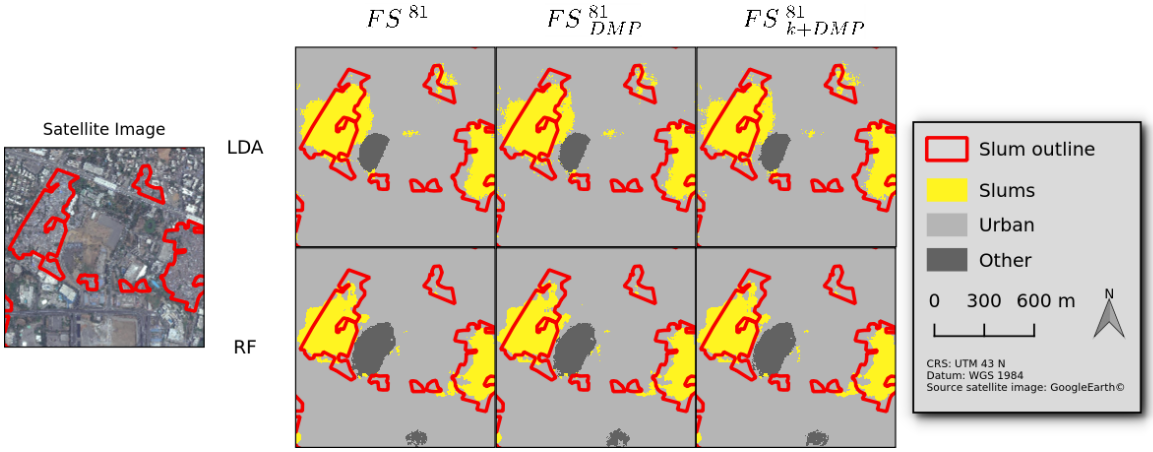


Figure 5.5: Exemplary classification results of the combined feature sets of LDA (top) and RF (bottom)

5.3 Importance of the Spatial Domain

By inspecting differently sized slums separately the assessment of the classification accuracy is extended to analyze its dependency on the spatial extent of the slums. As with the pixel based results before, the measures for this analysis also indicate a curved trend leaving the medium-sized kernels with the minimal ER values for all patch size classes. Whereas the LDA shows lower Error Rates throughout all kernel sizes in comparison to RF, the discrepancy between the three classes is larger than compared to RF. Since the PPV is assumed as static value, the variation of Sensitivity has to be higher between the small and large patches classified with LDA. The higher error of RF might result from the more conservative slum classification, so that the mutual overlap of T_O = 50% is not reached. Much like in the pixel based accuracy assessment, the slightly higher PPV of RF is not able to compensate the lower sensitivity.

Interestingly, the Error Rate of even the largest patches (25 to 198ha) increases with the highest kernel sizes. That is, for all the slum patch sizes the medium-sized kernels with an edge length of 41 and 81 pixels show the lowest Error Rate.

5.4 Possibility of Feature Reduction

The experimental feature reduction approach performed during this study is applied to decrease computational effort for the model creation and classification process. At the same time, the classification quality is expected to change. The feature reduction is performed by an initial ranking of the variable importance by the classification algorithms.

The results show that with reduced feature sets the classification qualities stay at a high level. Yet it is necessary to distinguish between the influences on overall and accuracy and the quality of slum discrimination. It occurs that the slums, which are underrepresented in the whole scene, need a higher number of image features until their classification quality stabilizes. While the overall accuracy measures suggest that very few textural features (e.g. 20 for RF and 40 for LDA) are needed to reach a sufficient quality of the results, a higher number of features is necessary if including the accuracy of the slum areas.

Interestingly, the composition of the 40 most important discriminating features (see figure 4.6) shows that the two classifiers deem different image features as more important. Based on these results a comparison of the importances of variables for the Random Forest classifier or the Linear Discriminant Analysis is not possible. Therefore, from this experimental approach an axiomatic set of most important image features necessary to classify slums from partially polarized SAR data cannot be identified conclusively.

