Automatic characterization of sub-urban environments by using VHR X-band COSMO-SkyMed images

Chiara Pratola

A.A. 2012-2013

Supervisor: **Dr. Fabio Del Frate**
Coordinator: **Prof. Giovanni Schiavon**

ii

Contents

Introduction

The world population of 7.2 billion is projected to increase by almost one billion people within the next twelve years, reaching 8.1 billion in 2025 and 9.6 billion in 2050 [1]. It has been foreseen that most of the population growth will occur in developing countries, while elsewhere no significant changes will be expected. Social, economic and environmental conditions are leading population to worldwide emigration, thus causing dramatic expansions of cities, and severe damages to the natural ecosystem due to the consumption of natural resources, disordered changes in land uses, worsening air and water quality.
In such a scenario, the authorities are urgently called to face new challenges in the urban planning and organization, and in resources management. Therefore, it is of vital importance the availability of suitable tools to monitor the urbanization process and get reliable information in populated areas, in order to contribute positively to the urban organization and strategic decision making, as well as to timely act when the quality of the environment is in danger. Knowledge about urban dynamics are usually derived from field surveys, aerial pictures or national censuses, which are costly and time consuming techniques, consequently leading to delays in maps updates and to lacking of detailed information. Alternatively, one of the best source of information related to the land cover and land uses changes in time is represented by the remote sensing, which allow to periodically collect images over wide areas. However, because of their large spatial and spectral variability, mapping human settlements is still one of the most challenging remote sensing task.

Observation in the optical band is broadly used to monitor land cover and its changes. Thanks to the recent launch of satellite equipped with multi-spectral and hyper-spectral sensors (e.g. WorldView-2, PROBA CHRIS, PRISMA), a significant improvement in discriminating between spectrally similar surfaces is expected, whereas the high and very high spatial resolution characterizing the acquired images allows to identify small-scale objects. However, unpredictable long periods of anomalous cloud cover or thick fog hinder regular use of optical images. Moreover, in case of emergency, the availability of images indicating the changes undergone by natural surfaces or man-made structures at the earliest convenience is vital, independently of sky conditions and illumination. As a consequence, to meet the requirements of promptness, timeliness and reliability, use of synthetic aperture radar (SAR) must be considered.

A crucial step forward in Earth observation has been allowed by the recent (2011) availability of SARs on the COSMO-SkyMed (CSK) satellite constellation, operated by the Italian Space Agency (ASI). In fact, the four CSK X-band SAR sensors now in orbit are able to provide images not only at 1 m spatial resolution, but also with a very short revisit time, presently down to 12 hours, irrespective of cloud cover and light conditions. To fully profit from the unique capabilities of the CSK observing system, adequate exploitation of the information contained by the metric-resolution multi-temporal SAR images is necessary. In particular, the large amount of data calls for the urgent development of suitable automatic techniques to manage in near-real time the information on land cover and its changes which are provided by the SAR observations.

The purpose of this work is to give a substantial contribution to the development of

fast, automatic and accurate algorithms performing the characterization of a sub-urban environment, through the land cover classification, buildings identification and the detection of changes, by using VHR SAR X-band images provided by the COSMO-SkyMed constellation. Basically, the features extraction has been carried out through both pixels-based techniques characterized by a supervised approach, and object-based methods, performed by unsupervised algorithms.

The thesis is divided into six parts:

Part I: *The satellite system*
After a general overview on SAR system and data, with a focus on the newest VHR SAR sensors (*Chapter 1*), the novelties introduced by the COSMO-SkyMed mission are described in *Chapter 2*.

Part II: *Algorithms and methodologies*
The different approaches adopted for each application are firstly described. Secondly, each algorithm and methodology, which has been implemented and opportunely tailored in order to perform the features extraction, is discussed in the following chapters. In particular, the Multi Layer Perceptron Neural Network (MLP-NN) exploited for the pixel-based classification is discussed in *Chapter 3*. Then, *Chapter 4* is dedicated to the unsupervised Pulse Coupled Neural Network (PCNN), implemented for the building detection task. Finally, a common methodology aiming at the extraction of texture information, that is the computation of the Gray Level Co-occurrence Matrix (GLCM), is presented in *Chapter 5*

Part III: *Dataset*
The test sites are described in *Chapter 6*. Subsequently, the list of exploited COSMO-SkyMed images is provided in *Chapter 7*, together with the characteristics of the data. The same information is given in *Chapter 8* for the Multispectral WorldView-2 and Quickbird images, which have been mainly used as a ground truth reference, or have been jointly exploited with the SAR acquisitions in data fusion experiments.

Part IV: *Supervised approach: pixel-based classification*
A supervised pixel-based approach has been adopted to perform the land cover classification of VHR COSMO-SkyMed images. In order to define a minimum number of classes to be discriminated by exploiting single polarized data, several pixels have been collected over different surfaces. The correspondent backscattering behaviors are statistically analyzed in *Chapter 9*. Then the classification exercise is presented in *Chapter 10*, where four different sets of input features are tested. Finally, in *Chapter 11* the joint use of VHR SAR and optical data is investigated, aiming at the enhancement of the classification of vegetated areas.

Part V: *Unsupervised approach: object-based features extraction*
The object based approach has been performed by an unsupervised Pulse Coupled Neural Network based algorithm, which has been implemented to quickly identify buildings in suburban environments imaged by COSMO-SkyMed SAR sensors. The description of the developed technique, as well as the attained results, are presented in *Chapter 12*.

Part VI: *Change detection*
　　The pixel-based and the object-based approaches have been combined in a novel change detection algorithm. The technique and the achieved results are shown in *Chapter 13.*

Part I

The Satellite System

Introduction

Synthetic Aperture Radar (SAR) systems have been widely used in many remote sensing applications for more than thirty years. The discovery and developments of SAR systems in the 50s and 60s were dominated by military research and use for survey purposes and man-made target detection. Successively, in the 70s and 80s several airborne systems were developed for civilian applications.

The first civilian SAR satellite was mounted on Seasat, which was launched in 1978. Since that moment, a huge step forward has been done in terms of used technology, image quality, spatial and temporal resolution. While formerly SAR images were mainly interesting from the science and engineering point of view, the last years have marked an important transition in the application of SAR. Today the public awareness of the usefulness of radar remote sensing beyond science is much higher, so that, thanks to their unique imaging capability, which is independent from daylight, cloud coverage and weather conditions [2], SAR satellites are meant to perform disaster monitoring, mitigation and damage assessment.

Nowadays more than fifteen spaceborne SAR sensors are being operated; among them, the recent TerraSAR-X, TanDEM-X, the COSMO-SkyMed constellation and Radarsat-2 belong to the new generation of SAR satellites, providing very high spatial resolution imagery. But the research keeps going on, and the launch of more than ten new improved SAR systems is foreseen within the next five years.

In the following chapters a summary of the SAR principles (§1.1) and SAR images characteristics (§1.2) will be provided. Furthermore, notions about SAR image pre-processing will be reminded in §1.3 . A general overview about the most recent SAR satellite systems, which are able to acquire very high resolution imagery, will be addressed in the §1.4. Finally, a more detailed description of the COSMO-SkyMed mission, and of the second generation of COSMO-SkyMed (CSG), will be provided in Chapter 2.

Chapter 1

Fundamentals

A radar (radio detection and ranging) system is an active instrument, which transmits electromagnetic pulses in the microwave frequencies, and receives the echoes of the scattered signal in a sequential way. The signal interacts with the Earth surface, and only a portion of the transmitted power is scattered to the receiving antenna, which can be the same as the transmitter (monostatic systems), or different (bi- or multistatic systems). In the following, only monostatic systems will be taken into account, so that the received scattering will be named *backscattering*. The amplitude and phase of the backscattered signal depend on the characteristics of the target, in terms of physical attributes, such as geometry or roughness, and electromagnetic properties (i.e. permittivity).

An isotropic radiator emits energy, which propagates uniformly in all directions. The non-directional power density S_N at the imaged object located at a distance r from the antenna, is given by:

$$S_N = \frac{P_T}{4 \cdot \pi \cdot r^2} \tag{1.1}$$

where P_T is the transmitted power. Radar antennas are normally directional, aiming at the increase of the power density in the direction of a small beam. Such augment is represented by the so called antenna gain G:

$$G = \frac{4 \cdot \pi \cdot A_A \cdot K_a}{\lambda^2} \tag{1.2}$$

where A_A represents the geometric antenna area, K_a the efficiency of the antenna, and λ the wavelength of the electromagnetic signal. The directional power density S_D at the target is obtained by:

$$S_D = S_N \cdot G \tag{1.3}$$

The power backscattered (P_R) by the target towards the receiver is a function of both the power density at the target itself, and the so called Radar Cross Section(RCS) σ:

$$P_R = S_D \cdot \sigma \tag{1.4}$$

being σ defined as:

$$\sigma = 4 \cdot \pi \cdot r^2 \cdot \frac{P_R}{P_I} \tag{1.5}$$

where P_I indicates the incidence power. According with the hypothesis of a monostatic radar, and assuming an equal antenna gain for the receiving and transmitting systems, the power received at the sensor can be expressed as:
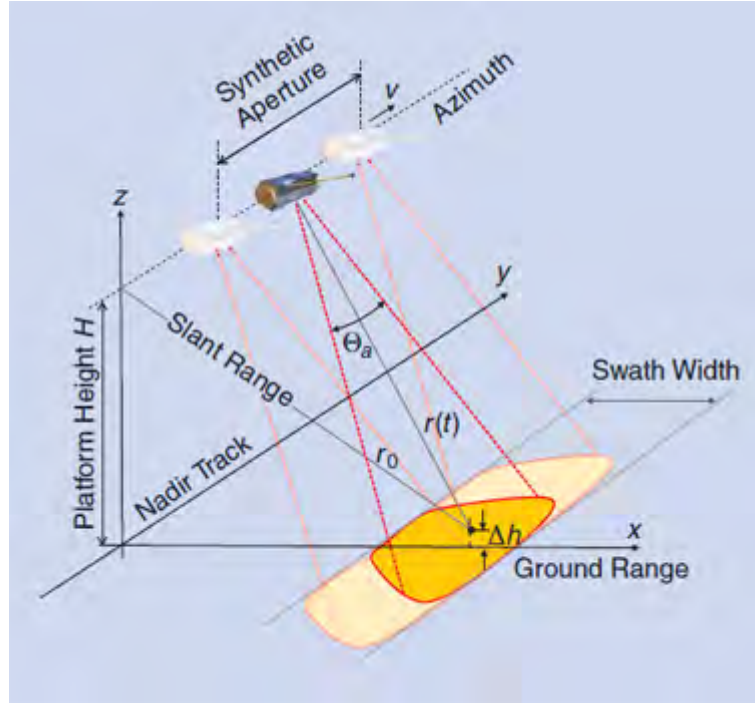
Figure 1.1. SAR imaging geometry.

$$P_E = \frac{P_R \cdot A_A \cdot K_a}{4 \cdot \pi \cdot r^2 \cdot L} = \frac{P_T \cdot G^2 \cdot \lambda^2 \cdot \sigma}{(4 \cdot \pi)^3 \cdot r^4 \cdot L} \qquad (1.6)$$

with L representing a measure for the system losses. The 1.6, which expresses the mutual relationship between transmitted power (P_T), power backscattered by the target (P_R), and power received by the antenna (P_E), is known as *radar equation*[2].

Both radar and synthetic aperture radar systems have a side looking imaging geometry. The simplest radar instrument provides a 2-D reflectivity map of the imaged area: high backscattered signal are identified as bright spots in the radar images, while flat smooth surfaces as dark areas. The flight direction is known as *azimuth* and the line-of-sight as *slant range* direction. Until the 50s imaging radars, denoted as SLAR (side-looking airborne radar), could provide from low up to moderate azimuth resolution, due to their specific operation mode. The azimuth resolution, given by:

$$\delta_a = \frac{\lambda}{d_a} \cdot r \qquad (1.7)$$

is the minimum distance between two points targets that can be detected by the radar. It depends on the signal wavelength λ, on the range distance r from the antenna to the targets, and it is inversely proportional to the dimension of the antenna d_a. It means that, in order to achieve high azimuth resolution, large antennas have to be used for satellite systems, where the distance to the Earth surface is very large. Such a characteristic of the SLAR systems has been its main drawback, which led to the development of modern techniques on which the Synthetic Aperture Radar (SAR) is based.

1.1 Synthetic Aperture Radar (SAR)

A Synthetic Aperture Radar (SAR) is an imaging radar mounted on a moving platform. Similar to a conventional radar, electromagnetic waves are sequentially transmitted and the

backscattered echoes are recorded by the radar antenna. Basically, by taking advantage of the Doppler history of the radar echoes generated by the forward motion of the spacecraft, the SAR synthesizes a large antenna, which allows high azimuth resolution in the resulting image.

Fig.1.1 depicts the typical SAR geometry: the platform moves along the *azimuth* (or *along track*)direction; the *slant range* is the direction perpendicular to the flight path of the radar; the *swath width* gives the ground-range extent of the imaged scene, while the length depends on the time interval in which the radar is turned on.

Given two targets in the same azimuth resolution cell, they can be detected only if their distance in range direction is larger than the spatial extent of a single electromagnetic pulse, preventing the overlapping of the echoes from the objects. Basically, the slant range resolution is limited by the bandwidth (B_r) of the transmitted signal :

$$\delta_r = \frac{c_0}{2 \cdot B_r} \tag{1.8}$$

with c_0 denoting the speed of light. A wide bandwidth can be achieved by a short duration pulse. However, the shorter the pulse, the lower the transmitted energy and the poorer the radiometric resolution. A proper signal processing is typically applied in order to preserve the radiometric resolution and to optimize the range resolution.

On the other hand, the azimuth resolution δ_a is provided by the construction of the synthetic aperture, which is the path length during which the radar receives echo signals from a point target. The beamwidth of an antenna with dimension d_a can be approximated by:

$$\Theta_a = \frac{\lambda}{d_a} \tag{1.9}$$

The corresponding synthetic aperture length is (Fig.1.1):

$$L_{sa} = \Theta_a \cdot r_0 = \lambda \cdot \frac{r_0}{d_a} \tag{1.10}$$

The virtual beamwidth can be expressed as:

$$\Theta_{sa} = \frac{\lambda}{2 \cdot L_{sa}} \tag{1.11}$$

Hence the azimuth resolution results:

$$\delta_a = r_0 \cdot \Theta_{sa} = r_0 \frac{\lambda}{2 \cdot L_{sa}} = \frac{d_a}{2} \tag{1.12}$$

The above equation suggests that a short antenna yields a fine azimuth resolution.

The received echo signal data form a bi-dimensional matrix of complex samples, expressed through a real and an imaginary part, hence representing an amplitude and a phase value.

Differently from optical images, raw SAR data do not give any useful information on the scene. The SAR images is obtained only after signal processing (*focusing*), which in principle consists of two separate matched filter operations along the range and azimuth dimensions. The transmitted chirp signals are firstly compressed to a short pulse. Each range line is multiplied in the frequency domain by the complex conjugate of the spectrum of the transmitted chirp. The resulting range compressed image gives information about the relative distance between the antenna and the targets. In the same way, the azimuth compression is carried out by a convolution of the signal with its reference function, that is the complex conjugate of the response expected from a point on the ground. While the range reference function is dependent only on the transmitted chirp waveform, the azimuth one depends on the geometry and is adapted to the range. After the azimuth compression, the energy backscattered by a single resolution cell on the ground is compressed in one pixel.

During focusing, the aforementioned steps are executed on the whole image, to obtain a complex image (Single Look Complex) where amplitude is related to the radar reflectivity, and phase to the acquisition geometry and on the ground topography.

However, SAR images are most typically displayed in terms of intensity values, where each pixel provides information about the reflectivity of a target on the Earth surface. In order to generate such kind of product, two further processing steps are necessary: *calibration* and *geocoding*. These issues will be addressed in the following §1.3.

1.2   Characteristics of SAR images

Scattering mechanisms

SAR images appear in gray scale tones, where dark and bright pixels represent low and high backscatter, respectively. In the latter case, a large percentage of energy is detected by the antenna, while dark features mean a very little amount of backscattered energy from the target. At a particular wavelength, the backscatter depends on the geometry of the scatterers within the target area, on their electrical properties and moisture content. However, the backscattering behavior of an object varies also with the wavelength and polarization of the SAR signal, and with the observation angle of the system.

The scattering of microwaves from a surface (*surface scattering mechanism*) is composed by a mix of specular and Lambertian scattering, depending on the roughness σ_s with respect to the wavelength [3]. Theoretically, an infinite large and perfectly smooth surface, with $\sigma_s << \lambda$, reflects all the energy in the specular direction. As a result, a flat surface is represented by a dark area in the SAR image. As much as the roughness increases with respect to the wavelength, as weaker are the specular components of the scattered energy, and stronger the Lambertian ones. A rough surface is hence represented by a bright area in the SAR image.

The backscattering is also a function of the relative dielectric constant ϵ_r characterizing the target material. The lower is ϵ_r, the lower is the reflectivity property of the material, and thus the higher is the penetration of the electromagnetic signal into the medium. Moreover, depending on the frequency and polarization, waves can penetrate into the vegetation and, on dry conditions, to some extent, into the soil (for instance dry snow or sand). Generally, the longer the wavelength, the stronger is the penetration into the target. With respect to the polarization, cross-polarized (VH/HV) acquisitions have a significant less penetration effect than co-polarized (HH/VV) ones.

In principle, all the real media (e.g. soil, snow, vegetation) are made of multiple particles which contribute to the backscattering through a so called *volumetric scattering* mechanism.

Surfaces inclined towards the antenna will have a stronger backscatter than surfaces which slope away from the radar, and will tend to appear brighter in a radar image. A particular scattering mechanism occurs when two surfaces (e.g. streets and buildings in an urban environment, or soil and trunks in a vegetated area) are lined up in such a way that the incoming radar pulses are able to bounce off the first surface (e.g. street) and then bounce again off the second one (e.g. building) and directly back towards the antenna. Such mechanism, called *double bounce*, generates very bright features in the SAR images.
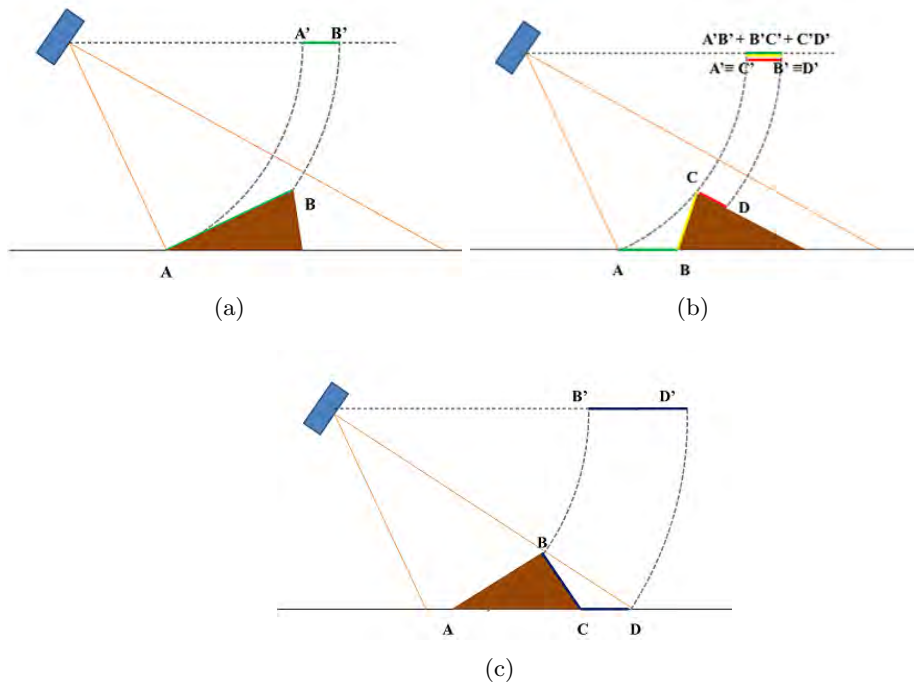
Geometric distortions

Since SAR measures the projection of a 3-D scene on the slant-range and azimuth coordinates, the resulting images are typically affected by geometrical distortions. Examples of such effects are shown in Fig.1.2 e described below.

Foreshortening. It is a dominant effect in SAR images of mountainous areas. Especially in the case of steep-looking spaceborne sensors, the projection along the slant-range direction of two points A and B located on the slope oriented toward the antenna, is shorter than they would be in flat areas (Fig.1.2(a)). This effect results in a slant-range compression of the radiometric information backscattered from foreslope areas which results brighter in the SAR image. Foreshortening effects may be compensated during the geocoding process (see §1.3) if a terrain model is available.

Layover. If a slope is steeper than the radar beam, targets in the valley have a larger slant range than related mountain tops, then the foreslope results reversed in the slant range image. It means that the ordering of surface elements on the radar image is the reverse of the ordering on the ground. As an example, in Fig.1.2(b) the signal propagating from the top of the mountain (point C) towards the antenna, is detected before the scattering from the bottom (point B). Generally, these layover zones, facing radar illumination, appear as bright features on the image, including the backscattering from targets hit in the valley, in the foreland of the mountain and in part of the backslope, as shown in Fig.1.2(b). Geocoding (see §1.3) can not resolve the ambiguities due to the representation of several points on the ground by one single point on the image.

Shadow. When surfaces are turned away from the sensor with an angle that is steeper than the SAR illumination, they result occluded from the radar beam and no backscattering is recorded at the sensor (Fig.1.2(c)). SAR shadows appear as dark areas, where fluctuations from the zero signal are mainly due to system noise, sidelobes, and other effects normally of small importance. It has to be observed that the shadows of two objects of the same height are longer in the far range than in the near range.


Figure 1.2. Geometric distortions in SAR imagery: a) Foreshortening; b) Layover; c) Shadow.

Speckle noise

The presence of multiple scatterers randomly distributed within a resolution cell causes the typical *speckle* effect, which gives a granular appearance to the SAR images. It is due to the constructive and destructive interference of the coherent SAR pulse by each elemental target. Such effect is particularly evident in those areas where the SAR wavelength is comparable to the surface roughness. The coherent sum of amplitudes and phases of the scattered waves results in strong fluctuations of the backscattering from resolution cell to resolution cell; therefore, the intensity and phase in the image are no longer deterministic, but follow an exponential and uniform distribution, respectively [2]. The total complex value measured at the sensor is given by:

$$\Phi = \sum_i \sqrt{\sigma_i} \cdot exp(i\phi_i^{scatt})) \cdot exp\left(-i\frac{4\pi}{\lambda}r_{0,i}\right) \tag{1.13}$$

where i is the number of scatterer belonging to the same resolution cell. In Fig.1.3, two examples of constructive and destructive interference are shown. While the former is associated with brighter pixels, the latter occurs in darker pixels. Although it is common to refer to the speckle as to a noise effect, indeed it is a physical measurement of the resolution cell structure at sub-resolution level. Moreover, having a multiplicative behavior, it is not possible to reduce it by increasing the transmit signal power. In fact, its variance increases with its intensity. A considerably amount of studies have been carried out with the purpose of characterize and reduce the speckle phenomenon [4], [5]. Some notions about despeckling processes will be given in §1.3.
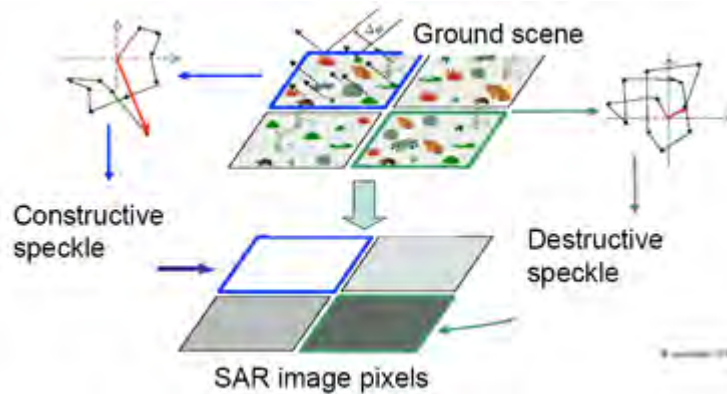
Figure 1.3. Examples of constructive and destructive speckle.

1.3   SAR image pre-processing

In order to make SAR imagery suitable for any remote sensing application, the focusing phase must be followed by a series of pre-processing operations.

Through the *calibration* the image pixels values are transformed into the radar cross section normalized to area σ_0. Actually, this processing step involves both internal instrument calibration as well as external SAR calibration by using targets of known reflectivity [6]. Though uncalibrated SAR imagery is sufficient for qualitative use, calibrated SAR images are essential to quantitative use of SAR data. Moreover, the radiometric correction is necessary for the comparison of SAR images acquired under different conditions (e.g. by using different sensors, or at different times, in different modes, or processed by different processors).

A further SAR image pre-processing consists in the *geocoding*, also known as georeferencing, geometric calibration, and ortho-rectification. It ensures that the location of any pixel in the SAR image (either slant range or ground range geometry) is directly associated to a map coordinate system. The process can be performed with or without the use of a Digital Elevation Model (DEM), carrying out an *ellipsoidal geocoding* or a *terrain geocoding*, respectively. In order to properly geocode SAR images, a range-Doppler approach must be adopted. In fact, SAR systems cause non-linear compressions (in particular in presence of topography), and thus they can not be corrected using polynomials as in the case of optical images, where (in the case of flat Earth) an affine transformation is sufficient to convert it into a cartographic reference system. Through the range-Doppler correction the relationship between the sensor, each single backscatter element and their related velocities is calculated, and therefore not only the illuminating geometry but also the processors characteristics are considered. This complete reconstruction of the imaging and processing geometry also takes into account the topographic effects (foreshortening, layover) as well as the influence of Earth rotation and terrain height on the Doppler frequency shift and azimuth geometry.

As previously asserted, speckle hampers the visual interpretation and image analysis of SAR acquisitions, even though, being related to the backscattering mechanism of the illuminated surface, it indirectly contains information about the type of land cover [7], [8]. According with the specific task to accomplish, different *despeckling* approaches can be adopted. Generally speaking there are two techniques of mitigating speckle noise: multi-look process and spatial filtering. The first is used at the data acquisition stage, while the second only once the data are stored. Any despeckle processing should meet some conditions, such as the preservation of radiometric information, objects edges and spatial resolution. Multi-look SAR processing is performed averaging neighboring pixels using a fixed window with the drawback of worsening the spatial resolution [5]. Alternatively, one can average different acquisitions taken at different times. In this case it will be the time resolution to be lowered. On the other side, spatial adaptive filters, such as the well known Frost, Lee, Kuan, Gamma-MAP, allow at the same time speckle reduction and preservation of the geometry and texture of the objects within the scene [9].

1.4 Very High Resolution SAR sensors

With the launch of the bi-static SAR satellite TerraSAR-X and TanDEM-X, the COSMO-SkyMed constellation and Radarsat-2 a new generation of SAR satellites was introduced. Beside being able to acquire data in all the illumination and weather conditions, the main peculiarity of such systems is their metrical spatial resolution.

TerraSAR-X and TanDEM-X are two missions developed by the German Aerospace Center (DLR). The former, launched in 2007, is equipped with a X-band (wavelength 31 mm, frequency 9.6 GHz) SAR, acquiring data with single or dual polarization in four acquisition modes: High-resolution Spotlight, Spotlight, StripMap and ScanSAR.
TanDEM-X (TerraSAR-X add-on for Digital Elevation Measurements) is a second, very similar spacecraft launched successfully in 2010. Since October 2010, TerraSAR-X and TanDEM-X fly in a close formation at distances of only a few hundred meters and record data synchronously. This unique twin satellite constellation will allow the generation of global digital elevation models (DEMs) of high accuracy, coverage and quality.

The COSMO-SkyMed mission was born from the cooperation of the Italian Space Agency (ASI) with the Italian Defence Ministry (MD), to provide an innovative constellation of satellites for the Earth's observation, conceived for a civilian and military use. The mission, started in 2007 with the launch of COSMO-SkyMed-1, was completed in 2010 with the launch of the fourth satellite. Similarly to TerraSAR-X, the constellation is equipped

with X-band sensors, operating in four different mode and different polarizations. The Spotlight mode, in particular, allow the acquisition of 1 meter spatial resolution imagery. Moreover, thanks to the configuration of the system, COSMO-SkyMed provides data of the same scene within a very short revisit time. More accurate details about the COSMO-SkyMed mission and the future second generation of COSMO-SkyMed will be given in Chapter 2

Differently from the aforementioned VHR SAR satellites, the Canadian Radarsat-2 acquires metrical resolution images by operating at C-band microwave frequencies. It supports multiple polarization modes, including the fully polarimetric. It was successfully launched in 2007 by the Canadian Space Agency as a follow-on to Radarsat-1. As an evolution of the Radarsat program, a new mission is planned. It foresees a constellation of three C-band SAR satellites (Radarsat Constellation) which will provide complete coverage of Canada's land and oceans, offering an average daily revisit, as well as daily access to 95% of the world. The mission development has begun in 2005, with satellite launches planned for 2018.

Nowadays, remote sensing community is still exploring the unique potentiality of VHR SAR data, whose content of information is expected to be extremely useful to accomplish several tasks. VHR SAR images have being successfully applied for civilian and security applications, damage assessment, land use mapping, sea monitoring and so on. However, the very high spatial resolution, allowing the detection of many small details in the illuminated scene, makes at the same time challenging the image interpretation of complex environments, especially when X-band microwave are used. In fact, the shorter wavelengths can augment the level of interference of the waves reflected from the many elementary scatterers, with the result of a possible increase of the speckle effect.

Chapter 2

COSMO-SkyMed Mission

Since 2007 a significant step forward for the remote sensing community has been accomplished thanks to the new COSMO-SkyMed (COnstellation of small Satellites for Mediterranean basin Observation) mission [10–12]. The project originated from the cooperation of the Italian Space Agency (ASI) and the Italian Defence Ministry (MD), to provide an innovative constellation of satellites for the Earth's observation, conceived for a civilian and military use. The mission consists of a constellation of four Low Earth Orbit (LEO) satellites, furnished with X-band (9.6 GHz frequency, 3.1 cm wavelength) Synthetic Aperture Radar (SAR), allowing the whole global coverage in every weather and illumination condition. Moreover, the possibility of acquiring images with a very high spatial resolution, as well as the system capability to pass over the same area of interest in few hours, make COSMO-SkyMed suitable for several applications such as risk management, natural disaster assessment, cartography updating, vegetation monitoring, sea ice monitoring etc.

Figure 2.1. COSMO-SkyMed constellation.

2.1 The mission

The COSMO-SkyMed mission, started with the launch of the first satellite on 8th June 2007 and completed with the fourth satellite on 6th November 2010, has been conceived to meet very high level requirement trough:

- Numerous daily acquisitions;

- Global satellite coverage;

- Capabilities of acquisition in all weather and illumination conditions;

- Short System Response Time (time interval between the finalization of the user request for the acquisition of the area of interest and the delivering of the product);

- Very High spatial and radiometric Resolution, which imply very fine image quality;

- Possibility of image spatial resolution trade-off with size, at most possible extent and including sub-meter resolution;

- Capability to cooperate with other EO missions and to expand towards other possible partners with different sensors aiming at the implementation of an integrated system providing Earth Observation integrated services to large user communities on a worldwide scale (IEM concept: Interoperability, Expandability and Multi-sensoriality).

Table 2.1. COSMO-SkyMed launch schedule

Mission	Launch date
COSMO-SkyMed 1	June 8, 2007
COSMO-SkyMed 2	December 9, 2007
COSMO-SkyMed 3	October 25, 2008
COSMO-SkyMed 4	November 6, 2010

The fully deployed constellation is able to collect up to 1800 images per day and, at the same time, the Ground Segment capacity can be scaled in order to archive and process a large amount of data, according with the Satellite Segment capacity.

2.2   Products

According with the international standards for Earth Observation, the COSMO-SkyMed Ground Segment can provide products through different standard processing levels, for several remote sensing applications, based on the direct exploitation of low level products. The standard processing can be divided into three main steps:

1. Pre-processing;

2. Processing;

3. Image geo-localization.

In the first stage (Level 0, *pre-processing*) all the necessary preliminary operations for data analysis and extraction of information are carried out. *Processing* stage involves radiometric and geometric corrections of the images (Level 1A and 1B processing). As last elaboration of the processing chain, the *image geo-localization* consists in projecting the images on known reference system (Level 1C and 1D processing). More specifically, four **Standard Levels Products** are available:

Level 0 (RAW): Raw data consists in unpacked echo data in complex in-phase and quadrature signal (I and Q) format, to which an internal calibration and an error compensation are also applied. The product includes the required auxiliary data to produce a higher level image.

Level 1A (SCS): Single look Complex Slant product is achieved by the focusing in slant range-azimuth projection of the RAW data.

Level 1B (MDG): Detected Ground Multi-look product is obtained through the detection, multi-looking, projection in zero-doppler ground range-azimuth onto a reference ellipsoid or on a DEM, and resampling at a regular grid on ground with associated ancillary data.

Level 1C (GEC): Geocoded Ellipsoid Corrected is obtained by projecting the SCS data onto the Earth ellipsoid in a regular grid obtained from a cartographic reference system chosen among a predefined set with associated ancillary data.

Level 1C (GTC): Geocoded Terrain Corrected is obtained by projecting the SCS data onto a Digital Elevation Model (DEM) in a regular grid obtained from a cartographic reference system chosen among a predefined set with associated ancillary data.
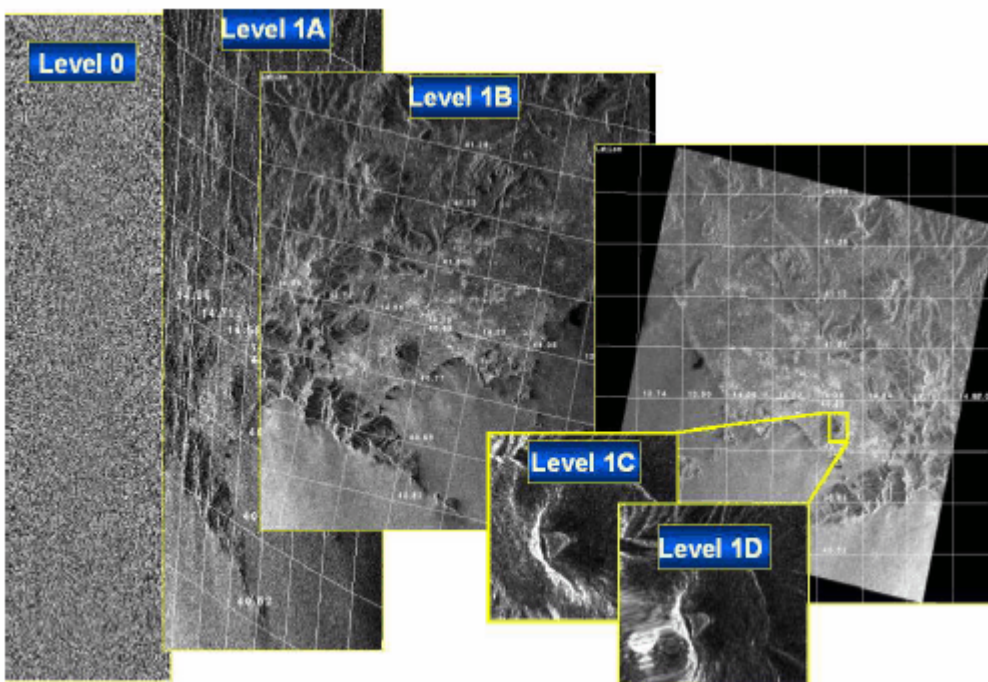
Figure 2.2. COSMO-SkyMed standard level products.

Higher Level Products are also available, for medium and high level remote sensing applications. For civilian applications high level products include:

Quicklook: image with a degraded spatial resolution for browsing aims, and integer pixels values scaled in the range 0÷255. These kind of products allow a fast and easy visualization of the images, which are ground-projected and geo-located.

Co-registered products: a set of images (a *master* used as reference, and one or more *slaves*) derived from the 1A or 1B processing level, are co-registered for remote sensing applications such as interferometry or change detection. Generally, due to the different geometry of the imagery, the slave image needs to be re-sampled into the master geometry to be superposed. The type of data is kept, hence real or complex images produce real or complex co-registered set of images. The co-registration process is bound by some conditions. For instance, images shall be taken in the same acquisition mode, look side and orbit

direction.

Mosaicked products: image obtained by merging a set of sub-images (Level 1B, geocoded or DEMs) in horizontal direction for large spatial coverage. Sometimes an overlap between contiguous frames or a gap occur, making not perfect the matching. Hence a proper feathering on the overlapping zones has to be carried out, as well as the exclusion or the filling of the gaps. Aiming at the production of high resolution mosaicked products, no undersampling process is applied to the input data.

Speckle filtered image: although standard level products provide speckle reduction by a multi-looking approach, it may be not suitable for high-level applications (classification, feature extraction, change detection, etc). Speckle filtered higher level data are derived from post-processing of 1A or 1B standard level products, to which a filter is applied, with a consequent increasing equivalent number of looks.

Interferometric products: derived by processing Level 1A co-registered images to generate interferometric coherence map and wrapped interferometric phase. Such products preserve the geometry of the master and slave input data, so they are in a slant range/azimuth projection. However, due to the interferometric multilooking, pixel spacing and size are reduced. Moreover, to generate interferometric products, input images must be co-registrable (acquired on the same subswaths, with the same look side, orbit direction and polarization).

DEM: it consists of the ellipsoidal height map and the corresponding height error map. DEM products are obtained by interferometric processing of the SAR Level 1A co-registered images, hence the same constraints for interferometric products generation also exist for the DEM generation.

Orbital and **Quality Control** products belong to the so-called ***Service Products***, which are only for internal uses. They are exploited as auxiliary data for production chain, or as tools for further analysis performed on all the described products. In particular, *Orbital data* are necessary to perform SAR images geo-location, whereas *Quality Control* products are used to assess the quality of SAR imagery derived form standard and non-standard processors.

2.3   Acquisition modes

The COSMO-SkyMed SAR sensors can operate in different modes, providing images of more or less wide areas, and characterized by medium or very high spatial resolution. Specifically the system can work in three main modes of acquisition:

1. **Spotlight mode (S2)** for very high spatial resolution over small regions. The acquisition is made by means of the antenna steering in the azimuth and the elevation plane, in order to illuminate the area of interest for a time period longer than the standard strip side view, enhancing the azimuth resolution through the increasing of the length of the synthetic antenna. Two different implementations of the Spotlight acquisition are foreseen:

   - **SMART (S1):** it is only for defense applications.
   - **Enhanced Spotlight (S2):** it allows the acquisition of images over an area of about 10x10 km^2, with a spatial resolution of 1x1 m^2 single look. The incidence

angle can be varied from 20° to 60°, while the signal polarization is selectable among HH or VV.

2. **Stripmap mode** for very high and high spatial resolution over areas of tenth of kilometers. This common acquisition mode is performed by pointing the antenna along a fixed direction ($\sim 20° \div 60°$ incidence angle), so that while the platform moves, the SAR acquires a strip on the illuminated area. Theoretically, the length of the strip along the azimuth track is unlimited. However, the SAR duty cycle of about 600 s, limits the strip length to about 4500 km. Two implementations are allowed for this mode of acquisition:

   - **Himage (HI):** this type of product is characterized by about 40 km either for the swath width and the azimuth extension (square frame). Stripmap HI images reach very high spatial resolutions (i.e. 3x3 m^2 single look), and can be acquired in HH, HV, VH or VV polarization.
   - **PingPong (PP):** is obtained by alternating the signal polarization between two possible ones: VV, HH, HV and VH. Operating in such way, the synthetic antenna length is not completely available in azimuth. For this reason the spatial resolution is reduced to about 15x15 m^2 single look). The swath width and the azimuth extension are both of about 30 km (square frame).

3. **ScanSAR mode** for images covering large swaths obtained by periodically stepping the antenna beam to neighboring sub-swaths, considering an incidence angle of $\sim 20° \div 60°$. The availability in azimuth of only a part of the synthetic antenna length, leads to the reduction of the azimuth resolution. Since the acquisition is performed in adjacent strip mode, it is virtually unlimited along the azimuth direction, but it is actually restricted because of the SAR instruments duty cycle of about 600 s. ScanSAR mode of acquisition can be implemented into two different way:

   - **WideRegion (WR):** acquisitions over three adjacent subswaths are grouped, allowing a ground coverage of about 100 km in the range direction, equal to the azimuth extension (100x100 km^2 coverage). The image spatial resolution is of about 30x30 m^2.
   - **HugeRegion (HR):** Six adjacent subswaths are grouped, allowing a ground coverage of about 200x200 km^2, with a spatial resolution of about 100x100 m^2.

Table 2.2. Characteristics of thee main COSMO-SkyMed acquisition modes.

	Spotlight	Stripmap		ScanSAR	
	S2	HI	PP	WR	HR
Polarization	Single	Single	Dual	Single	Single
Swath width (kmxkm)	10x10	40x40	30x30	100x100	200x200
Accessible swath	$\sim$ 620 km				
Geometric Resolution(m)	1	3	15	30	100

2.4 Orbital configuration

The nominal configuration of COSMO-SkyMed constellation is depicted in Fig.2.4, where four equi-phased satellites orbit in the same plane, guaranteeing the global Earth access
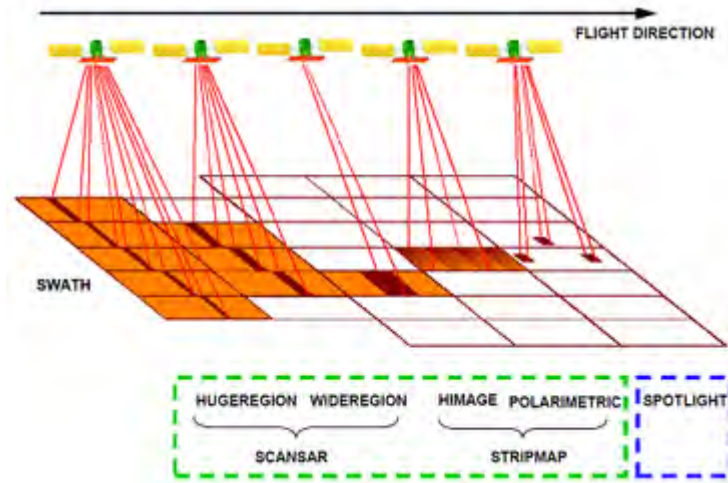
Figure 2.3. Three main modes of acquisition of COSMO-SkyMed satellites.
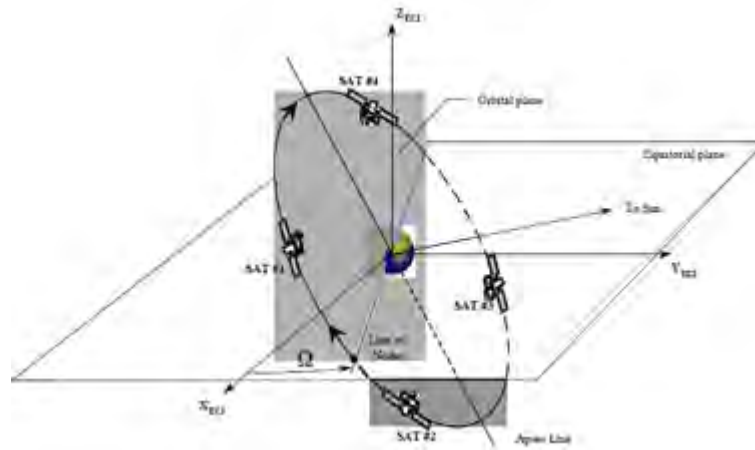
Figure 2.4. Nominal configuration of COSMO-SkyMed constellation.

in few hours, and at least two acquisitions of the same area of interest in one day, under different angles of incidence. In order to assure the "ground track repeatability", active spacecraft maneuvers for constellation geometry maintenance (compensation for the effects of in-plane and out-of-plane perturbations on the nominal orbit) are necessary to keep a given accuracy (tolerances of better than $\pm$ 1 km) with respect to the nominal ground track.

Specific remote sensing applications, such as near real time analysis of the dramatic changes due to environmental disasters (earthquakes, floods), are made possible by the interferometric configuration, which requires, besides the ground track control (analogous to the nominal configuration), also the maintenance of the interferometric baseline within an accuracy of tens of meters. This is made possible by the **nominal tandem** configuration, or by the **tandem-like** configuration, which allows two observations within a 24 hour delay. The tandem configuration is represented in Fig.2.5, where two satellites fly on different but close orbits, whose planes are 0.08° separated, corresponding to 151 km along-track. In such a configuration SARs can acquire interferometric pairs of images within a time interval of 20" from each other. The interval of the normal baseline can vary in a range of 100÷3500 m, with 20% of accuracy, according to the acquisition mode. In the tandem-like configuration the couple of satellites flies on the same orbital plane, and at a short distance each other (67.5°), in order to achieve the required baseline.
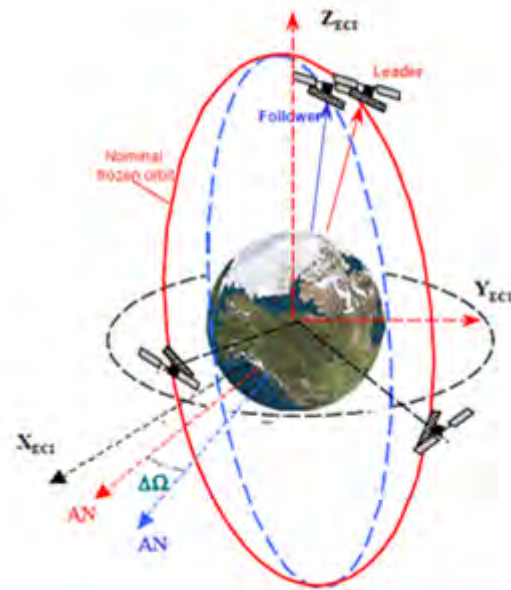
Figure 2.5. Tandem interferometric configuration of COSMO-SkyMed constellation.
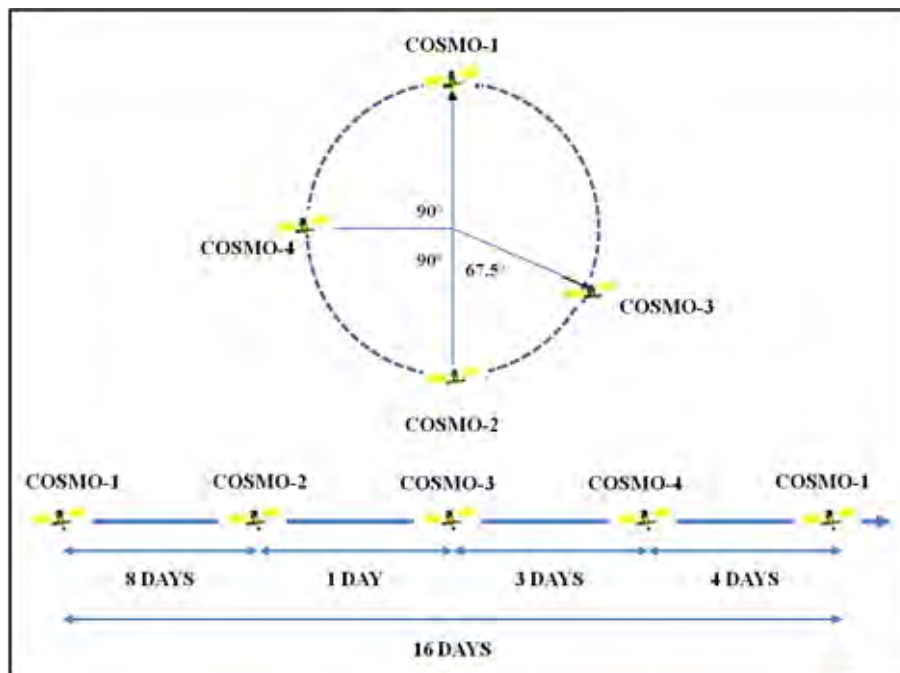
Figure 2.6. Current orbital configuration of COSMO-SkyMed constellation.

