ROYAL INSTITUTE
OF TECHNOLOGY

Multitemporal Remote Sensing for Urban Mapping using KTH-SEG and KTH-Pavia Urban Extractor

Alexander Jacob

June 2014

Licentiate Thesis in Geoinformatics

KTH Royal Institute of Technology
Department of Urban Planning & Environment
Drottning Kristinas Väg 30, 100 44 Stockholm

TRITA-SoM 2014-08
ISSN 1653-6126
ISNR KTH/SoM/14-08/SE
ISBN 978-91-7595-188-1

Printed by
Universitetsservice US AB
Stockholm, Sweden, 2014

Abstract

The objective of this licentiate thesis is to develop novel algorithms and improve existing methods for urban land cover mapping and urban extent extraction using multi-temporal remote sensing imagery. Past studies have demonstrated that synthetic aperture radar (SAR) have very good properties for the analysis of urban areas, the synergy of SAR and optical data is advantageous for various applications. The specific objectives of this research are:

1. To develop a novel edge-aware region-growing and -merging algorithm, KTH-SEG, for effective segmentation of SAR and optical data for urban land cover mapping;

2. To evaluate the synergistic effects of multi-temporal ENVISAT ASAR and HJ-1B multi-spectral data for urban land cover mapping;

3. To improve the robustness of an existing method for urban extent extraction by adding effective pre- and post-processing.

ENVISAT ASAR data and the Chinese HJ-1B multispectral , as well as TerraSAR-X data were used in this research. For objectives 1 and 2 two main study areas were chosen, Beijing and Shanghai, China. For both sites a number of multitemporal ENVISAT ASAR (30m C-band) scenes with varying image characteristics were selected during the vegetated season of 2009. For Shanghai TerraSAR-X strip-map images at 3m resolution X-band) were acquired for a similar period in 2010 to also evaluate high resolution X-band SAR for urban land cover mapping. Ten major landcover classes were extracted including high density built-up, low density built-up, bare field, low vegetation, forest, golf course, grass, water, airport runway and major road.

For Objective 3, eleven globally distributed study areas where chosen, Berlin, Beijing, Jakarta, Lagos, Lombardia (northern Italy), Mexico City, Mumbai, New York City, Rio de Janeiro, Stockholm and Sydney. For all cities ENVISAT ASAR imagery was acquired and for cities in or close to mountains even SRTM digital elevation data.

The methodology of this thesis includes two major components, KTH-SEG and KTH-Pavia Urban Extractor. KTH-SEG is an edge aware region-growing and -merging algorithm that utilizes both the benefit of finding local high frequency changes as well as determining robustly homogeneous areas of a low frequency in local change. The post-segmentation classification is performed using support vector machines. KTH-SEG was evaluated using multitemporal, multi-angle, dual-polarization ASAR data and multispectral

HJ-1B data as well as TerraSAR-X data. The KTH-Pavia urban extractor is a processing chain. It includes: Geometrical corrections, contrast enhancement, builtup area extraction using spatial stastistics and GLCM texture features, logical operator based fusion and DEM based mountain masking.

For urban land cover classification using multitemporal ENVISAT ASAR data, the results showed that KTH-SEG achieved an overall accuracy of almost 80% (0.77 Kappa) for the 10 urban land cover classes both Beijign and Shanghai, compared to eCognition results of 75% (0.71 Kappa) In particular the detection of small linear features with respect to the image resolution such as roads in 30m resolved data went well with 83% user accuracy from KTH-SEG versus 57% user accuracy using the segments derived from eCognition. The other urban classes which in particular in SAR imagery are characterized by a high degree of heterogeneity were classified superiorly by KTH-SEG. ECognition in general performed better on vegetation classes such as grass, low vegetation and forest which are usually more homogeneous.

It is was also found that the combination of ASAR and HJ-1B optical data was beneficial, increasing the final classification accuracy by at least 10% compared to ASAR or HJ-1B data alone. The results also further confirmed that a higher diversity of SAR type images is more important for the urban classification outcome. However, this is not the case when classifying high resolution TerraSAR-X strip-map imagery. Here the different image characteristics of different look angles, and orbit orientation created more confusion mainly due to the different layover and foreshortening effects on larger buildings. The TerraSAR-X results showed also that accurate urban classification can be achieved using high resolution SAR data alone with almost 84% for eight classes around the Shanghai international Airport (high and low density built-up were not separated as well as roads and runways).

For urban extent extraction, the results demonstrated that built-up areas can be effectively extracted using a single ENVISAT ASAR image in 10 global cities reaching overall accuracies around 85%, compared to 75% of MODIS urban class and 73% GlobCover Urban class. Multitemporal ASAR can improve the urban extraction results by 5-10% in Beijing. Mountain masking applied in Mumbai and Rio de Janeiro increased the accuracy by 3-5%.The research performed in this thesis has contributed to the remote sensing community by providing algorithms and methods for both extracting urban areas and identifying urban land cover in a more detailed fashion.

Keywords: KTH-SEG, ASAR, HJ-1B, Urban Land Cover Mapping, OBIA, Segmentation, Image Classification, KTH-Pavia Urban Extractor, Urban Extent

Acknowledgements

First of all I'd liked to thank my main supervisor Professor Yifang Ban for the opportunity to work on the research, and for her guidance, support and invaluable comments and suggestions during the course of this thesis research. I'd also like to thank my co-superviros Associate Professor Paolo Gamba at University of Pavia for the opportunity to spend three months at Pavia in 2013 and for his constructive comments and suggestions.

I also want to thank my mother Brigitta and Mark for their never ending support of my studies here in Stockholm. The stability and tranquility provided by them created the best foundation I could wish for!

I further liked to thank my fellow PhD student Jan Haas for all the nice time that we spend together in the office, sharing both the tough and happy moments.

Elisabetta Troglio I want to thank for all the nice lunches that we spend together. That helped me to switch off during work ours and of course was valuable free Italian language training!

Hans Hauska I like to thank for his valuable comments for improving the quality of the language in this thesis and all the upcheering talks and coffee breaks.

I also like to thank all other Phd students at our division for all the nice discussions both on work and other topics!

Alexander Jacob, Stockholm, June 2014

Table of Contents

List of Papers

Paper I:
Jacob, A., Ban, Y., 2012, Segmentation of Multi-Temporal ENVISAT ASAR and HJ-1B Optical Data Using an Edge-Aware Region Growing and Merging Algorithm, Proceedings, Dragon 2 Final Results and Dragon 3 Kick-Off Sympoisum, Beijing, P.R. China, June 25th -29th 2012

Paper II:
Ban, Y., Jacob, A., 2013, Object-Based Fusion of Multitemporal Multiangle ENVISAT ASAR and HJ-1B Multispectral Data for Urban Land-Cover Mapping, IEEE Transactions on Geoscience and Remote Sensing, Vol. 51, No. 4, April 2013, pp.1998-2006

Paper III:
Ban, Y., Jacob, A., Gamba, P., 2014, Spaceborne SAR Data for Global Urban Mapping at 30m Resolution Using a Robust Urban Extractor, submitted for publication at ISPRS Journal of Photogrammetry and Remote Sensing, Special Issue on Global Land Cover Mapping.

Paper IV:
Jacob, A., Ban, Y., 2014, Urban Land Cover Mapping with TerraSAR-X using and Edge-Aware Region-Growing and Merging Algorithm, IGARSS 2014, Quebec Canada, 13-18 July 2014

List of Figures

List of Tables

List of Acronyms

ANN – Artificial Neural Networks

GLCM - Grey Level Co-occurrence Matrix

KTH-SEG – KTH Segmentation

MDC – Maximum Distance Classifier

MLC – Maximum Likelihood Classifier

OBIA – Object Based Image Analysis

RADAR - Radio Detection And Ranging

RMS - Root Mean Square

SAR - Syntetic Aperture Radar

SVM - Support Vector Machine

SRTM – Shuttle Radar Topography Mission

1 Introduction

1.1 Rational for the Research

Urbanization is a phenomenon ongoing all around the globe. Since 2008, more than half of the worlds population was already living in cities (UN, 2011), with a tendency of a significant future increase also within the next decades. In 2030 more than 60% of the worlds population are expected to live in cities. This strong current of human beings into the cities has a strong impact on the structure of those cities and their growth rate.

This changes not only the urban structure but also alters the climate, influences the ecology and how life in society is organized, on local, regional and global scale (Schneider et al. 2009, Seto et al., 2011).

For understanding the detailed sources of those changes, urban mapping is a very important source of information being able to model the urban environment in a comprehensive way.

A very suitable source due to it's global accessibility is satellite imagery. A large variety of missions providing information in various part of the electromagnetic spectrum is currently available. Multi-spectral missions like the Landsat, Spot or Geoeye series cover the short-wave spectrum and different RADAR missions like Sentinel-1, TerraSAR-X or Radarsat cover longer wave lengths. These different sources can contribute with different kinds of information gain. We have e.g. information about color and texture, infrared and heat emission as well surface roughness and electrical properties. The large number of missions available is a blessing for civil research, but it also creates the need for handling large amounts of data preferably in a more and more automated way to keep analysis time and cost efficient and low. This raises a need for new analysis and algorithm development for handling this data.

With new missions also the resolution and hence minimum mapping unit increases greatly. From very coarse resolutions as provided from MODIS (250 – 1000m) over medium resolution as e.g. ENVISAT ASAR or Landsat (30m) to high and very high resolutions imagery as for example from Quickbird or TerraSAR-X (1m), everything is

available nowadays. How data is best analyzed changes drastically with the resolution, since also the physical surface of the earth that is contributing to the final observed value represented in one pixel changes. Pixel resulting from large surface of 10^{th} or 100^{th} of meters are often constructed by a multitude of different land-cover types especially in urban areas with it's high heterogeneity, whereas pixels from high resolution imagery of a couple of meters and below typically only represent one type of land cover. Here it is meaningful to convert from pixel-based image analysis to object based image analysis (Blaschke & Strobl, 2001, Blaschke, 2010). Especially in higher resolution, but also in mid resolution imagery object based image analysis can be beneficial due to the new features like geometrical and topological properties that it brings into the analysis process (Jensen, 2005).

The key to working with image objects lies in image segmentation which describes a number of different methodologies of how to extract image objects from raster data. This is not a new research fields (Haralick & Shapiro, 1984, Pal & Pal, 1993 or Hay & Castilla, 2006), but the advances in computation technology make the construction of more and more potent algorithms possibile and hence new algotihms are frequently presented.

1.2 Research Objectives

Paying tribute to the rapidly developing field of space-borne earth observation both in terms of quantity and quality the focus of this research lies in working with satellite imagery.

The objective of this licentiate thesis is to improve existing methods and develop novel methods and algorithms for urban extent extraction and urban land cover mapping using multi-temporal remote sensing imagery. Past studies have demonstrated that synthetic aperture RADAR (SAR) imagery is well suited for the analysis of urban areas. The synergy of SAR and optical data is advantageous for various applications. The specific objectives of this research are:

1. To develop a novel edge-aware region-growing and merging algorithm, KTH-SEG for effective segmentation of SAR and optical data for urban land cover mapping

2. To evaluate the synergistic effects of multi-temporal ENVISAT ASAR and HJ-1B multi-spectral data for urban land cover mapping

3. To improve the robustness of an existing method for urban extent extraction by adding effective pre- and post-processing.

1.3 Organization of Thesis

This thesis is organized in 6 main chapters. Chapter 1 gives a brief rational and introduces the main objectives for this research. Chapter 2 gives an overview over the literature touching the main research fields in this thesis. Chapter 3 describes the main study areas and data used during the course of the development of KTH-SEG and KTH-Pavia Urban Extractor. In Chapter 4 the methodology behind the previously named tools is inspected and described thoroughly. Chapter 5 presents and discusses the main findings of the different studies compiling this thesis and Chapter 6 gives round up of the whole work concluding with the contribution of this research and an outlook of where to go from here. In the end a reference list and the 4 papers written during the course of this licentiate are presented. Papers I, II & IV cover applications of KTH-SEG and paper III describes the KTH-Pavia Urban Extractor. Paper I focuses on data fusion aspects of SAR and optical data. Paper II explains the principles of KTH-SEG, compares it with eCognition and explores the suitability of different combinations of SAR data stacks for urban land cover mapping. Paper IV provides some insight into urban land cover mapping with high resolution X-band SAR data.

1.4 Contribution in Co-Authored Papers

In all co-authored papers of this thesis I Alexander Jacob have contributed to both writing and practical performance of the research conducted. All visualizations were co-designed by Yifang Ban and me, and were produced by me with the exception of the world map showing the study areas in paper III, which was done by Yifang Ban.

In papers II & III, where Yifang Ban is the first author, she initiated the ideas for the research and has done the main writing of the paper. I performed the implementations of the algorithms and have contributed to the writing of selected sections. In the case of paper II I have written a major part of the methodology, and co-written the sections on data dewscription, as well as results and discussion. In paper III I have co-written the methodology, results and discussion sections. Paolo Gamba contributed with the original algorithms and help with the revision of the paper.

In the paper I & IV the main drafting of the paper was done by me. Yifang Ban participated in the experimental design, data selection and discussions of the methods and results; co-wrote the abstract, introduction and conclusions and edited the rest of the paper prior to publication.