Furthermore, in the approach of feature reduction, as it is used in this study, the variable ranking is performed once with the complete set of variables. This is due to high computational effort needed especially for the creation of the *cforest* models. However, it is possible that the ranking of the remaining variables would change by eliminating the n least important features. Therefore, an iterative determination of the variable importance appears to be necessary to perform a reliable feature reduction. Even more degrees of freedom would be introduced by total randomization of the input variables. Only thusly, any variable interference could be eliminated. Yet, due to the focus of this study such a broad feature reduction approach is not applied. Therefore, a final evaluation of the importance of each individual image feature is not possible at this point.

In conclusion, feature reduction can be applied to the classification of GLCM image features in this study. Yet, while overall measures induce an approximation of the maximum values with only fractional part of all image features, the class specific accuracy measures show that especially the small land use class slums benefits from more features.

5.5 Comparison of the Used Classifiers

All experiments in this study were conducted using two classification algorithms of different nature. Overall the classification results of LDA and RF showed similar trends throughout all experiments. Despite small differences, GLCM and DMP feature spaces enable the discrimination of informal settlements and are able to describe intra-urban structures to a certain degree.

The most important difference between the two used classifiers is that the LDA classifies slums more extensively than RF. This is represented by higher Sensitivity values of all accuracy assessment approaches. However, Specificity and PPV as well as the visual impression

of the resulting images indicate that this effect also leads to a higher over-classification around the referenced slum patches. Therefore, it can be stated that the LDA is more prone to the smoothing effect of the focal textures used in this study. On the contrary, RF is able to distinguish the smoothed image content in a higher level of detail. However, this comes with an under-representation of the slum class at hand.

However, the experiments show that the LDA is able to detect very unique structures within the image which are not trained (see figure 5.3). The more conservative classification behavior of RF does not extrapolate the feature space enough to detect such untrained objects. Therefore, it might be considered taking other classification methods into account which are able to compensate such untrained image objects.

6 Conclusion

With the ongoing trend of urbanization, the pressure to the cities of the world is growing. Huge urban areas with more than 10 million inhabitants emerge. In the next two decades the number of such *megacities* is predicted to increase to 41 (United Nations 2014). Besides various challenges, for example to infrastructure, job market or health risks, the drift to the cities also affects the conditions of the living environment. Especially in developing countries with little possibilities to counteract these challenges, the urbanization leads to the development of illegal or informal settlements. Those settlements, which are also called “slums”, are predominantly located in unappealing or even polluted land and feature no durable housing. Until the year 2020 one and a half billion people worldwide live under such sub-standard living conditions (Arimah 2010). To counter this problem of urban areas, the United Nations (2013) formulated the need for analyzing, mapping and monitoring the development of slums.

The methodology of geographical remote sensing using satellite images enables land use / land cover mapping of large areas. Until now, such data and methodology was used to develop different approaches for slum detection. Oftentimes only small areas of investigation were chosen for methodological development but no exhaustive city-wide mappings were conducted. Most of these studies used (very) high resolution optical imagery. However, since most of the developing countries are located in the tropical regions, optical images are often affected by cloud cover. On the contrary, active SAR sensors are able to penetrate clouds and other atmospheric components and can also acquire images at night. Therefore, they could provide valuable data for large area land use / land cover mapping.

To investigate the capability of SAR data for extensive urban area and slum mapping, this study uses two scenes of partially polarized SAR data acquired with the TerraSAR-X (TSX) and TanDEM-X (TDX) satellites. The study area is located in the Indian megacity Mumbai which is home to millions of slum dwellers. In order to train and validate the classification models an area-wide reference data set is created. The informal settlements were delineated in a previous study by visual interpretation of very high resolution optical images (Taubenböck and Wurm 2015) and adapted to the acquisition date of the SAR imagery in this study. Furthermore, urban areas are derived from the Global Urban Footprint (GUF), a binary mask holding all built-up man-made structures (Esch et al. 2013). A third class representing all other land cover, e.g. water, vegetation or bare soil, is assigned to all remaining areas. This reference is used for the training of two different classifiers, Random Forest (RF) and Linear Discriminant Analysis (LDA).

