UNIVERSITA' DEGLI STUDI DI PAVIA
DOTTORATO IN INGEGNERIA ELETTRONICA,
INFORMATICA ED ELETTRICA

CICLO XXIX
A.A 2013-2016

Human settlement characterization at the global scale by fusing SAR and multispectral data sets at multiple resolutions

A DISSERTATION
SUBMITTED TO THE DEPARTMENT
OF ELECTRICAL, BIOMEDICAL AND COMPUTER ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF UNIVERSITY OF PAVIA
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Andreas Salentinig
2016

SUPERVISOR: Prof. Paolo Gamba

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma in any other University. To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

October 26, 2016

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Since 2008 more than 50% of the Earth's population is - for the first time in history - living in cities. Currently the world population continues to grow at tremendous rates and is projected to increase by more than one billion within the next 15 years, reaching 8.5 billion in 2030, and to increase further to 9.7 billion in 2050 and 11.2 billion by 2100 [1]. Even though urban areas cover only a small portion of the global land surface they are responsible for more than 60% of the global energy consumption, 70% of the greenhouse gas emissions and 70% of the global waste [2].

The ongoing rapid urbanization puts a lot of pressure on the geo-, bio-, and atmosphere and leads to a number of problems that need to be tackled immediately in order to guarantee a sustainable development of our planet. In this context it is very important to have detailed and accurate information about the location and extent of human settlements. Global urban maps can help us track urban expansion, improve our understanding of the influence of urban areas on the environment and climate and help us adapt when the biosphere/geosphere "pushes back" in the form of natural disasters [3]. Remote sensing offers a powerful tool to generate these urban area maps in a timely and cost-efficient manner.

In recent years the quantity as well as the quality of available Earth Observation (EO) data sets useful for urban mapping has increased significantly and their improved spatial, radiometric and temporal resolutions have opened unprecedented possibilities to extract detailed information about human settlements. Since multiple spatial scales, different materials and objects, two-dimensional and three-dimensional features are intermixed and closely located in urban areas, the use of multiple sensors, and the integration of their data with in situ measurements have become more and more widespread in urban related research using remote sensing sources. Spaceborne and airborne sensors are indeed able to capture spectral, spatial and geometrical properties of urban area constituents (e.g., roads, urban vegetation, buildings, infrastructures). However, each of them has its own limitations. For instance, very high spatial resolution (VHR) sensors provide fine object details, but with a limited capability to recognize materials because of their, usually small, number of spectral bands. Data collected from radar sensors mix geometrical and material properties into urban area images which are quite difficult to understand by a

non-expert. In order to exploit the wealth of today's satellite image availability and to take advantage of their partially complementary information it is necessary to design and apply sophisticated data fusion methods. A lot of effort has been made by the scientific community to develop efficient data fusion methods (see chapter 2 for a more detailed analysis of existing technical literature). However, in order to be prepared for future generations of EO data it is necessary to evaluate and optimize existing fusion methods as well as to develop new ones.

This thesis deals with the fusion of multi-source, multi-scale, and multi-temporal EO data for urban area extent mapping. An overview of already existing global urban maps and the presentation of state-of-the-art techniques used to obtain information about urban area extents from SAR and optical satellite imagery is provided in Chapter 2, along with a discussion of the basic concepts of data fusion.

The third chapter is devoted to multi-scale SAR data fusion. A set of novel data fusion approaches for SAR data of different spatial resolution are introduced, discussed and evaluated. This task is performed in a comparative study of four test sites, each one of them a different Megacity (i.e., a city with more than 10 million inhabitants). The developed fusion methods work at the pixel-, feature-, and decision-level and are intended to cover a variety of possible data availability scenarios. Accordingly, a general framework for multi-resolution SAR data fusion for urban area extraction has been formalized.

The possibility of fusing SAR and optical data is investigated in Chapter 4. Aiming at the incorporation of complementary information from active as well as passive sensors, two decision level fusion methods using pixel-, and object- based techniques are presented. Their accuracy is quantitatively assessed against manually extracted reference data sets for several urban areas around the world.

In the final chapter conclusions are drawn, summarizing the research findings.

All the research work described in this thesis has been designed, implemented and tested by me, under supervision and guidance by Prof. Paolo Gamba. Therefore, the novel ideas originally came from me, but were partly refined upon interaction with colleagues from the Telecommunications and Remote Sensing Laboratory at the University of Pavia and their partner institutions.

1.1 Initiatives and projects

The work presented in this thesis has been performed in the framework of three different international and interdisciplinary projects, briefly introduced in the following paragraphs.

1.1.1 ESA/MOST Dragon 3 Cooperation Programme

Dragon is an international Cooperation Programme funded by the European Space Agency (ESA) and the Ministry of Science and Technology of the People's Republic (P.R.) of China (MOST), and focuses on exploitation of ESA, third party missions and Chinese EO data for geo-science and applications development in land, ocean and atmospheric applications.

Following the successful completion of the Dragon 1 (2004-2007) and Dragon 2 (2008-2012) Programmes, the Dragon 3 Programme formally commenced with a Symposium in Beijing in 2012 and was recently concluded with a final results symposium (together with the Dragon 4 Kickoff Symposium) in July 2016 in Wuhan. The Programme brought together joint Sino-European teams to investigate 51 thematic projects.

Among those projects, one was dedicated to Urban Development and Climate (see red rectangle in figure 1.1), in which the University of Pavia played a major role investigating multi-temporal, multi-scale and multi-source satellite data for urbanization monitoring and climate impact analysis for sustainable urban development in China.

The specific objectives were:

- to understand how and where urban clusters/urban agglomerations develop in China using multi-temporal and/or multi-sensor satellite data;
- to design new algorithms that can rapidly detect urban clusters and land cover changes from multi-temporal and/or multi-sensor satellite data;
- to evaluate high spatial and spectral resolution spaceborne SAR and optical data for mapping the spatial configuration of selected Chinese urban landscapes;
- to evaluate what impact urban land use and land cover change has on the environment and climate in terms of local and regional temperature and precipitation;
- to investigate how climate change affect urban areas and to assess the vulnerability and risks of flooding, as well as extreme events in coastal urban areas.

1.1.2 ESA Climate Change Initiative

The Global Climate Observing System (GCOS), was formally established in 1992 by the World Meteorological Organization (WMO), Intergovernmental Oceanographic Commission (IOC), United Nations Environment Programme (UNEP), and International Council for Science (ICSU), as an international, interagency interdisciplinary framework for meeting the full range of national and international needs for climate observations. To meet the need for a systematic observation of climate, the GCOS programme developed the concept of the Essential Climate Variable (ECV). The ESA Climate Change Initiative (CCI) aims at the production of ECV datasets derived from Earth Observation data. Fifteen projects are producing 14 ECV datasets, whilst a Climate Modelling User Group offers a forum to ensure the products are relevant to the modelling and data assimilation community. Integrated teams involving university research groups, research institutions, government agencies and industry are engaged in delivering long-term time series of data needed for each of these. A part of the research of this thesis was performed in the CCI-Landcover (LC) framework, one of the 14 dedicated ECV projects (see figure 1.2).

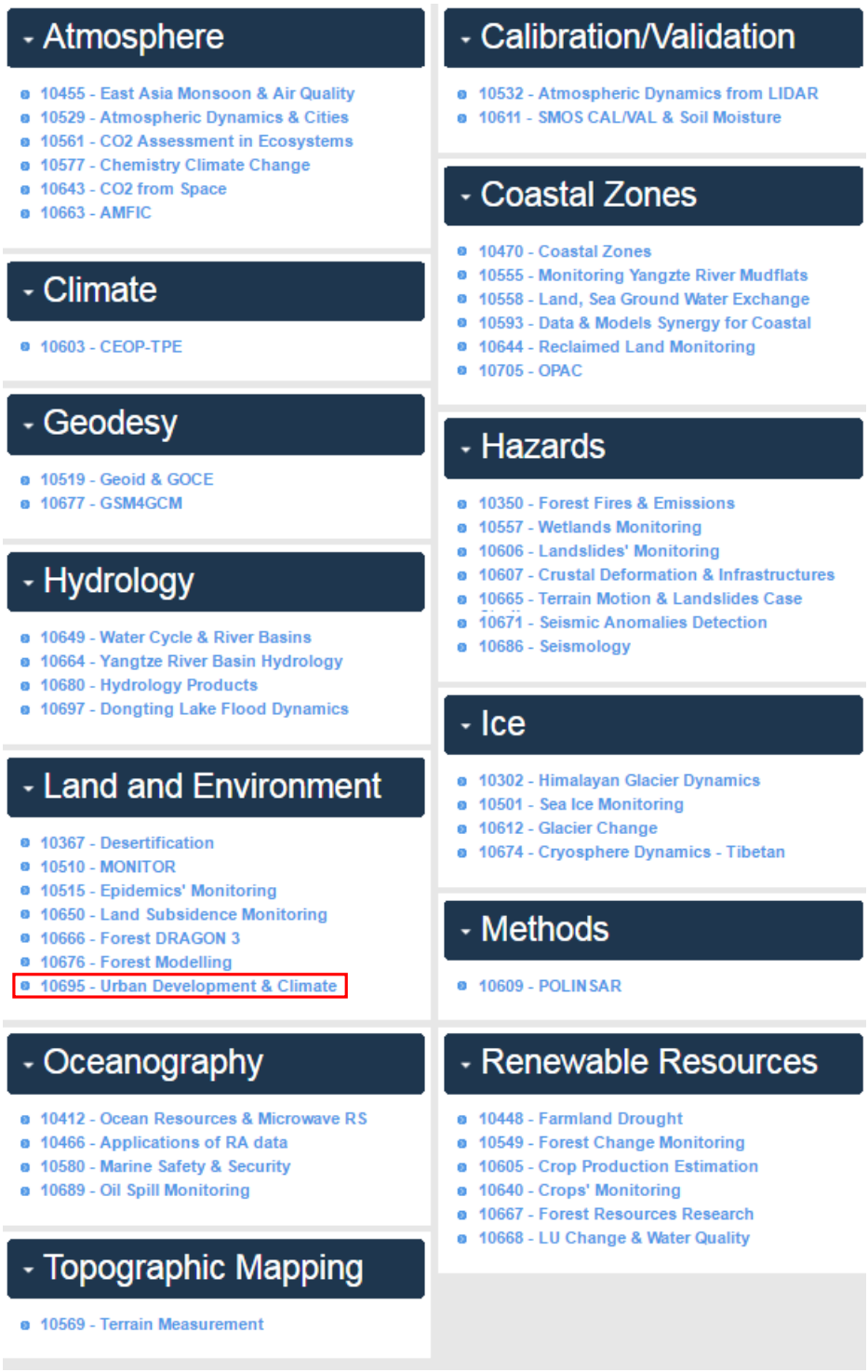

Figure 1.1: ESA/MOST Dragon 3 Cooperation Programme: 13 macro areas and 51 thematic projects; https://dragon3.esa.int/web/dragon-3/projects, accessed October 7, 2016.

1.1.3 TOLOMEO

The TOLOMEO (TOoLs for Open Mulit-risk assessment using Earth Observation data) project (http://tolomeofp7.unipv.it/, accessed October 8, 2016), funded under the Marie Curie International Research Staff Exchange Scheme (PIRSES-GA-2009), successfully

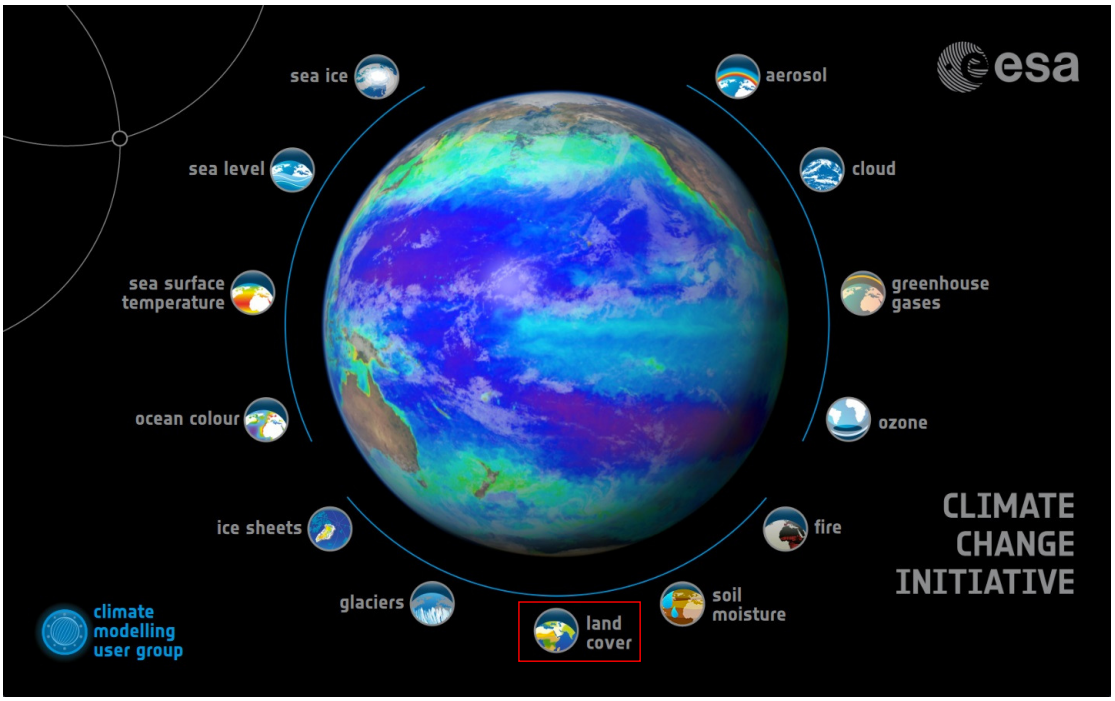

Figure 1.2: ESA CCI projects; http://www.esa-data-cci.org/showcase.html, accessed October 7, 2016.

completed in October 2015 with a final meeting in Rio de Janeiro, Brazil. The project, with the ultimate goal to develop a free software for the analysis of remote sensing data in the context of risk assessment and was managed by the University of Pavia and included the University of Extremadura Spain, the University of Hannover, Germany, the Pontifical Catholic University of Rio de Janeiro, Brazil, the Centre National d'Etudes Spatiales (CNES), France, and the Instituto Nacional de Pesquisas Espaciais (INPE), Brazil. The work was divided in different packages:

- WP2 - Tools for human exposure (HUEX) to multiple risks.
- WP3 - Tools for deforestation risk analysis (DERA).
- WP4 - Tools for earthquake physical (ERVU) vulnerability.
- WP5 - Tools for flood vulnerability (FLOV) characterization.
- WP6 - Tools for post-event risk (POER) management.
- WP7 - Training on parallel computing.

Parts of the research conducted on this thesis was performed in the context of TOLOMEO WP2.

Chapter 2

Global and regional urban area mapping by means of remote sensing

2.1 Human settlements and urban areas: a working definition

Defining the term "human settlement" may sound like a very easy task. A closer look, however, reveals that this is not the case, especially if it is intended to establish a globally valid description. A first definition of "human settlement" was officially introduced during the United Nations (UN) conference in Vancouver in 1976 as "the totality of the human community - whether city, town or village - with all the social, material, organizational, spiritual and cultural elements that sustain it" [4]. This definition is very broad and ambiguous and allows for very subjective interpretations. The term "human settlement" is sometimes used as a synonym or as an umbrella term for "urban area". As a matter of fact, a generally accepted global definition of the term urban - and its counterpart rural – does not exist and varies for each country, and even within countries it may not be consistent over time due to periodic reclassification. The UN Demographic Yearbook 2014 [5] contains a list of urban area definitions sorted by continent and country. Most approaches for urban area definition consider urban areas as a function of at least one of the following parameters: population size, area, population density or economic and social characteristics. The authors of [6] disclosed the global heterogeneity in defining urban areas with an illustrative example: "Of the 228 countries for which the United Nations compiles data, roughly half use administrative considerations — such as residing in the capital of the country or of a province — to designate people as urban dwellers. Among the other countries, 51 distinguish urban and rural populations based on the size or density of locales, 39 rely on functional characteristics such as the main economic activity of an area, 22 have no definition of "urban," and eight countries define all (Singapore, for example) or none (several countries in Polynesia) of their populations as living in urban

areas."

From an EO perspective none of the above mentioned definition approaches is suitable to describe the targets of remote sensing based urban mapping, because only physical characteristics can be derived directly from EO data. As a matter of fact, no general definition of urban area has been determined by the remote sensing community. Depending on specific applications and the used data type the actual features that are extracted vary significantly, from built-up areas [7]–[9], to artificial structures [10]–[12], to impervious surfaces [13]. In this thesis the results from fusion approaches exploiting SAR data only refer closest to built-up areas, whereas maps derived through fusion of SAR and optical data also include impervious surfaces.

2.2 Existing global urban maps and their limits

Although not completely overlapping with respect to the thesis contents, as discussed in the previous section, a number of global urban maps have been produced since the early 1990s. They vary significantly with respect to spatial resolution, used data and methodological approach. The following paragraphs are devoted to a brief introduction of global urban maps and a discussion of their advantages and drawbacks.

One of the earliest available global maps is the Vector Map Level 0 (VMAP0 [14]), a 1:1,000,000 scale GIS product that has been generated through digitizing a collection of navigational charts and maps. Usually the digitized urban patches trace the outer edge of urban areas and do not exclude inner-urban natural structures. Due to the fact that VMAP0 is generally a very conservative estimate of urban areas it is often used as part of input data for other global urban area extraction/refinement approaches.

MODIS1k [15] and its successive version MODIS500 [7], [16] are derived exploiting multi-spectral satellite imagery from the Moderate Resolution Imaging Spectroradiometer [17] using a stratified supervised classification approach based on training data visually collected from higher resolution optical imagery. Due to the fact that those maps rely on optical EO only, the urban areas are based on physical attributes, including a mix of human-constructed buildings and impervious surfaces.

The Global Landcover 2000 product (GLC2000 [10]) is based on similar remote sensing satellite data from the European Satellite Pour l'Observation de la Terre (SPOT)-VEGETATION instrument (1000m spatial resolution). It has been generated by 11 research teams, each applying distinct urban mapping approaches. The GLC2000 and the MODIS1K products have in common that they exploit Nighttime Lights data (LITES [18]) in order to constrain their classifications. LITES is a global layer of nighttime illumination intensity and therefore does not include binary information about human settlements, but provides a very important urban-related variable, because the intensity of emitted nighttime light is usually very high over urban areas, and therefore is a good indicator for human settlements.

Similar to GLC2000, Globcover (GLOBC [11]) is a multi-category global landcover

product containing a class that is dedicated to artificial surfaces and associated areas. However, the spatial resolution of GLOBC is with 300 m finer than GLC2000. GLOBC is generated through an automated landcover classification scheme based on spatio-temporal clustering of stratified product regions exploiting ENVISAT MERIS [19] multi-spectral data. A very similar global landcover product, the Climate Change Initiative - Landcover map (CCI-LC, [20]) capitalizes on the GLOBC unsupervised processing chain and improves it through adding machine learning classification steps and developing a multi-year strategy.

Other global urban products are based on a combination of remote sensing and ground based inputs. The result of the Global rural–urban mapping project (GRUMP [21]), a binary thematic urban extent map at a spatial resolution of 1000 m, is composed by integrating VMAP0, LITES, census data and a variety of GIS data sets. Landscan (LSCAN [22]) is a global layer that contains continuous spatial information about the global population distribution (1000 m spatial resolution) and is composed by integration of VMAP0, LITES, MODIS1K, census data and HR satellite imagery. Together with LITES, LSCAN is used to obtain another global map, the Global impervious surface area (IMPSA [13]). Finally, the History Database of the global environment (HYDE [23]) uses LSCAN, GLC2000, national census and land use statistics as well as administrative city gazetteers to estimate the fraction of urban land cover on a very coarse spatial resolution of 10,000 m.

More recently, and aiming at urban area mapping at the global scale at unprecedented levels of detail, three products exploiting medium and high resolution (HR) EO data have been produced.

The Globeland30 landcover map (GLC30 [12]) exploits multi-spectral Landsat data at a spatial resolution of 30 m. Among the 10 included landcover types, one class of GLC30 is devoted to artificial surfaces. The final product is generated through an integration of pixel-, object- and knowledge based classification methods.

The global urban footprint (GUF [24]) is based on two global HR SAR data sets acquired by the the satellite constellation of TerraSAR-X and TanDEM-X between 2011 and 2013. Textural information extracted from the SAR data is used in a fully automatic unsupervised classification that derives a built-up/non built up map at a spatial resolution of 12 m. Due to the fact that only SAR data is incorporated, the GUF output corresponds to built-up regions that feature vertical structures and excludes impervious surfaces without a vertical component.

A similar product, the global human settlement layer (GHSL [9]) incorporates multi-resolution (0.5-10 m), multi-sensFor (panchromatic, multi-spectral) and multi-temporal optical EO data to derive information about the global extent of urban areas. The information extraction is based on PANTEX [25], a rotation invariant, anisotropic textural measure based on the grey level co-occurence matrix [26] of the imagery. The GHSL (38 m final spatial resolution) is a continuous built up index that ranges between 0 and 1, but can be semantically translated to a binary built-up mask.

A summary of the global urban maps presented above including information about the

producers, the definition of urban areas, spatial resolution, used data, applied methods and global urban area extents is given in table 2.1.

Table 2.1: Overview of existing global urban maps (adopted and updated from [16] and [27]).

Abbr.	Time stamp	Producer	Definition of urban features	Res.	Data	Methods	Urban extent (km2)
VMAP0	1992	US National Geospatial Intelligence Agency	Populated places	1:1 mil.	Aeronautical charts, maps	Digitization	276,000
GLC 2000	1999/2000	European Commission Joint Research Center	Artificial surfaces and associated areas	1000 m	SPOT-Vegetation, Nighttime lights data	Unsupervised classification	308,000
GLOBC	2009	European Commission Joint Research Center	Artificial surfaces and associated areas	300 m	MERIS (partly GLC2000)	Unsupervised classification	313,000
HYDE	2000	Netherlands Environm. Assessment Agency	Urban area	10000 m	Landscan, UN census data, city gazetteers	Data fusion by aggregation, decision rules	532,000
IMPSA	2000/2001	US Nat. Geophysical Data Center	Density of impervious surface area	1000 m	Landscan, Nighttime lights data	Data fusion by linear regression	572,000
MODIS 500	2001/2002	University of Wisconsin, Boston University	Built environment	500 m	MODIS 500 m	Supervised classification	657,000
MODIS 1k	2000/2001	Boston University	Urban and built-up areas	1000 m	MODIS 1 km, Nighttime lights, population density	Supervised classification, data fusion	727,000
GRUMP	1995	Earth Institute at Columbia University	Urban extent	1000 m	VMAP, census data, Nighttime lights, maps	Data fusion by logarithmic regression	3,524,000
LITES	1992-2015	National Geophysical Data Center	Nighttime illumination intensity	1000 m	DMSP-OLS 2.2 km data	Data compositing	NA
LSCAN	1998-2014	US Oak Ridge National Laboratory	Ambient global population distribution	1000 m	Geocover maps, VMAP0, MODIS 1k, Landsat, census data, HR imagery	Data fusion	NA
CCI-LC	2000/2005/ 2010	European Space Agency	Urban areas	300 m	MERIS	Supervised classification	NA
GLC30	2000/2010	National Geomatics Center of China	Artificial	30 m	Landsat	Pixel-, object- and knowledge based classification	NA
GUF	2011-2013	German Aerospace Center	Built-up areas marked by the presence of vertical structures	12 m	TerraSAR-X /TanDEM-X	Unsupervised classification	NA
GHSL	2014	European Commission Joint Research Center	Built-up areas marked by the presence of buildings	38 m	multi-sensor optical data	PANTEX	NA

The evolution of the above mentioned global urban maps clearly shows that since the early 1990s their spatial resolution and level of detail has increased significantly due to the increasing availability, quality and accuracy of data sets usable for urban area mapping and the improved effectiveness of complex classification approaches. However, there are still major challenges associated with the usage of currently available global urban data sets. Looking at the rightmost column in table 2.1, it is obvious that the urban maps depict a high degree of variability and inconsistency in the quantification of the global urban area extent. Unfortunately this information is not available for all urban maps, but it may be assumed that those numbers would also be quite different from the ones in the table. As a consequence, it is impossible that all products depict urban areas correctly. Another issue with the usability and comparability of available global urban maps is that most of them exploit (very) coarse resolution data, which presumably leads to omissions

of small cities and errors along city edges due to mixed pixels.

The second major challenge in using global urban mapping approaches is that, as already mentioned in section 2.1, no consistent and unambiguous definition of the term "urban area" exists. The available global urban products exhibit information about human settlements either in a binary or continuous way and the features that are actually mapped include populated places, artificial surfaces and associated areas, percentage of urban areas, density of constructed impervious surfaces, nighttime illumination intensity and built-up areas marked by the presence of buildings or vertical structures. It has been shown that it is not uncommon that specific urban maps have been used in applications where other urban products or derivatives might have provided a better approximation [28], [29]. Additionally, the urban definition might be further blurred if data sets from very different sources are incorporated since the extraction of urban areas from spaceborne data sets is highly data dependent. The interested reader is referred to [16] and [27] for a more detailed comparison and quantitative evaluation of existing global urban products.

In the following sections a more detailed analysis of the state of the art of algorithms used for urban area extraction from SAR and multi-spectral satellite data is offered. Specific focus is on the techniques designed, implemented and optimized by the remote sensing research group at the University of Pavia.

2.3 SAR-based methods for urban mapping*

The big advantage of SAR for urban area mapping relates to the peculiarity of this active microwave sensor, able to acquire data in any weather and day/night conditions. The challenges are however as big, because the joint two- and three-dimensional features of urban area bring distortions to the SAR images and make their interpretation very difficult. The radar phenomenon that predominates in urban areas is the double bounce effect due to the presence of built up structures. This effect, sketched in figure 2.1, makes the return from artificial structures significantly higher than from most of other natural objects. Moreover, it is a feature that remains stable in time, while the backscattering from vegetated areas changes according to the season, and in water areas according to wind and wave patterns. The small subsets of two SAR images of Gothenburg (Sweden) and Beijing (P.R. China) in figure 2.1 show very clearly the bright pattern outlining urban areas. Although bright points (i.e., strong backscatterers) are visible in other portions of the scene, urban areas are characterized by a huge density of them. This feature can be used to discriminate urban areas even at rather coarse spatial resolution. As a matter of fact, at this scale it has been proven that SAR data can be more accurate than multispectral data with similar resolution [30] in urban area extent delineation.

*A large part of this section comes from: A. Salentinig and P. Gamba, "Combining SAR-Based and Multispectral-Based Extractions to Map Urban Areas at Multiple Spatial Resolutions," in *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 100-112, 2015 and G. Lisini, A. Salentinig, P. Du, and P. Gamba, "SAR-based global urban extents extraction: from ENVISAT to Sentinel-1", a paper submitted to *Remote Sensing of Environment* in March 2016

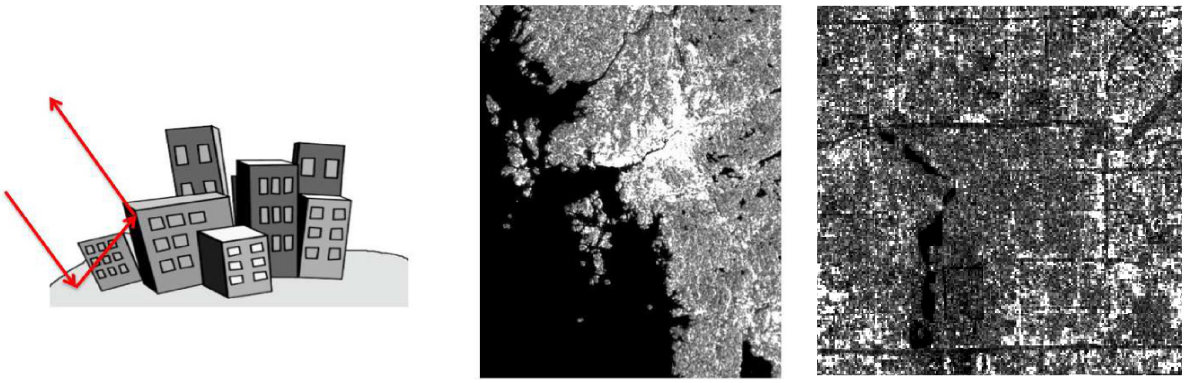

Figure 2.1: Graphical representation of the double bounce effect due to side-looking radars illuminating built-up structures (left), its effect (high density of bright pixels) in an ASAR WSM image of Gothenburg, Sweden (middle) and a typical pattern of urban and natural structures as seen on ASAR FM data (City Center of Beijing, P.R. China).