2 Literature Review

2.1 Remote Sensing for Urban Land Cover Mapping

2.1.1 Optical Data for Urban Land Cover Mapping

Urban land cover mapping from optical data can be seen as the traditional way and has been performed in numerous studies. In recent years the studies for urban land cover mapping and classification focus predominantly on high resolution satellite imagery (e.g IKONOS, Quickbird, Worldview, Geoeye , etc.). Weng et al. (2014) provided a thorough review on all sensors that are used for urban mapping and observations, including optical, SAR and night-time images. Hester et al. (2008) used multi-spectral Quickbird imagery that they pan-sharpened to 0.61m resolution. Using a hybrid classification approach based on a supervised MLC and ISODATA they achieved about 89% overall accuracy and 0.87 kappa for 6 classes over the urban area of suburban Raleigh, North Carolina, USA. Also Thapa and Murayama (2009) researched in a similar direction, combining MLC, ISODATA and fuzzy classification approaches. Evaluating each of the three methods individually, as well as the combined approach, they found that combination performs best with 89% overall accuracy and 0.87 kappa on ALOS AVNIR2 imagery over Tsukuba, Japan. Pacifici et al. (2009) developed a texture based neural network classification algorithm, which was tested over Rome and several American cities. Panchromatic Quickbird and WorldView-1 images were classified with up to 0.9 in kappa accuracy on ten to eleven classes. Bashkeran et al. (2010), Moran (2010) and Myint et al. (2011) compared object- vs. pixel-based classification accuracy in urban reas using IKONOS respectively Quickbird images. Both found that object based classification was superior to pixel-based classification with an advantage of at least 10% overall accuracy or 0.1 in kappa.

Novack et al. (2011) analyzed WorldView-2 and simulated Quickbird-2 data using four different classifiers, Decision Tree, Random Forest, SVM and Regression Tree. Random forest was found to be best in their object-based classification with on average 0.95 of kappa accuracy. They further found that more spectral bands have a positive impact on the final classification outcome, especially in shadowed ares. Pu et al.

(2011) compared ANN and MDC as well as object-based and pixel-based classification. Finding that the object-based ANN performed best on IKONOS imagery over Tampa Bay, Florida, USA. Zhang et al. (2009) analyzed the trajectories of urban land and industrial land in Shanghai over the 30 years using historical Corona, Landsat and Spot imagery. Using a hierarchical object based classification approach they achieved up to 86% overall accuracy. Hu *et al*. (2009) analyzed urbanization in the Yangtze River Delta using Landsat and CBERS imagery. For 4 eoochs the the urbanization index was calculated and compared. Furberg and Ban (2012) classified Landsat-5 imagery over 3 different epochs for monitoring urban sprawl. They used a combination of texture features and spectral features as input for a MLC reaching accuracies of 0.92-0.94 in kappa. Haas and Ban (2014) performed urban land cover classification for the purpose of post-classification change detection using random forests classification on Landsat-5 and HJ-1A&B imagery over three urban regions in China. For this mid-resolution imagery classification accuracies between 0.80 – 0.86 where reached having eight classes.

2.1.2 SAR Data for Urban Land Cover Mapping

That SAR imagery is suitable for urban land cover mapping has been demonstrated. A good summary of how to apply SAR sensors for this task, is given by Henderson and Xia (1997) and Xia and Henderson (1997). In these two papers they show for which tasks and with which information SAR can be applied for urban land cover mapping. In many cases they refer to airborne sensors, but in the present we have a number of spaceborne sensors available providing just as good data (e.g. TerraSAR X, Radarsat-2 and ALOS PALSAR). These sensors provide both, high resolution and images of different polarization, two important sources of information for successful mapping of urban areas. Other important features mentioned are images of different band (i.e. L, C or X-band), orbit orientation and angle of incidence (Xia and Henderson, 1997). The use of interferometric SAR has also been proved to provide reasonable results as well (e.g., Strozzi *et al*., 2000; Engdal and Hyyppä, 2003). They demonstrated this using a large number of ERS1&2 image pairs. Li et al. (2010) used polarimetric features derived from aerial high resolution L-Band SAR and could at least for the urban class of their classification scheme of 7 classes achieve over 80% accuracy. Alberga (2007) reached similar results with the same airborne

sensor and analysis based on polarimetric features finding some more urban classes but even here good results were only achieved for some classes.

Apart from which products of SAR imagery are used there is also a large variety in the methodology applied to these scenes. Dell'Acqua and Gamba (2006) used a multi-scale grey-level-co-occurrence-matrix (GLCM) approach reaching roughly 70% accuracy with ENVISAT ASAR and Stasolla and Gamba (2008) with ALOS PALSAR data achieved similar accuracies. Yang et al. (2009) used GLCM, Gabor and histogram of orientated gradients as input for an extreme random forest classifier on TerraSAR X spotlight imagery and reached 80% overall accuracy in their urban land cover map. Voisin et al. (2010) applied three different methods based on Markov random fields, GLCM and Power density functions (PDF) on TerraSAR X and Cosmo Skymed data for solving a simple land cover classification problem of three classes, finding that a combination of intensity and textural features yield the best results. In their study they obsevered slightly better results using Cosmo Skymed data than using TerraSAR X data. An improved version of their algorithm is presented in Voisin et al. (2012). Testing based on Cosmo Skymed data was very successful for their previous three class problem (98% overall accuracy). Esch et al. (2010) performed urban settlement detection using strip-mode imagery from TerraSAR X. They achieved 90% accuracy with their approach based on local speckle divergence and intensity of backscatter. It should be noted that the number of classes sought to be identified in those papers was always quite low (5 or less).

With very high resolution imagery the focus of techniques changes again. Here the SAR specific characteristics of the path of the signal and how it is reflected on different surfaces is of even higher importance. A study based on simulated data is performed by Maragerit et al. (2010). Another study (Reigber et al., 2007) uses sub-aperture decomposition of full polarimetric 1m x-band SAR from an airborne sensor. Hu and Ban (2012) analyzed Radarsat-2 C-Band imagery of Canada using an object-based and hierarchical knowledge-based classification approach. They were able to distingiush both urban land cover classes with an overall accuracy of almost 82% from 3 SAR scenes. In a series of papers Niu & Ban (2012, 2013a, 2013b & 2014) develop algorithms for urban classification using full polarimetric SAR

data from the Radarsat-2 mission. Outgoing from an stochastic expection-maximization (SEM) algorithm in combination with a markov random field (MRF) applied on a spatially variant finite mixture model (SVFMM), classification overall accuracies of up to 85% (0.83 kappa) where achieved on ten different classes. Introducing a rule-based classification scheme could improve the accuracy further to 89% overall accuracy. Switching from pixel-based to an object based approach and introducing SVM as the classifier the accuracies went up to 90% overall accuracies.

2.1.3 Fusion of SAR and Optical Imagery for Urban Land Cover Mapping

In the recently published book chapter, Ban *et al*. (2014) provided a comprehensive overview of the latest techniques and novel methods on fusion of SAR and optical data for urban land cover mapping and change detection.

A good introduction into the topic and a categorization of different approaches is given by Richards (2012) Chapter 12. He outlines four solution categories: image stacking, statistical merging, evidence theory and knowledge based approaches. The first two approaches he find only reasonable to be used with sensors of the same type of data, like for example two optical sensors or two radar sensors, the latter two can be used with arbitrary kind of sensors.

Another way of dividing fusion techniques is given by Pohl and Van Genderen (1998), which is also used by Zhang (2010). They divide the data fusion into pixel based, feature based and knowledge based fusion. Pixel based fusion is further divided into component substitution techniques such as PAN sharpening by IHS transformation, modulation-based fusion techniques and finally MRA (multi resolution analysis) fusion techniques. All these are mainly applied to fusion among one sensors bands or at least of very similar character. For high level fusion of sensors like SAR and optical very often either post classification or a combined classification are selected. In that case preferably a classifier independent of the underlying distribution function such as neural networks, support vector machines (SVM) or Markov random fields (MRF) should be used.

Zhang et al. (2010) created a block based regression model for fusing data of different sources, outperforming techniques like IHS, PPA, Wavelet-based and SVR in terms of correlation between fused data and original data sets. It proved to be suitable for SAR and optical data specifically. This technique falls in the feature based merging category.

Before actual image fusion can be performed however, images need to be correctly co-registered to each other. Traditionally this is performed by manually selecting identical points in image pairs and based on these calculate the transformation parameters to bring one image into the coordinate system of the other or the corresponding reference system if one of the images is geo-referenced. Classic approaches can be found in Jensen (2005) (Ch. 7) or Richards (2012) (Ch. 2), who treat also all other geometric distortions and effects there. This however is not very feasible when many images are involved. Here automatic approaches like in Zhang et al. (2000) should be applied. Their method is based on local grey level matching on image pyramids. A well of information on the topic in general in form of a very extensive literature review is given by Zitová and Flusser (2003). They covered all aspects of the topic for a number of different disciplines like remote sensing, computer vision and medical applications. They divided the methods into area based and feature based methods, where the classical approach of control point selection would be feature based. They also described the process as a whole, dividing it into 3 steps: feature or area selection, transformation and image resampling. Due to the difference in dynamic range in terms of grey levels between radar and optical data, the area based methods are less suited than the feature based approaches for these kinds of problems. A more recent approach by Yu et al. (2008), combines scale invariant feature transformation (SIFT) for coarse matching and piecewise linear transformation based on the Harris corner detector for matching Quickbird, Spot and Landsat imagery. Another approach by Müller et al. (2007) is based on edge detection, where the detected edges from several images can be used to bring them together or to be matched with vector based map databases for geo-referencing. In 2010 Wong and Clausi published an improvement on their ARRSI algorithm from Wong and Clausi (2007). This latest version is another feature based technique based on Wavelets, that has been proven to be useful for combinations of sensors of very different characters like optical and SAR as well as optical and LiDAR in a fully automatic manner. They got rms-errors of only a few pixels which is very good for an automatic

approach, but might still not be good enough for common processing where typically sub-pixel registration accuracy is required. One final example of a rather successful intensity based approach is from Chen et al. (2003). They were able to co-register quite well RADARSAT and Landsat TM images with their mutual information based algorithm.

Examples of successfully applied data fusion, can among others, be found by Ban (2003) where she combined a time series of ERS-1 and Landsat TM images for agricultural classification of 8 different crop types, finding that the fusion of sensors is superior to single sensor processing as well as an improved robustness of classification using artificial neural networks in comparison to a maximum likelihood classifier. A similar survey was undertaken by McNairn et al. (2009) using Radarsat 1, ENVISAT ASAR, Spot 4 & 5 and Landsat 5 data, explicitly pointing out the advantage of fused data sets over single sensor surveys. Ban et al. (2010) made use of Quickbird and Radarsat data for urban land cover mapping fusing the data on a decision level after segmentation, mainly to solve confusion problems of specific classes in the Quickbird image with the aid of Radarsat, which proved to be very successful for specific classes. Corbane et al. (2008) combined Radarsat with Spot 4 and ENVISAT ASAR with Spot 5 imagery for the purpose of rapid urban mapping. They started with individual texture analysis of the images and the actual fusion was performed afterwards using a fuzzy decision rule set. Their results also yield potential in fusing optical and radar data. Another interesting article is given Pacifici et al. (2007) summarizing the results of the GRSS Data Fusion Contest of 2007 where the participants were provided with ERS 1 &2 as well as Landsat 5 & 7 data. The winning algorithm was based on neural network, where the fused input was derived from principle component analysis of the individual datasets.

Except Ban et al. (2010) all other quoted papers in this section are making use of pixel based techniques. In recent years however object based approaches have taken over, due to their advantages by adding the information created by a larger group of pixels. Object Based Image Analysis

2.1.4 Segmentation

In contrast to pixel based analysis objects based image analysis can profit from additional features of objecst like texture, the shape and even the spatial context of several objects. The developed algorithms can roughly be divided into three groups Blaschke et al. (2005): Pixel based (e.g. thresholding techniques), edge based (i.e. determine boundaries of segments to define those) and region based segmentation techniques such as region growing, merging and splitting. The growing techniques start generating segments from the pixel domain by local homogeneity criteria. The splitting techniques work the other way around and in extreme cases start with one segment representing the whole image which is then subdivided. In both cases, growing and splitting, often a merging phase follows putting together adjacent segments if they are sufficiently alike based on statistical measures representing e.g. spectral and textural information. A good review of applications with optical data is given by Blaschke (2010). The most applied approach since it is implemented in eCognition Trimble (2011) and Berkely Image Segmentation BerkEnviro (2011), is based on Benz et al. (2004). Minor differences of their performance can only be due to propriety changes within the closed source code (Clinton et al., 2010). A variety of segmentation procedures has been proposed utilizing one or more of the previously stated properties for the segmentation of SAR images. Yang et al. (2008) proposed a method using a combination of Markov-Random-Field and Region Adjacency Graphs for segmentation of optical and SAR data, X. Zhang et al. (2008) used a combination of spectral clustering techniques for segmentation of SAR data, Quan et al. (2008) experimented with a combination of probabilistic neural networks and multiscale autoregression, Galland et al. (2009) segmented high resolution airborne SAR data as well as mid resolution ERS-1 using Fisher-Distributions, Tan et al. (2009) used a different approach based on MAP classification and anisotropic diffusion smoothing and Liu et al. (2010) created a multi-objective immune clustering ensemble technique (MICET) which is also performed for segmentation on SAR data. Further examples of specific interest for this work are given by Z. Wang et al. (2010), who proposed a method utilizing spectral information as well as shape measures for a statistical region merging process based on optical data. Barbieri et al. (2010) used an entropy based approach which works with the texture of an optical data object, Gu et al. (2008) used polarimetric SAR data in a

statistical region merging procedure where Y. Wang et al. (2010) used a clustering approach utilizing clustering techniques in Tensor space, Carvalho et al. (2009) used another method for SAR images also growing and merging regions. An enhanced region growing approach using edge penalties was presented by Yu and Clausi (2008). Testing on artificial and real SAR imagery showed very good agreement with manual derived segmentation. A further improvement of the algorithm was later presented by Qin and Clausi (2010). Most recently a new approach using stationary wavelet transforms (SWT) and algebraic multigrids has been developed and presented by Deng et al., (2014). A post segmentation classification over Toronto, Canada, was performed in 80% overall accuracy using Radarsat-2 imagery. A comparison to eCognition delivered results closte to 77% overall accuracy.