Using an existing framework for extracting Kennaugh-elements from the partially polarized data (Schmitt et al. 2015), information of different backscattering mechanisms of the actively sent microwaves is accessible with a high spatial resolution. From this image data different properties are derived. Texture is calculated using Gray Level Co-occurrence Matrix (GLCM) measures (Haralick et al. 1973) and morphological properties are extracted using Differential Morphological Profiles (DMP) (Benediktsson et al. 2003). The GLCM features are calculated in a focal approach using differently sized kernels. These are determined referring to the spatial extent of city block levels in order to investigate the influence of the spatial neighborhood on slum detectability. Hence, six square kernels are defined with varying edge lengths. The DMPs are calculated from the morphological image descriptors Opening and Closing by Reconstruction

with differently sized structuring elements.

In a broad experimental setup of mapping slums in the urban context, several feature sets are created. These are analyzed to determine the most suited image features for slum discrimination. At first, all GLCM textures are classified separately to identify the best discriminating GLCM kernel size. Afterwards, the most promising GLCM feature set is combined with the backscattering information from the original data and the DMP of the backscattering intensity layer. The classification accuracy is quantified using the area-wide reference data set in a spatially incoherent approach.

The experiments conducted in this study show that the size of the selected textural kernel affects the classification accuracy. This is indicated by a gradual change of all overall accuracy measures. Thus, GLCM textures derived from a 81 $\times$ 81 kernel are identified to achieve the highest overall classification qualities for both classifiers (LDA: OA = 87.90; RF: OA = 88.58). Both smaller and larger kernels show a decrease in accuracy. On the contrary, the class specific accuracy measures indicate that for the slums alone it is necessary to diversify the perspective, since not one single feature set holds the highest accuracies. Yet, classification quality of slums also follows a gradual change. By extending the feature set with the DMP and Kennaugh elements, no significant improvements to the classification quality are achieved.

An area based analysis of slums shows that larger patches are classified more accurately than smaller patches. In fact, the aforementioned gradual change of decreasing accuracies with very large GLCM kernels also applies to the largest slum patches with more than 25ha. Thus, once more the medium-sized kernels 41 $\times$ 41 and 81 $\times$ 81 can be identified as better suited for slum mapping than larger or smaller kernel sizes in this study.

In an experimental approach the number of image features are reduced based on variable importances of the two classifiers. This way, the influence of the number of features on the classification result was tested. It turned out that such a reduction of features mostly influences the detectability of slum areas, whereas overall measures showed more steady values. Yet, feature reduction has to be implemented using a recursive or even randomized approach, if more reliable results are to be provided.

In conclusion, it can be stated that textural image properties of partially polarized SAR imagery enable the discrimination of intra-urban structures. These image features provide information about the characteristic properties of the different urban landscapes. Therefore, SAR can be seen as a reliable data source for large-area slum mapping. In this study, GLCM textures with a kernel size of 81 $\times$ 81 return the best results in the applied experimental setup. Yet, although a gradual change of accuracies is detected, the experiments do not cover a more detailed examination of medium-sized GLCM kernels. In order to achieve a more reliable assertion about the best suited kernel, it is necessary to investigate medium-sized kernels with a finer spatial increment.

Besides the assessment of overall accuracy measures, it is important to consider the class specific accuracy as well as the spatial size of the slum patches. This study shows that the assessment of the accuracy should also include suitable measures for under- and over-classification, since slums are occupying only small shares of the whole urban area. In fact, overall measures are strongly biased towards the larger classes. Furthermore, the class specific

accuracy is strongly dependent on the chosen classifier. LDA tends to extend the slum areas and therefore introduce more over-classification. In contrary, RF showed a more conservative slum detection leading to fewer over-classification but also more under classification.

Even though it is possible to classify slums in a large area, the conducted experiments strongly rely on an extensive reference data set. It has also be shown that structures which are very unique cannot be detected with the certain classification methods. Also, the amount of misclassification should be reduced to generate more reliable results. Therefore, detailed investigations about unique properties of slum structures in multi-polarized SAR imagery are necessary to improve the classification success.