2.3.1 Existing methods in technical literature

As mentioned, the double bounce effect in urban areas can be exploited to recognize human settlements in SAR scenes. At a coarse spatial resolution this effect is not really dependent on the viewing angle (e.g., due to the sensor acquisition mode), or the satellite path (i.e., ascending/descending). The reason is that each pixel provides a combination of multiple double bounce effects due to multiple built up structures. Hence, in coarse resolution images, the total "brightness" of a pixel is related to the density of built up structures much more than their actual spatial pattern, which is not visible at that resolution. Accordingly a very effective algorithm to extract urban area extents from coarse resolution SAR data is the Urban Extractor (UEXT) method [31]. Since UEXT plays an essential role in this thesis its description is deferred to the next section (2.3.2).

At high and very high spatial resolution (VHR), although the double bounce effect is still valid as a peculiar feature of urban areas, the available finer details change the appearance of human settlements. Instead of having multiple reflections mixed up in one pixel, non-built-up areas between buildings become evident, revealing road networks, gardens, and courtyards. Additionally, layover and shadowing effects [32] cannot be avoided, and complex spatial patterns made by bright and dark pixels tend to appear.

Built-up areas in finer resolution SAR images are therefore better characterized using textural features to mathematically express the above mentioned spatial relations. Moreover, features representing spatial patterns should be able to capture local features at the building level as well as wider spatial scales (i.e., at the block level). Different textural features are thus mandatory.

The application of spatial analysis tools (e.g. texture measures) on VHR SAR allows a differentiation of even finer details of the heterogeneous inner urban mixture of artificial and natural surfaces. However, the computational cost as well as the needed data storage for urban extraction increases dramatically with finer resolutions and therefore global

VHR SAR based human settlement mapping is a very challenging task. Through the recently enabled availability of global VHR TerraSAR-X data sets and the application of a sophisticated urban area detection algorithm in a high level computing environment it is possible to map urban extents at an unprecedented level of detail. The work in [8] presents the Urban Footprint Processor (UFP), an urban extent mapping algorithm generating settlement masks from global data of the TanDEM-X Mission. UFP includes three main processing stages. First, "speckle divergence" [33], one of the possible textures features suitable for for detecting and delineating heterogenous urban structures from SAR data is extracted. The second step of UFP is devoted to the generation of binary settlement masks based on an unsupervised classification approach using the texture feature as well as the original backscattering amplitude. In a final post processing step the ASTER Global DEM [34] is used to discard false positives which might occur due to topographic effects in mountainous areas. The fully automated processing chain has been tested on a huge number of scenes covering wide areas. However, the final global urban product as a downscaled (from 12 to 50-75 m spatial resolution) public domain version is expected to be released soon.

In technical literature the main consensus of published papers dealing with SAR based urban area mapping procedures is that no single pixel measure is suitable alone for the task and that spatial analysis is required. Therefore, local as well as global spatial features are considered as a valid tool for human settlement mapping. In addition to the approaches mentioned above, further examples of SAR based urban area mapping approaches can be found in [35]–[42].

Another example of HR SAR data based urban mapping, developed at the University of Pavia, exploits a set of different texture measures and is described in detail below since this procedure has been used in several data fusion approaches in this thesis. As a matter of fact, the original approach in [43] eventually provided more sound results in [44] by jointly considering Local Indicators of Spatial Association (LISA) [45] and Grey-Level Co-occurrence Matrix (GLCM) [26] textures to try and capture this multi-scale textures. Indeed, LISA features (see figure 2.2) capture in the close neighborhood of a pixel the finest available details, such as checkerboard-like patterns due to alternating fill and void area in the 3D urban structure. GLCM textures, which are computed on a larger window, are more related to block patterns. A combination of these features, and the fusion of the decisions based on these combinations define a cascade of feature-based and decision-based data fusion techniques that is very effective when applied to ASAR Fine Mode (FM), PALSAR, Radarsat, TerraSAR-X and Cosmo/SkyMed images, from 30 to 1 m spatial resolution.

Specifically, the algorithm [44] computes LISA textures in the smallest possible 3×3 pixels window, while GLCM textures in a 21×21 window, with the clear intent to average the latter spatial information for blocks. LISA and GLCM features are first fused within the two sets to detect local patterns which are strong at the local or block level:

1. *a hard "AND" between points with high values of the Moran and Geary indexes*:

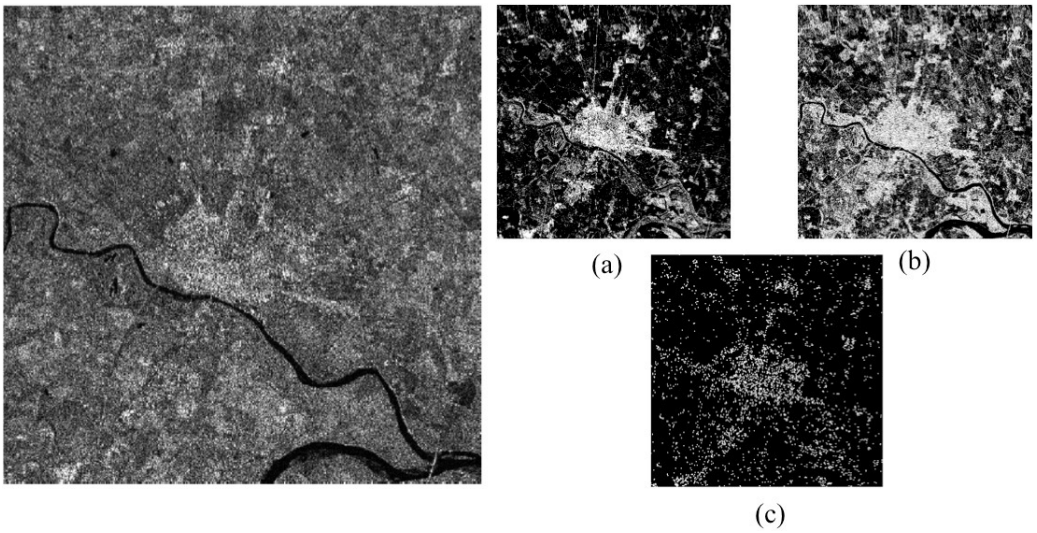

Figure 2.2: LISA textural features computed from a TerraSAR-X image of Pavia, Italy: (a) Getis-Ord's, (b) Geary's, and (c) Moran's indexes.

this is a way to select only areas that simultaneously exhibit positive and negative autocorrelation, like close buildings (bright targets) intertwined with roads or radar shadows (low/no reflection areas);

2. *the average between GLCM correlation and variance*: this step aims at finding areas in the SAR image where both high variance (due to targets and shadows) and correlation (e.g., strong spatial pattern) are found.

After the analysis at a single spatial scale, an object based analysis is performed by jointly considering the local density of points selected at either scales. Objects are selected by using a logical "OR", i.e. looking for all points that match the spatial pattern criteria at least at one scale. However, only objects predominantly characterized by pixels that are relevant at both scales (i.e., that results from a logical "AND") are finally selected.

This urban area extraction algorithm has recently been refined and tailored to the use of multi-temporal and -polarization ASAR FM data in collaboration with the Royal Institute of Technology (KTH) in Stockholm, resulting in a method called KTH-Pavia Urban Extractor [46]. The main modifications correspond to improved pre-processing of the input data including terrain correction using SRTM DEM and contrast enhancement as well as a post-processing step excluding misclassified urban structures in steep areas where human settlements are highly unlikely through exploitation of Digital Elevation Model (DEM) height information. Furthermore, the algorithm has been applied on a number of ASAR FM images acquired at different dates and with varying polarizations. The resulting urban extractions have been combined via logical 'OR' operation in order to reduce the omission rate.

2.3.2 The UEXT method

The UEXT algorithm is based on the following assumptions:

- thanks to the radar double/triple bounce effects urban areas are characterized by very bright pixels in SAR imagery; the dihedral structures formed by the walls of the buildings, oriented equally and randomly in any direction, ensure that very strong backscattering is experienced by any spaceborne SAR sensor from at least a few elements of an urban area, irrespectively of the ascending/descending orbit of the platform, the left/right viewing angle of the sensor, its polarization and frequency;

- at a spatial resolution of tens of meters, urban areas appear in SAR images as connected structures with an average backscatter level larger than the surroundings; even in residential blocks composed by small, isolated buildings with gardens the stability of the double bouncing effect, as opposed to the season variability of the diffusion by the vegetation, ensures that human settlements are brighter than the rest of the SAR scene;

- isolated and/or clusters groups of bright pixels in areas with large slope values are more likely to be rocks than built-up structures; as a mater of fact, people tend not to build in areas that are very difficult to access, unless they are close to existing larger settlements with more favorable terrain features. Similarly, bright pixels in a desert or within water areas are more likely to be false positives than correctly extracted human settlements.

In line with these assumptions, the UEXT approach is structured in three steps (graphically represented in figure 2.3):

1. a *seed selection* step, devoted to the selection of very bright points, expected to be part of a human settlement, if the amount of backscattered power is very large in absolute terms (e.g., with respect to a threshold whose value, as been discussed in [31]);

2. a *region growing* step, devoted to add more urban pixels to the original seeds. Neighboring and connected areas characterized by a similar level of backscattering are iteratively added to the starting points: by iteratively enlarging the set of pixels labeled as "urban", this step recognizes that urban areas are composed by clusters of buildings, interconnected by means of other artificial/natural structures;

3. A final *post-processing* step, devoted to fill into the gaps among built-up clusters and include non-built-up portions of an urban area (e.g., squares and parks, depending on the type of post-processing applied). Similarly, this step discards false extraction according to physical and logical rules, such as bright SAR points in rocky, desert or high sloped areas.

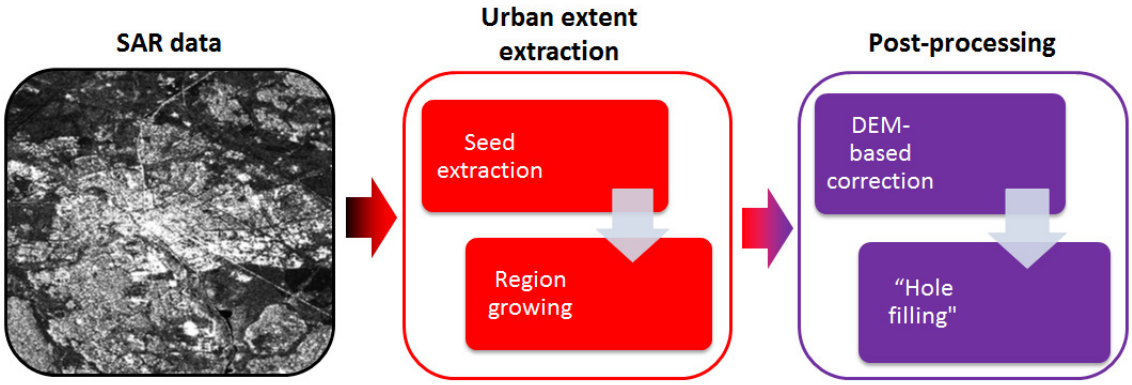

Figure 2.3: Graphical representation of the UEXT workflow.

Unfortunately, these assumptions are also the reasons for the limits of the original UEXT procedure in peculiar situations.

1. The notion of "bright" pixel depends on the environment in some cases the contrast between built-up areas and the natural background may be lower and depending on seasonal effect of the vegetation. The same comment is valid for human settlements in arid or rocky environments.

2. The notion of connected settlements is less and less valid, the finer the spatial resolution of the SAR data, and scattered and small settlements are likely to be missed, as they remain unconnected to major urban areas if do not possess a bright enough *seed*.

3. The rules for "non-urban" zones depend on the existence of settlements close to the extraction under scrutiny, and higher slope may be allowed, for instance, in areas close to mountain ranges or to the sea.

As a result, the UEXT algorithm, originally developed for coarse spatial resolution (75 m) data from the Envisat ASAR WSM [47] sensor, had to be adapted for data sets from Sentinel-1A (S1-A), one of the new European EO satellites, the so called Sentinels [48]. S1-A is equipped with a SAR sensor, which acquires 20/40 m spatial resolution data in a different image acquisition mode.

Indeed, while the ASAR sensor acquired imagery with the ScanSAR technique, S-1A uses Terrain Observation with Progressive Scans SAR (TOPSAR), which enables the beam to be steered not only in range, as in ScanSAR, but also from backward to forward in the azimuth direction for each burst. As a consequence, through this new SAR image acquisition method scalloping can be avoided and a more homogeneous image quality throughout the swatch can be achieved. However, due to the differences in SAR image acquisition, the resulting intensity values and ranges are different between ASAR and S-1A images. Therefore, the values responsible for the detection of very bright spots, and those ones which define the constraining value of the consecutive region growing procedure had

to be modified in order to extract reasonable information about built-up areas.

Additionally, parameters related to spatial size, such as the maximum number of pixels of an urban patch in mountainous areas, had to be adapted to the finest resolution. The threshold for the the number of pixels that selects urban patches, whose inside holes (caused by e.g. urban parks) are to be filled via a flooding procedure, had to be modified similarly. An exemplary result of the application of UEXT on Sentinel 1 data is shown in figure 2.4.

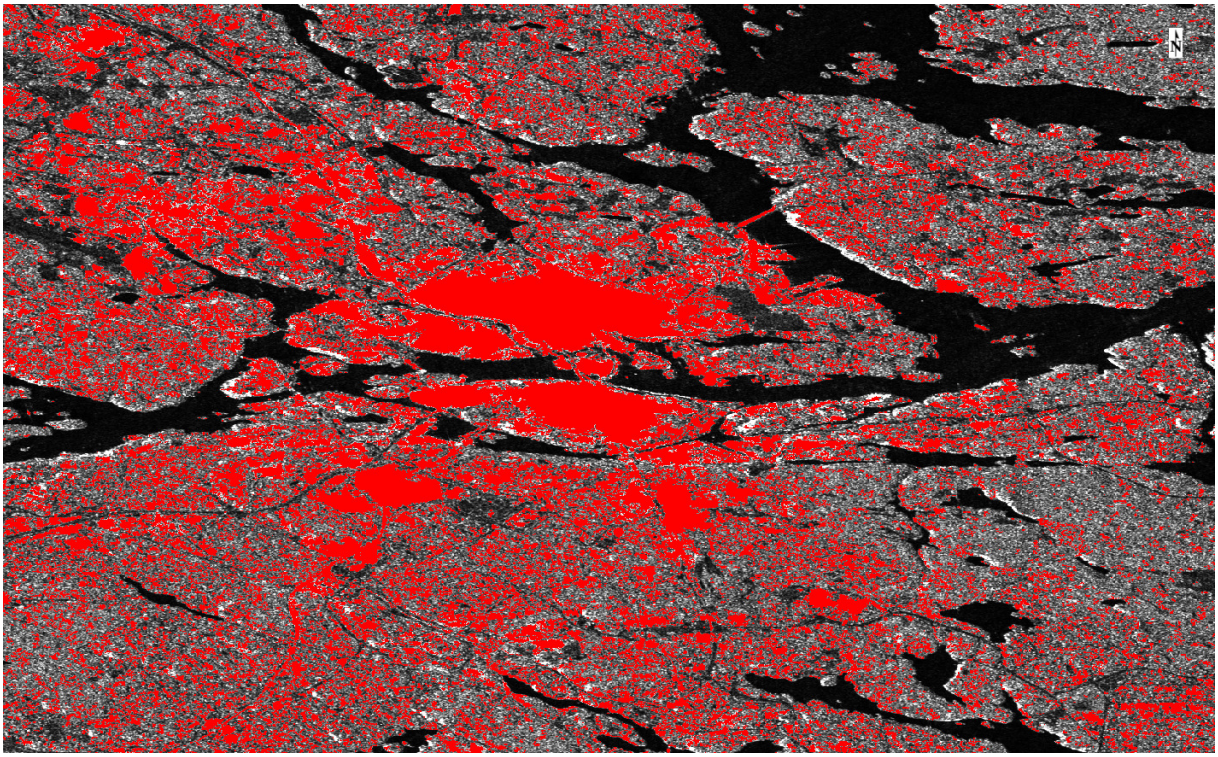

Figure 2.4: UEXT result for the city of Stockholm, Sweden, superimposed to Sentinel 1 data (VV polarization, acquired on May 15, 2016).

2.4 Optical-based methods for urban mapping*

While SAR is extremely useful to look for rather coarse information about human settlement extents, much more than optical data, the extraction of elements of the urban landscape at a finer scale is better obtained using multi-spectral data. Substantially, the availability of multi-spectral data for the whole world and for a time span longing more than 30 years, thanks to the NASA Landsat program, provides the users with an enormous amount of data and information for multi-temporal analysis. This is specifically important for human settlement characterization, and complements the information extracted

*A large part of this section comes from: A. Salentinig and P. Gamba, "Combining SAR-Based and Multispectral-Based Extractions to Map Urban Areas at Multiple Spatial Resolutions," in *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 100-112, 2015

by SAR data, which covers a much shorter time interval. It is worth noting that this is due to the fact that SAR sensors, although global missions are available since the late '80s, do not usually acquire in a continuous mode. Active sensors have higher energy demands than passive ones, and acquisition policies by national and international space agencies are different.

2.4.1 Existing methods in technical literature

Multi-spectral images of urban areas are quite various and there have been attempts to develop spectral indexes for urban area extraction. The authors of [49] introduced the Normalized Difference Built-up index (NDBI), which takes advantage of the unique spectral response of urban areas, using different bands of the near infrared (NIR) portion of the electromagnetic spectrum. The index based built-up index (IBI [50]) is distinguished from conventional indices by its first-time use of thematic index-derived bands to construct an index rather than by using original image bands. For its computation the already existing soil adjusted vegetation index (SAVI [51]), the modified normalized difference water index (MNDWI [52]) and the NDBI are used. Another index used to extract information about the extent of human settlements is the new built-up index (NBI [53]). In addition to two NIR bands, as in the computation for NDBI, the spectral response recorded in the red portion is included into the calculation of the NBI. Despite all the efforts to determine a single index to describe urban areas all indices are still far from being able to capture the vast complexity of urban material mixtures.

The need to manage this huge and various data sets in a consistent way has brought to the development of a methodology based on multiple indexes and semi-automatic training site selection. Specifically, as proposed in [54] it is possible to exploit a set of normalized indexes computing the differences among all possible bands of Landsat data, the so called Normalized Difference Spectral Vector (NDSV). The advantage of NDSV is that it has a very similar spectral behavior for urban areas all around the world, while natural materials, such as vegetation (see figure 2.5), show different features. Unfortunately, as the binary classification problem urban/non-urban must recognize a unique set of spectra within a wealth of very different and highly variable ones, the use of NDSV at the global level can be implemented only by selecting training areas all over the world, and using them as a geographically-tuned reference database. To this aim, existing global databases for urban area extents can be used, such as [55], as well as existing moderate resolution urban layers that are known for underestimating urban areas, such as GlobCover [11], [56]. By looking only for pixels well inside the border of these data sets, it is possible to extract a local sample set useful to run a binary supervised classifier, or a spectrum similarity matching algorithm.

As example of the extraction results for the Jinagsu province (P.R China) using the NDSV based approach is proposed in figure 2.6 for two dates (2000 and 2010). The extraction can be easily compared with a visual urban extent extraction, kindly provided by Tsinghua University. Note that this urban map includes only the most populated cities,

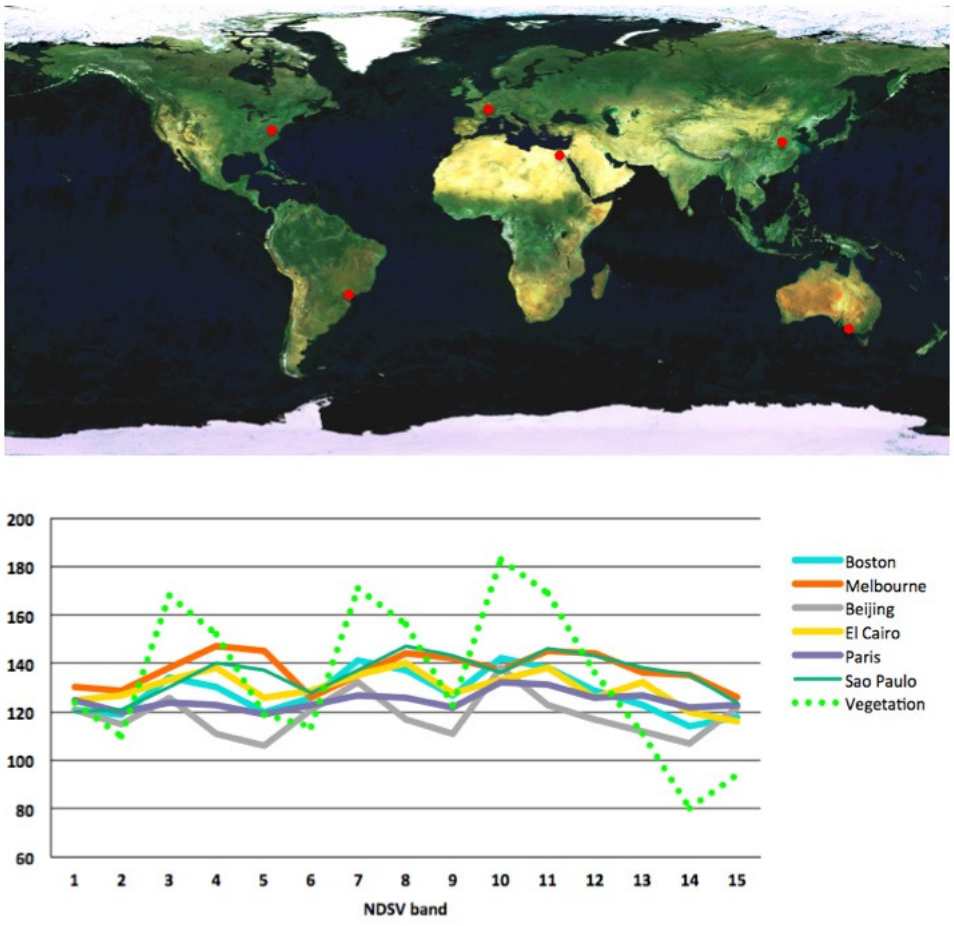

Figure 2.5: NDSV spectra for "urban" pixels in a few different parts of the world: all of them show a "flat" spectral behavior with respect to the typical NSDV spectrum of a vegetation pixel.

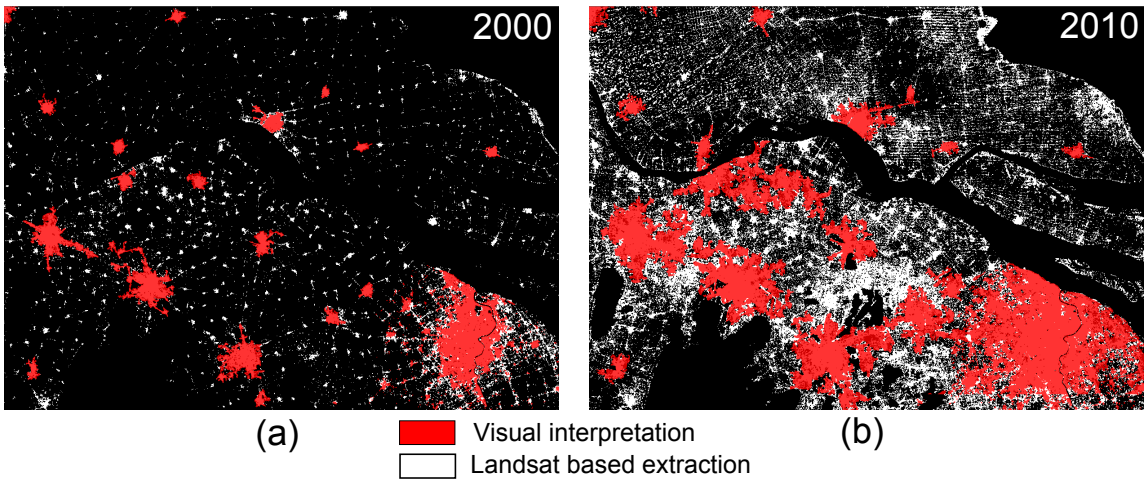

Figure 2.6: Extraction of human settlements extents using Landsat data for a part of the Jiang Su province compared with the extents visually obtained by Tsinghua University for (a) 2000 and (b) 2010.

hence discards small settlements.

The authors of [57] presented another promising approach for the exploitation of Landsat data for the purpose of urban area mapping. An "exclusion-inclusion" framework, which is comprised of two main steps has been developed. First, an inventory-based training set is optimized to obtain a number of statistical parameters required to build a non-settlement mask. In a second step the similarity of each image pixel to pre-collected sample points is calculated for the un-masked areas and a tresholding technique is applied in order to differentiate between urban and rural. However, the approach described above needs further validation in order to evaluate its usefulness for truly global analysis, as it has been tested for three regions in China only.

Moving to finer spatial resolution, the use of VHR optical data for the purpose of global urban mapping holds similar potentials and limitations as the ones discussed for VHR SAR data in section 2.3.1. In addition to the challenges regarding global data availability and consistency, the storage of huge data amounts and the computational cost to process them are aspects to be duly considered. Additionally, we must note that HR and VHR optical satellite data with spatial resolutions between 10 and 0.5 m allow a differentiation of small scale features, unnecessary for urban extent extraction purposes. For instance, information about features on a buildings' roof (e.g. chimneys, water tanks of windows) is not significant for the extraction of urban extents. Therefore, the very high spatial resolution and its inherited increased spectral variability is - in the context of wide area urban extractions - rather a disadvantage than an advantage. The Joint Research Center of the European Commission developed a general framework to address the challenges of global urban mapping using HR and VHR optical data [9]. The framework has been tested on a wide geographic areas with a variety of HR and VHR optical data sets. The fully automatic image information extraction, generalization and mosaic workflow includes

multiscale textural and morphological image feature extraction, image feature compression and optimization and novel learning and classification algorithms and is able to achieve results beyond current levels of detail of global urban maps. For a detailed description of the processing chain see [9].

2.5 Data fusion basic concepts

The urban area extraction approaches discussed in sections 2.3 and 2.4 have in common that only one data type, either from active or passive sensors, is used for analysis. Furthermore, multi-scale aspects are not exploited, as most approaches do not integrate multi-resolution data. However, any natural or artificial environment is so complex and challenging that it would be simplistic to assume that a single sensor is able to provide all the information that we may require for its characterization. Moreover, all these environments keep changing, although with different temporal scales (from minutes to decades), and sensors are steadily improving, gaining an increasing capability to provide multi-temporal/multi-wavelength/multi-view images of the same area. No doubt in the future a huge amount of data sets will be available on wider and wider geographical areas, with ground resolutions that may range from several tens of meters to below one meter. The challenges posed by such a large amount of information are first of all related to the establishment of suitable application driven criteria for data choice and use. A further challenge is related to the development of automatic or semi-automatic techniques capable of extracting vector layers from raster data, in order to collect them in a Geographic Information System (GIS). As a matter of fact, the use of remotely sensed information in geographic applications is increasing, applications are often located at the high resolution end of the current data ground resolution range, many interesting characterizations are feasible using data from existing sensors. The same amount of data will also pose issues with data storage, compression, visualization, search, and transmission, besides interpretation, all fields where data fusion concepts may be and are already exploited to provide more efficient and effective data management. An example is pansharpening [58]–[60], equally useful for visualization and for classification, a type of fusion based on the injection of the finer spatial details of panchromatic images into the valuable spectral details of multispectral/hyperspectral data sets. In a few words, the continuously increasing amount of multiple sensor data offers opportunities to address new questions and to solve existing problems more effectively, but the amount of possibilities is sometimes frightening. To exploit the full potential offered by these diverse data sources, several questions need to be investigated, for example how to select the best data sets and how to implement and exploit the algorithms and procedures originally developed for single sensor data processing. Another important question is how to combine the data and/or the information derived from these different sensors, for example at which level to fuse the individually processed data. Moreover, often data is not available within the most suitable timing window, but something else is achievable. This is especially true for very high spatial resolution sys-

tems, both for optical sensors like Quickbird-2, Worldview-3 [61], and Pleiades [62] and more recently for radar systems like Cosmo/SkyMed [63] and TerraSAR-X [64]. If different sensors have to be considered, different spatial resolutions and frequencies (polarizations for radar systems), as well as different types of noise, positional and geometrical errors must be taken into account for each data set. A flexible data fusion system is therefore required, user-oriented and able to adapt to multiple applications. To reach this goal, research on data fusion has been considering different typologies of fusion [65], as briefly reported here below.