Evaluation of segmentation results is discussed by H. Zhang et al. (2008). They discuss problems of current techniques and give hints for future improvements by, for example combining some of the existing methods. Marpu et al. (2010) presented a method for estimating over- and under-segmentation, which was tested on 12 different segmentation implementations and might be a promising method for segmentation evaluation in general.

2.1.5 Classification

The topic of classification for the purpose of land cover mapping is of particular interest. Commonly used techniques in the field of remote sensing are, e.g. Maximum-Likelihood-Classifier (MLC), Nearest-Neighbor-Classifier, Neural-Networks (NN) of different kinds, decision and rule based methods as well as Support-Vector-Machines (SVM). Due to characteristics of the dataset as well as the classes sought to be found, classifiers based on specific assumptions on data distribution,such as the widely used MLC, are not used. This decision was supported by Huang et al. (2002), who compared SVM with neural networks, MLC and decision tree classifiers on multispectral Landsat and MODIS images. Key findings were that SVM, though the slowest method, performed best especially with large input data sets in terms of bands. Pal & Mather (2003) performed an evaluation of classifiers for land cover mapping as well, finding that the SVM performed

significantly better than neural network or MLC applied on different scenes of ETM+ data. Melgani & Bruzzone (2004) evaluated the performance of various SVM classifiers in comparison to nearest neighbor and radial base function neural network classifiers. Additionally, they also explored different voting schemes for multi class SVMs. Their obtain best performance of a SVM using a RBF kernel. The best voting scheme was found to be the one against one approach. Further comparison of different classifiers using multi-temporal SAR data can be found in Waske & Braun (2009). Based on the literature review we decided, to mainly utilize SVM. Mountrakis et al. (2010) give an overview regarding the use of SVM in remote sensing applications. Among many other they quote Waske & van der Linden (2008), which show post-classification data fusion of SAR and optical data using SVM, or Hu & Ban (2008) who successfully applied SVM on a segmented set from RADARSAT fine-beam images. The previously quoted paper of Ban et al. (2010) is also interesting in this context, since they apply a rule based scheme to merge two SVM classified data sets for improved accuracies of specific classes. Hu & Ban (2011) applied SVM and NN to Radarsat-2 data, finding that SVM in combination with a rule-based classification scheme is able to produce satisfactory classification results. Niu & Ban (2013b) could also show the distinct benefit of SVM classification as a post-processing of image segmentation which helped to further improve their image classification results from previous studies of pixel-based classification approaches. In the latter case SVM not always performed best, but based on objects it was found to be superior.

A good entry point in the actual use of SVMs is Hsu et al. (2010), giving a good and practical introduction to SVM and the LibSVM software package. A deeper and more general description and tutorial of SVM can be found in Burges (1998).

Finally Foody (2002) gives some important clues about how to perform and report classification accuracy assessment.

2.2 Remote Sensing for Urban Extent Mapping

2.2.1 Optical for Urban Extent Mapping

The urban extent has been extracted from a number of optical satellite missions. On a global scale the MODIS 500 (Schneider et al. (2010)) and glob cover (Defourny et al. (2006)) land cover map provide a coarse extraction level of about 500m per pixel from MODIS 500 and 300m per pixel from Glob Cover.

Recently with the advent of high resolution satellite imagery more and more studies have been undertaken using this much more detailed source of information for ubran extraction. Pesaresi et al. (2011) developed an algorithm that utilized texture features for urban extraction of 40 cities around the globe from Ikonos and Quickbird panchromatic images. The algorithm is based on rotation-invariant anisotropic textural analysis computed using GLCM features. A thresholding procedure is applied to distinguish urban from non-urban areas. It can be selected automatically or based on reference samples. On average 80% accuracy were achieved for the automatic unsupervised threshold and 87% using the supervised version. A main challenge with the global urban area extraction from optical images is the data availability. Due to the dependency on clear sky weather conditions in daylight, often suitable imagery with complete coverage of the globe is not available. Gong et al. (2013) presented recently the first global land cover map in 30m resolution. The map is based on a large number of landsat TM and ETM+ imagery. It has 2010 as the reference date, but due to large problems creating a complete cloud free map the imagery had to be taken from the years 2006-2011. Especially in the urban context, with its fast changing and growing pace, it is not a surprise that the achieved accuracy on impervious surfaces, where the urban areas are represented, falls between 6 -30% user accuracy depending on the classifier. This is not satisfactory and needs further improvement. That clouds is part of their classification, which would not usually count as land cover, highlights once more the major problem with optical imagery based classification approaches. Here a clear advantage in using SAR data can be seen, which is treated in the following section. A very good source for further information about global urabn mapping can be found from Gamba & Herold (2009).

2.2.2 SAR Data for Urban Extent Mapping

The much lower dependency on weather conditions and indepence of sunlight makes SAR imagery an optimal candidate for global urban area extraction. Gamba et al. (2009) compared ALOS PALSAR, TerraSAR X and Cosmo Skymed data for their capability of human settlement detection. They used their newly developed urban extractor and validated the results against two different global land cover models. They were finding that the two tested algorithms and data from 3 satellites are suitable for the task but with no outstanding individual candidate. An improved version of their algorithm including GLCM features is given by Gamba et al. (2011). In addition they tested the tool with ENVISAT ASAR and TerraSAR X data. Surprisingly the ASAR result was better with the applied methodology than the TerraSAR X result (around 80% vs. around 65%). The algorithm described is the foundation for the KTH-Pavia Urban Extractor. For global extraction Gamba and Lisini (2013) developed another method as a compromise between spatial accuracy and computational requirements. 75m ASAR wideswath data is globally compiled in time-series image stacks and processed to derive a 300m resolved global urban extent map. Multi-temporal SAR stacks are speckle filtered and the filtered images are searched for urban "seeds" characterized by very bright backscatter. A region growing step to delineate urban areas follows, slopes derived from gobal DEM help to correct for errors of mountainous regions. Achieving promissing accuracy the major problem here is the dependency on a large number of images (more than 10 at any given spot on earth). At DLR in Germany Esch et al. (2010 &2013) developed another processing chain to extract the urban extent from imagery of the TanDEM-X mission. The urban extent is mapped in 12m resolution. The Urban Footprint Processor used by DRL consist of three major steps, feature extraction, urban non-urban classification and post-editing and mosaicking. Features are extracted based on a speckle divergence. Using the speckle divergence and backscatter image a adaptive thresholding algorithm depending on image dynamics classifies the image. A logical or between results of different images serves for mosaicking purposes. Finally also here a DEM helps to identify misclassification due to high backscatter in mountainous regions.

3 Study Areas and Data Description

3.1 Study Areas

Due to the rapid urban development in China it makes an excellent case for testing our algorithms (Ban et al., 2011). For the classification of urban areas two main study areas have been selected in China, the megacities Beijing and Shanghai (Zhang et al.,2011). Both cities consist of a vast number of different types of buildings and numerous configurations of land cover and land use. There are commercial districts with skyscrapers, industrial areas consisting of large halls, harbor ports, large residential building blocks, small villas with gardens, satellite suburban towns and villages etc. The diversity and complexity of these cities makes them excellent study objects for the investigation of urban classification methodology.

For the urban extraction a larger number of cities was selected covering all continents but Antarctica. In total eleven urban areas were selected, situated in different landscapes and of different character. We included megacities such as Mexico City, Rio de Janeiro and Beijing, we included cities in a tropical environment such as Jakarta and Lagos, typical European cities such as Berlin and Stockholm, smaller towns south of the alps in the north of Italy in the Lombardy region, cities in dry environment such us Mumbai and Sydney and - last but not least - New York City, with the Skyscrapers of Manhattan and endless low density built-up areas along the Hudson river. See Figure 1 in Paper III.

3.2 Data Description

The main data source for this research was synthetic aperture RADAR (SAR) imagery. A large number of ENVISAT ASAR scenes (30m C-band imagery) was acquired, for the main study areas for urban classification as well as for the globally distributed study areas of the urban extension. For the comprenhensive studies of how to combine SAR stacks optimally a large variety of images was acquired. Data in different polarization, orbit orientation, incidence angle as well as time-series. For the data fusion experiments also images from the Chinese HJ-1B mission were acquired over Shanghai and Beijing. For the experiments with high resolution SAR data, TerraSAR-X data was ordered over Shanghai. A list of all images used can be seen in Table 1.

Table 1 Satellite Imagery used for this research

Location	Date	Sensor	Bands / Polarization	Incidence Angle	Orbit Orientation	Image Mode
Beijing	2009-05-07	ENVISAT ASAR	HH/VV	39.1-42.8	Ascending	APP
Beijing	2009-05-12	HJ-1B	NIR,R,G,B			
Beijing	2009-05-17	ENVISAT ASAR	HH/VV	19.2-26.7	Ascending	APP
Beijing	2009-05-27	ENVISAT ASAR	HH/VV	19.2-26.7	Descending	APP
Beijing	2009-06-08	ENVISAT ASAR	HH/VV	30.6-36.2	Ascending	APP
Beijing	2009-06-11	ENVISAT ASAR	HH/VV	39.1-42.8	Ascending	APP
Beijing	2009-08-04	ENVISAT ASAR	VV	42.5-45.2	Ascending	IMP
Beijing	2009-09-08	ENVISAT ASAR	VV	42.5-45.2	Ascending	IMP
Beijing	2009-10-13	ENVISAT ASAR	VV	42.5-45.2	Ascending	IMP
Shanghai	2009-05-11	ENVISAT ASAR	VV	19.2-26.7	Ascending	IMP
Shanghai	2009-05-22	HJ-1B	NIR,R,G,B			
Shanghai	2009-07-20	ENVISAT ASAR	VV	19.2-26.7	Ascending	IMP
Shanghai	2009-08-24	ENVISAT ASAR	VV	19.2-26.7	Ascending	IMP
Shanghai	2009-09-28	ENVISAT ASAR	VV	19.2-26.7	Ascending	IMP
Shanghai	2010-06-30	TerraSAR-X	HH	22.3-25.6	Ascending	Strip
Shanghai	2010-08-02	TerraSAR-X	HH	22.3-25.6	Ascending	Strip
Shanghai	2010-08-18	TerraSAR-X	HH	39.8-42.3	Ascending	Strip
Shanghai	2010-08-22	TerraSAR-X	HH	41.7-43.9	Descending	Strip
Shanghai	2010-08-22	TanDEM-X	HH	41.7-43.9	Descending	Strip
Shanghai	2010-09-04	TerraSAR-X	HH	22.3-25.6	Ascending	Strip
Berlin	2010-09-25	ENVISAT ASAR	VV	19.2-26.7	Descending	IMP
Jakarta	2009-09-06	ENVISAT ASAR	VV	19.2-26.7	Ascending	IMP
Lagos	2010-08-18	ENVISAT ASAR	VV	19.2-26.7	Descending	IMP
Lombardy	2010-07-20	ENVISAT ASAR	VV	19.2-26.7	Ascending	IMP
Lombardy	2010-08-24	ENVISAT ASAR	VV	19.2-26.7	Descending	IMP
Mexico City	2010-05-14	ENVISAT ASAR	VV	19.2-26.7	Descending	IMP
Mumbai	2010-06-27	ENVISAT ASAR	VV	19.2-26.7	Descending	IMP
New York City	2010-07-26	ENVISAT ASAR	VV	26.0-31.4	Ascending	IMP
Rio De Janeiro	2010-01-15	ENVISAT ASAR	VV	19.2-26.7	Ascending	IMP
Stockholm	2006-06-16	ENVISAT ASAR	VV	19.2-26.7	Descending	IMP
Sydney	2010-09-03	ENVISAT ASAR	VV	31.0-36.3	Descending	IMP

In addition to the data named in Table 1 elevation data from the SRTM

mission was acquired for terrain correction in the preprocessing of the Shanghai and Beijing urban classification studies as well as for the mountain masking for the urban extraction study.

4 Methodology

4.1 Data Preprocessing

Preprocessing is an essential part of every analysis in order to make sure that the data used is compatible and comparable with each other as well as optimized for the analysis purpose at hand. The most important step when using images from multiple sources is to make sure that they are geo-referenced or at least referenced relative to each other so that the content of two different sources is correctly aligned when overlayed with each other. This is necessary regardless of the data type, ie. SAR or optical data or both. Additionally, the effects of the terrain have to be taken into account orthorectifiying (Jensen, 2005) the image. For geocoding the orthoengine in Geomatica (PCI Geomatics, 2014) was used.

Especially in the case of SAR imagery there is always the question of how to treat the speckle (Henderson , 1998) generated from the way the radio signals are processed during the caption of a SAR image. Speckle is not always noise. Hence it has to be considered if one tries to correct for it by filtering the image or to use it as a source of information. Both has been done, during the research conducted here. For the object based urban classification speckle has been treated as noise and was filtered out; for the urban extraction the texture characteristics of the speckle have been preserved as a feature to distinguish between urban and non-urban areas. When correcting for speckle two different approaches were used depending on the number of images available. Less than three images covering the same area, an enhanced Lee-filter (Lee et al., 2009) was used to correct for speckle; if a larger number of images in sufficient temporal vicinity (within one year) covering the same areas was available instead a multi-temporal speckle filter was applied (Quegan et al., 2000). In contrast to classic speckle filters the multi-temporal filter makes use of the temporal stack averaging not only the area under the filter mask, but also over time. The filter was applied using the NEST software (ESA NEST, 2014).