With the aim of mapping, analyzing and monitoring slum areas in megacities, further research could focus on transferring the results obtained in this study to other test sites including other cities and especially other cultural regions. Only thus the influence of the morphological differences between slums around the globe on the detectability can be assessed. Furthermore, one focus can be on taking other image features, such as lacunarity, into account. Moreover, combinations of multiple sensor types and other geoinformation like Volunteered Geographic Information, mobility patterns or geo-tagged social media data can be used for the development of automated slum detection.

References

Altman, D. G. and Bland, J. M. (1994a). “Diagnostic tests. 1: Sensitivity and specificity.” *British Medical Journal* 308 (6943), p. 1552.

Altman, D. G. and Bland, J. M. (1994b). “Statistics Notes: Diagnostic tests 2: predictive values”. *British Medical Journal* 309 (6947), p. 102.

Arimah, B. C. (2010). “The face of urban poverty: Explaining the prevalence of slums in developing countries”. *United Nations University, World Institute for Development Economics Research – Working paper* (2010, 30).

Asha, K (2006). “Urban slums in India – The myths and the reality”. *Asha-Seattle’s Quarterly Newsletter* 12 (2). Online; accessed December 30, 2016, pp. 1–3. URL: http://data.ashanet.org/files/Chapters/Seattle/Newsletters/Newsletter-2006Q1.pdf.

Backhaus, K., Erichson, B., Plinke, W. and Weiber, R. (2011). *Multivariate Analysemethoden – Eine anwendungsorientierte Einführung*. 13th ed. Berlin, Heidelberg: Springer Verlag. 538 pp.

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

Beijma, S. van, Comber, A. and Lamb, A. (2014). “Random forest classification of salt marsh vegetation habitats using quad-polarimetric airborne SAR, elevation and optical RS data”. *Remote Sensing of Environment* 149, pp. 118–129.

Belgiu, M. and Drăguţ, L. (2016). “Random forest in remote sensing: A review of applications and future directions”. *ISPRS Journal of Photogrammetry and Remote Sensing* 114, pp. 24–31.

Bellens, R., Gautama, S., Martinez-Fonte, L., Philips, W., Chan, J. C.-W. and Canters, F. (2008). “Improved classification of VHR images of urban areas using directional morphological profiles”. *IEEE Transactions on Geoscience and Remote Sensing* 46 (10), pp. 2803–2813.

Benediktsson, J. A., Pesaresi, M. and Amason, K. (2003). “Classification and feature extraction for remote sensing images from urban areas based on morphological transformations”. *IEEE Transactions on Geoscience and Remote Sensing* 41 (9), pp. 1940–1949.

Breiman, L. (1996). “Bagging predictors”. *Machine learning* 24 (2), pp. 123–140.

Breiman, L. (2001). “Random forests”. *Machine learning* 45 (1), pp. 5–32.

Breiman, L., Friedman, J., Stone, C. J. and Olshen, R. A. (1984). *Classification and regression trees*. Boca Raton: Chapman & Hall / CTC. 368 pp.

Chaabouni-Chouayakh, H. and Datcu, M. (2010). “Backscattering and statistical information fusion for urban area mapping using TerraSAR-X data”. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 3 (4), pp. 718–730.

Cohen, J. (1960). “A coefficient of agreement for niminal scales”. *Educational and Psychological Measurement* 20 (1), pp. 37–46.

Congalton, R. G. and Green, K. (2008). *Assessing the accuracy of remotely sensed data: principles and practices*. Boca Raton, London, and New York: CRC press. 183 pp.

Davis, M. (2006). *Planet of Slums*. London, New York: Verso. 228 pp.

Dell'Acqua, F. and Gamba, P. (2003). "Texture-based characterization of urban environments on satellite SAR images". *IEEE Transactions on Geoscience and Remote Sensing* 41 (1), pp. 153–159.

Dietterich, T. G. (2000). "An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization". *Machine learning* 40 (2), pp. 139–157.

Du, P., Samat, A., Waske, B., Liu, S. and Li, Z. (2015). "Random Forest and Rotation Forest for fully polarized SAR image classification using polarimetric and spatial features". *ISPRS Journal of Photogrammetry and Remote Sensing* 105, pp. 38–53.

Duque, J. C., Patino, J. E., Ruiz, L. A. and Pardo-Pascual, J. E. (2015). "Measuring intra-urban poverty using land cover and texture metrics derived from remote sensing data". *Landscape and Urban Planning* 135, pp. 11–21.