- Fusion at the **pixel level**: information extraction by joint classification of combinations at the pixel level of data sets coming from different sensors (pros: very fast approach, cons: requires non-parametric classifiers, usually less user-friendly).
- Fusion at the **feature level**: comparison of features (objects) extracted from different imagery to improve temporal resolution or fill in the gaps of useful time series (pros: interesting for specific features, such as roads, buildings in urban areas, and for target detection, cons: difficult to generalize).
- Fusion at the **decision level**: post-classification combination using maps coming from different data sets, after classification processes based on their peculiar characteristics (pros: results at a higher level of abstraction, hence more easily interpreted, with class transitions in multi-temporal data sets, for instance, cons: computationally less efficient and potentially requiring different classification techniques for multiple data sets).

2.6 Chapter conclusions

This chapter is devoted to the discussion of several aspects of global urban area mapping based on EO data.

First, it has been highlighted that up to now no general definition of the term "urban" exists and that its specifications significantly vary among countries. From a remote sensing perspective most of the definitions are not suitable to describe the targets of remote sensing since the exploitation of EO data delivers physical characteristics only. However, until there is a uniform and generally valid definition, maps depicting information about human settlements will always be heterogeneous and difficult to compare since depending on the used data type different features are actually mapped. Since SAR-based maps refer mostly to built-up, while optical-based maps depict more easily impervious surfaces, the fusion of these data and/or the corresponding urban extent maps looks promising to cope with this issue, providing additional flexibility to the user.

In recent years a lot of effort has been made by the urban remote sensing community to extract accurate global urban maps or to improve existing ones with newly available, high-quality EO data. Good results could be achieved with multi-spectral as well as SAR data. However, both data types inherit very specific advantages and drawbacks with respect to

human settlement characterization. Due to the measurement of surface reflectance in a variety of regions of the electromagnetic spectrum multi-spectral EO data is generally well suited to differentiate between artificial and natural materials and structures. However, the quality of the data strongly depends on the atmospheric conditions at the time of image acquisition. Satellite-based active microwave measurements, on the other hand, are daylight and weather independent, but the resulting images are often more difficult to interpret due to the joint of two- and three-dimensional features of urban areas and data suitable for urban area extraction are usually acquired upon request or as a by-product of other researches. As a consequence neither of those data types is globally available in a consistent quality and/or quantity. Thus, a suitable fusion approach may be able to reduce the issues related to each specific sensor, improving the accuracy of the final fused map.

Finally, current products depict global human settlements at unprecedented levels of detail. However, due to the fact that these maps are generated exclusively using either one single optical or SAR sensor, they are not exploiting the complementary information of all available sensors. Once again, in order to take advantage of the full potential of today's wealth of multi-source EO the development and application of data fusion techniques are necessary. In this context not only information from different sensors, but also data from sensors with different spatial resolutions need to be exploited, leading to a framework offering solutions at the pixel-, feature-, and decision level.

Chapter 3

SAR data fusion for urban area mapping

3.1 Introduction

SAR data is without doubt very useful for urban area extraction purposes. However, depending on the location of the area of interest, the intended spatial resolution of the generated map, and project-related financial constraints, data sets precisely meeting all requirements for an urban area mapping task at hand might not be available. This is especially true for the pre-Sentinel era, when SAR data - apart from the notable exception of Envisat ASAR - have not been acquired globally in a continuous mode. As a consequence there is a strong need for efficient SAR data fusion techniques in order to make use of the prosperous, but also heterogeneous SAR data availability.

Additionally, the SAR data processing chains discussed in section 2.3.1 (KTH-Pavia Urban Extractor and UEXT) are tailored to specific input data, and are affected by different detection errors. Although SAR images at different spatial resolution depicting the same urban area cannot be considered as independent, the extracted patterns representing the settlements at coarse or fine spatial resolutions are different enough to allow assuming as an advantage the combination of intermediate and/or final results of the two chains.

Indeed, the idea discussed in this thesis chapter is to design data fusion techniques able to exploit SAR data at different spatial resolutions for urban extent extraction. Input data to such a procedure may have been recorded by the same sensor using different acquisition modes, or by different sensors. The fusion goal can be described as the idea to use finer resolution data to locally improve an originally coarse resolution extraction, hence increasing the level of detail of the final human settlement map.

The first part of this chapter is devoted to the presentation and discussion of an improved version of the UEXT algorithm (see section 2.3.2) considering multi-temporal, pixel level data fusion. Multi-scale data aspects are treated in the second part of this chapter, and a general framework for multi-resolution SAR data fusion for the purpose of urban area extraction is proposed.

3.2 Fusing multiple SAR images at the pixel level*

When multiple (SAR) images of the same area are available, classification algorithms can be applied to multiple data stacks. Although this is a very quick way to exploit data sequences, it requires more and more sophisticated algorithms, able to deal with data space with more and more dimensions, hence relatively sparse data clouds. With increasing data dimensionality classic remote sensing classification algorithms like maximum likelihood [66] tend to fail to exploit the full potential of the data. More complex approaches (e.g. Support Vector Machines [67],[68], neural networks [69], [70] or Random Forest [71],[72]) on the other hand, are better suited to obtain accurate mapping results with hyperdimensional data.

Another relatively simple class of approaches, best applied to images from the same sensor and/or captured with the same image mode, includes methods to combine these images into a single one using selection or combination strategies in order to either filter out disturbing features or to highlight objects of interest. These fusion approaches are manifold and strongly depend on the used data type and on the application objective. The most commonly used techniques to combine images from the same sensor and resolution into a single fused image range from statistical/numerical methods [73],[74],[75], to adaptive pixel selection rules [76], to multi-temporal filtering approaches [77].

If the focus is on multiple SAR data sets with the same resolution, looking angle and polarization (i.e., basically data from the same sensor acquiring in the same mode), a pixel level data fusion technique may be applied. First, each input image is smoothed using a speckle filtering technique (e.g. Lee filter [78]) in order to reduce the noise within the data. Then, a fused image is built, by computing the per pixel average among all speckle filtered input images.

Figure 3.1 illustrates the smoothing and highlighting effect of the applied multi-temporal pixel level fusion technique. It shows Sentinel 1 SAR data scenes captured in different phenological seasons (a-c), as well as the combined multi-temporal data stack from the same sensor. A visual comparison shows that the stacked image contains less noise, and that artificial structures (bright areas) are more discriminated with respect to surrounding natural areas that undergo different phenological phases throughout a year. For instance, the building structures located in the lower left part of the SAR data subsets can be easier identified in the multi-temporal stack. Also, the water body in the upper left area and the small river across the scene are barely recognizable in the single SAR images while the landuse for these specific objects can be easily determined by looking at the fused data set. A similar effect can be observed for agricultural fields located around the bright artificial structures and in the right part of the image. The single date images contain a lot of noise due to the environmental characteristics at the time of image acquisition. In the fused image, on the other hand, the same area appears to be more recognizable.

*A large part of this section comes from: G. Lisini, A. Salentinig, P. Du, and P. Gamba, "SAR-based global urban extents extraction: from ENVISAT to Sentinel-1", a paper submitted to *Remote Sensing of Environment* in March 2016

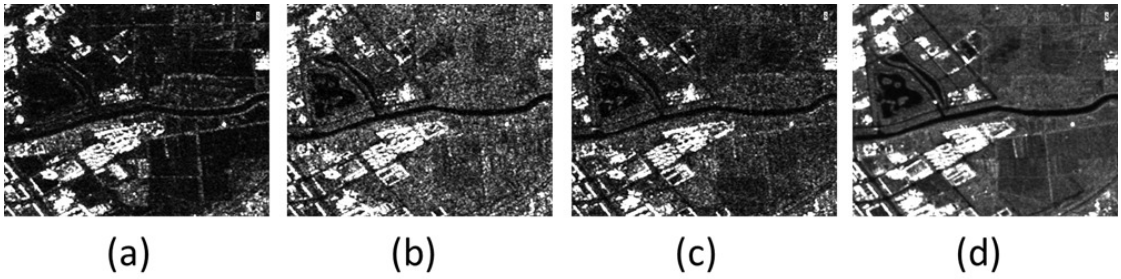

Figure 3.1: Sentinel 1 SAR data over a sparse urban area: (a) February 2015, (b) June 2015, (c) November 2015, (d) multi-temporal stack February-December 2015.

However, this approach requires to apply a preprocessing step in which a number of SAR data are merged into a single image to improve the discriminability of natural and artificial objects. In fact, this is the most crucial step for the successful application of the algorithm since the method is based on the assumption that, in a multi-temporal stack of SAR imagery, urban areas are highlighted because they are structures much more stable in time than the surrounding vegetation, which usually undergoes significant changes throughout the annual phenological cycle.

3.2.1 Improving UEXT considering multi-temporal data

To overcome the issues presented in section 2.3.2 for the UEXT algorithm, this technique needs to be made more adaptable to different situations and environments, and in an automatic way. The proposed solution is to use the number of acquisitions that are combined to obtain the input image as a proxy to the level of discriminability between urban areas and the surrounding natural environment. Specifically, the input SAR layer is analyzed in a three different ways. The first case (or "Case 1" for short) corresponds to areas where a large number of acquisitions are combined, and significantly large discrimination of urban areas with respect to the surrounding natural landscape is achieved. On the contrary, the "Case 3" corresponds to a small number of acquisitions, down to a single SAR image, and correspond to the most challenging scenes. Finally, all other scenes are labelled as "Case 2" areas. For instance, since the input ASAR WSM data sets corresponds to a maximum of 120 combined images, to define these three cases the two values of 40 ("acquisition-high") and 20 ("acquisition-low") may be selected.

3.2.1.1 Adaptive seed selection and aggregation

In all three cases the extraction procedure (composed of the seed selection and the aggregation steps) follows the preliminary version of the extraction procedure in [31], but with different sets of parameter values. The urban seed selection procedure requires two thresholds, one set to a (relatively) high value to select backscatter values, and a lower one, to stop the region growing procedure. Additionally, the pixels around the seed(s) can be disjoint: the extent of the area to be considered while looking for other parts of the

same settlement is defined by the "kernel-around" parameter. Finally, it is assumed that the maximum slope on which an urban area may be located is limited and the portion of urban areas on steep areas is small.

Table 3.1: Procedure parameters

Parameter	**Case**	**Default value**	**Unit**	**S1 value**
max-tile	1-2-3	1500	pixels	
tile-overlap	1-2-3	20	pixels	
kernel-acq	1-2-3	5	pixels	
kernel-around	1-2-3	3	pixels	
seed-th-max-slope	1-2-3	38.7	°	
slope-th-high	1-2-3	16.7	°	15
slope-th-low	1-2-3	11.3	°	7
slope-th-sea	1-2-3	0.004	°	
slope-th-histog	1-2-3	5.7	°	
percentage-slope	1-2-3	50	%	
perc-slope-sea	1-2-3	80	%	
upper-th-histog	1-2-3	0.4	SAR intensity	
blob-fill-th	1-2-3	100	pixels	1400
blob-dilate-pixel	1-2-3	2	pixels	
max-blob-mountain	1-2-3	50	pixels	
acquisition-high	1-2	40	scenes	0
acquisition-low	2-3	20	scenes	0
seed-th-high-1	1	1.5	SAR intensity	0.18/0.04
seed-th-high-2	1	0.5	SAR intensity	
around-seed-high	1	0.35	SAR intensity	0.05/0.015
seed-th-middle-1	2	1.2	SAR intensity	
seed-th-middle-2	2	1.0	SAR intensity	
around-seed-middle	2	0.4	SAR intensity	
seed-th-lowest-1	3	1.7	SAR intensity	
seed-th-lowest-2	3	1.5	SAR intensity	
around-seed-lowest	3	0.5	SAR intensity	

The extraction for "Case 1" areas is performed sequentially by using two different seed values ("seed-th-high-1&2"), but the same value for the region growing parameter ("around-seed-high"), leading to two distinct urban classes, referring to wider and smaller human settlement, respectively. These urban extractions are very reliable, because of the high settlement discriminability due to the large number of combined SAR scenese.

For "Case 2", the same scheme is used (three parameters for two urban classes), while the parameters for the region growing are selected automatically according to the average backscatter in the neighborhood of the pixels to be considered. This *local* average value is obtained considering only backscattering values of non-urban areas, selected according to the following rules:

- the backscatter is low enough (smaller than the "upper-th-histog" parameter), AND

- the corresponding slope value is small (smaller than the "slope-th-histog" parameter).

These rules aims at excluding pixels in correspondence of candidate urban centers (whose average backscatter value is not of interest here), as well as of mountainous or marine areas that may lower (or raise in some cases) the average backscatter value.

Finally, for "Case 3" scenes, the same procedure than for the "Case 2", only with different seed and stop thresholds (see table 3.1), is performed.

The routine applies the extraction procedure to a specific area according to the selected case. Since this means that different parameters are considered from area to area, discontinuity effects may arise on the boundaries between geographical areas with (very) different acquisition numbers. In oder to reduce this issues, a moving window median filter is eventually applied.

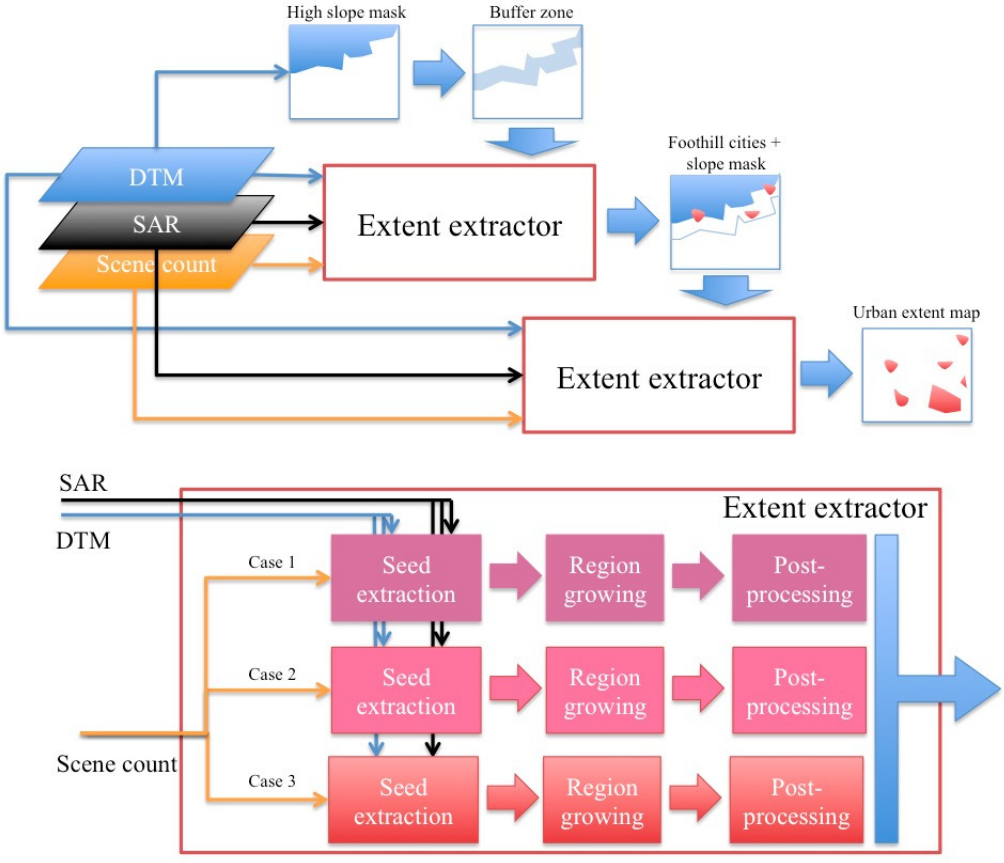

Figure 3.2: Graphical representation of the workflow of the improved UEXT procedure.

3.2.1.2 Post-Processing

Once a candidate settlement has been extracted, and if a DEM is available, its portion on high slopes is computed, to be compared with a maximum value stored in the "max-blob-mountain" parameter. A first way to recognize whether this is a real urban area is to compute whether this correspond to a large percentage of the urban area, compared with

a "percentage-slope" parameter, this candidate settlement is discarded. However, there are cities, especially large ones, with a portion close to a similarly large mountainous area. For these cities (e.g., Genoa in Italy, Rio de Janeiro in Brazil, and Chongqing in China) it may be possible that the ratio between urban and mountainous areas exceeds the "percentage-slope" threshold, resulting into a large amount of false negatives.

To reduce the impact of this issue, we are forced to include a step which is actually a pre-processing, as opposed to a post-processing step. Indeed, as shown in figure 3.2, the extent extraction procedure is run twice: initially only areas close to steep slopes are considered. Then they are masked out, and the algorithm works on relatively flat areas. Specifically, initially only the candidate areas originating from seeds in a frame around steep slopes, defined by the "slope-th-low" parameter, are considered. These candidate areas are checked more accurately to avoid discarding small towns on the slopes of hills or mountains, and to better discriminate them from false positives caused by, e. g. , mountain spurs. To this aim, a candidate settlement is retained and actually assigned to the urban class if the extension of the portion outside the steep slope is more than 50 % of the total area, or if the area of this part is small in absolute terms (i. e. smaller than the above mentioned "max-blob-mountain" value), but not in relative terms, i. e. the ratio between the rest of the candidate area and this portion is below the "percentage-slope" value. The first condition is also useful to avoid that limited steep slope areas inside extensive urban centers affect too much the extraction results. At the end of this procedure, pixels in steep slope areas, their framing area, and the candidate areas recognized as urban settlements are masked out. Accordingly, other candidate cities and large mountain ranges are disconnected. To these candidate areas, only the much simpler check on the "percentage-slope" value is applied.

The last part of the post-processing procedure is used to select whether "urban areas" should be equated to "built-up areas" or they include parks, ponds, and bare soil areas within their boundaries. If the latter definition is considered, two parameters are considered. The first one, "blob-fill-th", defines the minimum size of an extracted settlements to be subject to the procedure; the second parameter, "blob-dilate-pixel", defines the size of the structuring element for a dilate morphological operation, to select the area including the initially set city boundaries within which the previously mentioned internal elements will be integrated (via flooding) into the final settlement extents. These parameters are the same for all classes, and this step is therefore applied uniformly everywhere.

As a final note on computational costs and parallelization, the overall procedure has been designed to manage files of considerable size by means of tile splitting. Specifically, the procedure works on portions (tiles) of the original data set, saving and combining the partial results, and in the future running on parallel processors. The procedure automatically computes the optimal tile size considering the number of rows and columns of the image and according to the average extension of human settlements in the area(s) under test. In order to avoid edge phenomena (e. g., due to truncations by the flooding routine), each tile is enlarged by a framing area. The results eventually copied to the file will only

be those contained in the tile, excluding the framing area. Indicatively, to process a scene of about 100000 x 30000 pixels on an Intel i7 processor and 1 Gb of RAM (although this is not a limiting factor), the software runs for about 2 h.

3.2.2 Results and discussion

The UEXT procedure has been tested extensively on the multi-temporal pixel-level fused data for a number of different test areas. The first set of experiments was run to visually assess the new version of the procedure with respect to the one presented in [31] and the urban extents available from the latest release of the ESA CCI Land Cover database [56]. To this aim, Moscow, Mumbai, Paris, and Beijing were considered. Figure 3.3 provides a visual comparison between the previous extractions (in blue) and the new one (in red), superimposed to the SAR input layer. The new extraction is significantly larger and more compact than the original one, and the extraction includes finer details of the urban structure, as well as a better characterization of small settlements and the urban-rural fringes. This is evident, for instance, for Beijing (figure 3.3d), where the new procedure improves the extraction of the outskirts and reveals more details of the inner urban structure. An additional visual comparison of the new procedure extractions with the CCI urban layer, obtained from MERIS multispectral data, for the same four cities (figure 3.3e-h) allows to appreciate the fine characterization of the extents extracted using ASAR WSM data.

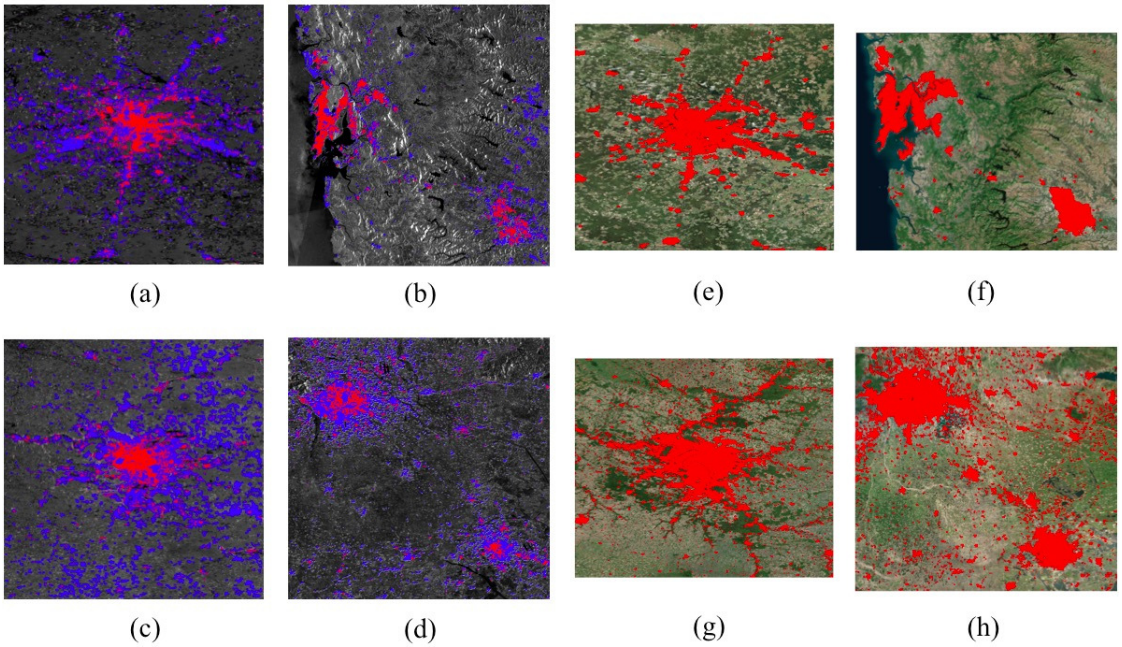

Figure 3.3: A comparison between the original (blue) and the improved (red) urban extent extraction results for a) Moscow, b) Mumbai, c) Paris, and d) Beijing and Tianjin, superimposed on the original ASAR WSM data. For visual comparison, in (e-h) the extents of the latest CCI urban layer, superimposed to optical data, are reported.

As a first quantitative test of the difference between the original and the new proce-

Table 3.2: Accuracy values for the preliminary versus current global urban extents release in the test areas of the 2011 ESA RR.

	From [31]		Improved version	
Test site	**OA (%)**	**Omiss. (%)**	**OA (%)**	**Omiss. (%)**
North America	78.5	43.0	82.1	35.0
South East Asia	94.8	11.8	95.8	7.4
Western Europe	56.2	85.8	57.1	84.4
Africa	60.0	80.0	89.5	24.0
South America	68.1	63.8	86.0	23.8
Northern Asia	77.0	46.0	85.1	29.8

dures, the test areas selected for the 2011 RR, whose extraction results are presented in [31] were considered. These test sites refer to six locations, in Western Europe, Canada (North America), Brazil (South America), Africa, Siberia (Northern Asia) and South East Asia, and the evaluation of the extracted human settlement extents was run again using the new version of the UEXT algorithm. Table 3.2 presents the overall accuracy for these test areas compared with the original one reported in [31]. It is recalled here that the validation sets are composed for each location by 1000 randomly placed points in a 200 by 200 km area in the above mentioned regions (see figs 2, 4, 5 in [31]), manually checked in order to obtain 500 urban and 500 non-urban validated samples. As an exception due to the low urbanization of the test area, in Africa only 100 urban and 100 non-urban points were considered. Due to the fact that the manual validation of random samples is extremely cumbersome and also because the availability of HR satellite data necessary for the validation of the samples is limited in the test sites, the number of reference points is relatively low. However, since the same reference data have been used for the accuracy assessment of the old as well as the new version of UEXT they are valid for the purpose of evaluating the improvement of the algorithm.

The numerical results show that the new procedure outperforms the original one in all locations, and specifically in Northern Asia and South America, where the issues with the original approach related to the extraction of a limited portion of the total urban area (for an example in a different area but with similar properties, see the case of Moscow in figure 3.3a). For the test site in Africa, instead, the issue solved by the new approach refers to the low value of the backscattered field in urban areas. The better performance of the new release is confirmed by looking at the differences in the commission error percentages of the original and new extractions, both reported in table 3.2. It must be noted, however, that the urban extent extraction in Western Europe (a semi-arid area across the border between Spain and Portugal) is still poor. This issue is the main rationale behind the selection of the test sites for the 2015 Urban RR, as discussed later in this section.

A second set of quantitative experiments was then performed on a much wider area, i. e. the whole P. R. China. UEXT was applied to the input SAR layer of the whole country, using the original and the improved UEXT algorithms. In addition to showing

a visual example in figure 3.5, a quantitative comparison was performed by computing overall accuracy values at the province level. For each of the 31 provinces of Mainland China, the urban extents manually obtained from 2010 Landsat images by a team at Tsinghua University were used as reference urban areas . These extents do not correspond to every urban area in these provinces, as they depict only the most populous 663 cities of the country in the year 2010 [79]. Therefore, the overall accuracy values reported in figure 3.4 cannot be considered in absolute terms, because most of the commission errors are not actual errors at all. Instead, it is important to observe that there is a significant improvement between the results preliminary and the current versions of the UEXT layers.

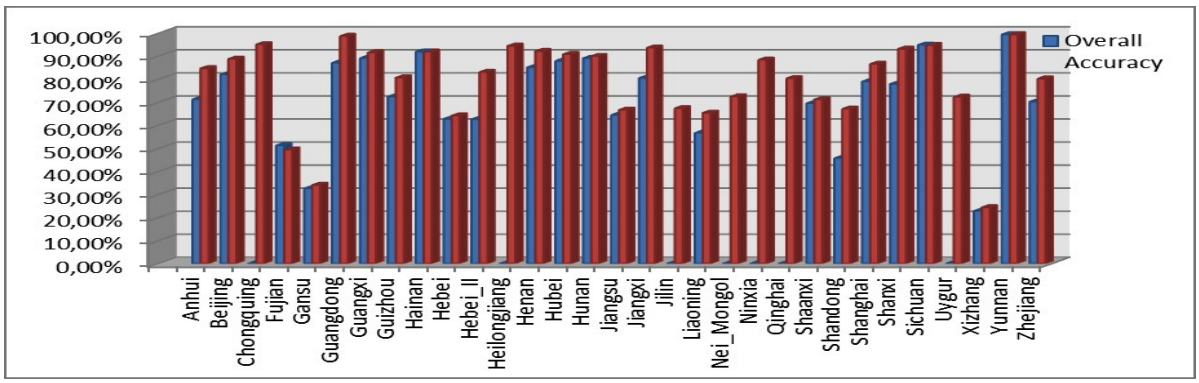

Figure 3.4: A comparison between the original (blue) and improved (red) UEXT extraction results for 31 provinces of P R. China, compared against validation data in [79]. Province boundaries and names according to the Global Administrative Areas project (http://www.gadm.org, accessed February 2, 2016).

Specifically, the graphical comparison at the province level proposed in figure 3.4 shows that the overall accuracy improves almost everywhere, with respect to the reference data, using the new version of the algorithm. Improvements are sometimes dramatic, because for some of the provinces no extracted settlements were available in the first version of the urban layer. As previously mentioned, figure 3.5 shows some details for a portion of the Guangdong province, where the extraction (in red) and the reference data (in white) are in agreement for the major towns, while the smaller ones are detected by the proposed procedure but are missing in the reference data. In figure 3.5 b we report a map of the same area as available from Google Maps ^{TM} for comparison.