Currently, in order to optimize the time performance of the system, COSMO-1, COSMO-2 and COSMO-4 satellites are placed with a displacement of 90° while COSMO-2 and COSMO-3 are positioned in the tandem-like configuration (Fig.2.6). Such a scheme has been designed with the aim of satisfying the acquisitions requests under various conditions of emergency. Indeed, three operative modes are foreseen for the constellation: *routine*, *crisis* and *very urgent*. Consequently, different levels of mission planning are conceived: *Long Term Plan* (LTP), *Medium Term Planning* (MTP) and *Short Term Planning* (STP). The LTP and the corresponding coarse schedule cover a mission time period multiple of 16 days repeat cycle of a single satellite, for regional monitoring or routine (low priority) Programming Requests. The MTP is associated to a coarse schedule, covering a time span
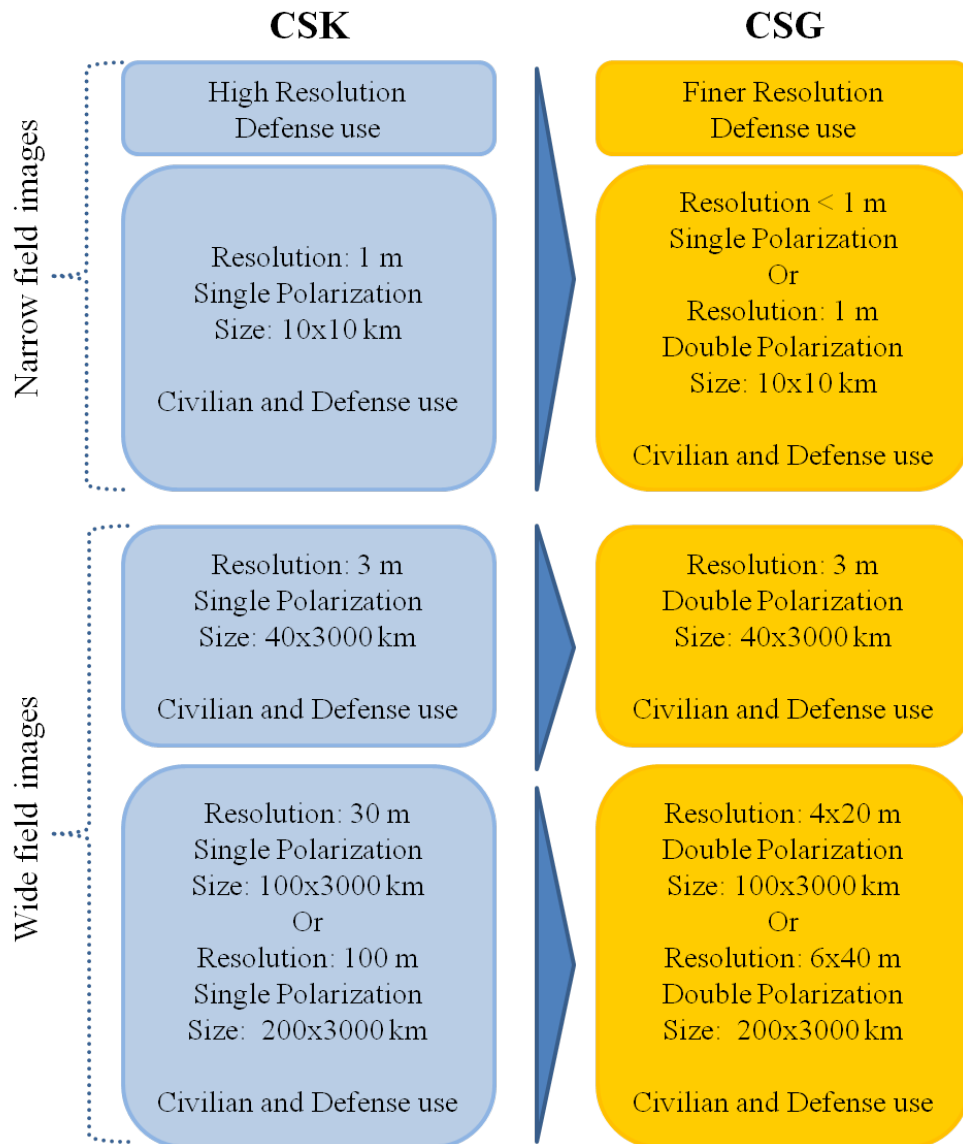
which varies, according with the number of active satellites, from four days (4-satellites constellation), to eight days (2-satellites constellation) or sixteen days (1 or 3-satellites constellation). Finally, STP refers to fine scheduling, for acquisitions within a time interval of 24 hours.

2.5   COSMO-SkyMed Second Generation

COSMO-SkyMed was designed foreseeing a nominal lifetime of five years (seven years envisaged "extended" lifetime) for each satellite of the constellation. Since the mission was completed in 2010, the space segment has been fully operational for three years (2013). The processing and delivering of SAR images will be assured even after the satellites disposal, thanks to the longer lifetime of the ground segment (fifteen years foreseen).

In order to guarantee a continuity with COSMO-SkyMed, and the enhancement of the SAR services, the COSMO-SkyMed Second Generation (CSG) mission has been planned. It will be ready for operations timely to replace the previous generation satellites whenever they are being progressively phased out at the end of their lifetime, starting from 2015 onward. With the deployment of two new satellites, both equipped with more performing SAR instruments, and the evolution of the existing ground segment, COSMO-SkyMed Second Generation will meet the requirements of quick response time, security rules, data confidentiality and type, enhancement of quality and number of images per orbit and per day [13]. CSG, besides keeping and improving the operative modes of its precursor sensors, will also support new sensor modes and functionality, not provided by the current COSMO-SkyMed. As for the first generation, CSG will provide SAR images with different size and resolution (Fig.2.7), which will be improved by a factor of three in narrow field imagery. One of the most important novelty of CSG will be the capability of SAR sensors of taking images at double polarization in each acquisition mode.

The Ground Segment will play a key role to allow a smooth transition from the first to the second generation, supporting CSG system deployment and qualification, maintaining at the same time the CSK functionality at operations in both Defence and Civilian different domains. The combination of CSG Ground Segment and Space Segment increases the system performances in terms of number of products, double data volume of a single product, and global time needed for data processing which will be reduced of 30% - 50% with respect to the first generation system. Thus, the Ground Segment will require significant upgrades in order to better manage, control and exploit its enhanced capabilities, as well as take advantage of the improved overall performance of the CSK-CSG system.

CSK

CSG

Narrow field images

CSK
High Resolution
Defense use

Resolution: 1 m
Single Polarization
Size: 10x10 km

Civilian and Defense use

CSG
Finer Resolution
Defense use

Resolution < 1 m
Single Polarization
Or
Resolution: 1 m
Double Polarization
Size: 10x10 km

Civilian and Defense use

Wide field images

Resolution: 3 m
Single Polarization
Size: 40x3000 km

Civilian and Defense use

Resolution: 3 m
Double Polarization
Size: 40x3000 km

Civilian and Defense use

Resolution: 30 m
Single Polarization
Size: 100x3000 km
Or
Resolution: 100 m
Single Polarization
Size: 200x3000 km

Civilian and Defense use

Resolution: 4x20 m
Double Polarization
Size: 100x3000 km
Or
Resolution: 6x40 m
Double Polarization
Size: 200x3000 km

Civilian and Defense use

Figure 2.7. Comparison between CSK and CSG SAR imaging typologies.

Part II

Algorithms and Methodologies

Introduction

The availability of an increasing amount of VHR SAR data calls for the developing of effective, fast and as much automatic as possible techniques for the understanding of the information embedded within the satellite images. Therefore, this work focused on the implementation of algorithms for the features extraction in sub-urban environments imaged by COSMO-SkyMed X-band sensors. Monitor a sub-urban scenario, which is likely to be subject to a progressive increase of the settlements density, basically means to know the land uses and their changes in time. When metrical spatial resolution data are exploited, the complexity of the imagery, which are characterized by a larger number of details (lamps along the streets, cars, other small manmade structures), together with the typical speckle noise, geometrical distortions, multiple backscattering contributions, which commonly affects SAR images, makes harder the interpretation of the scenario. Moreover, as it will be proved afterward in the Part IV, the exploitation of the only single polarized backscattering intensity may be not efficient to discriminate different land covers. Hence, further sources of information, such as the texture parameters stemmed from the computation of the Gray Level Co-Occurence Matrix (GLCM) [14], should be investigated and possibly productively exploited by the classification algorithms.

The study of COSMO-SkyMed data has been carried out by applying two different approaches: pixel-based and object-based (Fig. I).
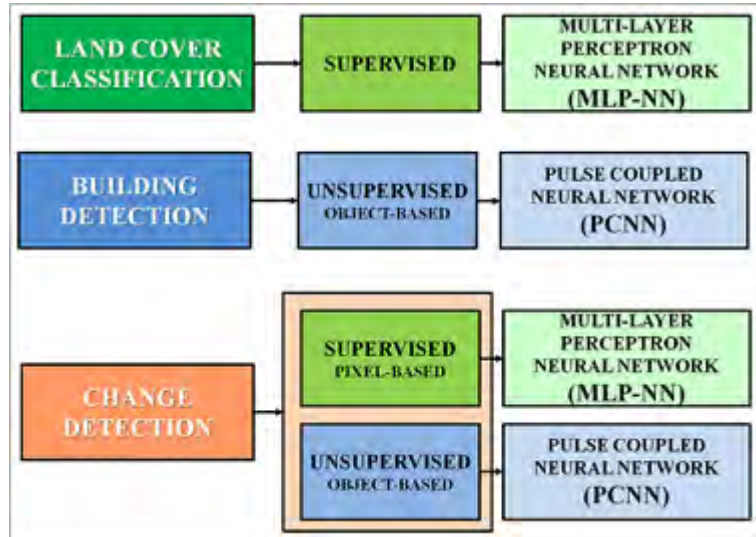
Figure I. Different approaches and algorithms adopted for the sub-urban environment monitoring.

The pixel-based methods have been performed by the supervised Multi Layer Perceptron Neural Network (MLP-NN) algorithm, aiming at the production of thematic maps. In the following Chapter 3 this classical architecture of artificial neural network will be described, while the sub-urban classification issue will be addressed in the Part IV. Here

it will be also shown how the use of textural parameters can enhance the classification maps in terms of accuracy and number of classes. An overview about the adopted Gray Level Co-Occurrence Matrix (GLCM) and the most effective textural features which can be stemmed from them, will be presented in Chapter 5.

Given the particular interest to the urbanization process, an object-based investigation has been accomplished for the buildings extraction in the satellite images, by adopting an unsupervised procedure called Pulse Coupled Neural Network (PCNN), which will be discussed in Chapter 4.

Finally, the pixel-based and the object-based approaches have been jointly applied to implement an automatic change detection algorithm (Chapter 13).

Chapter 3

Multi Layer Perceptron Neural Networks (MLP-NNs)

Since their invention, artificial neural networks (ANNs) have been and are still of great interest to the scientific and industrial community, finding wide use in various fields, such as remote sensing, medicine, robotics, economics. Historically, many concepts in neural computing have been inspired by studies of biological networks, which can be modeled as an ensemble of relatively simple processing units interconnected each other [15].

The most appealing characteristic of the neural network systems is their ability in solving even complex problems, by learning the input-output relationship through a number of examples and regardless of the underlying physical mechanisms, whereas the conventional procedures require for a deep knowledge of the phenomenon under study and the development and implementation of a mathematical model to describe it [16]. In particular, the versatility of neural networks proves to be very handy for function approximation, control and signal processing, satellite images classification [17]. The latter has probably been one of the most investigated neural network applications, due to the possibility of exploiting multi-dimensional data, irrespective of the knowledge of their probability distributions, which is instead necessary for the more traditional approaches such as the Bayesian methods. Moreover, many studies confirmed that neural networks, besides representing a new and easy way of machine learning, exhibit interesting properties, such as the capability of capturing subtle dependencies among the data, an inherent fault tolerance due to their parallel and distributed structure, and a capability of positively merging pieces of information stemming from different sources [18].

However, neural networks drawbacks lead sometime researchers to prefer different machine learning systems, such as the Support Vector Machines (SVM). One of the major problems to consider when one handles with the ANNs is the possibility of the common backpropagation learning algorithm of reaching a local minimum. Moreover, the performance of an ANN depends on its architecture and on the method of presenting the data and carrying out the training. An ANN correctly implemented generally presents a better performance than the standard statistical approaches. Nevertheless, a network incorrectly designed or trained gives poorer results than standard methods such as the maximum likelihood method [17]. By addressing such issues, some researchers focus their study on the automatic selection of the parameters characterizing the NN algorithm, or on the improvement of the learning phase [19]. Other authors explore the potential of the ANNs in solving new and more challenging tasks, that can arise from the opportunity of exploiting data provided by the newest Earth Observation missions.

In this context, VHR COSMO-SkyMed SAR images have been classified in this work by using the common feed-forward Multi-Layer Perceptron (MLP) topology [20], which is described in the following sections.

3.1 Neuron structure

The basic element of a NN is the neuron, which can be seen as an information processing unit, generally characterized by several input and one output [20]. The stimulus coming from the closest neighbors, or from an external source, is elaborated by the single neuron in order to provide an output signal, which is propagated to other units. The artificial neuron (Fig. 3.1) reproduces the biological one by modeling the mechanism through three components:

- weight vector;

- network function;

- activation function.
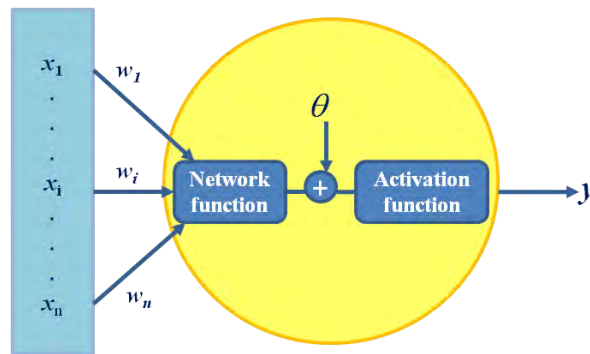
Figure 3.1. Artificial neuron scheme: the input vector is combined with the weight vector by the network function. The activation function is then applied to the resulting signal and the neuron output is provided.

The weight vectors represent the synapses: higher weights correspond to stronger connections between an input and a unit. The neuron activity is described by the network and activation functions. The former combines together the inputs by taking into account the respective weights. A typical network function is the following linear combination:

$$y = \sum_{i=1}^{N} x_i w_i + \theta \qquad (3.1)$$

where y is the network function output, x_i the input vector elements, w_i the weights vector, and θ the bias. The latter has been introduced in order to avoid the activation of the neuron if the cumulative effect of the inputs does not overcome the threshold θ. The activation function controls the amplitude of the output, whose acceptable values generally range between 0 and 1, or -1 and 1. The network function output is modified by the activation function through a linear or non-linear transformation, whose result is the neuron output. Usually, the activation functions are characterized by saturation at a minimum and a maximum value, and by being non-decreasing functions.

3.2 Network topology

The Multi Layer Perceptron Neural Network (MLP-NN) takes its origin from the simpler perceptron model proposed in the early 50's, where a single neuron utilizes a linear network function and a threshold activation neuron function. However, although such a model can solve simple information processing functions, it suffers the linear separability

limitation, which makes it not suitable to more complex problems, such as the pattern recognition. To overcome the perceptron weakness, the MLP-NN was designed to have continuous value inputs and outputs, and nonlinear activation functions. The success of the MLP, which is the most widely used neural network, lies in its high generalization capability, by approximating an unknown input-output relationship, and hence providing a nonlinear mapping between its inputs and outputs. The MLP architecture is made of layers of neurons (Fig.3.2), namely:

- one input layer;

- one or more hidden layers;

- one output layer.

The number of the input and output neurons is directly related to the dimension of the correspondent input space and desired output space. While the first level merely distributes the inputs to the internal stages of the network, in the hidden layers the information is elaborated and propagated to the output layer, which provides the final result.
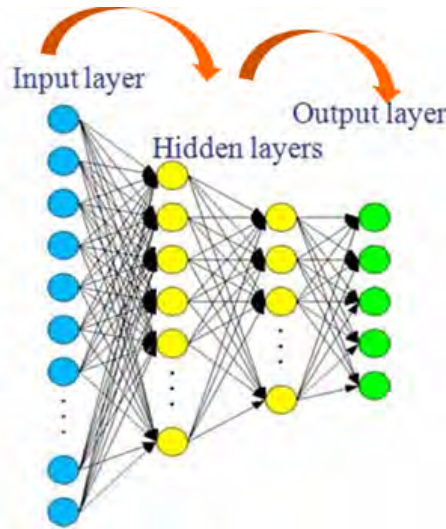
Figure 3.2. Multi Layer Perceptron neural network scheme.

Each neuron of a layer is connected to all the neurons belonging to the following layer, and the signal is propagated in only one direction (*feed-forward* architecture), that is from the input layer to the output layer. Being feed-forward, the MLP is *static*, in fact it provides only one set of output values rather than a sequence of values from a given input. It means that its response to an input does not depend on the previous network status.

The design of the optimal topology of a MLP-NN is generally a difficult task, consisting in the choice of the number of hidden layers and neurons, which affects the performance of the neural network. If the number of units is too small, the input-output associative capabilities of the network are too weak. On the other side, if a large number of neurons is used, a lack of generality can be exhibited by networks that are too much tailored to the training set. Generally speaking, a single hidden layer should be sufficient for most problems, especially for classification tasks; in fact, a MLP with one layer can approximate any continuous function. However, a network with two hidden layers may solve some tasks much more efficiently than a network with only one hidden layer. The majority of the ANNs used in remote sensing are based upon a single hidden layer MLP but some authors have reported the use of network with two hidden layers in land cover classification and retrieval of biophysical parameters [17]. In [21], seven commonly used heuristics to compute

the optimum number of hidden layer nodes have been evaluated. The authors found that the majority of the heuristics produced similar results; therefore, they can be taken into account in order to compute a number that can be used as a starting point for a search towards the optimum number of hidden layer nodes.

Another approach to identify an efficient network architecture is to remove useless inter-neuron connections of a network larger than the size normally considered necessary (network pruning).

3.3 Learning phase

As previously asserted, the success and the interest aroused by the artificial neural networks, lie in the learning capability of such systems. Specifically, the network learns how to approximate an unknown input-output relationship and at the end of the training phase, it is able to generalize the problem, providing a reliable output when it is fed by new input data. Being the MLP-NN a supervised algorithm, the training set includes both the input vectors and the desired outputs, which are used to minimize an error function during the training phase. Usually, the NN performance is evaluated by the *Sum of Squares Error* (SSE):

$$SSE = \sum_{p \in patterns} \sum_{j \in output} (t_{pj} - y_{pj})^2 \tag{3.2}$$

where t_{pj} is the desired output of neuron j for pattern p, and y_{pj} is its actual output. In this work, the cost function has been evaluated as the *Mean Sum Squared Error* (MSE), defined as:

$$MSE = \sum_{p \in patterns} \frac{\sum_{j \in output} (t_{pj} - y_{pj})^2}{N} \tag{3.3}$$

where N is the number of outputs. The minimization of the error is carried out by changing iteratively the weights of the connections between the neurons through the following equation:

$$w(t+1) = w(t) + \Delta w(t) \tag{3.4}$$

where $\Delta w(t)$ is the correction applied to the weights $w(t)$ at the epoch t. There are several algorithms suitable for the minimization of the error, which differ in terms of $\Delta w(t)$. Beside the most known *back-propagation* algorithm [20], the *Scaled Conjugate Gradient* (SCG) ([22]) approach is widely used, as it shows to be effective and faster than the former. Hence it has been chosen for the image classifications carried out in this study. The SCG belongs to the class of conjugate gradient methods (CGMs), which are second order techniques, since they make use of the second derivatives of the error function, whereas the first order ones, like the back-propagation, only use the first derivative. Differently from the back-propagation, which proceeds down the gradient of the cost function, the CGM reaches the minimum by following a direction which is conjugate to those of the previous steps. Therefore, the CGMs generally find a better way to a (local) minimum than a first order technique, but at a higher computation cost. However, the SCG method has the advantage of being faster than the other CGMs, by applying a scaling approach to the step size.

As the network is trained to minimize the error on the training set, a major issue is overlearning or overfitting [17]. Overfitting means that the network adjusts to very specific random features of the training data, that have no causal relation to the target function. Such problem is more likely to occur when complex network architectures are used, or when
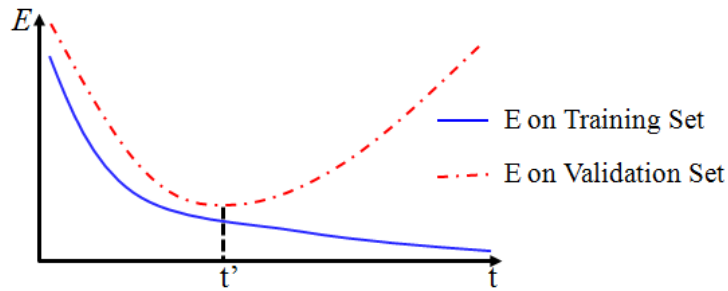
Figure 3.3. Typical error function trend for the training and validation set.

the learning process is performed for too long, or also when the feeding training examples are rare [20].

For what concerns the last issue, the selection of a proper number of training samples, which must be also representative of the correspondent classes, is not a trivial task. The use of too few training samples in neural networks means that characteristics of the classes cannot be derived, while the use of a too large a number of training samples requires more time for learning.

Several rules have been proposed to determine the optimum number of training samples. For example, in [23] the authors recommend the use of approximately 5÷10% of the image to train a network with a satisfactory performance. However, it has to be highlighted the fact that the required size of the training set depends upon the complexity of the network. Therefore, many techniques have been based on the dimensions of the ANN.

For instance, in [24] the authors derived the choice of the number of training samples from the number n_i of input nodes. They found that $60 \cdot n_i(n_i + 1)$ samples allowed a near-optimal performance.

Another criterion introduced in [25] suggests that the number of training samples should be about 10 times the number of weights in the network.

Even though the size and representativity of the training data are of considerable importance, also the distribution of the data can affect the accuracy of the results, since the ANN minimizes the overall error. In the particular case of the images classification, many studies address the issue by setting the size of the training set per class on the basis of the extension of the area covered by each class. However, this approach can lead the network to bias its decision towards the more represented classes, as this allows the algorithm to lower the overall error, which is much more heavily influenced by the more common cases [17]. Thus, since the class accuracy depends upon the size of the training set per class, the data set may need to be modified by replicating the less numerous cases, or removing some of the samples from the most represented classes [26].

Another critical issue in the use of ANNs is to determine the point at which the learning process should be terminated before the overfitting occurs.

One of the easier criterion is that training should stop after a set number of iterations. Another stopping criterion is when the estimated error for the training data is lower than a user-specified level. However, such approaches require for some a priori knowledge of the expected minimum error value that the network can achieve.

A second set of criteria is based on the magnitude of the gradient of the error. The learning phase is stopped when the magnitude of the gradient is small, being zero at the minimum point. On the other hand, this method could lead to erroneously stop the training in a local minimum or in a plateau of the error function, since also in these cases the magnitude of the gradient will be small.

Finally, a more direct and widely accepted way is cross-validation, which consists in

testing the algorithm on an independent validation set of data. Generally the error function decreases monotonically towards zero. If the testing error stops dropping, or starts to rise (Fig.3.3), this indicates that the network is starting to overfit the data. In order to get the best predictive and fitted model, the training phase should be stopped as soon as the minimum of the error function of the validation dataset is reached (*early stopping rule*). However, it can be misleading to stop the learning process at the point where the first rise in the error on the validation data occurs, as error usually starts to decrease again after the first increase. Therefore, determining the best point to stop using cross-validation is not straightforward, and it requires careful design of the learning process [17].

Chapter 4

Pulse Coupled Neural Networks (PCNNs)

Pulse Coupled Neural Network (PCNN) is a relatively novel model, based on the implementation of the mechanisms underlying the visual cortex of small mammals. A first prototype was designed in the early nineties by Eckhorn *et. al* [27], but some modifications of the original model were worked out to fit the PCNN to each specific application. In this regard in literature several papers dealing with the use of PCNN for satellite images edge detection [28], segmentation [28–30], object extraction [31–33] or recently for change detection purposes [34], [35] can be found.

4.1 PCNN model

PCNNs belong to the class of unsupervised neural networks, hence no training must be carried out.
Differently from the most common neural network architectures, the PCNN model is not conceived as a multi-layers structure, but it consists of multiple nodes coupled together with their neighbors within a definite distance, forming a bi-dimensional grid. The schematic representation of the PCNN is depicted in Fig.4.1. Each pixel of the image to process represents a neuron of the net, characterized by two input compartments. Indeed the so-called *Feeding* compartment receives both an external and a local stimulus, whereas the *Linking* partition only receives a local input. A third compartment integrates the PCNN structure: it is a dynamic threshold which is compared to the internal activity. When the internal activity, obtained by the combination of the feeding and linking functions, becomes larger than the threshold, the neuron fires, while the threshold rapidly increases. Afterward it decays, epoch by epoch, until its value is again exceeded by the internal activity value. As a consequence, when applied to image processing, the PCNN yields a series of binary pulsing signals, each associated to one pixel or to a cluster of pixels. From the pulsing neuron (or group of neurons) an auto-wave emanates, and propagates without any reflection or refraction property [36].

Formally the PCNN algorithm is described by the following expressions:

$$F_{ij} = e^{-\alpha_F} \cdot F_{ij}[n-1] + S_{ij} + V_F \sum_{kl} M_{ijkl} Y_{kl}[n-1] \tag{4.1}$$

$$L_{ij} = e^{-\alpha_L} \cdot L_{ij}[n-1] + V_L \sum_{kl} W_{ijkl} Y_{kl}[n-1] \tag{4.2}$$
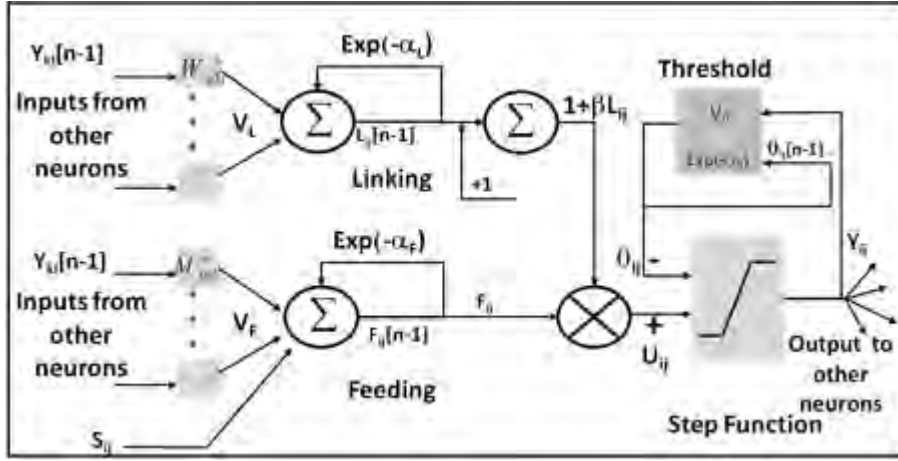
Figure 4.1. PCNN model scheme.

where S_{ij} is the input, i.e. the normalized gray level of the pixel, to the neuron (i,j), F_{ij} and L_{ij} the values of its Feeding and Linking compartments respectively. As the equations highlight, each neuron communicates with its neighbors (kl) by means of the weights given by M and W kernels, which typically refer to the Gaussian weight functions with the distance. Y represents the output of a near neuron, related to the previous iteration $[n$-$1]$. Both the Feeding and the Linking modules have a memory of the previous state, which decreases in time by the exponent term. V_F and V_L are two normalizing constants. The internal activity of the neuron, U, derives from the combination of the feeding and linking states, through the linking strength:

$$U_{ij}[n] = F_{ij}[n] \cdot \{1 + \beta L_{ij}[n]\} \qquad (4.3)$$

The condition to yield a pulse in the neuron is defined by:

$$Y_{ij}[n] = \begin{cases} 1 & \text{if } U_{ij}[n] > \theta_{ij}[n] \\ 0 & \text{otherwise} \end{cases} \qquad (4.4)$$

The variation in time of the threshold is given by the equation:

$$\theta_{ij}[n] = e^{-\alpha_\theta} \cdot \theta_{ij}[n-1] + V_\theta Y_{ij}[n] \qquad (4.5)$$

where V_θ is a constant, generally one order larger than the average value of U [36].

The original PCNN algorithm consists of iteratively computing (4.1) through (4.5) until the user decides to stop.
At the beginning of the image processing F, L, U and Y are set to zero, while θ can be set to zero or to a larger value depending upon the user's needs. If it is initialized to zero then any neurons, having a larger internal activity, will pulse in the first iteration. The consequent sharp increase of θ value makes necessary several further iterations before they will fire again. On the contrary, if the threshold is initially set to a high value, first iterations may not produce any pulses since θ needs to decay, but the frames with useful information will be produced in the earlier epochs than in the 'initially 0' scenario.

The result of a PCNN image processing depends on many parameters, that can be varied in order to perform a specific task. Specifically, α_F and α_L control the rate of decay of the feeding and linking stimulus, respectively. V_F and V_L can enlarge or reduce the influence of surrounding neurons. The user can act on V_F in order to avoid the auto-wave to enter any region where the stimulus is null. It is possible by setting V_F to zero. However, there is a range of values that allows the auto-wave to travel within a limited distance [36].

M and W refer to the mode of inter-connection among neurons in the feeding and linking receptive field. The values of such matrices determine the synaptic weight strength.

The coefficient β, representing the weight of the linking channel in the internal activity of the neuron, significantly affects the image segmentation. If much influence from the linking channel is expected, β should be given larger value.

The time constants, and the offset parameter of the firing threshold can be properly set to adjust the conversions between pulses and magnitudes. The dimension of the convolution kernel affects the propagation speed of the auto-wave. The pulse behavior of a single neuron strongly depends on α_θ, which controls the decay of the threshold, and on V_θ, which affects the height of the threshold increase after the neuron fires. Smaller α_θ can make the PCNN work more meticulous but it would take much time to finish the processing. But if the neuron is expected to fire only one time, α_θ may be set to a large value.

4.2 PCNN properties

The PCNN architecture exhibits a sort of synchronizing behavior, which leads the segments in the image to pulse together. Such a synchronicity, which is particularly evident in the first iterations, occurs when the internal activity of a neuron is very close to the dynamic threshold, but lower than θ, and its neighbor fires. The stimulus coming from the neighbor causes the U increase thus allowing the neuron to fire prematurely. However, as the iterations progress, the segments tend to de-synchronize. This behavior is particularly evident in more complex images, due to residual signals. Indeed, during the image processing, since the neurons begin to receive information indirectly from farther neurons, the behavior of the net is modified and the synchronicity fails [36].

(a)   (b)
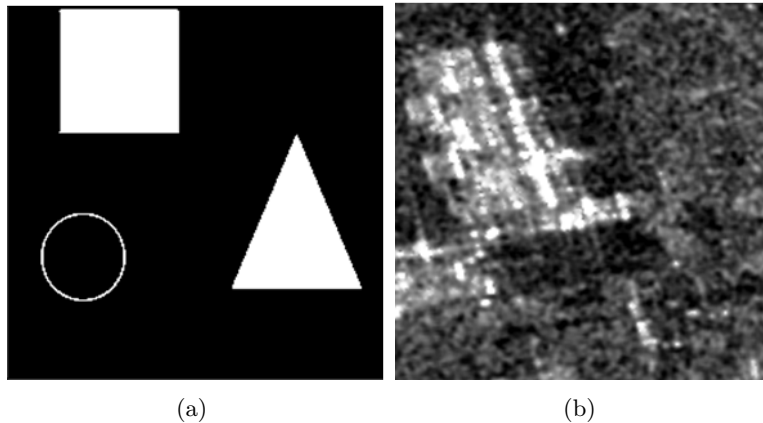
Figure 4.2. a) Binary image (200x200 pixels); b) X-band SAR image (200x200 pixels) (©ASI).

In Fig.4.3 and Fig.4.6 are shown two examples which clarify the pulsing nature and the synchronicity property of the PCNN. The first one is the result of a simple binary image processing (Fig.4.2(a)), while the second one is obtained from a X-band SAR frame (Fig.4.2(b)), which is typically affected by a strong speckle noise. During the image processing, neighbor pixels having very similar internal activity identify segments which pulse together when U overcomes the local threshold. A sort of synchronicity is evident during the PCNN processing of the simplest picture (Fig.4.2(a)), whereas the complexity of the SAR image (Fig.4.2(b)) leads to an earlier lack of synchronicity. Fig.4.4 and Fig.4.7 show the states of a single neuron, which fires (Y=1) when the internal activity U overcomes the

dynamic threshold. Afterward, θ undergoes a large growth and it takes several iterations before the threshold decays enough, allowing the neuron to pulse again.

In order to translate the pulse images into a single vector of information, the *time signal* can be computed as:

$$G[n] = \frac{1}{N} \sum_{ij} Y_{ij}[n] \qquad (4.6)$$

It represents the average number of neurons firing at the epoch $[n]$, in an image of N pixels. Although both the wave signatures which are plotted in Fig.4.5 and Fig.4.8 exhibit a quite periodic trend, the time signal derived from the binary image is more regular than that one related to the more complex SAR example. An interesting property of the PCNN algorithm is the wave signature invariance to changes in rotation, scale, shift, or skew of an object within the scene. Such a behavior makes the PCNN a suitable tool to handle the processing, features extraction or change detection of VHR SAR images, where the view angle of the sensor may play an important role.
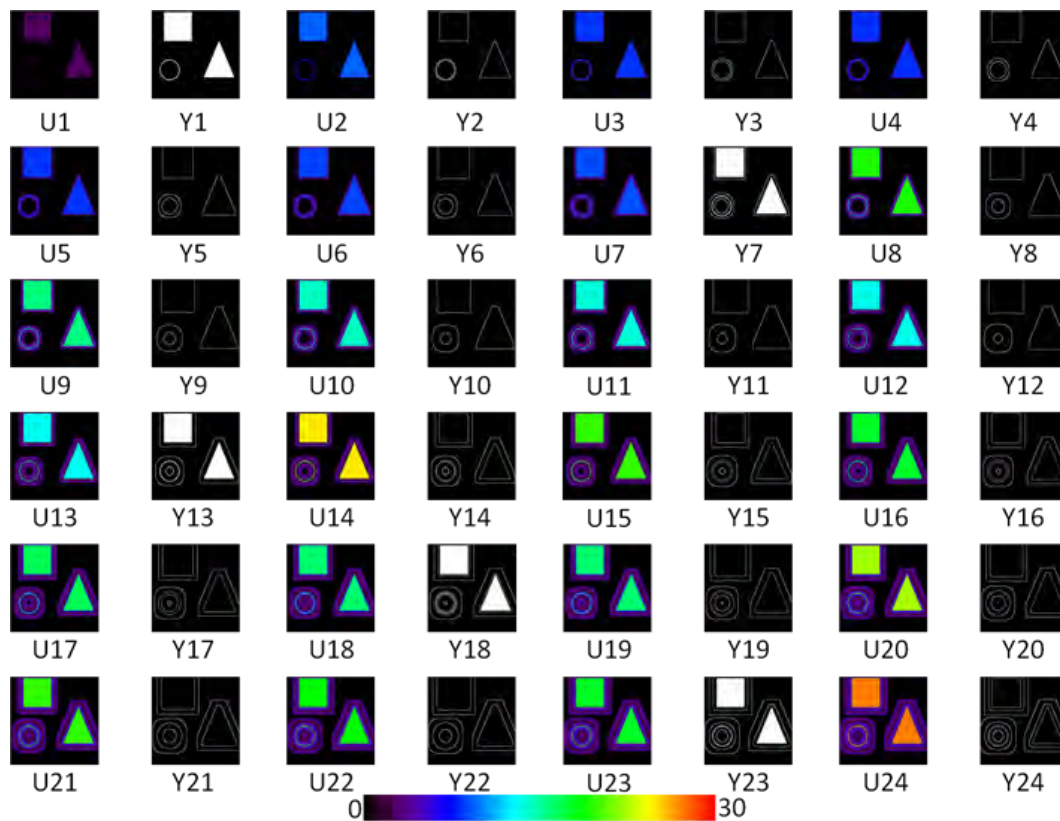
Figure 4.3. PCNN processing of the image in Fig.4.2(a). U: internal activity; Y: PCNN output.

Figure 4.4. Progression of the states of a single neuron referred to the image processing in Fig.4.3.

Figure 4.5. PCNN time signal referred to the image processing in Fig.4.3.

Figure 4.6. PCNN processing of the image in Fig.4.2(b). U: internal activity; Y: PCNN output.
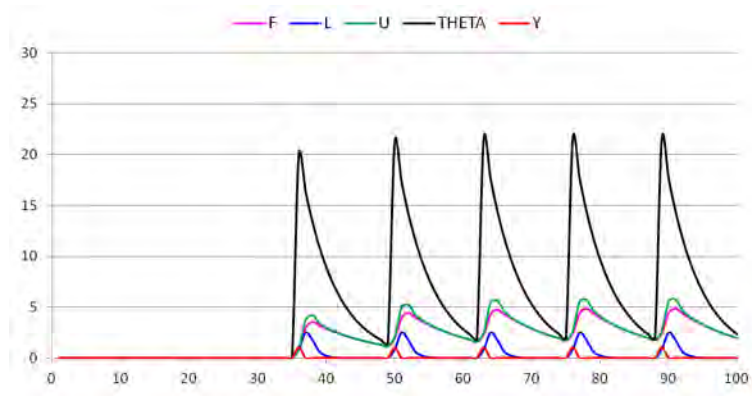
Figure 4.7. Progression of the states of a single neuron referred to the image processing in Fig.4.6.
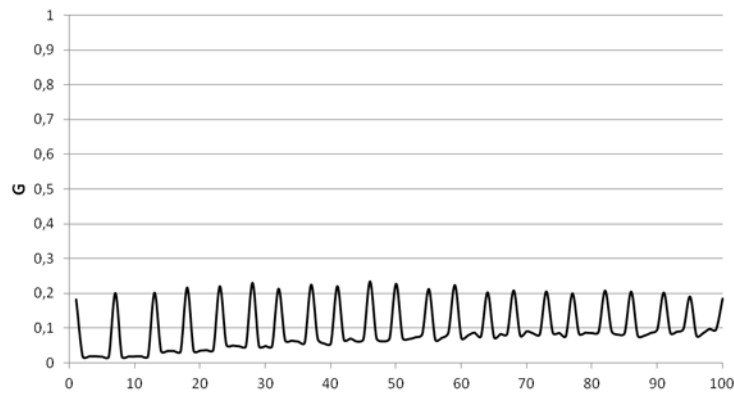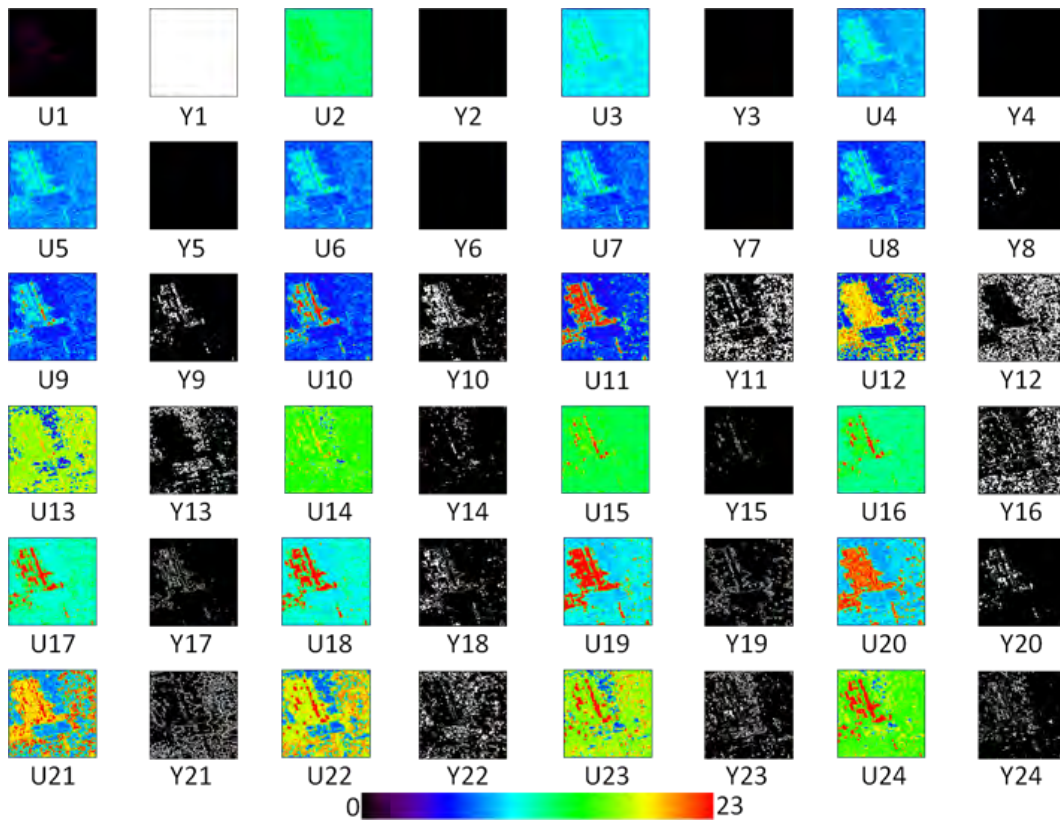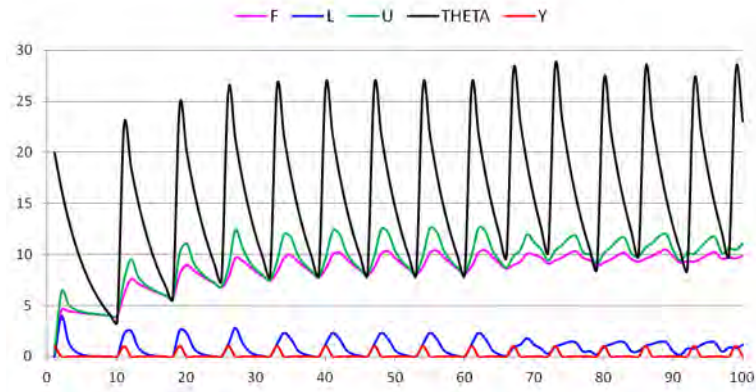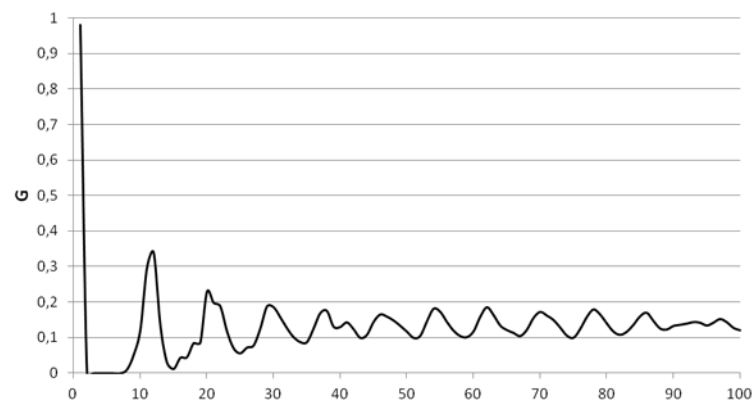

Figure 4.8. PCNN time signal referred to the image processing in Fig.4.6.

Chapter 5

Gray Level Co-occurrence Matrix (GLCM)

Identification of objects or regions of interest in an image, and therefore classification tasks are normally addressed by investigating on the spectral, textural and contextual features of the image itself. Spectral information regard the average tonal variation in the electromagnetic spectrum bands, whereas textural features describe the spatial distribution of tonal variations over relatively small areas, and the contextual characteristics of an image take into account the information derived from areas surrounding the one under investigation. In each specific case of study a class of features can dominate over the others, or it may occur that all the typology of information are equally effective for the image interpretation. In the framework of the present research, the only information carried by the backscattered signal may not be sufficient to discriminate different land uses within single polarization SAR data. On the other hand, many authors demonstrated that texture features can be useful for the interpretation of SAR images ([37–41]). This study investigated on the possibility of exploitation of textural parameter stemmed from the computation of the Gray Level Co-Occurrence Matrix (GLCM) introduced by Haralick [14] in 1973. In this chapter the GLCM and the most common textural features stemmed from them will be described.

5.1   GLCM definition

Since its introduction by Haralick, the Gray Level Co-Occurrence Matrix (GLCM) technique has been widely applied in many different research fields, from biomedical to remote sensing image interpretation. In [14], Haralick assumes that the texture information in an image is contained in the overall or "average" spatial relationship, which the gray tones in the image have each other. The GLCM is the matrix of relative frequencies P_{ij} with which two neighboring resolution cells (Fig.5.1), having i and j gray tone respectively, and located at a specified distance d along the direction θ , occur in the image, or in a smaller window of it.

Let be I an image of dimension $N_x \times N_y$, and $G = 1, 2, ..., N_g$ the set of N_g quantized gray tones. The image I can be represented as a function which assigns some gray tone in G to each resolution cell:

$$I : L_y \times L_x \rightarrow G \tag{5.1}$$

where $L_y \times L_x$ is the set of resolution cells of the image. Let consider two neighboring pixels $I(k,l)$ and $I(m,n)$, characterized by i and l gray level, respectively. The spatial relationship between the resolution cells, in terms of gray tone unnormalized frequencies, can be defined as:
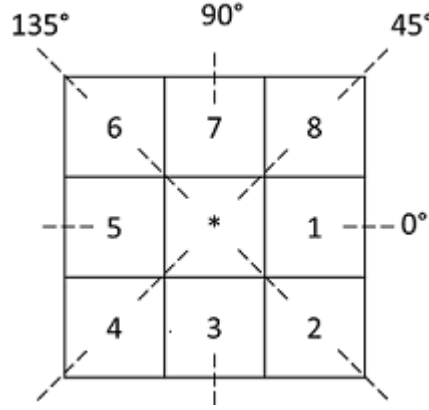
Figure 5.1. Neighborhood resolution cells.

$$P(i, j, d, 0°) = \#\{((k,l)), (m,n) \in (L_y \times L_x$$
$$\times (L_y \times L_x|k - m = 0, |l - n| = d, I(k,l) = i, I(m,n) = j\} \qquad (5.2)$$

$$P(i, j, d, 45°) = \#\{((k,l)), (m,n)) \in (L_y \times L_x$$
$$\times (L_y \times L_x)|(k - m = d, l - n = -d) or (k - m = -d, l - n = d), \qquad (5.3)$$
$$I(k,l) = i, I(m,n) = j\}$$

$$P(i, j, d, 90°) = \#\{((k,l)), (m,n) \in (L_y \times L_x)$$
$$\times (L_y \times L_x)||k - m| = d, l - n = 0, I(k,l) = i, I(m,n) = j\} \qquad (5.4)$$

$$P(i, j, d, 135°) = \#\{((k,l)), (m,n) \in (L_y \times L_x)$$
$$\times (L_y \times L_x)|(k - m = d, l - n = -d) or (k - m = -d, l - n = -d), \qquad (5.5)$$
$$I(k,l) = i, I(m,n) = j\}$$

where $\#$ indicates the number of elements in the set. A characteristic of the GLCM is its symmetry, so that: $P(i, j; d, \theta) = P(j, i; d, \theta)$. In some cases it could be useful to normalize the GLCM by dividing the co-occurrence frequencies by R, which is the number of possible neighbor pixel pairs in a given direction θ, and at a given distance d. If $d = 1$ (Fig.5.2), the number of nearest neighbor pairs is equal to $2N_y (N_x - 1)$ in the horizontal direction ($\theta = 0°$), $2N_x (N_y - 1)$ in the vertical direction ($\theta = 90°$), $2 (N_y - 1)(N_x - 1)$ in each diagonal direction by symmetry ($\theta = 45°$ or $\theta = 135°$). In Fig.5.2 it is represented an image of dimension 4×4 pixels, as well as the computed GLCM ($d = 1, \theta = 0°$) and its normalized form.