Dutt, A., G., P., Islam, I. and Chatterjee, I. (2016). "Cities of South Asia". In: Brunn, S. D., Graybill, J. K., M., H.-M. and J., Z. D., eds. *Cities of the World – Regional Patterns and Urban Environments*. 6th ed. Rowman & Littlefield, pp. 369–410.

Elachi, C. (1987). *Introduction to the Physics and Techniques of Remote Sensing*. New York: John Wiley & Sons. 413 pp.

Engstrom, R., Sandborn, A., Qin, Y., Burgdorfer, J., Stow, D., Weeks, J. and Graesser, J. (2015). "Mapping slums using spatial features in Accra, Ghana". In: *Joint Urban Remote Sensing Event (JURSE) 2015*. IEEE, pp. 1–4.

Esch, T., Marconcini, M., Felbier, A., Roth, A., Heldens, W., Huber, M., Schwinger, M., Taubenböck, H., Müller, A and Dech, S. (2013). "Urban footprint processor—Fully automated processing chain generating settlement masks from global data of the TanDEM-X mission". *IEEE Geoscience and Remote Sensing Letters* 10 (6), pp. 1617–1621.

Fauvel, M., Benediktsson, J. A., Chanussot, J. and Sveinsson, J. R. (2008). "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles". *IEEE Transactions on Geoscience and Remote Sensing* 46 (11), pp. 3804–3814.

Fernández-Delgado, M., Cernadas, E., Barro, S. and Amorim, D. (2014). "Do we need hundreds of classifiers to solve real world classification problems?" *The Journal of Machine Learning Research* 15 (1), pp. 3133–3181.

Fisher, R. A. (1936). "The use of multiple measurements in taxonomic problems". *Annals of eugenics* 7 (2), pp. 179–188.

Geiß, C., Klotz, M., Schmitt, A. and Taubenböck, H. (2016). "Object-based morphological profiles for classification of remote sensing imagery". *IEEE Transactions on Geoscience and Remote Sensing* 54 (10), pp. 5952–5963.

Ghosh, A., Fassnacht, F. E., Joshi, P. K. and Koch, B. (2014). "A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales". *International Journal of Applied Earth Observation and Geoinformation* 26, pp. 49–63.

Gong, P., Pu, R. and Yu, B. (1997). "Conifer species recognition: an exploratory analysis of in situ hyperspectral data". *Remote Sensing of Environment* 62 (2), pp. 189–200.

Graesser, J., Cheriyadat, A., Vatsavai, R. R., Chandola, V., Long, J. and Bright, E. (2012). “Image based characterization of formal and informal neighborhoods in an urban landscape”. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5 (4), pp. 1164–1176.

Graybill, J. K., Hays-Mitchell, M., Zeigler, D. J. and Brunn, S. D. (2016). “World Urban Development”. In: Brunn, S. D., Graybill, J. K., M., H.-M. and J., Z. D., eds. *Cities of the World – Regional Patterns and Urban Environments*. 6th ed. Rowman & Littlefield, pp. 3–46.

Habermeyer, M. and Schmullius, C. (1997). “Ein Algorithmus zur wissensbasierten Klassifikation multitemporaler Radar-Fernerkundungsdaten”. *Photogrammetrie, Fernerkundung, Geoinformation (PFG)*, pp. 313–323.

Haralick, R. M., Shanmugam, K. and Dinstein, I. H. (1973). “Textural features for image classification”. *IEEE Transactions on Systems, Man and Cybernetics* (6), pp. 610–621.

Haralick, R. M., Sternberg, S. R. and Zhuang, X. (1987). “Image analysis using mathematical morphology”. *IEEE transactions on pattern analysis and machine intelligence* (4), pp. 532–550.

Hariharan, S., Tirodkar, S. and Bhattacharya, A. (2016). “Polarimetric SAR decomposition parameter subset selection and their optimal dynamic range evaluation for urban area classification using Random Forest”. *International Journal of Applied Earth Observation and Geoinformation* 44, pp. 144–158.