Since the numbers in figure 3.4 are not representative of the true accuracy of the proposed approach, in order to quantitatively compare the extraction results, two of the provinces have been selected for a more detailed analysis. For these provinces an accurate ground truth was manually extracted from visual interpretation of freely available Landsat data by local experts. More specifically, for Jiangsu a total of randomly placed 700 samples, each one referring to an area of nearly 0.66 km^2, were considered. Half of these samples were automatically and randomly selected in rural areas, and half of the them within "likely urban" areas, obtained by means of the intersection among all available urban

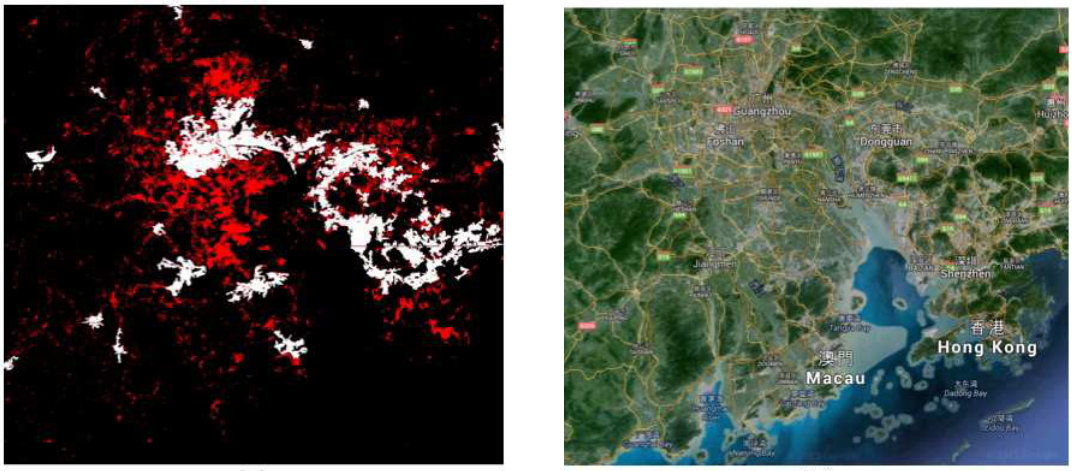

Figure 3.5: A more detailed comparison between the improved human settlement extent extraction results (red) and the extents in [79] (white) for a portion of the Guangdong province (a) and a map of the same area from Google Maps^{TM} for visual comparison (b).

global layers we were able to gather in these locations. Then, each of the "likely urban" samples was manually checked and assigned, according to the percentage of built up area in the sample, to one among 5 classes: no built-up area at all, built-up area covering less than 25 %, from 26 to 50 %, from 51 to 75 %, or more than 75 % of the sample. Finally, to validate the urban extent extractor, all the urban samples with 50 % or more of built-up area were assigned to a generic "urban" class, and a corresponding number of rural area samples were randomly selected. Eventually, 118 urban and 118 rural area samples were compared with the results of UEXT. Finally, a comparison between the sample classification by visual extraction and by automatic urban extent extraction using UEXT (both the original and the improved version) lead to the numbers reported in table 3.3, compared with those obtained under the same conditions for the Globcover and the MODIS 1000 urban layer. To avoid any regional bias, the same procedure was also applied to the Hubei province, considering a validation set of 123 urban and 123 rural samples, and the results are shown in the same table. It must be noted that the number of samples is different than for Jiangsu, because of the different size of the two provinces.

Table 3.3: Accuracy values of urban extent extraction for the Jiangsu and Hubei provinces

	Jiangsu		**Hubei**	
	OA (%)	$\hat{k}$	**OA (%)**	$\hat{k}$
Original UEXT	73.3	0.47	78.0	0.56
Improved UEXT	85.5	0.65	84.1	0.68
Globcover	65.3	0.31	69.5	0.39
MODIS 1k	71.6	0.43	76.0	0.52

This table shows that the SAR global urban layers perform very well in terms of overall accuracy values, which are always larger than the ones by the latest version of the Globcover urban layer, as well as the MODIS 1000 layer. Still, the new layer improves with respect to the preliminary results in absolute terms in both provinces.

3.2.2.1 ASAR versus S1-A

Since the new version of the urban area mapping approach is meant to work on ASAR but also on S-1A data, the last set of tests was designed to check the performance of the UEXT algorithm on S-1A data sets. The aim is not only to check that UEXT works on these data, but also to understand which are the improvements that can be obtained by considering data from this sensor as opposed to ASAR data. As test sites, data sets provided by ESA in the framework of the 2015 RR on Urban Mapping have been used. The 2015 RR was focused on arid areas, exactly were the original extractions starting from ASAR data were not as good as expected. Test areas were selected in Portugal, Tunisia, Egypt, Israel and Turkey. S-1A data sets collected between October and December 2015 in the Interferometric Wide Swath mode (spatial resolution of 20 m) were considered, and subject to the same pre-processing steps mentioned mentioned in [31]. Figure 3.6 shows a visual comparison of the multitemporal input layers for the Nile river delta and the city of Cairo (Egypt) computed from a combination of ASAR WSM and S-1A data.

The quantitative evaluation for these test sites has been performed considering manually extracted samples, like for the Jiangsu and Hubei province data sets. The number of validation points is 239, 235, 488, 274, and 153 for Portugal, Tunisia, Egypt, Israel and Turkey, respectively, and the overall accuracy and $\hat{k}$ values are reported in table 3.4. The results for the ASAR WSM data sets refer to the improved UEXT version only.

Table 3.4: Accuracy values of urban extent extraction for the 2015 RR test sites

	ASAR WSM				**S1-A**			
	OA(%)	$\hat{k}$	**Comm.(%)**	**Omis.(%)**	**OA(%)**	$\hat{k}$	**Comm.(%)**	**Omis.(%)**
Portugal	72.8	0.46	2.5	51.9	86.8	0.74	11.7	14.6
Tunisia	52.8	0.06	0.85	93.6	90.0	0.80	5.5	14.5
Egypt	54.9	0.10	2.25	87.9	80.0	0.60	26.0	13.9
Israel	76.8	0.54	2.92	43.4	82.8	0.66	17.2	17.2
Turkey	64.4	0.29	0.0	71.2	93.1	0.86	4.6	9.2

The numbers show that, as expected, the use of finer resolution SAR from S1-A improves the overall accuracy by reducing the omission error percentages. Instead, commission error is getting worse, but the net advantage of using a finer resolution sensor in these test areas is substantial. It must be added that, although not reported in the table, the values for the most recent Globcover urban layer are in line with (actually, slightly worse than) those of the ASAR WSM urban extraction, hence the extraction from S1-A definitely improves with respect to the current data sets from ESA.

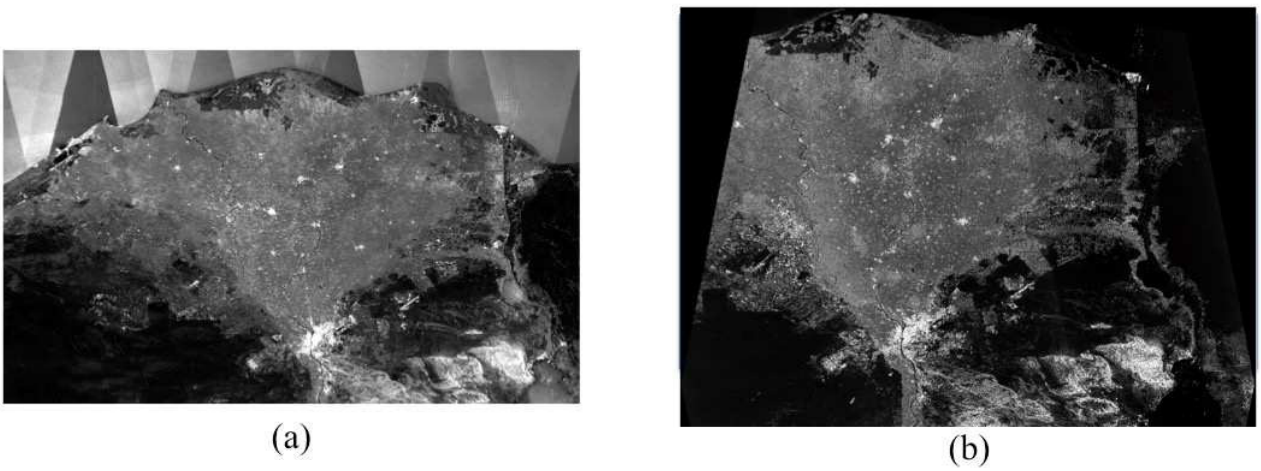

Figure 3.6: Input SAR layers obtained from as a multi-temporal combination of ASAR WSM (a) or S-1A (b) data sets for the Nile river delta and the city of Cairo (Egypt).

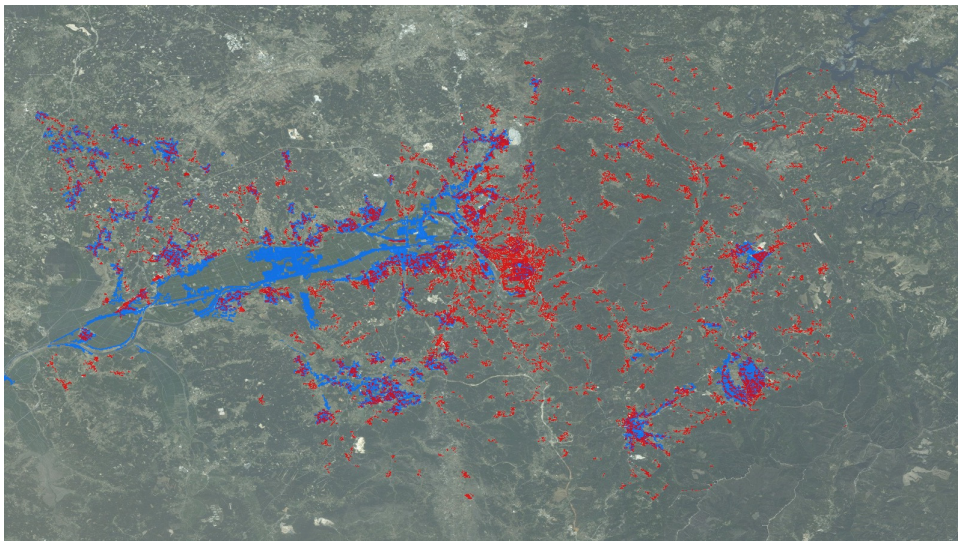

Figure 3.7: Urban extent for the town of Coimbra (Portugal) from S-1A data in blue, superimposed on the urban extents from European Urban Atlas in red.

Finally, and in order to test more precisely the extraction at the town level, a last test was performed, considering the small Portuguese cities of Faro, Coimbra and Aveiro, whose detailed urban land cover maps, at a finer resolution than any of the products considered in this work, can be obtained thanks to the European Urban Atlas (*http://www.eea.europa.eu/data-and-maps/data/urban-atlas*). After translating Urban Atlas land cover maps into binary urban/rural maps, 500 urban and 500 non-urban randomly placed samples were extracted for each location. Since the resolution of the remotely sensed data used for the Urban Atlas is 5 m, there is no point in evaluating ASAR-based maps, and only the comparison with the extents extracted from S-1A data was performed. The overall accuracy of S-1A

data processed by the improved UEXT algorithm is 78.5 %, 69.5 %, and 78.8 %, for the three cities of Faro, Coimbra and Aveiro, respectively. As usual, commission errors are far lower than omission errors, but this not always the case, as for Alveiro for instance the opposite is true. A visual analysis of the results for the worst case (Coimbra), compared with the Urban Atlas data set in figure 3.7, shows clearly this tendency.

3.3 Fusing SAR data at multiple spatial resolutions*

A more general problem than the one discussed in the previous portion of this chapter is the need to develop methodologies to fuse SAR data coming from the same or different SAR sensors with different spatial resolution, i.e. multi-resolution SAR data fusion methods.

Let's therefore focus on the case when, instead of having the same spatial resolution, the available data sets contain a variety of images at different spatial resolutions. As a first statement, it is possible to note that, depending on the research question and data availability, multi-scale data fusion approaches vary a lot.

To this aim, different results using different data fusion algorithms at the pixel, feature or decision level are expected [80],[65]. The simplest and most straightforward approaches to pixel level SAR fusion are statistical/numerical methods. They range from arithmetic combinations to difference and ratio images. For this kind of fusion it is especially important to harmonize the statistics of the input imagery in order to produce reasonable fusion results. A very effective pixel-based numerical multi-temporal fusion approach for SAR data has already been presented in section 3.2, where Envisat ASAR WSM or Sentinel 1 data data are fused through filtering and averaging, aiming at highlighting urban areas. Alternatively, statistical methods may be applied (e.g. Principal Component Analysis (PCA) in [81] and [82]), and in this case thoughtfully determined weights based on data reliability and quality can significantly improve the fusion. In multi-scale SAR fusion every image typically serves as input for PCA, which transforms the multivariate data set into a new one, consisting of uncorrelated combinations of the original variables. Other pixel-based fusion techniques exploiting multiple scales are based on pyramidal signal representations [83]-[84]. Finally, multi-scale Kalman filter has proved to be a very powerful tool for multi-scale SAR data fusion, where spatial resolution is considered as the independent variable as opposed to time [85], [86].
All these methods suffer from a few limitations:

- if harmonization of the image statistics is conducted in the pre-processing step, the original Digital Numbers (DN) of at least one image are modified and this may result in loss of information;

- depending on the intended scale of analysis, the inclusion of coarser or finer resolution data on the pixel level could lead to unwanted effects like an increase of the salt and

*A large part of this chapter comes from: A. Salentinig, and P. Gamba, "A general framework for urban area extraction exploiting multi-resolution SAR data fusion", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 9, Issue 5, pp. 2009-2018, 2016

pepper effect or smoothing of typically heterogeneous urban patterns in HR SAR imagery;

- the accuracy of the urban area results from fused multi-resolution SAR data depends not only on the fusion product but also on its suitability to the specific urban area extraction algorithm.

Pixel level methods have also some critical advantages, since they are straightforward and easy to implement. Due to the fact that information is combined before the data is processed in order to differentiate between urban and rural areas, error propagation caused by defective classification is effectively avoided.

Moving from pixel-based to feature-based fusion, multi-scale feature level fusion approaches focus on more abstracted features generated from input SAR images. In the proposed framework, aimed at urban area extraction, a feature can be anything that adds useful information about the appearance of human settlements. The most common ones are textural measures [87]-[88], as well as fuzzy sets depicting the probability that a pixel belongs to the urban class [89]. While there are standard image processing techniques to compute texture measures, approaches to define more abstract spatial features like urban probabilities significantly vary. Hence, the final built-up area detection result depends on the performance and reliability of the methods which generate the spatial features. Another drawback is that hard thresholds might be necessary for a classification of the fused features into urban and non urban areas. This limitation is especially valid for applications where thresholds are selected by human analysts whose qualification, experience and familiarity with the data and the study area directly impact human settlement characterization results. As an advantage, however, spatial features are generally easier to understand, as it is more natural to interpret more abstracted information instead of digital numbers. In literature the most common methods are algebraic combinations [89], probability theory based approaches [90], Dempster-Shafer evidence theory [91], neural networks [92], and knowledge based approaches [93].

Finally, human settlement extent extraction from multiple resolution SAR data may be improved by decision fusion. i.e. by the process of combining already existing urban/non urban maps at different resolutions. The main advantage of this approach is that completely harmonized information about the occurrence of human settlements is available through binary classifications. The methods mentioned for feature level fusion approaches may be also used to fuse urban/non urban maps, considered as binary features. Otherwise, other approaches, such as logical operators [94] or voting strategies [95] may be used.

3.3.1 Overall structure of the proposed approach

Given the number of possibilities for multi-scale data fusion, the overall structure of a common framework for the fusion of SAR data sets in the context of urban area extraction may be extremely complex.

The one proposed in this thesis, formulated and published in [96], is graphically presented in figure 3.8. It includes several processing steps, and relies on fusion techniques at pixel, feature and decision level, as discussed in the previous section. Within the proposed framework five different SAR data fusion approaches have been explored and compared. Therefore, the following paragraphs are devoted to a description of this framework, to the presentation of the developed SAR data fusion techniques, as well as to a comparative evaluation and discussion of the data fusion approaches and results. The resulting conclusions may definitely help in the selection of an appropriate SAR data fusion method for the problem of a more accurate delineation of human settlement extents.

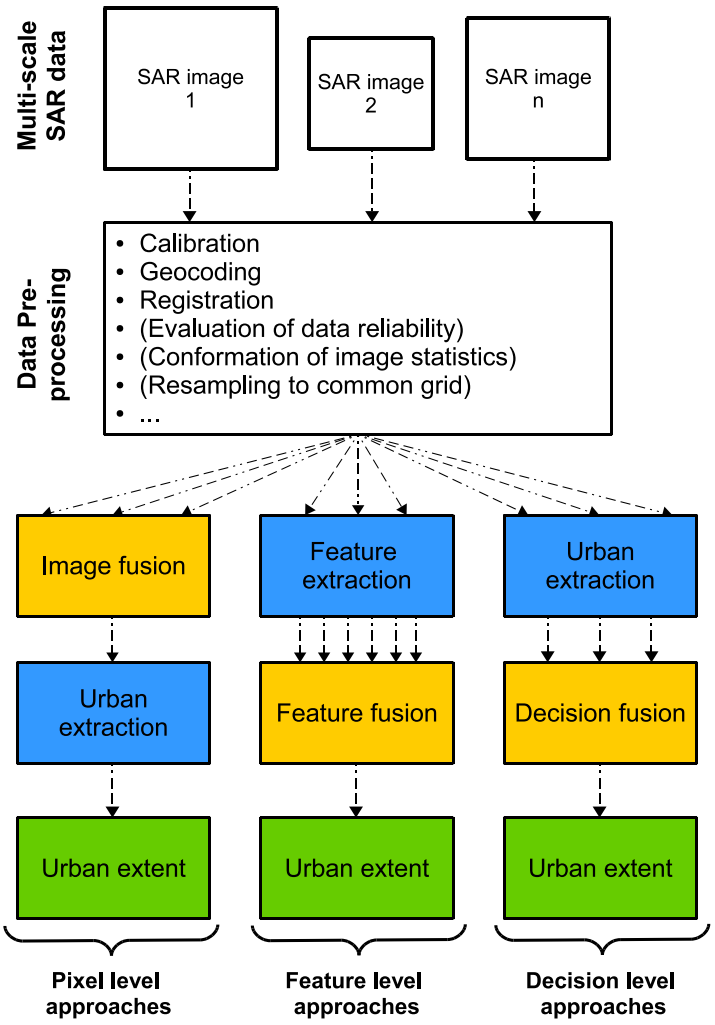

Figure 3.8: Graphical representation of the general framework for multi-scale SAR fusion for the purpose of built-up area extraction.

3.3.2 Pre-processing steps

First of all, due to the differences among multi-resolution SAR data, a number of pre-processing steps are necessary for an effective combination of the imagery. In this work it is assumed that steps applied to each of the data by itself are performed at the data

providers premises. Instead, steps required because of the fusion procedure are performed within the current framework.

Among the first group, the most important steps are calibration and geocoding. Calibration is typically performed during the generation of a SAR product. In the context of urban area extraction it is not advisable to use raw data due the inconsistency of the uncalibrated signal. Before fusing SAR data sets coming from different sensors and at different spatial resolutions all input data sets need to be thoroughly calibrated with respect to their intrinsic sensor and signal acquisition properties.

Similarly, data providers may include accurate geocoding of the data sets, and an important prerequisite for a reasonable fusion of SAR data is that this step is performed, too. It is worth noting, however, that the differences in spatial resolutions of scenes makes this task demanding, especially for SAR images. Nevertheless, in order to avoid misclassifications due to poor co-registration this task needs to be conducted with due care. In this work no co-registration procedure was necessary because the used data were already precisely geolocated and therefore fit each other perfectly with respect to geographic location at data delivery (see section 3.3.8 for further details on used data).

The methods discussed in the following sections refer to multiple data fusion techniques. Specifically, the following methods are considered:

- pixel level fusion techniques, such as Histogram Matching/Weighting, Discrete Wavelet Transforms and Multi-scale Kalman Filters;
- feature level fusion approaches, such as Fuzzy Fusion of urban probabilities;
- decision level fusion techniques, such as logical operators for the combination of urban maps at different scales.

The rationale for this selection of data fusion methods lies in the fact that they are reasonably straightforward and easy to understand and at the same time in line with the state of the art of multi-scale SAR data fusion approaches.

In order to harmonize the notation among all these algorithms, let's denote with $\mathbf{X}^{(m)}$ a SAR image at the m-th resolution level in a multi-scale data set, where $m = 1$ denotes the finest scale level, and higher values imply coarser resolutions. Furthermore, let's call $\mathbf{Y}$ the image fusion result and $\mathbf{M}$ the final urban map. Note that $\mathbf{Y}$ and $\mathbf{M}$ are assumed at the finest possible resolution level (hence no m, since $m = 1$).

3.3.3 Histogram matching/weighting (HMW)

As described above, it is important to harmonize the image statistics of input SAR images in order to conduct reasonable pixel level fusion. In this work, a standard histogram matching method (for details see [97]) has been applied in order to generate statistically harmonized images $\dot{\mathbf{X}}^{(m)}$ resampled on a common (fine) grid. Due to this common grid, the combination of information is easy to implement and the fused image is obtained by a weighted sum: $\mathbf{Y} = \sum_m w_m \dot{\mathbf{X}}^{(m)}$, where $\sum_m w_m = 1$.

To select the weights (see table 3.5 for the weights used in this work), it has been found that, for multi-temporal SAR image stacks, the number of acquisitions is the most important factor. Specifically, a positive correlation between number of acquisitions in the multi-temporal stack and the weight of this input into the HMW fusion has been observed. Furthermore, experimental checks revealed that topography, climate region, and SAR polarization only moderately influence the results, while incidence angle, ascending/descending orbits, and the noise equivalent sigma zero value have a negligible impact.

3.3.4 Discrete Wavelet Transform (DWT)

In general, DWT is a common way of introducing multi-scale signal representation and is built upon a filter bank, consisting of two filters. An image $\mathbf{X}^{(m)}$ is filtered first along rows and then along columns by means of a high-pass and a low-pass filter. After the filtering, the output is down-scaled by a factor of two. The result of this procedure is a set of four sub-images containing horizontal, vertical, and diagonal details, as well as a low pass version of the original image (the context image, or $\mathbf{CI}^{(m+1)}$). Note that the $m+1$ index is due to the downsampling procedure, and implies that this techniques is valid for input multi-scale SAR data sets whose spatial resolution differs by an integer power of 2.

A way to merge two images by means of DWT fusion is to start from the finer resolution image, apply the DWT and substitute the context image with a combination of itself and a histogram matched copy of the coarser image. In a consecutive step, inverse DWT is applied in order to generate a fused image product at the finer level. In case of a multi-resolution data set, this procedure can be applied recursively starting from the two coarsest resolution images, assumed at scale i and j, respectively ($i < j$). Specifically, the i-th scale is decomposed using DWT with the required number of levels, until the j-th resolution scale is reached. The resulting contest image $\mathbf{CI}^{(j)}$ is then substituted by the average of this value with $\dot{\mathbf{X}}^{(j)}$, the histogram matched version (with respect to $\mathbf{CI}^{(j)}$) of $\mathbf{X}^{(j)}$. Then, inverse DWT is performed in a consecutive step in order to generate a fusion result at at the i-th scale. The procedure is then iterated using the fused result and the input image with the finest resolution among those still to be fused. This process ultimately results in a fused image $\mathbf{Y}$, input to urban area extraction algorithms.

3.3.5 Multi-scale Kalman filter (MKF)

The MKF algorithm belongs to the group of pyramidal techniques based on multilevel trees. Here we follow [85] and [86] where, in contrast to traditional applications of Kalman filters, spatial resolution is the independent variable as opposed to time. At the coarsest resolution the signal is sampled at the pixel level, that corresponds to 4 pixels at the next finer level and, in general, to $4m$ values at the m-th resolution level. MKF is capable of obtaining optimal estimates of the state $\mathbf{S}^{(m)}$ described by the multi-scale model using

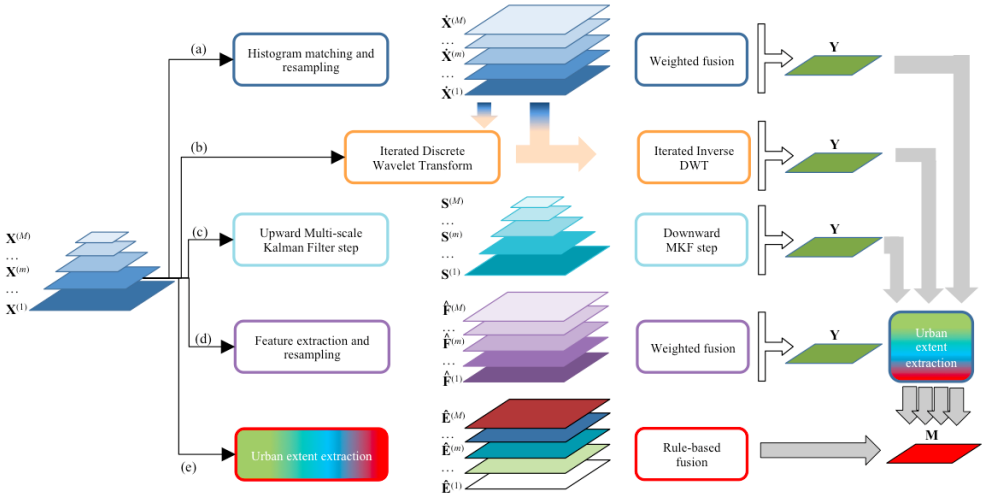

Figure 3.9: Graphical representation workflows of the different fusion techniques implementing the framework of figure 3.8: (a) HMW; (b) DWT; (c) MKF; (d) Fuzzy fusion; (e) logical operators.

observations $\mathbf{X}^{(m)}$ at a hierarchy of resolutions. The general scheme consists of two steps: a downward step (from coarse to fine resolution) and an upward step (from fine to coarse resolution). The downward model is given by $\mathbf{S}^{(m)} = \mathbf{S}^{(\gamma m)} + \mathbf{W}^{(m)}$ and $\mathbf{X}^{(m)} = \mathbf{S}^{(m)} + \mathbf{V}^{(m)}$, where γ_m specifies the parent level with respect to the m-th one.

Since $\mathbf{S}^{(\gamma m)}$ represents the scene at a coarser resolution than $\mathbf{S}^{(m)}$, it can be considered as a prediction term for the finer level. New knowledge is transferred from one scale to the next, and this is expressed in the model by $\mathbf{W}^{(m)}$. The noisy measurements of the state $\mathbf{S}^{(m)}$ and the equations from the downward model form the state estimation problem. In order to conduct state estimation, covariance matrices of the state and measurements are necessary and can be computed according to $\mathbf{P}_x^{(m)} \equiv E\left[\mathbf{S}^{(m)} \cdot \left(\mathbf{S}^{(m)}\right)^T\right]$. The upward step on the other hand is $\mathbf{S}^{(\gamma m)} = \mathbf{D}^{(m)} \cdot \mathbf{S}^{(m)} + \mathbf{W}^{(m)}$, where $\mathbf{D}^{(m)} = \mathbf{P}_x^{(\gamma m)} + \left(\mathbf{P}_x^{(m)}\right)^{-1}$.

3.3.6 Fuzzy fusion

Feature-level fusion needs spatial features extracted using specific feature extraction algorithms from each data set $\mathbf{X}^{(m)}$ at its own resolution. In this approach, the extracted features $\mathbf{F}^{(m)}$ contain information about the likelihood of a pixel/object/area being urban, like, for instance. the membership degree of a pixel to the urban class. A combination of these features may be obtained according to the very simple formula:

$$\mathbf{Y} = \sum_m w_m \, \hat{\mathbf{F}}^{(m)}.$$

In this work, the $\mathbf{F}^{(m)}$ are selected, for each SAR input data set, as one or more intermediate products of the urban area extraction algorithms in [44] and [31]. Each of these

intermediate products is obtained at the scale of the corresponding input, and $\mathbf{Y}$ is obtained at the finest resolution by resampling each $\mathbf{F}^{(m)}$ to $\hat{\mathbf{F}}^{(m)}$ via a nearest neighbours technique. Eventually, a binary urban extent map is generated by simple thresholding, while the weights are -in contrast to the HMW technique- set as proportional to the overall accuracy of the stand-alone urban maps extracted from the input data (see again table 3.5 for the weights used in the test cases discussed in next Section).

