UNIVERSITÀ DEGLI STUDI TOR VERGATA
ROMA

Dipartimento di Informatica, Sistemi e Produzione

GeoInformation Doctorate

High Resolution Urban Monitoring by Neural Network and Transform Algorithms

A thesis submitted in partial fulfillment for the PhD degree
(*Dottore di Ricerca*)

Candidate: **Chiara Solimini**

Mentor (Uni. Tor Vergata): **Fabio Del Frate**
Mentor (CU-Boulder): **William J. Emery**

January 2007

Contents

1 Introduction **3**

1.1 Relevance of classification and change detection in monitoring urban areas: evolution and processes 9

1.2 Type of surfaces to be classified 14

2 Remote Sensing Data **16**

2.1 Optical/IR and radar available today and in near future 16

2.2 Relevance of Very-High-Resolution (VHR) optical data 19

2.3 Features of data available today 20

3 Classification Methods of High Resolution Remote Sensing Data **21**

3.0.1 Unsupervised Neural Network Classification Algorithms: The Self-Organizing Map 25

3.1 Summary of methods used for classification by spectral radiances 26

3.1.1 The Neural Net method: Performance and comparison with other method 26

3.1.2 The Neural Network algorithm 27

3.1.3 Single image classification: MLP near-optimal structure 29

4 Methods to "globally" characterize a (large) urban area **40**

4.1 The Discrete Fourier Transform 40

4.1.1 The Discrete Spectrum 40
4.1.2 Discrete Fourier Transform Formulae 41
4.1.3 Properties of the Discrete Fourier Transform 42
4.2 The Discrete Fourier Transform of an Image 43
4.2.1 Definition 43
4.2.2 Evaluation of the Two Dimensional, Discrete Fourier Transform 44
4.2.3 The Concept of Wavenumber 45

5 Applications 48
5.1 Detailed classification of "elementary" surfaces using the optical/IR spectra 48
5.2 Global characterization of large cities of parts of cities 51
5.3 High Resolution Change Detection 51
5.3.1 Transform algorithms to characterize urban areas 54

6 Conclusions 96

7 Appendix 116

Chapter 1

Introduction

Urban areas occupy a relatively small portion of the earth surface. At the same time they represent one of the most complex, intricate, and variable of all land covers and land use and are also among the most rapidly expanding and changing elements of the landscape. A timely manner of monitoring urban areas is needed to be able to accurately assess the impact of human activities on the environment; in particular, monitoring the existence, distribution and changing patterns of cities which play a crucial role in the allocation and conservation of natural resources, environmental and ecosystem management, and economic development. Also, an understanding the dynamics of the urban area and its impact of the human activities on the environment is needed to assess and to assess the supporting land carrying capacity. Usually land cover and classification maps are provided by government administration and field surveys, collecting, for example, aerial pictures and national censuses.

However, the data collection provided by these techniques is costly and time consuming, often not updated and lacking in detailed information. Realistically, the only feasible source of information on land cover over large areas which allows data to be acquired in a regularly repeatable manner is satellite remote sensing. Indeed, in principle, remote sensing systems could measure en-

ergy emanating from the earth's surface in any reasonable range of wavelenghts (Richards (1993)). However, in spite of the great potential of remote sensing as a source of information on land cover and the long history of research devoted to the extraction of land cover information from remotely sensed imagery, many problems have been encountered, and the accuracy of land cover maps derived from remotely sensed imagery has often been viewed as too low for operational users (Bernard *et al.* (1997), Binaghi *et al.* (1996) and Foody (2002)).

Many factors may be responsible for these problems. These include the nature of the land cover classes (e.g. discrete or continuous), the properties of the remote sensor (e.g., its spatial and spectral resolutions), the nature of the land cover mosaic (e.g., degree of fragmentation), and the methods used to extract the land cover information from the imagery (e.g., classification methods) (Foody and Ajay (2004)). These various problems have driven research into a diverse range of issues focused on topics such as image analysis techniques. Many of the problems in mapping land cover noted in the literature are associated with the methods used to extract the land cover information from the imagery. This has motivated a considerable amount of research into classification methods and supervised classifications in particular. Early work based on basic classifiers such as the minimum distance to means algorithm prompted the adoption of more sophisticated statistical classifiers such as the maximum-likelihood classification. Problems associated with satisfying the assumptions required by such classification methods has driven research into nonparametric alternatives including techniques such as evidential reasoning (Peddle and Franklin (1992), Wilkinson and Megier (1990)) and more recently neural networks (Benediktsson *et al.* (1990), Kanellopoulos and Wilkinson (1997), Liu *et al.* (2003), and Del Frate *et al.* (1999)), decision trees (Goel *et al.* (2003), McIver and Friedl (2002) and Friedl and Brodley (1997)) and genetic algorithms (Nabeel *et al.* (2006)). Indeed, the accuracy with which land cover may be classified by these techniques has often been found to be higher than that de-

rived from the conventional statistical classifiers (e.g. Peddle *et al.* (1994), Rogan *et al.* (2002), Li *et al.* (2003) and Pal and Mather (2003)).