Herold, M., Liu, X. and Clarke, K. C. (2003). “Spatial metrics and image texture for mapping urban land use”. *Photogrammetric Engineering & Remote Sensing* 69 (9), pp. 991–1001.

Hothorn, T., Hornik, K. and Zeileis, A. (2006). “Unbiased recursive partitioning: A conditional inference framework”. *Journal of Computational and Graphical statistics* 15 (3), pp. 651–674.

Japkowicz, N. and Stephen, S. (2002). “The class imbalance problem: A systematic study”. *Intelligent data analysis* 6 (5), pp. 429–449.

Kerwien, N. (2007). “Zum Einfluss von Polarisationseffekten in der mikroskopischen Bildentstehung”. PhD thesis. Universität Stuttgart.

Khalilia, M., Chakraborty, S. and Popescu, M. (2011). “Predicting disease risks from highly imbalanced data using random forest”. *BMC medical informatics and decision making* 11 (1), p. 1.

Klotz, M., Kemper, T., Geiß, C., Esch, T. and Taubenböck, H. (2016). “How good is the map? A multi-scale cross-comparison framework for global settlement layers: Evidence from Central Europe”. *Remote Sensing of Environment* 178, pp. 191–212.

Kohli, D., Sliuzas, R., Kerle, N. and Stein, A. (2012). “An ontology of slums for image-based classification”. *Computers, Environment and Urban Systems* 36 (2), pp. 154–163.

Kuffer, M., Pfeffer, K. and Sliuzas, R. (2016). “Slums from space – 15 years of slum mapping using remote sensing”. *Remote Sensing* 8 (6), pp. 455–484.

Kulkarni, A. D. and Lowe, B. (2016). “Random Forest Algorithm for Land Cover Classification”. *International Journal on Recent and Innovation Trends in Computing and Communication* 4 (3), pp. 58–63.

Kulkarni, V. Y. and Sinha, P. K. (2012). “Pruning of random forest classifiers: A survey and future directions”. In: *2012 International Conference on Data Science & Engineering (ICDSE)*. IEEE, pp. 64–68.

Landis, J. R. and Koch, G. G. (1977). “The measurement of observer agreement for categorical data”. *Biometrics* 33, pp. 159–174.

Liaw, A. and Wiener, M. (2002). “Classification and Regression by randomForest”. *R News* 2 (3), pp. 18–22. URL: http://CRAN.R-project.org/doc/Rnews/.

Lisini, G., Gamba, P. and Dell’Acqua, F. (2012). “A novel extension of the anisotropic rotation-invariant built-up presence index to SAR data”. *European Journal of Remote Sensing* 45 (1).

Masjedi, A., Zoej, V., Javad, M. and Maghsoudi, Y. (2016). “Classification of Polarimetric SAR Images Based on Modeling Contextual Information and Using Texture Features”. *IEEE Transactions on Geoscience and Remote Sensing*.

Moore, R. K., Chastant, L. J., Porcello, L. and Stevenson, J. (1983). “Imaging Radar Systems”. In: Colwell, R. N., Simonett, D. S. and Ulaby, F. T., eds. *Manual of Remote Sensing*. 2nd ed. Falls Church: American Society of Photogrammetry, pp. 429–474.

Mosley, L. (2013). “A balanced approach to the multi-class imbalance problem”. PhD thesis. Iowa State University.

Niebergall, S., Loew, A. and Mauser, W. (2008). “Integrative assessment of informal settlements using VHR remote sensing data—The Delhi case study”. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 1 (3), pp. 193–205.

Nobrega, R., O’hara, C. and Quintanilha, J. (2008). “An object-based approach to detect road features for informal settlements near Sao Paulo, Brazil”. In: Blaschke, T., Lang, S. and Hay, G., eds. *Object-based image analysis*. Springer, pp. 589–607.

Pacifici, F., Chini, M. and Emery, W. J. (2009). “A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification”. *Remote Sensing of Environment* 113 (6), pp. 1276–1292.

Pacione, M. (2006). “City Profile-Mumbai”. *Cities* 23 (3), pp. 229–238.

Pal, M. (2005). “Random forest classifier for remote sensing classification”. *International Journal of Remote Sensing* 26 (1), pp. 217–222.