3.3.7 Logical operators

As its name implies, in decision-level SAR fusion the decision whether a pixel belongs to the urban class of not has already been made. Through combination of existing urban extractions it is expected to improve the accuracy of built-up area extraction and delineation results. To understand the relevance of fusing maps from more sensors, two simple combinations of the urban extents maps have been performed, i. e. the union (logical OR) and the intersection (logical AND) of the binary stand alone extractions $\mathbf{E}^{(m)}$ - once again resampled to the finest spatial resolution and denoted by $\hat{\mathbf{E}}^{(m)}$:

$$\mathbf{M}_i = \prod_m \hat{\mathbf{E}}^{(m)}$$

and

$$\mathbf{M}_u = \begin{cases} 1 & if \quad \sum_m \hat{\mathbf{E}}^{(m)} > 0 \\ 0 & \text{otherwise} \end{cases}$$

3.3.8 Experimental results on four megacity test sites

For our experiments four *megacities* showing tremendous growth rates were selected as test sites: Beijing, Guangzhou and Shanghai (P.R. of China), and Sao Paulo (Brazil). Those cities are characterized by a heterogeneous inner-urban compilation of natural and artificial surfaces (e.g. parks, lakes) and are therefore suitable to assess the performance and robustness of the applied SAR fusion and urban area extraction approaches. Specifically, in this work a multi-scale SAR data set consisting of two spatial resolution levels has been used. The ASAR instrument on board of the already retired ENVISAT satellite operated in C-band and acquired imagery in five different modes, aiming at various applications and scales of analysis. In urban remote sensing, data acquired in Wide Swath Mode (WSM) as well as imagery captured in Image Mode Precision (IMP) have proven their usability at the global [31] and local scales [44], respectively. While IMP scenes between 2007 and 2010 were singularly considered, due to the possible different acquisition parameters, WSM data have been acquired in the same way between 2008 and 2010, and can be merged into a multi-temporal composite, with a higher discriminating power between artificial and natural landscape elements [31]. Table 3.5 summarizes the characteristics of the data used in this work.

The main reason for the selection of the above mentioned test areas is the availability of reliable reference data from independent investigations [79]-[98]. These data set were

Table 3.5: Data specifications

Test site	Acronym	Acquisition date	Num. of images	Res. [m]	Polariz.	Size [pixel]	w_m HMW	w_m Fuzzy
Beijing	BJ-WSM	2008-2010	17	75	VV/HH	512 × 512	0.6	0.52
	BJ-IMP1	01/07/10	1	30	VV	1024 × 1024	0.4	0.48
	BJ-IMP2	05/08/10	1	30	VV	1024 × 1024		
Guangzhou	GZ-WSM	2008-2010	11	75	VV/HH	512 × 512	0.5	0.5
	GZ-IMP1	14/03/2009	1	30	VV	1024 × 1024	0.5	0.5
	GZ-IMP2	18/05/2009	1	30	VV	1024 × 1024		
Sao Paulo	SP-WSM	2008-2010	33	75	VV/HH	512 × 512	0.7	0.55
	SP-IMP1	01/02/07	1	30	VV	1024 × 1024	0.3	0.45
	SP-IMP2	16/01/2007	1	30	VV	1024 × 1024		
Shanghai	SH-WSM	2008-2010	13	75	VV/HH	512 × 512	0.5	0.5
	SH-IMP1	11/05/09	1	30	VV	1024 × 1024	0.5	0.5
	SH-IMP2	20/07/2009	1	30	VV	1024 × 1024		

obtained by means of visual interpretation of Landsat data in order to analyze urban expansion in China and the State of Sao Paulo in Brazil, respectively.

Additionally, these manual urban area extractions were refined, whenever available, through the inclusion of Open Street Map (OSM) data. Natural features, which are stored in the OSM database, have been clipped out of the reference data in order to be able to assess the accuracy of the results with respect to the inner urban heterogeneity. Whenever multi-temporal information was available, such as for the Chinese test sites, only the human settlement delineation in year 2010 was considered.

Finally, a representative and relatively small (ca. 85 square km) test site has been selected in Beijing and Sao Paulo to extract a more accurate ground truth by visual interpretation of VHR data optical images freely available via Google Earth^{TM}.

Thanks to all these reference data, a quantitative assessment of urban extraction results was possible at two different levels: a coarse-scale city level (spatial resolutions of 10 to 30 m), as well as a fine-scale inner-urban level (spatial resolution around 1 m). In order to avoid a biased accuracy assessment caused by predominance of built-up areas in the test scenes, the same number of random points has been generated for the urban as well as for the rural class (10000 at the city level and 2000 at the inner-urban level). Commission and omission errors, overall accuracy (OA) and K-HAT were eventually computed. In order to prove that the obtained results are statistically relevant and the difference between original and fused urban maps most likely reflects a real difference, a statistical significance test (one sample t-test [99] with 99 degrees of freedom) has been performed. The Null Hypothesis (p-value below 0.05) has been rejected since all calculated p-values report values smaller than 0.03. As a consequence it can be stated that all results presented in this work are statistically very significant.

The best results for all fusion methods are reported in table 3.6 and 3.7 for the city and inner-urban level, respectively. To understand these numbers and provide some guidelines about the usage of the multiple fusion techniques included in our framework, these numbers

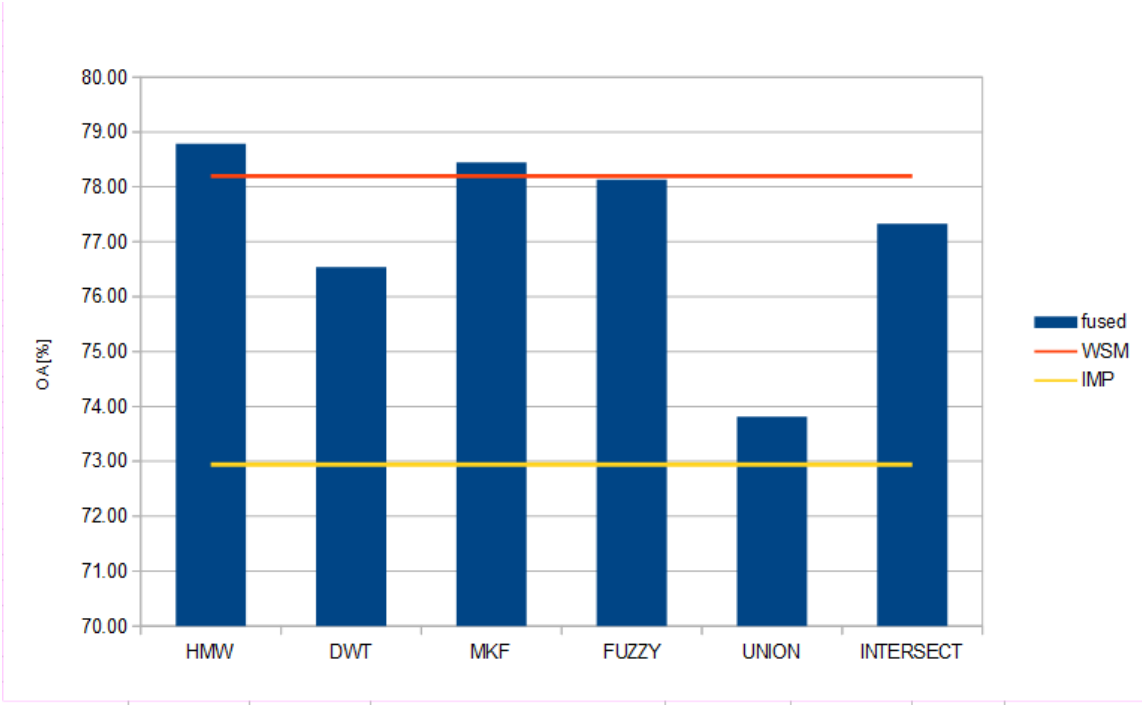

Figure 3.10: Average OA values per method at the city scale.

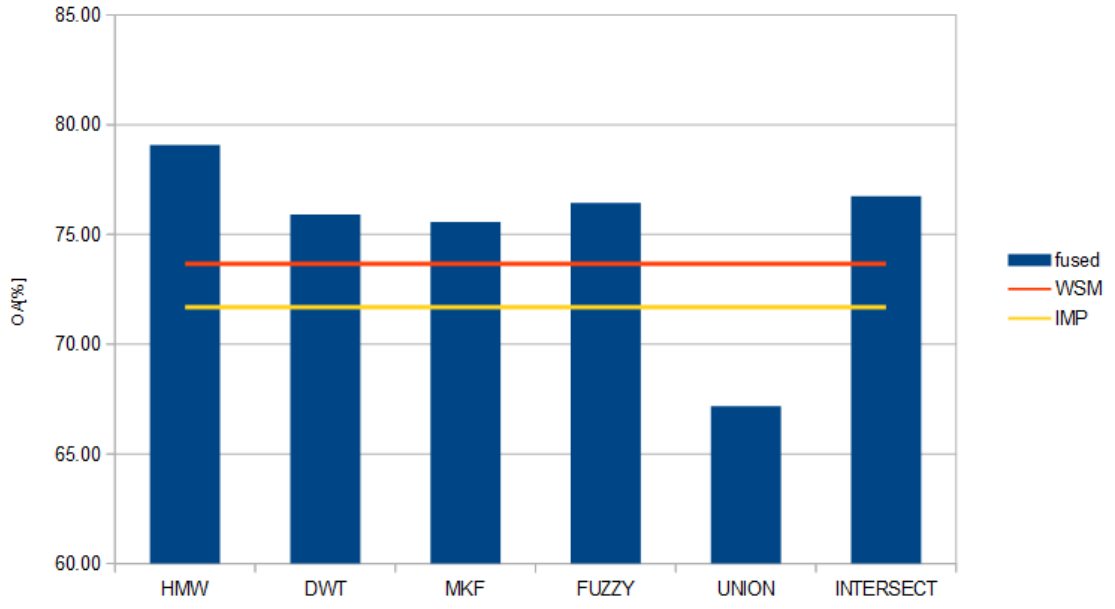

Figure 3.11: Average OA values per method at the inner-urban scale.

have been analyzed by methodology and by test area separately.

3.3.8.1 Results by fusion method

In order to get a general idea of how well the data fusion algorithms perform irrespectively of the test area, the overall accuracy values of urban maps derived from each approach have been averaged among all test sites.

Judging only from the average values, HMW delivers the best results for both the city level and the inner urban level. While the improvement for the whole city is marginal (0.6 %) the accuracy of the urban map in the small test sites within Beijing and Sao Paulo increases significantly (6 % and 4.85 %, respectively). Additionally, the results reveal that the decision level fusion approach using UNION is a very simple but inefficient way to fuse urban extractions at different scales. In the INTERSECT operator, it is not surprising that the omission error rates are very low, whereas the commission errors are very high and therefore the overall result is not acceptable. The UNION operator, on the other hand, leads to a more accurate differentiation (improvement of 3.08 % at the inner-urban level) of urban and natural areas with respect to the stand alone urban extractions based only on either WSM or IMP data.

It is also interesting to observe that, even though the MKF and the fuzzy fusion approaches belong to different levels of SAR data fusion, their overall results are very similar (78.45 % and 78.13 % at the city level, 75.56 % and 76.43 % at the fine scale). DWT results are good at the inner-urban level, but are inferior at the city level (76,54 %). Figures 3.10 and 3.11 provide a graphical comparison of the proposed SAR fusion and urban area extraction algorithms at the city- and inner-urban level, respectively. To visually help the understanding, figure 3.12 and 3.13 show the results of the proposed methods in comparison to reference data and stand alone extractions for the Beijing test site, and at the city and inner-urban level, respectively.

3.3.8.2 Results by test site

Since the experiments have been conducted in different geographic regions with varying local standard urban structures, diverse climate and vegetation zones and SAR imagery acquired in different seasons, it is impossible to make a plenary statement about which of the proposed methods is best for the purpose of built-up area extraction at the global level. Instead, it is useful to investigate the performance of the methods for each test site according to the environmental circumstances and the day of acquisition. Indeed, SAR imagery acquired in wet seasons are usually less suitable for land cover classifications due to the increased soil moisture content [100]. Hence, SAR imagery acquired in dry seasons would be the best option in order to get optimal urban mappings results. According to the Koeppen-Geiger climate classification [101] the test sites of Sao Paulo, Guangzhou and Shanghai are located in the Cfa climate zones, characterized by humid warm temperate climates without dry seasons, mild winters and hot summers. The climate in Beijing can instead be described as a humid continental climate (Dwa after Koeppen-Geiger) with dry

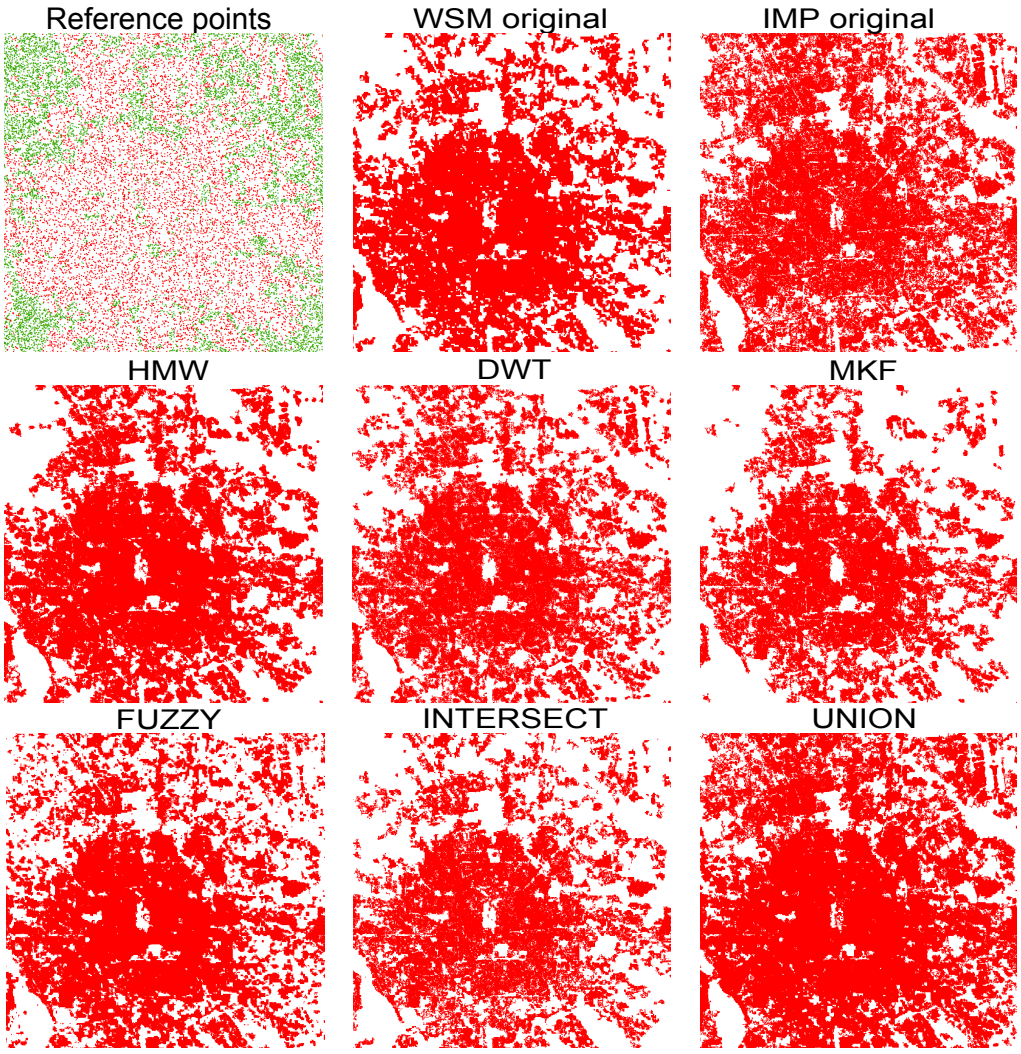

Figure 3.12: Original and fused urban area extractions for Beijing at the city level, to be visually compared with the manually extracted ground truth (top left) from [79].

winters and hot summers with almost 75 % of annual precipitation during the months of July and August. In this work, unfortunately, only the imagery covering Guangzhou was acquired in spring, a relatively dry season in this area, whereas the remaining three test sites were captured in summer. Therefore, it is not surprising that the best results for the IMP stand-alone classification have been reached in Guangzhou. For WSM based extractions using the UEXT algorithm a positive correlation between the reached accuracy and the number of acquisitions included in the multi-temporal composite can be observed at the city level. UEXT performed best in Sao Paulo reaching 84.17 % OA with 33 available images. In Beijing 17 acquisitions were used for the multi-temporal composite. leading to an OA of 79.8 % for the resulting urban map. For Shanghai and Guangzhou, 13 and 11 images were included in the WSM stack, respectively, and OA values of of 71.87 % and 76.95 % were obtained. Here, the interested reader might have noticed the exception to the before mentioned positive correlation. Indeed, the latter result can be explained by the

different nature of the Shanghai reference data, obtained by a semi-automatic extraction rather than the visual analysis of Landsat data [79].

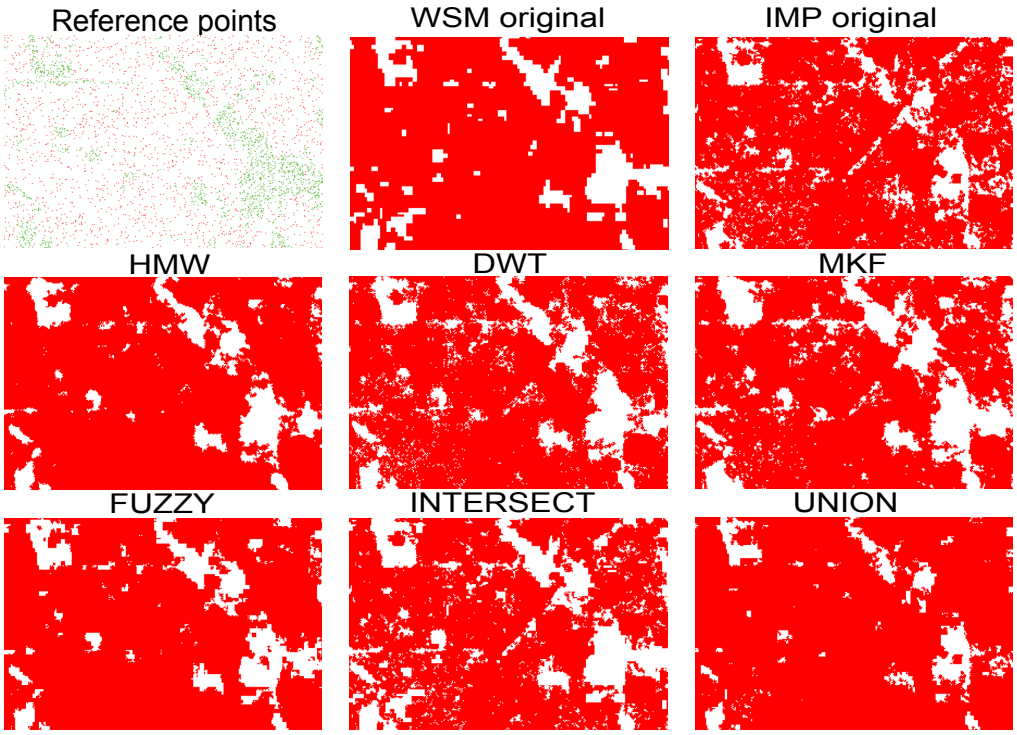

Figure 3.13: Original and fused urban area extractions for Beijing on the inner urban level. Reference points are shown for urban and rural areas as red and green points, respectively.

At the inner-urban level of validation, the number of acquisition apparently has no impact on the urban extraction result, and OA values for Beijing and Sao Paulo are similar. This could be explained by the fact that the selected small test sites are located in very stable parts of the city where generally only slight differences between annual phenological stages occur. Among the four test sites, the results in Sao Paulo results show the highest level of accuracy. While the WSM based extraction is very accurate, the performance of the LISA/GLCM method delivers poor results due to the inappropriate time of acquisition. However, the OA of the urban map derived by HMW are superior to the WSM-based extraction at the city as well as at the inner-urban level. This point proves that even low quality information can be used in order to improve urban maps when used by appropriate fusion methods.

In summary, although simple and straightforward, HMF outperforms the other methods presented in this paper due to the inclusion of carefully determined image weights into analysis. This statement is valid for all test sites even though in Beijing HMF does not seem to be the best choice at the city level, whereas at the fine level it produces the most accurate urban map. As described in section 3.3.3, the weights have been determined empirically in the course of the HMW appoach. The final weights are therefore only valid for this specific method and cannot be used for any other approach. However, another, more general method to determine image weights, based on OA of stand alone

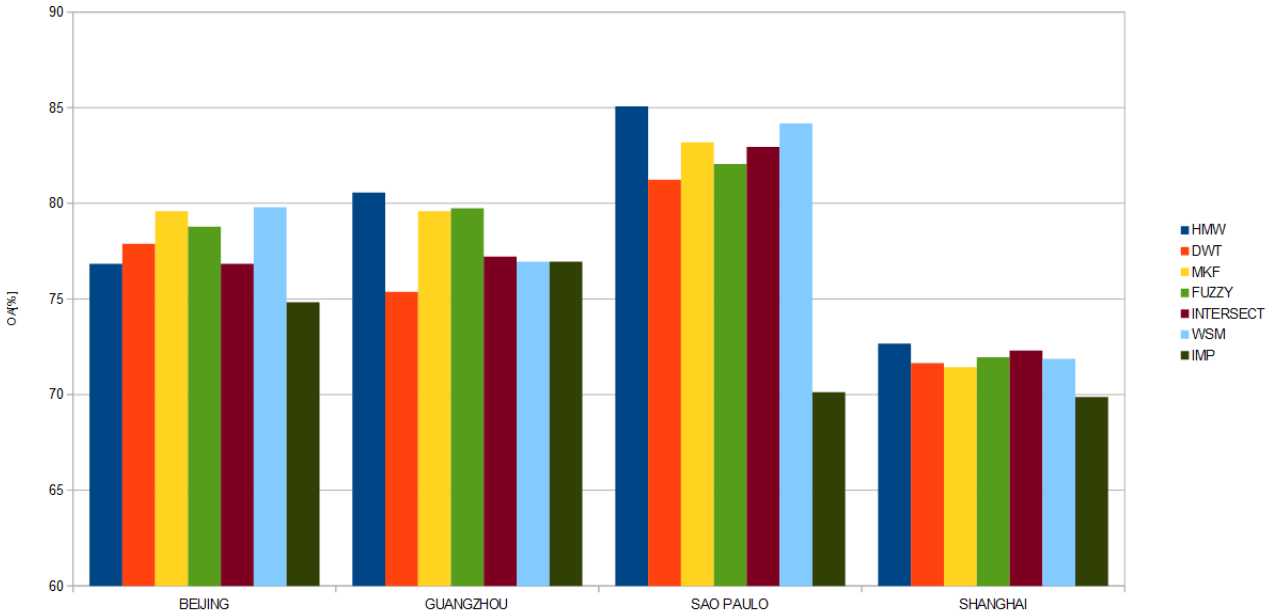

Figure 3.14: OA values per method and test site at the city level.

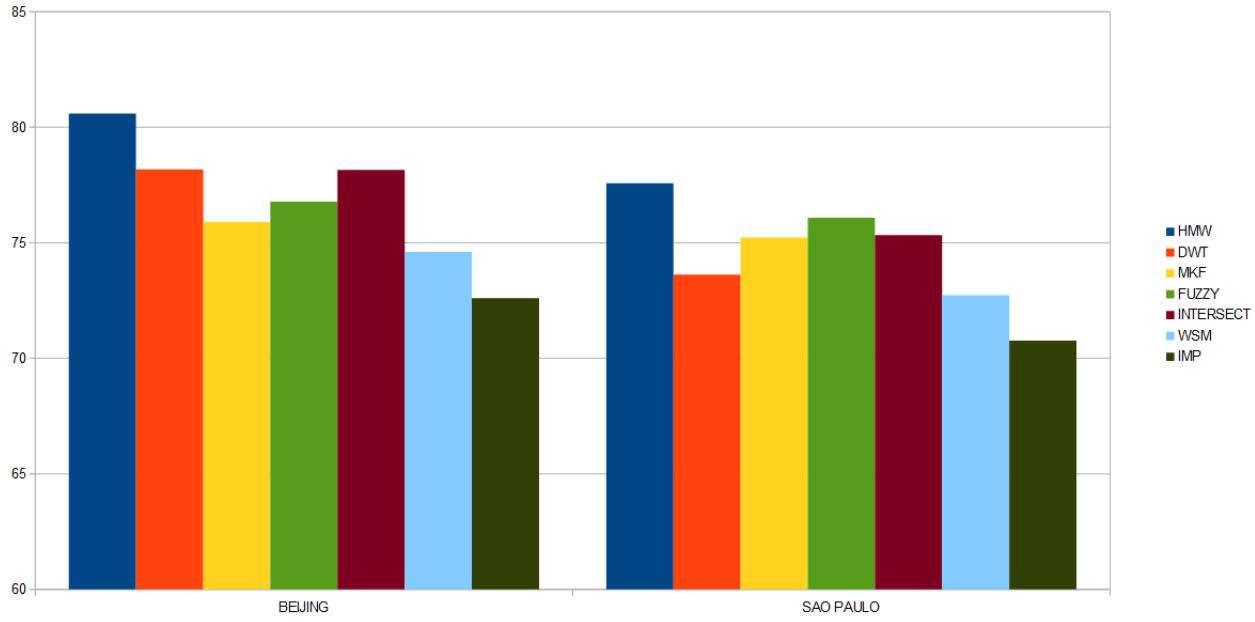

Figure 3.15: OA values per method and test site at the inner-urban level.

urban area extractions of the input data has been used in this work (see section 3.3.6 . Since their usage showed an improvement in final urban map accuracy only for the fuzzy fusion approach, they have not been used in all other methods.

It is noticeable that at the city level the decision-level fusion method using an intersection operation delivers urban maps superior to most other approaches for the Sao Paulo test site. Instead, the commission error is very high in comparison to other cities, which means that the result show a strong over-classification of urban areas. Due to the high quality of the WSM product and the applied AND operator, the strong over-classification of the IMP product basically just confirms the WSM result and therefore does not incorporate in the final fusion result. However, the over-estimation mainly takes place in the outskirts of the city and is reflected by a large amount of false urban positives, while the separation of inner urban natural and artificial surfaces is still captured well and at fine detail (especially in comparison to the WSM results at 75 m spatial resolution). As a result, the fused urban maps are in fact more accurate than the stand-alone extractions at the finer scale of analysis. Similar comments apply to the comparison on the other test sites. Figure 3.14 reports OA values for all test sites and fusion methods at the city level, and figure 3.15 at the inner-urban level only for Beijing and Sao Paulo.

3.4 Chapter conclusions

In the first part of this chapter the latest release of the global urban extent layer extracted from ASAR WSM data using the UEXT algorithm has been presented and discussed, together with a description of the improved algorithm used to obtain it. Results have been validated and compared with reference data sets at several scales, from aggregated province level for the whole P.R. China, to square km level in two of these provinces, down to 300 m resolution in the 2011 and 2015 RR test sites, and finally to the finest level of the European Urban Atlas for three selected cities in Portugal.

Both visual and quantitative assessments of the results allow to come to the following conclusions.

- The new version of the urban layer extracted from ASAR WSM dramatically improves with respect to the first version presented in [31], and achieves an average overall accuracy across both the 2011 and 2015 ESA RR test sites larger than 69 %. Considering the visual and quantitative assessment performed on the same areas plus the Chinese provinces of Jiangsu and Hubei, this performance is higher than what is achieved in the same areas by the existing ESA Globcover and CCI urban layers. Accordingly, SAR-based extraction at the original spatial resolution of 75 m, resampled at 150 m, is at least at the same level than existing multi-spectral-based extractions.