Finally depending once more on the application a contrast enhancement was applied. On the hand improved contrast between different image objects should help to improve the quality of the segmentation result,

on the other hand the contrast is to be preserved for not changing the histogram characteristics for urban extraction.

4.2 KTH-SEG for Image Segmentation

Processing images for image classification is divided into a number of tasks working with KTH-SEG. Preprocessing has to be performed outside of KTH-SEG, with the exception of contrast enhancement which can also be done inside of KTH-SEG. A schematic overview of a classification task in KTH-SEG is given in Figure 1.

4.2.1 Edge Detection

Two methods for edge detection are implemented in the latest version of KTH-SEG. The first method described in Paper 2 applies a 5x5 Sobel-filter for gradient detection and then uses thresholding and logical operators to find the edges from a multi-image stack. This process marks image regions that are highly heterogeneous.

The second process is an enhanced canny edge detector for the use on multi image stacks. It is based on a 3x3 Sobel-filter for gradient detection. The gradient images are then processed according to the scheme suggested by Canny (1986). The gap-filling procedure that is using the two binarized channels derived by thresholding on the gradient images is improved looking not only at the two binary images from one date but also looking at those from other scenes in the image stack. In this way one edge layer representing the whole stack is derived. This is particularly beneficial when using images of completely different kind like optical and SAR imagery, which are very likely to highlight and find different edges.

The derived edge layer is added as additional information to the region growing and merging algorithm that is responsible for segmenting the image stack into a set of homogeneous objects.

4.2.2 Region Growing and Merging

The goal of the region growing and merging algorithm developed at KTH is to segment an image composed of an arbitrary number of bands

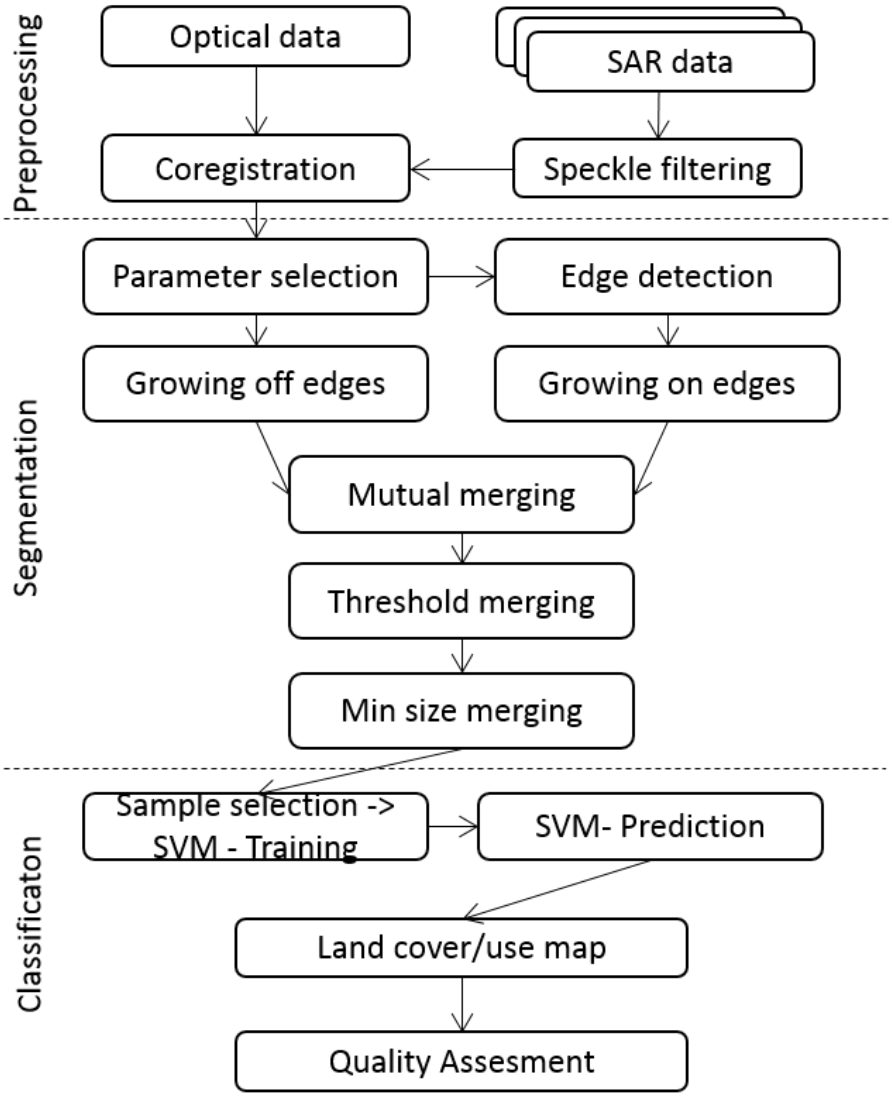

Figure 1 Processing phases of KTH-SEG

into objects. Note that the derived algorithms is universally applicable to any kind of raster data.

An object is understood as a homogeneous area who's outline is defined by a polygon of arbitrary shape, but not self-intersecting. An object is also supposed to approximate real-world objects detected in the image

scene as good as possible. Those real world objects in the context of urban mapping could e.g., be houses, roads, rivers, parks, agricultural fields, etc. How detailed these objects are depends mainly on the quality of the input imagery, i.e. its spectral and spatial resolution.

The algorithm is divided into a series of processing steps (see Figure 1). Segmentation starts at the pixel level. Initially every individual pixel is considered to be a segment of dimensions equal to the extent of the given pixel. The first step is called the growing phase, which allows adjacent pixels to grow into larger segments if they are homogeneous. In KTH-SEG homogeneity is defined by a weighted sum of change in mean and change in standard deviation, with respect to the situation before and after a growing or merging step (for details see paper II and Jacob (2011)).

Growing is limited by a size parameter. For the growing phase this parameter is called minimum segment size. This is also the smallest mapping unit in the resulting classification, since no object smaller than this size will be extracted. The growing phase also takes into accounts edges. Growing is performed separately on pixels marked as part of an edge and pixels not part of an edge. This basically divides the segmentation into segmentation over homogeneous areas and segmentation over heterogeneous areas, since edges are areas of high local change. This separation makes sure that the statistics of especially homogeneous areas are more stable and reduces the risk of undesired merges in later stages, but without introducing computationally costly comparisons of edges on a vector level. The growing phase is finished when all segments have reached minimum segment size, or as close to it as possible. Patches of edge pixels smaller than the minimum segment size could exist. In that case growing will stop when all pixels of this edge patch are part of one segment, even if it is smaller than the minimum segment size.

From now on segments are not growing any more, but are merged with each other, that means all neighboring segments are checked for potential candidates to merge with. Even this phase is split into two different merging techniques. Firstly, only segments that are each other's best matches with respect to the previously mentioned change in homogeneity are merged. Then in a second step a threshold is defined to also merge segments that are not each other's best matches. This

threshold is derived automatically from the global average of difference in homogeneity between all neighboring segments (details are explained in paper II). Empirically a threshold of one third of this global measure has been set in the program. It has been tested on a large number of different images even outside the research presented in the papers of this thesis, including optical and SAR imagery, e.g. Landsat 4,5,7 &8, Spot 4&5, Geoeye, Quickbird, ERS 1&2, ENVISAT ASAR, Cosmo Skymed, TerraSAR-X and Radarsat-2. Only the sensor data used in the papers compiling this thesis has been tested using ground truth data. The remaining missions have only been inspected visually.

Merging is generally limited by the maximum segment size. In the merging phase both edge based and non-edge based segments are allowed to merge. It can be assumed, that the edge which usually has to belong to one of if its neighbors will have the smallest change of homogeneity with respect to exactly that member or a continuous part of the same edge. Only edges of objects having a width comparable to the pixel resolution such as roads in mid resolution imagery, where the width of the road typically is one pixel or even less, don't follow that assumption.

The final step of the segmentation is then to assign all those segments smaller than the minimum segment size to its best neighbor with no constraints on the homogeneity other than assigning it to the one that generates the smallest change of the merging criteria.

For the processing of the ENVISAT ASAR data the parameters within KTH-SEG are set to be 40 and 4000, for minimum and maximum segment size and the weights between mean and standard deviation where set equally to 0.5. These parameters are chosen not to find very small objects. As we did not expect to find individual houses from 30m resolved data and hence one can expect to obtian clusters of buildings in one object yielding the definition of high and low density built-up as major urban classes.

For the processing of TerraSAR-X imagery the parameters were set differently due to the much higher resolution. 16 to 4000 for the size parameters and equal weights for the homogeneity parameters were found to be reasonable. The smaller segment size in combination with

the higher resolution makes the identification of individual buildings possible.

4.3 Urban Land Cover Classification using Support Vector Machines

Once the whole image stack has been segmented into objects, the nature of the objects has to be determined, since it is only known that those objects found are different from its neighbors. In the current version of KTH-SEG it is done using a supervised machine learning classification approach, the support vector machine (SVM). The SVM can be trained either interactively by the operator or class definitions in the form of labeled polygons from an external source can be imported. In the second case, objects generated from the above described algorithm are superimposed with those imported training objects and the label of the segment that has the highest overlap from the generated objects gets the label of the imported training object assigned.

The SVM is based on a radial base function (RBF) kernel, which requires two parameters often denoted as C and γ, which control the margin width and complexity of the decision boundary of the SVM. The optimal solution for those two parameters is determined performing a grid-search which tests all combinations of those two values within a user provided range. Each parameter set is evaluated by a folding procedure where a subset of the training objects is used to predict the remaining training objects. The process is five folded meaning that 5 times 20% are used to predict the other 80% of the training set. The overall accuracy of the five folds is the score used for finding the most effective parameters. The grid search is repeated twice - first on a coarse grid and then on a finer grid. Around the best value of the first coarse grid a fine grid is defined and then searched. The optimal values from both searches are taken for the prediction of all objects generated from the segmentation using all training samples provided. The values that go into the SVM as features are the mean and standard deviation for each band, i.e., an image stack with 4 bands will have 8 features that are used for the classification. Additional features that are available but not currently used are geometric properties of the object, the area and perimeter.

4.4 KTH-Pavia Urban Extraction

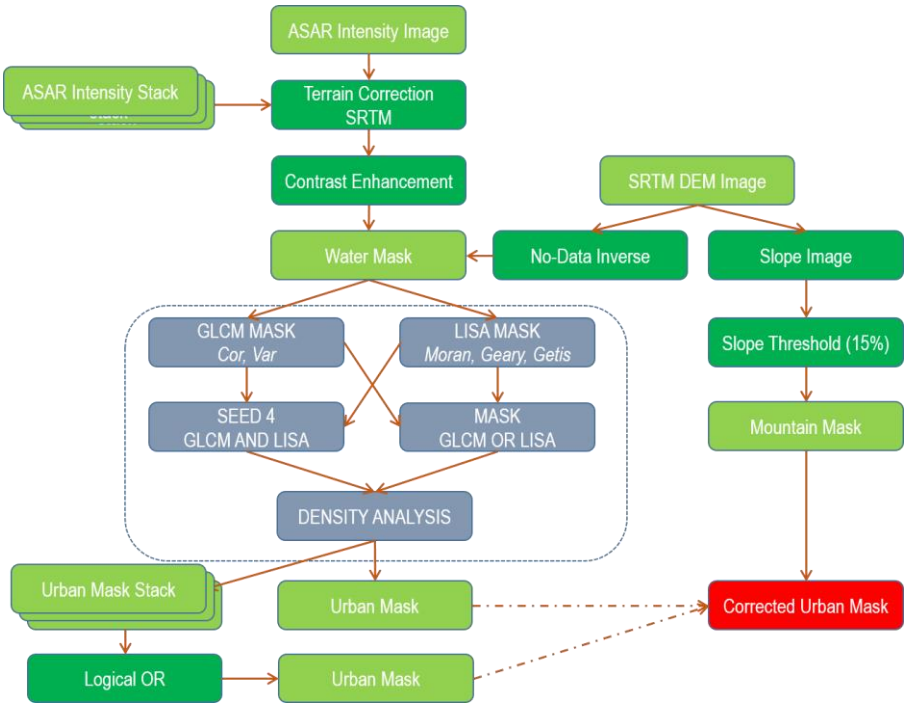

Figure 2 KTH-PAVIA Urban Extractor – processing overview (Blue is the original processing, Green are new processing steps, Red marks the final result)

This second algorithm was originally designed by researchers of the University of Pavia (Stasolla, Aldrighi, and Gamba) and has been improved in a joint research project. The algorithm is designed to extract urban areas from SAR imagery using textural features as well as geospatial statistical indices. In contrast to the previously described algorithm this one utilizes the presence of speckle as a feature and hence images are not speckle-filtered prior to processing. The algorithm prior to our improvement is best described in Gamba et al. (2011). In summary, the geospatial statistics based on Moran's I, Geary's C and Getty's are used to select first those areas with extreme high or low backscatter and little speckle, which are assumed to most likely be the urban areas, using a so called density analysis. In a second step the grey-

level-co-occurrence matrix (GLCM) features variance and covariance are calculated and once again density analysis is used to extract the urban areas. Both results are connected via a logical AND-operation.