5.2   Textural Features stemmed from the GLCMs

In his work, Haralick [14] introduced a set of fourteen textural features which can be derived from the computation of the GLCM. Some of these parameters describe specific textural characteristics of the image (i.e. homogeneity, contrast) and possibly reveal the presence of organized structure within the image. Other features relate to the complexity and nature of the occurring gray-tone transitions. In many remote sensing applications, and especially
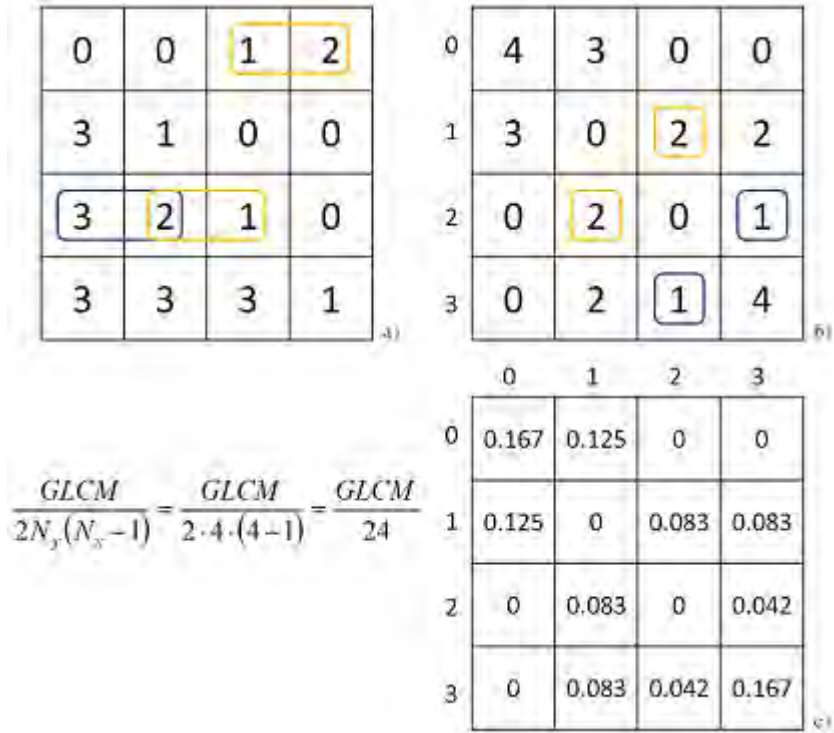
Figure 5.2. a) Example of 4×4 image. Each pixel is characterized by a quantized gray tone. b) GLCM evaluated for d=1 and $\theta = 0°$. c) Normalized GLCM.

in image classification tasks, eight out of fourteen features are commonly successfully exploited. However, it is worth noting that not all these parameters are independent each others [37], hence a careful analysis is necessary to select the most suitable ones, without any redundancy of information. In the following sections the aforementioned eight textural parameters will be described by dividing them into three groups:

- Contrast group;

- Orderliness group;

- Stats group.

5.2.1 Contrast group

Contrast

$$f_1 = \sum_{i,j=0}^{N_g-1} P_{i,j}(i-j)^2 \tag{5.6}$$

Contrast is an index of the gray-level variations between two neighbor pixels. The GLCM derived from a low contrast image presents the highest elements values around the principal diagonal ($|i-j| = 0$). A low contrast image is not necessarily characterized by a low variance of the gray tones; on the other hand, to keep the contrast low, it is sufficient to maintain variance low [37]. In addiction, high contrast values imply high contrast texture. It means that the GLCM *contrast* feature tends to be highly correlated with spatial frequencies (i.e. difference between the highest and lowest values of a contiguous pair of pixels) when the module of the spatial displacement tends to one.

Dissimilarity

$$f_2 = \sum_{i,j=0}^{N_g-1} P_{i,j}|i-j| \tag{5.7}$$

Differently from the *contrast* feature, in the *dissimilarity* measure the weights $|i-j|$ do not increase exponentially but linearly. *Contrast* and *dissimilarity* are very similar parameters, hence strongly correlated.

Homogeneity or Inverse Difference Moment

$$f_3 = \sum_{i,j=0}^{N_g-1} \frac{P_{i,j}}{1+(i-j)^2} \tag{5.8}$$

By assuming larger values for smaller gray tone differences, the *inverse difference moment* measures the homogeneity of the image. Therefore this parameter is more sensitive to low contrast, organized texture structures, whose GLCM is characterized by the presence of near diagonal elements [37], [42]. By their intrinsic meaning, it is clear that *homogeneity* and *contrast* are closely and inversely correlated.

5.2.2   Orderliness group

Energy or Angular Second Moment or Uniformity

$$f_4 = \sum_{i,j=0}^{N_g-1} (P_{i,j})^2 \tag{5.9}$$

Energy is a measure of the uniformity of the image texture. When only similar gray level are present, or there is a small number of gray tone transitions for a given displacement vector, few elements of the GLCM will be larger than 0 and close to 1, while most of the entries will be close to 0. Hence, high *energy* values occur in orderly images, where the gray level spatial distribution is constant or periodic [42]. For this reason *energy* is strongly uncorrelated to first order statistical variables (e.g. contrast and variance) [37].

Entropy

$$f_5 = \sum_{i,j=0}^{N_g-1} P_{i,j}(\ln P_{i,j}) \tag{5.10}$$

If on one side the GLCM *energy* measures the uniformity of an image, on the other side the *entropy* textural feature estimates its disorder. Maximum values of *entropy* occur when the image is characterized by pixels with random gray tones. In fact in this case the GLCM elements are represented by the same value. As much as the disorder increases, as much the number of GLCM entries having very small number becomes larger. Conceptually, *entropy* and *energy* parameters are strongly and inversely correlated [37].

5.2.3   Stats group

Mean

$$f_6 = \sum_{i,j=0}^{N_g-1} i(P_{i,j}) \tag{5.11}$$

In the GLCM *mean* equation, the pixel value is weighted not by its frequency of occurrence by itself (as in the well known mean equation) but by its frequency of its occurrence in combination with a certain neighbor pixel value. The P_{ij} value is the probability value from the GLCM, i.e. how many times that reference value occurs in a specific combination with a neighbor pixel. It is also worth to notice that, since the combinations of pixels are different for the horizontal and vertical GLCM, *mean* features are different in the two cases.

Variance

$$f_7 = \sum_{i,j=0}^{N_g-1} P_{i,j}(i - \mu_i)^2 \tag{5.12}$$

Variance is a measure of the dispersion of the values around the mean. Given the symmetry property of the GLCM, this parameter will be the same if i is replaced by j. In texture measures it performs the same task as does the common descriptive statistic called variance. However, GLCM *variance* deals specifically with the dispersion around the mean of combinations of reference and neighbor pixels. Moreover it is strongly correlated to the first order statistical variables such as standard deviation [37], [43]. In the specific case of a square image, for $d = 1$ and $\theta = 0°$ or $\theta = 90°$, the first order statistical variance will assume the same value of the GLCM *variance* parameter. It has to be pointed out the fact that *variance* is not correlated to the GLCM *contrast* [37]. In fact, when the displacement tends to 1, low spatial frequencies and low contrast value may be associated, at the same time, to either a high or low *variance* measure.

Correlation

$$f_8 = \sum_{i,j=0}^{N_g-1} \frac{(i - \mu_i)(j - \mu_j) \cdot P_{ij}}{\sqrt{\sigma_i^2 \cdot \sigma_j^2}} \tag{5.13}$$

Correlation expresses the linear dependency between the gray tones within the image [44]. The maximum value, equal to 1, is reached when a linear relationship occurs between the gray levels of pixel pairs, therefor regardless of pixel pairs occurrence. By its definition, *correlation* is independent from either the GLCM *energy* and *entropy*, as well as from the GLCM *contrast* [37].

Part III

Dataset

Chapter 6

Test sites

Most of the research activities included into this thesis has been carried out in the framework of the CAMLAND (Computer-assisted monitoring of land cover and its changes by SAR images) project, partially funded by ASI (AO project N. 1484, agreement No. I/061/09/03). The main goal was to develop numerical tools aiming at the extraction of land cover information from COSMO-SkyMed images, with particular reference to metric-resolution data acquisitions. The design of the tools is essentially driven by the endeavor to attain automation of the image processing for producing land cover and change maps in the suburb of large cities, in order to monitor the occurring urbanization. The evaluation of the performance of the processing tools has been carried out against the high and very-high resolution multi-spectral QuickBird and WorldView-2 images provided by DigitalGlobe Co. [45], [46].

Because of its current and continuous dramatic conversion of land cover from rural to urban, the south-eastern outskirt of Rome (Italy) was chosen as a suitable test site for the validation of the implemented algorithms. A views of the examined scenario, acquired by the QuickBird satellite is shown in Fig.6.1.

This area, centered on the Tor Vergata University campus, covers residential districts in the north and west side, as well as industrial and business zones mainly located in the

Figure 6.1. QuickBrid image of Tor Vergata, Rome (Italy). Date of acquisition: 12th April 2011. ©DigitalGlobe
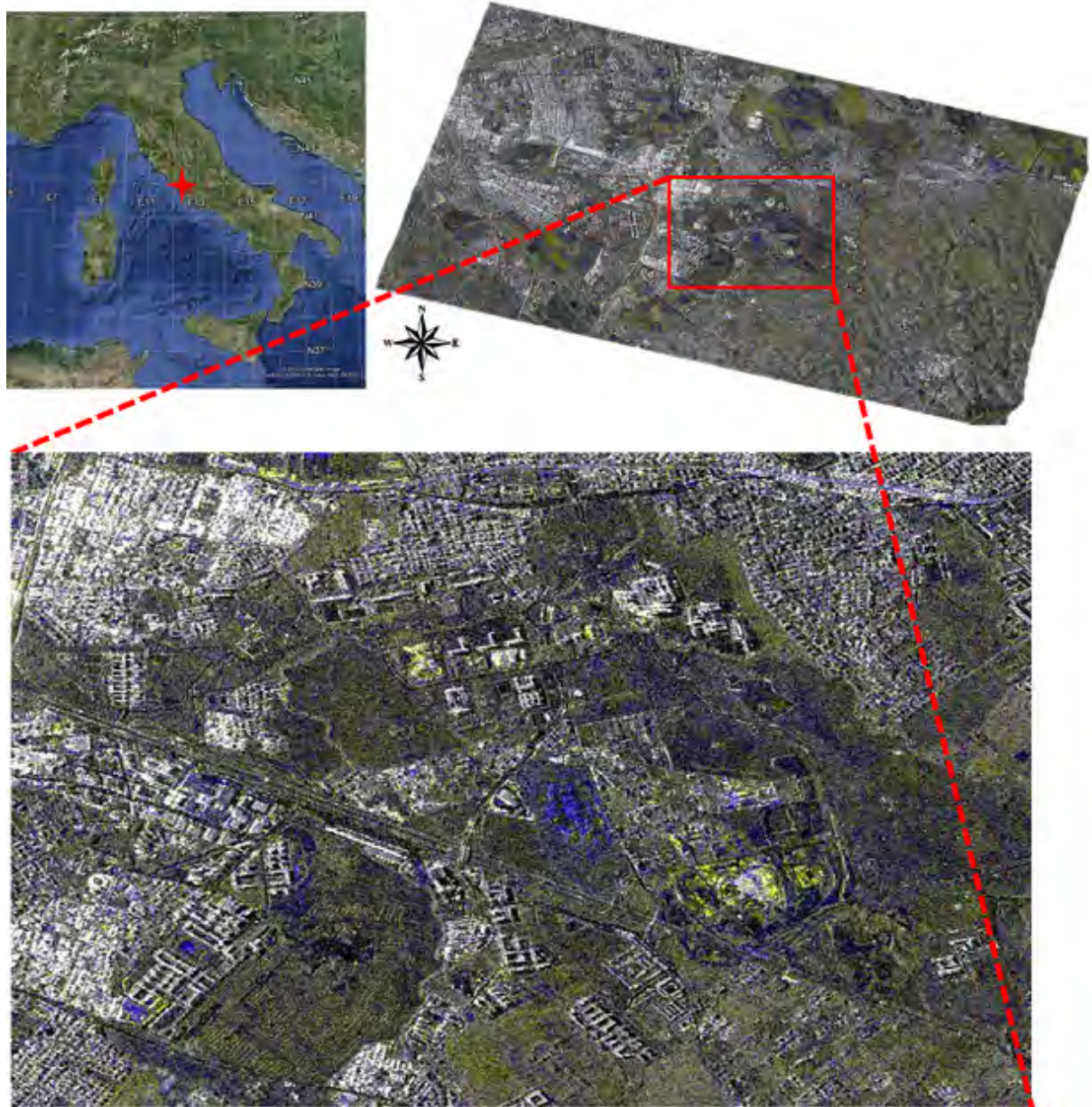
Figure 6.2. False color composition of a pair of Stripmap HImage COSMO-SkyMed ©ASI images. R and G: 20th September 2008; B: 22nd April 2011. The objects in blue represent manmade structures which have been built up during the three years between the two acquisitions. The yellow objects mainly identify construction sites. The unchanged areas are made out by white and dark pixels.

south, and it extends up to the western countryside of Frascati (Rome, Italy). In such a scenario, a rather large variety of surfaces and features may be potentially identified by exploiting very high resolution satellite images. As can be observed in Fig.6.1, the area of interest is characterized by different land uses and typology of buildings, such as large structures (e.g. hospital, business buildings with glass walls, industrial constructions whose roofs are made of different materials, or bus shed, shopping malls), and residential buildings in high or low density urban environments, as well as isolated smaller houses. Moreover the road network here consists of large streets, including the motorway that crosses the scenario represented in Fig.6.1, and narrower ones. In spite of the current and constant urbanization process that is affecting the Tor Vergata area, wide green spaces are still present: meadows, cultivated fields, tree-lined avenues, groves, cover a rather big percentage of this outskirt. Nevertheless, in Fig.6.2 it is pretty evident how the replacement of vegetated zones with new constructions is going to change rather quickly the scenario. In Fig.6.2 a pair of Stripmap HImage COSMO-SkyMed imagery acquired at three years apart each other (2008 - 2011) has been assembled in a false color composite picture. By filtering the old acquisition in the red and green bands, and the new one in the blue channel, it is possible to observe, besides white and dark unchanged zones, also blue and yellow objects. The former refer to areas characterized by high backscattering in the most recent acquisition, and low backscattering in the old one. Therefore, blue pixels mainly represent new buildings. On the other side, yellow areas put in evidence the presence of high backscattering from objects in the old acquisition, which turns into a low backscattering in the new image. Typically these features identify areas under construction, where the presence of vehicles, cranes, scaffolding, containers, can vary continuously during the period of work.

As it will be shown in the following chapters, the use of a lonely single polarized COSMO-SkyMed image is not sufficient to identify accurately so many different typology of surfaces. However, it will be proven that the use of more than one image, or the exploitation of further information stemmed from the SAR acquisitions, such as textural features and, when available, the degree of the interferometric coherence, is an effective approach aiming at the improvement of the land cover and change detection maps, in terms of both accuracy and number of classes. Hence, being aware of the challenge carried by the very high spatial resolution of COSMO-SkyMed data, and also considering the most typical problems affecting the SAR products, the objective of the classification exercises in the sub-urban environment represented by the Tor Vergata area has been set at the identification of at least three land covers: *manmade structures*, *asphalt* and *vegetation*. Actually, in this work, four classes have been successfully distinguished by separating the *low vegetated* areas from the *trees* category. Moreover, focusing the study on a cultivated region of the area of interest, and by the joint exploitation of the SAR and optical properties of COSMO-SkyMed and WorldView-2 sensors respectively, it has been possible to correctly identify up to six different natural surfaces.

Some assessment of the designed algorithms has been also carried out in the Great Denver area, Colorado, USA (Fig.6.3). Although the general landscape of this region is similar to the Tor Vergata site, it is characterized by different specific features. Specifically, besides the above mentioned four land covers, the class *water* has been added because of the large extension of this type of surface in the area.

In the following chapters, a summary of the characteristics of the satellites products which have been used in this research study will be presented.

Figure 6.3. ©Google Earth image of the Great Denver area, Colorado, USA.

Chapter 7

VHR X-Band SAR dataset

In principle, the study has been focused on COSMO-SkyMed images (Chapter 2), with a particular interest for the VHR Spotlight data (1 m spatial resolution). However, the developed algorithms have been tested also on Stripmap HImage COSMO-SkyMed data (3 m spatial resolution). In the frame of the CAMLAND project, a main requirement was the availability of an extensive dataset of images of Tor Vergata, acquired as much as possible frequently and with the same geometry of acquisition, in order to perform a continuous and automatic monitoring of the urbanization process. In spite of the large amount of raw data (level 1A) provided by ASI, a smaller set has been selected to assess the performance of the implemented techniques of *Classification* (C), *Data Fusion* (DF), *Building Detection* (BD) and *Change Detection* (CD). It has to be pointed out the fact that, as it will be discussed in the Chapter 13, the developed change detection system requires a first classification phase. Therefore, all the pairs of images (Spotlight and Stripmap) exploited for the change detection study have been also classified at the same time.

All the raw data have been undergone to a pre-processing phase before being used for the various applications. Specifically, the images have been geocoded by performing the range Doppler terrain correction and calibration (see §1.3). Afterward, the adaptive Enhanced Lee filter has been used to mitigate the speckle noise. This method is an adaptation of the Lee filter and similarly uses local statistics (coefficient of variation) within individual filter windows. Each pixel is put into one of three classes: homogeneous, heterogeneous, or point target. Each class type is treated differently. For the homogeneous class, the pixel value is replaced by the average of the filter window. For the heterogeneous class, the pixel value is replaced by a weighted average. For the point target class, the pixel value is not changed [5],[47]. Subsequently, when required by the specific application, pairs of images have been co-registered.

7.1   Spotlight COSMO-SkyMed images

In Table 7.2 the list of Spotlight COSMO-SkyMed dataset is reported. The images of the Tor Vergata area, depict the same scenario from June 2010 to October 2011. As it can be noted, all the used images are acquired by the SAR looking to the right side of the ascending orbit (RA), and with the same incidence angle ($\sim 25.3°$). Moreover, nine out of eleven images are in HH polarization. Such a choice allowed for the testing of the automatic classification algorithm, as well as of the change detection technique. On the other side, the use of images with a different polarization (VV), but same geometry of acquisition, has been useful in order to enhance the vegetation mapping in a cultivated region of the site under investigation (Chapter 11). A further Spotlight acquisition has been exploited for the classification of the Gran Denver area in Colorado (USA).

Table 7.1. Spotlight COSMO-SkyMed dataset (1 m spatial resolution). Each image has been used for one or more applications: Classification (C), Data Fusion (DF), Change Detection (CD) and Building Detection (BD).

SPOTLIGHT COSMO-SkyMed						
Site	**Satellite**	**Date**	**Pol.**	**Orbit**	**Inc. Angle**	**Task**
Tor Vergata, Rome, Italy	CSK3	08/06/2010	HH	RA	24.8÷25.7	C + DF
	CSK2	23/06/2010	VV	RA	24.8÷25.7	DF
	CSK2	09/07/2010	HH	RA	24.7÷25.6	C + CD
	CSK3	10/07/2010	HH	RA	24.8÷25.6	C
	CSK3	26/07/2010	VV	RA	24.9÷25.8	DF
	CSK2	16/12/2010	HH	RA	24.9÷25.7	C
	CSK3	01/01/2011	HH	RA	24.0÷25.7	C
	CSK3	02/01/2011	HH	RA	24.8÷25.7	DF + BD
	CSK2	23/04/2011	HH	RA	24.9÷25.8	DF
	CSK3	11/06/2011	HH	RA	24.8÷25.7	C+ CD
	CSK1	08/10/2011	HH	RA	24.7÷25.6	DF
Great Denver, Colorado, USA	CSK3	11/12/2010	HH	RA	25.3÷26.1	C

7.2 Stripmap SAR images

A set of five Stripmap HImage COSMO-SkyMed images, four of which taken over the Tor Vergata district, and one over the Great Denver area, have been used as input to the classification algorithm (Table 7.2). All of them are acquired in HH polarization, with an incidence angle ranging in the interval $36° \div 39°$ (Tor Vergata), and $25° \div 28°$ (Great Denver). As it can be observed in Table 7.2, the change detection exercise has been carried out on a pair of images acquired by the SAR system looking to the left along the descending orbit (LD). A TerraSAR-X acquisition of Tor Vergata has been also included into the dataset, since it has been used to tailor the PCNN based algorithm to perform the buildings detection task. Similarly to the others, the considered TerraSAR-X image was acquired in Stripmap mode (3 m spatial resolution), in HH polarization and with an incidence angle of about $33.7°$.

Table 7.2. Stripmap HImage COSMO-SkyMed (3 m spatial resolution) and Stripmap TerraSAR-X (3 m spatial resolution) dataset. Each image has been used for one or more applications: Classification (C), Change Detection (CD) and Building Detection (BD).

STRIPMAP HIMAGE COSMO-SkyMed						
Site	**Satellite**	**Date**	**Pol.**	**Orbit**	**Inc. Angle**	**Task**
Tor Vergata, Rome, Italy	CSK2	20/09/2008	HH	LD	36.2÷38.9	C + CD
	CSK2	24/05/2010	HH	RD	36.1÷38.8	C
	CSK3	25/05/2010	HH	RD	36.1÷38.8	C
	CSK2	22/04/2011	HH	LD	36.1÷38.8	CD
Great Denver, Colorado, USA	CSK2	10/12/2010	HH	RA	24.9÷28.3	C
STRIPMAP TerraSAR-X						
Tor Vergata, Rome, Italy	TSX	24/11/2007	HH	RA	∼ 33.7	BD

Chapter 8

VHR Multispectral sensors dataset

Although this work has been developed with the purpose of investigating on the potentiality of the information embedded within the VHR X-band COSMO-SkyMed product, VHR multispectral optical data contributed positively to reaching the goal. Indeed, despite the metrical spatial resolution of the SAR images, their interpretation is still quite tough. Therefore QuickBird and WorldView-2 [46] imagery have been used to understand the different coverages of the area of study. Based on their visual interpretation, the ground truth (GT) has been collected to assess the accuracy of the final products of the building detection, classification and change detection algorithms. Moreover, taking advantage of the eight bands of the WorldView-2 imagery [46], data fusion (DF) experiments have been carried out to enhance the vegetation mapping by jointly exploiting mutlispectral and SAR data (Chapter 11). A list of the used optical images provided by DigitalGlobe is reported in Table 8.1.

Table 8.1. Multispectral QuickBird and WorldView-2 dataset. Each image has been used for the visual interpretation of the scenario, and hence as Ground Truth (GT), or for the Data Fusion (DF) experiments.

MULTISPECTRAL QUICKBIRD					
Site	**Satellite**	**Date**	**N. Bands**	**Spatial Res.**	**Task**
	QB	08/12/2008	4	2.4 m	GT
Tor Vergata,	QB	30/09/2010	4	2.4 m	GT
Rome, Italy	QB	12/04/2011	4	2.4 m	GT
	QB	20/05/2011	4	2.4 m	GT
MULTISPECTRAL WORLDVIEW-2					
Tor Vergata,	WV2	04/07/2010	8	2.0 m	GT + DF
Rome, Italy	WV2	10/10/2011	8	2.0 m	GT + DF

Part IV

Supervised approach: pixel-based classification

Introduction

The possibility of periodically collecting images over wide areas, makes the remote sensing one of the best source of information related to the land cover and land uses changes in time. Particularly complex environments are represented by the sub-urban areas, where rural zones are progressively enveloped in metropolitan regions, with the consequent deep transformation and possible severe damages of the original natural ecosystem.

The high performance of the most recent VHR optical and SAR sensors enable a better management of urban areas. To this aim, satellite images must be transformed into usable geographic products, through the classification of the different surfaces imaged in the scene. Remote sensing in the optical band and at very high spatial resolution has proven to be a useful tool for land cover and change maps production. However, since optical sensors are affected by the weather and illuminations conditions, the use of SAR imagery can become necessary, especially in those regions which are covered by clouds for most of the year, or characterized by long dark periods. In addition, SAR systems represent the best tool in case of emergency, when the availability at the earliest convenience of images carrying information of the changes possibly undergone by natural surfaces or manmade structures is vital. Metric-resolution SAR data, such as those delivered by the COSMO-SkyMed constellation [12] or by the TerraSAR-X satellite [48], [49], are expected to foster substantial progress [50], allowing the detection of small-scale objects in the urban environment (e.g.elements of residential housing, commercial buildings, transportation systems and utilities). On the other hand, the very high spatial and temporal resolution capabilities of these space platforms make unextendible the deadline for providing processing tools as much as possible automatic so as to prevent the accumulation in the archives of never seen data.

The classification of urban and sub-urban environment by using VHR SAR data, is currently one of the most appealing challenge for the remote sensing community. While in the previous decametric resolution SAR systems acquisitions the electromagnetic and geometric effects was smeared by the coarser spatial resolution, VHR SAR images are strongly affected by the complexity and variety of scattering mechanisms, even for a single isolated building. This behaviour makes the association pixel-class not obvious. Typical classification method are based on object-oriented approaches, exploiting texture and contextual properties [51],[52]. On the contrary, the pixel-based classification of VHR SAR images has been faced by few authors. In [53] fire scars mapping and suburban/agricultural land cover classification has been performed by an unsupervised neural network algorithm, the Textural Self-Organizing Map, based on the textural features of the radar image. In [54] a complex technique has been implemented in order to classify urban areas acquired by the VHR TerraSAR-X and COSMO-SkyMed sensors. The algorithm, based on a supervised approach, combines a finite mixture technique to estimate class-conditional probability density functions, Bayesian classification, and Markov random fields (MRFs). Moreover, an improvement in the discrimination of urban areas has been achieved through the inclusion of the information carried by the textural features extracted by the GLCM.

In this work a supervised pixel-based classification approach based on MLP-NN (Chapter 3) has been adopted to classify Spotlight and Stripmap COSMO-SkyMed images. The

effectiveness of neural networks for pixel based satellite image classification has been already demonstrated for both optical images and medium-high resolution SAR images ([41, 55–58]). Indeed, many authors have reported considerable advantages of ANNs over conventional methods [17]. In particular, since the neural networks set up the recognition of each class on the characteristics of the whole ensemble of training samples, they have the ability of performing supervised classification using less training data than the maximum probability [59]. Moreover, when properly trained, the NNs are rather competitive if compared with other techniques, such as the Bayesian or Support Vector Machines [60], and are able to automatically classify different images, as long as the statistics of these latter is contained in the training set.

As it has been already mentioned in Chapter 3, the performance of the MLP-NN, and hence the classification accuracy, mostly depends on the suitability of the input features, on the quality of the training set and on the neural network architecture.

In principle, the information provided by the input samples must represent the distinctive behaviour of the correspondent class to detect. An intensive analysis of the available dataset has been performed, aiming at the selection of the most effective ensamble of features to be exploited as input to the classifier. Note that most of the new SAR systems acquires only single polarization images. Since the mere use of the backscattering amplitude is likely to be not efficient to discriminate various surfaces, the lack of multi-band data can be overcome by adding spatial information, such as morphological [61] or textural [14] features. This study mainly focused on texture parameters stemmed from the computation of the GLCM(Chapter 5). Indeed, several papers in literature alreay proved that methods exploiting the GLCM features in optical and decametric resolution SAR images, are particularly successful in the urban areas classification [39, 62–67].
In order to take advantage of the unique capability of the Cosmo-SkyMed constellation, which can acquire images of the same scene with a very short revisit time, multi-temporal information has been also taken into account. An investigation on the possible improvement of the maps accuracy has been carried out by exploiting the texture information stemmed from two acquisitions, as well as the interferometric choerence.

For what concern the selection of the MLP-NN training set, in order to make the patterns as much as possible statistically representative, samples of pixels have been collected randomly over each land cover by taking in consideration all the different types of surfaces. A cross-check with optical images allowed the interpretation of each point. The ensemble of pixels has been successively partitioned, by dedicating about the $60 \div 70\%$ to the training set, and about the $30 \div 40\%$ to the validation set, which have been employed for the network design. A third independent set of pixels has been collected for the final assessment of the obtained classification maps.

As regards the design of the optimal NN topology, a heuristic approach has been adopted in each of the experiments carried out in this work. Basically, the number of hidden layers and neurons has been defined by progressively adding units and evaluating the performance of the algorithm, through the estimate of the MSE (Chapter 3). The minimization of the error has been worked out by using the SCG algorithm. The NN implementation has been carried out by the Stuttgart Neural Network Simulator package [22].

In the following chapters it will be firstly presented the analysis of the data, which led to the identification of the main types of land covers in the study site (Chapter 9). Results achieved by exploiting different input features (simple local texture, GLCM textural parameters, interferometric coherence) will be presented in Chapter 10. Finally multi-sensor, multi-temporal, and multi-polarization data have been jointly exploited in Chapter 11, where data fusion experiments have been carried out aiming at the enhancement of the accuracy of vegetation maps.

Chapter 9

Data analysis

The classification task has been carried out by using the MLP-NN algorithm. As already-highlighted in Chapter 3, the accuracy of the final map is affected by several factors, and in particular by the quality of the training samples. Depending on the variety of surfaces within the area under investigation, a crucial issue is the selection of set of representative pixels which must comprehensively describe each class of land cover. Indeed, a single category may be represented by surfaces characterized by different geometry, roughness, orientation with respect to the SAR, or different materials. As a consequence, the range and distribution of values of the backscattering can vary significantly even within the same class. In order to investigate on the possible number of land covers, which the only use of the intensity of the backscattering allows to distinguish, a first data analysis has been performed by considering the Spotlight COSMO-SkyMed image in Fig.9.1. Bearing in mind the variety of objects included into the examined region of Tor Vergata (Chapter 6), several examples have been selected, and their statistics have been computed on a set of randomly chosen pixels, to analyze all possible behaviors in terms of backscattered signal. In principle, three main categories of objects can be distinguished:

1. *manmade objects*

2. *asphalted surfaces*

3. *natural areas*

Manmade objects

Regarding the manmade structures, nine buildings, characterized by different shapes, sizes, materials and environment of settlement, have been identified (Fig. 9.2, Fig.9.3 and Fig.9.4). Two thousands of pixels have been randomly taken for each building, and the results of the statistical analysis are shown in Table 9.1. The intensity of the backscattering from a building, is due to the occasionally joint contribution of all the parts that constitute it, and eventually of others objects (nearby houses, man-made structures or parking areas) contained by the area assigned to the building [68]. This fact explains the quite wide range of values belonging to a single edifice, as pointed out by the standard deviation measures (Table 9.1).
For a better understanding of the data, the nine objects have been divided into three classes of similarity: the first includes large buildings, such as the hospital in Fig. 9.2(a), the company building Fig.9.2(b), and the shopping malls in Fig.9.2(c); the second one is composed by sheds for industrial or bus garage uses (Fig. 9.3(a) and Fig.9.3(b)); the third group contains three types of residential houses, that is tall and close to each other (Fig.9.4(a)), high and separated (Fig.9.4(b)), and isolated smaller houses (Fig.9.4(c)). The histograms

Figure 9.1. Spotlight COSMO-SkyMed image of Tor Vergata, Rome (Italy). Date of acquisition: 9th July 2010. ©ASI.

of the backscattering coefficients referred to these three categories are shown in Fig.9.5. As expected, they show a strong variability of the backscattering values, even within the same aforementioned category. In particular, the shape of the shopping mall, and its orientation with respect to the satellite, makes the associated pixels values quite low with respect to the other buildings (Fig.9.5(a)). Buildings belonging to the second category display similar backscattering (Fig.9.5(b)). The shift between the histograms is basically due to the different materials of which the sheds are made. While the differences between the residential buildings are mainly related to the environment in which they are located, backscattering values are higher in densely urbanized areas, because of the superposition of many contributions, double bounce and trihedral effects (Fig.9.5(c)).

Table 9.1. Statistics of the backscattering coefficient (dB) of a sample of pixels belonging to the buildings in Fig.9.2, Fig.9.3 and Fig.9.4.

MANMADE OBJECTS					
Type	**Min**	**Max**	**Mean**	**St. dev.**	**N. pixels**
Hospital	-17.4	22.3	-2.9	5.2	2000
Company	-22.1	18.0	-4.7	5.7	2000
Mall	-27.8	27.3	-10.3	7.5	2000
Industrial 1	-15.3	13.4	-4.2	3.1	2000
Industrial 2	-24.5	23.1	-10.1	5.2	2000
Bus shed	-28.1	15.9	-7.4	5.8	2000
Residential	-20.6	28.3	4.5	6.9	2000
Hostel	-21.9	21.3	-2.7	5.2	2000
Houses	-21.7	14.9	-5.8	5.2	2000

Asphalted surfaces

Completely different behavior is expected by smooth surfaces [3], such as asphalted areas, that should be characterized by low backscattering and high contextual homogeneity. However, the analysis of the selected objects in Fig.9.6 has stressed the influence of the roughness on the return of the X-band SAR signal. Table 9.2 shows that the mean value of backscattering of most streets is about -17 dB, whereas the motorway is characterized by higher values, as well as the parking area. But others exceptions occur, such as the smooth asphalt in Fig.9.6(c) that has low backscattering values. Such differences are probably related to the state and texture of the surface, more or less smooth, and to the presence of sidewalks, light poles and vehicles. The behavior of the analyzed various asphalted surfaces are highlighted by the histograms shown in Fig.9.7, computed by considering 200 pixels randomly selected in each ROI in Fig.9.6.

Natural surfaces

A similar investigation has been carried out for the natural areas depicted in Fig.9.8 and Fig.9.9, which represent two categories, the former including low o scarcely vegetated surfaces and bare soil, the second referred to high vegetation coverages. The corresponding analysis based on histograms are shown in Fig.9.10 and Fig.9.10(b). Low vegetation samples have generally a lower backscattering value with respect to the tall vegetation, as can be expected, given the larger elements and the more important double bounce effect in arboreous vegetation. On the other side, the ensemble of the two vegetation classes, on average, shows backscattering values that fall in a range bounded by those of the buildings

Table 9.2. Statistics of the backscattering coefficient (dB) of a sample of pixels belonging to the asphalted surfaces in Fig.9.6.

ASPHALTED SURFACES						
Fig.	Type	Min	Max	Mean	St. dev.	N. pixels
a)	Bridge	-26.2	-9.3	-17.3	2.9	200
b)	Motorway	-22.3	-5.7	-13.5	2.6	200
c)	Toll gate	-26.0	-6.9	-18.8	2.9	200
d)	Roundabout	-26.3	-11.6	-17.7	2.9	200
e)	Parking area	-24.3	-6.7	-15.5	2.8	200
f)	Street	-23.3	-10.9	-17.6	2.5	200

class (upper bound) and the asphalted ones (lower bound). Among all vegetation classes, the *tree* class displays the highest amplitude value, since scattering from large elements in the crown can be rather significant. For both low and high vegetation the trend of the histograms has a behavior relatively close to that of the buildings, with a more regular distribution around the maximum value. However, at the same time, in the image under investigation the range of measures representing the motorway is very similar to those recorded over the dig in Fig.9.8(e) and the grasslands in Fig.9.8(d).

Table 9.3. Statistics of the backscattering coefficient (dB) of a sample of pixels belonging to the low and scarcely vegetated surfaces in Fig.9.8.

LOW VEGETATED SURFACES and BARE SOIL						
Fig.	Type	Min	Max	Mean	St. dev.	N. pixels
a)	Field	-22.9	-4.1	-12.9	2.6	1000
b)	Grass	-19.4	-2.6	-9.7	2.5	1000
c)	Round garden	-21.5	-4.5	-12.0	2.9	1000
d)	Grass	-23.0	-6.1	-14.0	2.8	1000
e)	Dig	-25.9	-5.5	-13.5	2.8	1000

Table 9.4. Statistics of the backscattering coefficient (dB) of a sample of pixels belonging to the high vegetated surfaces in Fig.9.9.

HIGH VEGETATED SURFACES						
Fig.	Type	Min	Max	Mean	St. dev.	N. pixels
a)	Field	-21.0	-0.4	-9.2	3.0	2000
b)	Grove	-22.6	13.9	-9.7	3.8	2000
c)	Field	-23.0	-2.2	-11.2	2.8	2000
d)	Row of trees	-26.4	5.7	-9.1	3.7	2000

The preliminary analysis of the land cover has put in evidence the presence of a wide variety of surfaces, even within the same class (manmade structures, asphalt, natural), whose discrimination, by means of the exploitation of only the intensity of the backscattering, may be a hard task. To correctly map COSMO-SkyMed imagery, is therefore necessary to identify and to extract further suitable information from the data.
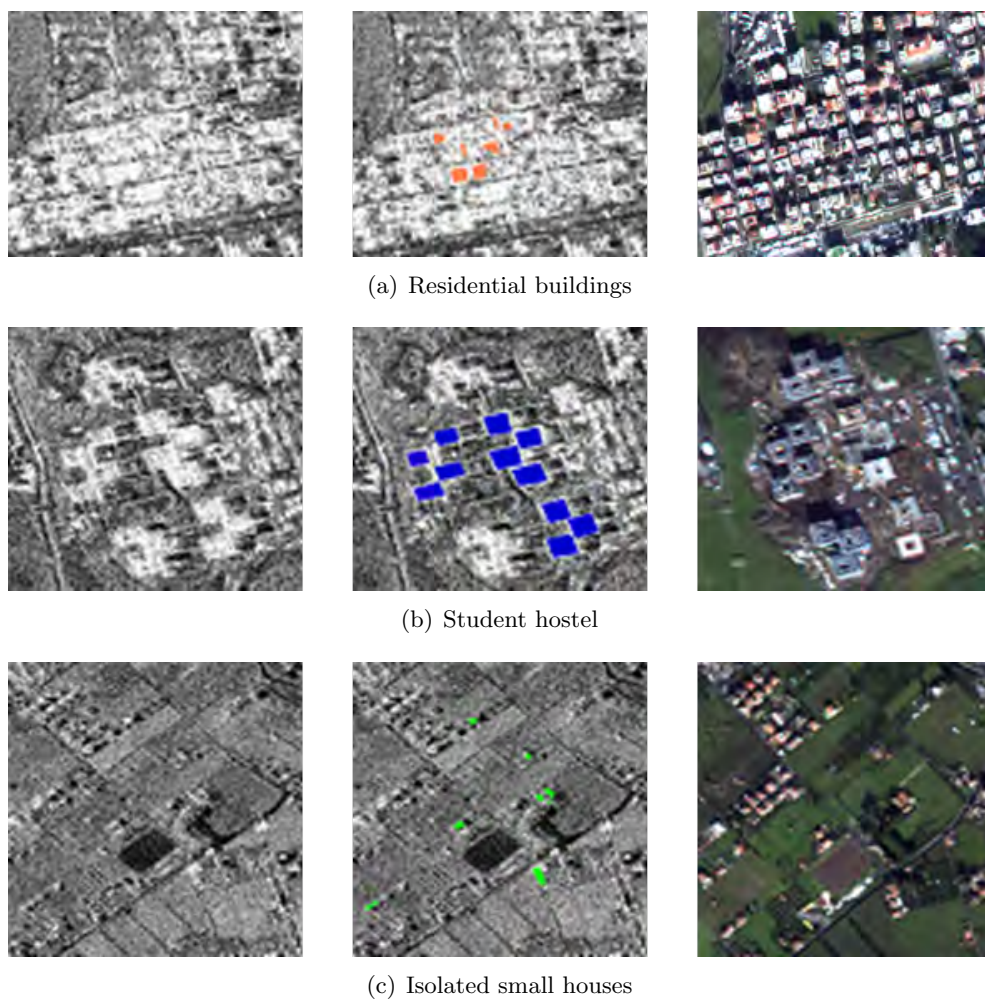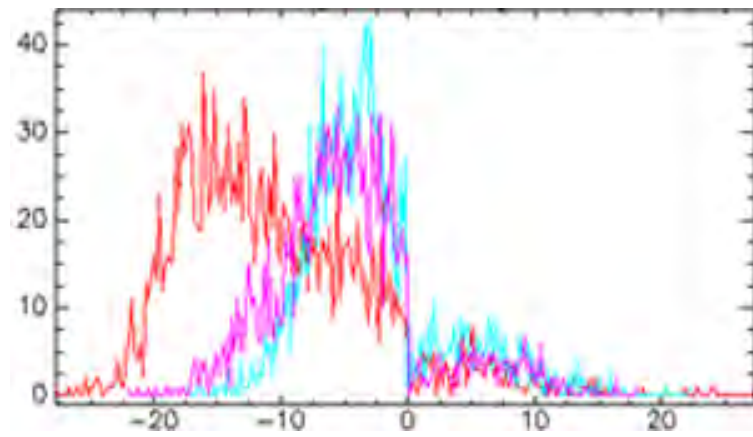
(a) Hospital

(b) Company building

(c) Mall
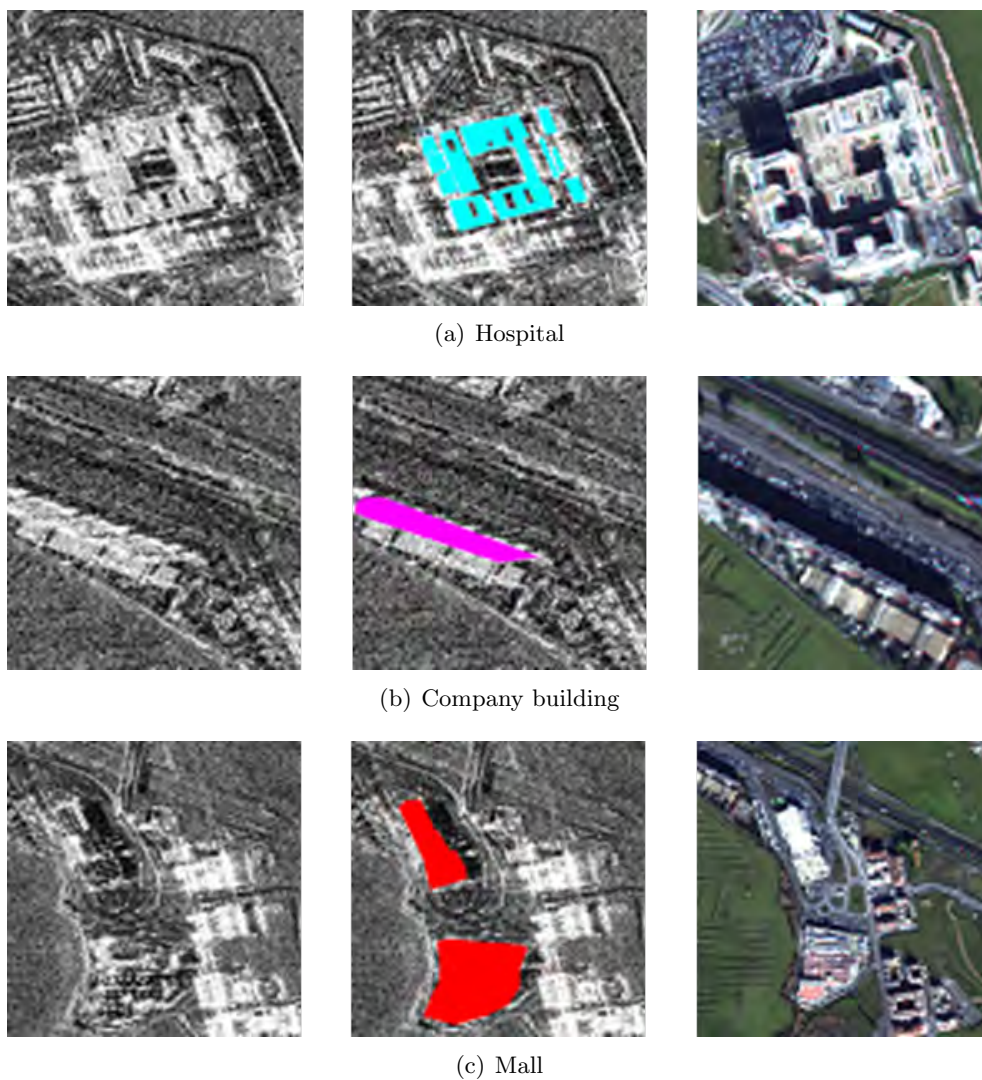
Figure 9.2. Examples of large buildings in a Spotlight CSK image (left) ©ASI, in a WV-2 (right) image ©DigitalGlobe, and ROI selected over the CSK image (middle).
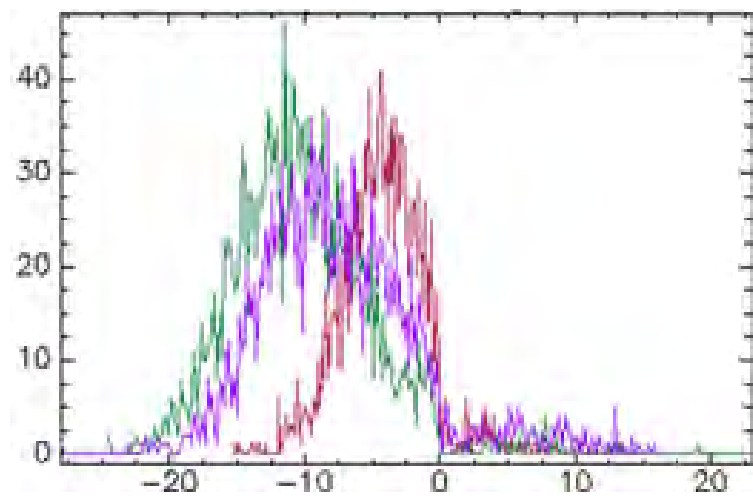
(a) Industrial sheds

(b) Bus sheds
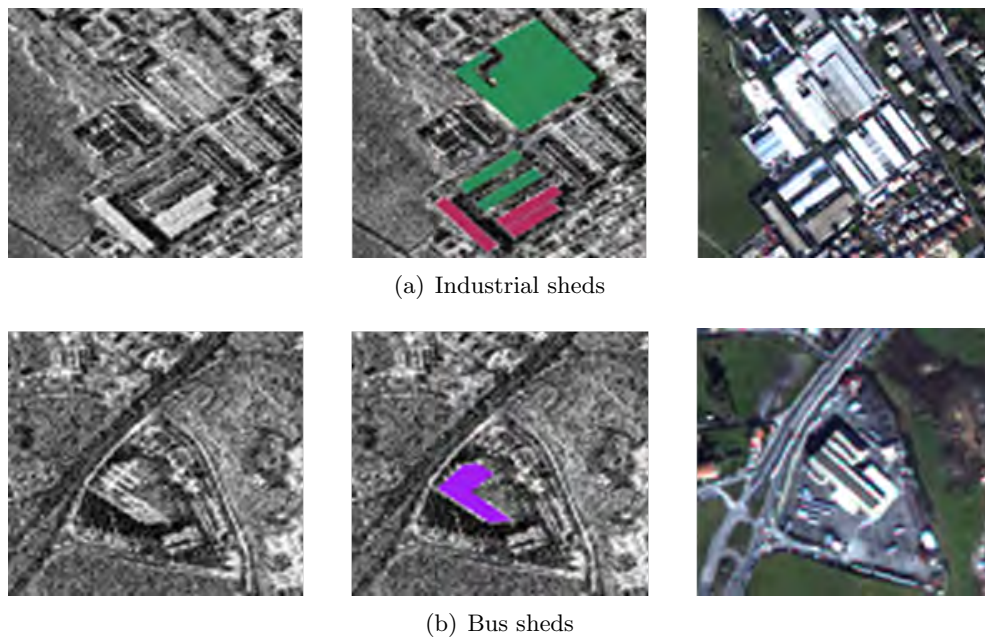
Figure 9.3. Examples of sheds in a Spotlight CSK image (left) ©ASI, in a WV-2 (right) image ©DigitalGlobe, and ROI selected over the COSMO-SkyMed acquisition (middle).