- A detailed comparison of the results from ASAR WSM and S1-A data in arid regions shows that the latter outperforms the former, by hugely decreasing the omission errors while only marginally increasing the commission error. Although a complete world coverage of S1-A data with enough images to obtain a layer similar to the one used in the test areas is not available yet, these results show a very promising performance of S1-A data. A urban layer extracted from these data in the future is

Table 3.6: Comparison of accuracy assessment results on the city level

Testsite	Data		Method	Commission	Omission	OA	$\hat{k}$
Beijing	WSM	Original	UEXT	14.03	26.31	79.8	0.6
	BJ-IMP1	Original	LISA/GLCM	25.1	23.88	75.51	0.51
	BJ-IMP2	Original	LISA/GLCM	28.62	23.12	74.14	0.48
	BJ-IMP2	Fused	HMW	9.43	36.43	77.01	0.54
	BJ-IMP1	Fused	DWT	12.75	31.11	78.03	0.56
	BJ-IMP1	Fused	MKF	15.9	24.88	79.59	0.59
	BJ-IMP1	Fused	FUZZY	15.59	26.52	78.92	0.58
	BJ-IMP1	Fused	UNION	31.28	11.59	78.61	0.57
	BJ-IMP2	Fused	INTERSECT	8.02	37.89	76.98	0.54
Guangzhou	WSM	Original	UEXT	27.35	18.73	76.95	0.54
	GZ-IMP1	Original	LISA/GLCM	24.16	23.73	76.05	0.52
	GZ-IMP2	Original	LISA/GLCM	20.82	23.5	77.84	0.56
	GZ-IMP2	Fused	HMW	19.57	18.49	80.97	0.62
	GZ-IMP2	Fused	DWT	12.4	36.37	75.63	0.51
	GZ-IMP2	Fused	MKF	20.33	19.24	80.22	0.6
	GZ-IMP2	Fused	FUZZY	23.32	16.93	79.87	0.6
	GZ-IMP2	Fused	UNION	38.04	7.14	77.39	0.55
	GZ-IMP2	Fused	INTERSECT	10.13	35.1	77.4	0.55
Sao Paulo	WSM	Original	UEXT	16.62	15.05	84.17	0.68
	SP-IMP1	Original	LISA/GLCM	50.58	13.38	68.18	0.36
	SP-IMP2	Original	LISA/GLCM	41.72	14.34	72.08	0.44
	SP-IMP2	Fused	HMW	11.21	17.28	85.73	0.71
	SP-IMP2	Fused	DWT	14.52	17.99	83.73	0.67
	SP-IMP2	Fused	MKF	11.92	18.73	84.65	0.69
	SP-IMP2	Fused	FUZZY	9.95	25.08	82.42	0.65
	SP-IMP2	Fused	UNION	48.75	5.81	72.91	0.46
	SP-IMP2	Fused	INTERSECT	9.6	23.59	83.35	0.67
Shanghai	WSM	Original	UEXT	41.68	14.67	71.87	0.44
	SH-IMP1	Original	LISA/GLCM	32.98	26.34	70.35	0.41
	SH-IMP2	Original	LISA/GLCM	36.92	24.33	69.39	0.39
	SH-IMP1	Fused	HMW	24.57	29.35	73.03	0.46
	SH-IMP1	Fused	DWT	27.33	29.21	71.73	0.43
	SH-IMP2	Fused	MKF	18.92	37.22	71.9	0.44
	SH-IMP1	Fused	FUZZY	30.39	25.64	71.99	0.44
	SH-IMP1	Fused	UNION	51.12	9.3	69.86	0.4
	SH-IMP1	Fused	INTERSECT	23.55	31.71	72.36	0.45

very likely to achieve higher accuracy thanks to the finer spatial resolution and the suitability of the UEXT algorithm to SAR data at different resolutions.

- Detailed extraction at the city level are not achievable by considering the (relatively) simple UEXT algorithm. Although very preliminary, a comparison of S1-A extraction with Urban Atlas data sets show that the level of commission/omission of S1-A

Table 3.7: Comparison of accuracy assessment results on the inner urban level

Testsite	Data		Method	Commission	Omission	OA	$\hat{k}$
Beijing	WSM	Original	UEXT	43.9	8.36	74.6	0.48
	BJ-IMP1	Original	LISA/GLCM	43.96	10.47	73.47	0.46
	BJ-IMP2	Original	LISA/GLCM	48.22	9.87	71.74	0.43
	BJ-IMP1	Fused	HMW	27.61	11.78	80.63	0.61
	BJ-IMP2	Fused	DWT	33.08	11.33	78.24	0.56
	BJ-IMP1	Fused	MKF	42.37	6.6	76.25	0.52
	BJ-IMP1	Fused	FUZZY	39.91	7.0	77.22	0.54
	BJ-IMP1	Fused	UNION	60.42	3.12	69.41	0.37
	BJ-IMP1	Fused	INTERSECT	27.45	15.71	78.66	0.57
Sao Paulo	WSM	Original	UEXT	46.81	9.44	72.73	0.44
	SP-IMP1	Original	LISA/GLCM	31.54	33.01	67.69	0.35
	SP-IMP2	Original	LISA/GLCM	37.02	16.28	73.83	0.47
	SP-IMP2	Fused	HMW	35.79	10.1	77.64	0.55
	SP-IMP2	Fused	DWT	38.65	11.33	75.64	0.51
	SP-IMP1	Fused	MKF	22.93	26.43	75.24	0.51
	SP-IMP2	Fused	FUZZY	23.71	19.90	78.28	0.56
	SP-IMP1	Fused	UNION	65.55	4.13	66.57	0.31
	SP-IMP2	Fused	INTERSECT	31.49	15.15	77.05	0.54

urban extent extraction obtained from UEXT is far from being comparable with the one obtainable from VHR optical imagery. The use of VHR SAR and more complex approaches, considering textural patterns in addition to back scattering values as in [46], [102], is definitely required to meet urban delineation needs at finer spatial resolutions.

In the second part of this chapter a general framework for multi-resolution SAR data fusion is presented. In this framework, five different multi-scale SAR fusion techniques and two urban extent extraction methods suitable for different SAR spatial resolutions have been applied to multi-resolution SAR data sets. The application of the presented data processing chains in four megacities reveals that -in the present experimental setup- the HMW method outperforms the stand-alone urban extractions and all other proposed methods at the city as well as at the inner-urban level of analysis. The main advantage of the HMW method lies in the inclusion of carefully determined, method specific weights based on quality of the input data and the environmental conditions in the study areas. Results also prove that even though a single SAR image might not be suitable for good quality urban area extraction by itself, it can still contain important information to contribute to the fused product and as a consequence improve the accuracy of fused urban maps. The best result obtained by HMW does not prevent the other tested fusion methods to deliver improved urban area extractions as well, especially at the inner-urban level of analysis. Accuracy assessment results reveal that the final quality of urban maps shows only moderate differences among most of the approaches presented in this paper.

Only the decision level fusion using 'OR' as logical combination failed in improving the urban mapping accuracy. However, with respect to a good compromise between urban mapping accuracy, complexity and implementability of the approach, computational cost and time expenditure, HMW is the best choice for the given data sets and study areas. In conclusion it must be highlighted again that the final selection of the most suitable SAR fusion method has to be done for each study individually. One important factor for the selection of an appropriate method are environmental characteristics of the study site. As already discussed in section 3.3.8.2 the study areas of this work are located in temperate mid-latitude climate regions. However, if the area under investigation lies within arid climates, the backscatter characteristics of SAR signals significantly change, typically resulting in an increased confusion between urban structures and bare soil or rock formations. For tropical regions a similar effect can usually be observed: the increased soil moisture content is decreasing the usefulness of SAR data sets for urban area extractions. The aforementioned effects normally lead to an overestimation of urban areas when SAR based built-up areas extraction algorithms are applied. As a consequence, pixel-level SAR fusion approaches might not be the best choice for urban mapping in those regions. In order to counteract overestimation, instead, decision level data fusion using the logical "AND" operator will most likely deliver the best urban map in this case. In addition to geographic location and climate region, time also plays an important role in the selection of an appropriate SAR fusion method. First, data availability throughout the annual phenological cycle strongly influences the initial situation of analysis and therefore also the selection of a fusion method. In case data acquisitions from different phenological seasons are available, pixel-level fusion methods are a meaningful way to combine SAR data. Similar to the generation of a multi-temporal ASAR WSM layer stack [31], HMW, DWT or MKF are capable of reducing significantly the disturbing effects of phenology and therefore would most likely deliver superior results in comparison to feature- and decision-level approaches. However, in case data availability is restricted with respect to the annual cycle of vegetation, decision level fusion approaches would probably be the best choice in order not to risk strong misclassification of human settlements. The second aspect of time which drives the selection of reasonable SAR data fusion algorithms is the date of data acquisition. If one is interested in the extent of human settlements in a specific point of time it is necessary to use EO data with acquisition dates as close as possible to the intended one. If the time span between the most current and the oldest available data sets exceeds a couple of years and one is interested in the urban area extent at a specific time in between, decision level fusion approaches are the best choice.

With respect to data availability, not only the acquisition date, but also the spatial resolution of each available SAR image has to be considered . In case the available multi-scale SAR data set does not match a pyramidal structure, DWT and MKF cannot be applied for data fusion purposes. Also, in case the multi-scale data set consists of SAR imagery with very different spatial resolutions (e.g. abruptly decreasing from very high to very coarse resolution without intermediate scales) pixel- and feature-level fusion methods

might fail. Thus, decision level data fusion is most likely the best choice if only a very scale-heterogeneous data set is available.

Finally, another aspect that has to be considered to select the most suitable SAR data fusion method is the availability of computing resources, especially if the analysis has to be conducted on the regional, continental or even global scale. Among the presented methods, MKF is the most challenging one with respect to complexity and computational costs. However, in case numerous SAR data sets fitting a pyramidal structure and sufficient computing power are available, MKF might be the best choice for SAR data fusion because the Kalman Filter is known to be one of the best estimators of stochastic processes when measurements about the process evolution are available. With sufficient data availability and the possibility to run an arbitrary number of iterations, MKF eventually would lead to a fused image that incorporates all the available information from the multi-scale data set and therefore would be most suitable for SAR based urban extent extraction.

Chapter 4

Multi-source decision level data fusion for urban mapping

4.1 Introduction*

This chapter aims at designing and validating techniques for the combination of SAR and multi-spectral data sets, with focus on urban area extent extraction. Due to their complimentary nature, the joint use of both data types in a uniform processing chain bears huge potentials for an improvement of thematic details and spatial as well as temporal resolutions of the urban mapping products available today. Methods exploiting optical data might lack in consistency and reliability in areas where cloud cover is a common phenomenon due to climatic conditions. SAR data on the other hand is limited with respect to the detailed extraction of natural, vegetated areas. Only the combination of both data sets enables the possibility to make up for their particular drawbacks.

Joint SAR and multi-spectral data analysis carries a big potential for a better discrimination of urban areas against natural elements with similar spectral and different spatial properties, such as bare soil areas, or inundated fields, that may be sometimes confused with human settlements. This may be easily exemplified by the example in figure 4.1 for the area of Ribeirao Preto (Sao Paulo state, Brazil). The fused result has been generated using a map extracted from Landsat data (30 m spatial resolution) and a map from ASAR Fine mode SAR data (75 m spatial resolution) applying a logical "AND" operator for large settlement, and an "OR" for small ones (e.g., less than 200 pixels at the WSM resolution). The better quality of the fused product can visually be assessed looking, for instance, to the red rectangle. The major towns are better characterized thanks to the SAR data set, while the smaller human settlement extents come mainly from the optical finer resolution extraction.

In addition to the above mentioned, and somehow trivial, fusion strategy based on logical operations of individually produced urban products, fusion approaches on the decision

*A large part of this introductory section comes from: A. Salentinig and P. Gamba, "Combining SAR-Based and Multispectral-Based Extractions to Map Urban Areas at Multiple Spatial Resolutions," in *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 100-112, 2015.

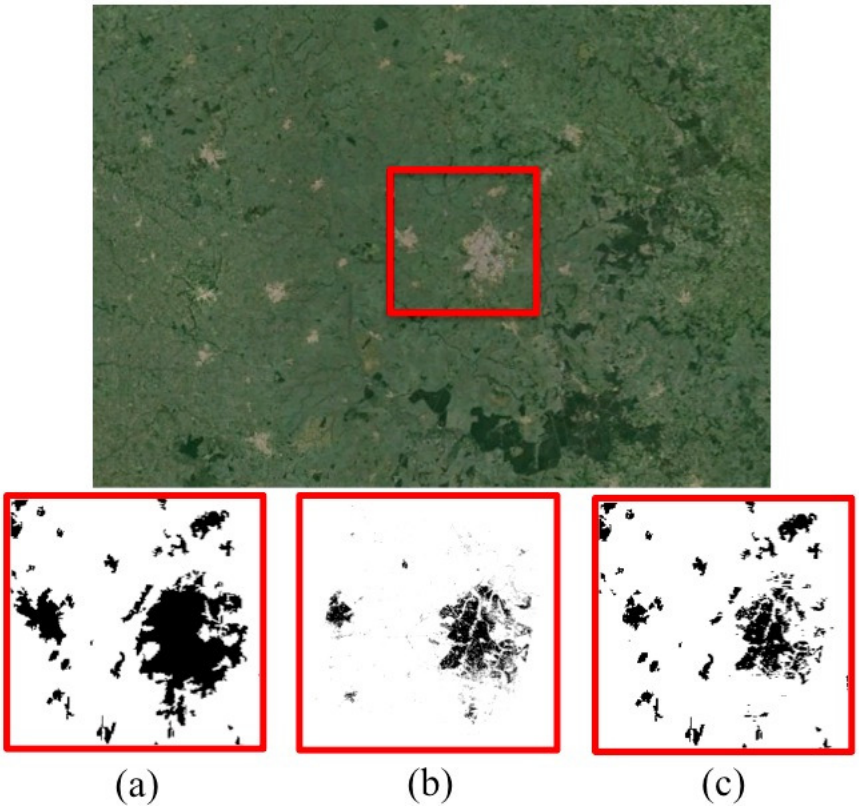

Figure 4.1: A visual comparison of the urban area extent extraction in a portion of Ribeirao Preto, Sao Paulo state, Brazil, starting from (a) WSM data or (b) Landsat data, to be compared with the result of a very simple decision-level data fusion approach using logical operators. The image on the top is a Landsat scene of the area.

level as well as on the raw-level need to be further investigated. Indeed, promising methods for the combination of microwave and multi-spectral data already exist. Specifically, two main research trends in the field of multi-source data fusion can be observed. The first one belongs to the family of extended Maximum likelihood classifier based approaches applied on single image vectors which are concatenations of signals from different sensors (e.g. [103]). The main challenge with these methods is the establishment of a good model for the multi-source data distribution. Evidential reasoning approaches are applied to classify data from optical and SAR data sets by taking into account the uncertainties caused by the varying data sources (e.g. [104], [105]). The second trend is towards more sophisticated Neural Network classifier based fusion techniques (e.g [106], [107]), which offer a powerful tool to effectively integrate different data types through statistical learning.
Other approaches at the decision level fusion are based upon the idea to use an ensemble of classifiers and subsequently merge them to improve classification performance (e.g. [108]–[111]). However, all these approaches have been tested only on small geographic areas, with very specific data setups, ahence their applicability on wide geographical areas is still lacking. From a more general perspective a number of important aspects and parameters that influence the fusion result as well as a consecutive classification need to be considered.

Not only the the selection and application of the actual fusion and classification method, but also factors including appropriate data acquisition/selection, precise georeferencing and co-registration, feature extraction, and quality assessment of the results greatly impact the final results [112]. To try and match all these criteria is a very challenging task, especially when aiming at accurate urban extent extractions for the whole Earth surface.

In this chapter two novel decision level data fusion methods for the purpose of urban area detection and delineation are presented. The first one (section 4.2) exploits two independent SAR based urban area extraction algorithms and uses their outputs in an combined pixel- and object based approach to update an existing optical EO data based urban map. The second method, discussed in section 4.3, spatially regularizes multi-source urban maps from different spatial resolutions and combines the results in a weighted fusion approach.

4.2 Pixel- and object-based fusion of urban maps

In the framework of the Land Cover CCI project mentioned in section 2.2 an urban RR (see also section 3.2.2.1) was initiated focusing on the development of a processing chain to update / improve the existing urban class (based on optical MERIS data at a spatial resolution of 300 m) of the CCI-LC global land cover products and get prepared for the next generation of LC maps using Sentinel 1 SAR imagery.

This section contains a summary of the collaborative contribution by the University of Pavia (UPV) and the Friedrich-Schiller-University of Jena (FSU), Germany, to the urban RR. First, the used data and the test sites are presented, followed by a detailed description of the processing chain for updating the current version of the urban CCI-LC class using two independent SAR data based urban area extraction algorithms. The first phase of the proposed map updating procedure is devoted to the pixelwise generation of two separate binary urban maps from Sentinel 1 data. Those maps, together with the urban CCI-LC class form the input data set for a subsequent decision level fusion strategy, which exploits object-based image processing concepts to generate a fused urban product. The section concludes with a discussion of the obtained results.

4.2.1 Data and test sites

The urban RR sites are located in the Mediterranean and in North Africa, namely in Portugal, Turkey, Israel, Egypt and Tunisia, encompassing an area between 60.000 and 122.000 km^2 (see figure 4.2). The test sites were selected as optical and SAR data face major issues for urban area mapping in semi-arid and arid regions.

The urban RR data package consists of multi-temporal Sentinel-1 C-Band SAR data acquired between October 1 and December 31, 2015 in the Interferometric Wide Swath mode (GRDH product, 20 m spatial resolution). For each test site a time series of 14 to 16 VV- and VH-polarized Sentinel-1 images including both, ascending and descending orbit, were available. The conducted data processing steps included image calibration,

Figure 4.2: Location of the ESA urban RR 2015 test sites.

terrain geocoding, topographic normalization, multi-temporal filtering and the calculation of multi-temporal mean values (MMean, see pixel level image fusion method described in chapter 3.2). For the validation of the results the reference data sets mentioned in section 3.2.2.1 which have been generated following the procedure described in section 3.2.2 were used.

4.2.2 Processing chain

Starting from the above mentioned multi-temporal SAR data sets, the proposed urban extent extraction procedure first follows two separate approaches. Both groups, from UPV and FSU have successfully been working on SAR data based urban area mapping in the past and developed different SAR data urban mapping algorithms. However, both methods exhibit advantages and drawbacks leading to different, but complementary results with varying performances based on the environmental characteristics in the area under investigation and the size, composition and structure of present human settlements. In order to make use of the the advantages of both approaches the resulting urban maps are combined in a second step applying an object based decision level fusion approach.

4.2.2.1 Used urban mapping methods

The urban area mapping approach designed at FSU is based on the assumption that due to the interplay of artificial and natural structures, urban areas exhibit a very specific texture in SAR imagery. To exploit this fact a texture measure called UADP (Urban Area Detection Parameter) has been developed [113]. The UADP measure calculates for each image pixel the mean radar backscattering difference to its surrounding pixels with a distance of 2 pixels and aims at highlighting built-up areas. After the computation of UADP the FSU algorithm follows a combined pixel- and object-based approach. The im-

age objects are extracted using the eCognition software. The eCognition Server software provides a processing environment for batch execution of image analysis jobs. This software is designed specifically for the analysis of large scenes as well as large quantities of data and can automatically process thousands of images in a single, fully automated run. In order to improve the segmentation results and to speed up the segmentation process significantly, the S1-A MMean values are transformed to "urban probabilities", stretched to 8 bit and then used as input for the segmentation process. The urban probability maps used for image segmentation are processed as follows. The S1-A MMean values covered by the urban class of the CCI-LC map are sorted in ascending order (lowest 20 % were excluded from further analysis) and divided into 100 equal intervals. For each interval step a corresponding threshold value is specified. These thresholds are applied to the S1-A MMean values and the ratio of the S1-A MMean values covered by the urban class of the CCI-LC map and the total pixel number exceeding the threshold is computed. Regions with a slope over 5 % are not considered in this analysis. Generally, a high radar backscatter value corresponds to a high urban probability. Areas with a mix-up of the urban class with other land use / land cover classes are characterized by a low overall urban probability. For example, in Egypt many agricultural used areas have a radar backscatter similar to urban areas. The S1-A data of Israel show significant striping effects in certain parts of the scenes. For both test sites, the calculated maximum urban probabilities are low. Thus, the urban probability statistics could be valuable information to identify image scenes where most probably problems exist and the classification accuracy might be low. The update of the CCI-LC map is performed by simple thresholding. The thresholds used are not fixed and are assessed for each scene/region separately in an automatic manner. The update process itself is subdivided in two main processing steps. The first step deals with the overestimation of the urban class in the CCI-LC product. All objects covered by the urban class of the CCI-LC product and having a mean texture at VV or VH below 0.4 times the maximum urban probability are assigned to the rural class. In the second main processing step urban areas missing in the CCI-LC map are added. The estimation of meaningful threshold is based on statistics derived from all urban objects identified in the first processing step. In principle, the processing sequence used for the calculation of the urban probability maps described above is applied. The process stops, once the number of pixels outside the urban areas exceeds those of urban areas (with a maximum threshold corresponding to the 0.8 interval). Next the thresholds are applied to the MMean values and the texture measures. Each image segment with a proportion of potential urban pixels exceeding 10 % are assigned as new urban region in the final update map. For regions with a slope of less than 2.5 degrees the MMean values and the UADP texture measure are considered. For slope values ranging from 2.5 to 10 degrees the UADP texture measure only is used, as this parameter is less affected by topography. In a final step all regions having a slope above 10 degrees are masked.

In addition to the urban map delivered by FSU using UADP, the Sentinel 1-optimized version of the UEXT algorithm, which has already been described in detail in section 2.3.2

has been applied on the input data and the obtained results have been used as input to the combined map updating approach.

4.2.2.2 Combined approach

In order to update the CCI urban map and deliver improved information about human settlement extents the results from the FSU as well as the UPV processing chains - which are complementary to some extent - have been used in an integrative manner. The UEXT results for VV and VH data as well as the FSU result and the CCI 2010 map are used as input in an object-based decision level data fusion strategy. The aim of this fusion strategy is to make use of the advantages of each input data while trying to minimize the effects of extraction specific drawbacks. The CCI 2010 urban map has been generated through exploitation of coarse scale optical satellite imagery. While big urban patches are extracted reliably (at the coarse scale), small urban settlements are not detected due to the limited spatial resolution. Generally, the processing chains proposed by FSU and UPV are both suited for reliable urban map updating, but still have some limitations. The FSU approach sometimes lacks in the detection of very small settlements. Misclassifications between urban areas and regions of strong topography are quite low. UEXT on the other hand is capable of capturing even very small urban structures but shows weaknesses in areas with advanced landform configurations, which often causes ambiguities between urban areas and bare soil/rock formations in SAR data. Taking those considerations into account the following fusion strategy has been developed:

- urban objects extracted from the UPV or FSU processing chain are considered urban in the final fused urban map only if at least 20 % of its area is overlapped by the CCI 2010 urban map;

- urban objects extracted from the UPV or FSU processing chain are considered urban in the final fused urban map only if at least 20 % of its area is overlapped by one of the other inputs of the fusion approach;

- objects with average slope values above 2.5 degrees have to be urban in UEXT VH or UEXT VV as well as in the FSU result in order to be considered urban in the final map.

The final urban map includes a mixture of FSU and UPV extraction based elements which depicts not only a refinement of the delineation of mid-size and big cities, but also represents a significantly improved detection of small human settlements. Since the intended output resolution of the CCI-LC map is 300 m the map updating results (20 m spatial resolution) generated by the combined UPV/FSU approach have been resampled using the nearest neighbor technique. A graphical representation of workflow of the map updating approach is shown in figure 4.3.

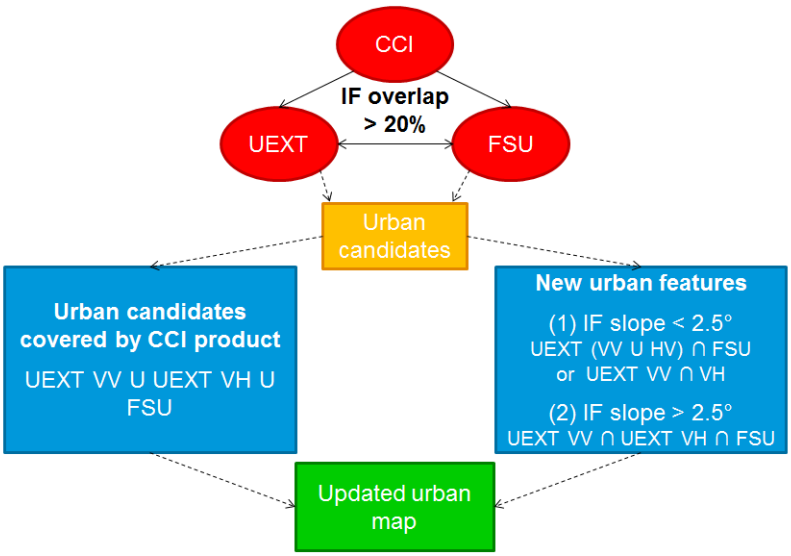

Figure 4.3: Graphical representation of the presented map updating strategy.

4.2.3 Results and discussion

The results achieved by applying this combined processing chain demonstrate that the classification accuracy of the current version of the CCI-LC map could be improved significantly by the proposed update methodology. The accuracy assessment results are reported in table 4.1. However, in few regions some problems could be identified. In the course of the update process, some agricultural used areas are assigned to the urban class. This problem is evident especially in the Egypt test area, where the Nile delta is composed by intensively used agricultural areas and a intertwined net of small to medium human settlements. Agricultural fields characterized by a very high radar backscatter are usually covered by dense and/or mature vegetation, whereas the radar backscatter from bare fields and fields in a very early growing stage is quite low. To reduce misclassifications between urban areas and agricultural fields the selection of meaningful acquisition dates is quite important. In regions with distinct seasonality, the time series should include SAR data acquired at the early beginning of the main growing season and after the main harvest time. Acquisitions from the winter months should be avoided. In all the other test sites, where agricultural landuse in the vicinity or urban patches is less apparent and the unfavorable acquisition dates of the SAR data are less influential the urban area extractions show tremendously higher accuracy values (ranging between 5.23 and 20.07 % OA improvement with respect to the 2010 CCI-LC map). Figure 4.4 shows a comparison of the CCI-LC 2010 urban maps and the updated result using the UPV/FSU approach for the small towns of Sousse and Monastir in Tunisia.

Table 4.1: Accuracy values of the results from the combined UPV/FSU approach, generated in the framework of the ESA urban RR 2015

	CCI-LC 2010				Updated map using S-1 data			
	OA(%)	$\hat{k}$	**Comm.(%)**	**Omis.(%)**	**OA(%)**	$\hat{k}$	**Comm.(%)**	**Omis.(%)**
Portugal	83.89	0.68	31.38	0.84	89.12	0.78	12.55	9.21
Tunisia	77.66	0.55	31.06	13.62	91.28	0.83	5.96	11.49
Egypt	91.92	0.83	14.55	2.87	80.84	0.62	9.43	28.89
Israel	68.98	0.38	42.34	19.71	89.05	0.78	16.42	5.47
Turkey	74.84	0.50	47.06	3.27	88.24	0.76	14.38	9.15

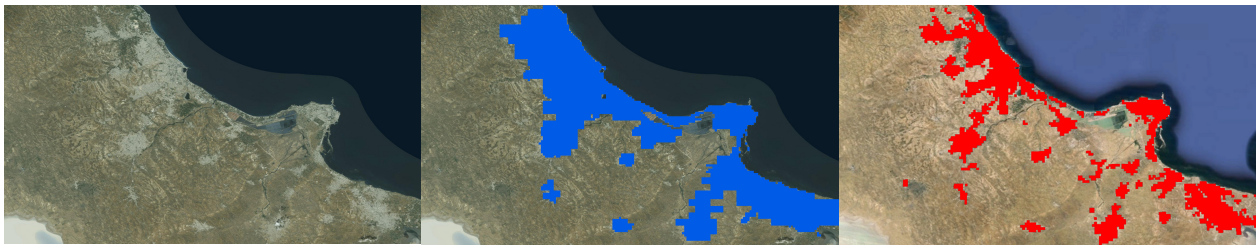

Figure 4.4: Urban maps for Sousse and Monastir, Tunisia; left:Google Earth^{TM} , middle: CCI-LC 2010 urban map, right: result from the proposed map updating procedure.

4.3 Decision fusion by spatial regularization and image weighting*

A different solution to the problem of combining multiple urban extent maps at different spatial resolutions is proposed in this section. It is assumed that multiple urban area extraction maps, either binary or fuzzy (i.e., featuring a urban probability value) are considered and combined into a final product. Starting from the assumptions that these urban maps might lose valuable information when they are resampled to another resolution, the method tries to improve the upsampling and downsampling procedures. Subsequently, the rescaled products are fused using image weights, computed according to considerations about scale-dependent advantages and drawbacks, as well as the environmental characteristics at the time of image acquisition in the area under investigation.

4.3.1 Data and study area

The satellite images used for the experiments in this chapter are very heterogeneous with respect to sensor type and spatial resolution, as well as data acquisition time.

*This chapter is to a large extent composed of two conference contributions: A., Salentinig, and P. Gamba, "Towards global human settlement characterization exploiting multi-scale optical and SAR data", in *Proc. of* Dragon 3 Final Results Symposium, 2016, in press and A., Salentinig, and P. Gamba, "Multi-scale decision level data fusion by means of spatial regularization and image weighting", submitted as an invited paper for the Joint Urban Remote Sensing Event 2017.