This version of the algorithm has some known issues. E.g. mountainous regions can be misclassified as urban areas due to their very high backscatter. When the contrast in the image is too low however, urban areas are often omitted. To tackle these two main known issues the preprocessing of the data used for urban extraction was changed. An overview of the updated processing scheme is given in Figure 2. For the first the SAR images were corrected for terrain effects using the Range-Doppler algorithm with Shuttle Radar Topography Mission (SRTM) data. To better adapt to the typical histogram of a SAR image, which tends to be skewed to the left with a long tail towards the high values, image scaling was performed. The original procedure was based on a 1% clipping on each end of the histogram. This was changed to 1% on the left and 4% on the right side when converting the data from 16 to 8 bit. To maintain the characteristics of the original data as much as possible and to not lose texture features, a linear scaling function was used. The original algorithm was limited to work with one image band only. To utilize the information that e.g. can be found from polarimetric imagery of two or more bands the possibility to merge several bands was implemented and tested. From the experience with the original version it was known that the algorithm tends to omit rather than to commit with too much urban areas, with the known exception of the mountains. This behavior made the data merge easier because it can generally be assumed that adding more images together is much more likely to improve the results than to introduce new erros. The merge is based on a logical OR connection of the binary classified urban - non-urban layers. Mountainous areas where afterwards taken care of by applying a mask in dependence of the slope. Areas that are steeper than a given threshold will be filtered out. For this dataset the slopes where derived from the SRTM digital elevation model (DEM).

4.5 Accuracy Assessment

4.5.1 Urban Land Cover Classification

Accuracy assessment is considered an integral part of every analysis performed and all classification results in this thesis are evaluated using

independent testing areas for evaluation of the classification. Independent means that areas selected for training and areas selected for testing do not overlap.

Test samples are selected in patches of small polygons and are distributed over the whole scene. They are further selected in a way to cover all aspects of a class, so that all kinds of shortcomings in detection can be found. E.g. in the case of buildings there are areas, that represent small villas, skyscrapers, industrial halls, large residential complexes etc. In this case around 5000 pixels where labeled per class using ground truth information from google earth imagery.

Another applied strategy was to randomly assign pixels with classification labels. In this case around 2000 pixels where labeled per class.The random assignment is much less observer dependent since here the observer only needs to decide the class in a specific location, and not the location itself.

Classification results and test samples are then processed together to generate confusion matrices. Those matrices represent the number of pixels that are classified correctly as defined from the test dataset or if not classified correctly with which class they were confused. From these matrices, widely used accuracy measures such as the overall accuracy and the kappa value can be derived, giving a quality benchmark in one number for the whole classification result. Especially the kappa value is of great interest since it also takes the inter-class confusion, i.e. the off diagonal elements into account.

For further investigation of the quality of the results similar processing has also been applied using Trimble´s eCognition for image segmentation. For this comparison the segmentation was performed in eCognition according the following procedure. To reduce the effect of differences in scale several sets of objects were created in eCognition to find a similar number of objects as KTH-SEG. Results were created with scale levels 40, 25 & 20 in eCognition. The weights in eCognition where set to 0.9 favoring color and 0.5 for compactness, meaning that spectral information has a 90% weight and the geometry of an object only 10%. Compactness and fragmentation were set to have equal influence on the shape of the final objects. The objects obtained from eCognition have then been exported to a shape file together with the

mean and standard deviation of every individual band. In this way it was made sure that the same information as provided by KTH-SEG is available when doing classification. To make sure that any deviation of the classification results from KTH-SEG and eCognition can only be derived from differences in the segmentation, we performed the classification inside KTH-SEG. In that way both sets of objects, the ones from eCognition and the ones from KTH-SEG, are processed in exactly the same way apart from the segmentation itself.

4.5.2 Urban Extent Extraction

The accuracy assessment for the urban extent was performed similar to the accuracy assessment for the urban land cover classification. Here small patches of around 50-100 pixels were labeled. Around 100-150 polygons where selected per class and image and it was made sure that at least 10000 pixels per class where labeled in the end. Also for the urban extent mapping ground truth was derived from google earth. Here a random assignment was just too time consuming due to the large number of study sites. In the previous chapter the testing set had to be defined only once and could then be used for all comparative studies over the same area. In this case however, a lot of different study areas, eleven to be accurate, had to be labeled.

Since there are only two classes to compare to each other and the main focus is on urban areas, in addition to overall accuracy and kappa the omission and commission errors were compared. The latter two accuracy measures describe two different kind of errors. The omission error measures those cases where an area that in reality is urban was classified as non-urban and the commission error measures those cases where a true non-urban area was classified as urban. This information is very useful to understand if the results over- or underestimate the quantity of urban areas.

5 Results and Discussion

The algorithms described in the previous chapter have been applied on a number of study sites using a large number of images.The following discussion compares and analyzes the results from the individual papers in a wider scope. The results from paper I & II are interesting to compare due to the fact that a similar methodology was applied to two different study sites. Paper I & IV are looking at the same study area but fundamentally different data. Paper III might on first sight look a little bit off with respect to the others but there is great potential for future integration which will be discussed in more detail in the next chapter.

5.1 Multi-temporal SAR for Urban Land Cover Mapping

When looking at different data stack compositions of the SAR data it is evident that not all combinations yield equally good results. One aspect is the comparison of time series stacks and stacks composed of imagery with different observations characteristics in terms of orbit orientation (ascending or descending track), viewing angle or inclination and also the question of the SAR band selected (in this thesis C and X-band). When looking at the time-series present in the ENVISAT ASAR data

Table 2 Image stack comparison. Envisat ASAR, Shanghai

HJ-1B CCD2 20090522	ASAR IS2 200905011	ASAR IS2 20090720	ASAR IS2 20090824	ASAR IS2 20090928	Overall Accuracy	Kappa Value
x					66,06	0,600
	x	x	x	x	57,67	0,504
x	x				**79,66**	**0,761**
x		x			76,95	0,731
x			x		79,08	0,756
x				x	72,13	0,674
x		x	x		73,21	0,687
x	x			x	74,68	0,704
x	x	x	x	x	76,47	0,724

Table 3 SAR stack composition comparison. Envisat ASAR, Beijing

IS 2 A	IS 2 D	IS 4	IS 6	IS 7	IS 7	Over. Acc.	Kappa
x	x					47,29	0,41
	x		x			48,32	0,42
	x			x	x	50,55	0,44
		x		x	x	53,40	0,48
		x	x			53,90	0,48
x	x		x	x		54,47	0,49
			x	x	x	53,08	0,47
	x	x				54,56	0,49
x		x	x	x	x	54,43	0,49
x	x		x	x	x	56,85	0,52
	x	x	x	x	x	56,86	0,52
	x		x	x	x	56,52	0,51
	x	x		x	x	58,96	0,54
x	x	x	x	x	x	65,14	0,61

over Shanghai, the results show actually not much difference with regard to the composition of the stack. Table 2 shows the results were not satisfying in the data stack with the largest temporal difference between the optical image and the SAR scene (from late September). What was interesting to see here, was that one image close in time performed just as well as all images together or even better. A similar observation could also be made in the Beijing case where a subset of 4 images performed almost as well as all 8 images taken together.

From the Beijing study it is evident that the more different ASAR scenes the stack contained the better was also the accuracy. For urban studies higher incidences angles (i.e. IS4-IS7) are favorable as can be seen from Table 3. When using the high resolution TerraSAR-X data different image stacks performed differently. Contrary to the ENVISAT ASAR results over Shanghai the time-series results performed better than an image stack composed from a variety of different types of images. Figure 3 shows how the data is represented in ENVISAT ASAR and TerraSAR-X. The two RGB composites containing images from ascending and descending orbit as well as from high and low inclination angles. Buildings or other objects of higher elevation are

Figure 3 Time series overlays. Top: ENVISAT ASAR, Bottom: TerraSAR-X

depicted very different in the different channels of the image composition of TerraSAR-X. The cyan and red parallel lines are the backscatter resulting from the same buildings but in different location due to the different orbit properties and viewing angles. In the

Table 4 Comparison of Urban and Vegetation Classification for Envisat ASAR

	Average Shanghai	Optical	SAR	ALL	Best Merge
Vegetation	57,75	52,49	33,69	66,01	70,70
Urban	69,10	56,82	54,95	72,95	73,69
Low backscatter	64,95	58,05	45,81	72,74	72,72

	Average Beijing	Optical	SAR	ALL	4-Date
Vegetation	66,53	63,37	63,20	68,49	71,07
Urban	76,63	68,06	73,03	84,90	80,54
Low backscatter	73,00	75,60	57,74	80,22	78,47

ENVISAT ASAR images – resolution is one order of magnitude lower - this effect is not that pronounced since the pixel size is large with respect to the object size.

An increase in quality of the classification with the number of images used within one data stack could also be observed from the given studies, the dependence is however not as clear as in the previously named examples. In the TerraSAR-X analysis (paper III) this also seems to be the case when looking at the results with larger segments, but at smaller segment size there is no clear connection. In the classification of SAR data over Beijing we can see a similar trend. In the case of Shanghai using ENVISAT ASAR, the best results were actually coming from fusion of one SAR and one optical image.

Another interesting comparison is the performance of urban classification versus vegetation classification. In all cases urban classes (high density built-up, low density built-up, roads and runways) were mapped with a higher accuracy than vegetation classes (low vegetation, forest, golf course and grass/pasture). See Table 4.

One thing that became evident during the processing of the TerraSAR-X imagery was that the exact same classification scheme as applied for processing the ENVISAT ASAR data would not hold. Instead of searching for high and low density built-up which has a distinct

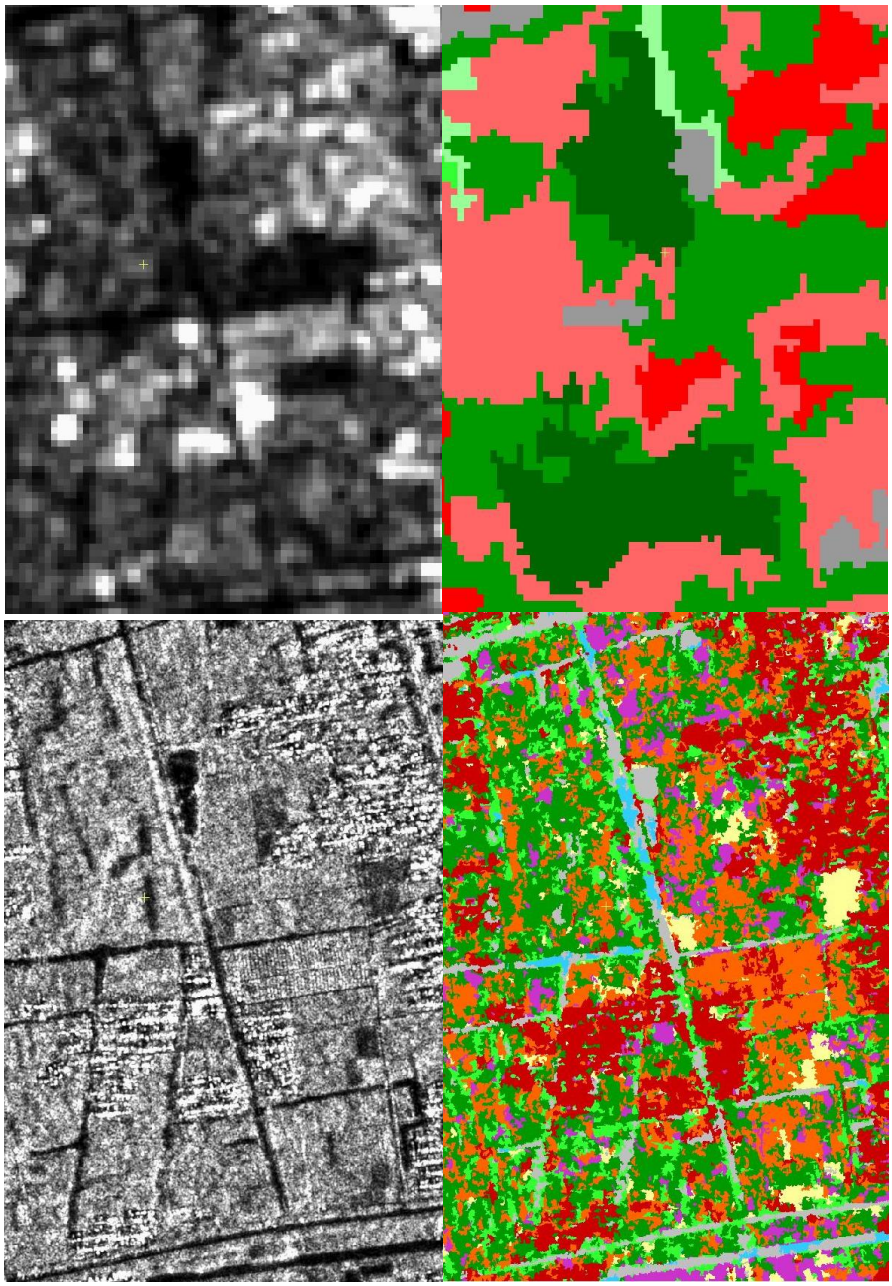

Figure 4 Low Density Built-Up Comparison, Shanghai. Top: ENVISAT ASAR, Bottom: TerraSAR-X

signature when looking at the ENVISAT ASAR, only one class was used here (building). The reason is that the high resolution of TerraSAR-X, 3m in the case of the strip map date we used, does not create the same sort of mixed pixels, as was the case in the 30m data. Instead a low density built-up area is characterized by a topological

relationship of building objects and adjacent vegetation objects (see Figure 4). In order to get to the same kind of classification scheme as applied before another layer of processing needs to be introduced, investigating the topological relationship between classified segments, in order to achieve a higher order of classification. This has yet to be done. Additionally, airport runways and roads were kept in one class, due to the absence of optical data. When looking only on backscatter the difference between these classes is so small that a meaningful differentiation is unreasonable without introducing other errors.

Overall, the examples show that TerraSAR-X data can provide for accurate urban classification. The eight classes that were sought from up to six TerraSAR-X and one TanDEM-X image were classified with 84% overall accuracy without using any optical data. When using a time-series of only three images a classification accuracy of up to 77% was achieved using KTH-SEG.