(a) Residential buildings
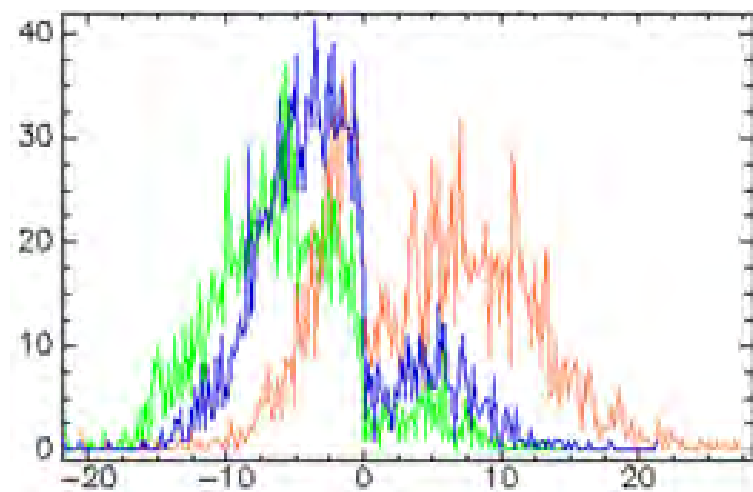

(b) Student hostel

(c) Isolated small houses

Figure 9.4. Examples of different residential buildings in a Spotlight CSK image (left) ©ASI, in a WV-2 (right) image ©DigitalGlobe, and ROI selected over the CSK acquisition (middle).
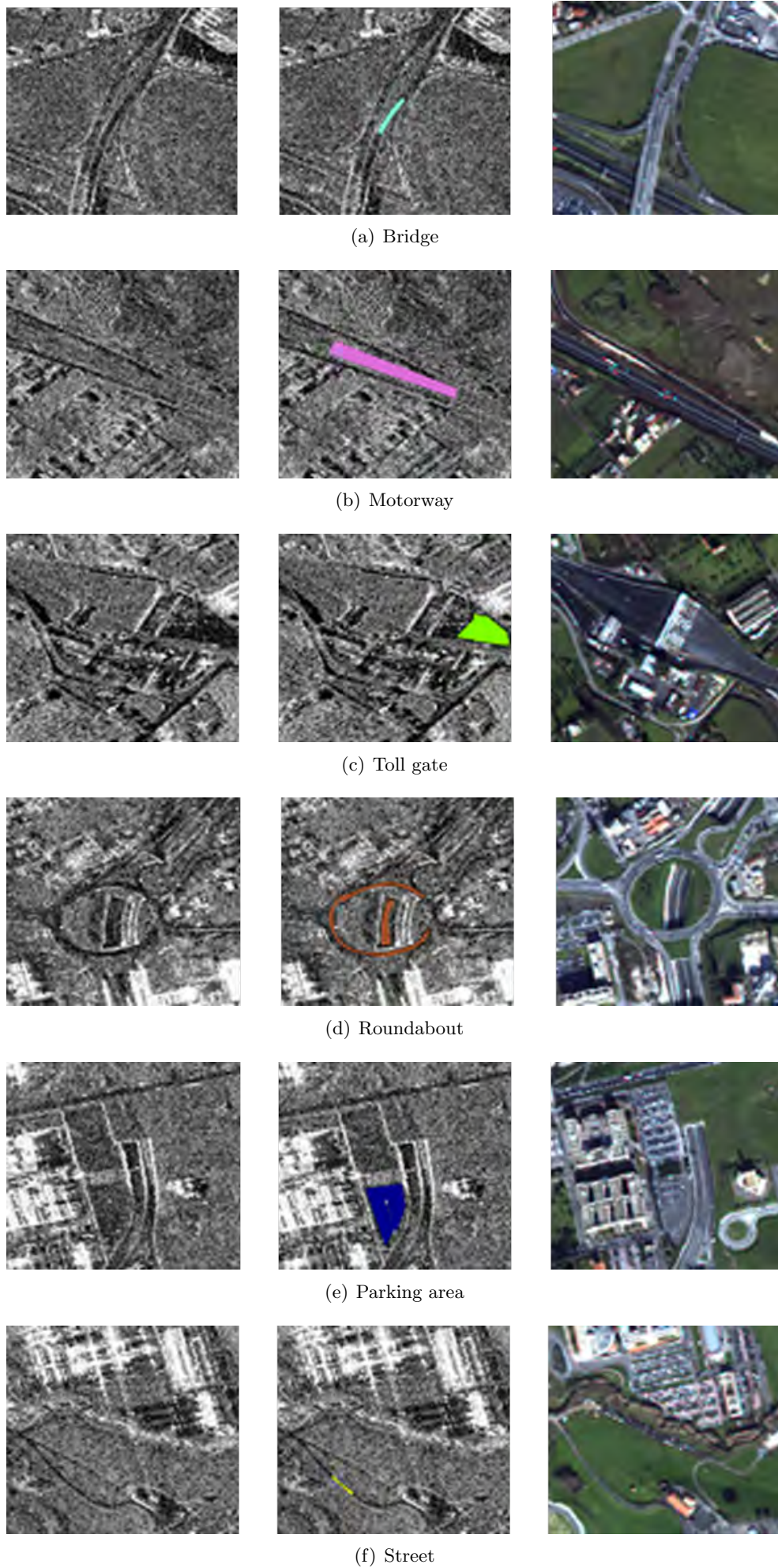
Figure 9.5. Histograms of the backscattering coefficient of 2000 random pixels belonging to the buildings selected respectively in: a) Fig.9.2: hospital (cyan), company building (magenta), shopping mall (red); b): Fig.9.3: industrial 1 (maroon), industrial 2 (dark green), bus shed (purple); c): Fig.9.4: residential (orange), hostel (blue), houses (green).

(a) Bridge

(b) Motorway

(c) Toll gate

(d) Roundabout

(e) Parking area

(f) Street

Figure 9.6. Examples of different asphalted surfaces in a Spotlight CSK image (left) ©ASI, in a WV-2 (right) image ©DigitalGlobe, and ROI selected over the CSK acquisition (middle).

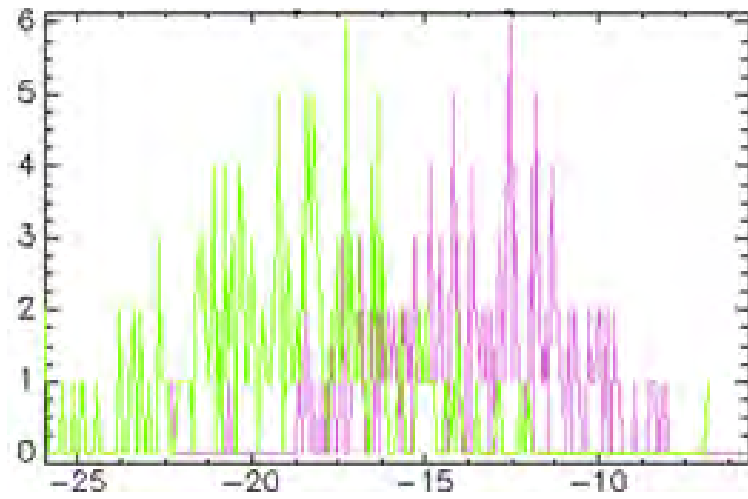**Figure 9.7.** Histograms of the backscattering coefficient of 200 random pixels belonging to the asphalted surfaces in Fig.9.6: a) motorway (magenta) and toll gate (green); b) motorway (magenta) and roundabout (maroon); c) parking area (blue) and street (yellow).

(a) Field

(b) Grass

(c) round garden

(d) Grass

(e) Dig

Figure 9.8. Examples of different low vegetated areas in a Spotlight CSK image (left) ©ASI, in a WV-2 (right) image ©DigitalGlobe, and ROI selected over the CSK acquisition (middle).

(a) Field

(b) Grove
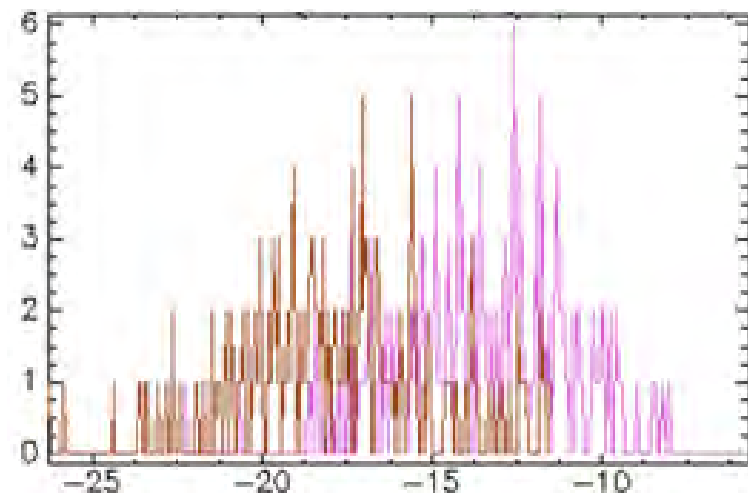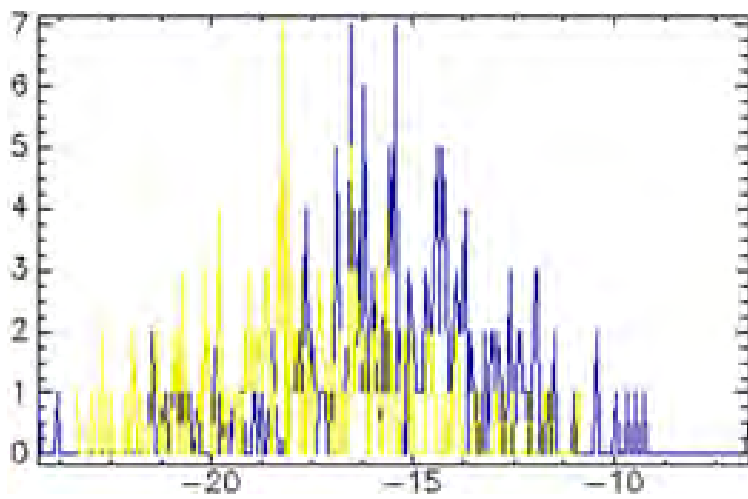

(c) Field

(d) Row of trees
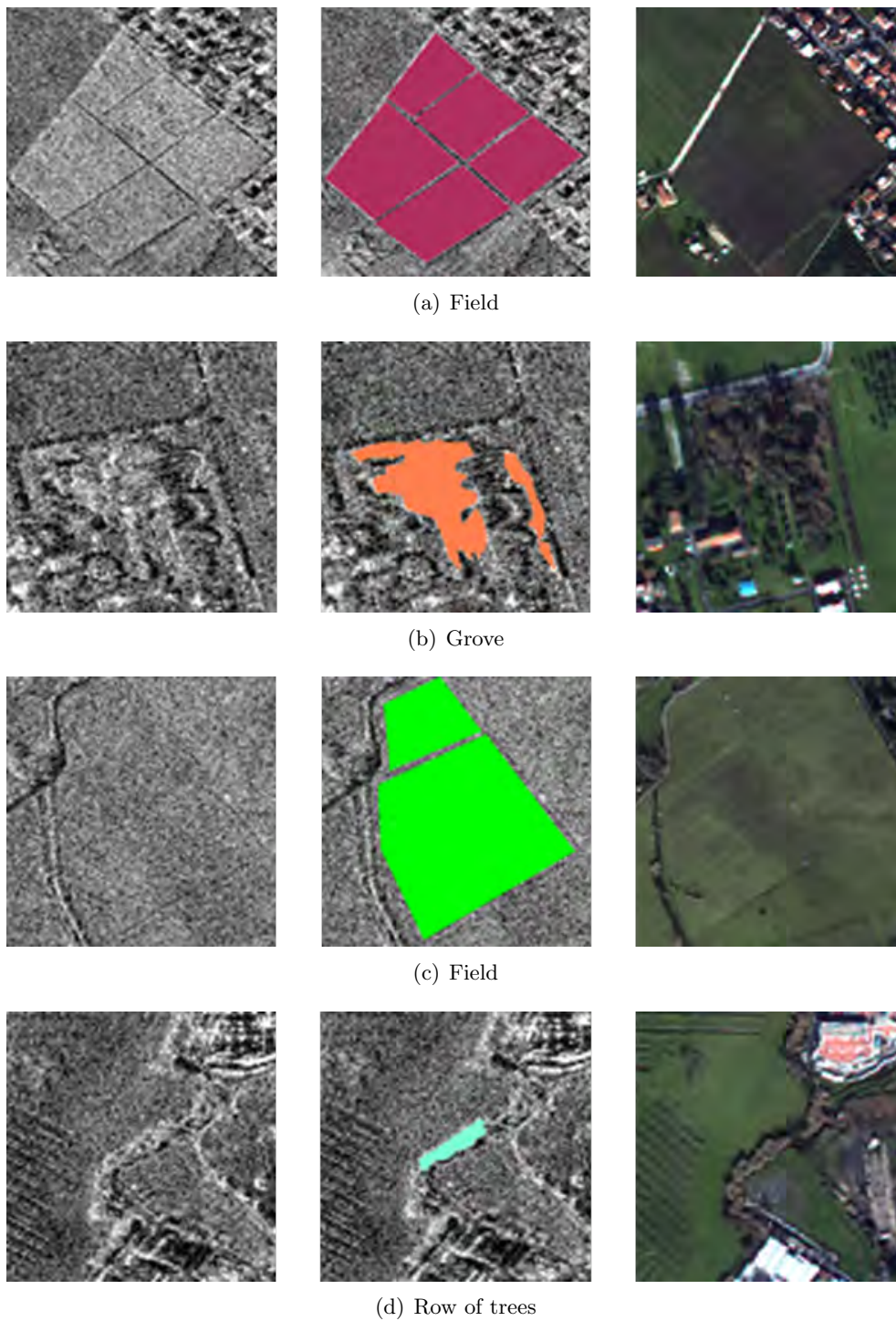
Figure 9.9. Examples of different high vegetated areas in a Spotlight CSK image (left) ©ASI, in a WV-2 (right) image ©DigitalGlobe, and ROI selected over the CSK acquisition (middle).
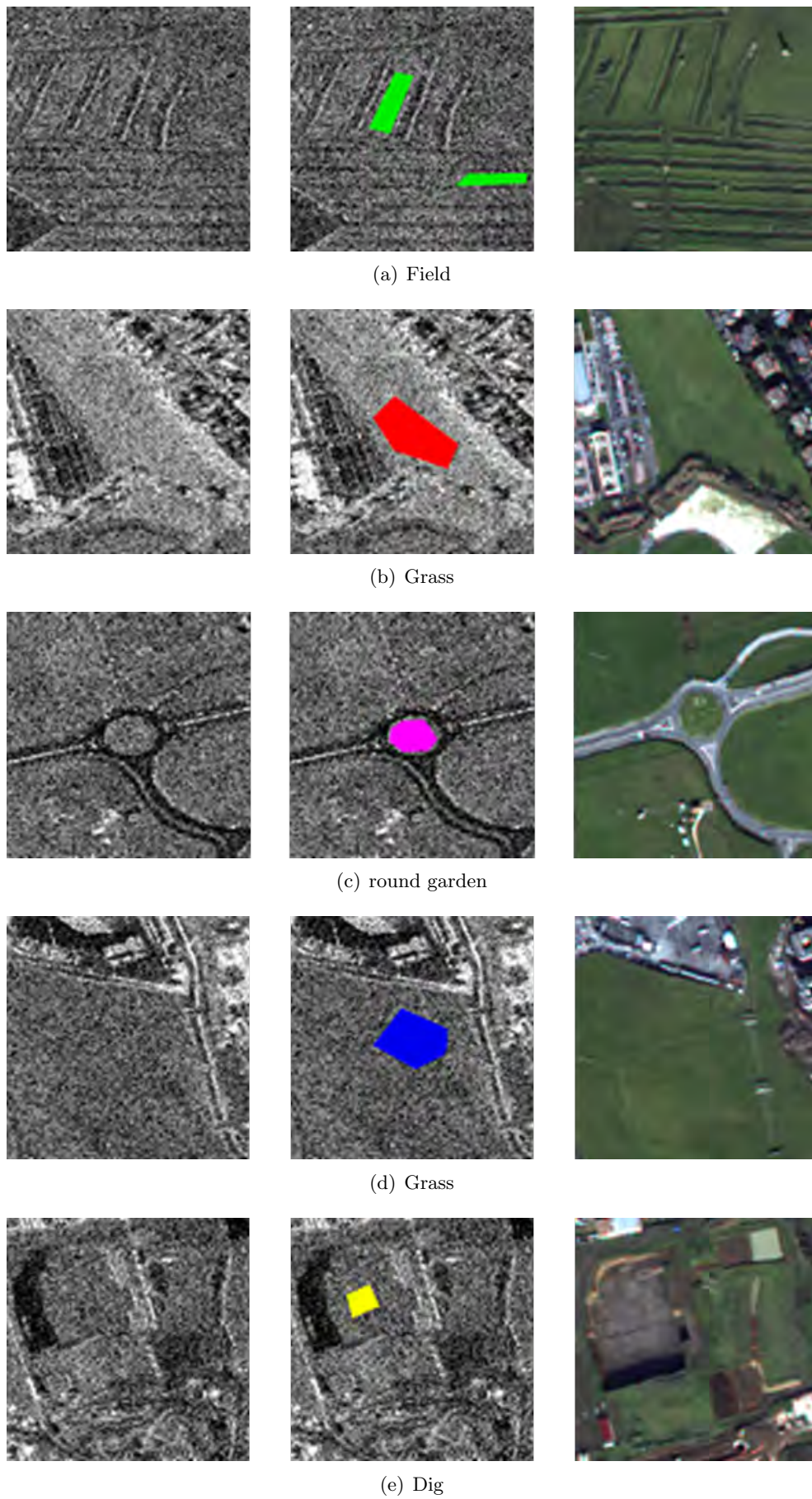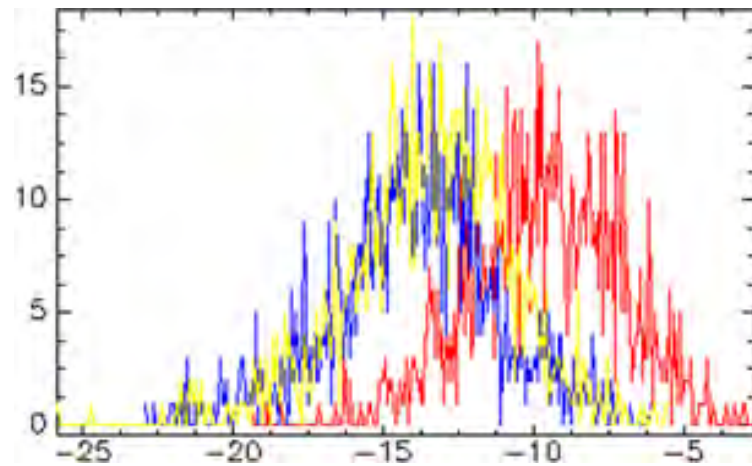
(a)
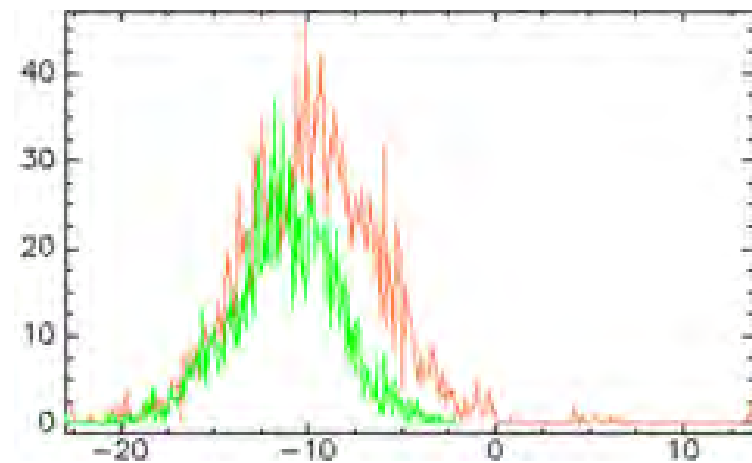
(b)

Figure 9.10. Histogram of the backscattering coefficient of: a) 1000 random pixels belonging to two different low-vegetated areas (red and blue) and to bare soil (yellow) in Fig.9.8(b), Fig.9.8(d) and Fig.9.8(e); b) 2000 random pixels collected over a region covered by trees (orange) and a cultivated field (green) in Fig.9.9(b) and Fig.9.9(c).

Chapter 10

Land cover classification

This chapter addresses the land cover pixel-based classification task, which has been performed by exploiting different type of information embedded within the VHR COSMO-SkyMed images. The identification of the most suitable input features is not a trivial issue. As previously mentioned, textural and contextual parameters play an important role for the interpretation of single polarized SAR images. Hence, they have been considered for the classification purpose. A first experiment, based on the exploitation of local texture features evaluated on sliding windows of pixels, will be described in §10.1. Secondly, in §10.2 an investigation on the possibility of improving the quality of the land cover mapping, especially in terms of number of classes, will be carried out by taking into account textural parameters stemmed from the GLCM computation. Finally, in §10.3 it will be shown how the interferometric coherence retrieved from pairs of images, can positively contribute to the improvement of the identification of some classes. In each case it will be demonstrated the efficiency of the neural network approach to correctly and automatically classify new images, which have not been used during the training samples selection.

10.1 Exploitation of simple local texture

The first classification exercise has been driven by the attempt to develop an efficient and fast algorithm, to quickly provide a response to particular events that may rapidly change the land cover of an urban area. Therefore, a limited number of data and information have been tested. Specifically, the method followed in this study considers partially inde-

---

Part of this chapter's contents is extracted from:

1. F. Del Frate, D. Loschiavo, C. Pratola, G. Schiavon and D. Solimini,*Sub-urban landscape characterization by very-high resolution X-band COSMO-SkyMed SAR images: first results*, SPIE Remote Sensing Event, 2010, Toulouse, France.

2. C. Pratola, F. Del Frate, G. Schiavon, D. Solimini and G. A. Licciardi,*Characterizing land cover from X-band COSMO-SkyMed images by neural networks*, Urban Remote Sensing Joint Event, JURSE 2011, Munich, Germany.

3. F. Del Frate, C. Pratola, G. Schiavon and D. Solimini, *Automatic features extraction in sub-urban landscape using very high resolution COSMO-SkyMed SAR images*, Proc. IEEE, International Geoscience and Remote Sensing Symposium , IGARSS 2011, pp.3614-3617, 2011.

4. C. Pratola, F. Del Frate, G. Schiavon and D. Solimini, *Automatic monitoring of land cover in sub-urban areas by X-band VHR COSMO-SkyMed imagery*, Proc. 17th Ka and Broadband Communications Navigation and Earth Observation Conference, 2011, Palermo, Italy.

:

pendent sources of information carried by a single SAR image. Besides the backscattering amplitude, two local contextual properties have been included into the input vector. The first feature has been evaluated by averaging the backscattering amplitude in sliding boxes, whose dimension has been set according with the spatial resolution of the CSK images: 5x5 pixels for the Spotlight images, and 3x3 pixels for the Stripmap acquisitions. The second contextual parameter, that is the standard deviation, has been calculated on the same moving windows. Mean and standard deviation values have then been assigned to the central element of each box.

The classification exercise consisted in mapping asphalted surfaces, natural areas (vegetation and bare soils) and manmade structures (buildings and other artificial elements, such as lamps and guard-rail along the streets). As it has been shown in Chapter 9, a large variability of surfaces within the same class is likely to occur. However, it can be assumed that asphalted areas are characterized by low backscattering and high contextual homogeneity. Hence low backscattering amplitude, *mean*, and *standard deviation* values characterize asphalted surfaces. On the other side, higher variations of local amplitude are expected over natural areas, whereas pixels with high backscattering are generally associated with buildings and metallic objects. Nevertheless, manmade constructions imaged by the VHR SAR system, result from multiple contributions related to the elements belonging to the building itself, and to different objects (e.g. trees, other manmade structures) in its immediate surrounding. In addition, the presence of layover, double bounce effect and shadow, make largely variable the backscattering values associated to the buildings.

Results

Single Spotlight image classification

Fig.10.1(a) shows the optical image of the test area seen by the WorldView-2 VHR instrument, while Fig.10.1(b) report the Spotlight COSMO-SkyMed acquisition taken on 8th of June 2010 [69]. An overall number of 735 pixels has been considered for the training set, while 315 examples have been collected for the validation set. The optimal neural network topology chosen for the exercise consisted of two hidden layers of eight neurons each (3x8x8x3). The training phase took only a few minutes to reach the early stopping condition that occurred after 600 epochs.

As it can be observed in Fig.10.1(c), the classification result is fairly correlated with the ground-truth. The confusion matrix in Table 10.1, evaluated considering a random sample of 861 ground truth points, shows an overall accuracy of 91.3%. In particular, the main roads and traffic circles are quite well delineated. However, some portions of the streets are confused with vegetated surfaces, probably because of the presence of trees and bushes along the roads, or of a particularly rough surface. Moreover, where the bare soil dominates the dry natural surfaces with respect to vegetation, the classifier erroneously assigned the pixels to the *asphalt* class.

Table 10.1. Confusion matrix referred to the classification map in Fig.10.1(c).

CLASS	A	N	MM	TOT
A	79.05	5.04	0.00	21.95
N	20.95	94.08	2.05	55.40
MM	0.00	0.88	97.95	22.65
TOT	100.00	100.00	100.00	100.00

Overall Accuracy: 91.30%

K coefficient: 0.86

(a)

(b)
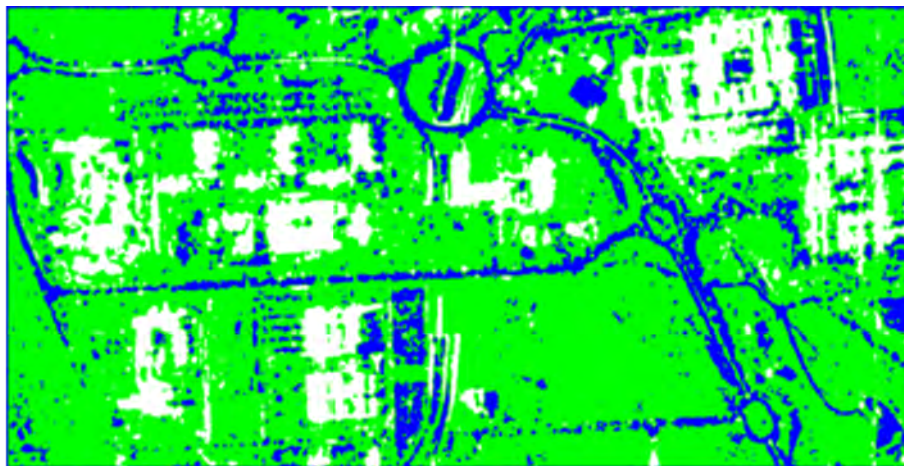
(c)

Figure 10.1. Tor Vergata area, Rome (Italy) imaged by a)WorldView-2 sensor ©DigitalGlobe (acquisition date: 4th July 2010) and by b) COSMO-SkyMed operating in Spotlight mode (994x506 pixels), in HH polarization, ©ASI (acquisition date: 8th June 2010). c) Classification map achieved by using a 3x8x8x3 MLP-NN.

Single Spotlight image fully automatic classification

To test a fully automatic scheme that does not need the selection of pixels extracted from the image to be classified, a set of 12600 training samples and 5400 validation pixels have been randomly selected in two Spotlight images (4230x2500 pixels) acquired on 8th June 2010 and 9th July 2010, respectively [70]. It is worthy to highlight the fact that the latter image does not differ significantly from the former, and from a third Spotlight acquisition (Fig.10.2(b)) taken by COSMO-SkyMed on 10th July 2010, which has been used only for evaluating the features extraction result. Note that all the images were acquired with the same geometry (Chapter 7) and in the same season. The last aspect should imply similar distributions of the backscattering values, which is a main requirement for the automatic classification of different imagery by using trained neural networks. With respect to the previous classification exercise, in this case a larger area (4230x2500 pixels) has been considered, aiming at the investigation on the performance of the algorithm in quickly and automatically providing maps of even more complex environments, characterized by a larger variety of surfaces. As for the first experiment, backscattering amplitude, local mean and standard deviation have been included into the input vectors in order to identify three main classes: *asphalt* (A), *natural* surfaces (N) and *manmade* structures (MM).

A two-hidden layer NN scheme, each one with 9 neurons (3x9x9x3), has proved to be the best topology in terms of MSE minimization. It has to be highlighted the fact that, although a huge quantity of data has been elaborated, once properly trained, the NN has been able to produce in only few seconds and automatically the map shown in Fig.10.2(c). By the visual interpretation of the result, it is possible to assert that the spatial distribution of the different types of pixels is highly correlated with the ground-truth represented by the multispectral WorldView-2 image in Fig.10.2(a). Such statement is confirmed by the obtained overall accuracy of 80.9%, as computed by a comparison with a randomly selected sample of 6000 ground-truth pixels. In particular, the delineation of most of the main roads and roundabouts seems to be rather correctly achieved. However, some confusion can be again observed between asphalted surfaces and natural areas, but this is explained with the presence, within the latter class, of sparsely or no vegetated regions, the fraction of which prevails over that of densely vegetated areas. Moreover, trees are sometimes interpreted as manmade structures and vice-versa. The detail of the confusion matrix is given in Table 10.2.

Table 10.2. Confusion matrix referred to the classification map in Fig.10.2(c).

CLASS	A	N	MM	TOT
A	64.80	7.00	0.10	23.90
N	34.40	86.50	8.40	43.10
MM	0.80	6.50	91.50	33.00
TOT	100.00	100.00	100.00	100.00

Overall Accuracy: 80.90%

K coefficient: 0.71

(a)

(b)

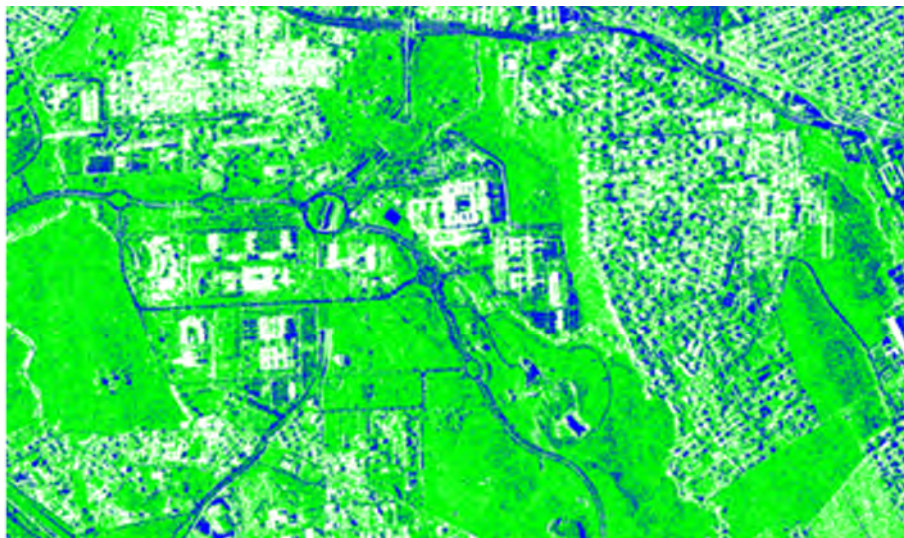Asphalt (A)   Natural (N)   Manmade (MM)

(c)

Figure 10.2. Tor Vergata area, Rome (Italy) imaged by a) WorldView-2 sensor ©DigitalGlobe (acquisition date: 4th July 2010) and by b)COSMO-SkyMed operating in Spotlight mode (4230x2500 pixels), in HH polarization ©ASI (acquisition date: 10th July 2010). c)Classification map achieved automatically by means of the previously trained 3x9x9x3 MLP-NN.

Single Stripmap image classification

The approach adopted for the Spotlight images has been tested also on the Stripmap acquisition in Fig.10.3 taken by COSMO-SkyMed on 24th of May 2010. As in the previous exercise, the three elements input vector included: the amplitude of the backscattering, its local mean and local standard deviation computed on sliding windows whose size, given the lower spatial resolution (3 meters), has been set to 3x3 pixels. The algorithm has been trained by 735 input patterns, while the performance of the learning phase has been assessed through 315 validation samples. The optimal neural network architecture has been found to be made of two hidden layers of eight neurons each (3x8x8x3). The algorithm has been subsequently applied to the whole image, which has been processed in few seconds, providing the map in Fig.10.4(b). A visual interpretation of the result put in evidence the capability of the NN to identify correctly the manmade structures and the majority of the streets. Yet, most of the wide vegetated areas turns out to be erroneously covered by little objects classified as asphalt. The reason might be due to the presence of small patches of bare soil, or shadow in closeness of trees, as well as to the effect of residual speckle. The qualitative evaluation of the classification has been confirmed by the computation of the confusion matrix in Table 10.3, obtained by a cross-check with 744 pixels of ground truth randomly collected in each land cover, and interpreted by means of the WorldView-2 image in Fig.10.4(a).

Figure 10.3. Tor Vergata, Rome (Italy) imaged by COSMO-SkyMed operating in Stripmap mode (836x520 pixels), in HH polarization ©ASI (acquisition date: 24th May 2010).

Table 10.3. Confusion matrix referred to the classification map in Fig. 10.4(b).

CLASS	A	N	MM	TOT
A	90.00	14.16	0.00	31.85
N	10.00	80.53	0.00	39.52
MM	0.00	5.31	100.00	28.63
TOT	100.00	100.00	100.00	100.00
Overall Accuracy: 88.31%				
K coefficient: 0.82				

(a)

(b)
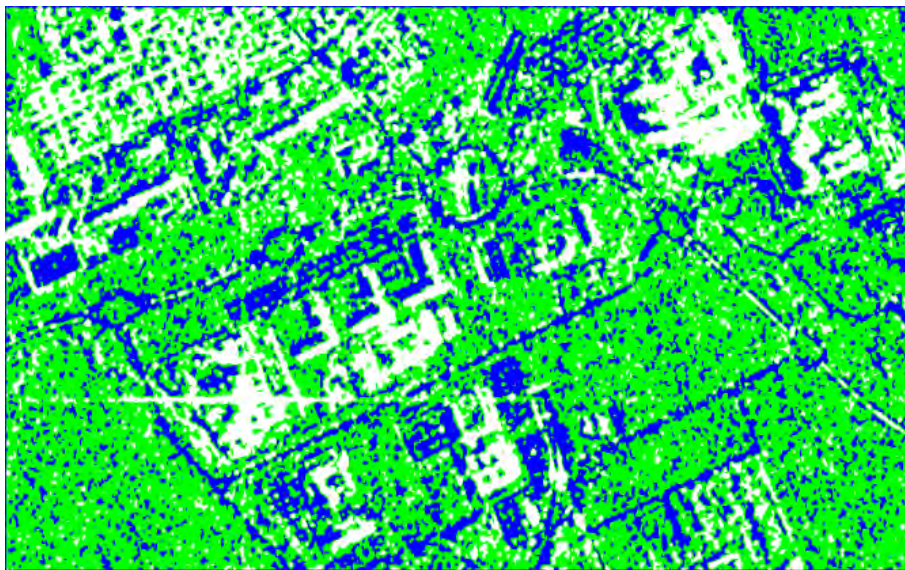
■ Asphalt (A)   ■ Natural (N)   □ Manmade (MM)

Figure 10.4. Tor Vergata area, Rome (Italy) imaged by a) WorldView-2 sensor ©DigitalGlobe (acquisition date: 4th July 2010). b) Classification map of the image in Fig.10.3, achieved by means of a 3x8x8x3 MLP-NN.

10.2   Exploitation of GLCM textural parameters

Given the promising results achieved in basic land covers discrimination by exploiting single polarized images and simple local texture, a further effort has been made to improve the classification in terms of number of classes. By considering the Tor Vergata landscape, the *natural* class has been divided into *low vegetation* (LV) class, including sparse and low vegetated surfaces, as well as bare soil, and *trees* (T) class, which includes tall plants and dense vegetated areas. The issue has been addressed by carrying out an analysis to investigate on the possibility of including more complex textural parameters into the input vector of the MLP-NN algorithm. Specifically, the study has been focused on the selection of the optimum set of textural features stemmed from the computation of the GLCM.

Such approach requires a proper setting of several key parameters on which the derived indexes depend: the distance between two positions, the distance direction, and the window considered for the GLCM computation.

The distance may provide a tool to discriminate among textures through the spacing characteristic. Although in coarse spatial resolution SAR images a very small distance can be sufficient to differentiate various surfaces, a larger parameter should be taken into account for the classification of VHR imagery.

Direction is particularly important in case of evident anisotropy in the texture of different land covers [71]. However, when no preferred direction can be observed, a common approach is the calculation of the mean values of the textural parameters estimated in each direction.

On the other side, the window size plays a main role in the land cover discrimination capability. Indeed, it defines the area around a pixel within which it is assumed that texture patterns are statistically steady. Several authors in literature addressed the problem of the selection of the best dimension by adopting a trial-and-error method. Most of them conclude that larger textural windows lead to lower classification accuracies. However, aiming at an empirical definition of the relation between window size and classification accuracy of VHR panchromatic images of urban environments, in [72] it has been demonstrated the possibility to obtain high classification performances with very wide-area textural windows, which allow for the detection of even small textural patches in the image. As a matter of fact, the box dimension is connected to the mean size of textured areas, such as settlements blocks in urban environments [73]. Nevertheless, the block size is usually unknown, and likely different, for large towns, in different parts of the same urban area. Thus, in principle, the best and most commonly adopted approach is a trial-and-error method. Alternatively multi-scale analysis, as those suggested in [73] and [74], may be valuable for the classification of a sub-urban environment.

Assuming that not all the aforementioned eight textural measures (see §5.2) are effective to discriminate the land cover surfaces, in this work it has been studied the behavior of 5000 pixels collected for each class in the Spotlight image in Fig.9.1. The graphs in Fig.10.5 are referred to the textural parameters evaluated over windows of five different sizes (7x7, 9x9, 11x11, 13x13 and 15x15), chosen on the basis of the spatial resolution and the dimension of the main objects within the scene. Each texture image has been obtained by averaging the pixel values computed considering the 0°, 45° and 90° directions, with a spatial shift equal to 15. As it can be observed in Fig.10.5, the general trend of the mean values of the parameters stemming from the GLCM computation does not change significantly with varying dimension of the window. However, in all the cases, only four out of eight measures appear to be relevant for discriminating the different land covers. Therefore, in addition to the backscattering intensity, also the *mean*, *variance*, *contrast* and *correlation* features, have been identified as the set of most suitable input information for the NN training.

Even if it has been demonstrated that larger windows may lead to more stable texture
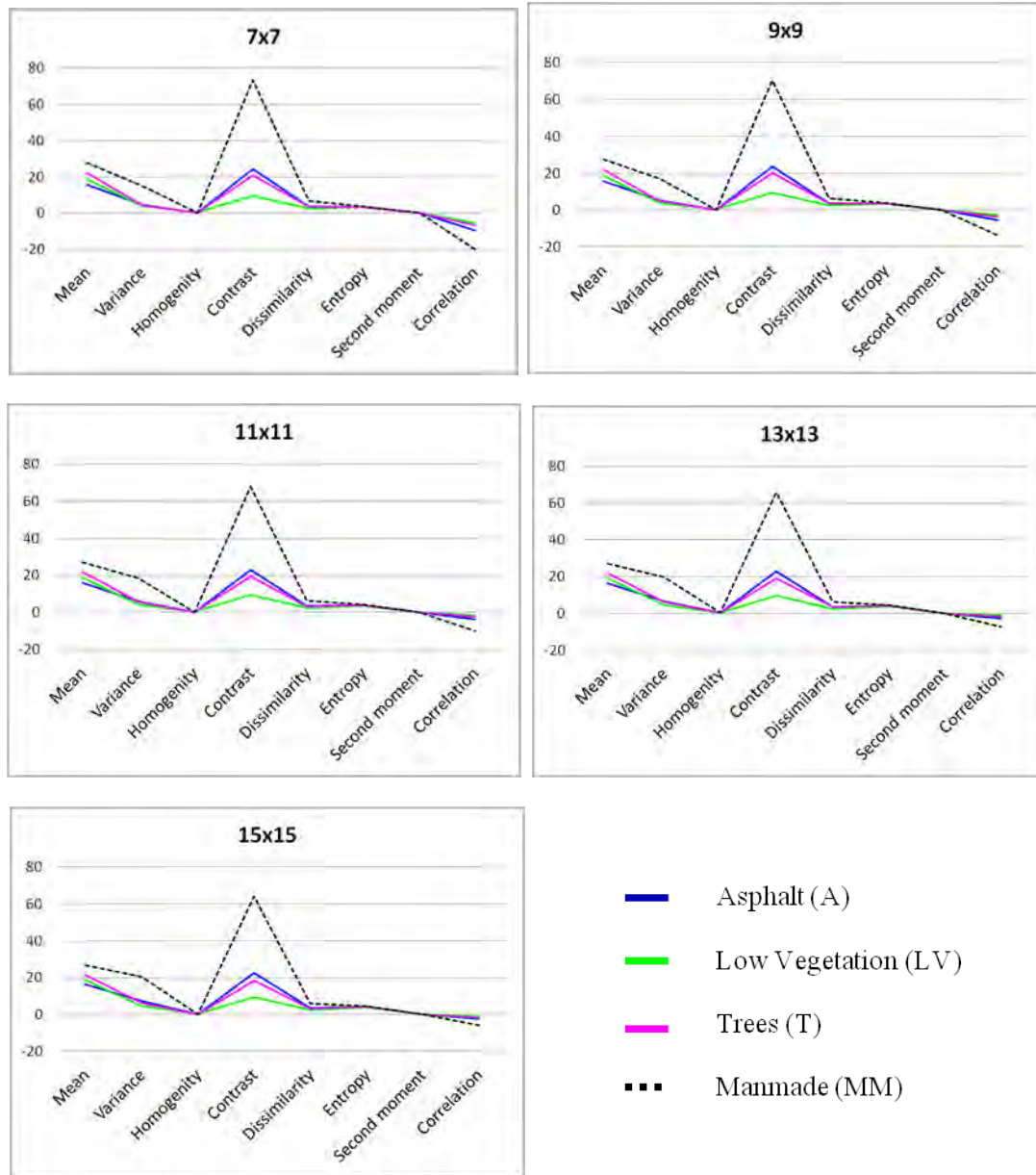
Figure 10.5. Textural parameters mean values evaluated for 5000 pixels collected in each class. The GLCM has been computed by considering different window sizes: 7x7, 9x9, 11x11, 13x13, 15x15 pixels; the spatial distance has been fixed to 15; the textural features have been firstly estimated in three main directions, that are 0°, 45° and 90°, and then the average has been taken into account.

features, they can cause blurring of edges in VHR SAR images, while smaller cells can make the objects boundary wrong [74]. To get the appropriate window size, the Transformed Divergence (TD) [75], [76] has been computed for each pair of classes of interest, in order to figure out their degree of separability. The TD is defined as follows:

$$TD(i,j) = 2 \cdot \left[1 - \exp^{\frac{-D_{i,j}}{8}} \right] \tag{10.1}$$

where $D_{i,j}$ is the divergence between classes i and j. The higher $D_{i,j}$, the higher is the separability expressed by TD, the maximum value of which is 2. A pair of classes are considered well-distinguishable if the Transformed Divergence is larger than 1.9.
Table 10.4 and its corresponding plot in Fig.10.6 show a good separability for each pair of classes, even when, besides the intensity of the backscattering, the only four most significant GLCM parameters have been taken into account. In general the separability increases with the window size, with exceptions for the *asphalt-trees* pair, and for the *asphalt-manmade* one. The latter exhibits a slight decrease in the TD value, starting from the 13x13 dimension. Notwithstanding, the Transformed Divergence still remains higher than 1.9. On the other hand, the classes *asphalt* and *trees* are less easy to be distinguished, as highlighted by the graph in Fig.10.7.
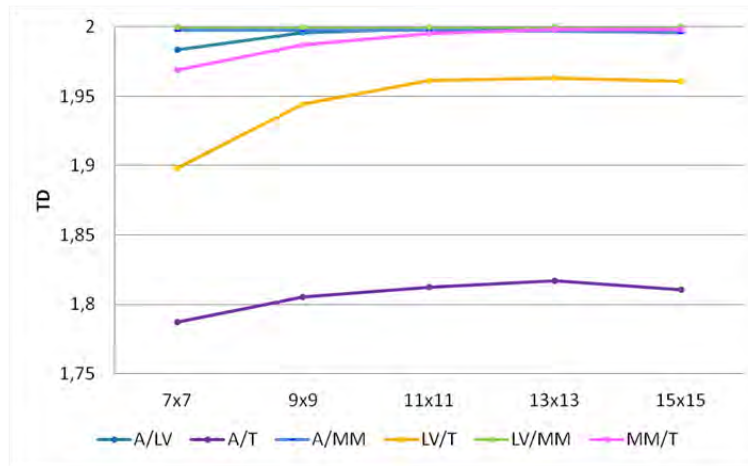
Figure 10.6. Class separability as a function of window size.

Table 10.4. Transformed Divergence evaluated for each pair of classes and different window sizes, based on Intensity, Mean, Variance, Contrast, Correlation pixel values.

	7x7	9x9	11x11	13x13	15x15
A/LV	1.9833	1.9959	1.9988	1.9996	1.9998
A/T	1.7875	1.8054	1.8124	1.8172	1.8105
A/T	1.7875	1.8054	1.8124	1.8172	1.8105
A/MM	1.9973	1.9975	1.9974	1.9966	1.9955
LV/T	1.8981	1.9445	1.9610	1.9631	1.9606
LV/MM	2.0000	2.0000	2.0000	2.0000	2.0000
MM/T	1.9691	1.9869	1.9949	1.9979	1.9983

Based on such analysis, the 13x13 window size has been finally selected for the classification of Spotlight imagery.
By adopting the same rationale, the textural analysis has been replicated for the Stripmap images. Consistent outcomes have been achieved in terms of suitable GLCM features and
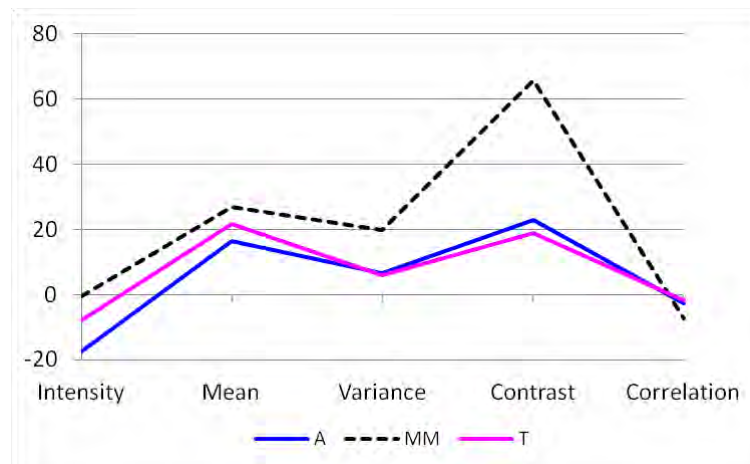
Figure 10.7. Mean values of the intensity of backscattering and textural parameters obtained from the GLCM computation (window size: 13x13 pixels; spatial distance: 15; direction: 0°, 45°, 90°(average of the estimated textural parameters).

relationship between window size and classes separability. Therefore, the same five parameters (intensity of backscattering, mean, variance, contrast and correlation) have been used to train the MLP-NN classifier. Given the coarser spatial resolution of the Stripmap acquisitions, the best set of textural parameters have been obtained by computing the GLCM on windows made of 11x11 pixels, and considering a distance equal to 13.

Results

Single Spotlight image classification

The couple of Spotlight data in Fig.10.8(a) and Fig.10.8(b) (9700x5300 pixels), acquired in a range time of about one year, have been considered for the classification exercise. In every image, 45825 training pixels and 17625 validation pixels have been randomly selected in correspondence of each class of interest, taking into account all the possible surfaces previously statistically analyzed. All these samples, carrying five sources of information (intensity of backscattering, mean, variance, contrast and correlation), have been used to train the net and to assess the efficiency of the learning. The topology of the neural network which better minimizes the MSE has been, for both the images, that one with two hidden layers containing 12 neurons each (5x12x12x4). The mapping results are shown in Fig.10.9(a) and Fig.10.9(b).

For the validation phase, the maps have been compared with the two optical images represented in Fig.10.10, acquired by WorldView-2 on 4th July 2010, and by QuickBird on 20th May 2011, respectively. A first qualitative interpretation of the results reveals a fairly good correlation with the ground truth; in particular, manmade structures are generally recognized with good accuracy, as well as the road network. As for the first classification experiments, also in these cases some confusion between A and LV classes occurs, basically due to the large percentage of bare soil whose behavior, in terms of intensity of backscattering and texture, can be very similar to that of parking lots. As the statistical analysis has pointed out, the correct classification of the motorway has been hard to achieve, therefore most of the correspondent pixels have been interpreted as vegetation. Moreover, the typical geometrical artifacts affecting the SAR images, especially layover and shadowing, can make the buildings recognition incorrect: flat roofs may be classified as asphalt, and parts of a single manmade structure may be interpreted as a tree. Such a confusion is much more evident in high-density urban areas, where the backscattered signal comes from a variety of elements very close to each other (buildings, other manmade structures, trees, roads, etc.).

As far as the differences between the two maps are concerned, the main and most visible one occurs in the percentage of pixels classified as asphalt, which is higher in the July image. The hypothesized reason of a larger percentage of bare soil or scarcely vegetated surfaces in the older image, and of the presence of taller lawn in the more recent acquisition, finds a possible confirm in the analysis of the false color composite optical images and their respective NDVI maps (Fig.10.10). In fact, the image acquired in 2011 exhibits wider patches characterized by high reflectivity in the NIR band (represented as bright red areas), and high NDVI index. However, somewhere the difference between the two acquisitions, in terms of NIR reflectivity and NDVI value, is not very strong. Yet, the classifier demonstrated to be able to distinguish the *low vegetation* from the *trees* class, since the SAR backscattering mostly depends on the geometry of the imaged targets.

Through the comparison with 28200 ground truth pixels randomly collected within each COSMO-SkyMed acquisition, the accuracy of the results has been evaluated equal to 83.34% for the 9th July 2010 acquisition, and to 84.92% for the 11th June 2011 image. The details of each confusion matrix are reported in Table 10.5 and Table 10.6.