In the SAR domain the finest resolution that has been used corresponds to Alos PALSAR data (12 m). Even though the imagery dates back to 2007 it is still a useful source for urban mapping due to its fine level of detail. Moreover, since urban areas usually do not diminish over time, especially in highly urbanized environments, archived data still has a right to exist in data fusion based approaches for urban area mapping. On a slightly coarser scale (20 m) Sentinel 1, ESA newest generation SAR data, has been exploited in this work, too. Used data correspond to a multi-temporal composite of images acquired between February and November 2015. Moving to coarser resolution data, the ASAR sensor, mounted on the ENVISAT satellite, provided SAR data sets in a variety of resolutions and modes between 2002 and 2012. In this work ASAR Image Mode Precision (IMP) imagery at a spatial resolution of 30 m and Wide Swath Mode data (75m) were used. While the IMP data sets have been acquired between May and August 2010 the WSM data correspond to a multi-temporal composite containing imagery from 2008 – 2010. The same pixel level fusion approach as for the generation of the Sentinel 1 composite has been applied, whose description can be again found in section 3.2.

In the multi-spectral domain a Landsat 8 data multi-temporal composite from 2015 has been exploited in the proposed fusion method. Based on the NDVI values, only the "greenest" pixels in the time series have been used in order to generate a product with as little cloud cover as possible and to guarantee the best distinguishability between natural and artificial surfaces. The method used to produce the multi-temporal composite is described in detail in [114].

In order to include information about the state and condition of the vegetation cover in the study areas at the time of image acquisition maximum annual Landsat 7/8 NDVI layers have been used. A summary of data characteristics of the used EO imagery is given in table 4.2.

Table 4.2: Characteristics of the data used in this chapter

Sensor	**Type**	**Resolution [m]**	**Date**
Alos PALSAR	SAR	12	01/2007 - 07/2007
Sentinel 1	SAR	20	02/2015 - 11/2015
Landsat 8	optical	30	2015
Landsat 7/8	optical - NDVI	30	2007, 2008, 2010, 2015
Envisat ASAR IMP	SAR	30	05/2010 - 08/2010
Envisat ASAR WSM	SAR	75	2008 - 2010

Two Chinese Megacities, with a population exceeding 10 million inhabitants have been selected as test sites: Beijing and Guangzhou. While both locations share an exceptionally high urbanization rate, they differ with respect to climate regions and vegetation typologies. The climate in Beijing can be described as a humid continental (Dwa after Koeppen-Geiger) with dry winters and hot summers, and almost 75 % of annual precipitation during the months of July and August. Guangzhou, on the other hand, is located in the Cfa climate zone, characterized by a humid warm temperate climate without dry

seasons, with mild winters and hot summers. As a consequence the flora within the two test sites, including the manifestation of distinct phenological seasons, is very different.

4.3.2 Methodology

Let's start from the general framework for multi-scale SAR data fusion for the purpose of urban area extraction described in section 3.3. Within this formalized framework a number of different SAR data fusion techniques have been developed and compared against each other with respect to accuracy and location-specific performance differences. One of the major conclusions of chapter 3.3 is that a reasonable and elaborate selection of image weights can significantly improve the performance of fusion algorithms and that environmental characteristics at the time of image acquisition need to be taken into account within the fusion procedure because they strongly influence usefulness of EO data for urban area extraction. To that aim a new fusion strategy has been developed and is described below.

4.3.2.1 Urban area extraction

The idea proposed in this chapter is to build upon the already extracted urban extent maps by the single data sets (or composite) that are considered. Those maps have been generated through application of already existing, sophisticated and intensively validated urban area extraction algorithms developed at the University of Pavia. More specifically, the LISA/GLCM algorithm (see chapter 2.3.1) has been applied in the Alos PALSAR and the Envisat ASAR IMP data sets. The UEXT method (see chapter 2.3.2) has been used to extract urban maps from the Sentinel 1 and the Envisat ASAR WSM data. Finally, the NDSV based urban mapping procedure discussed in chapter 2.4.1 has been used for the Landsat data based urban products.

4.3.2.2 Spatial regularization

In order to combine information from multi-scale urban maps it is necessary to resample all input data sets to a common grid and spatial resolution. In this chapter two different intended output resolutions (20 and 75 m) are considered. In [115] it has been shown that valuable information might be lost due to rescaling processes and that - depending also on the method used to spatially homogenize the data on the same grid - different results will be produced. In this context "Pixel aggregation" is considered as an interesting alternative to classic rescaling/resampling methods (e.g. nearest neighbor, cubic convolution). "Pixel aggregation" means that the average of all values inside a pixel of the final grid at the intended resolution is computed, considering a weight related to the spatial contribution (area) by each pixel to the final aggregated pixel. The rescaled pixel values are therefore computed as

$$\mathbf{F}(n)_{i,j} = \frac{\sum_i^p w_p * \mathbf{U}(m)_{i',j'}}{p},$$

where $\mathbf{F}(n)_{\mathrm{i,j}}$ is the pixel value of the rescaled image at the generic n-th spatial resolution, $\mathbf{U}(m)_{\mathrm{i',j'}}$ corresponds to the pixel values of the image at another generic spatial resolution (m) to be rescaled to the intended resolution n, and w_{p} is the weight for each pixel of the image to be resampled and corresponds to the fraction of the pixel that intersects the extent of the overlapping pixel from the intended output raster. In case of binary urban maps, pixel aggregation basically computes the urban fraction for each new pixel based on the finer resolution information. Thus, it somehow gives information on the urban probability, which is even better characterized when considering already urban probability maps as input to the process. The procedure described above has been applied to each data set with finer spatial resolution than the intended one.

In the upscaling scenario, i.e. when the starting resolution is coarser than the intended one, a pixel aggregation procedure is not possible without an intermediate step. Specifically, the coarser resolution image is upscaled first to an even finer resolution (the smallest common factor of the input and intended resolutions). This has to be done in order not to lose any information in the upscaling. Subsequently the upscaled image can be resampled to the intended resolution via pixel aggregation. In doing this, the urban fraction within each pixel of the new raster image is computed on the basis on the originally coarser resolution image, eventually leading to an "urban probability map" (like the one for in the previous – downscaling – case).

4.3.2.3 Image weighting

As highlighted in chapter 3.3 the introduction of reasonable image weights can significantly improve the performance of data fusion methods. In this approach image weights are computed under consideration of spatial resolution and environmental characteristics at the time of image acquisition. The final image weight $\mathbf{w}(n)$ for each input map is calculated according to

$$\mathbf{w}(n) = \kappa(n) * \lfloor 1 - \mathbf{NDVI}(n) \rfloor .$$

The factor $\kappa(n)$ accounts for the influence of rescaling of input maps and corresponds to a preliminary evaluation of the goodness-for-fit of the urban map at the n-th spatial resolution. This preliminary evaluation is obtained by computing the Kappa index, commonly used to describe the agreement of remote sensing products with a set of reference points, on a set of validation points randomly selected as 50 % of the training set (for details see table 4.3). To compute these values, binary urban maps at each scale are automatically generated by means of the image preserving thresholding technique introduced in [116]. The aim of this preliminary evaluation is clearly to understand about how well a specific input map is suited for urban mapping at a peculiar scale. As mentioned above, this information is directly used as a part of the final image weight.

The second factor of the image weight is a function of the normalized difference vegetation index $\mathbf{NDVI}(n)$ at the spatial scale n. Its introduction in the formula above aims at taking into consideration the environmental characteristics of the test area at the time of image acquisition. As a matter of fact, NDVI has been used intensively by the remote

sensing data analysis community to retrieve information about the status of vegetation. In this approach maximum annual NDVI is used as an ancillary information source about the state of the vegetation in the year of each input data. Since the relationship between the existence of urban fabric and the NDVI is clearly a negative correlation, the complements to 1 of the NDVI values are used.

In an independent experiment the SAR-based urban area extraction algorithm presented in [44] has been applied to a Sentinel 1 data time series in order to investigate the correlation between the spatial average (in the test sites) of the maximum monthly NDVI values in a 3 year timeframe and the chronologically fitting urban area extraction results. Results revealed that, especially in the Beijing area, where distinct phenological seasons are apparent, the best extractions are achieved when the NDVI values are low and vice versa. Figure 4.5 shows the course of urban area extraction accuracy and mean NDVI values throughout a year in the test sites of Beijing and Guangzhou, respectively.

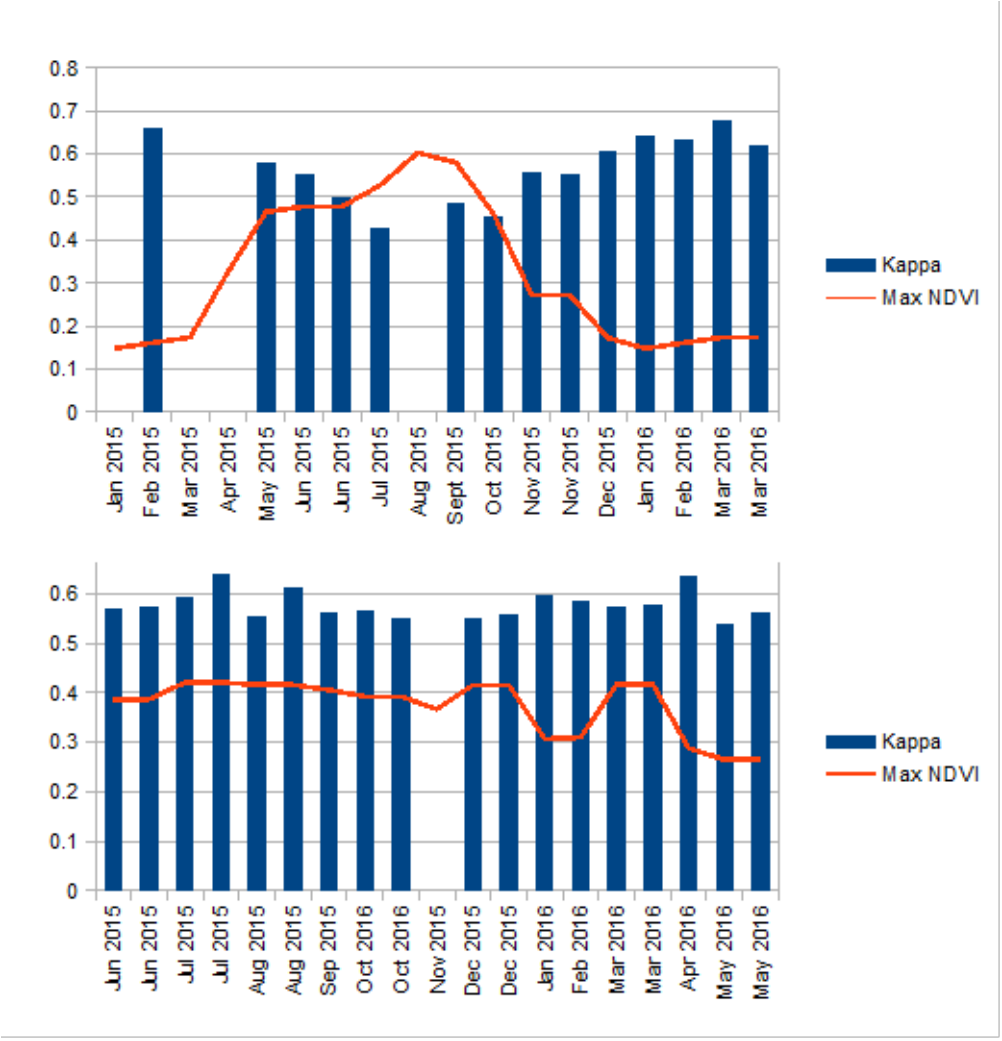

Figure 4.5: Urban area extraction accuracy versus average maximum NDVI: Beijing area (top), Guangzhou area (bottom).

In this approach the used data sets were acquired in the course of a couple of years

instead of months. As a consequence, the conclusions from the experiments mentioned above are transferred to a coarser temporal scale, using maximum annual NDVI values as a part of the final image weights.

4.3.2.4 Information fusion

Eventually the input data sets are integrated into a single image fusion result $\mathbf{Y}(n)$ via a simple summation of the product of the urban fraction maps $\mathbf{F}(n)$ and their corresponding image weight $\mathbf{w}(n)$:

$$\mathbf{Y}(n) = \sum \mathbf{w}(n) * \mathbf{F}(n).$$

In the final step the fused urban fraction maps are transformed into a binary result $\mathbf{M}(n)$ using the same moment preserving thresholding technique mentioned in section 4.3.2.3. Figure 4.6 shows the fused urban fraction map and the final result at a resolution of 20 m for the Beijing test site. A graphical representation of the complete workflow

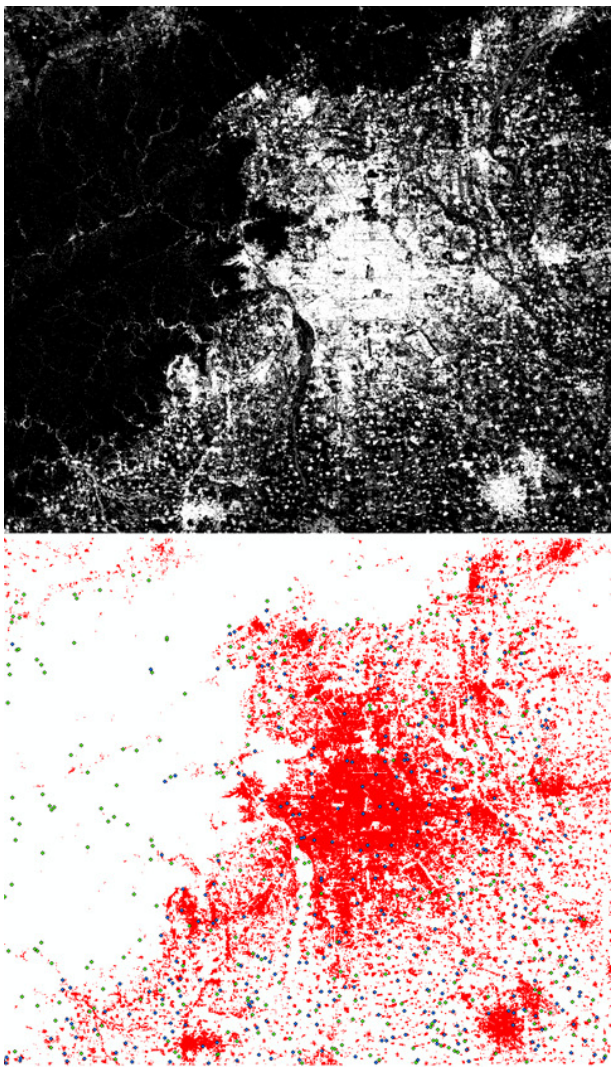

Figure 4.6: Top: Fused urban fraction map Y(n); bottom: final binary map M(n) at 20 m spatial resolution superimposed by urban (blue) and rual (green) test points.

presented in this chapter is shown in figure 4.7.

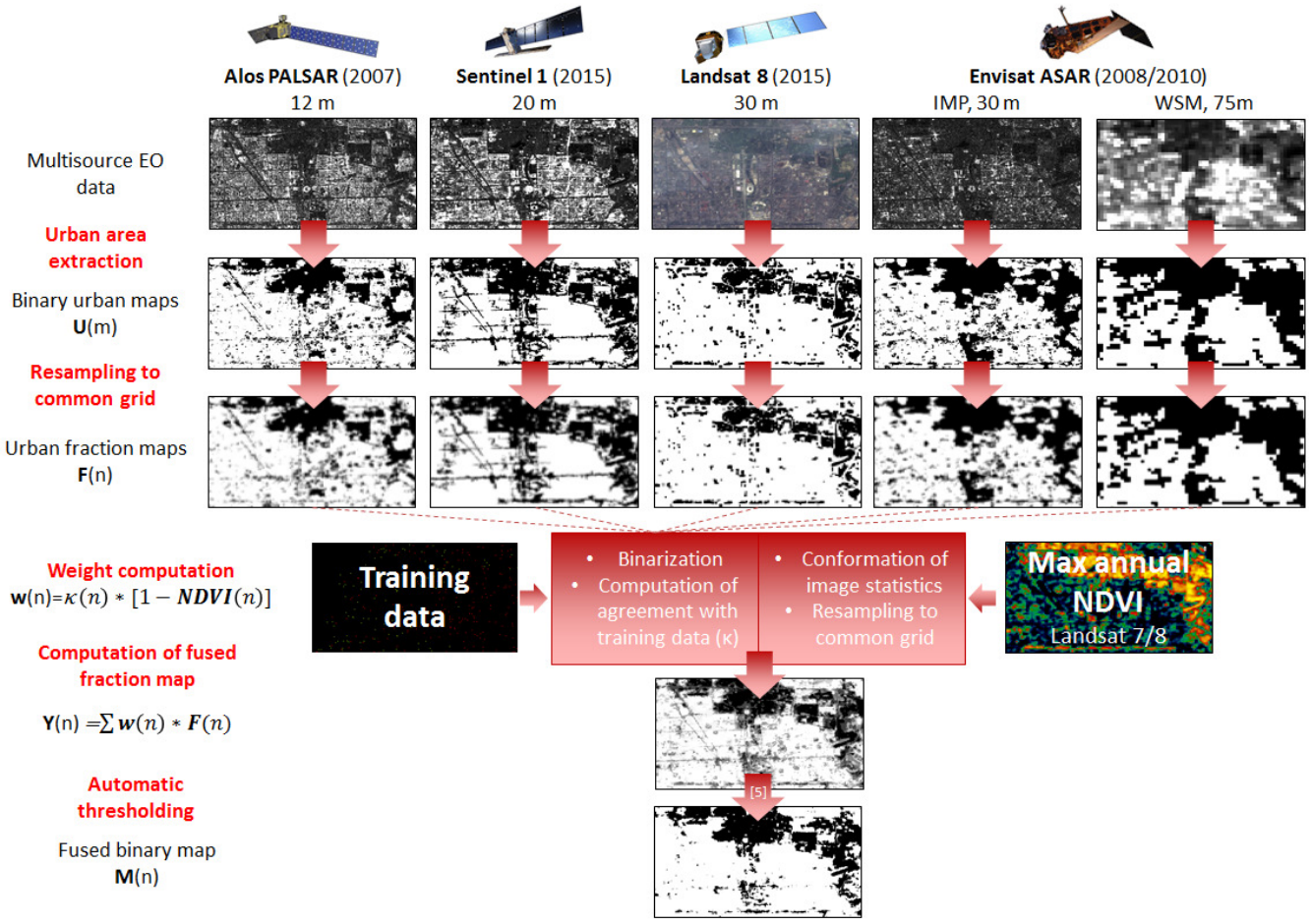

Figure 4.7: Graphical representation of the spatial regularization and image weighting fusion approach.

4.3.2.5 Accuracy assessment methodology

A quantitative accuracy assessment was obtained first with reference to an existing set of reference points (10000 points), which also has been used for evaluation in chapter 3.3. These points have been obtained from the results discussed in [79], i.e. maps of the urban expansion in P.R. China from 1990 to 2010 by human interpretation of Landsat data sets. This extraction was refined through the inclusion of Open Street Map (OSM) data. Natural features, which are stored in the OSM data base, were clipped out of the reference data in order to be able to assess the accuracy of the results with respect to the inner urban heterogeneity. In order to avoid biased accuracy assessment numbers caused by predominance of built-up areas, from the original maps the same number of random points (10000, as mentioned before) was generated for the urban as well as for the rural class. These reference data sets cover the cities of Beijing and Guangzhou and are therefore characterized by a heterogeneous mix of dense urbanization and natural

structures. In order to evaluate the results on a wider geographical area, including small human settlements in the greater metropolitan areas, a second set of reference points was generated. Starting from the urban class of GLC30 [12], a global landcover classification at a resolution of 30 m 1700 and 894 urban and rural points were randomly selected and manually checked for the greater areas of Beijing and Guangzhou, respectively. In this context urban candidate points have been discarded in case they were located in areas without any building existence, meaning that 'built-up' has been used as definition for an urban area. The rationale for this decision lies in the fact that buildings are usually the target of SAR based urban mapping procedures and they are also included in optical-based urban area extractions, in addition to artificial surfaces without a vertical component. Based on the size of the study area, the number of total reference points has been adapted appropriately for the Guangzhou test site in order to guarantee the same point density among the study areas. The set of reference points has subsequently been separated into training points (80 %), half of which are randomly picked and used for the calculation of image weights, and test points (20 %). Table 4.3 shows a summary of the reference data set.

Table 4.3: Summary of reference data sets

	Beijing	**Guangzhou**
Area of test site (km^2)	11165	6079
Total number of reference points	3400 (1700 urban, 1700 rural)	1788
Number of training points (80%)	2720	1430
Number of test points (20%)	680	358

4.3.3 Experimental results

The numbers in table 4.4 show that at both levels, the OA values are improved by the proposed methodology in a significant manner with respect to the original, stand-alone extractions. Looking at the wider geographical test set, OA values range between 95.2 and 96.1 % for the fused urban maps, while the original urban maps do not exceed 89.7 %, with improvements not smaller than 5.5 %. At the level of the finer features of the inner city, OA values are in general smaller than in the previous case, due to the high number of validation points which rely on external data sources and have not been checked manually. However, also on the city level it is clear that the proposed method delivers significantly improved urban maps. As a matter of fact, the OA values for the fused maps range between 82.3 and 85.5 % while the best original extraction reaches only 77 %.

In case training data and maximum annual NDVI maps are not available for the test area, the first step of the proposed approach can still be used to integrate the information from the input data and produce a fused urban map. The input data are rescaled to a common grid using the pixel aggregation technique and the corresponding urban fraction

Table 4.4: Accuracy assessment results of the spatial regularization and image weighting approach

			Metropolitan area		City core area	
Test area	**Product**	**Resolution [m]**	**OA** [%]	$\hat{k}$	**OA** [%]	$\hat{k}$
Beijing	Palsar	12	84.85	0.70	74.83	0.50
	Sentinel1	20	86.47	0.73	76.09	0.52
	Landsat8	30	89.71	0.79	73.63	0.47
	ASAR IMP	30	76.76	0.54	76.41	0.53
	ASAR WSM	75	61.76	0.24	75.28	0.51
	Fused - majority	20	93.24	0.86	78.99	0.58
	Fused - mean	20	93.97	0.88	80.66	0.61
	Fused - $\kappa(n)$	20	95.03	0.90	82.53	0.63
	Fused - w(n)	20	95.98	0.92	84.59	0.66
	Fused SAR only - w(n)	20	89.62	0.80	84.24	0.65
	Fused- majority	75	94.26	0.89	78.61	0.57
	Fused- mean	75	93.97	0.88	81.94	0.64
	Fused- $\kappa(n)$	75	95.12	0.90	84.03	0.66
	Fused - w(n)	75	96.09	0.92	85.52	0.68
	Fused SAR only- w(n)	75	92.18	0.84	84.54	0.67
Guangzhou	Palsar	12	78.49	0.57	68.81	0.38
	Sentinel1	20	80.17	0.60	70.81	0.42
	Landsat8	30	88.27	0.77	76.99	0.54
	ASAR IMP	30	79.89	0.60	74.33	0.49
	ASAR WSM	75	74.58	0.49	74.12	0.48
	Fused- majority	20	89.66	0.79	75.70	0.57
	Fused- mean	20	89.11	0.78	80.66	0.61
	Fused- $\kappa(n)$	20	93.02	0.86	81.10	0.62
	Fused- w(n)	20	95.47	0.91	82.33	0.64
	Fused SAR only- w(n)	20	89.89	0.78	80.97	0.60
	Fused- majority	75	87.99	0.76	78.32	0.57
	Fused- mean	75	89.11	0.78	81.22	0.62
	Fused- $\kappa(n)$	75	91.34	0.83	82.33	0.63
	Fused- w(n)	75	95.15	0.90	84.27	0.66
	Fused SAR only- w(n)	75	91.79	0.83	83.08	0.63

maps are then averaged in a pixel wise manner in order to produce a fused raster. Since this intermediate result contains continuous values it has been translated into a binary map, again using the same image moment preserving thresholding technique [116]. In addition – and for the sake of comparison – the input data sets have also been rescaled using the nearest neighbor technique, the standard resampling technique for classified data. The resulting, common resolution maps have subsequently been combined using majority voting. The average threshold that has been computed for the fused urban fraction maps (≥ 0.39) has been adopted for the majority voting approach adequately so that a pixel has the classified as urban at least in two of the input data sets in order to be urban

in the fused result. Accuracy assessment results show that on the metropolitan level no significant difference between the two methods can be observed. However, on the city core the OA values reached by results delivered by spatial regularization and averaging are significantly larger, resulting into an improvement between 1.7 and 5 %. Thus, even though the methods deliver similar results with respect to the detection of big cities and small settlements, the pixel aggregation is apparently more capable to capture the spatially finer heterogeneous mix of artificial and natural structures in the core of urban areas.

Additionally, another experiment was conducted, with the aim to single out the impact of the reference data based algorithm training on the final data fusion product. Accordingly, only the first factor $\kappa(n)$ of the final image weight $w(n)$ was applied on the input data, effectively considering in this test only the urban fraction maps generated by the preceding spatial regularization method. The merged and binarized weighted urban fraction maps reach OA values between 91.3 and 95.1 % at the metropolitan level and between 81.1 and 84 % at the city core level, respectively. The average improvement due to the training phase of the proposed approach is 2.1 % for the metropolitan area and 1.4 % for the inner city with respect to the urban map generated through simple averaging of the urban fraction maps.

As described in section 4.3.2.3 the final weight of each input image is determined using $\kappa(n)$ as well as annual maximum NDVI maps to account for the environmental characteristics in the study area at the time of image acquisition. A comparison between the fused urban maps using only $\kappa(n)$ and the final weight $w(n)$ shows that due to the inclusion of the NDVI the accuracy of the maps could be further improved by 2.1 % in the metropolitan areas and by 1.7 % in the city core level. Therefore, it can be concluded that maximum annual NDVI maps can be used directly as image weight in order to improve the fused urban area extraction results.

Finally, a last experiment was performed with the aim to find out if similar urban map accuracy can be achieved by using only SAR data. This question is especially interesting since, as clear from the numbers in table 4.4, the best original input urban map are obtained by means of the exploitation of the multi-spectral Landsat 7/8 composite. The results of this test show that, while the final accuracy of SAR-only fusion results do not reach the same accuracy values than the fused maps when all input data are considered, it is still possible to observe a significant improvement. Therefore, it is proved that the proposed approach is useful to enhance the quality of urban maps, irrespectively of the data types used to produce the input data sets.

To further visually stress this discussion, figure 4.6 shows the final fused urban map for Guangzhou as well as the intermediate results discussed in this section.

As a final remark it must be noted that the results of the conducted experiments indicate that even low quality data can contain valuable information and contribute to an improvement of the mapping results. Thus, it is valid to state that the more data is being incorporated into analysis, the better the results will be. However, due to several possible reasons including limited data availability or a lack of computing resources not

all the data used in this chapter might be incorporated on the global scale.

4.4 Chapter conclusions

A pixel- and object-based approach to fuse urban maps from multi-source EO data is proposed in the first part of this chapter. Two separate and independent methods to extract information about urban area extents are used in order to update the 2010 version of the CCI-LC map. First of all, the UEXT algorithm, fine tuned for Sentinel 1 data, has been applied, and the resulting maps serves as one of the inputs to the fusion approach. Then, the UADP method, developed at the University of Jena, has been applied to the same data and the resulting urban probability maps have been used to update the urban class of the CCI-LC product, eventually leading to a binary urban map, incorporated into the fusion scheme as well. Based on the method-specific advantages and drawbacks, a knowledge based fusion strategy has been developed and applied in five test sites located in arid areas. Results show that the proposed decision level data fusion approach is capable of significantly improving the urban area maps with respect to the urban class of the CCI-LC product. Due to the fact that dense and/or mature vegetation exhibits very similar SAR backscatter characteristics as the ones from built-up areas, the fused results are prone to confusion between urban and agricultural areas. In order to minimize this problem it is necessary to include SAR data sets whose acquisition times spread throughout the whole annual phenological cycle.