5.2 Data Fusion of SAR and Optical Data

The results compiled from ENVISAT ASAR and HJ-1B data over Beijing and Shanghai, show a strong synergy effect when merging data from these two sensors. When processing SAR data alone we achieved accuracies of only about 58% in the Shanghai case and 66% for Beijing. The difference can be explained by the fact that the imagery over Beijing was mainly composed of dual polarized (HH/VV) SAR imagery, whereas the Shanghai area consisted only of single polarized imagery (VV). From the confusion matrices of these processing runs (Table 5) it is evident that the problem lies in the detection of vegetation. Low density built-up is confused with high density buildup, bare fields and forest. There is a difference in representation of vegetated areas between HH and VV polarization. This contrasts with

urban areas, which are more homogeneously represented in these two polarizations.

Table 5 Confusion of Envisat ASAR processing. Top: Shanghai, Bottom: Beijing

Name	ID	1	2	3	4	5	6	7	8	9	10	User
High density	**1**	83,1	10,1	0	1,3	4,7	0	0	0	0,8	0	**83,1**
Low density	**2**	43,4	36,6	0,1	6,4	11,2	0,6	0,1	0	0,3	1,2	**36,6**
Bare field	**3**	0	25,2	32,4	17,4	2,4	0,7	2,6	12,5	0	6,9	**32,4**
Low veg.	**4**	7,3	5,5	10	40,7	8,7	3,7	17,3	6,9	0	0	**40,7**
Forest	**5**	5,9	22,2	1,3	16,8	46,1	1,2	0	0	0	6,4	**46,1**
Golf course	**6**	0	3,7	29,8	1,6	0	18,3	6,5	1,9	26,5	11,8	**18,3**
Pasture	**7**	0,1	1,3	20,6	3,6	3,3	20,2	29,7	8,7	3,6	8,9	**29,7**
Water	**8**	0,5	2,8	11,4	1,4	0,7	0,7	0,4	78,5	3,2	0,5	**78,4**
Airport	**9**	0	3,9	5,2	0,7	0	3,9	0,1	3,8	79,6	2,8	**79,6**
Major Roads	**10**	6	20,1	19,5	11,1	2	8,3	6	6,1	0,5	20,4	**20,4**
Producer's Acc.		**75,8**	**35,8**	**20,5**	**36,2**	**42,2**	**29,6**	**47,3**	**87,2**	**65,1**	**32,5**	

Name	ID	1	2	3	4	5	6	7	8	9	10	User
High density	**1**	85,2	14,5	0	0	0	0	0	0	0	0,3	**85,2**
Low density	**2**	23,3	72,6	0	0	0,8	0	0	0	0	3,3	**72,6**
Bare field	**3**	0,1	0	47,2	7	0	5,7	28,1	0	2,4	9,4	**47,2**
Low veg.	**4**	0	0,1	11,9	38,7	6,7	4,5	37,6	0	0	0,6	**38,6**
Forest	**5**	0,2	4,3	0,1	10,2	79,9	0	4,8	0	0	0,5	**79,9**
Golf course	**6**	0	0	10,2	1,8	0	64,5	6,7	7,5	5,1	4,2	**64,5**
Pasture	**7**	0	1	9,4	8,6	0,6	2,3	69,8	0	1,4	7	**69,7**
Water	**8**	0	0	0,9	0	0	10,6	0	42,6	41,6	4,3	**42,6**
Airport	**9**	3,1	1,3	1	1,8	0,1	0	2,7	7,5	75	7,4	**75,1**
Major Roads	**10**	7	10,3	0,8	2	4,2	6,6	0,2	9,6	0	59,2	**59,3**
Producer's Acc.		**78,7**	**71,7**	**58,7**	**54,5**	**86,5**	**70,2**	**45,4**	**64,1**	**58,4**	**58,7**	

The classification results of the HJ-1B imagery for Beijing and Shanghai are in better agreement with each other, having only a minor difference of 2% in overall accuracy (68% for Beijing and 66% for Shanghai). When merging SAR and optical data, the accuracy for both

classifications rose by at least 10%. Especially low density built-up, bare field and roads were classified better when adding SAR to the optical images. From a SAR perspective the confusion within low-backscatter classes such as water roads, runways and golf courses with their extremely fine-trimmed grass could be reduced by adding the optical data into the processing. Overall, the confusion matrix got a lot more balanced after merging the two datasets with much lower differences in between the individual matrices and hence creating a better overall reliability of the resulting classification map.

5.3 KTH-SEG in Comparison with eCognition

The results of the processing using KTH-SEG have been compared to results produced by eCognition applying a similar processing scheme. In the example over the new Olympic park in Beijing (Figure 5), some of the differences can be seen. Firstly, the large segments in vegetated park areas and that the shape of the segments are much simpler as those produced by KTH-SEG, due to fact that compactness and fragmentation are considered while shaping the objects. At a first glance the objects are visually appealing, since they are easy to recognize and understand. On a closer look, it becomes evident that a large number of errors is introduced into the objects. If one is interested in nice object delineation this might be a desirable feature, but when doing fully- or semi-automatic classification in a subsequent step, in our experience this seems to create problems. The errors due to the shape of the object affect the statistical properties of the segments in terms of mean and standard deviation. If too many pixels that should not belong to a specific object are merged into that object, this changes the mean value of that object and hence makes it less likely to be classified correctly in the next step. In the top-left part of Figure 5 we can see large polygons in the north east of the park. This means that some of the water bodies have been merged with the grass or park surfaces. That leads to a lowered mean value and finally a misclassification into road. In these rather homogeneous areas, it seems that eCognition is not able to detect subtle changes.

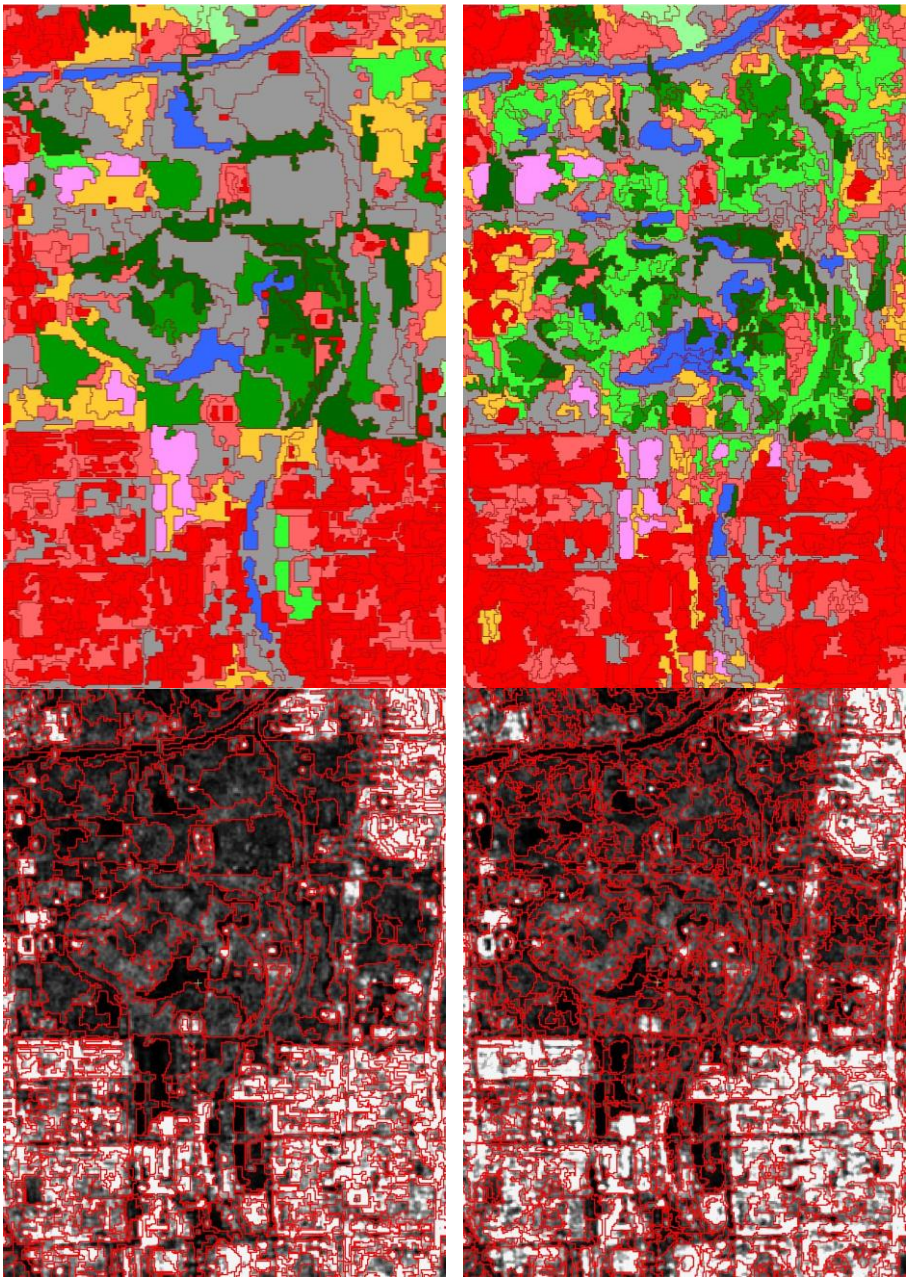

Figure 5 Segmentor Comparison, Olympic Park, Beijing. Left eCognition, Right: KTH-SEG, top: classification, bottom: SAR overlayed with segments.

In the south east of the bottom-left part of Figure 5, a high density built-up area, one can see that the average segment size generated from eCognition decreased drastically in this highly heterogeneous

environment. The rather rectangular shape of the objects is visible. In the case of the buildings this is not creating much of a problem since they tend to have more rectangular shapes. The region seems to be a bit over-segmented and it is not clear sometimes why certain segments, that obviously belong to the same object, have not been merged. KTH-SEG produces larger segments on average, collecting larger building complexes into one object. In terms of roads KTH-SEG and eCognition perform approximately equally at the highway in the south-west and the major roads south of the park. The roads directly adjacent to the park where better defined by KTH-SEG.

5.4 Urban Extent Mapping

In addition to urban area classification, identification of urban areas as a whole was performed. The KTH-PAVIA urban extractor was tested on eleven test sites globally distributed. These test sites were also evaluated using the MODIS 500 and GLOB COVER urban areas. For the evaluation independent training data, as described in section 4.5, was used. On average our method achieved 85% overall accuracy in contrast to 76% accuracy from MODIS 500 and 73% accuracy from GLOB COVER, see also Table 6. We obtained approximately 24% omission and roughly 5% commission error. This means our method is generally underestimating the urban area, hence having a more conservative estimate of the urban area. It is interesting to look at the standard deviation for those eleven test sites with respect to the overall accuracy. The KTH-PAVIA urban extractor has a deviation of 4%. MODIS and GLOB-COVER more than 10%. This says something about the reliability and constancy of our results when applied to a large number of cities with completely different character in terms of surrounding landscape but also in the design of the cites. For an even more robust extraction of the urban areas there are still some issues to be fixed however. We could for example see some detection problems

Table 6 Average performance comparison of all eleven test sites

Average values	Kappa	Overall Accuracy	Std. Dev.	Comission	Omission
KTH - UNIPV	0,707	85,36%	4%	5,47	23,75
GlobCover	0,471	72,67%	13%	17,10	40,47
Modis 500	0,525	76,31%	11%	20,03	31,12

in low density buitlup areas in the west of Rio de Janiero and along the Hudson River north of New York City. Other improvements like the mountain masks could improve the results by up to 5% and introducing multiple images over the same area brought improvements of up to 10% in the Beijing test area (details in paper III).

6 Conclusion & Future Research

6.1 Conclusions

This licentiate thesis combines the societal need for understanding the patterns and consequences of urbanization with the scientific need for more advanced and refined methods to analyze this globally present phenomenon. While working on this licentiate thesis, methods for locating the urban areas as well as analyzing the internal configuration of those areas, have been developed, improved and tested.

KTH-SEG, a new algorithm for image segmentation, has been developed, implement and successfully been applied to SAR & optical imagery for urban land cover mapping in Beijing and Shanghai. Compared with eCognition, the most commonly used segmentation software, the results obtained with our algorithm can be considered as slightly better than those obtained with eCognition, in particular when classifying linear features.

SAR data has been found to have very good properties for the analysis of urban areas. The combination of both, mid resolution optical and SAR data for urban land cover mapping has been investigated by merging data from both ENVISAT ASAR and the Chinese HJ-1B mission and found to be very useful. A diverse SAR data stack has been found to be especially beneficial for urban land cover classification. With higher resolution SAR imagery such as the TerraSAR-X stripmap imagery even SAR data alone without additional optical data we could achieve promising urban land cover mapping results over Shanghai.

With the improved version of the KTH-Pavia Urban Extractor we could further show, that the extraction of urban areas using ENVISAT ASAR data alone is giving good results. This gives an interesting outlook of what can be achieved, once the Sentinel-1 mission is operational.

6.2 Future Research

The research performed and described in this thesis is by no means arrived at its end and many possibilities for future study exist. The most important improvement for KTH-SEG, that is planned, is the

integration of multi-resolution segmentation and classification in order to improve the detection and classification of objects of different scales.

High on the list is the integrated use of the two tools for urban mapping. Using the KTH-Pavia Urban Extractor for finding the urban areas and classify the interior of the urban areas with KTH-SEG seems very reasonable. In this way the amount of data can be reduced to those areas that are of eminent interest and improve performance. First steps towards this integration have already been undertaken by enabling KTH-SEG to operate under a masked region only.