Pau, G., Fuchs, F., Sklyar, O., Boutros, M. and Huber, W. (2010). “EBImage – an R package for image processing with applications to cellular phenotypes”. *Bioinformatics* 26 (7), pp. 979–981.

Peerbhay, K. Y., Mutanga, O. and Ismail, R. (2015). “Random Forests Unsupervised Classification: The Detection and Mapping of Solanum mauritianum Infestations in Plantation Forestry Using Hyperspectral Data”. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (6), pp. 3107–3122.

Pesaresi, M. and Benediktsson, J. A. (2001). “A new approach for the morphological segmentation of high-resolution satellite imagery”. *IEEE transactions on Geoscience and Remote Sensing* 39 (2), pp. 309–320.

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

Prasad, A. M., Iverson, L. R. and Liaw, A. (2006). “Newer classification and regression tree techniques: bagging and random forests for ecological prediction”. *Ecosystems* 9 (2), pp. 181–199.

Puissant, A., Hirsch, J. and Weber, C. (2005). “The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery”. *International Journal of Remote Sensing* 26 (4), pp. 733–745.

Richards, J. A. and Jia, X. (2006). *Remote sensing digital image analysis*. 4th ed. Berlin, Heidelberg: Springer. 439 pp.

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

Schmitt, A. and Brisco, B. (2013). “Wetland monitoring using the curvelet-based change detection method on polarimetric SAR imagery”. *Water* 5 (3), pp. 1036–1051.

Schmitt, A., Wendleder, A. and Hinz, S. (2015). “The Kennaugh element framework for multi-scale, multi-polarized, multi-temporal and multi-frequency SAR image preparation”. *ISPRS Journal of Photogrammetry and Remote Sensing* 102, pp. 122–139.

Serra, J. (1982). *Image analysis and mathematical morphology, v. 1*. Orlando, FL, USA: Academic Press, Inc. 610 pp.

Shanmugan, K. S., Narayanan, V., Frost, V. S., Stiles, J. A. and Holtzman, J. C. (1981). “Textural features for radar image analysis”. *IEEE Transactions on Geoscience and Remote Sensing* (3), pp. 153–156.

Soh, L.-K. and Tsatsoulis, C. (1999). “Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices”. *IEEE Transactions on geoscience and remote sensing* 37 (2), pp. 780–795.

Sternberg, S. R. (1986). “Grayscale morphology”. *Computer vision, graphics, and image processing* 35 (3), pp. 333–355.

Strobl, C., Boulesteix, A.-L., Zeileis, A. and Hothorn, T. (2007). “Bias in random forest variable importance measures: Illustrations, sources and a solution”. *BMC bioinformatics* 8 (1), p. 1.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T. and Zeileis, A. (2008). “Conditional variable importance for random forests”. *BMC bioinformatics* 9 (1), p. 1.

Stumpf, A. and Kerle, N. (2011). “Object-oriented mapping of landslides using Random Forests”. *Remote Sensing of Environment* 115 (10), pp. 2564–2577.

Taubenböck, H. and Kraff, N. J. (2015). “Das globale Gesicht urbaner Armut? Siedlungsstrukturen in Slums”. In: Taubenböck, H., Wurm, M., Esch, T. and Dech, S., eds. *Globale Urbanisierung*. Heidelberg and Berlin: Springer, pp. 107–119.

Taubenböck, H. and Kraff, N. (2014). “The physical face of slums: a structural comparison of slums in Mumbai, India, based on remotely sensed data”. *Journal of Housing and the Built Environment* 29 (1), pp. 15–38.

Taubenböck, H. and Wurm, M. (2015). “Ich weiß, dass ich nichts weiß – Bevölkerungsschätzung in der Megacity Mumbai”. In: *Globale Urbanisierung*. Heidelberg and Berlin: Springer, pp. 171–178.

Tuia, D., Pacifici, F., Kanevski, M. and Emery, W. J. (2009). “Classification of very high spatial resolution imagery using mathematical morphology and support vector machines”. *IEEE Transactions on Geoscience and Remote Sensing* 47 (11), pp. 3866–3879.