(a)

(b)

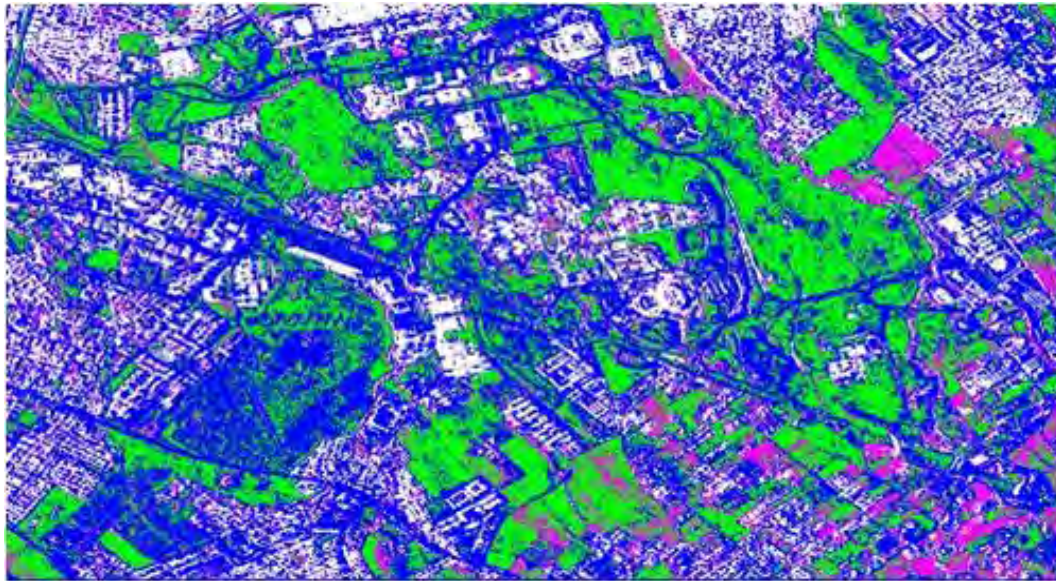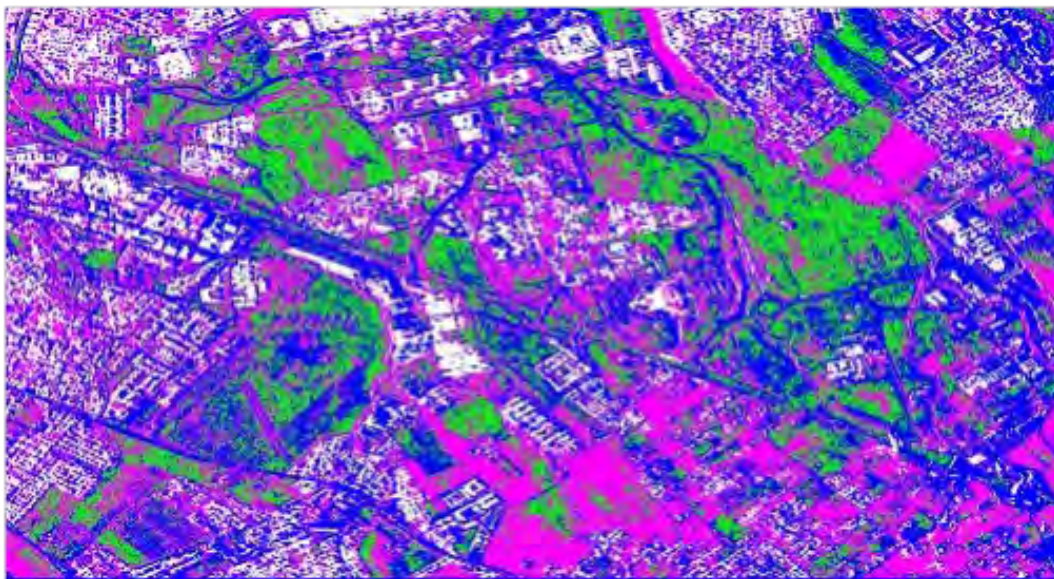**Figure 10.8.** Tor Vergata area, Rome (Italy) imaged by COSMO-SkyMed operating in Spotlight mode (9700x5300 pixels), in HH polarization ©ASI. a) Acquisition date: 9th July 2010. b) Acquisition date: 11th June 2011.

(a)

(b)

Figure 10.9. Classification results of the Spotlight COSMO-Skymed images (9700x5300 pixels) in a) Fig. 10.8(a) and b) Fig. 10.8(b), achieved by means of a 5x12x12x4 MLP-NN.

Table 10.5. Confusion matrix referred to the classification map in Fig. 10.9(a).

CLASS	A	LV	T	MM	TOT
A	90.16	23.25	2.37	1.30	29.27
LV	8.30	75.42	6.37	0.62	22.68
T	0.86	0.84	75.59	5.70	20.74
MM	0.68	0.49	15.67	92.38	27.31
TOT	100.00	100.00	100.00	100.00	100.00

Overall Accuracy: 83.34%

K coefficient: 0.78

Table 10.6. Confusion matrix referred to the classification map in Fig. 10.9(b).

CLASS	A	LV	T	MM	TOT
A	86.85	10.24	1.79	1.02	24.97
LV	9.11	70.62	1.11	0.18	20.25
T	3.91	19.00	87.00	3.59	28.38
MM	0.13	0.14	10.10	95.21	26.39
TOT	100.00	100.00	100.00	100.00	100.00

Overall Accuracy: 84.92%

K coefficient: 0.79

(a)

(b)

(c)

(d)
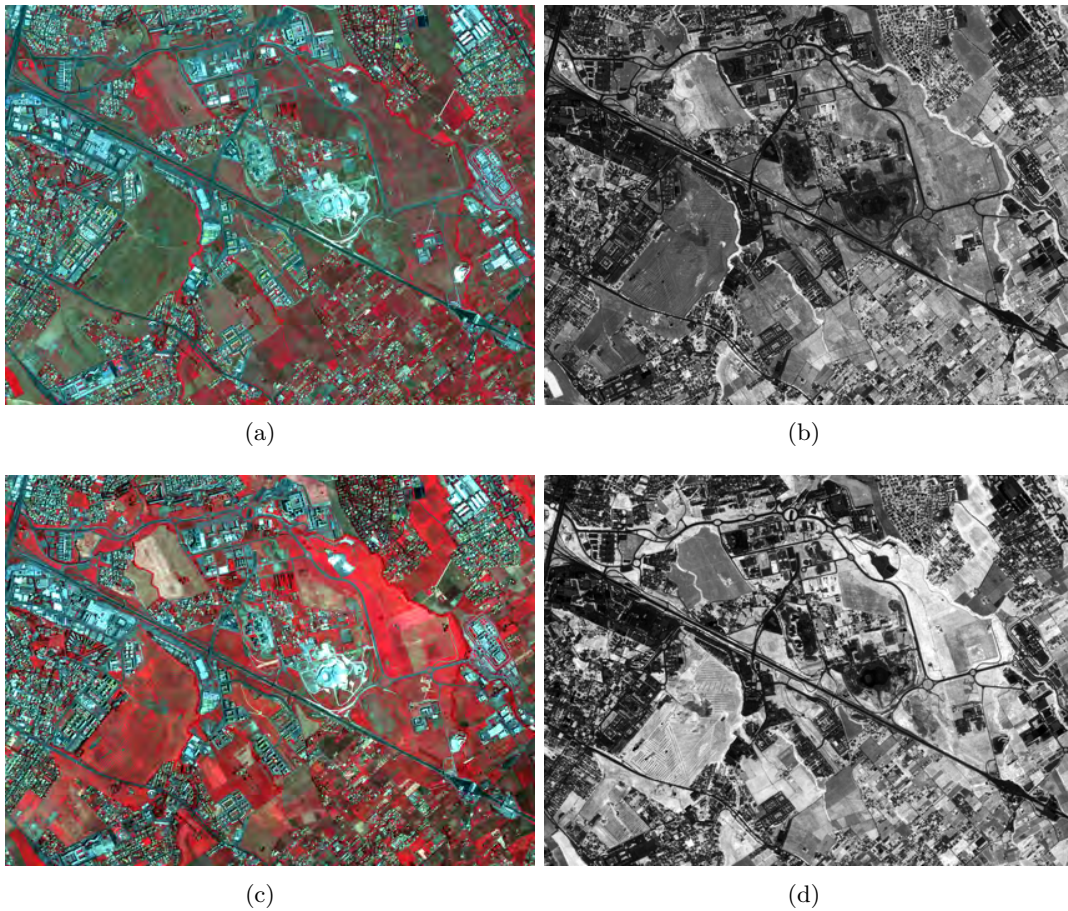
Figure 10.10. a) WorldView-2 image (4th July 2010) ©DigitalGlobe (R: NIR1, G: Red, B: Green) and b) its correspondent NDVI map. c) QuickBird image (11th June 2011) ©DigitalGlobe (R: NIR, G: Red, B: Green) and d) its correspondent NDVI map.

Single Stripmap image classification

By replicating the textural analysis previously discussed for the Spotlight images, it has been found out that the best class separability for the Stripmap acquisitions is obtained by combining the intensity of the backscattering with the mean, variance, contrast and correlation information obtained from the GLCM computation, on windows of 11x11 pixels and considering a spatial shift equal to 13. Hence, a 5-input neural network has been designed to identify four possible outputs. The best performance of the algorithm has been achieved by using 7200 training samples (and a set of 3200 validation pixels) to feed a 5x10x10x4 MLP-NN. The experiment has been performed on the 2800x1800 pixels Stripmap image in Fig.10.11. The confusion matrix in Table 10.7, which exhibits an overall accuracy of 92.45%, has been evaluated by means of 4000 ground truth pixels, which have been randomly selected and interpreted through a comparison with the QuickBird image in Fig.10.12(a). High performance of the algorithm occurs in mapping almost all the manmade structures and areas covered by scarce or low vegetation. The result shows again some confusion between asphalt surfaces and tree-covered areas, but these errors mainly occur along the tree-lined streets, where it could be very hard to distinguish the land cover pixel by pixel, so that the training set could include wrong examples.

Figure 10.11. Tor Vergata, Rome (Italy) imaged by COSMO-SkyMed operating in Stripmap mode (2800x1800 pixels), in HH polarization ©ASI (acquisition date: 20th September 2008)

Table 10.7. Confusion matrix referred to the classification map in Fig.10.12(b).

CLASS	A	LV	T	MM	TOT
A	87.60	0.80	1.70	0.20	22.58
LV	4.60	97.70	6.70	0.00	27.25
T	7.20	1.50	85.40	0.70	23.70
MM	0.60	0.00	6.20	99.10	26.47
TOT	100.00	100.00	100.00	100.00	100.00

Overall Accuracy: 92.45%

K coefficient: 0.89

(a)

(b)
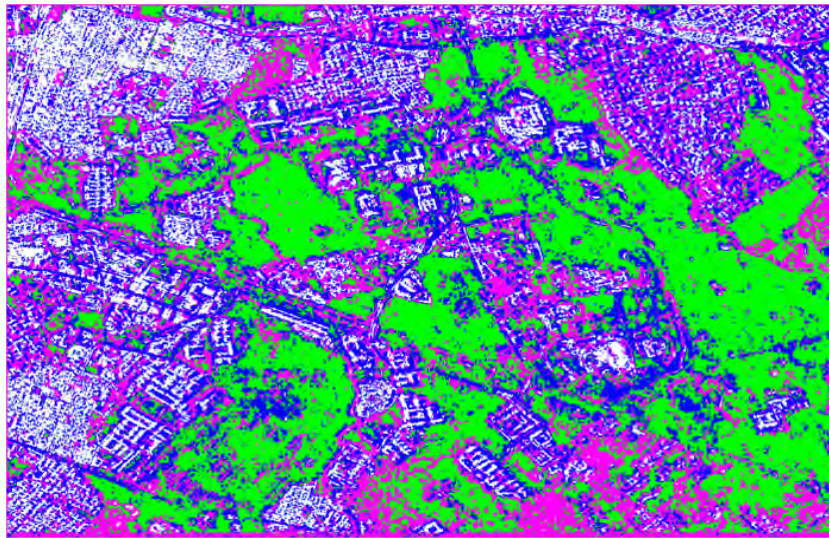
Figure 10.12. Tor Vergata area, Rome (Italy) imaged by a) QuickBird sensor ©DigitalGlobe (acquisition date: 8th December 2008). b) Classification map of the image in Fig.10.11 achieved by means of a 5x10x10x4 MLP-NN.

Algorithm robustness tests

The use of textural information derived from the GLCM computation proved to be effective in improving the quality of the classification maps. However, as previously highlighted, the ability of the MLP-NN classifier mostly depends on the training set, in terms of size and representativeness. In order to test the robustness of the algorithm, two different exercises have been carried out. In the first case, the image has been classified by using MLP-NN trained by an increasing number of samples, and the accuracies of each result have been evaluated. The goal of the second test has been to verify the efficiency of the algorithm in providing maps of a different site, specifically the Greater Denver area, Colorado, USA.

Robustness test on the training set size

As previously asserted, in order to make the net able to correctly distinguish the classes of interest, it is of crucial importance the exploitation of statistically representative training patterns. Aiming at the assessment of the sensitivity of the method with respect to small training sets, an experiment has been carried out by feeding the NN by an increasing number of samples (from 200 to 1800 pixels per class), randomly collected in correspondence of each class of interest over the Stripmap acquisition in Fig.10.11. Moreover, to assess the efficiency of the learning phase, the algorithm has been validated on different and smaller validation sets (from 80 to 800 pixels per class). Afterward, the trained NNs have been applied to the whole image and the accuracies of the classification maps have been quantified through the cross-check with 4000 pixels representing the ground truth, provided by the comparison with the optical QuickBird image acquired on 8th December 2008 (Fig.10.12(a)). As Fig.10.13 shows, by increasing the training pattern size a slight enhancement in the overall accuracy can be observed. However, the experiment proved the effectiveness of the neural network classifier, since a quite precise map has been obtained even by exploiting only 200 training pixels per class (88.9% overall accuracy), while the best result, for which an overall accuracy of 92.5% has been evaluated, was achieved by feeding the net by means of 1800 samples per class.
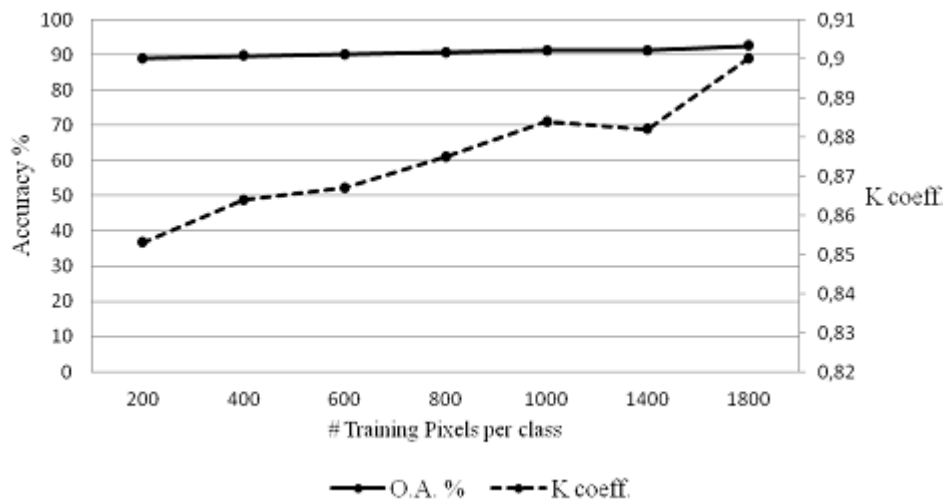
Figure 10.13. Overall accuracy and K coefficient achieved by increasing the size of the training set feeding the NN classifier.

Image classification of a different test site

The pixel classification methodology shown for COSMO-SkyMed Spotlight images acquired over Rome, has been validated on the Greater Denver area, Colorado, USA (Fig.10.14). In this case the class of *water* (W) has been added because of the large extension of this type of surface in the area. Differently from the analyzed italian site, here the urbanized areas are mainly characterized by single houses with gardens and yards, all organized in regular blocks. The identification of each single element in such dense residential settlements imaged by the VHR SAR, is not a trivial task. Indeed, the only buildings which can be visually recognized, are the larger and isolated ones, as well as the smaller houses oriented along the azimuth direction. In this cases, a stronger backscattering results in bright objects or lines, corresponding to the facades of the buildings.
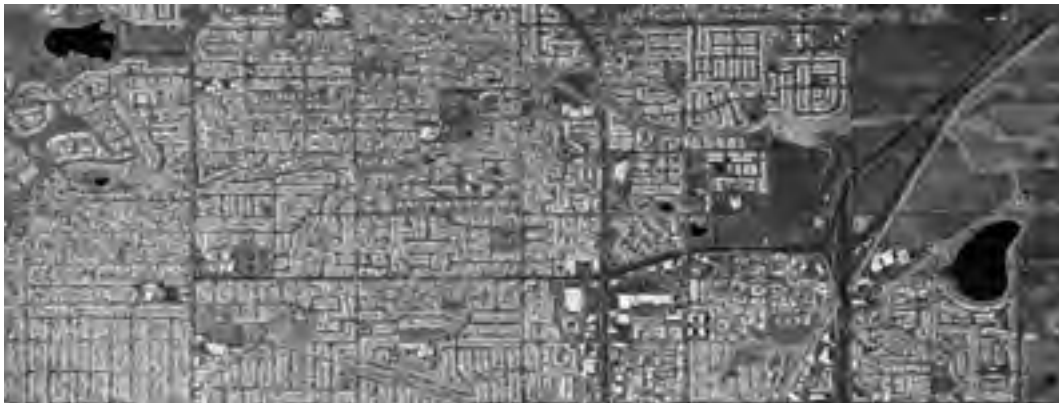
(a)

(b)

Figure 10.14. Two frames of the Greater Denver area, Colorado, USA, selected from a Spotlight COSMO-SkyMed acquisition (HH polarization; acquisition date: 11th December 2010) ©ASI. a) Training image (9428x3563 pixels). b) Image exploited for the fully automatic classification exercise (9282x4584 pixels).

Regarding the input quantities, backscattering coefficient, plus the four GLCM features (mean, contrast, variance and correlation) have been considered. The results for an image of size 9428x3563 pixels, and for an image of size 9282x4584 pixels, are shown in Fig.10.15(b) and Fig.10.16(b), respectively. The interpretation of the SAR pixels that have

been collected to build the training, validation and ground truth sets, has been performed by exploiting the only available optical source, which has been provided by GoogleEarth. The training phase has been carried out by feeding a 5x12x5 MLP-NN by means of 29000 samples, and its efficiency has been evaluated by considering 21100 validation pixels randomly collected in each land cover. Note that the classification of the second image has been carried out in a fully automatic mode, hence no pixels belonging to this image have been considered in the training phase of the classifying network.

From the figures it can be seen that the class *water* has been detected with high accuracy in the first example, while in the second some confusion occurs, likely due to the presence of different material on the surface. In both the maps the main roads seem to be well classified, while the narrow streets, crossing the residential blocks, are often confused with trees or building, because of the layover effect and the presence of vegetation along the avenues. As expected, the worst performance of the algorithm occurs in the detection of manmade structures. Actually, single and/or larger buildings are correctly classified, whereas the residential districts are basically classified as *trees*. The assessment of the achieved land cover maps is summarized in the confusion matrices in Table 10.8 and Table 10.9, which have been computed through a cross-checking with 25000 ground truth samples.

(a)

(b)

| | A | | LV | | T | | MM | | W |
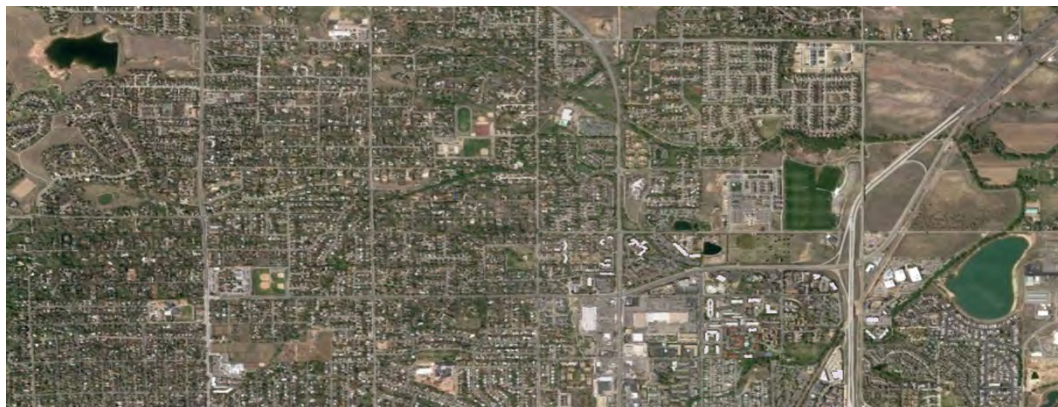
Figure 10.15. a)©Google Map Ground Truth. b)Classification map of the image in Fig.10.14(a) achieved by means of a 5x12x5 MLP-NN.

(a)

(b)

| | A | | LV | | T | | MM | | W |

Figure 10.16. a)©Google Map Ground Truth. b) Classification map of the image in Fig.10.14(b) achieved automatically by using the previously trained 5x12x5 MLP-NN.

Table 10.8. Confusion matrix referred to the classification map in Fig.10.15(b).

CLASS	A	LV	T	MM	W	TOT
A	69.54	7.82	4.58	1.14	4.02	17.42
LV	19.02	81.96	7.02	0.00	0.00	21.60
T	5.68	10.22	84.28	36.84	0.00	27.40
MM	0.00	0.00	4.12	62.02	0.00	13.23
W	5.76	0.00	0.00	0.00	95.98	20.35
TOT	100.00	100.00	100.00	100.00	100.00	100.00

Overall Accuracy: 78.75%

K coefficient: 0.73

Similarly to what has been performed for Spotlight data, a test of the classification on a different area, again the Great Denver area, Colorado, has been carried out (Fig.10.17).

The result shown in Fig.10.18(b) has been achieved by training a 5x10x10x5 MLP-NN by using 33800 samples and 14600 validation pixels. A qualitative evaluation of the result put in evidence a pretty good agreement with the ground truth. In fact the class *water*

Table 10.9. Confusion matrix referred to the classification map in Fig.10.16(b).

CLASS	A	LV	T	MM	W	TOT
A	82.42	15.58	1.02	0.52	28.02	25.51
LV	13.08	83.42	20.74	0.00	0.24	23.50
T	2.42	0.76	76.50	34.36	0.00	22.81
MM	0.00	0.00	1.74	65.12	0.00	13.37
W	2.08	0.24	0.00	0.00	71.74	14.81
TOT	100.00	100.00	100.00	100.00	100.00	100.00

Overall Accuracy: 75.84%

K coefficient: 0.70

Figure 10.17.   Great Denver area, Colorado, USA imaged by COSMO-SkyMed operating in Stripmap mode (4676x4087 pixels), in HH polarization ©ASI (date of acquisition: 10th December 2010).

has been correctly classified, as well as the vegetation classes. The main roads have been detected with a quite good accuracy, whereas some confusion occurs for the narrow streets, which are hidden because of the SAR geometrical distortions affecting trees and buildings located along them. The algorithm demonstrated to be able to recognize most of the largest buildings, and some of the smaller houses belonging to dense urban districts. Yet, the closeness of the houses, as well as the presence of trees and gardens led the neural network to erroneously classify the manmade structures as trees. In Table 10.10 it is reported the confusion matrix computed by using 10000 ground truth samples.
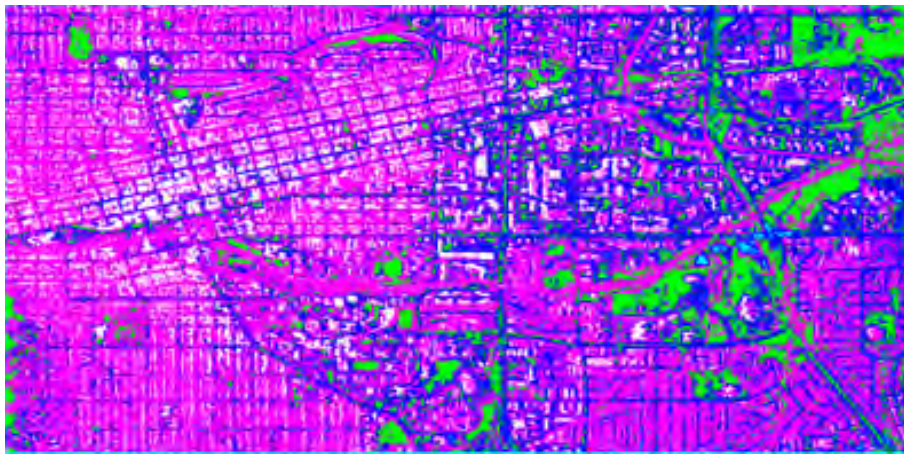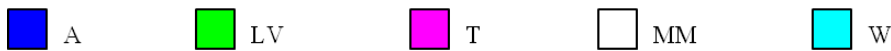
(a)

(b)
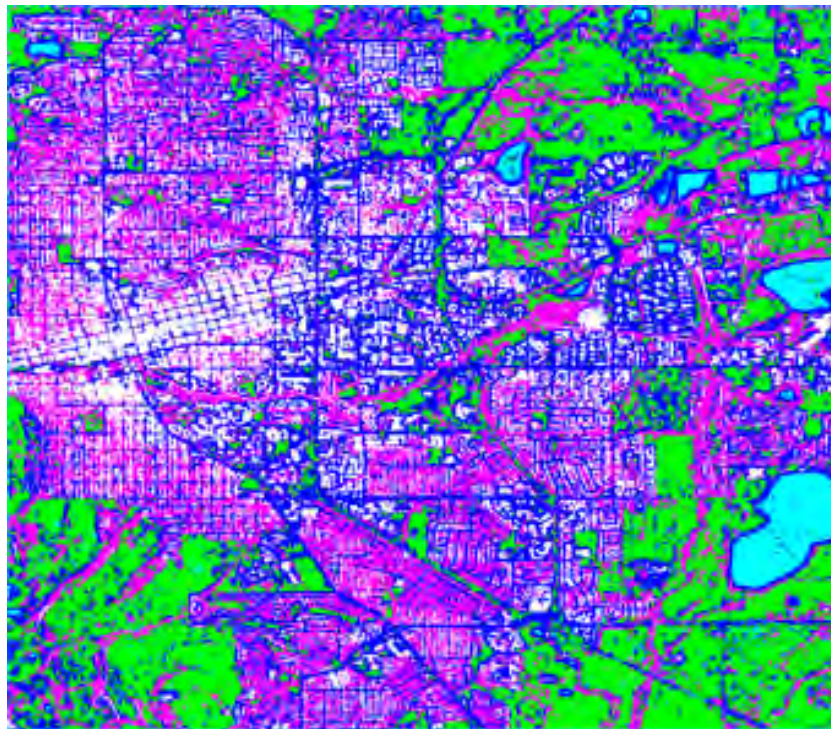
| A | LV | T | MM | W |

Figure 10.18. a)©Google Map Ground Truth. b)Classification map of the image in Fig.10.17 achieved by means of a 5x10x10x5 MLP-NN.

Table 10.10. Confusion matrix referred to the classification map in Fig.10.18(b).

CLASS	A	LV	T	MM	W	TOT
A	82.50	9.30	5.05	2.55	4.85	20.85
LV	15.45	86.70	9.10	0.00	0.85	22.42
T	0.55	0.20	81.45	16.75	0.00	19.79
MM	1.50	0.00	4.10	80.70	0.00	17.26
W	0.00	3.80	0.30	0.00	94.30	19.68
TOT	100.00	100.00	100.00	100.00	100.00	100.00

Overall Accuracy: 85.13%

K coefficient: 0.81

10.3   Exploitation of the interferometric coherence

Although promising results have been achieved by exploiting backscattering and its textural and contextual properties extracted from a single COSMO-SkyMed image, the algorithm exhibited some weakness in discriminating between artificial surfaces and natural areas. In order to increase the accuracy of the classification, the possibility to benefit from the availability of interferometric images pairs has been investigated. Indeed, when two images are taken on the same area within a limited time window, the modulus of the degree of coherence can be added as an additional piece of information besides the backscattering and texture characteristics. The rationale is based on the observed higher values over artificial surfaces and objects, such as streets and buildings, in comparison with the degree of coherence values evaluated over vegetated areas.

The following experiments have been carried out by means of the information stemmed from two couples of Spotlight and Stripmap images in order to identify three broad classes (asphalt, natural areas and manmade structures). In addition, a further test has been performed by using a pair of Spotlight acquisitions for the classification of *asphalt*, *low vegetation*, *trees* and *manmade* classes. In the first exercises, the MLP-NN have been fed by seven elements input vectors, including the degree of coherence and the backscattering amplitude, local mean and local standard deviation, estimated in both the images. Aiming at the discrimination between four classes, eleven features have been considered. In fact, besides the interferometric coherence, for each image, the same five features already exploited in the experiments discussed above, have been included into the input vector.

Results

Three classes mapping by using two Spotlight images and simple local texture.

The first test has been carried out by exploiting the pair of Spotlight images taken on 8th June 2010 by COSMO-SkyMed 3 satellite and on 9th July 2010 by COSMO-SkyMed 2 [69], [77]. For sake of comparison, the same scene classified in Fig.10.1 has been processed by training a 7x9x9x3 MLP-NN by using the same set of 735 pixels. The performance of the learning algorithm has been tested on the same 315 validation samples as well. The obtained land cover map in Fig.10.19 shows an actual enhancement of the quality of the classification, which has been assessed by a cross-checking of 861 ground truth pixels.

The confusion matrix in Table 10.11 shows an overall accuracy of 92.91%. In particular, the main roads and traffic circles are now better delineated. However, since dry bare soil exceeds vegetation, some confusion can be still observed between asphalted surfaces and natural surfaces. It has also to be noted that no filtering to remove high spatial frequency effects (such as small isolated high-backscattering pixels) has been applied to refine the output map. In fact, if in some cases the very small objects observed in the classified map
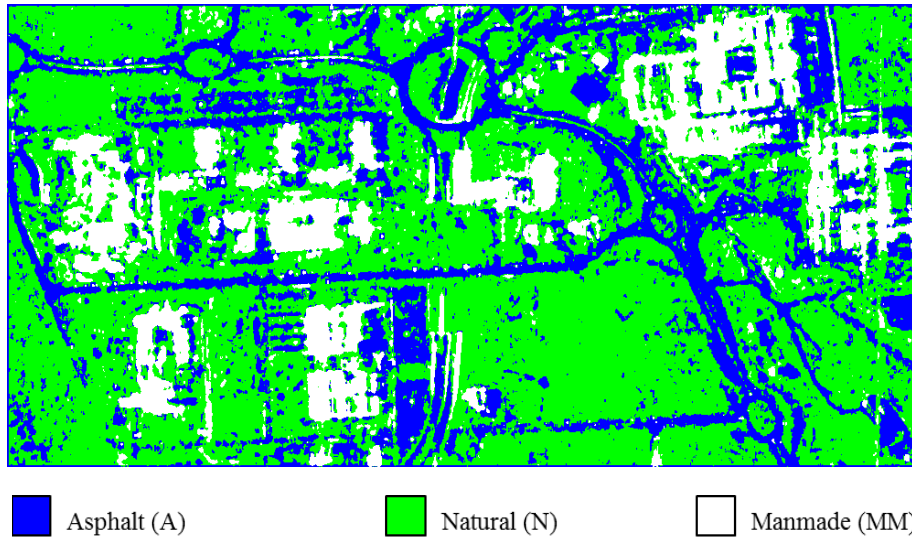
Figure 10.19. Classification map (994x556 pixels) achieved by exploiting the backscattering amplitude, the local mean, the local standard deviation and the interferometric coherence derived from two Spotlight COSMO-SkyMed images acquired on 8th June 2010 and 9th July 2010 over Tor Vergata, Rome (Italy). MLP-NN architecture: 7x9x9x3.

are artifacts of speckle, in other cases they are in agreement with the real scenario, which includes the appearance of cars, lamps on the road sides or other discontinuities in the locally dominating type of surface. An interesting feature is noted in parts of the biggest roads, where the algorithm is capable of detecting the presence of the small grass and of the guardrail separating the two ways of direction.

Table 10.11. Confusion matrix referred to the classification map in Fig.10.19.

CLASS	A	N	MM	TOT
A	94.76	9.21	0.00	27.99
N	5.24	89.91	2.05	49.36
MM	0.00	0.88	97.95	22.65
TOT	100.00	100.00	100.00	100.00
Overall Accuracy: 92.91%				
K coefficient: 0.88				

By the analysis of Table 10.12, it is reasonable to assert that the addition of the second image and of the interferometric coherence plays a significant role in improving the discrimination between asphalted and natural surfaces. This is evident in Fig.10.20 which compares the results of the single-image with that of the double-image algorithm. It can be observed that the roundabout and the parking area behind the School of Engineering (Fig.10.20(a)), as well as the parking area near the building of Faculty of Economics (Fig.10.20(d)), are much better detected by the two images classification. Differently, the additional information carried by the second image did not alter significantly the performance on the built surfaces. In fact, the high backscattering characteristics of this class seem to make it discernible using only one image.

(a) QuickBird image   (b) Two images   (c) Single image

(d) QuickBird image   (e) Two images   (f) Single image
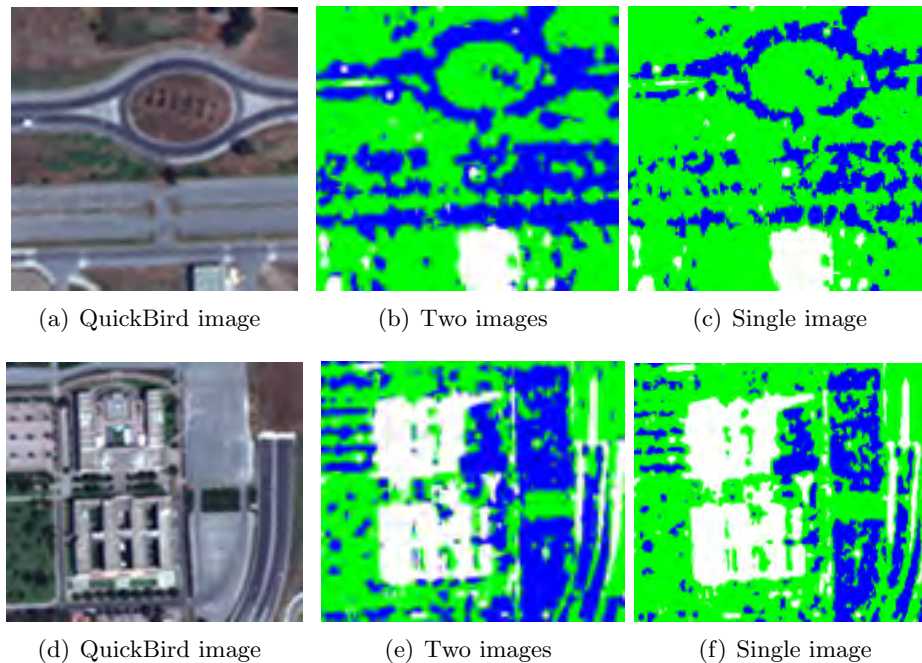
Figure 10.20. Improvement in the discrimination of the asphalted surfaces using the 2 images (Fig.10.20(b) and Fig.10.20(e)) instead of the 1-image algorithm (Fig.10.20(c) and Fig.10.20(f)).

Table 10.12. Comparison of per-class accuracies (%) considering the single image and the two-images schemes.

	Per-class accuracies (%)	
CLASS	**1 IMAGE**	**2 IMAGES**
A	79.05	94.76
N	94.08	89.91
MM	97.95	97.95
Overall accuracies	91.30	92.91

Three classes mapping by using two Stripmap images and simple local texture.

Two Stripmap acquisitions, taken on 24th and 25th May 2010 by COSMO-SkyMed 2 and COSMO-SkyMed 3, respectively have been considered for a second experiment. As for the Spotlight case, the same frame classified by exploiting the single image in Fig. 10.3 has been processed in this exercise. Given the closeness in time of the acquisitions, no significant variations are present in the second image. The same training and validation set previously collected has been used to feed a 7x9x9x3 MLP-NN which provided the land cover map represented in Fig.10.21.

The visual assessment of the result, as well as the confusion matrix computed through 744 ground truth samples (Table 10.13) lead to state and confirm the utility of the backscattering amplitude, its contextual properties, and of the interferometric coherence as an efficient set of exploitable information carried by two SAR acquisitions. In particular, the main benefit due to the exploitation of the degree of the interferometric coherence is in the better delineation of the streets. Moreover, a comparison with the map achieved by using a single acquisition (Fig.10.4(b)), put in evidence the improvement in the classification of natural surfaces, where the strong "noise" due to the confusion with the *asphalt* class has been definitely mitigated. As a consequence, the overall accuracy increased from 88.31% to 94.35% (Table 10.14).
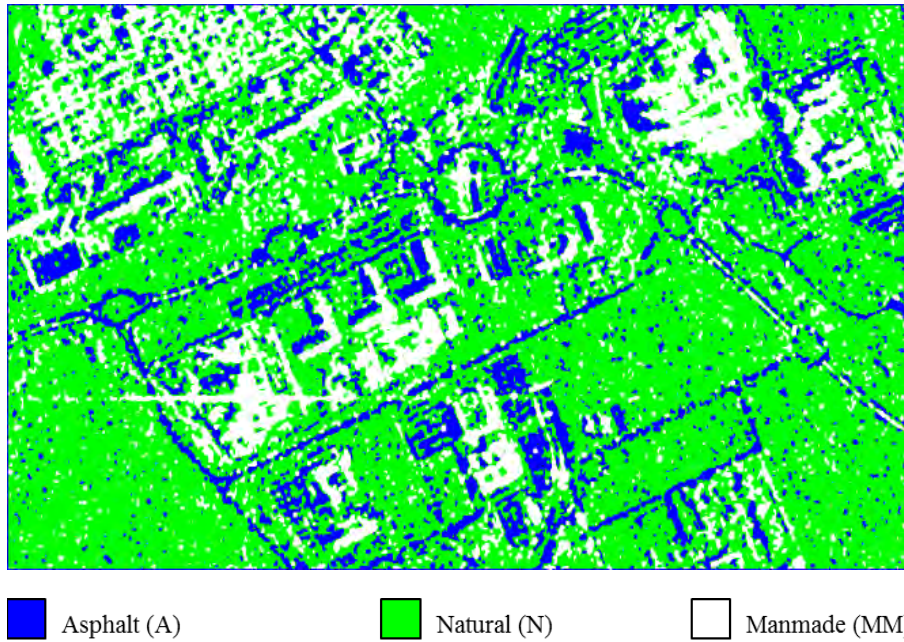
Figure 10.21. Classification map (836x520 pixels) achieved by exploiting the backscattering amplitude, the local mean, the local standard deviation and the interferometric coherence derived from two Stripmap COSMO-SkyMed images acquired on 24th and 25th May 2010 over Tor Vergata, Rome (Italy). MLP-NN architecture: 7x9x9x3.

Table 10.13. Confusion matrix referred to the classification map in Fig.10.21.

| CLASS	A	N	MM	TOT
A	92.86	2.65	0.00	27.42
N	7.14	92.04	0.00	43.95
MM	0.00	5.31	100.00	28.63
TOT	100.00	100.00	100.00	100.00
Overall Accuracy: 94.35%				
K coefficient: 0.91				

Table 10.14. Comparison of per-class accuracies (%) considering the single image and the two-images schemes.

	Per-class accuracies (%)	
CLASS	1 IMAGE	2 IMAGES
A	90.00	92.86
N	80.53	92.04
MM	100.00	100.00
Overall accuracies	88.31	94.35

Four classes mapping by using two Spotlight images and GLCM textural features.

The experiment has been carried out by considering the pair of Spotlight images of Tor Vergata, taken on 16th December 2010 and 1st January 2011. The scene under study is represented in false color composition in Fig.10.22. The image has been obtained by filtering the former acquisition in the red channel, the second in the green one, and the coherence in the blue band. As it can be observed, this simple method of visualization

enhance the interpretation of the SAR image. In fact, because of their high backscattering values and high degree of coherence, the manmade structures can be easily identified as white objects. On the other side, the asphalted surfaces, which are characterized by low backscattering values but high degree of coherence, results in blue pixels. Elsewhere, over the vegetated surfaces, the degree of coherence is lower and more variable. However, higher backscattering values, resulting in brighter green/yellow patches, can be mainly associated to the class *trees*.


Figure 10.22. False color composite image of Tor Vergata, Rome (Italy)(8298x5822 pixels). R: Spotlight COSMO-SkyMed image. HH polarization. Date of acquisition: 16th December 2010 ©ASI. G: Spotlight COSMO-SkyMed image. HH polarization. Date of acquisition: 1st January 2011 ©ASI. B: Interferometric coherence image.

The classification task has been performed by training and testing an 11x11x4 MLP-NN by means of 51000 training samples and 30500 validation pixels. The output map is shown in Fig.10.23(b). For the result assessment, the closest optical image available was the QuickBird in Fig.10.23(a), which was acquired more than four months after the second CSK image. A qualitative comparison reveals a quite high accuracy of the classification. The exploitation of the coherence information allowed the correct classification of the road network, and in particular of the motorway, whose recognition has not been successful in the previously discussed experiments. Moreover, both single large manmade structures, and buildings located in denser urban environments, have been identified with a good accuracy. The differentiation among the two natural surfaces seems to be well correlated with the ground truth. The quantitative assessment of the classification exercise has been carried out by using 24000 ground control points. The confusion matrix in Table 10.15 agrees to the qualitative interpretation of the result. An overall accuracy of 87.80% has been reached in spite of some misclassification between the *low vegetation* and *trees* classes, as well as between the latter and the *manmade* class.

The achievement is particularly satisfactory if compared with that obtained by using the single image acquired on December. As it can be observed in Fig.10.24, the exploitation of eleven features, including the GLCM textural parameters and the interferometric coherence, enhances significantly the correct classification of asphalted surfaces and of manmade constructions in dense urban areas. Moreover, such an approach has led to a general increasing of the overall and per-class accuracies (Table 10.16).

(a)

(b)
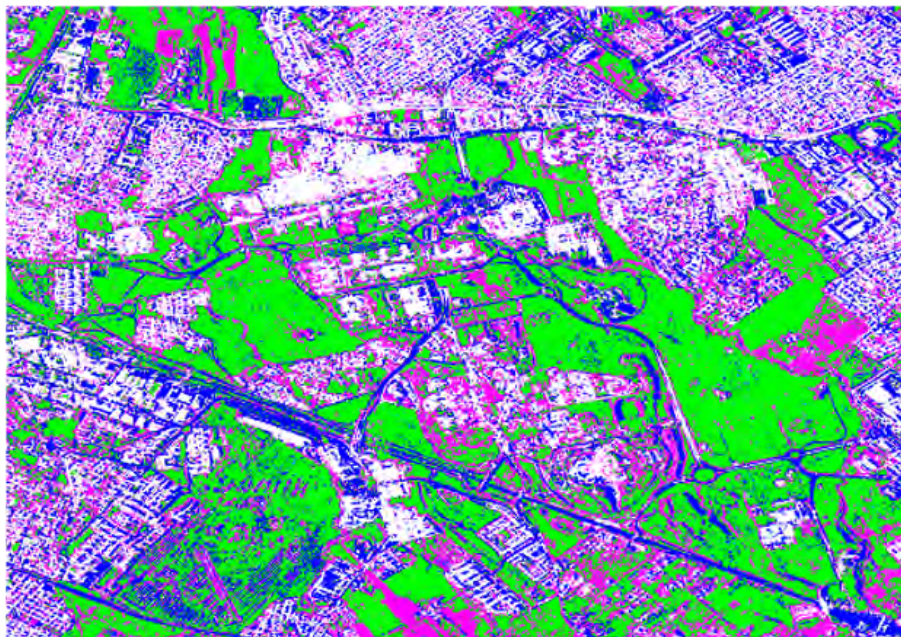
Figure 10.23. Tor Vergata area, Rome (Italy) imaged by a) QuickBird sensor ©DigitalGlobe (Acquisition date: 12th April 2011). b) Classification map achieved by exploiting the intensity of backscattering, mean, variance, contrast and correlation stemmed from the GLCM computation, and interferometric coherence derived from two Spotlight COSMO-SkyMed images acquired on 16th December 2010 and 1st January 2011. MLP-NN architecture: 11x11x4.

Table 10.15. Confusion matrix referred to the classification map in Fig.10.23(b).

CLASS	A	LV	T	MM	TOT
A	89.70	6.11	2.97	0.00	24.70
LV	4.12	86.67	9.90	0.00	25.17
T	1.03	6.11	79.21	4.40	22.69
MM	5.15	1.11	7.92	95.60	27.45
TOT	100.00	100.00	100.00	100.00	100.00

Overall Accuracy: 87.80%

K coefficient: 0.84

(a) QuickBird image　　　　(b) Two images　　　　(c) Single image

(d) QuickBird image　　　　(e) Two images　　　　(f) Single image
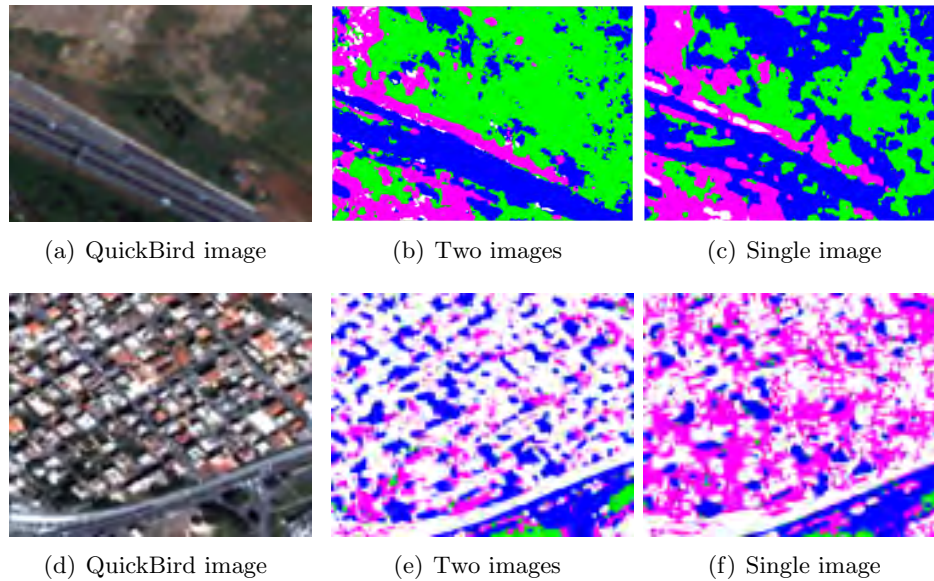
Figure 10.24. Improvements in the classification of asphalted surfaces (Fig.10.24(a)) and of dense urban environment (Fig.10.24(d)) by using two images (classification details in Fig.10.24(b) and Fig.10.24(e)) instead of a single acquisition (classification details in Fig.10.24(c) and Fig.10.24(f)).

Table 10.16. Comparison of per-class accuracies (%) considering the single image and the two-images schemes.

	Per-class accuracies (%)	
CLASS	1 IMAGE	2 IMAGES
A	81.08	89.70
LV	73.42	86.67
T	76.45	79.21
MM	79.98	95.60
Overall accuracies	77.73	87.80

10.4　Conclusions

Aiming at the monitoring of land cover in suburban areas by using VHR Spotlight and Stripmap COSMO-SkyMed imagery, the presented study has been focused on the development of an automatic processing chain based on neural networks. The analysis of the several surfaces belonging to a complex scenario, such as that which characterizes the Tor Vergata area, put in evidence the necessity of addressing the issue of investigating on further possible sources of information to be derived from single polarized SAR data.

A first attempt concerned the exploitation of simple local texture (local mean and standard deviation) in order to achieve thematic maps, in which were possibly identified three main classes: *asphalt*, *natural* surfaces and *manmade* structures. High accuracies (larger than 80 %) have been obtained for both the Spotlight and Stripmap products.

Afterward, the analysis of the information carried by the textural feature stemmed from the computation of the GLCM, led to the selection of four parameters (*mean*, *variance*, *contrast* and *correlation*) which have been included into the input vector of the MLP-NN. Such approach allowed the enhancement of the classification maps, especially in terms of number of recognized land covers. In fact, the algorithm proved to be able to discriminate four classes: *asphalt*, *low vegetation* surfaces, *trees* and *manmade* structures. Generally speaking, the achieved maps reproduce the ground-truth patterns, reaching overall accuracies higher than 80%. However, some confusions occurred in the discrimination between scarcely vegetated areas and asphalted surfaces. The reason might be due to the presence of a large percentage of bare soil whose backscattering and textural characteristics may be similar to those of relatively rough streets.

Tests on the generalization capability of the MLP-NN based algorithm have been carried out by applying the trained neural networks to new images, which have not been used for the learning phase. It has to be noted that, since such experiments provided promising results in a fully automatic way, the neural network approach seems to be valuable towards the implementation of fast and effective processing chain to perform the land cover classification of a large amount of SAR data, without any human intervention.

Tests on the robustness of the algorithm have been also accomplished. By increasing the size of the training set, and by evaluating the accuracies of the achieved classification maps, it has been proved the capability of the algorithm in providing high quality products, even when only few hundreds of samples are exploited.