Further, with the increase in data size due to the higher resolution, more advanced implementation towards parallel operation are desirable in order to keep processing time on a reasonable level. In that domain first parts of the algorithm have already been parallelized. Ideas go in two directions cloud & cluster computation for very large datasets and heavy parallelization, processed on the local graphic card, using technologies like openCL and CUDA for single high resolution image processing.

A lot more test with different datasets could be of interest, e.g. data fusion of high resolution SAR & optical data like Geoeye, Quickbird, SPOT-6, TerraSAR-X, and Radarsat-2.

We are already investigating SAR-based change detection for urban environments at our institute (Ban & Yousif, 2012, Yousif & Ban, 2013 and Hu and Ban, 2014). In form of a bachelor thesis (Bergsjö, 2014) recently we started looking into the applicability of KTH-SEG on change-detection matters using optical imagery. So object based change detection is definitely another field of research that could be further explored using and extending the algorithms presented in this thesis.

The KTH-Pavia Urban Extractor also still has potential for further improvements. Including water masks is one of the possible extensions. Another is to solve the issues with low density builtup areas. Here it might be beneficial to use multi polarized imagery as well as from different orbit orientation. Additionally speckle filtering could be considered to get a better contrast to purely vegetated areas.

References

Alberga, V., 2007, A Study of Land Cover Classification Using Polarimetric SAR Parameters, International Journal of Remote Sensing, Vol. 28, No.17, pp. 3851 – 3870

Baatz, M., Schäpe A., 2000, Multiresolution segmentation – an optimization approach for high quality multi-scale image segmentation, Angewandte Geographische Informationsverarbeitung, Wichmann-Verlag, Heidelberg, 2000, pp. 12-23

Ban, Y, 2003, Synergy of multitemporal ERS-1 SAR and Landsat TM data for classification of agricultural crops, Canadian Journal of Remote Sensing, 2003, 29, pp 518-526

Ban, Y., Hu, H. and Rangel, I. M. , 2010, Fusion of Quickbird MS and RADARSAT SAR data for urban land-cover mapping: object-based and knowledge-based approach, International Journal of Remote Sensing, Vol. 31, No.6, pp. 1391 – 1410

Ban, Y., P. Gamba, P. Gong and P. Du. 2011. Satellite Monitoring of Urbanization in China for Sustainable Development, The Dragon 'Urbanization' Project. Earthzine.

Ban, Y, Yousif, O. A., 2011, Unsupervised Change Detection Using Multitmeporal Spaceborne SAR Data: A Case Study in Beijing, JUSRE 2011 - Joint Urban Remote Sensing Event, Munich, Germany, April 2011

Ban, Y. and O. A. Yousif, 2012. Multitemporal Spaceborne SAR Data for Urban Change Detection in China. IEEE Journal on Selected Topics in Applied Earth Observations and Remote Sensing, 5(4): 1087-1094.

Ban, Y., Yousif, O., Hu, H., 2014, Fusion of SAR and optical Data for Urban Land Cover Mapping and Change Detection, Global Urban Monitoring and Assessment through Earth Observation. Ed. Q. Weng. Taylor and Francis Group, LLC, In press.

Barbieri, Andre L., et al, 2010, An entropy-based approach to automatic image segmentation of satellite images. Physica A 390, pp. 512-518.

Benz, U., et al., 2004, Multi-resolution, object oriented fuzzy analysis of remote sensing data for GIS-ready information, ISPRS Journal of Photogrammetry and Remote Sensing, 2004, 58, pp. 239-258

Bergsjö, J., 2014, Object-based change detection in urban areas using KTH-SEG, Bachelor of Science Thesis in Geoinformatics, TRITA-GIT EX 14-xxx, Royal Institute of Technology (KTH), Stockholm Sweden

BerkEnviro (2011), Berkely Image Segmentation, available at: http://berkenviro.com/berkeleyimgseg/ . Accessed at 20 September 2010.

Bhaskaran, S., et al., 2010, Per-pixel and object oriented classification methods for mapping urban features using Ikonos satellite data, Applied Geography 30 (2010), pp. 650-665

Blaschke, T., 2010, Object based image analysis for remote sensing, ISPRS Journal of Photogrammetry and Remote Sensing 65, 2010, pp. 2-16

Blaschke, T., et al., 2005, Remote Sensing Image Analysis: Including the Spatial Domain, Kluwer Academic Publishers, 2005, pp. 211-236

Bontemps, S., et al., 2008, An object-based change detection method accounting for temporal dependencies in time series with medium to coarse spatial resolution, Remote Sensing of the Environment, 112, 2008, pp.3181-3191

Burges, C., 1998, A Tutorial on Support Vector Machines for Pattern Recognition. Kluwer Academic Publishers, Boston. pp. 1-43.

Burini, A., Putignano, C., Del Frate, F., Licciardi, G., Pratola, C., Schiavon, G. and Solimini, D., 2008, TerraSAR-X/SPOT-5 Fused Images for Supervised Land Cover Classification, Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International

Canadian Space Agency, 2013, RADARSAT Constellation, last accessed May 23, 2014: http://www.asc-csa.gc.ca/eng/satellites/radarsat/

Carvalho, E.A. et al., 2009, SAR imagery segmentation by statistical region growing and hierarchical merging. Digital Signal Processing 20, pp. 1365-1378.

Chen, H.-M., Akora, M. K. and Varshney, P. K., 2003, Mutual Information-based Image Registration for Remote Sensing Data, International Journal of Remote Sensing, Vol. 24, No. 18, pp. 3701-3706

Clinton, N., et al., 2010, Accuracy Assessment Measures for Object-based Image Segmentation Goodness, Photogrammetric Engineering and Remote Sensing, March 2010, pp. 289 – 299

Corbane, C., et al., 2008, Rapid Urban Mapping using SAR/Optical Imagery Synergy, Sensors 2008, 8, pp. 7125-7143; DOI: 10.3390/s8117125

Deforny, P., Vancutsem, C., Bicheron, P., Brockmann, C., Nino, F., Schouten, L., Leroy, M., 2006, Globcover: A 300M Global Land Cover Product For 2005 Using ENVISAT Meris Time Series, ISPRS Commission VII Mid-Term Symposium: Remote Sensing: from Pixels to Processes, Enschede (NL), 8-11 May 2006

Dell'Acqua, F. and Gamba, P., 2006, Discriminating urban environments using multiscale texture and multiple SAR images, International Journal of Remote Sensing, Vol. 27, No. 18, pp. 3797 – 3812

Desclée, B., et al., 2006, Forest change detection by statistical object-based method, Remote Sensing of Environment, 102, 2006, pp.1 – 11

Engdal, M. E. and Hyyppä, J. M., 2003, Land-Cover Classification Using Multitemporal ERS-1/2 InSAR Data, IEEE Transactions on Geoscience and Remote Sensing, Vol. 41, No. 7, pp. 1620 – 1628

ESA NEST, Online Documentation, http://corp.array.ca/nest-web/help/index.html last accessed 2014-06-08.

Esch, T., Thiel, M., Schenk, A., Roth, A., Müller, A. and Dech, S., 2010, Delineation of Urban Footprints from TerraSAR-X data by analyzing speckle characteristics and intensity values, IEEE Transactions on Geoscience and Remote Sensing, Vol. 48, No. 2, pp. 905 – 916

Esch, T., Marconcini, M., Felbier, A., Roth, A., Heldens, W., Huber, M., Schwinger, M., Taubenböck, H., Müller, A., 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, Vol. 10, No. 6, pp. 1617-1621

Foody, G. M., 2002, Status of Land Cover Classification Accuracy Assessment, Remote Sensing of Environment, Vol 80, 2002, pp. 185.201

Frankling, S. E., Wulder, M. A., 2002, Remote Sensing Methods in Medium Spatial Resolution Satellite Data Land Cover Classification of Large Areas, Progress in Physical Geography, 2002, Vol 26, No 2, pp. 173-205

Furberg, D., Ban, Y., 2012, Satellite Monitoring of Urban Sprawl and Assessment of its Potential Environmental Impact in the GTA between 1985 and 2005, Environmental Management, Vol. 50, pp. 1068-1088

Galland, F., et al., 2009, Unsupervised Synthetic Aperture Radar Image Segmentation Using Fisher Distributions, IEEE Transactions on Geoscience and Remote Sensing, 2009, Vol. 49, No 8, pp. 2966-2972

Gamba, P. , Aldrighi, M. and Stasolla, M., 2011, Robust Extraction of Urban Area Extents in HR and VHR SAR Images, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 4, No. 1, pp. 27 – 34

Gamba, P., Aldrighi, M., Stasolla, M. and Sirtori, E., 2009, A Detailed Comparison Between Two Fast Approaches to Urban Extent Extraction in VHR SAR Images, Proceedings Joint Urban Remote Sensing Event 2009, Shanghai, China

Gamba, P. and M. Herold, Eds. 2009. Global Mapping of Human Settlement—Experiences, Datasets, and Prospects. Boca Raton, FL: CRC Press.

Gao, Y., et al., 2011, Optimal region growing segmentation and its effect on classification accuracy, International Journal of Remote Sensing, 2011, Vol. 33, No 13, pp. 3747-3763

Gu, H. Y., et al., 2008, Object-oriented Classification of Polarimetric SAR Imagery based on Statistical Region Merging and Support Vector Machine.

2008 International Workshop on Earth Observation and Remote Sensing Applications, pp. 1-6.

Haas, J.,Ban, Y., 2014, Urban growth and environmental impacts in Jing-Jin-Ji, the Yangtze River Delta and the Pearl River Delta, International Journal of Applied Earth Observation and Geoinformation, Vol. 30, pp. 42–55.

Haralick, R. M., Shapiro, L. G., 1984, Image Segmentation Techniques, Computer Vision, Graphics and Image Processing, 29, 1985, pp. 100-132

Hay, G. J., Castilla, G., 2006, Object-Based Image Analysis: Strength, Weaknesses, Opportunities and Threats (SWOT), 1st International Conference on Object-based Image Analysis (OBIA 2006), Salzburg

Henderson, F. M. and Xia, Z.-G., 1997, SAR Applications in Human Settlement Detection, Population Estimation and Urban Land Use Pattern Analysis: A Status Report, IEEE Transactions on Geoscience and Remote Sensing, Vol. 35, No. 1, pp. 79 – 85

Henderson, F. M., Lewis, 1998,A. J., Principles & Applications of IMAGING RADAR, Manual of Remote Sensing, Third Edition, Volume 2, 1998, John Wiley & Sons, Inc.

Hester, D. B., Cakir, H. I., Nelson, S.A.C., Khorram, S., 2008, Per-pixel Classification of High Spatial Resolution Satellite Imagery for Urban Land-cover Mapping, Photogrammetric Engineering & Remote Sensing, Vol. 74, No. 4, pp. 463-471

Hsu, C.-W., et al, 2010, A Practical Guide to Support Vector Classification, available from http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf

Hu, H. and Ban, Y., 2012, Multitemporal RADARSAT-2 Ultra-fine Beam SAR Data for Urban Land Cover Classification, Canadian Journal of Remote Sensing, Vol. 38, No. 2

Hu, H., Ban, Y., 2008a, Urban Landuse/Land-cover Mapping with High-resolution SAR Data Using Neural Network and Rule-Based Classifiers, SPIE Europe Remtote Sensing Conference, Cardiff, U.K., 2008

Hu, H., Ban, Y., 2008b, Urban Land-cover Mapping and Change Detection with Radarsat SAR Data Using Neural Network and Rule-Based Classifiers, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, ISPRS Congress 2008, Beijing, China, Vol XXXVII, pp.1549-1553

Hu, H., Ban, Y., 2011, Multitemporal RADARSAT-2 Ultra-Fine-Beam SAR Data for Urban Land Cover Classification, Canadian Journal of Remote Sensing, Accepted for printing 2011-12

Hu, H., Ban, Y., 2014, Unsupervised Change Detection in Multitemporal SAR Images over Large Areas, IEEE Journal on Selected Topics in Applied Earth Observations and Remote Sensing, Accpeted for publication.