United Nations (2010). *State of the World’s Cities 2010/2011: Bridging the Urban Divide*. Tech. rep. United Nations, Human Settlements Programme, Nairobi.

United Nations (2013). *The Millennium Development Goals Report 2013*. Onilne; accessed December 30, 2016. URL: http://www.un.org/millenniumgoals/pdf/report-2013/mdg-report-2013-english.pdf.

United Nations (2014). *World Urbanization Prospects 2014: Highlights*. Tech. rep. United Nations, Department of Economic and Social Affairs, Population Division, New York.

UNPD (2011). *World Urbanization Prospects: The 2011 Revision*. Tech. rep. New York: United Nations Population Division.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Fourth. New York: Springer. URL: http://www.stats.ox.ac.uk/pub/MASS4.

Vincent, L. (1993). “Morphological grayscale reconstruction in image analysis: applications and efficient algorithms”. *IEEE transactions on image processing* 2 (2), pp. 176–201.

Waske, B., Linden, S. van der, Benediktsson, J. A., Rabe, A. and Hostert, P. (2009). “Impact of different morphological profiles on the classification accuracy of urban hyperspectral data”. In: *2009 First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*. IEEE, pp. 1–4.

Weiss, G. M. (2004). “Mining with rarity: a unifying framework”. *ACM SIGKDD Explorations Newsletter* 6 (1), pp. 7–19.

Williams, D. P., Myers, V. and Silvious, M. S. (2009). “Mine classification with imbalanced data”. *IEEE Geoscience and Remote Sensing Letters* 6 (3), pp. 528–532.

Wright, C. and Gallant, A. (2007). “Improved wetland remote sensing in Yellowstone National Park using classification trees to combine TM imagery and ancillary environmental data”. *Remote Sensing of Environment* 107 (4), pp. 582–605.

Wurm, M. (2013). “Verknüpfung von Fernerkundungsdaten und Survey-Daten (SOEP und BASE-II) in städtischen Räumen für sozialwissenschaftliche Analysen”. PhD thesis. Technische Universität Graz.

Wurm, M., Taubenböck, H., Roth, A. and Dech, S. (2009). “Urban structuring using multisensoral remote sensing data: By the example of the German cities Cologne and Dresden”. In: *2009 Joint Urban Remote Sensing Event*. IEEE, pp. 1–8.

Wurm, M., Taubenböck, H., Schardt, M., Esch, T. and Dech, S. (2011). “Object-based image information fusion using multisensor earth observation data over urban areas”. *International Journal of Image and Data Fusion* 2 (2), pp. 121–147.

Wurm, M., d’Angelo, P., Reinartz, P. and Taubenböck, H. (2014). “Investigating the applicability of Cartosat-1 DEMs and topographic maps to localize large-area urban mass concentrations”. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (10), pp. 4138–4152.

Wurm, M., Schmitt, A. and Taubenböck, H. (2016). “Building Types’ Classification Using Shape-Based Features and Linear Discriminant Functions”. *IEEE Journal of Selected Toppics in Applied Earth Observation and Remote Sensing*.

Wurm, M., Taubenböck, H., Weigand, M. and Schmitt, A. (2017). “Slum mapping in polarimetric SAR data using spatial features”. *Remote Sensing of Environment* 194, pp. 190–204.

Zhu, Z., Woodcock, C. E., Rogan, J. and Kellndorfer, J. (2012). “Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using Landsat and SAR data”. *Remote Sensing of Environment* 117, pp. 72–82.

Zvoleff, A. (2016). *glcm: Calculate Textures from Grey-Level Co-Occurrence Matrices (GLCMs)*. R package version 1.6.1. URL: https://CRAN.R-project.org/package=glcm.

Eidesstattliche Erklärung

Ich versichere, dass ich die vorliegende Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe, und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat. Alle Ausführungen der Arbeit, die wörtlich oder sinngemäß übernommen wurden sind als solche gekennzeichnet.

Ich bin einverstanden, dass die von mir angefertige Masterarbeit einer breiten Öffentlichkeit zugänglich gemacht wird.

☐ Nein.
☐ Ja, nach Abschluss des Prüfungsverfahrens.

Augsburg, den 13.02.2017

Matthias Weigand