Moreover, besides the sub-urban area of Tor Vergata, a different type of urban environment has been taken into account for the classification task. Specifically, single polarized Spotlight and Stripmap COSMO-SkyMed images of the Great Denver area have been classified by feeding the MLP-NNs by means of the same five input features (intensity of the backscattering, plus four GLCM textural parameters: *mean*, *variance*, *contrast* and *correlation*). Differently from the previous exercises, the human settlements are mainly represented by dense residential areas, organized in blocks of lined small houses surrounded by gardens; moreover, the investigated scenario included a number of lakes, which made necessary to add the class *water* into the output vector of the classifier. The results showed a good interpretation of the main road networks, wide vegetated areas and water surfaces. The algorithms proved to be still able to identify large or isolated manmade constructions, whereas the multiple backscattering contributions occurring in high density residential districts, led the net to mis-interpret the buildings, which have been classified as trees.

Thanks to the configuration of the COSMO-SkyMed constellation, the availability of interferometric pairs of acquisitions led to the investigation on the possibility of employing the coherence information to improve the discrimination among artificial and natural surfaces. Actually, it has been demonstrated the effectiveness of such approach which especially allowed the enhancement of the *asphalt* and *natural* (or *low vegetation*) classes accuracy.

To conclude, it is worth to highlight the fact that the developed computational tool can readily and effectively incorporate additional information stemming from images taken at different polarizations or by other kinds of sensors, in order to improve the discrimination of several land covers.

Chapter 11

Data fusion

The recent development of new-generation satellite systems equipped with VHR SAR and optical sensors, leads scientists to explore the potentialities of the use of the combined information contents. While optical remote sensing has proven its usefulness in land cover and change detection maps production, SAR images have the advantage of high confidence/certainty in observing the scene. Data fusion is expected to be a suitable tool both to overcome the limitations of single sensors (mainly due to the difficult interpretation of SAR images, especially if acquired in single polarization, and to the optical sensors dependency on weather and lighting conditions),and to enhance identification and monitoring of various land covers.

This study mainly regards the synergy of different sensors in discriminating within vegetated landscapes, i.e., within a single land cover type. Different combinations of Spotlight COSMO-SkyMed data, either at HH or at VV polarization, and WorldView-2 multispectral images have been tested in order to evaluate the mapping improvement brought by the fusion of the information, in comparison with the single kind of data. Combining optical and microwave features is expected to yield enhanced land cover maps, with higher overall accuracy and increased number of vegetation classes.

Below, the exploited multi-data images will be described, as well as their necessary pre-processing. Afterward, the experiment will be explained and the results discussed.

11.1 Vegetation mapping by combining VHR SAR and optical data

The benefits from fusing SAR and optical observations has been already confirmed by several authors [78–82]. In particular, Pierdicca et al. [82] found that the separability of agricultural classes was enhanced by simultaneous use of optical (AVNIR-2 and SPOT) and microwave (COSMO-SkyMed, PalSAR, ERS-2) sensors.

The rationale is that different plants may be characterized by different geometries of vegetation canopies, though having close water and chlorophyll content. Hence, the similar spectral response would not allow discrimination by optical imaging. Instead, SAR images carry the differentiating information, being the microwave backscattering sensitive to the


Part of this chapter's contents is extracted from:

1. C. Pratola, G. A. Licciardi, F. Del Frate, G. Schiavon and D. Solimini, *Fusion of VHR Multispectral and X-band SAR data for the enhancement of vegetation maps*, Proc. IEEE, International Geoscience and Remote Sensing Symposium , IGARSS 2012, pp. 6793-6796, 2012.

:

geometry of scatterers. On the other hand, different plants may happen to have quite similar geometric characteristics, but a diverse spectral response. In this case, microwave imagery would not be able to discriminate, whereas optical observations would carry the useful information.

In the example in Fig.11.1 three different natural surfaces acquired by CSK in HH and VV polarizations, and by WV2 in multispectral mode have been analyzed. As it can be observed, the SAR sensor operating in VV polarization is able to differentiate overhead trellis vineyard from shrubs, whereas the optical signatures appear very similar, hindering the discrimination between the two land covers. On the other hand, shrubs and bare soil exhibit, in this particular case, a similar VV backscattering, probably due to the speckle, the Bragg effect, the roughness, or to the soil moisture; however the multispectral information may make easier their classification. In HH polarization, the less penetration into the canopies of the X-band signal, does not permit to distinguish shrubs from vineyard, while it is possible to discriminate between both of them and bare soil. When different natural surfaces show at the same time similar backscattering values and spectral signature, the classification task requires for the identification of further suitable features, such as texture parameters, vegetation indexes etc...

(a)   (b)

(c)   (d)   (e)
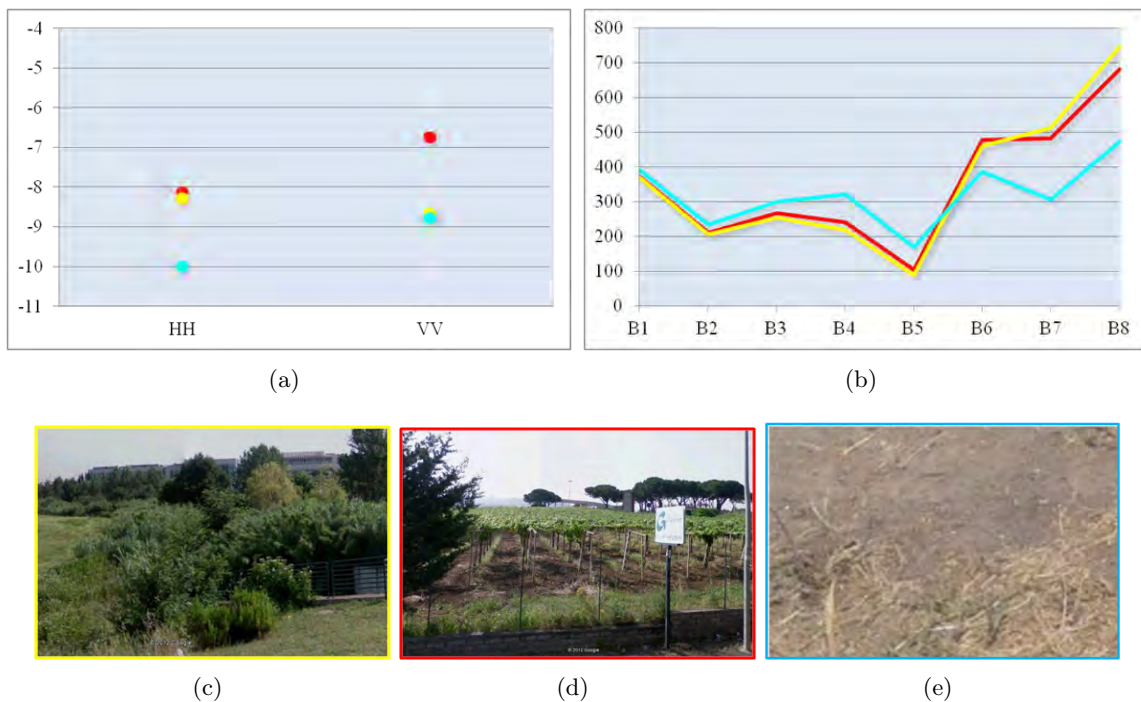
Figure 11.1. a)Example of backscattering behavior and b) multispectral signature recorded by CSK and WV-2 respectively, of three different natural land covers: c) shrubs (yellow), d) vineyard (red) and e) bare soil (cyan).

Fusing optical and SAR metric images, such as those acquired by CSK and WV2, is still an open issue, particularly in urban environments ([83], [84]), where the localization effects of the substantially different acquisition features make the co-registration arduous at the pixel level. However, this effect becomes negligible over flat areas essentially covered by vegetation.

Image pre-processing has been necessary towards the joint exploitation of the available multi-sensors dataset. Hence each SAR acquisition has been orthorectified, geocoded, calibrated and filtered to reduce the speckle noise. Because of the different spatial resolution, a further processing step was required to co-register the images, so that the higher COSMO-

SkyMed spatial resolution of 1 m was degraded to the lower 2 m resolution of the optical acquisitions. Finally, to enhance the information content of the SAR data, textural images derived from the GLCM computation [14] have been produced. Also based on previous experience [41], the best discrimination between the classes of interest has been achieved by using mean and homogeneity textures, derived from the GLCM with a window size of 3x3 and 7x7 pixels, respectively. The joint use of multi-sensor data in vegetation mapping has been tested over an area in the southeast outskirts of Rome, where cultivated fields (overhead trellis or rows vineyards and olive yards) alternate with pastures, bare soil and shrubs (Fig.11.2).

The data set consists of the multi-temporal optical and SAR acquisitions reported in Table 11.1. The multispectral WorldView-2 set includes two images taken on the 4th of July 2010 and on the 10th of October 2011. Six Spotlight-mode COSMO-SkyMed images, covering all the seasons, have been jointly exploited. Four SAR images are acquired at HH polarization (8th June 2010, 2nd January 2011, 23rd April 2011 and 8th October 2011), while the others at VV polarization (23rd June 2010 and 26th July 2010).

(a)   (b)
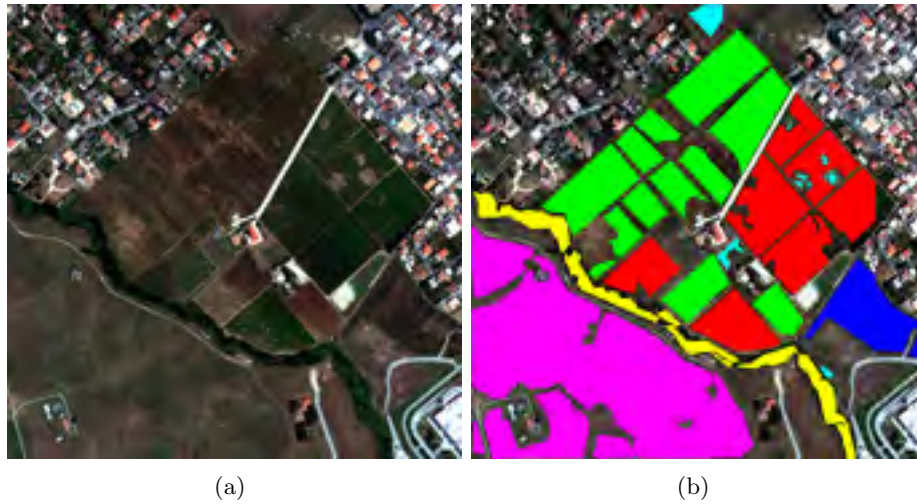
Figure 11.2. a) WV2 image of the area of interest (acquisition date: 4th July 2010) ©DigitalGlobe; b) Ground Truth: Overhead trellis vineyards (Red); Row vineyards (Green); Olives (Blue); Shrubs (Yellow); Bare Soil (Cyan); Pastures (Magenta).

Table 11.1. Dataset

Image	Sensor	Date	Polarization	Information
A	WV2	04-07-2010	-	8 bands
B	WV2	10-10-2011	-	8 bands
C	CSK	08-06-2010	HH	Mean+Hom
D	CSK	23-06-2010	VV	Mean+Hom
E	CSK	26-07-2010	VV	Mean+Hom
F	CSK	02-01-2011	HH	Mean+Hom
G	CSK	23-04-2011	HH	Mean+Hom
H	CSK	08-10-2011	HH	Mean+Hom

11.2 Results

As for the previously discussed pixel-based classification experiments, also the vegetation mapping has been carried out by the commonly used MLP-NN, and the efficiency of the training has been evaluated by the MSE function. A few thousands (9925) of training pixels representing the two types of vineyards (VT: cultivated in overhead trellis; VR: cultivated in rows), olive groves (O), shrubs (S), bare soil (BS) and pastures (P), as well as a smaller validation set (5950 pixels), were randomly collected over known areas to train the net and to evaluate its generalization capability during the learning phase. It has to be noted that the net topology was tailored to each combination of images chosen for the classification task. Once properly trained, the neural network has processed the images, and the accuracy of the result has been assessed by cross-checks with the ground truth shown in Fig.11.2(b). It has to be pointed out that no significant changes in land covers occurred between the dates of acquisition, with the only exception of bare soil which somewhere became covered by grass. Such varying areas have not been included into the training/validation set, nor into the ground truth.

In a first experiment, data acquired in summer 2010 have been considered, by joining the information embedded within one multispectral WorldView-2 image (A in Table 11.1), with four textural features extracted from two Spotlight CSK acquisitions at HH and VV polarization (C and D in Table 11.1).
A similar exercise has been repeated for a set of acquisitions taken in autumn and winter 2011, by stacking in a twelve-components vector the previously co-registered B, F and H images (Table 11.1).

A third classification has been performed through the fusion of the twenty-eight features derived from the whole available dataset.
As a comparison, two more experiments have been carried out: in one case only mean and homogeneity textures extracted from the HH and VV CSK data (C and D in Table 11.1) have been taken into account, while in the second one the information content embedded within the eight bands of the single multispectral image A (Table 11.1) has been exploited.
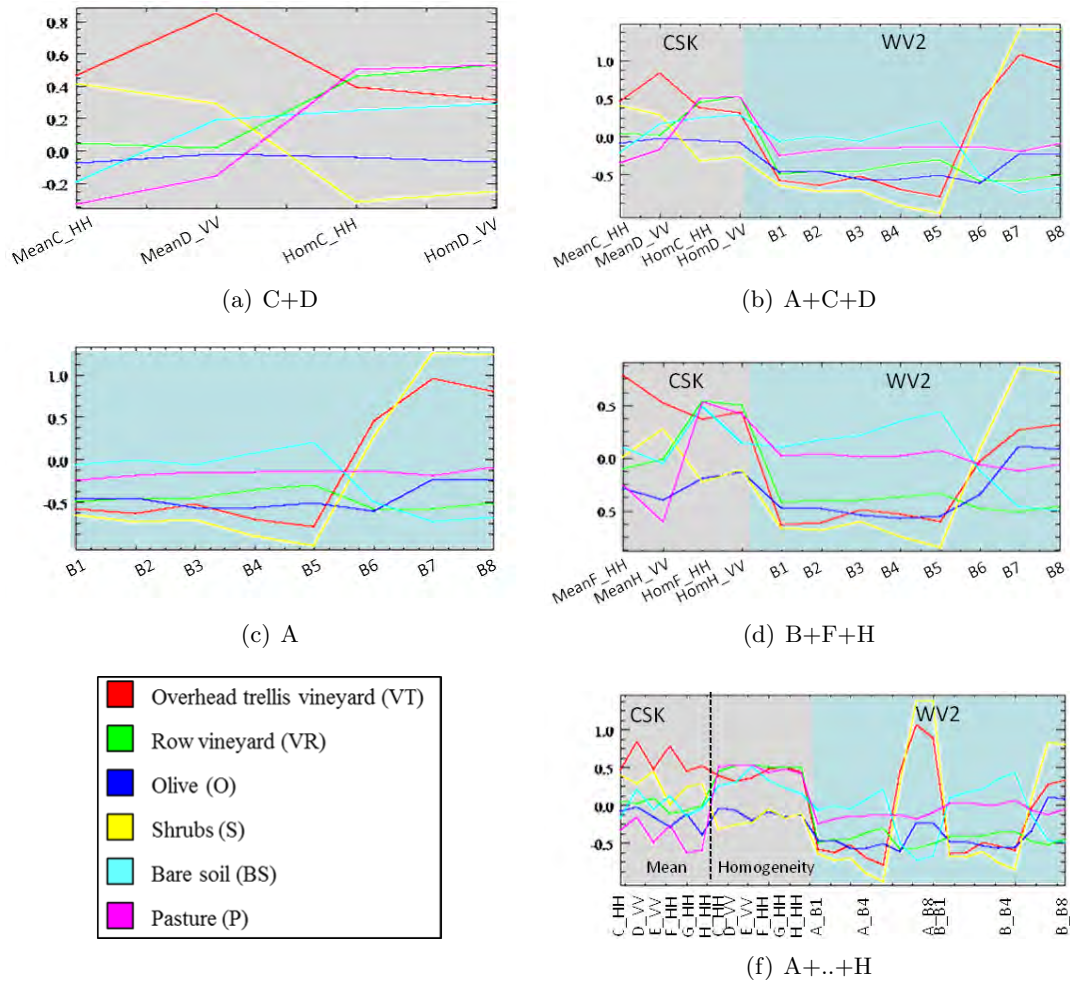
In Fig.11.3 the means of the normalized values of the features exploited for the neural network training put in evidence how the joint use of multi-temporal data, SAR derived textural parameters and multispectral information can be useful to improve the discrimination between some natural surfaces. In particular rather different values of mean and homogeneity, even considering a single polarization SAR image, allow the identification of overhead trellis vineyard and shrubs. Note that, since in all the exercises it has been observed that the exploitation of the amplitude of the backscattering did not bring any significant contribution to the discrimination between the classes, it has been neglected with the advantage of a lower computational cost.
Fig.11.4 shows the vegetation maps resulting from each classification experiment [85]. Comparison with the ground truth confirms the benefits brought by the joint use of multi-sensors data.

The map in Fig.11.4(a), characterized by a general confusion among all the natural surfaces (bare soil is missing), puts in evidence the lack of suitable information content in the textural features extracted only from a pair of SAR images.

On its side, exploiting a single multispectral image yields a better result (Fig.11.4(b)), but confusion occurs between overhead-trellis vineyards and shrubs, as well as between pasture and bare soil.

The classification results shown in Fig.11.4(c) and Fig.11.4(d), obtained by fusing SAR and optical data acquired in different periods of the year, exhibit a slight further improvement, in spite of some misclassification of the pasture areas, mainly attributable to mixed pixels. In Fig.11.4(d), relative to the winter-autumn 2011 acquisitions, we note the correct

(a) C+D

(b) A+C+D

(c) A

(d) B+F+H

Overhead trellis vineyard (VT)
Row vineyard (VR)
Olive (O)
Shrubs (S)
Bare soil (BS)
Pasture (P)

(f) A+..+H

Figure 11.3. Means of the normalized values of the features exploited as input to the neural network classifier.

identification of newly planted olive trees in the area adjacent to the already existent olive grove.

Finally, the highest correlation with the ground truth appears in the map of Fig.11.4(e), obtained by including into the classifier all the features in eight multi-sensors images.

It has to be mentioned that an *urban* class has not been included into the classification exercises. When data refer to manmade structures, the classifier attributes the built surfaces mainly to the *shrub* or *bare soil* classes. This error does not affect the accuracies of each class reported in Table 11.2. In general, increasing the input vector dimension by adding information carried by both SAR and optical acquisitions, enhances the vegetation mapping in terms of both producer and overall accuracies. The highest overall accuracy, larger than 90%, is attained by processing the twenty-eight features image by a 28x40x40x6 MPL. The classification result is also quite accurate for each class. The lowest producer's accuracy, less than 84%, occurs for the olive grove class and is probably caused by the different geometric acquisition features of the satellite systems.

11.3 Conclusions

This study has been carried out aiming at the enhancement of vegetation mapping, attainable through the fusion of COSMO-SkyMed X-band VHR SAR images and multispectral

WorldView-2 data. Different combinations of the multi-sensors data have been investigated, and the results have been compared with those obtained by single-sensor acquisitions. The classifier has been implemented by an MLP-NN algorithm, fed by the eight values of spectral reflectivity provided by the optical sensor and/or by *mean* and *homogeneity* textural parameters derived from the SAR images. The checks of the classification results against the ground truth highlight the benefits from the data fusion, generally resulting in a rather significant improvement of the producer's and overall accuracies. As expected, widening the data set by including different acquisition systems and dates, enhances the vegetation mapping, up to accuracies over 90%. Such results look promising towards a possible continuous monitoring of agricultural surfaces by the available different multi-sensor products.

Table 11.2. Comparison of the producer accuracies (%) related to the maps achieved by the use of: CSK HH and VV imagery (C+D); single WV2 image (A); CSK and WV2 summer acquisitions (A+C+D); CSK and WV2 winter-autumn acquisitions (B+F+H); fusion of the whole multi-sensors dataset (A+..+H). VT: overhead trellis vineyards; VR: vineyards in rows; O: olive groves; S: shrubs; BS: bare soil; P: pastures.

Ex.	C+D	A	A+C+D	B+F+H	A+..+H
Input neurons	4	8	12	12	28
Hidden neurons	2x16	2x20	2x24	2x24	2x40
VT	86.40	81.12	91.25	92.31	87.12
VR	57.51	87.22	95.67	95.09	97.84
O	20.92	59.76	82.56	71.57	83.40
S	41.11	48.56	75.45	80.33	84.06
BS	0.00	57.89	55.84	83.1	88.61
P	60.01	88.03	83.99	78.45	90.59
OA	60.89	81.96	87.56	85.43	90.86
K	0.46	0.75	0.83	0.80	0.87
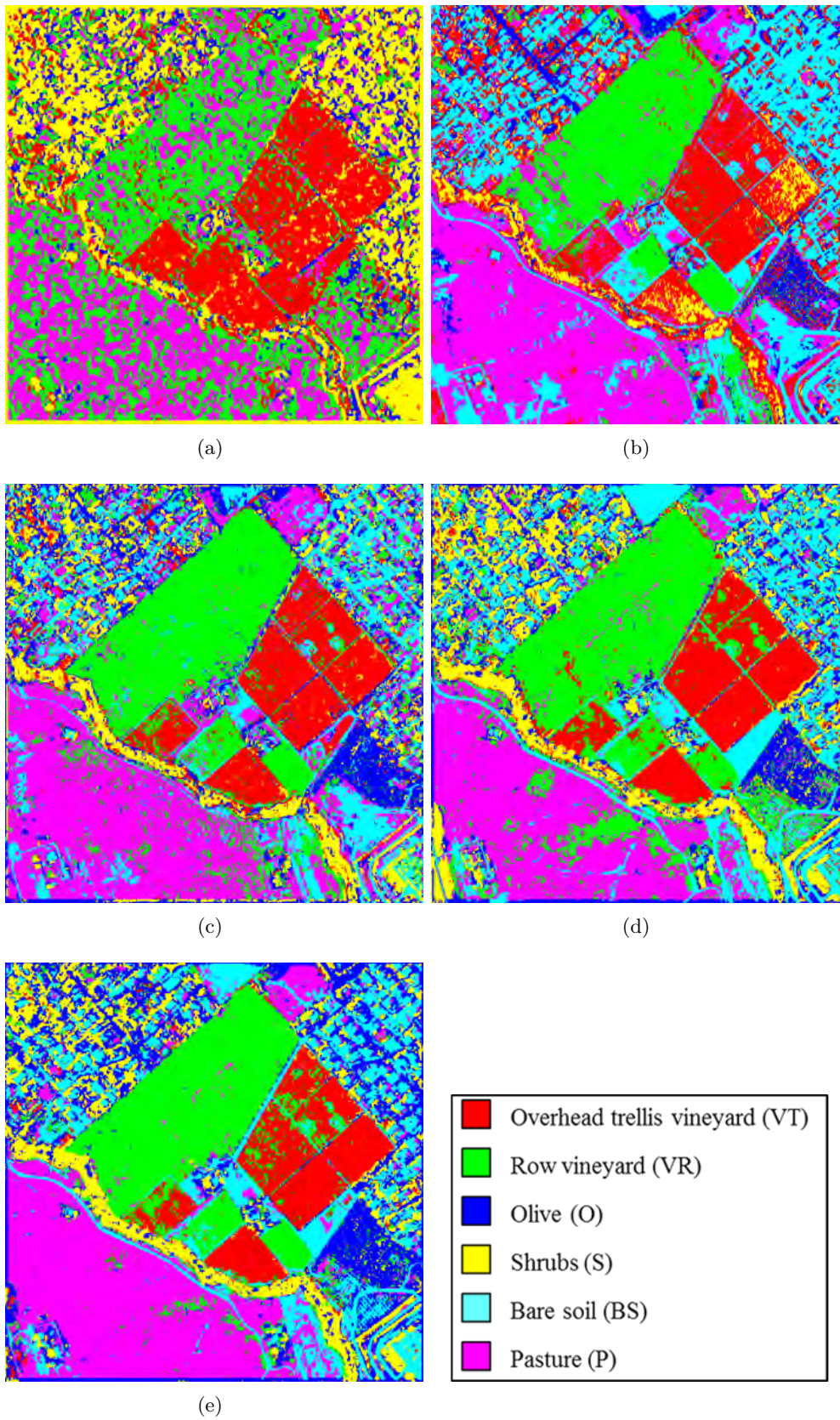
Figure 11.4. Vegetation maps achieved by combining different multi-sensors and multi-temporal data. a) C+D; b) A; c) A+C+D; d) B+F+H; e) A+B+C+D+E+F+G+H.

Part V

Unsupervised approach: object-based features extraction

Introduction

Very high spatial resolution satellite images are a powerful tool for the urban areas monitoring and planning or for the characterization of settlements. Moreover, the capability of COSMO-SkyMed SAR sensors of acquiring a huge amount of images over the same area with a short revisit time, irrespective of weather and light conditions, makes such data particularly useful for the near real-time assessment of damages occurring after natural disasters or human conflicts. For these reasons the development of automatic, fast and effective algorithms for buildings detection is an important and active research field. However, reaching such a goal can be rather difficult when dealing with VHR SAR images. Actually, the appearance of typical urban structures depends on several variables. For instance, materials of the walls, shapes and dimensions of the buildings, typology of roofs, affect the backscattered signal, which generally results from multiple contributions related to the elements belonging to the construction itself, and to different objects (e.g. trees, buildings) in its immediate surrounding. The viewing angle of the antenna, as well as the orientation of the buildings with respect to the SAR, has a strong impact on the objects imaging. Moreover, the geometry of acquisition of the system leads to typical distortions such as layover and shadowing. In very dense urban environments, buildings can be completely or partially hidden by the shadow caused by the presence of others close manmade structures.

In last decades, the interest of remote sensing community for buildings detection yielded a variety of methods, exploiting both optical and SAR imagery.
Focusing the attention on the latter kind of data, the majority of works in literature is based on the analysis of the information carried by interferometric SAR data or by multiple polarimetric images. In order to identify urban structures, a typical approach concerns the detection of edges for the building footprints reconstruction [86]. In [87] a proper edge detection algorithm is applied to multiple PolSAR data. The Hough transform is also taken into account to characterize the retrieved edges and generate footprints hypothesis. A further approach consists on the shadow areas analysis, which can give information about the building dimensions [88]. In [89] the authors developed an iterative technique for buildings reconstruction, achieved by the identification of the combined occurrence of a bright line (high backscattering value) and a shadowed region in an elevated area.

The interpretation of urban environments by exploiting a single VHR SAR image is more challenging, and few related works have been published so far. In [90] the understanding of built-up area is carried out by means of a stochastic geometrical model and a-posteriori probability maximization. An integration of the concepts of basic features extraction and their composition to buildings candidate, by considering the semantic meaning of the extracted features, is carried out in [91]. Different object extractors are fused with a coarse-to-fine approach in [51], where an image segmentation is followed by the selection of bright and dark lines; finally the interpretation of the urban area is performed by considering contextual knowledge.

The following chapter concerns the implementation of an unsupervised object-based technique to automatically detect buildings in VHR SAR images, acquired in a sub-urban

environment. The algorithm, which is based on the Pulse Coupled Neural Network model (see Chapter 4), has been originally developed for the identification of buildings in TerraSAR-X images. However, in the following sections, it will be demonstrated its efficiency in handling COSMO-SkyMed imagery, as well. The performance of the algorithm, which is described in Chapter 12, has been tested on three examples of VHR SAR images acquired on different areas in the Tor Vergata district, where the buildings are not very close each other, but differ for shape, dimension and orientation. In §12.1, the result obtained by testing the implemented technique on a Stripmap TerraSAR-X image (3 m spatial resolution), is showed and discussed in comparison with the output achieved by using another known edge detection method, based on Roberts convolution filter. Afterwards, two higher spatial resolution Spotlight COSMO-SkyMed images (1 m spatial resolution) have been exploited to confirm the capability of the algorithm to automatically identify buildings in a sub-urban scenario.

Chapter 12

Automatic PCNN based algorithm for buildings detection

As previously asserted, the PCNN architecture demonstrated to be a suitable tool for several remote sensing applications, mainly providing image segmentation or edge detection. However, even if the original model is based on an unsupervised approach, it requires the human intervention to establish when to stop the image processing, and to identify the best PCNN output, according with the specific goal to reach. In this study, the simultaneous analysis of the time signal G and of the resulting pulsing images, led to the formulation of an "*automatic stopping*" rule.

As highlighted in Chapter 4, the less complex is the image, the longer is the net progress period in which the synchronization property is kept, allowing the simultaneous pulsing of clusters of similar and close pixels (segments). Generally, X-band SAR imagery are very complex and affected by a strong speckle noise, which can drastically reduce the performance of the algorithm. Hence, as first step of the chain of processing, the adaptive despeckle filtering has to be applied, achieving smoother and more homogeneous objects.

A crucial point of the PCNN algorithm lies on the proper parameters setting. Since the PCNN is a complex and non-linear system, whose relationship between model parameters and network outputs is rather hard to understand, an empirical approach has been adopted to establish the most suitable variables. Finally it has been observed that a quite accurate buildings edge detection is performed by setting the PCNN parameters in Table 12.1.

The mode of inter-connection among neurons has been performed thorough two identical 3x3 feeding and linking kernels, whose elements values decrease with the distance from the central node. They are described by the following matrix:

Part of this chapter's contents is extracted from:

1. F. Del Frate, G. Licciardi, F. Pacifici, C. Pratola and D. Solimini, *Pulse Coupled Neural Network for automatic features extraction from COSMO-SkyMed and TerraSAR-X imagery*, Proc. IEEE, International Geoscience and Remote Sensing Symposium , IGARSS 2009, vol. 3, pp. 384-387, Jul. 2009

2. F. Del Frate, D. Latini, C. Pratola, *Automatic object extraction from VHR satellite SAR images using Pulse Coupled Neural Networks*, SPIE Remote Sensing Event, Toulouse, France, Sept. 2010.

3. F. Del Frate, D. Latini, C. Pratola, F. Palazzo, *PCNN for automatic segmentation and information extraction from X-band SAR imagery*, International Journal of Image and Data Fusion, vol. 4, no. 1, pp. 75-88, Sept. 2012.

   :

Table 12.1. PCNN parameters setting.

PCNN parameters	Value
α_F	0.5
α_L	1.0
α_θ	0.2
V_F	0.5
V_L	0.5
V_θ	20.0
β	0.1

$$M = W = \begin{bmatrix} \frac{1}{r} & 1 & \frac{1}{r} \\ 1 & 1 & 1 \\ \frac{1}{r} & 1 & \frac{1}{r} \end{bmatrix} \tag{12.1}$$

where r is the euclidean distance from the central neuron.

At the beginning of the PCNN processing all the functions, included the dynamic threshold, are initialized to zero. Therefore, for each neuron, the first iteration of the algorithm provides an internal activity only depending on the local stimulus, derived from the pixel value which is previously scaled in the $[0\div1]$ range:

$$F_{ij}[n = 1] = S_{ij} \tag{12.2}$$

$$L_{ij}[n = 1] = 0 \tag{12.3}$$

$$U_{ij}[n = 1] = S_{ij} \tag{12.4}$$

Being the local threshold equal to zero, all the neurons will fire at the first epoch. Afterward, θ increases up to V_θ. As the PCNN progresses, only the neurons with the highest internal activity, larger than the local threshold, will generate a pulse.

Such design of the PCNN scheme demonstrated to be suitable for the buildings detection in SAR images. Actually, since the manmade structures are represented by backscattering values which are in general higher than those coming from the background or different objects, the related pixels correspond to the first and more frequently pulsing neurons. Indeed, as the iterations progress, the autowaves emanate from the original pulse regions and the shapes of the buildings evolve through the epochs due to the pulsing nature of the PCNN. After few iterations, the net is able to detect the edges of the objects of interest. However, to make the chain of processing fully automatic a further analysis has been carried out. By analyzing the pulsing signature G of several examples, it has been observed that the best output image is always achieved in a moment between the epoch in which the time signal starts to increase from the zero value (t_0) and its first maximum value (t_{max}) (Fig.12.1). It has also been noted that from t_0 to the first G inflection point t_i (Fig.12.1), not significant changes occurs in the output images. On the basis of such observations the rule for the automatic PCNN stopping has been established.

Summarizing, the implemented PCNN method runs until the time signal reaches a maximum. Then the inflection point is evaluated, and the output achieved at the first iteration where G assumes a value greater than its mean, calculated within the interval $[t_0 \div t_i]$, is taken as the best output (t_b). The resulting image identifies the edges of the objects of interests. However, some mistakes can occurs, mainly due to the presence of artifacts within the SAR image, or caused by the high backscattering values of tall vegetation.
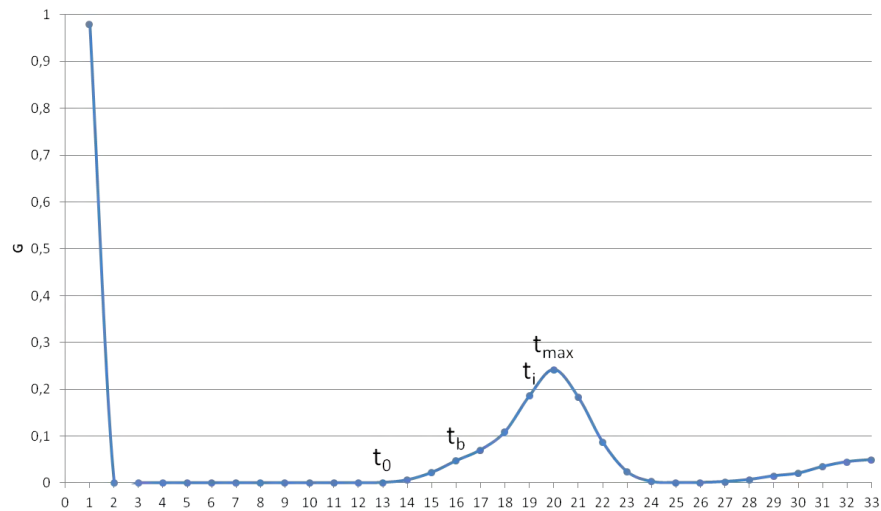
Figure 12.1. General trend of the PCNN time signal. Excluding from the analysis the first processing iteration, when G exhibits the highest possible value, t_0 is the first iteration in which G is larger than 0, t_{max} is the number of the epoch in which G reaches the first maximum value, t_i is the moment in which the G function presents an inflection point, and finally t_b represents the number of the iteration in which the best PCNN output is achieved.

In order to improve the accuracy of the output, the resulting PCNN image is filtered to remove those elements with perimeter values smaller than a certain threshold, viewed as not significant. The average dimension of ground truth objects has been taken as reference to this end. Finally, a region growing technique, using as seed points the pixels belonging to the edges of the survived objects, has been used to refine the previous results. Note that that this region growing technique has been applied considering the backscattering values of the despeckled SAR image. The whole chain of processing is schematised in Fig.12.2.
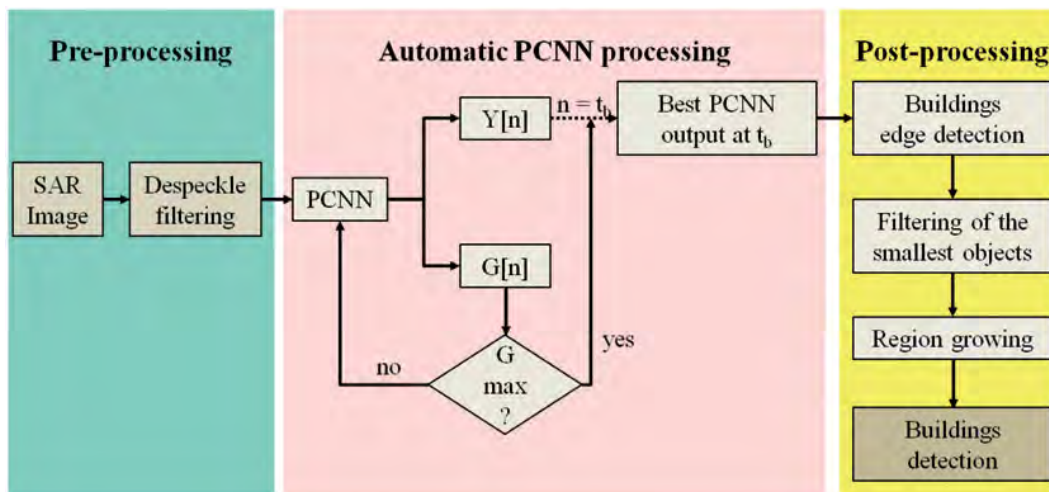
Figure 12.2. Fully automatic building detection algorithm.

In next section, the effectiveness of the technique will be demonstrated by showing and assessing the results obtained by exploiting Stripmap TerraSAR-X (3 m spatial resolution) and Spotlight COSMO-SkyMed (1 m spatial resolution) images.

12.1   Results

**Building detection in a Stripmap TerraSAR-X image**

A first experiment has been carried out by applying the fully automatic building detection algorithm to the Stripmap TerraSAR-X image shown in Fig.12.3(a). The SAR acquisition resulted to be hardly interpretable by a simple visual analysis. Actually, despite the despeckle processing, a residual noise affects the image. However, to avoid a loss of information concerning the buildings footprints, no further image enhancement method has been applied. Moreover, very high resolution effects, which include double bounces, layover, shadows, as well as the presence of other artificial objects (e.g. cars, lamps, guard rail), make challenging either the human interpretation and the automatic building detection. Therefore, the ground truth, represented by red ROIs in Fig.12.3(c), has been roughly taken on the Google Earth optical image in Fig. 12.3(b), where big buildings corresponding to shopping malls, business centers and science department of Tor Vergata University appear together with small houses in a mainly extra-urban landscape with large bare soil or uncultivated areas.

After the speckle filtering, the image has been processed by the PCNN algorithm, by setting the parameters as in Table 12.1. The progression of the net has been stopped after 100 iterations, and the wave signature has been plotted (Fig.12.4), while the pulsing outputs have been analyzed. The behavior of the time signal exhibits an almost periodic trend, providing cyclic similar outputs (Fig.12.5). However, as the PCNN progresses the pulsing objects synchronicity runs out and the resulting binary images do not contribute to any significant information. Nevertheless, by analyzing the results in Fig.12.5, it is possible to observe that buildings are identified after few iterations, before the first G peak is reached.

For the considered TerraSAR-X image, the algorithm implemented in this work automatically stopped after 22 iterations, identifying the best output at the 17th epoch. Hence, the aforementioned post-processing has been carried out, finally providing the result in Fig. 12.6(a).

Building extraction employing PCNN has also been compared with another known edge detection method, based on Roberts convolution filters [92] made available within the ENVI software processing libraries. In this case, the automatic PCNN processing module has been replaced with the non-automatic edge detection by the Roberts filter, while the pre-processing and post-processing steps were the same of the automatic algorithm. Fig. 12.6(b) represents the final building extraction.

In both the achieved results the main built structures in the scene are detected. A more quantitative evaluation of the accuracy of the algorithms can be carried out by comparing the output with the ground survey in terms of objects. In fact, a pixel-based analysis of the PCNN based technique is likely to give a low value of the overall accuracy, as most of the firing neurons, identifying an edge, are those characterized by a strong backscattering especially due to a double bound effect. Such an effect is usually stronger over some parts of man-made features and not over the entire buildings. It can be seen that, as far as the detection of the actual ground truth objects is concerned, the performance of the PCNN (58 out of 64 detected buildings) is comparable to that of the benchmark algorithm (61 out of 64 detected buildings). However, the difference of the results is significantly in favor of the PCNN if the number of false alarms is considered (Table. 12.2).
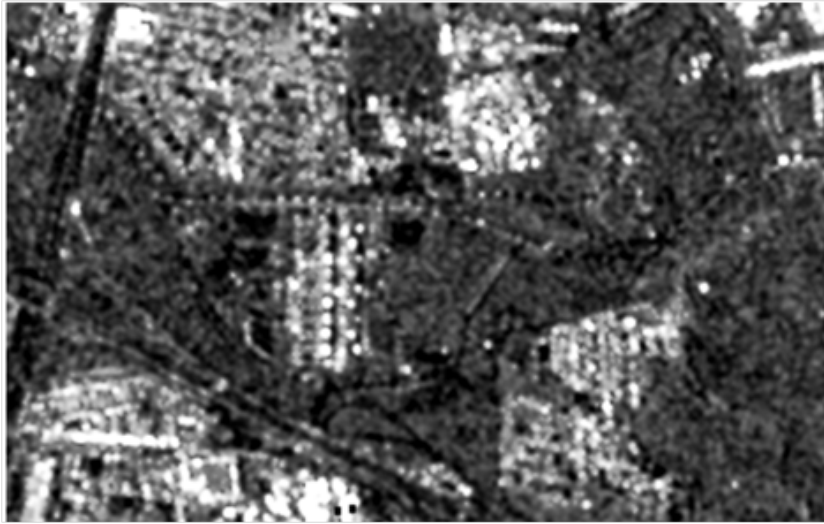
(a)

(b)
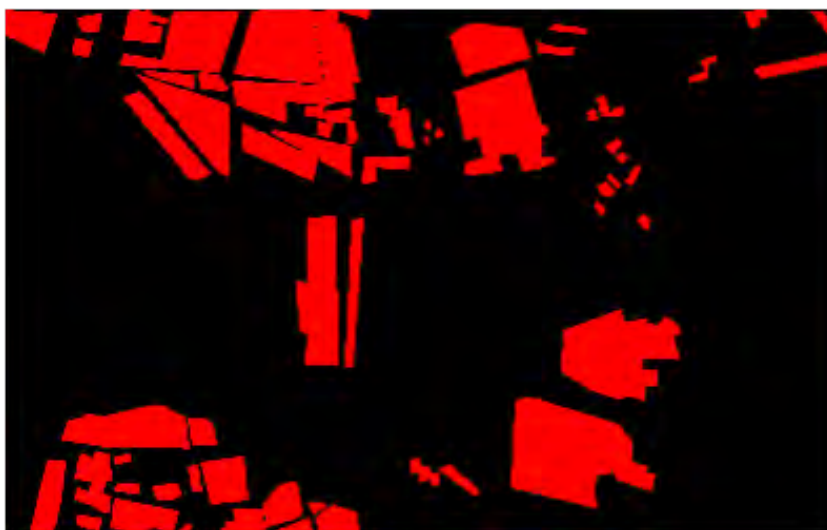
(c)

Figure 12.3. (Tor Vergata area, Rome (Italy) imaged by a) TerraSAR-X operating in StripMap mode (395x248 pixels), in HH polarization (date of acquisition: 24th November 2007). b) ©GoogleEarth optical image.c) Ground truth.
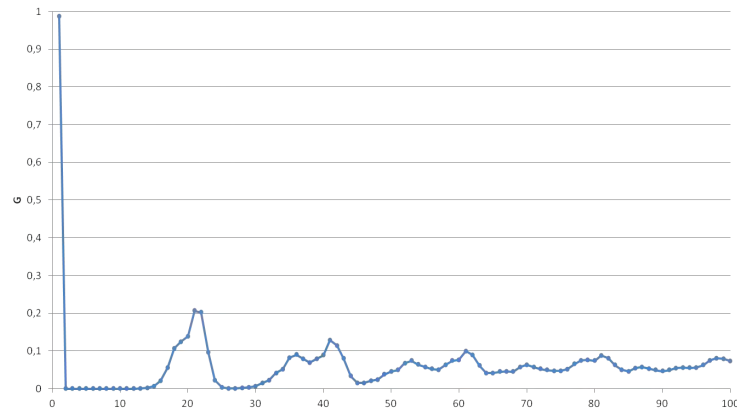
Figure 12.4. Time signal G referred to the PCNN processing of the TSX image.
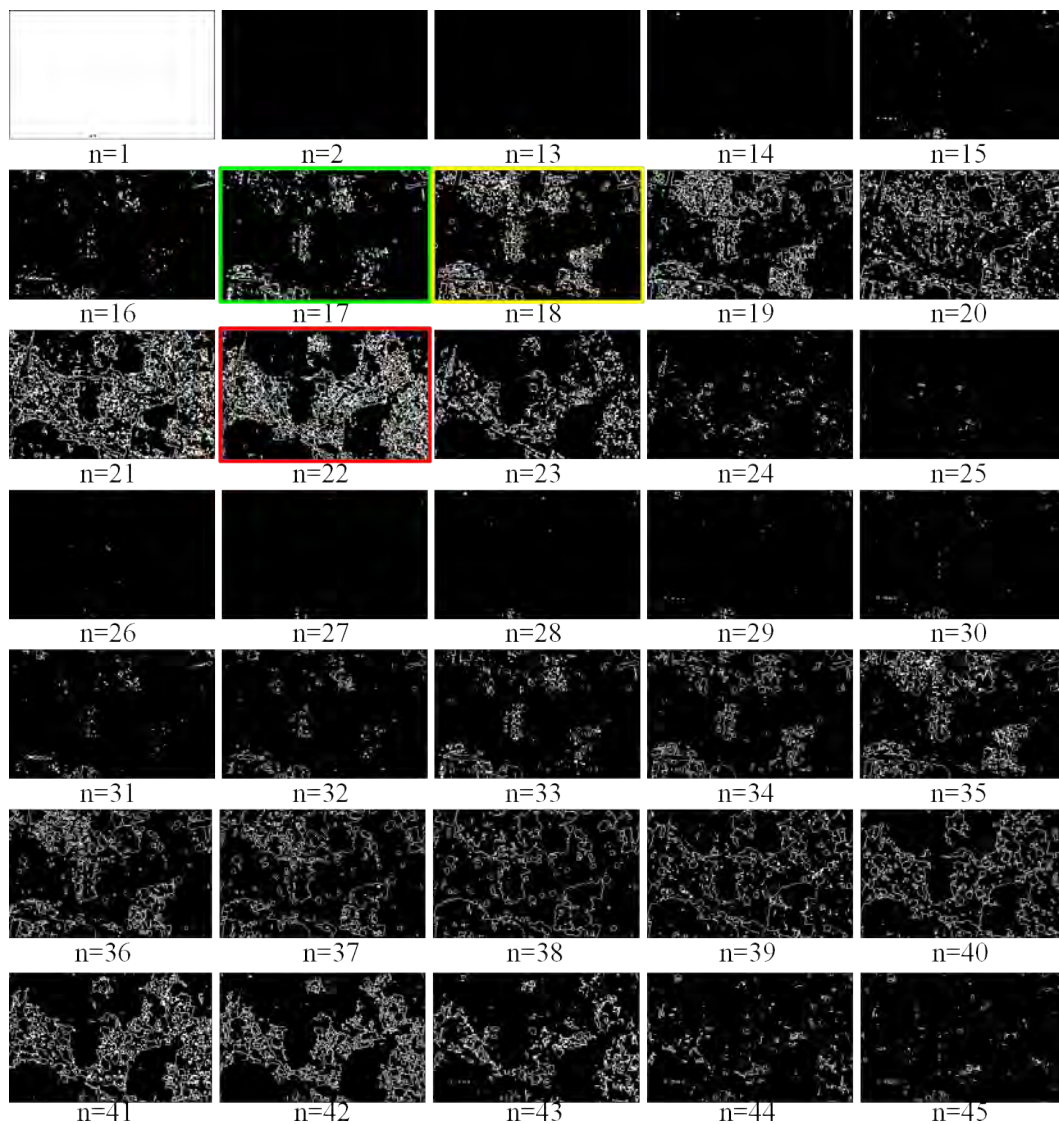
Figure 12.5. PCNN outputs. Red: last iteration of the automatic building detection algorithm. Yellow: PCNN iteration to which corresponds the G inflection point. Green: best PCNN output automatically selected.

(a)

(b)

Figure 12.6. Building detection results achieved by: a) the automatic PCNN based algorithm, b) the non automatic Roberts filter based technique.

Table 12.2. Comparison of the performance of the automatic PCNN based algorithm and the non-automatic technique based on Roberts edge detector.

	1 PCNN	2 Roberts
True positive	58	61
False positive	6	3
True negative	6	48
TOT detected	70	114
Ground truth	64	64

Building detection in Spotlight COSMO-SkyMed images

The fully automatic algorithm has been tested on different regions within the sub-urban area of Tor Vergata, represented in Fig.12.7(b) and Fig.12.7(c).