Hu, Y., Ban, Y., Zhang, Q., Liu, J., 2009, The Trajectory of Urbanization Process in the Yangtze River Dela During 1990 to 2005, 2009 Urban Remote Sensing Joint Event

Huang, C., et al., 2002, An Assessment of Support Vector Machines for Land Cover classification, International Journal of Remote Sensing, 2002, Vol 23 No 4, pp. 725-749

Im, J., et al., 2008, Object-based change detection using correlation image analysis and image segmentation, International Journal of Remote Sensing, Vol. 29, No. 2, 2008, pp. 399-423

Jacob, A., 2011, Radar and Optical Data Fusion for Object Based Urban Land Cover Mapping, Master of Science Thesis in Geoinformatics, TRITA-GIT EX 11-009, Royal Institute of Technology (KTH), Stockholm Sweden

Jensen, J. R., 2005, Introductory Digital Image Processing, A remote Sensing Perspective, Third edition, 2005, Pearson Prentice Hall, Pearson Education Inc., Upper Saddle River NY

Lee, J-S., Wen, J-H., Ainsworth, L., Chen, K-S., Chen, A. J., 2009, Improved Sigma Filter for Speckle Filtering SAR Imagery, IEEE Transactions on Geoscience and Remote Sensing, Vol. 47, No. 1, pp. 202-213

Li, X., Pottier, E., Guo, H. and Ferro-Famil, L., 2010, Urban land cover classification with high-resolution polarimetric SAR interferometric data, Canadian Journal of Remote Sensing, Vol. 36, No. 3, pp. 236-247

Liu, R., et al. 2010, A Multiobjective Immune Clustering Technique Applied to Unsupervised SAR image Segmentation, CIVR 2010, Proceedings of the ACM International Conference on Image and Video Retrieval

Margarit, G., Mallorquí, J. J. and Pipia, L., 2010, Polarimetric Characterization and Temporal Stability Analysis of Urban Target Scattering, IEEE Transactions on Geoscience and Remote Sensing, Vol. 48, No. 4, pp. 2038 – 2048

Marpu, P. R., et al., Enhanced Evaluation of Image Segmentation Results, Journal of Spatial Science, June 2010, Vol 55, No 1, pp. 55-68

Matinfar, H.R., et al., 2007, Comparison of Object-Oriented and Pixel-Based Classification of Land Use/Land Cover Types on Landsat7, Etm+ Spectral Bands (Case Study:Arid Region of Iran), American-Eurasion Journal of Agriculture & Environment Science 2, pp. 448-456

McNairn, H., et al., 2009, Integration of Optical and Synthetic Aperture Radar (SAR) Imagery for Delivering Operational Annual Crop Inventories, ISPRS Journal of Photogrammetry and Remote Sensing, 2009, Vol 64, pp. 434-449

Melgani, F., Bruzzone, L., 2004, Classification of Hyperspectral Remote Sensing Images With Support Vector Machines, IEEE Transaction On Geoscience and Remote Sensing, Vol 42, No 8, 2004, pp. 1778- 1790

Moran, E.F., 2010, Land Cover Classification in a Complex Urban-Rural Landscape with Quickbird Imagery, Photogrammetric Engineering Remote Sensing, Vol. 76, No. 10, pp. 1159-1168

Mountrakis, G., et al., 2010, Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing (2010), doi:10.1016/j.isprsjprs.2010.11.001

Myint, S.W., Gober, P., Brazel, A., Grossman-Clarke, S., 2011, Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery, Remote Sensing of Environment, Vol. 115, pp. 1145-1161

Müller, M., Krüger, W. and Saur, G., 2007, Robust Image Registration for Fusion, Information Fusion, No. 8, pp. 347-353

Niu, X. and Y. Ban. 2012. An adaptive SEM algorithm for urban land cover mapping using multitemporal high-resolution polarimetric SAR data, IEEE Journal on Selected Topics in Applied Earth Observations and Remote Sensing, 5(4):1129-1139.

Niu X. and Y. Ban. 2013 a. Multitemporal Polarimetric RADARSAT-2 SAR Data for Urban Land Cover Mapping Through a Dictionary-based and a Rule-based Model Selection in a Contextual SEM Algorithm. Canadian Journal of Remote Sensing, 39(02): 138-151.

Niu, X., and Y. Ban, 2013 b. Multitemporal RADARSAT-2 Polarimetric SAR Data for Urban Land Cover Classification using Object-based Support Vector Machine and Rule-based Approach, International Journal of Remote Sensing, 34(1):1-26.

Niu X. and Y. Ban, 2014. A Novel Contextual Classification Algorithm for Multitemporal Polarimetric SAR Data. IEEE Transaction on GeoScience and Remote Sensing Letters, 11(3): 681-685.

Novack, T., Esch, T., Kux, H., Stille, U., 2011, Machine Learning Comparison between WorldView-2 and QuickBird-2-Simulated Imagery Regarding Object-Based Urban Land-Cover Classification, Remote Sensing, Vol. 3, pp. 2263-2282

Pacifici, F., et al., 2007, Outcome of the 2007 GRSS Data Fusion Contest, IEEE Geoscience and Remote Sensening Letters, Vol 5, No. 3, pp. 331 – 335

Pacifici, F., Chini, M., 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, Vol. 113, pp. 1276-1292

Pal, M., Mather, P. M., 2003, Support Vector classifiers for Land Cover Classification, Map India Conference 2003

Pal, N. R., Pal, S. K., 1993, A Review on Image Segmentation Techniques, Pattern Recognition, Vol 26 No 9, 1993, pp . 1277-1294

PCI Geomatica, 2014, http://www.pcigeomatics.com/software/geomatica2013/, last accessed 2014-06-08

Pesaresi, M., Ehrlich, D., Caravaggi, I., Kauffmann, M., Louvrier C., 2011, Toward Global Automatic Build-Up Area Recognition Using Optical VHR Imagery, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 4, No. 1, pp. 16-26

Pesaresi, M., Gerhardinger, A., 2011, Improved Textural Built-Up Presence Index for Automatic Recognition of Human Settlements in Arid Regions With Scattered Vegetation, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 4, No. 1, pp. 16-26

Pohl, C., Van Genderen, J. L., 1998, Review Article Multisensor Image Fusion in Remote Sensing: Concepts, Methods and Applications, International Journal of Remote Sensing, Vol. 19, No. 5, pp. 823-854

Pu, R., Landry, S., Yu, Q., 2011, Object-based urban detailed land cover classification with high spatial resolution IKONOS imagery, International Journal of Remote Sensing, Vol. 32, No. 12, pp. 3285-3308

Quan, J.-J., et al., 2008, Multiscale Probilistic Neural Network Method for SAR Image Segmentation, Applied Mathematics and Computation, 2008, Vol. 205, pp. 578-583

Quegan S., Le Toan, T., Yu, J. J., Ribbes, F., Floury, N., 2000, Multitemporal ERS SAR Analysis Applied to Forest Mapping, IEEE Transactions on Geoscience and Remote Sensing, Vol. 38, No. 2, pp. 741-753

Qin, Y., Z. Niu, F. Chen, B. Li & Y. Ban. 2013. Object-based land cover change detection for cross-sensor images. International Journal of Remote Sensing, 34(19): 6723-6737.

Radke, R. J., et al., 2005, Image Change Detection Algorithms: A Systematic Survey, IEEE Transactions on Image Processing, Vol. 14, No. 5, 2005, pp. 294 - 307

Reigber, A., Jäger, M., He, W., Ferro-Famil, L., Hellwich, O., 2007, Detection and classification of urban structures based on high-resolution SAR imagery, Proceedings Joint Urban Remote Sensing Event 2007, Paris, France

Richards, J.A., 2012, Remote Sensing Digital Image Analysis, DOI 10.1007/978-3-642-30062-2_12, Springer-Verlag Berlin Heidelberg 2013

Schneider, A., Friedl, M. A., Potere, D., 2010, Mapping global urban areas using MODIS 500-m data: New methods and datasets based on ‘urban

ecoregions', Remote Sensing of Environment, Vol. 114, No., 8, pp. 1733-1746

Stasolla, M. and Gamba, P., 2008, Spatial Indexes for the Extraction of Formal and Informal Human Settlements from High-Resolution SAR Images, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 1, No. 2, pp. 98 – 106

Strozzi, T., Dammert, P. B.G., Wegmüller, U., Martinez, J. M., ASkne , J. I. H., Beaudoin, A., Hallikainen, M. T., 2000, Landuse Mapping with ERS SAR Interferometry, IEEE Transactions on Geoscience and Remote Sensing, Vol. 38, No. 2, pp. 766-775

Tan, S. R., et al., 2009, Unsupervised SAR image segmentation method based on MAP classification criterion and anisotropic diffusion smoothing, MIPPR 2009, Automatic Target Recognition and Image Analysis, Proc. of SPIE Vol. 7495 74951J-1

Thapa, R.B., Muruyama, Y., 2009, Urban mapping, accuracy & image classification: A comparison of multiple approaches in Tsukuba City, Japan, Applied Geography, Vol. 29, pp.135-144

Torresa, R., Snoeija, P., Geudtnera, D., Bibbya, D., Davidsona, M., Attemaa E., Pierre Potina, et al., 2012, GMES Sentinel-1 mission, Remote Sensing of Environment, Vol., 120, pp. 9-24

Trimble 2011, eCognition available at: http://www.ecognition.com/ Accessed at 20 September 2011

Walter, V., 2004, Object based classification of remote sensing data for change detection, ISPRS Journal of Remote Sensing, No 58., 2004, pp. 225-238

Wang, Y., Han C., 2010, PolSAR Image Segmentation by Mean Shift Clustering in the Tensor Space. Acta Automatica Sinica No 6. June 2010, Vol. 36.

Wang, Z., Jensen, John R., Im, J. 2010, An automatic region-based image segmentation algorithm for remote sensing applications. Environmental Modelling and Software 25, pp. 1149-1165.

Waske, B., Braun, M., 2009, Classifier Ensembles for Land Cover Mapping Using Multitemporal SAR Imagery, ISPRS Journal of Photogrammetry and Remote Sensing 64, 2009, pp. 450 - 457

Waske, B., Linden, van der, S., 2008, Classifying Multilevel Imagery From SAR and Optical Sensors by Decision Fusion, IEEE Transactions on Geoscience and Remote Sensing, Vol. 46, No. 5, pp. 1457 – 1466

Weih, Robert C. Jr., Riggan, Norman D. Jr., 2009, A Comparison of Pixel-based versus Object-based Land Use/Land Cover Calassification Methodologies, Journal of the Arkansas Academy of Science Vol. 63, 2009 152

Weng Q, Gamba P., Mountrakis, G., Pesaresi, M., Lu, L., Kemper, T., Xian, G., Jin, H., Miyazaki, H., Xu, B., Quresh, S., Keramitsoglou, I., Ban, Y., Roth, A., Elvidge, C. D., 2014, Urban Obsvering Sensors, Global Urban Monitoring and Assessment through Earth Observation. Ed. Q. Weng. Taylor and Francis Group, LLC, In press.

Voisin, A., Krylov, A. V., Moser, G., Serpico, S. B. and Zerubia, J., 2012, Multiscale Classification of Very High Resolution SAR Images of Urban Areas by Markov Random Fields, Copula Functions and Texture Extraction, Riunione Annuale GTTI 2012, Cagliari e Villasimius, 25-27 giugno 2012

Voisin, A., Moser, G., Krylov V., Serpico, S., and Zerubia, J., 2010, Classification of Very High Resolution SAR Images of Urban Areas by Dictionary-Based Mixture Models, Copulas And Markov Random Fields Using Textural Features, Proceedings of Signal Processing for Remote Sensing (SPIE) XVI, Vol. 7830, 78300O-1

Wong, A. and Clausi, D. A., 2007, ARRSI: Automatic Registration of Remote-Sensing Images, IEEE Transactions on Geoscience and Remote Sensing, Vol. 45, No. 5, pp. 1483-1493

Wong, A. and Clausi, D. A., 2010, AISIR: Automated Inter-Sensor/Inter-Band Satellite Image Registration Using Robust Complex Wavelet Feature Representations, Pattern Recognition Letters, No. 31, pp. 1160-1167

Xia, Z.-G. and Henderson, F. M., 1997, Understanding the Relationships Between Radar Response Patterns and the Bio- and Geophysical Parameters of Urban Areas, IEEE Transactions on Geoscience and Remote Sensing, Vol. 35, No. 1, pp. 93 – 101

Yang, W., Zou, T., Dai, D., Shuai, Y., 2009, Supervised Land-cover Classification of TerraSAR-X Imagery over Urban Areas Using Extremely Randomized Clustering Forests, Proceedings Joint Urban Remote Sensing Event 2009, Shanghai, China

Yang, Y., et al., 2008, AMarkov Random Field Model-Based Fusion Approach to Segmentation of SAR and Optical Images, IGARSS 2008, IEEE International Geosciene and Remote Sensing Symposium

Yousif, O. and Y. Ban, 2013. Improving Urban Change Detection from Multitemporal SAR Images Using PCA-NLM. IEEE Transaction on GeoScience and Remote Sensing, Vol. 51, No. 4, pp 2032-2041.

Yu, L., Zhang, D., Holden, E.-J., 2008, A Fast and Fully Automatic Registration Approach Based on Point Features for Multi-Source Remote-Sensing Images, Computers & Geosciences, No. 34, pp. 838-848

Zhang, H., et al., 2008, Image Segmentation Evaluation: A Survey of Unsupervised Methods, Computer Vision and Image Understanding, 2008, Vol 110, pp. 260-280

Zhang, J., 2010, Multi-source Remote Sensing Data Fusion: Status and Trends, International Journal of Image and Data Fusion, Vol. 1, No. 1, pp. 5-24

Zhang, J., Yang, J., Zhao, Z., Li, H. and Zhang, Y., 2010, Block-regression Based Fusion of Optical and SAR Imagery for Feature enhancement, International Journal of Remote Sensing, Vol. 31, No. 9, pp. 2325-2345

Zhang, Q, Ban, Y., Hu, Y, Liu, J., 2009, The Trajectories of Urban Land and Industrial Land in Shanghai over the Past 30 Years, 2009 Urban Remote Sensing Joint Event

Zhang, Q, Y. Ban, J. Liu and Y. Hu. 2011. Simulation and Analysis of Urban Growth Scenarios for the Greater Shanghai Area, China. Computers, Environment and Urban Systems, 35(2): 126-139

Zhang, X., et al., 2008, Spectral Clustering Ensemble Applied to SAR Image Segmentation, IEEE Transaction on Geoscience and Remote Sensing, 2008, Vol 46 No7, pp. 2126-213

Zhang, Z. , Zhang, J., Uao, M. and Zhang, U., 2000, Automatic Registration of Multi-Source Imagery Based on Global Image Matching, Photogrammetric Engineering & Remote Sensing, Vol. 66, No. 5, pp. 625-629

Zhou, W., et al., 2008, Object-based Land Cover Classification and Change Analysis in the Baltimore Metropolitan AreaUsing Multitemporal High Resolution Remote Sensing Data, Sensors, No. 8, 2008, pp. 1613-1636

Zitová, B. and Flusser, J., 2003, Image Registration Methods: A Survey, Image and Vision Computing, Vol. 21, pp. 977-1000