In the second part of the chapter a novel multi-scale and -source data fusion approach for the purpose of urban area extraction is presented. Experiments in the test areas of Beijing and Guangzhou reveal that the proposed method delivers significantly enhanced results with improvement rates of at least 5 % with respect to the original stand-alone extractions on the metropolitan as well as on the city core level. The first step of the proposed method is spatial regularization using the "pixel aggregation" rescaling technique, which is capable of delivering improved urban area mapping results in comparison to standard nearest neighbor resampling and majority voting. The training phase of the algorithm aims at the inclusion of image weights based on the performance of the spatially regularized versions of the input data sets. The corresponding factor $\kappa(n)$ is capable of improving the OA of urban maps by 2.09 and 1.38 % OA on the metropolitan and city scale, respectively, when used as weight for the urban fraction maps (i.e. results of pixel aggregation). Through additional inclusion of maximum annual NDVI maps into the final image weights the accuracy and reliability of the extracted urban area maps is further enhanced by similar rates. Another advantage of the proposed method, namely the fact that it is robust with respect to the location of the study area as long as the quality of the input data sets is similar, has been highlighted in this work. Finally, it has been proven that the method works irrespective of the data source of the input urban maps since significantly improved urban maps could be produced also with input data coming from SAR data only.

Chapter 5

Conclusions

Information about location and extent of human settlements is valuable for a variety of research areas and applications. In this context remote sensing offers a very powerful tool to generate this information in a time- and cost-efficient manner. Even though the quality of global urban maps has tremendously improved since the early 1990s, none of today's global layers is flawless and their usefulness for specific applications might be limited due to a number of reasons. Most of the available global products exploit very coarse single scale EO and/or ground based data, are restricted to a specific time or do not make use of complementary information coming from different sensor types. To overcome this issues and to take advantage of the wealth of available EO data, a variety of multi-scale, multi-temporal and multi-source data fusion methods for the purpose of urban area extraction have been developed and are presented in this thesis.

A first important result of this thesis is a technique to exploit multi-temporal SAR data of the latest generation, maintaining continuity and back-comparability with results obtained with coarser resolution SAR data. Indeed, the UEXT algorithm has proven its effectiveness and robustness for the extraction of urban areas from coarse scale Envisat ASAR WSM data. However, since the Envisat mission ended in 2012 and new satellites equipped with state-of-the-art SAR sensors are in orbit, data with increased spatial, radiometric and temporal resolution are available. In order to respond to the new data situation the UEXT algorithm has been adapted and tuned for an application on Sentinel 1 data. The same algorithm has also been improved in order to overcome issues connected to topography and the number of acquisitions combined into a multi-temporal data stack. The work of this thesis shows that the new version of UEXT applied on multi-temporal data composites is very robust with respect to different climate regions and related environmental characteristics and that it delivers improved urban area maps in comparison to the previous version of UEXT applied to ASAR WSM data.

A second highlight of this thesis is the general framework for multi-scale SAR data fusion for urban area mapping. Within this framework 5 different fusion methods at the pixel-, feature, and decision level, offering solutions to different data availability scenarios, have been implemented and tested on four megacity testsites in P.R. China and Brazil. The main conclusion that can be drawn from these experiments is that the inclusion of

suitably designed image weights can significantly improve the extraction results. However, the best results obtained from the HMW method (see section 3.3.3) does not prevent the other approaches from generating improved urban mapping results in comparison to stand-alone extractions exploiting each input data separately. The experiments clearly demonstrate the advantage of incorporating multi-scale data sets into the analysis, and achieve more accurate human settlement maps.

Finally, this thesis faced the issue that, depending on the sensor type and sensor geometry the actual urban features that are being mapped from space are slightly different, because the information content of the satellite images is to some extent complementary. Data fusion methods of imagery from active and passive sensors are needed in order to exploit the full potential of available data. To that aim two multi-source data fusion approaches have been developed.

The first method is designed to update an existing very coarse resolution (300m) optical-based urban map using new generation Sentinel 1 SAR data. To that aim two independent SAR-based urban area extraction algorithms - namely UEXT from the University of Pavia and UADP from the University of Jena, Germany - are used to locate and delineate urban areas. Subsequently, the resulting urban maps are used to update the existing optical-based urban extent map using a pixel- and object based decision level data fusion approach. The advantages and drawbacks of each input data set have been evaluated and decision rules have been defined based on them. Results show that not only big urban agglomerations are reflected with more detail in the fused results, but also small human settlements are correctly detected due to the fine spatial resolution of the SAR data.

A second method exploits multi-scale and multi-source EO data using a spatial regularization and image weighting approach. Pixel aggregation is used as an alternative to classic methods to rescale urban area extraction results from multi-scale SAR and optical EO data to a common grid. Subsequently, the resulting urban fraction maps are weighted and combined based on their scale depended performance as well as environmental characteristics at the time of image acquisition. The fused results show that the proposed method is capable of integrating the complementary information from SAR- and optical-based urban maps and exploiting the advantages of each image type due to carefully determined image weights.

As a final remark, the research conducted in this thesis aimed at a better understanding of the integration of multi-scale EO data for the purpose of urban area mapping on the global scale. The possibilities to fuse EO data sets are manifold and the main challenge is to take advantage of each input data set while minimizing their disturbing effects. The proposed multi-scale and multi-source data fusion techniques offer options for a variety of data availability scenarios and are well suited to be expanded in the future. Upcoming satellite missions and cloud computing infrastructures will open new perspective in the field of EO data fusion, and will enable analyses of global urbanization at an unprecedented level of detail.

Bibliography

[1] *United Nations world population prospects: The 2015 revision, key findings and advance tables*, http://www.un.org/en/development/desa/population/events/other/10/index.shtml, [Online; accessed October 2016].

[2] *United Nations Habitat 3 conference: The new urban agenda*, https://habitat3.org/the-new-urban-agenda, [Online; accessed October 2016].

[3] D. Potere and A. Schneider, "A critical look at representations of urban areas in global maps," *GeoJournal*, vol. 69, no. 1-2, pp. 55–80, 2007.

[4] *United Nations Habitat Conference: The Vancouver declaration on human settlements*, 1976.

[5] *United Nations Demographic Yearbook 2014*, http://unstats.un.org/unsd/demographic/products/dyb/dyb2014.htm, [Online; accessed October 2016].

[6] M. Brockerhoff, *An urbanizing world*. Population Reference Bureau, 2000.

[7] A. Schneider, M. A. Friedl, and D. Potere, "A new map of global urban extent from MODIS satellite data," *Environmental Research Letters*, vol. 4, no. 4, p. 044 003, 2009.

[8] T. Esch, M. Marconcini, A. Felbier, A. Roth, W. Heldens, M. Huber, M. Schwinger, H. Taubenböck, A. Müller, and S. Dech, "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, 2013.

[9] M. Pesaresi, X. B. H. Guo, D. Ehrlich, S. Ferri, L. Gueguen, M. Halkia, M. Kauffmann, T. Kemper, L. Lu, M. Marin-Herrera, G. Ouzounis, M. Scavazzon, P. Soille, V. Syrris, and L. Zanchetta, "A Global Human Settlement Layer From Optical HR/VHR RS Data: Concept and First Results.," *IEEE J. Selected Topic Appl. Earth Obs. Remote Sens.*, vol. 6, no. 5, pp. 2102–2130, 2013.

[10] E. Bartholomé and A. Belward, "GLC2000: A new approach to global land cover mapping from earth observation data," *International Journal of Remote Sensing*, vol. 26, no. 9, pp. 1959–1977, 2005.

[11] O. Arino, D. Gross, F. Ranera, M. Leroy, P. Bicheron, C. Brockman, P. Defourny, C. Vancutsem, F. Achard, L. Durieux, *et al.*, "Globcover: ESA service for global land cover from meris," in *2007 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 2007, pp. 2412–2415.

[12] J. Chen, J. Chen, A. Liao, X. Cao, L. Chen, X. Chen, C. He, G. Han, S. Peng, M. Lu, *et al.*, "Global land cover mapping at 30m resolution: A POK-based operational approach," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 103, pp. 7–27, 2015.

[13] C. D. Elvidge, B. T. Tuttle, P. C. Sutton, K. E. Baugh, A. T. Howard, C. Milesi, B. Bhaduri, and R. Nemani, "Global distribution and density of constructed impervious surfaces," *Sensors*, vol. 7, no. 9, pp. 1962–1979, 2007.

[14] D. M. Danko, "The digital chart of the world project," *Photogrammetric engineering and remote sensing*, vol. 58, no. 8, pp. 1125–1128, 1992.

[15] A. Schneider, M. A. Friedl, D. K. McIver, and C. E. Woodcock, "Mapping urban areas by fusing multiple sources of coarse resolution remotely sensed data," *Photogrammetric Engineering & Remote Sensing*, vol. 69, no. 12, pp. 1377–1386, 2003.

[16] A. Schneider, M. A. Friedl, and D. Potere, "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, 2010.

[17] T. S. Pagano and R. M. Durham, "Moderate resolution imaging spectroradiometer (MODIS)," in *Optical Engineering and Photonics in Aerospace Sensing*, International Society for Optics and Photonics, 1993, pp. 2–17.

[18] C. D. Elvidge, M. L. Imhoff, K. E. Baugh, V. R. Hobson, I. Nelson, J. Safran, J. B. Dietz, and B. T. Tuttle, "Night-time lights of the world: 1994–1995," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 56, no. 2, pp. 81–99, 2001.

[19] G. Baudin, R. Bessudo, and J.-L. Bezy, "Medium-resolution imaging spectrometer," in *Garmisch-DL tentative*, International Society for Optics and Photonics, 1994, pp. 115–125.

[20] S. Bontemps, P. Defourny, J. Radoux, E. Van Bogaert, C. Lamarche, F. Achard, P. Mayaux, M. Boettcher, C. Brockmann, G. Kirches, *et al.*, "Consistent global land cover maps for climate modelling communities: Current achievements of the ESA's land cover CCI," in *Proceedings of the ESA Living Planet Symposium, Edimburgh*, 2013, pp. 9–13.

[21] CIESIN, "Ciat global rural-urban mapping project (GRUMP)," *Alpha Version: Population grids*, 2004.

[22] B. Bhaduri, E. Bright, P. Coleman, and J. Dobson, "Landscan," *Geoinformatics*, vol. 5, no. 2, pp. 34–37, 2002.

[23] K. K. Goldewijk, "Estimating global land use change over the past 300 years: The HYDE database," *Global Biogeochemical Cycles*, vol. 15, no. 2, pp. 417–433, 2001.

[24] T. Esch, M. Marconcini, A. Felbier, A. Roth, W. Heldens, M. Huber, M. Schwinger, H. Taubenböck, A. Müller, and S. Dech, "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, 2013.

[25] M. Pesaresi, A. Gerhardinger, and F. Kayitakire, "A robust built-up area presence index by anisotropic rotation-invariant textural measure," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 1, no. 3, pp. 180–192, 2008.

[26] R. M. Haralick, K. Shanmugam, *et al.*, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, no. 6, pp. 610–621, 1973.

[27] M. Klotz, T. Kemper, C. Geiß, T. Esch, and H. Taubenböck, "How good is the map? a multi-scale cross-comparison framework for global settlement layers: Evidence from central europe," *Remote Sensing of Environment*, vol. 178, pp. 191–212, 2016.

[28] S. Sterling and A. Ducharne, "Comprehensive data set of global land cover change for land surface model applications," *Global Biogeochemical Cycles*, vol. 22, no. 3, 2008.

[29] K. Trusilova, M. Jung, G. Churkina, U. Karstens, M. Heimann, and M. Claussen, "Urbanization impacts on the climate in europe: Numerical experiments by the PSU-NCAR Mesoscale Model (MM5)," *Journal of Applied Meteorology and Climatology*, vol. 47, no. 5, pp. 1442–1455, 2008.

[30] P. Gamba, G. Lisini, G. Trianni, and E. Angiuli, "Urban area mapping using ASAR wide swath mode and Landsat data: A comparison on eastern china," in *Urban Remote Sensing Event (JURSE), 2013 Joint*, IEEE, 2013, pp. 127–130.

[31] P. Gamba and G. Lisini, "Fast and Efficient Urban Extent Extraction Using ASAR Wide Swath Mode Data.," *IEEE J. Selected Topic Appl. Earth Obs. Remote Sens.*, vol. 6, no. 5, pp. 2184–2195, 2013.

[32] U. Soergel, U. Thoennessen, and U. Stilla, "Visibility analysis of man-made objects in SAR images," in *Remote Sensing and Data Fusion over Urban Areas, 2003. 2nd GRSS/ISPRS Joint Workshop on*, IEEE, 2003, pp. 120–124.

[33] T. Esch, A. Schenk, T. Ullmann, M. Thiel, A. Roth, and S. Dech, "Characterization of land cover types in TerraSAR-X images by combined analysis of speckle statistics and intensity information," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 1911–1925, 2011.

[34] *NASA: ASTER Global Digital Elevation Model Version 2*, `http://asterweb.jpl.nasa.gov/gdem.asp`, [Online; accessed October 2016].

[35] F. Dell'Acqua, M. Stasolla, and P. Gamba, "Unstructured human settlement mapping with SAR sensors," in *Proceedings of the IEEE Geoscience and Remote Sensing Symposium, IGARSS*, 2006.

[36] T. Esch and A. Roth, "Semi-automated classification of urban areas by means of high resolution radar data," in *ISPRS 2004 Congress*, Citeseer, 2004, pp. 478–482.

[37] N. Herold, B. Haack, and E. Solomon, "Radar spatial considerations for land cover extraction," *International Journal of Remote Sensing*, vol. 26, no. 7, pp. 1383–1401, 2005.

[38] F. Dell'Acqua, P. Gamba, and G. Trianni, "Semi-automatic choice of scale-dependent features for satellite SAR image classification," *Pattern Recognition Letters*, vol. 27, no. 4, pp. 244–251, 2006.

[39] F. M. Henderson and Z.-G. Xia, "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, 1997.

[40] D. J. Weydahl, F. Bretar, and P. Bjerke, "Comparison of RADARSAT-1 and IKONOS satellite images for urban features detection," *Information fusion*, vol. 6, no. 3, pp. 243–249, 2005.

[41] A. Wiesmann, U. Wegmuller, M. Honikel, T. Strozzi, and C. L. Werner, "Potential and methodology of satellite based SAR for hazard mapping," in *Geoscience and Remote Sensing Symposium, 2001. IGARSS'01. IEEE 2001 International*, IEEE, vol. 7, 2001, pp. 3262–3264.

[42] D. Maktav, F. Erbek, and C. Jürgens, "Remote sensing of urban areas," *International Journal of Remote Sensing*, vol. 26, no. 4, pp. 655–659, 2005.

[43] M. Stasolla and P. Gamba, "Spatial indexes for the extraction of formal and informal human settlements from high-resolution SAR images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 1, no. 2, pp. 98–106, 2008.

[44] P. Gamba, M. Aldrighi, and M. Stasolla, "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, 2011.

[45] J. Ping, C. Green, R. Zartman, and K. Bronson, "Exploring spatial dependence of cotton yield using global and local autocorrelation statistics," *Field Crops Research*, vol. 89, no. 2, pp. 219–236, 2004.

[46] Y. Ban, A. Jacob, and P. Gamba, "Spaceborne SAR data for global urban mapping at 30m resolution using a robust urban extractor," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 103, pp. 28–37, 2015.

[47] Y.-L. Desnos, C. Buck, J. Guijarro, G. Levrini, J.-L. Suchail, R. Torres, H. Laur, J. Closa, and B. Rosich, "The ENVISAT advanced synthetic aperture radar system," in *Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International*, IEEE, vol. 3, 2000, pp. 1171–1173.

[48] M. Berger, J. Moreno, J. Johannessen, P. Levelt, and R. Hanssen, "ESA's sentinel missions in support of Earth system science.," *Remote Sensing of Environment*, vol. 120, pp. 84–94, 2012.

[49] Y. Zha, J. Gao, and S. Ni, "Use of normalized difference built-up index in automatically mapping urban areas from TM imagery," *International Journal of Remote Sensing*, vol. 24, no. 3, pp. 583–594, 2003.

[50] H. Xu, "A new index for delineating built-up land features in satellite imagery," *International Journal of Remote Sensing*, vol. 29, no. 14, pp. 4269–4276, 2008.

[51] A. R. Huete, "A soil-adjusted vegetation index (SAVI)," *Remote sensing of environment*, vol. 25, no. 3, pp. 295–309, 1988.

[52] X. Han-qiu, "A study on information extraction of water body with the modified normalized difference water index (MNDWI)," *Journal of Remote Sensing*, vol. 5, pp. 589–595, 2005.

[53] C. Jieli, L. Manchun, L. Yongxue, S. Chenglei, and H. Wei, "Extract residential areas automatically by new built-up index," in *2010 18th International Conference on Geoinformatics*, IEEE, 2010, pp. 1–5.

[54] G. Trianni, E. Angiuli, G. Lisini, and P. Gamba, "Human settlements from landsat data using Google Earth Engine," in *2014 IEEE Geoscience and Remote Sensing Symposium*, IEEE, 2014, pp. 1473–1476.

[55] S. Angel, J. Parent, D. L. Civco, and A. M. Blei, *Atlas of urban expansion*. Lincoln Institute of Land Policy, 2012.

[56] O. Arino, J. Ramos, V. Kalogirou, P. Defourny, and F. Achard, "Globcover 2009.," *Proc. of 2010 ESA Living Planet Symposium, Bergen, Norway*, pp. 686–689, 2010.

[57] X. Li and P. Gong, "An "exclusion-inclusion" framework for extracting human settlements in rapidly developing regions of china from Landsat images," *Remote Sensing of Environment*, vol. 186, pp. 286–296, 2016.

[58] M. Ehlers, S. Klonus, P. Johan Åstrand, and P. Rosso, "Multi-sensor image fusion for pansharpening in remote sensing," *International Journal of Image and Data Fusion*, vol. 1, no. 1, pp. 25–45, 2010.

[59] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald, "A critical comparison among pansharpening algorithms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2565–2586, 2015.

[60] L. Loncan, L. B. de Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Licciardi, M. Simoes, *et al.*, "Hyperspectral pansharpening: A review," *IEEE Geoscience and remote sensing magazine*, vol. 3, no. 3, pp. 27–46, 2015.

[61] N. Longbotham, F. Pacifici, B. Baugh, and G. Camps-Valls, "Prelaunch assessment of worldview-3 information content," in *IEEE GRSS Workshop on Hyperspectral Image and Signal Processing (WHISPERS)*, 2014, pp. 479–486.

[62] M.-A. Gleyzes, L. Perret, and P. Kubik, "Pleiades system architecture and main performances," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 39, B1, 2012.

[63] F. Covello, F. Battazza, A. Coletta, E. Lopinto, C. Fiorentino, L. Pietranera, G. Valentini, and S. Zoffoli, "COSMO-skymed an existing opportunity for observing the earth," *Journal of Geodynamics*, vol. 49, no. 3, pp. 171–180, 2010.

[64] S. Buckreuss and M. Zink, "The missions TerraSAR-X and TanDEM-X: Status, challenges, future perspectives," in *General Assembly and Scientific Symposium, 2011 XXXth URSI*, IEEE, 2011, pp. 1–1.

[65] C. Pohl and J. Van Genderen, "Multisensor image fusion in remote sensing: Concepts, methods and applications," *International Journal of Remote Sensing*, vol. 19, no. 5, pp. 823–854, 1998.

[66] J. Settle and S. Briggs, "Fast maximum likelihood classification of remotely-sensed imagery," *International Journal of Remote Sensing*, vol. 8, no. 5, pp. 723–734, 1987.

[67] M. Pal and P. Mather, "Support vector machines for classification in remote sensing," *International Journal of Remote Sensing*, vol. 26, no. 5, pp. 1007–1011, 2005.

[68] C. Huang, L. Davis, and J. Townshend, "An assessment of support vector machines for land cover classification," *International Journal of remote sensing*, vol. 23, no. 4, pp. 725–749, 2002.

[69] D. M. Miller, E. J. Kaminsky, and S. Rana, "Neural network classification of remote-sensing data," *Computers & Geosciences*, vol. 21, no. 3, pp. 377–386, 1995.

[70] P. M. Atkinson and A. Tatnall, "Introduction neural networks in remote sensing," *International Journal of remote sensing*, vol. 18, no. 4, pp. 699–709, 1997.

[71] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognition Letters*, vol. 27, no. 4, pp. 294–300, 2006.

[72] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31, 2016.

[73] T. M. Pellizzeri, P. Gamba, P. Lombardo, and F. Dell'Acqua, "Multitemporal/multiband SAR classification of urban areas using spatial analysis: Statistical versus neural kernel-based approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 10, pp. 2338–2353, 2003.

[74] B. Tso and P. Mather, "Crop discrimination using multi-temporal SAR imagery," *International Journal of Remote Sensing*, vol. 20, no. 12, pp. 2443–2460, 1999.

[75] H. Skriver, F. Mattia, G. Satalino, A. Balenzano, V. R. Pauwels, N. E. Verhoest, and M. Davidson, "Crop classification using short-revisit multitemporal SAR data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 4, no. 2, pp. 423–431, 2011.

[76] T. Nagler and H. Rott, "Retrieval of wet snow by means of multitemporal SAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 2, pp. 754–765, 2000.

[77] S. Quegan, T. Le Toan, J. J. Yu, F. Ribbes, and N. Floury, "Multitemporal ERS SAR analysis applied to forest mapping," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 2, pp. 741–753, 2000.

[78] J.-S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE transactions on pattern analysis and machine intelligence*, no. 2, pp. 165–168, 1980.

[79] L. Wang, C. Li, Q. Ying, X. Cheng, X. Wang, X. Li, L. Hu, L. Liang, L. Yu, H. Huang, *et al.*, "China's urban expansion from 1990 to 2010 determined with satellite remote sensing," *Chinese Science Bulletin*, vol. 57, no. 22, pp. 2802–2812, 2012.

[80] M. Costantini, A. Farina, and F. Zirilli, "The fusion of different resolution SAR images," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 139–146, 1997.

[81] J. R. Orlando, R. Mann, and S. Haykin, "Classification of sea-ice images using a dual-polarized radar," *IEEE Journal of Oceanic Engineering*, vol. 15, no. 3, pp. 228–237, 1990.

[82] Y. Jin, Y. Ruliang, and H. Ruohong, "Pixel level fusion for multiple SAR images using PCA and wavelet transform," in *2006 CIE International Conference on Radar*, IEEE, 2006, pp. 1–4.

[83] G. Simone, A. Farina, F. C. Morabito, S. B. Serpico, and L. Bruzzone, "Image fusion techniques for remote sensing applications," *Information fusion*, vol. 3, no. 1, pp. 3–15, 2002.

[84] A. Salentinig, "Multi-scale SAR image fusion at the feature level for enhanced urban area extraction," in *Dragon 3Mid Term Results*, vol. 724, 2014.

[85] G. Simone, F. Morabito, and A. Farina, "Radar image fusion by multiscale kalman filtering," in *Information Fusion, 2000. FUSION 2000. Proceedings of the Third International Conference on*, IEEE, vol. 2, 2000, WED3–10.

[86] P. Kumar, "A multiple scale state-space model for characterizing subgrid scale variability of near-surface soil moisture," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 1, pp. 182–197, 1999.

[87] A. Ismail, X. Gao, and C. Deng, "SAR image classification based on texture feature fusion," in *Signal and Information Processing (ChinaSIP), 2014 IEEE China Summit & International Conference on*, IEEE, 2014, pp. 153–156.

[88] F. Dell'Acqua and P. Gamba, "Discriminating urban environments using multiscale texture and multiple SAR images," *International Journal of Remote Sensing*, vol. 27, no. 18, pp. 3797–3812, 2006.

[89] A. Salentinig and P. Gamba, "A fuzzy fusion approach for improved urban area detection in multi-resolution SAR data," in *2015 Joint Urban Remote Sensing Event (JURSE)*, IEEE, 2015, pp. 1–4.

[90] A. H. S. Solberg, A. K. Jain, and T. Taxt, "Multisource classification of remotely sensed data: Fusion of Landsat TM and SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 4, pp. 768–778, 1994.

[91] A. Bouakache, R. Khedam, N. Abbas, Y. A. Abdesselam, and A. Belhadj-Aissa, "Multi-scale satellite images fusion using dempster shafer theory," in *Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008. 3rd International Conference on*, IEEE, 2008, pp. 1–6.

[92] A. Makarau, G. Palubinskas, and P. Reinartz, "Multi-sensor data fusion for urban area classification," in *2011 Joint Urban Remote Sensing Event*, IEEE, 2011, pp. 21–24.

[93] M. C. Dobson, L. E. Pierce, and F. T. Ulaby, "Knowledge-based land-cover classification using ERS-1/JERS-1 SAR composites," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 1, pp. 83–99, 1996.

[94] A. Salentinig and P. Gamba, "Fusing SAR urban extent extractions at multiple spatial resolutions," in *2014 IEEE Geoscience and Remote Sensing Symposium*, IEEE, 2014, pp. 199–202.

[95] Y. Ban and A. Jacob, "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, pp. 1998–2006, 2013.

[96] A. Salentinig and P. Gamba, "A general framework for urban area extraction exploiting multiresolution SAR data fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 5, pp. 2009–2018, 2016.

[97] R. C. Gonzalez and E. Richard, "Woods, digital image processing," *Ed: Prentice Hall Press, ISBN 0-201-18075-8*, 2002.

[98] B. F. T. Rudorff, D. A. Aguiar, W. F. Silva, L. M. Sugawara, M. Adami, and M. A. Moreira, "Studies on the rapid expansion of sugarcane for ethanol production in são paulo state (brazil) using Landsat data," *Remote sensing*, vol. 2, no. 4, pp. 1057–1076, 2010.

[99] J. Rice, *Mathematical statistics and data analysis*. Nelson Education, 2006.

[100] S. S. Saatchi, J. V. Soares, and D. S. Alves, "Mapping deforestation and land use in Amazon rainforest by using SIR-C imagery," *Remote Sensing of Environment*, vol. 59, no. 2, pp. 191–202, 1997.

[101] M. C. Peel, B. L. Finlayson, and T. A. McMahon, "Updated world map of the Köppen-Geiger climate classification," *Hydrology and earth system sciences discussions*, vol. 4, no. 2, pp. 439–473, 2007.

[102] P. Gamba and M. Aldrighi, "SAR Data Classification of Urban Areas by means of Segmentation Techniques and Ancillary Optical Data.," *IEEE Journal on Selected Topics in Applied Earth Observation and Remote Sensing*, vol. 5, no. 4, pp. 1140–1148, 2012.

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

[104] S. Le Hegarat-Mascle, I. Bloch, and D. Vidal-Madjar, "Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 1018–1031, 1997.

[105] A. Bendjebbour, Y. Delignon, L. Fouque, V. Samson, and W. Pieczynski, "Multisensor image segmentation using Dempster-Shafer fusion in markov fields context," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 8, pp. 1789–1798, 2001.

[106] S. B. Serpico and F. Roli, "Classification of multisensor remote-sensing images by structured neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, no. 3, pp. 562–578, 1995.

[107] L. Bruzzone, D. F. Prieto, and S. B. Serpico, "A neural-statistical approach to multitemporal and multisource remote-sensing image classification," *IEEE Transactions on Geoscience and remote Sensing*, vol. 37, no. 3, pp. 1350–1359, 1999.

[108] G. J. Briem, J. A. Benediktsson, and J. R. Sveinsson, "Multiple classifiers applied to multisource remote sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2291–2299, 2002.

[109] B. Waske and J. A. Benediktsson, "Fusion of support vector machines for classification of multisensor data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 3858–3866, 2007.

[110] Y. Ban, H. Hu, and I. Rangel, "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, 2010.

[111] Y. Ban, O. Yousif, and H. Hu, "Fusion of SAR and optical data for urban land cover mapping and change detection," 2014.

[112] G. Palubinskas, "How to fuse optical and radar imagery?" In *2012 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 2012, pp. 2171–2174.

[113] C. Schmullius, C. Pathe, T. Riedel, P. Gamba, G. Lisini, and M. Santoro, "Deriving a global urban extent map from Envisat ASAR wide swath mode," in *Proceedings of the Living Planet Symposium 2013*, ESA, 2013.

[114] G. Trianni, G. Lisini, E. Angiuli, E. Moreno, P. Dondi, A. Gaggia, and P. Gamba, "Scaling up to national/regional urban extent mapping using Landsat data .," *IEEE J. of Selected Topics in Applied Earth Observation and Remote Sensing*, vol. 8, no. 7, pp. 3710–3719, 2015.

[115] A. Salentinig and P. Gamba, "Multiscale multisensor decision level data fusion for urban mapping," in *Fourth International Workshop on Earth Observation and Remote Sensing Applications*, 2016, in press.

[116] W.-H. Tsai, "Moment-preserving thresolding: A new approach," *Computer Vision, Graphics, and Image Processing*, vol. 29, no. 3, pp. 377–393, 1985.