As a first exercise, the area in Fig.12.8(a) has been selected. It includes the Engineering, Letters and Business colleges of the Tor Vergata University of Rome.
After 21 epochs the PCNN processing stopped automatically, identifying the 17th binary image as the best result. Actually, other small objects are detected, many of which represent different man-made structures, such as lamps and guard-rail along the roads, or a metal railing along the underpass on the top right of the picture. After the post-processing filter and the region growing step, many of such "false alarms" have been removed, finally

(a)

(b)   (c)

Figure 12.7. Tor Vergata, Rome, Italy imaged by a) COSMO-SkyMed operating in Spotlight mode (date of acquisition: 2nd January 2011, dimension: 9700x5300 pixels). ©ASI. Rectangles represent the regions processed by the automatic PCNN based algorithm for buildings detection, and showed in the QuickBird frames in b) and c) (date of acquisition: 30th September 2010) ©DigitalGlobe.

achieving the result in Fig.12.9(c). In spite of some residual errors (37 false alarms), basically due to the established filtering threshold, whose choice has been based on a trade-off between accuracy of buildings detection and number of false alarms, the objects based assessment puts in evidence the effectiveness of the algorithm, since all the constructions have been automatically, quickly and correctly identified.

(a)   (b)

(c)
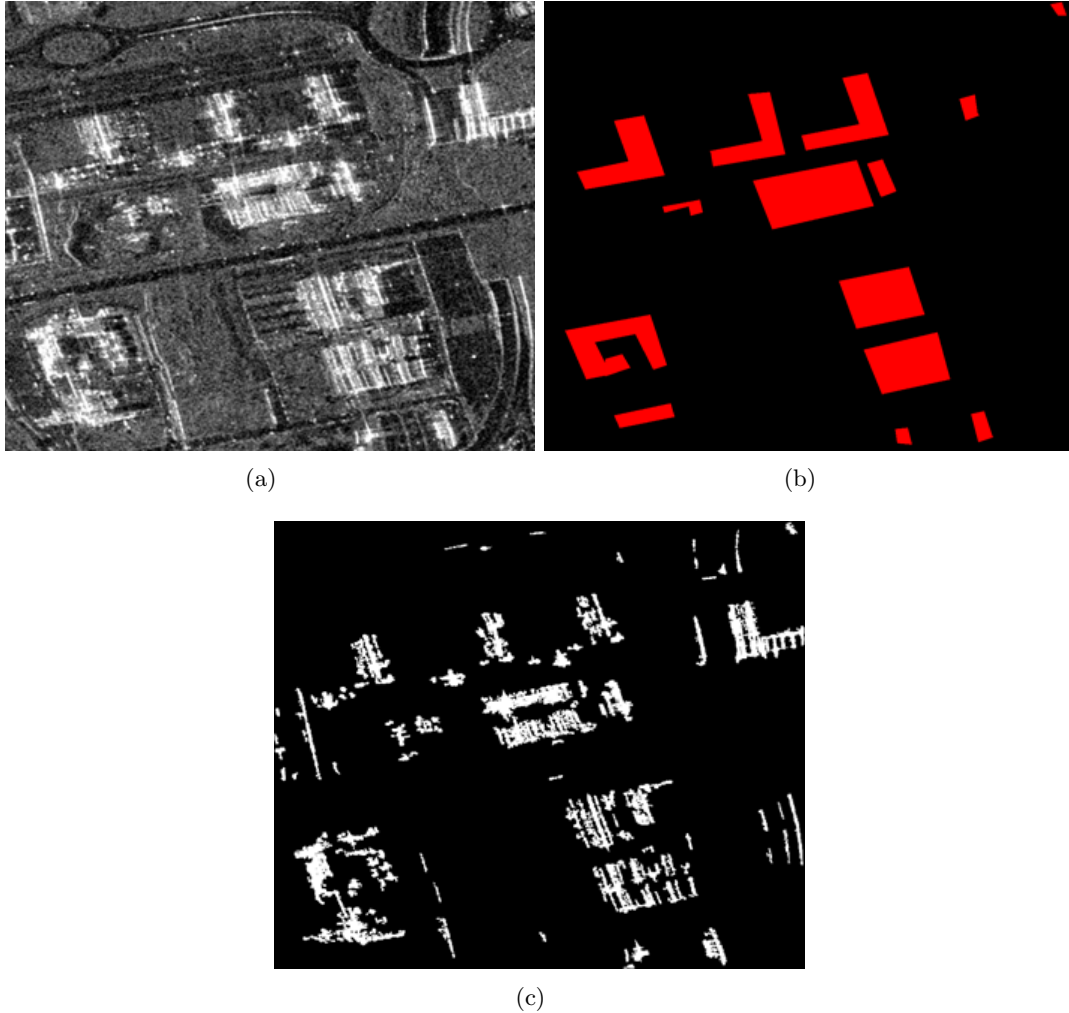
Figure 12.8. Tor Vergata University Campus imaged by a) COSMO-SkyMed, date of acquisition: 2nd January 2011 (1091x924 pixels) ©ASI; b) ground truth; c) building detection result.

A second test has been carried out by considering the region represented by the COSMO-SkyMed acquisition in Fig.12.9(a), including the complex of buildings of the Science Faculty of Tor Vergata University (top left), shopping malls (middle-bottom left) and a residential area (middle right). 22 PCNN iterations have been necessary to identify the 18th output as the best result. Given the variability of the buildings dimension, it has been necessary to set a smaller threshold in order to allow the identification of the smaller houses. In spite of the complexity of the scenario, such an approach led to the efficient detection of 57 out of 57 objects (Fig.12.9(c)). Nevertheless, larger number of false alarms (126) affected the overall accuracy of the final result.

(a)   (b)

(c)
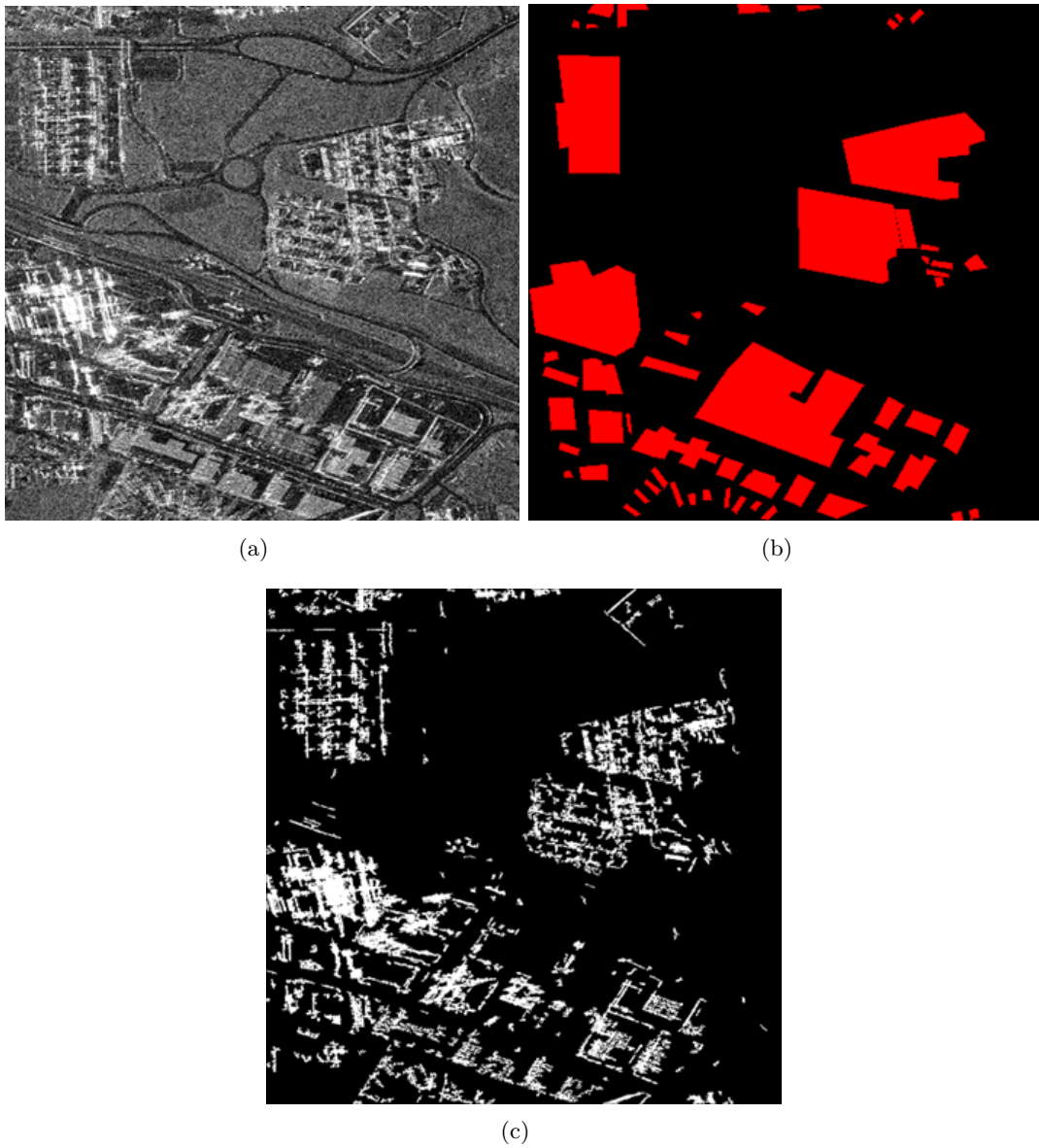
Figure 12.9. Tor Vergata area imaged by a) COSMO-SkyMed, date of acquisition: 2nd January 2011 (1730x1730 pixels) ©ASI; b) ground truth; c) building detection result.

12.2   Conclusions

In this chapter, the results obtained by using a PCNN technique for building detection have been presented. The efficiency of the PCNN model has been tested on different VHR satellite images acquired by either TerraSAR-X and COSMO-SkyMed. The performance of the implemented algorithm is interesting under different points of view if compared with standard alternative approaches. Firstly, the PCNN significantly increases the level of automation in the processing scheme, due to its unsupervised nature. Secondly, the accuracy of the results can be recognized as better, or at least comparable, with that of other suitable edge detection techniques. Indeed, in all the experiments more than the 90% of the buildings has been correctly detected. However, few residual objects have been also identified. Such false alarms, that mainly corresponds to different manmade structures or artifacts, depend on the dimension threshold, which has to be set in the post-processing phase in order to filter the smallest objects. Additionally, it has to be noted that the performance of the PCNN does not seem to be affected by the spatial resolution of the elaborated satellite SAR image. In fact, investigations carried out on a Stripmap TerraSAR-X (3 m spatial resolution) and two Spotlight COSMO-SkyMed images (1 m spatial resolution) showed that the PCNN approach owns good properties of robustness. Finally, the processing time is rather short, requiring only a few seconds for the buildings extraction.

Part VI

Change detection of sub-urban areas

Introduction

Social, economic and environmental conditions are leading to worldwide relocation of human populations, thus causing dramatic expansions of cities, with accompanying increased consumption of natural resources, disordered changes in land uses, worsening air and water quality and exposing people to severe dangers both of natural and of human origin. In such a scenario, the monitoring of land cover changes in urban areas has become a major problem for land management and security authorities. Satellite images offer quite effective monitoring means, thus providing valuable support for decision-making processes. As already highlighted, although remote observations in the optical part of the spectrum are generally used to monitor land cover and its changes, in order to meet the requirements of promptness, timeliness and reliability, use of SAR must be considered.

To take advantage of the unique capabilities of the CSK observing system, adequate exploitation of the information contained by the metric-resolution multi-temporal SAR images is necessary. In particular, the large amount of data contained in each image calls for the urgent development of suitable automatic techniques to manage in near-real time the information on land cover changes which are provided by the SAR observations.

In such a context change detection in multi-temporal acquisitions can be basically performed through two different approaches: supervised and unsupervised.

The algorithms based on a supervised approach are post-classification techniques, exploiting the knowledge of the area of interest to provide multi-temporal thematic maps, which are subsequently compared to identify modifications in land cover [93], [94]. The change detection accuracy strongly depends on the performance of the supervised classifiers, as the analysis of land cover transition must be preceded by the implementation of effective classification algorithms. Errors in the supervised detection of changes may be mainly attributed to unexploited pieces of temporal information contained in the images of the same area acquired at two different times. To overcome this problem, several authors propose multi-date supervised classification techniques, which simultaneously take into account related features in both images ([95], [96]), or they exploit the temporal correlation in a compound mapping [97] to directly identify the transitions from a type of land cover to another.

Regarding the most common unsupervised change detection methods, procedures can vary from the simple subtraction (or log-rationing in case of SAR data [98]), on a pixel basis, of two images acquired at different times, to more sophisticated techniques, such as those based on a Markov Random Field (MRF) approach [99], or on a combination of the use of a MRF and a maximum a posteriori probability (MAP) measure to achieve enhancement in change detection accuracy [100].

When dealing with VHR SAR data, change detection is more difficult due to the layover and shadowing that typically affect the accuracy of the co-registration of multi-temporal images. Moreover the speckle noise, which is particularly strong in the X-band VHR images, may produce false alarms in the change map. To overcome these problems, in [101] the authors apply a feature-based approach to correct the map of transitions previously obtained by an area-based method. In [102] an adaptive technique was developed to identify

modifications in VHR SAR images. On one hand, the method models the spatial-context information by considering homogeneous parcels and, on the other, it allows the analysis of changes in complex objects by exploiting theoretical similarity measures. In recent work [103] the authors addressed change detection in VHR COSMO-SkyMed images, by combining undecimated discrete wavelet transform (DWT), MRF and generalized Gaussians, resulting in an accuracy improvement in spite of the presence of speckle noise. Coherent change detection is another approach that exploits the magnitude of the complex cross correlation of a pair of interferometric images to identify changes. A coherence threshold is usually applied to highlight changed and unchanged areas. However, due to a bias in the coherence estimate, changed regions in COSMO-SkyMed images do not exhibit a complete loss of coherence [104]. As a consequence, setting the optimal threshold may be a difficult task.

In this chapter, a novel change detection method, based on the joint use of different neural networks architectures is presented and discussed. The effectiveness of neural networks (NNs) in classifying optical and SAR satellite images is confirmed by several authors [55–57, 105], and previously demonstrated in this thesis (see Chapter 10). Given the satisfactory performances of NNs, a simple post-classification comparison can provide change detection maps with an appreciably high accuracy [41]. Further improvement has been achieved through a more sophisticated architecture, called NAHIRI (Neural Architecture for HIgh Resolution Imagery) [106], where three MLP neural networks provide respectively the master land cover map, the slave land cover map, and a change mask. The last one is the output of a multi-temporal-information fed NN and is instrumental to correct the change map, stemmed from the classified master-slave comparison. Although NAHIRI was demonstrated to be an effective tool for VHR optical images, its structure might be less suitable for the analysis of changes in VHR SAR urban imagery. Indeed, first of all, the information in Spotlight and Stripmap COSMO-SkyMed images is carried by a single polarization, then the aforementioned layover and shadowing in urban environment acquisitions, as well as the strong speckle, join to make hard the attainment of high classification accuracies. Use of log-ratio of the intensity of backscattering and of the textural features ([39–41, 70]), or the inclusion of the coherence information into the input vector [69], is generally not sufficient to overcome the errors caused by the pixel misregistration in correspondence of buildings. Moreover, in a suburban environment, where built-up areas alternate with more or less wide vegetated zones, the degree of interferometric coherence ([69], [107]) may lead to confusion among changed and unchanged land covers.

Given the above drawbacks, Pulse Coupled Neural Networks have been addressed in this study, and an investigation has been carried out on the possibility of inserting them into the change detection scheme. Indeed, PCNNs are commonly applied in image segmentation, edge detection [28–30] or objects extraction [31], [32], but they were found to perform successfully also in VHR optical images change detection [34]. In principle, the algorithm generates, step by step in an iterative scheme, a specific signature of the scene, depending both on the values associated with single pixels and on the contextual information. A measure of the correlation between the preceding and the subsequent signatures is able to suggest intervened changes. On this background, a novel algorithm based on the combined use of MLP-NNs and of unsupervised PCNNs has been designed, implemented and tested on VHR COSMO-SkyMed images to monitor changes in land cover. The more traditional Post Classification Comparison (PCC) technique has been considered as a benchmark for the performance analysis. The implemented algorithm, which has been applied to three examples referred to one Stripmap image acquisition, and two Spotlight frames, will be described in the following chapter. Results achieved after each step of the chain of processing will be showed and discussed in next sections.

Chapter 13

Automatic change detection algorithm

In this work a novel change detection algorithm has been designed to investigate on the changes occurred in a suburban environment, exploiting VHR multi-temporal data delivered from COSMO-SkyMed constellation.

The architecture of the presented change detection algorithm has been derived from the NAHIRI model developed in [106], which proved its efficiency on VHR optical imagery processing. However, the idea of running multiple neural networks has been modified to work with the VHR SAR data exploitation. Actually, although the MLP-NN may be still a valuable method for the land cover characterization, its usefulness may decay when the desired output is a change mask. In this case the net could erroneously associate to changed regions those areas which are affected by a non perfect co-registration of the old and new images, or vegetated surfaces which are characterized by low values of the interferometric coherence, or artifacts which are rather common in CSK images. For these reasons the third MLP-NN of the NAHIRI model has been replaced by a double image processing performed by the automatic and unsupervised PCNN algorithm. A scheme of the implemented change detection system is showed in Fig.13.1.

The input to the whole scheme consists of a couple of images, acquired at two different times, which are previously orthorectified, despeckled by means of an adaptive filter, i.e. the enhanced Lee filter, and co-registered. Basically four image processing NN algorithms run in a parallel way: two of them are aimed to the production of land cover maps on a pixel basis, while the others provide the localization of "hot spots" [34], where changes probably happened.
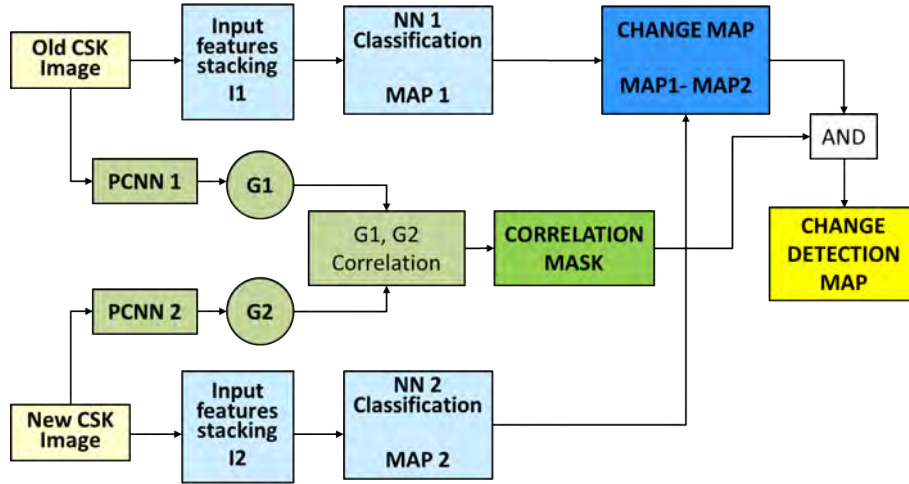
More in detail, a supervised approach is adopted to obtain the old and the new thematic maps, called *MAP1* and *MAP2* respectively. To this end the training and validation patterns must be collected within each image, forming the sets of input vectors (*I1* and *I2*), that will successively feed the MLP-NNs. Since the considered acquisitions are taken over

---

Figure 13.1. Flowchart of the change detection algorithm. The old and new images are firstly orthorectified, despeckled and co-registered.

the same area of interest by the same kind of SAR sensor and with the same polarization and view angle, the topology of both *NN1* and *NN2* is identical, having an equal number and type of input features and output classes. A simple post-classification step follows: a comparison between the achieved MAP1 and MAP2 puts in evidence the changes in land uses. However false alarms, caused by the aforementioned limits of the X-band VHR SAR images, are likely to occur.

The duty of the second unsupervised module is the reduction of these errors and the enhancement of the accuracy of the change detection map. Such a compartment includes two PCNNs, one for each image (*PCNN1* and *PCNN2*), whose neurons receive the intensity of the backscattering of the corresponding pixels of the image as external signal S_{ij}. As explained in Chapter 4, the PCNN yields binary outputs, with a pulsing behavior of the neurons, depending on several parameters. By focusing on the possible changes in built-up areas, a preliminary study has been carried out, aiming at the research of the best configuration of the variables, which may make the PCNN able to identify objects within VHR X-band SAR images. The mode of inter-connection among the neurons has been performed through two identical 3x3 feeding and linking kernels, whose elements values decrease with the distance from the central node. They are described by the following matrix:

$$M = W = \begin{bmatrix} \frac{1}{r} & 1 & \frac{1}{r} \\ 1 & 1 & 1 \\ \frac{1}{r} & 1 & \frac{1}{r} \end{bmatrix} \tag{13.1}$$

An empirical approach has been adopted by varying the parameters values and assessing the PCNN object detection performance. Once all the variables have been properly set (Table 13.1), the PCNN was ready to be used for the change detection of every pair of VHR X-band SAR images, without any further human intervention.

The idea of identifying significantly changed regions by exploiting the PCNN model, originates from the measurement of the similarity between the PCNN signals associated to the old and new image respectively. To this end, it is necessary to convert the pulse imagery into a single component information, which is the time signal G, defined by the expression (4.6). In the developed scheme, the time signals *G1* and *G2* are evaluated over a certain number of iterations, and a measure of the similarity between the *PCNN1* and *PCNN2* pulsing behavior is given by the correlation value.

When a significant change occurs, the combined use of the PCNNs allows to catch those areas, as they exhibit low correlation values. In Fig.13.2 two examples of changed and

Table 13.1. PCNN parameters setting.

PCNN parameters	Value
α_F	0.1
α_L	1.0
α_θ	0.2
V_F	0.5
V_L	0.5
V_θ	20.0
β	0.1
θ_n	1

non changed scenes are showed. The pulsing imagery yielded by the PCNN algorithm has been converted into the time signal G, whose trend is plotted in Fig.13.3. While the pulsing behavior of the images in Fig.13.2(a) is rather different form the very first epochs (Fig.13.3(a)), the trend of the G functions in Fig.13.3(b) is almost identical. This is confirmed by the computation of the correlation value between the time signals related to the old and new acquisitions. For the "hot spots" example, it quickly decays in about 30 iterations, reaching a value lower than 0.7. When no changes occur, the correlation value does not vary significantly, being in the 0.96÷1 range. Several tests over changed and unchanged scenes, sampled from different datasets (VHR SAR and optical multi-temporal acquisitions [34]), confirmed that a correlation value lower than 0.75 occurs in presence of "hot spots".

(a)

(b)
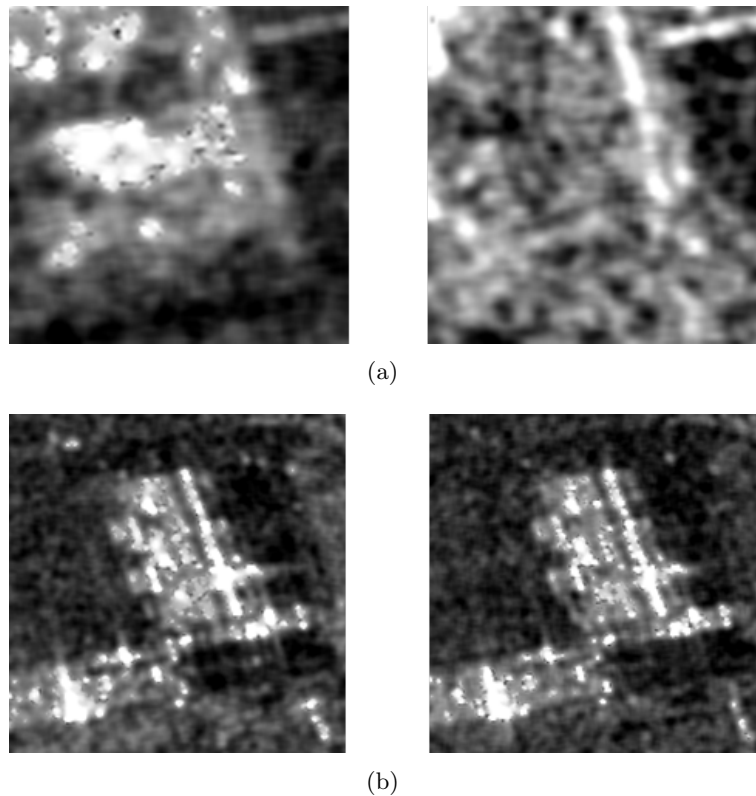
Figure 13.2. Examples of two test sites imaged at two different times (left: old, right: new). a) "Hot spot" area where a changed occurred. b) Non changed area.

Aiming at a finer but fast detection, a PCNN approach based on the pursuit of changed
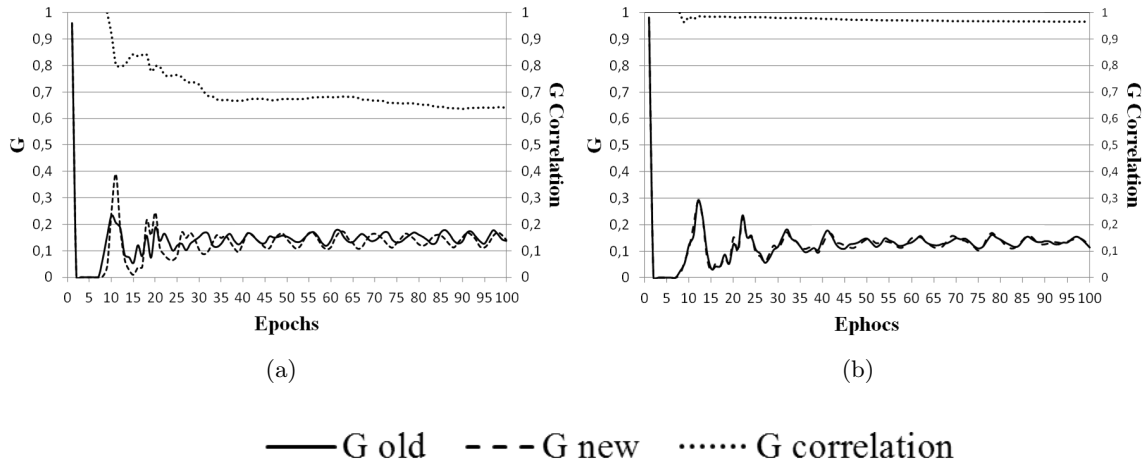
Figure 13.3. PCNN signals of the images acquired at two different times, and trend of the G correlation value. a) "Hot spot" example; b) Non changed area example.

areas is carried out. This is possible by running the PCNN on sub-images extracted by a moving window, whose dimension has to be established according to the spatial resolution of the images and the kind and/or extent of changes to be detected. Afterward, the boxes of pixels which exhibit a correlation measure lower than the established threshold are classified as "hot spot", which are located in a correlation mask. Finally, only the objects that are identified by the change map and, at the same time, totally or partially included within these "hot spots", are highlighted in the resulting change detection map.

13.1 Dataset

The designed change detection system has been tested on pairs of VHR X-band COSMO-SkyMed images acquired at different times on Tor Vergata site (Fig.13.4).

In this work Stripmap (3 meters spatial resolution) and Spotlight (1 meter spatial resolution) pairs of images have been exploited.

As regards the former, the available data were acquired on 20th September 2008, and on 22nd April 2011 by CSK-2. Both the images were taken in HH polarization and with the same geometry of acquisition (see Chapter 7). In Fig.13.5 the processed test site, is shown in a false color composition image, together with the ground truth, which is highlighted on a couple of temporally consistent optical images. The false color composition in Fig.13.5(a) highlights the changes which occurred during the considered time interval. In principle, transition from natural areas to artificial surfaces appear in blue, whereas the opposite changes result in yellow objects.

The Spotlight dataset consists of two frames shown in Fig.13.6 and Fig.13.7, respectively extracted from an image acquired on 9th July 2010 by CSK-2, and from a more recent one taken on 11th June 2011 by CSK-3, both in HH polarization and looking on the left side on the descending orbit (see Chapter 7). Again, the false color composition images allow a prompt identification of the occurred changes within the scenes.

13.2 Results

In this section there will be shown the achievement of the application of the designed processing chain on the 3 meters spatial resolution Stripmap images, and the 1 meter
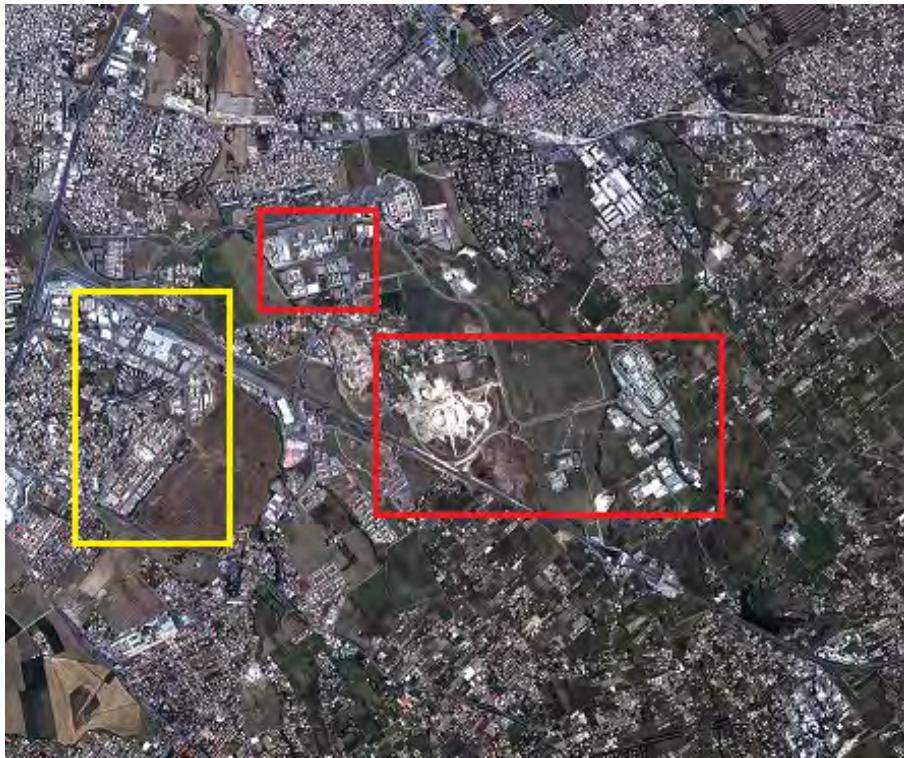
Figure 13.4. WorldView-2 image of Tor Vergata, Rome, Italy. Date of acquisition: 4th July 2010, ©Digital Globe. Rectangles refer to the sites of interest selected for the change detection algorithm test (Yellow: Stripmap frame; Red: Spotlight frames).

spatial resolution Spotlight ones. The photo-interpretation of multi-temporal optical images allowed to identify those areas where changes occurred within the considered time intervals. Consequently, the sub-image in Fig.13.5, and two frames in Fig.13.6 and Fig.13.7, have been selected for the first and second exercise, respectively. Since the urbanization process mainly led to the replacement of natural surfaces with built-up areas, the study has been focused only on such land-use modifications.

In the following paragraphs the change detection scheme is divided into three modules: the supervised pixel-based classification, the unsupervised identification of "hot spots" by the PCNNs, and the final mapping of changes.

Change detection by exploiting Stripmap COSMO-SkyMed images

Pixel-based classification by means of MLP-NNs

The goal of the pixel-based classification phase is to discriminate between four classes: *asphalt* (roads and parking areas), *low vegetated* surfaces (including short grass and bare soil), *trees* (also including tall plants, vineyards, olives and shrubs) and *manmade* structures. On the basis of the analysis already discussed in Chapter cap: classification, the land cover identification has been performed by feeding the MLP-NNs by a set of input features taking into account, besides the intensity of the backscattering, also textural parameters stemmed from the Gray Level Co-occurrence (*GLCM*) matrices computation.

Since the land cover classification plays an important role within the change detection chain of processing, a particular attention has been given to the MLP-NN training phase. On the basis of the outcome of the experiments already illustrated in §10.2, both the NN1 and NN2 of the change detection algorithm have been trained by the 1800 samples per class. The topology of the neural network optimal with respect to the MSE minimization has been,
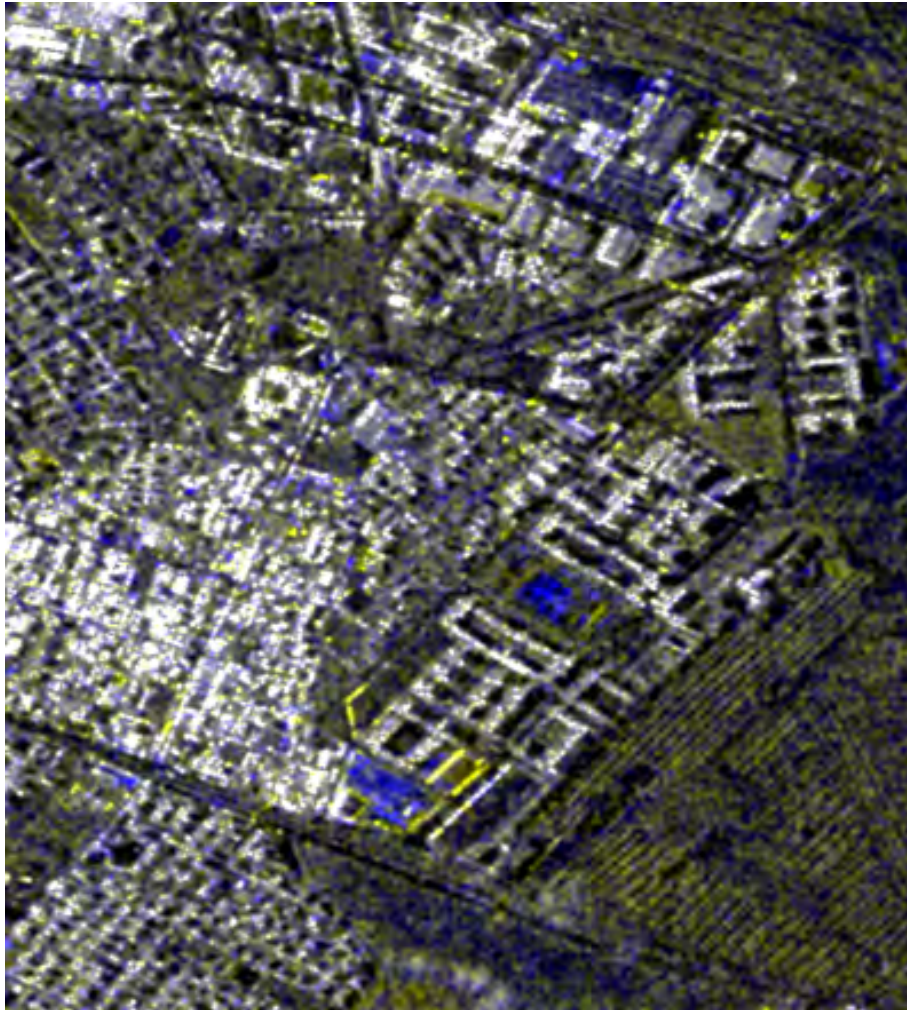
(a)

(b)   (c)

Figure 13.5. COSMO-SkyMed Stripmap dataset ©ASI (810 X 900 pixels): a) False color composition: R and G: 20th September 2008 acquisition; B: 22nd April 2011 acquisition; b) 8th December 2008 QuickBird image ©Digital Globe; c) 12th April 2011 QuickBird image ©Digital Globe. Red circles represent the ground truth of changed objects.
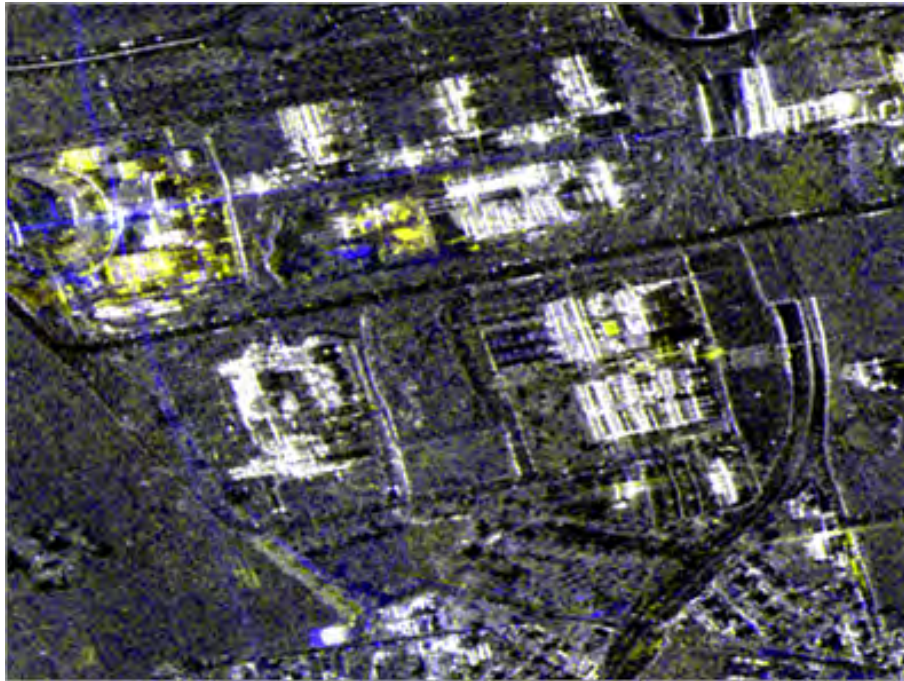
Figure 13.6. First COSMO-SkyMed Spotlight dataset ©ASI (1546 X 1156 pixels): a) False color composition: R and G: 9th July 2010 acquisition; B: 11th June 2011 acquisition. b) 4th July 2010 WorldView-2 image ©Digital Globe; c) 20th May 2011 QuickBird image ©Digital Globe. Red circles represent the ground truth of changed objects.

(a)

(b)

(c)

Figure 13.7. Second COSMO-SkyMed Spotlight dataset ©ASI (4319 X 2083 pixels): a) False color composition: R and G: 9th July 2010 acquisition; B: 11th June 2011 acquisition. b) 4th July 2010 WorldView-2 image ©Digital Globe; c) 20th May 2011 QuickBird image ©Digital Globe. Red circles represent the ground truth of changed objects.

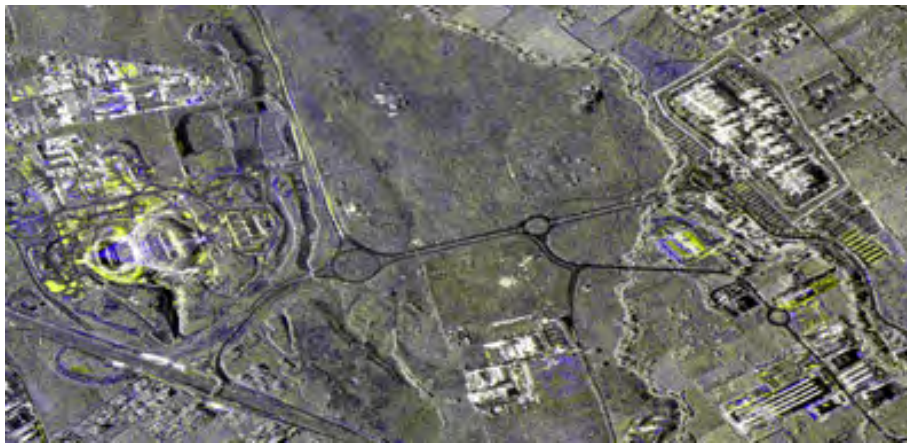for both the images, the one with two hidden layers containing 10 neurons each(5x10x10x4). The so trained *NN1* and *NN2* have been successively fed by the whole images, providing the *MAP1* and *MAP2*. After the selection of the test site in Fig.13.5, the *manmade* class has been extracted from each map, and a comparison has been worked out to produce the change map in Fig.13.8. It has to be noted that the test site is characterized by a dense built-up area. Here the complexity of the received backscattering signal, arising from multiple contributions coming from different elements (buildings, roads, trees, gardens) which are very close each other, makes the understanding of the scene particularly hard. Hence misclassification errors are likely to occur. Moreover, in such urban scenario, the layover and shadowing effects decrease the precision of the co-registration. As a consequence a large amount of errors can be observed in the post classification change detection.

Figure 13.8. Change map referred to the Stripmap images pair in Fig.13.5(a).

"Hot spots" identification by PCNNs

As previously asserted, the goal of the unsupervised PCNN module, is the enhancement of the pixel-based change map, through the reduction of the number of false alarms. To this end, each image has been processed by the PCNNs, applied to sliding windows, whose size of 45x45 pixels has been established according to the CSK spatial resolution (3 meters), to the extent of the possible changes, and to the dimension of the buildings.
According with the phenomenon of the loss of correlation between the time signals generated by the PCNNs, a threshold of 0.75 has been established to select the "hot spots". The correlation mask depicted in Fig.13.9 is the output of the automatic and unsupervised module of the change detection chain of processing.

Change detection map

The last step of the change detection chain of processing is concerned with the combination, through the logical operator "and", of the *change map* and the*correlation mask*, resulting from the supervised processing and the unsupervised one respectively. It means that all the objects identified by the change map, which are completely or in part included into

Figure 13.9. Correlation mask achieved by setting the threshold value of the G function correlation to 0.75.

the "hot spots", are classified as changes, while the others are removed. The comparison between the change map and the changed areas is represented in Fig.13.10(a). A post-processing has been also carried out to filter the single scattered pixels or the smallest objects (Fig.13.10(b)). The assessment of the result has been carried out on an object basis, by means of a cross-check against the ground truth highlighted in the red circles in the QuickBird images in Fig.13.5. The final change detection map in Fig.13.10(b) proves that this technique is suitable to enhance the accuracy of the change detection, in fact 6 out of 9 real changes have been recognized, while the amount of false alarms is drastically decreased to 38 objects. Note that, if the same filtering was directly applied to the simple PCC result, the number of false positives would remain quite high, equal to 98.

The analysis of the three missing objects showed in Fig.13.11 puts in evidence the reasons which led to the error of the algorithm. The change in Fig.13.11(a) is recognized by both the change map and the correlation mask, being the correlation value equal to 0.65. However, because of the object dimension, it has been removed from the detection map after the post-processing. By analyzing the other two examples, a non significant difference in terms of backscattering value and texture can be observed between the old and new acquisition. Such a behavior has led to rather correlated pulsing signals during the PCNN processing. Indeed the objects in Fig.13.11(b) and Fig.13.11(c) are included in a box characterized by a correlation value of about 0.83 and 0.92, respectively. This is the reason why their "hot spots" have been excluded from the correlation mask, and therefore from the final change detection map.

Change detection by exploiting Spotlight COSMO-SkyMed images

Pixel-based classification by means of MLP-NNs

As far as concern the pixel-based classification of the pair of Spotlight images, the results already discussed in § for the 9700x5300 pixels imagery, have been considered to generate the change maps of the frames under investigation.

In spite of the rather good correlation between the *manmade* class and the ground truth provided by the optical acquisitions, some confusion can be observed especially in corre-spondence of the tall vegetation. Moreover a slight spatial shift between the old and new

(a) (b)
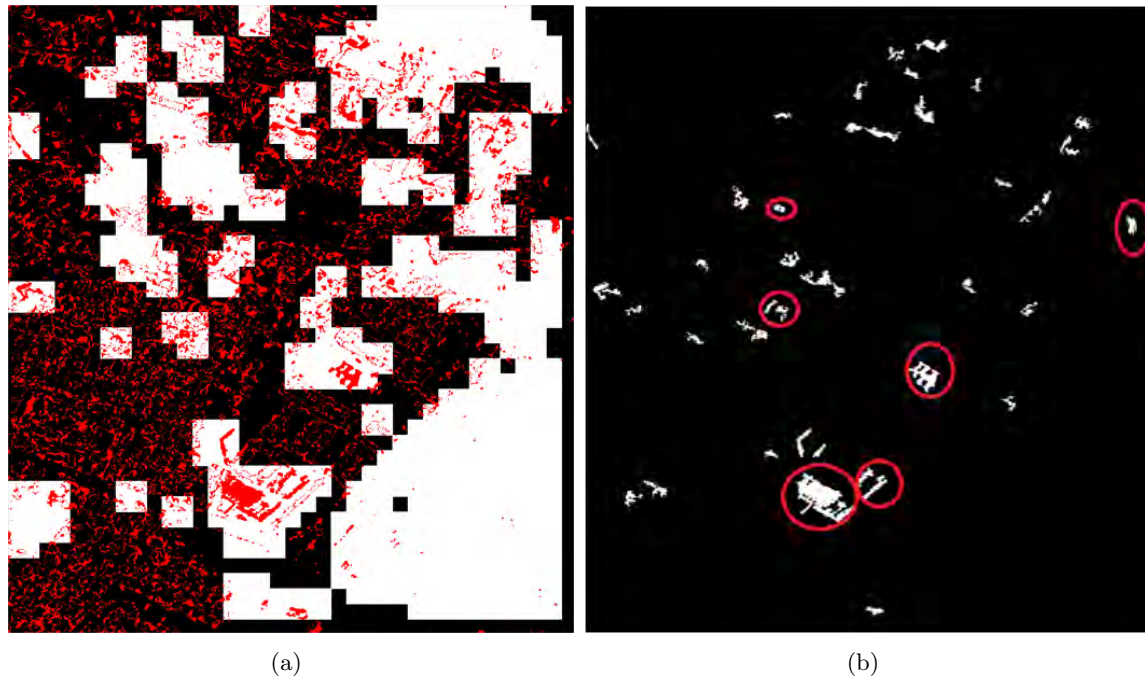
Figure 13.10. a) Change map (red) and Correlation mask (white) comparison. b) Change detection map. Red circles put in evidence the correct changed objects.

(a)

(b)

(c)
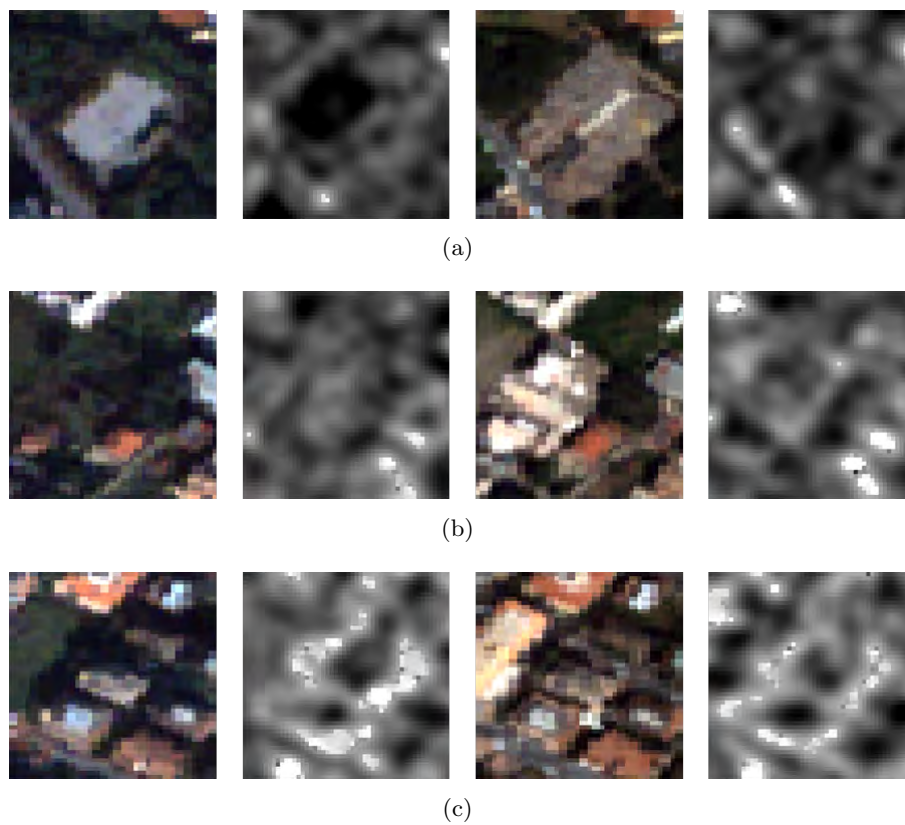
Figure 13.11. Three missing changes (a-c) in the change map achieved by exploiting the Stripmap images pair. QuickBird and COSMO-SkyMed images are distributed in columns in a temporal order.

images yields false alarms, as can be noted in Fig.13.12 and Fig.13.13, where the edges of the most of the buildings are classified as changes.
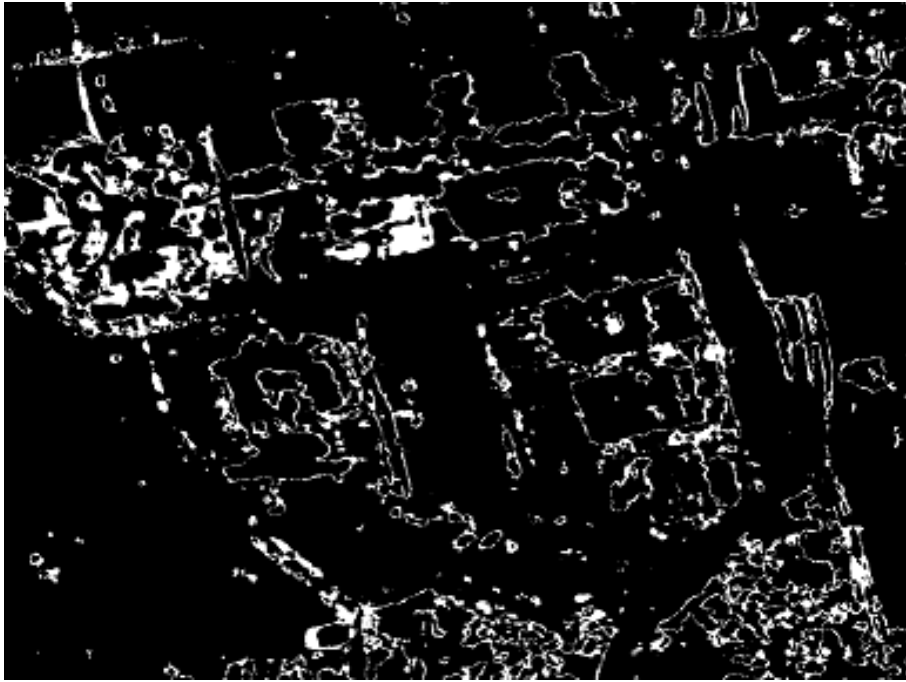
Figure 13.12. Change map referred to the Spotlight images pair in Fig.13.6(a).
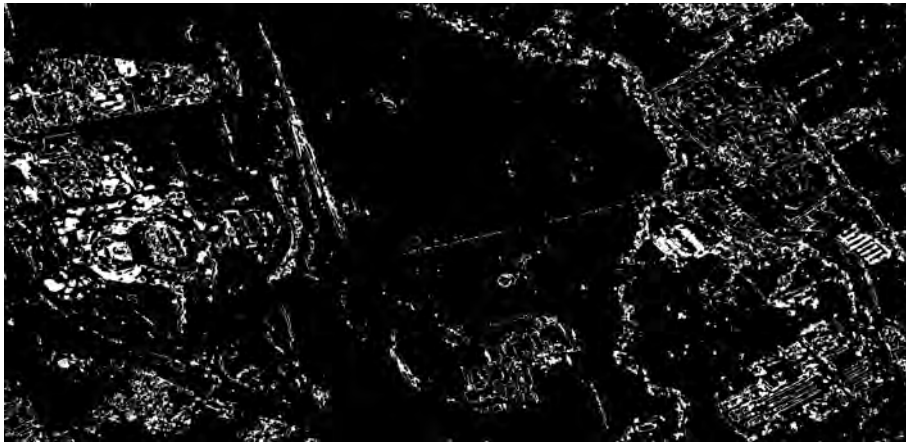
Figure 13.13. Change map referred to the Spotlight images pair in Fig.13.7(a).

"Hot spots" identification by PCNNs

According to the very high spatial resolution of the data (1 meter), and considering the extent of the changed objects, the PCNN algorithm has been applied to sub-images by means of a sliding window of 100x100 pixels, identifying "hot spots" of the same dimension when the G function correlation value is lower than 0.75.

Change detection map

The comparison between the change map and the changed areas, followed by a post-processing, provided the change maps represented in Fig.13.14 and in Fig.13.15. The object

based assessment of the results confirmed the rather good performance of the change detection system. Actually, in both the examples, the algorithm recognized all the changes, considerably reducing the amount of false positives. It should be noted that both the analyzed areas include construction sites (top left in Fig.13.6 and middle left in Fig.13.7), where the presence of vehicles, cranes, scaffoldings, containers, can vary continuously during the period of work; for this reason all the little changes detected by the algorithm might be correct. Elsewhere, in the middle and bottom right of Fig.13.7, there are several parking areas, where the presence or absence of cars produces a positive output in the final map of changes. Moreover, because of the strong speckle noise, some errors in land cover maps, as well as artifacts (especially visible in Fig.

Figure 13.14. Change detection map referred to the Spotlight images pair in Fig.13.6(a). Red circles put in evidence the correct changed objects.

Figure 13.15. Change detection map referred to the Spotlight images pair in Fig.13.7(a). Red circles put in evidence the correct changed objects.

Table 13.2. Number of false alarms detected by the PCC and by the implemented change detection algorithm.

Types of image pair	PCC	MLP+PCNN Change detection
Stripmap	98	38
Spotlight Ex.1	66	28
Spotlight Ex.2	1157	227

13.3   Conclusions

This work aims at developing fast, automatic and accurate algorithms for detecting changes of land cover from VHR SAR X-band images, such as those provided by recent Earth Observation missions.

The method is based on two different NN architectures: a supervised MLP-NN provides the land cover map for each acquisition, whereas an unsupervised and automatic PCNN model allows identifying "hot spots" by measuring the correlation value between the pulsing signature of the signals generated from each image. A comparison between the change map obtained by a post-classification analysis and the correlation mask produced by the PCNN module yields the final change map. Further post-processing is useful to enhance the accuracy of the resulting product.

The change detection scheme has been tested on the urbanization process occurring in the Tor Vergata University area. To this end a first experiment has been carried out on an area imaged at 3 meters spatial resolution by the COSMO-SkyMed X-band SAR in Stripmap mode at a time interval of about three years. The exercise has been repeated on two different portions of two COSMO-SkyMed Spotlight images acquired at 1 meter spatial resolution in 2010 and 2011, respectively.

Six out of nine true changes have been detected from the pair of Stripmap images, while all the transitions into the *manmade* class have been identified from the Spotlight data. It is worth highlighting that the inclusion of the PCNN into the change detection processing chain resulted in enhanced robustness against co-registration errors.

Among the advantages brought by joining the MLP and PCNN modules, it has to be especially mentioned the reduction of the number of false alarms (mainly caused by misclassification, misalignment, artifacts and speckle) if a comparison with the standard PCC is performed (Table 13.2). In addition, it has to point out that the end-to-end algorithm is quite fast and the whole processing chain can be readily designed for a fully automatic implementation of the change detection between pairs of any land cover classes.

Conclusions

The recent advent of the last generation of VHR SAR systems opened new challenges in remote sensing field. In particular, the unique capabilities of the COSMO-SkyMed constellation of acquiring images at metrical spatial resolution and with a very short revisit time, make this system especially appealing as a possible efficient tool for many Earth Observation applications, from risk management, to damage assessment or environment monitoring. However, the novelties brought by COSMO-SkyMed are leading the scientific community to intensively investigate on the most suitable methods to retrieve information from the still partially unknown X-band SAR data, by addressing the issues related to the very high spatial resolution, as well as to the availability of single polarized images. Moreover, the large amount of data calls for the urgent development of efficient automatic techniques to manage in near-real time the information on land cover and its changes, which is of vital importance in case of emergency.

In such a scenario, this study has been carried out with the aim of providing a significant contribution toward the development of fast, automatic and accurate algorithms performing the characterization of a sub-urban environment, through the land cover classification, building identification and detection of eventual changes. Basically, the features extraction has been worked out through both pixels-based techniques characterized by a supervised approach, and object-based techniques, performed by unsupervised algorithms.

A pixel-based approach has been adopted for the classification of single polarized Spotlight and Stripamp COSMO-SkyMed images of the Tor Vergata outskirt of Rome, Italy. The method, based on the MLP-NN model, makes use of the backscattering information content, as well as contextual and/or textural properties of the explored types of surfaces. As a matter of fact, even the exploitation of simple local texture, such as the *local mean* and *local standard deviation* evaluated on small boxes of pixels, proved to allow the discrimination between asphalted surfaces, natural areas, and manmade structures. On the other hand, it has been demonstrated the benefit of considering more complex textural features, such as those derived from the GLCM computation, which results in the enhancement of the thematic maps, in terms of number of recognized classes. Indeed, the exploitation of the GLCM *mean*, *variance*, *contrast* and *correlation* yielded the identification of four classes (*asphalt*, *low vegetation*, *trees* and *manmade*), and overall accuracies higher than 80%. Such results have been proved to be attainable, even when only few hundreds of training pixels are exploited. On the other side, some errors occurred in the classification of natural surfaces, in which a wide presence of bare soil might have caused the confusion with the *asphalt* class.

Thanks to the configuration of the COSMO-SkyMed constellation, close interferometric pairs of acquisitions are available, providing further sources of information which might be valuable for a better understanding of the imaged scenes. Within the framework of the presented research, the possibility of improving the discrimination among artificial and natural surfaces by employing the interferometric coherence estimates, has been investigated. Actually, it has been demonstrated the effectiveness of such approach, which especially allowed the enhancement of the *asphalt* and *natural* (or *low vegetation*) classes accuracy.

An important achievement has been reached by applying previously trained MLP-NNs to new images, to perform the classification task in a fully automatic way. High accuracies, at least comparable with those achieved on the training images, have been obtained in all the experiments, proving the good generalization capabilities of the MLP-NN based algorithm.

Tests on a different kind of urban environment, represented by the Great Denver area, still gave reason to the effectiveness of the exploited sources of information, in identifying large and relatively smooth asphalted surfaces, as well as natural areas, and big and/or isolated buildings. However, the correct classification of dense residential settlements is a topic that should be addressed in future researches, bearing in mind the limits due to the SAR system of acquisition. Actually, the orientation of the buildings with respect to the azimuth direction, as well as their geometry and the presence of many artificial and natural elements in the neighboring area, may completely hinder the detection of a single construction. In these contexts, the developed MLP-NN schemes erroneously associated the manmade pixels to the trees class.

Moving the interest on the enhancement of the vegetation mapping, a set of experiments have been carried out by exploiting VHR COSMO-SkyMed and Multispectral WorldView-2 images, as input data to a MLP-NN based classifier. Indeed, the multi-sensors, multi-temporal and multi-polarization data fusion allowed the improving of the producer's and overall accuracies, which have been evaluated higher than 90%. Such results look promising towards a possible continuous monitoring of agricultural surfaces by the available different multi-sensor products.

As far as concern the unsupervised object-based features extraction, the PCNN model has been properly adjusted in order to perform the edge detection task, as a crucial step of the fully automatic processing chain, which has been implemented for the buildings detection in VHR X-band SAR images. Originally developed for handling TerraSAR-X imagery, the algorithm succeeded also when applied to Spotlight COSMO-SkyMed acquisitions. The performance of the implemented technique is interesting under different points of view if compared with standard alternative approaches. Firstly, the PCNN significantly increases the level of automation in the processing scheme, due to its unsupervised nature. Secondly, the accuracy of the results can be recognized as better, or at least comparable, with that of other suitable edge detection techniques. Indeed, despite the identification of few false alarms, basically corresponding to different manmade structures or artifacts, in all the experiments more than the 90% of the buildings has been correctly detected. Additionally, it has to be noted that the performance of the PCNN does not seem to be affected by the spatial resolution of the elaborated satellite SAR image. In fact, investigations carried out on a Stripmap TerraSAR-X (3 m spatial resolution) and two Spotlight COSMO-SkyMed images (1 m spatial resolution) showed that the PCNN approach owns good properties of robustness. Finally, the processing time is rather short, requiring only a few seconds for the buildings extraction.

The last objective of the work aimed at developing fast, automatic and accurate algorithms for detecting changes in sub-urban areas, where the replacement of natural surfaces with new constructions frequently occurs, with the possible consequent worsening of the quality of the environment. The method makes use of both the partially supervised pixel-based approach, and the unsupervised object-based one. Basically, two different NN architectures are jointly applied: a supervised MLP-NN provides the land cover map for each acquisition, whereas an unsupervised and automatic PCNN model allows identifying "hot spots" by measuring the correlation value between the pulsing signature of the signals generated from each image. A comparison between the change map obtained by a post-classification analysis and the correlation mask produced by the PCNN module yields the final change map. The change detection scheme provided rather accurate results. In fact, six out of nine true changes have been detected from the pair of Stripmap images, while

all the transitions into the *manmade* class have been identified from the Spotlight data. It is worth highlighting that the inclusion the PCNN into the change detection processing chain resulted in enhanced robustness against co-registration errors. Among the advantages brought by joining the MLP and PCNN modules, it has to be especially mentioned the reduction of the number of false alarms (mainly caused by misclassification, misalignment, artifacts and speckle) if a comparison with the standard PCC is performed.

In conclusion, it is worthwhile to highlight the fact that each single task which has been discussed in this study, has been performed by an algorithm conceived as an independent module of a fast and automatic chain of processing, where the human intervention may be limited at the selection of representative training and validation samples for the first training of the MLP-NN classifier. Actually, once properly trained, the neural networks are able to automatically process stacks of data acquired at different time. Simultaneously, the unsupervised PCNN algorithm would identify the objects of interest, and/or scan pairs of images with the purpose of detecting "hot spots" for the following change detection task.

Appendix

With the advent of the new generation of VHR SAR sensors, perspective and objectives of the analysis of the acquired imagery have been revolutionised. The investigation on the potentiality of such data extends from the global study of the main scattering mechanisms and texture properties of patterns of urban settlements, to the analysis of the scattering properties of individual buildings. In this work it has been proved the potentiality of COSMO-SkyMed imagery to characterize urban scenarios by providing information in 2-D thematic maps. However, one of the future and most appealing applications of VHR SAR data, concerns the possibility of interpreting a single object from a 3-D point of view. Although the very high spatial resolution allows the identification of many details, the complexity of the urban areas, and of the multiple backscattering mechanisms which tipically occur in such environments, make the task rather challenging. In fact, even a simple isolated building may appear as an ensemble of many architectural details and it may be surrounded by gardens, trees, roads, social and technical infrastructure and many temporary objects. In Fig.A1 and Fig.A2 are shown two examples of buildings imaged by COSMO-SkyMed operating in the Spotlight mode (1 m spatial resolution), and by the Multispectral sensor of WorldView-2 (2 m spatial resolution).

The typical layover effect due to the SAR systems (see §1.2) is visible in both the images. Specifically, all the three buildings belonging to the Faculty of Engineering of the Tor Vergata University (Fig.A1a) display few dotted-lines of high backscattering lying parallel to the brightest continuous line. This is particularly visible in the yellow circle where two dotted-lines can be clearly distinguished (Fig.A1b). The explanation for such a backscattering behavior is that a significant trihedral corner-reflector mechanism occurs in each window on the walls facing the satellite, so that the aligned windows result in a bright dotted line. This means that the number of dotted-lines for each object might be associated to the number of floors in each building, and hence to their vertical dimension. Such assumption is confirmed by the picture in Fig.A1d.

Fig.A2 shows how the phase of the measured signal can produce additional information on the viewed object, which, in this case, is the Tor Vergata University sport center under construction. In Fig.A2a the interferometric phase is superimposed onto the backscattering intensity, whereas Fig.A2b shows how the interference fringes denote the 3-D structure, while in Fig.A2c it can be observed that the high degree of interferometric coherence carries information on the metallic structure.

This simple analysis of the COSMO-SkyMed images puts especially in evidence the potentiality of the VHR SAR data, which can be exploited for the interpretation of manmade structures. The scientific community is already working on the development of techniques for the backscattering mechanisms interpretation in urban scenarios, aiming at the accurate 3-D reconstruction of buildings [108]. Promising studies based on the exploitation of the interferometric information will allow the retrieval of the third dimension of manmade structures ([109–111]). Such knowledge is expected to contribute significantly to the urban planning, illegal building monitoring, risk management or damage assessment of manmade constructions.

(a)

(b)

(c)

(d)
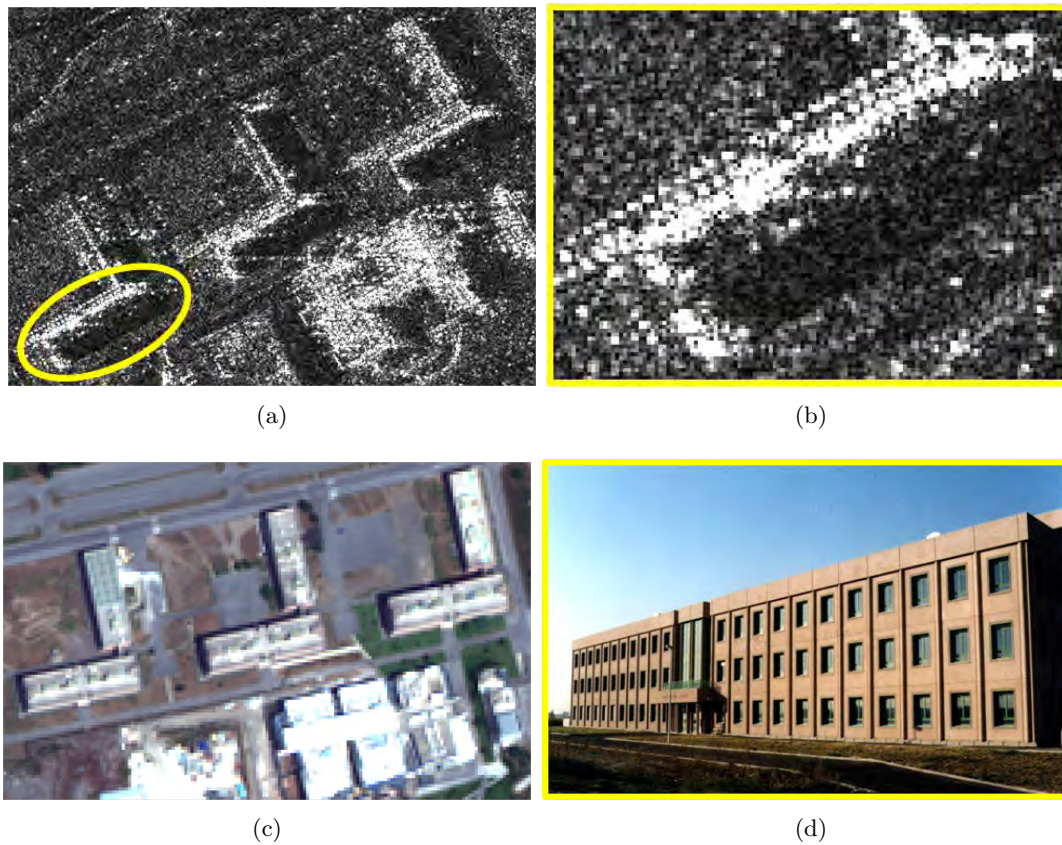
Figure A1. a)Spotlight COSMO-SkyMed image representing the buildings of the School of Engineering of the Tor Vergata University ©ASI. The yellow circle indicates the building whose details are showed in the zoom in b). The bright dotted lines correspond to the windows characterizing each floor of the building. In c) the same view is taken by the Multispectral sensor of WorldView-2 ©DigitalGlobe. d)Photograph of the building under study.
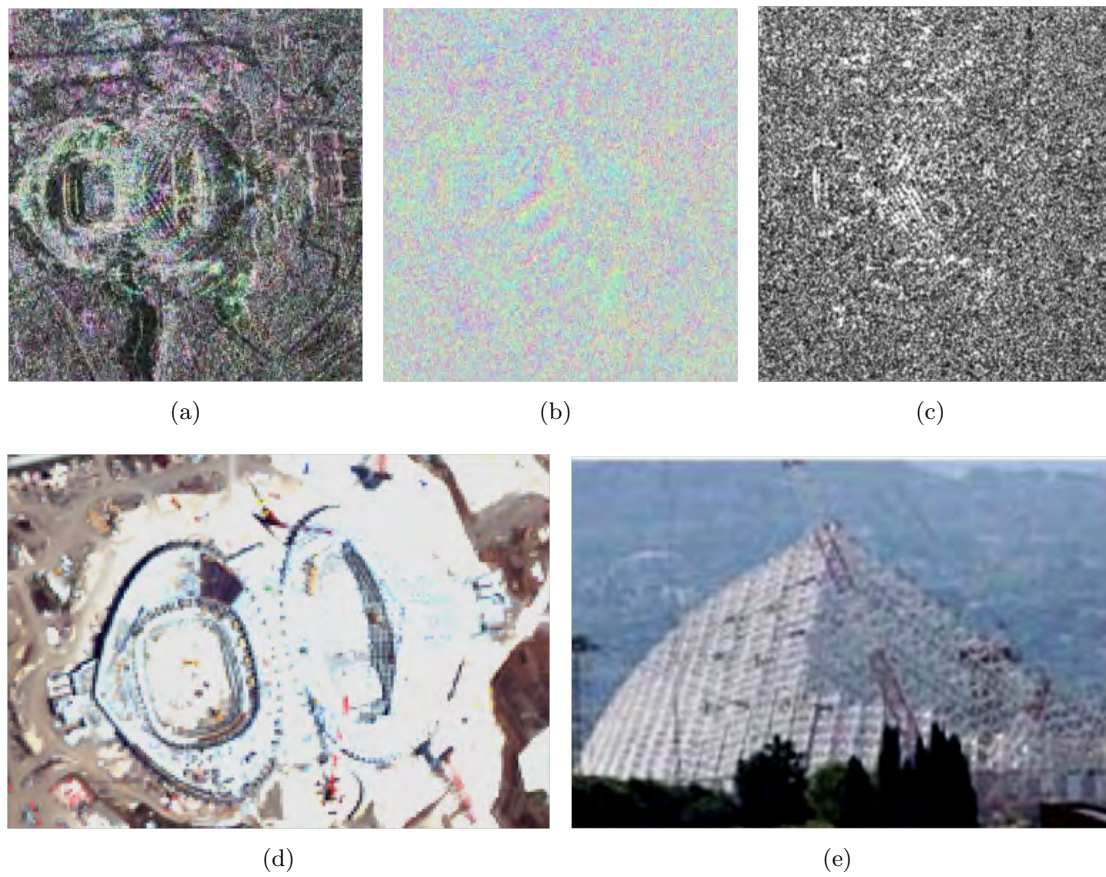
Figure A2. Tor Vergata University sport building under construction: a) superimposition of the interferometric phase to backscattering intensity of a spotlight COSMO-SkyMed image ©ASI; b) interferogram over the same frame. The fringes denote the 3-D structure of the construction; c) coherence image. The highest degree of coherence occurs over the metallic structure. d) Multispectral WorldView-2 image of the building ©DigitalGlobe; e) Photograph of the structure under study.

Bibliography

[1] United Nation. World Population Prospects: The 2012 Revision. Technical report, UE, June 2013.

[2] C. Oliver and S. Quegan. *Understanding Synthetic Aperture Radar Images*. Artech House, Norwood, MA, 1998.

[3] F. T. Ulaby, R. K. Moore, and A. K. Fung. *Microwave Remote Sensing: Active and Passive*, volume vol.2. Artech House, 1982.

[4] J. W. Goodman. "Some fundamental properties of speckle". *J. Opt. Soc. Amer.*, vol. 66(11):1145–1150, Nov. 1976.

[5] J. S. Lee. "Speckle suppression and analysis for synthetic aperture radar images". *Opt. Eng.*, vol. 25:636–643, 1986.

[6] A. Freeman. "SAR calibration: an overview". *IEEE Trans. Geoscience and Remote Sensing*, vol. 30(6):1107–1121, Nov. 1992.

[7] M. C. Dobson, F. T. Ulaby, and L. E. Pierce. "Land cover classification and estimation of terrain attributes using synthetic aperture radar". *Remote Sensing of Environment*, vol. 51:199–214, 1995.

[8] 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 Trans. Geoscience and Remote Sensing*, vol. 49:1911–1925, June 2011.

[9] M. Mansourpour, M.A. Rajabi, and J.A.R. Blais. "Effects and performance of speckle noise reduction filters on active radar and sar images". *Proc. ISPRS, XXXVI-1/W41*, 2006.

[10] Agenzia Spaziale Italiana. *COSMO-SkyMed System Description & User Guide*. ASI, Agenzia Spaziale Italiana, 2007.

[11] Agenzia Spaziale Italiana. *COSMO-SkyMed SAR Products Handbook*. ASI, Agenzia Spaziale Italiana, 2007.

[12] G. Valentini, A. Capuzi, S. Fagioli, R. Leonardi, A. Coletta, G. Angino, F. Battazza, F. Caltagirone, and F. Impagnatiello. "COSMO-SkyMed program: utilization and description of an advanced space EO dual-use asset". In *Envisat Symposium 2007*, Montreux, Switzerland, 23-27 April 2007. ESA SP-636.

[13] F. Caltagirone et. al. "COSMO-SkyMed Second Generation characteristics and development plan". *Proc. 17th Ka and Broadband Communications Navigation and Earth Observation Conference*, 3-5 October Oct. 2011.

[14] R. M. Haralick, K. Shammugam, and I. Dinstein. "Textural features for image classification". *IEEE Trans. Syst., Man, Cybernet.*, vol. 3:610–621, 1988.

[15] I. A. Basheer and M. Hajmeer. "Artificial neural networks: fundamentals, computing, design, and application". *Journal of Microbiological Methods*, vol.43:3–31, 2000.

[16] K. Hornik, M. Stinchcombe, and H. White. "Multilayer feedforward networks are universal approximators". *Neural Networks*, vol.2(5):359–366, 1989.

[17] J.F. Mas and J. J. Flores. "The application of artificial neural networks to the analysis f remotely sensed data". *Int. Journal of Remote Sensing*, vol.29(3):617–663, Feb. 2008.

[18] F. Pacifici. *Novel Neural Network-based algorithms for urban classification and change detection from satellite imagery.* PhD thesis, Ph.D. thesis, Tor Vergata University, Rome, Italy, 2010.

[19] G. B. Huang, Q. Y. Zhu, and C. K. Siew. "Extreme learning machine: Theory and applications". *Neurocomputing*, vol. 70:489–501, May 2006.

[20] C. Bishop. *Neural Networks for Pattern Recognition.* New York: Oxford University Press, 1995.

[21] T. Kavzoglu and P. M. Mather. "The use of backpropagating artificial neural networks in land cover classifications". *Int. Journal of Remote Sensing*, vol. 24:4907–4938, 2003.

[22] A. Zell et al. *SNNS Stuttgart Neural Network Simulator User Manual.* Report N6/95, University of Stuttgart, Institute for Parallel and Distributed High Performance Systems, Stuttgart, Germany, 1995.

[23] X. Zhuang, B.A. Engel, D. F. Lozano-Garcia, R. N. Fernandez, and C.J. Johannsen. "Optimization of training data required for neuro-classification". *Int. Journal of Remote Sensing*, vol. 15:3271–3278, 1994.

[24] D. R. Hush. "Classification with neural networks: a performance analysis". In *IEEE International Conference on Systems Engineering*, pages 277–280, Dayton, OH, Aug. 1989.

[25] E.B. Baum and D. Haussler. "What size net gives valid generalization?". *Neural Computation*, vol. 1:151–160, 1989.

[26] G. A. Carpenter, S. Gopal, S. Macomber, S. Martens, C. E. Woodcock, and J. Franklin. "A neural network method for efficient vegetation mapping ". *Remote Sensing of Environment*, vol. 70:326–338, 1999.

[27] R. Eckorn, H. J. Reitboeck, M. Arndt, and P. Dicke. "Feature linking via synchronization among distributed assemblies: simulation of results from cat visual cortex". *Neural Computation*, vol. 2:293–307, 1990.

[28] K. Waldemark, T. Lindblad, V. Becanovic, J. L. L. Guillen, and P. Klingner. "Patterns from the sky satellite image analysis using pulse coupled neural networks for pre-processing, segmentation and edge detection". *Pattern Recognition Letters*, vol. 21:227–237, 2000.

[29] J. A. Karvonen. "Baltic sea ice sar segmentation and classification using modified pulse-coupled neural networks". *IEEE Trans. Geoscience and Remote Sensing*, vol. 42(7):1566–1574, Jul. 2004.

[30] G. Kuntimad and H. S. Ranganath. "Perfect image segmentation using pulse coupled neural networks". *IEEE Trans. Neural Networks*, vol. 10(3):591–598, May 1999.

[31] H. S. Ranganath and G. Kuntimad. "Object detection using pulse coupled neural networks". *IEEE Trans. Neural Networks*, vol. 10(3):615–620, May 1999.

[32] F. Del Frate, D. Latini, and C. Pratola. "Automatic object extration from vhr satellite sar images using pulse coupled neural networks". *SPIE Remote Sensing Event*, Sep. 20-23 2010.

[33] F. Del Frate, D. Latini, C. Pratola, and F. Palazzo. "PCNN for automatic segmentation and information extraction from x-band sar imagery". *Int. Journal of Image and Data Fusion*, vol. 4(1):75–88, Sep. 2012.

[34] F. Pacifici and F. Del Frate. "Automatic change detection in very high resolution images with pulse coupled neural networks". *IEEE Geoscience and Remote Sensing Letters*, vol. 7(1):58–62, Jun. 2009.

[35] C. Pratola, F. Del Frate, G. Schiavon, and D. Solimini. "Towards fully automatic detection of changes in suburban areas from vhr sar images by combining multiple neural-network models". *IEEE Trans. Geoscience and Remote Sensing*, vol. 51(4):2055–2066, Apr. 2013.

[36] T. Lindblad and J. M. Kinser. *Image processing using pulse-coupled neural networks*. Springer - Verlag., 2005.

[37] A. Baraldi and F. Parmeggiani. "An investigation of the textural characteristics associated with gray level co-occurrence matrix statistical parameters". *IEEE Trans. Geoscience and Remote Sensing*, vol. 33(2):293–304, Mar. 1995.

[38] F. Dell'Acqua and P. Gamba. "Texture-based characterization of urban environments on satellite sar images". *IEEE Trans. Geoscience and Remote Sensing*, vol. 41(1):153–159, Jan. 2003.

[39] F. Dell' Acqua and P. Gamba. "Discriminating urban environments using multiscale texture and multiple sar images". *Int. Journal of Remote Sensing*, vol. 27(18):3797–3812, Sep. 2006.

[40] Y. Ban and H. Hu. "RADARSAT fine-beam sar data for land-cover mapping and change detection in the rural-urban fringe of the greater toronto area". *Urban Remote Sensing Joint Event*, 2007.

[41] D. Solimini F. Del Frate, F. Pacifici. "Monitoring urban land cover in rome, italy, and its changes by single-polarization multitemporal sar images". *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, vol. 1(2):87–92, Jun. 2008.

[42] M. E. Shokr. "Evaluation of second-order textural parameters for sea ice classification in radar images". *Journal of Geophysical Research*, vol. 96(C6):10625–10640, 1991.

[43] P. Gong, J. D. Marceau, and P. J. Howarth. "A comparison of spatial feature extraction algorithms for land-use classification with spot hrv data". *Remote Sensing of Environment*, vol. 40:137–151, 1992.

[44] R. Cossu. "Segmentation by means of textural analysis". *Pixel*, vol. 1(2):21–24, 1988.

[45] DigitalGlobe. *DigitalGlobe's Core Imagery Product Guide, v.1.1.* DigitalGlobe, May, 2013.

[46] DigitalGlobe. "The benefits of the 8 spectral bands of worldview-2". *www.digitalglobe.com*, March 2010.

[47] A. Lopes, R. Touzi, and E. Nezry. "Adaptive speckle filters and scene heterogeneity". *IEEE Trans. Geoscience and Remote Sensing*, vol. 28(6):992–1000, 1990.

[48] A. Roth, J. Hoffmann, and T. Esch. "TerraSAR-X: how can high-resolution sar data support the observation of urban areas? ". In *Joint Symposia URBAN - URS 2005*, Tempe, AZ, USA, March 14-16 2005.

[49] DLR German Aerospace Center. *TerraSAR-X Ground Segment - Basic Product Specification.* DLR - German Aerospace Center, 2008.

[50] M. Liao, T. Balz, L. Zhang, Y. Pei, and H. Jiang. "Analyzing terrasarx and cosmoskymed high-resolution sar data of urban areas ". In *JISPRS Hannover Workshop 2009 High-Resolution Earth Imaging for Geospatial Information*, Hannover, Germany, June 2009.

[51] H. Chaabouni-Chouayakh and M. Datcu. "Coarse-to-fine approach for urban area interpretation using terrssar-x data". *IEEE Geoscience and Remote Sensing Letters*, vol. 7:78–82, Jan. 2010.

[52] T. Esch and A. Roth. "Semi-automated classification of urban areas by means of high resolution radar data". In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS), Part B7, Commission VII*, Istanbul, Turkey, July 2004.

[53] A. Burini, C. Putignano, F. Del Frate, M. Lazzarini, G. Licciardi, G. Schiavon, D. Solimini, F. De Biasi, and P. Manunta. "TerraSAR-X imaging for unsupervised land cover classification and fire mapping". *IEEE, International Geoscience and Remote Sensing Symposium , IGARSS 2008*, 2008.

[54] A. Voisin, G. Moser, V. A. Krylov, S. B. Serpico, and J. Zerubia. "Classification of very high resolution sar images of urban areas by dictionary-based mixture models, copulas and markov random fields using textural features". *SPIE Remote Sensing Event*, Sep. 20-23 2010.

[55] C. Chen and H. McNairn. "A neural network integrated approach for rice crop monitoring". *Int. Journal of Remote Sensing*, vol. 27(7):1367–1393, Apr. 2006.

[56] D. Stathakis and A. Vasilakos. "Comparison of computational intelligence based classification techniques for remotely sensed optical image classification". *IEEE Trans. Geoscience and Remote Sensing*, vol. 44(8):2305–2318, Aug. 2006.

[57] F. Del Frate, F. Pacifici, G. Schiavon, and C. Solimini. "Use of neural networks for automatic classification from high resolution imagery". *IEEE Trans. Geoscience and Remote Sensing*, vol. 45(4):800–809, Apr. 2007.

[58] F. Pacifici, M. Chini, and W. J. Emery. "A a neural network approach using multiscale textural metrics from very high-resolution panchromatic imagery for urban landuse classification". *Remote Sensing of Environment*, vol. 113:1276–1292, 2009.

[59] J. D. Paola and R. A. Schowengerdt. "A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification". *IEEE Trans. Geoscience and Remote Sensing*, vol. 33:981–996, 1995.

[60] F. Pacifici, F. Del Frate, W. J. Emery, P. Gamba, and J. Chanussot. "Urban mapping using coarse sar and optical data: outcome of the 2007 grss data fusion contest". *IEEE Geoscience and Remote Sensing Letters*, vol. 5(3):331–335, Jul. 2008.

[61] P. Soille. *Morphological image analysis*. Springer-Verlag, Berlin-Heidelberg, 2004.

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

[63] Q. Zhang and I. Couloigner. "Benefit of the angular texture signature for the separation of parking lots and roads on high resolution multi-spectral imagery". *Pattern Recognition Letters*, vol. 27(9):937–946, Jul. 2006.

[64] A. Puissant, J. Hirsch, and C. Weber. "The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery". *International Journal of Remote Sensing*, vol. 26(4):733–745, Feb. 2005.

[65] T. Kurosu, S. Uratsuka, H. Maeno, and T. Kozu. "Texture statistics for classification of land use with multitemporal jers-1 sar single-look imagery". *IEEE Trans. Geoscience and Remote Sensing*, vol. 37(1):227–235, Jan. 1999.

[66] L. Kurvonen and M. Hallikainen. "Textural information of multitemporal ers-1 and jers-1 sar images with application to land and forest type classification in boreal zone". *IEEE Trans. Geoscience and Remote Sensing*, vol. 37(1):680–689, Jan. 1999.

[67] S. Arzandeh and J. Wang. "Texture evaluation of radarsat imagery for wetland mapping". *Canadian Journal of Remote Sensing*, vol. 28:653–666, 2002.

[68] G.Franceschetti, A.Iodice, and D.Riccio. "A canonical problem in electromagnetic backscattering from buildings". *IEEE Trans. Geoscience and Remote Sensing*, vol. 40:1787–1801, 2002.

[69] C. Pratola, F. Del Frate, G. Schiavon, D. Solimini, and G. A. Licciardi. "Characterizing land cover from x-band cosmo-skymed images by neural networks". *Urban Remote Sensing Joint Event*, 2011.

[70] F. Del Frate, C. Pratola, G. Schiavon, and D. Solimini. "Automatic features extraction in sub-urban landscape using very high resolution cosmo-skymed sar images". *Proc. IEEE, International Geoscience and Remote Sensing Symposium , IGARSS 2011*, pages 3614–3617, 2011.

[71] D. G. Barber and E. F. LeDrew. "SAR sea ice discrimination using texture statistics: a multivariate approach". *Photogramm. Eng. Rem. Sens.*, vol. 57:385–395, 1991.

[72] M. Pesaresi. "Textural classification of very high-resolution satellite imagery: empirical estimation of the interaction between window size and detection accuracy in urban environment". In *International Conference on Image Processing, ICIP 99*, volume vol. 1, pages 114–118, 1999.

[73] F. Dell'Acqua, P. Gamba, and G. Trianni. "Semi-automatic choice of scale-dependent features for satellite sar image classification". *Pattern Recognition Letters*, vol. 27:244–251, 2006.

[74] C. Wen, Y. Zhang, and K. Deng. "Urban area classification in high resolution sar based on texture features". In *Proc. ISPRS Working Groups ICWG IV/VIII, II/1, IV/8, IV/4, VIII/1*, volume vol. XXXVIII-7/C4, pages 281–285, 2009.

[75] L. A. Bartolucci, M. E. Dean, and P. E. Anuta. "Evaluation of the radiometric quality of the tm using clustering and multispectral distance measures". *Proc. Landsat-4 Scientific Characterization Early Results Symp., NASA Goddard Space Flight Center*, Feb. 22-24 1983.

[76] A. Singh. "Some clarifications about the pairwise divergence measure in remote sensing". *International Journal of Remote Sensing*, vol. 5(3):623–627, 1984.

[77] F. Del Frate, D. Loschiavo, C. Pratola, G. Schiavon, and D. Solimini. " Sub-urban andscape characterization by very-high resolution x-band cosmo-skymed sar images: first results". *SPIE Remote Sensing Event*, Sep. 20-23 2010.

[78] B. Haack and M. Bechdol. "Integrating multisensor data and radar texture measures for land cover mapping". *Computer & Geosciences*, vol. 26:411–421, May 2000.

[79] M. J. Hill, C. J. Ticehurst, J. S. Lee, M. R. Grunes, G. E. Donald, and D. Henry. "Integration of optical and radar classifications for mapping pasture type in western australia". *IEEE Trans. Geoscience and Remote Sensing*, vol. 43:1665–1681, 2005.

[80] A. Burini, G. Schiavon, and D. Solimini. "Fusion of high resolution polarimetric sar and multi-spectral optical data for precision viticulture". *Proc. IEEE, International Geoscience and Remote Sensing Symposium , IGARSS 2008*, Jul. 2008.

[81] M. Stasolla and P. Gamba. "Fusion of sar and optical data for urban extent extraction improvement". *Proc. IEEE, International Geoscience and Remote Sensing Symposium , IGARSS 2009*, Jul. 2009.

[82] N. Pierdicca, F. Pelliccia, and M. Chini. "Thematic mapping at regional scale using siasge radar data at x and l band and optical images". *Proc. IEEE, International Geoscience and Remote Sensing Symposium , IGARSS 2011*, Jul. 2011.

[83] G. Oller, D. Petit, and J. Inglada. "On the use of sar and optical images combination for scene interpretation". *Revue francaise de photogrammétrie et de télédétction*, vol. 182:87–91, 2006.

[84] G. Lehureau, F. Tupin, C. Tison, G. Oller, and D. Petit. "Registration of metric resolution sar and optical images in urban areas". *Proc. Synthetic Aperture Radar (EUSAR)*, Jul. 2008.

[85] C. Pratola, G. A. Licciardi, F. Del Frate, G. Schiavon, and D. Solimini. "Fusion of vhr multispectral and x-band sar data for the enhancement of vegetation maps". In *International Geoscience and Remote Sensing Symposium (IGARSS)*, Munich, Germany, 22-27 July 2012.

[86] A. Thiele, E. Cadario, K. Schulz, U. Thoennessen, and U. Soergel. "Building recognition from multi-aspect high-resolution insar data in urban areas". *IEEE Trans. Geoscience and Remote Sensing*, vol. 45:3583–3593, Nov. 2007.

[87] F. Xu and Y.-Q. Jin. "Automatic reconstruction of building objects from multiaspect meter-resolution sar images". *IEEE Trans. Geoscience and Remote Sensing*, vol. 45:2336–2353, Jul. 2007.

[88] R. Hill, C. Moate, and D. Blacknell. "Estimating building dimensions from synthetic aperture radar image sequences. *IET Radar Sonar Navig.*, vol. 2:189–199, Jun. 2008.

[89] U. Soergel, U. Thoennessen, and U. Stilla. "Reconstruncion of buildings from interferometric sar data of built-up areas". *ISPRS Conference "Photogrammetric image analysis"*, pages 59–64, Sep. 2003.

[90] M. Quartulli and M. Datcu. "Stochastic geometrical modeling for built-up area understanding from a single sar intensity image with meter resolution". *IEEE Trans. Geoscience and Remote Sensing*, vol. 42:1996–2003, Sep. 2004.

[91] A. Ferro, D. Brunner, and L. Bruzzone. "Automatic detection and reconstruction of building radar footprints from single vhr sar images". *IEEE Trans. Geoscience and Remote Sensing*, vol. 51(2):935–952, Feb. 2013.

[92] J. Xiao, J. Li, and A. Moody. "A detail-preserving and flexible adaptive filter for speckle suppression in sar imagery". *Int. Journal of Remote Sensing*, vol. 24(12):2451–2465, 2003.

[93] J. E. Colwell and F. P. Weber. "Forest change detection". *Proc. 15th Int. Sym. Remote Sensing Environment*, pages 839–852, 1981.

[94] F. Del Frate, G. Schiavon, and C. Solimini. "Application of neural networks algorithms to quickbird imagery for classification and change detection of urban areas". *Proc. IEEE, International Geoscience and Remote Sensing Symposium , IGARSS 2004*, pages 1091–1094, Sep. 2004.

[95] A. Singh. "Digitalchange detection techniques using remotely-sensed data". *Int. Journal of Remote Sensing*, vol. 10(6):989–1003, 1989.

[96] L. Bruzzone and S. B. Serpico. "An iterative technique for detection of land cover transition in multispectral remote sensing images". *IEEE Trans. Geoscience and Remote Sensing*, vol. 35(4):858–867, Jul. 1997.

[97] L. Bruzzone, D. Férnandez Prieto, and S. B. Serpico. "A neural statistical approach to multitemporal and multisource remote-sensing image classification". *IEEE Trans. Geoscience and Remote Sensing*, vol. 37:1350–1359, 1999.

[98] E. J. Rignot and J. J. van Zyl. "Change detection techniques for ers-1 sar data". *IEEE Trans. Geoscience and Remote Sensing*, vol. 31(4):896–906, Jul 1993.

[99] L. Bruzzone and D. F. Prieto. "An adaptive semiparametric and contextbased approach to unsupervised change detection in multitemporal remote sensing images". *IEEE Trans. Image Processing*, vol. 11(4):452–466, Apr. 2002.

[100] T. Kasetkasem and P. K. Varshney. "Image change detection algorithm based on markov random field models". *IEEE Trans. Geoscience and Remote Sensing*, vol. 40(8):1815–1823, Aug. 2002.

[101] P. Gamba, F. Dell' Acqua, and G. Lisini. "Change detection of multitemporal sar data in urban areas combining feature-based and pixel-based techniques". *IEEE Trans. Geoscience and Remote Sensing*, vol. 44(10):2820–2827, Oct. 2006.

[102] F. Bovolo and L. Bruzzone. "An adaptive technique based on similarity measures for change detection in very high resolution sar images". *Proc. IEEE, International Geoscience and Remote Sensing Symposium , IGARSS 2008*, pages 158–161, 2008.

[103] G. Moser and S. B. Serpico. "Unsupervised change detection with very high-resolution sar images by multiscale analysis and markov random fields". *Proc. IEEE, International Geoscience and Remote Sensing Symposium , IGARSS 2010*, pages 3082–3085, 2010.

[104] A. Bouaraba, A. Younsi, A. Belhadj-Aissa, M. Acheroy, N. Milisavljevic, and D. Closson. "Robust techniques for coherent change detection using cosmo-skymed sar images". *Progress in Electromagnetics Research M.*, vol. 22:219–232, 2012.

[105] C. Pratola, M. Del Greco, F. Del Frate, G. Schiavon, and D. Solimini. "Towards fully automatic generation of land cover maps from polarimetric and metric-resolution sar data". *Proc. IEEE, International Geoscience and Remote Sensing Symposium , IGARSS 2010*, pages 3102–3105, 2010.

[106] F. Pacifici, F. Del Frate, C. Solimini, and W. J. Emery. "An innovative neural-net method to detect temporal changes in high-resolution optical satellite imagery". *IEEE Trans. Geoscience and Remote Sensing*, vol. 45(9):2940–2952, Sep. 2007.

[107] L. Bruzzone, M. Marconcini, U. Wegmüller, , and A. Wiesmann. "An advanced system for the automatic classification of multitemporal sar images". *IEEE Trans. Geoscience and Remote Sensing*, vol. 42(6):1321–1334, Jun. 2004.

[108] Xiao Xiang Zhu and R. Bamler. "Very High Resolution spaceborne sar tomography in urban environment". *IEEE Trans. Geoscience and Remote Sensing*, vol. 48(12):4296–4308, Dec. 2010.

[109] A. R. Brenner and L. Roessing. "Radar imaging of urban areas by means of very high-resolution sar and interferometric sar". *IEEE Trans. Geoscience and Remote Sensing*, vol. 46(10):2971–2982, Oct. 2008.

[110] M. Eineder, N. Adam, R. Bamler, N. Yague-Martinez, and H. Breit. "Spaceborne spotlight sar interferometry with terrasar-x". *IEEE Trans. Geoscience and Remote Sensing*, vol. 47(5):1524–1535, May 2009.

[111] G. A. Giardino, G. Schiavon, and D. Solimini. "An approach for improving building height estimation from interferometric sar data". *IEEE, International Geoscience and Remote Sensing Symposium , IGARSS 2013*, 2013.

List of Acronyms and Abbreviations

A Asphalt class

ANN Artificial Neural Network

ASI Agenzia Spaziale Italiana

AVNIR Advanced Visible and Near Infrared Radiometer

BD Building Detection

BS Bare Soil class

C Classification

CAMLAND Computer-Assisted Monitoring of Land cover

CD Change Detection

CGM Conjugate Gradient Method

CHRIS Compact High Resolution Imaging Spectrometer

COSMO-SkyMed COnstellation of small Satellites for Mediterranean basin Observation

CSG COSMO-SkyMed Second Generation

CSK COSMO-SkyMed

DEM Digital Elevation Model

DF Data Fusion

DLR Deutschen Zentrums für Luft und Raumfahrt

EO Earth Observation

ERS European Remote Sensing satellite

GEC Geocoded Ellipsoid Corrected product

GLCM Gray Level Co-Occurence Matrix

GTC Geocoded Terrain Corrected product

HI HImage mode

HR Huge Region mode

IEEE	Institute of Electrical and Electronics Engineers
IEM	Interoperability, Expandability and Multi-sensoriality
LEO	Low Earth Orbit
LTP	Long Term Plan
LV	Low Vegetation class
MAP	Maximum a posteriori Probability
MD	Defence Ministry
MDG	Detected Ground Multi-look product
MLP-NN	Multi Layer Perceptron Neural Network
MM	Manmade class
MRF	Markov Random Fields
MSE	Mean Sum Squared Error
MTP	Medium Term Planning
N	Natural class
NAHIRI	Neural Architecture for HIgh Resolution Imagery
NDVI	Normalized Difference Vegetation Index
NN	Neural Network
O	Olive class
OA	Overall Accuracy
P	Pasture class
PalSAR	Phased Array type L-band Synthetic Aperture Radar
PCC	Post Classification Comparison
PCNN	Pulse Coupled Neural Network
PP	Ping Pong mode
PRISMA	PRecursore IperSpettrale della Missione Applicativa
PROBA	Project for On-Board Autonomy
QB	QuickBird
RADAR	RAdio Detection And Ranging
RCS	Radar Cross Section
S	Shrubs class
S2	Spotlight mode

SAR	Synthetic Aperture Radar
SCG	Scaled Conjugate Gradient
SCS	Single look Complex Slant product
SLAR	Side-Looking Airborne Radar
SLC	Single Look Complex
SNR	Signal to Noise Ratio
SSE	Sum of Squares Error
STP	Short Term Planning
SVM	Support Vector Machine
T	Trees class
TanDEM-X	TerraSAR-X add-on for Digital Elevation Measurement
TD	Transformed Divergence
TSX	TerraSAR-X
VHR	Very High Resolution
VT	overhead Trellis Vineyard class
VR	Vineyards cultivated in Rows class
W	Water class
WR	Wide Region mode
WV2	WorldView-2