The advent of the recent generation of very high spatial resolution satellites has lead to a new set of applications made possible by the geometrical precision and high level of thematic detail in these images. In particular monitoring urban areas has captured the researchers attention. In fact, urban analysis using high spatial resolution involves a large number of applications such us road network mapping, government survey map updating, monitoring of urban growth and over building. These kinds of applications were not feasible with the previous generation of moderate-resolution satellites (e.g. Landsat Thematic Mapper). However, further improvements in the accuracy of automated classification algorithms are needed to satisfy the end-user requirements in all application domains. In particular, the number of classes extracted by classification algorithms is relatively low when compared with the number often required by environmental mapping agencies to describe land cover uses at regional, national and continental scales (e.g. the European Environment Agency's hierarchical CORINE Land Cover Data Set: 44 classes; the U.S. National Land-Cover Data 2001: 26 classes at level II). This lack of classification efficiency still leads to time consuming and expensive photointerpretation procedures.

For these reasons, it is important that the remote sensing community invests more energy to define advanced and effective methods to address the classification problem. Thus, research into new classification methods continues, and support vector machines (SVMs) have recently attracted the attention of the remote sensing community (Huang *et al.* (2002), Brown *et al.* (1999) and Halldorsson *et al.* (2003)). A key attraction of the SVM-based approach to image classification is that it seeks to fit an optimal hyperplane between classes and may require only a small training sample (Huang *et al.* (2002), Mercier and Lennon (2003) and Belousov *et al.* (2002)). Although the potential of

SVM is evident and early studies have demonstrated considerable success in using them to map land cover accurately, there are problems in their usage (Foody and Ajay (2004)). One of the main concerns is that SVMs were originally defined as binary classifiers and their use for multiclass classifications is problematic, requiring strategies that reduce the multiclass problem to a set of binary problems. Therefore, the researchers have sought to extend the basic binary SVM approach to form a multiclass classifier (Perez-Cruz and Artes-Rodriguez (2002), Angulo *et al.* (2003), Lee *et al.* (2003), Zhu *et al.* (2003)) and recently an approach for "one-shot" multiclass SVM classification has been reported (Hsu and Lin (2002)). However, at the present state of the art there is no discernible evidence in the classification accuracy advantages between the neural and non-neural approaches (Wilkinson (2005)).

In this context, one of the major problem related to high resolution image processing is to handle the extremely large set of data, therefore challenging, the general ability of the chosen classification algorithm. With respect to classical methods, neural networks represent a fundamentally different approach to problems like pattern recognition. They do not rely on probabilistic assumptions and do not need assumptions about normality in data sets. Moreover, they show a considerable ease in using multi-domain data sources. The effectiveness of neural networks is related to their self-adaptive characteristic; in particular, they can adjust themselves to the input data without any *a priori* assumptions (e.g., they do not require any explicit specification of the data distribution), and can approximate any function with arbitrary accuracy (Bishop (1995)). Therefore, neural network algorithms are well suited to the classification of remote sensing images.

A crucial issue which the classification accuracy depends on, is the decision-making process related to the assignment of the classes (e.g, in a pixel based classification, to assign a pixel to one class rather than a different one). The class assignment process performed by a classifier based on neural network al-

gorithms has been found to be effective (Del Frate *et al.* (1999), Del Frate *et al.* (2004)). For this reason Neural Networks (NN) have received considerable attention, as a tool in the field of remote sensing, after a new training scheme was developed. This new principle of a *back-propagation* algorithm was initially proposed by Werbos (1974) and rediscovered by Rumelhart *et al.* (1986). As explained above, since the early 1990s, several researchers have compared the performance of NN with conventional statistical approaches for remote sensing image classification. Benediktsson *et al.* (1990) evaluated the both methods for multi-source remote sensing data classification. They noted that a neural network has a great potential as a pattern recognition method for multi-source remotely sensed data due to its underlying distribution-free nature. Bishof *et al.* (1994), Paola and Schowengerdt (1995) compared methods for multispectral classification of Landsat TM data and both found that with proper training, a neural network was able to perform better than the maximum-likelihood classifier. However, even if these studies seem to show that NN performance is comparable or better than those provided by other techniques, they were mainly focused on medium resolution Landsat images and on the use of a single neural network for classifying and/or extracting specific features from a single image, namely the same image from which the examples training the network are taken.

A detailed analysis of the pixel-based classification yielded by this type of NN algorithm on very high resolution images such as those provided by the QuickBird (QB) or Ikonos platforms is still lacking. Moreover, the potential of a single neural network as a tool for automatic and sequential processing of images contained in high-resolution image archives has been scarcely investigated. With sequential processing we mean that the network might be able to retrieve from the archive all the images that contain or do not contain a specific class of land cover, or where the ratio between areas corresponding to different classes is within/out predefined ranges. In other words the NN allows the iden-

tification of high-level (object or region of interest) spatial features from the pixel representation contained in a raw image or image sequence, hence being able to address scientific issues characteristic of the image information mining field (Datcu *et al.* (2003) and Hsu *et al.* (2002)).

As a first step, in this study, we want to assess and optimize the neural network approach for the pixel-based classification of a single very high resolution image, such as those provided by the QB satellite.

Next we move to the conceptually most innovative part of the study which is to investigate the capabilities of supervised NN in providing automatic classification of a collection of images, therefore assessing their potentialities from an image information mining point of view. This stresses their general capabilities to adapt to new input patterns different from the patterns on which the nets have been trained. Several factors interfere with the objective of designing a NN able to be generalized for their use with images not used in the training phase. Examples the influences the different incidence angles and/or atmospheric conditions or the fact that different types of material may characterize the same class. In spite of these problems, the robustness of the spectral information has to be investigated and such an analysis needs to concur with the design of the NN. Addressing this point, we consider both very high (QB) and moderate (Landsat) resolution images and a specific application domain which is the feature extraction and information discovery applied to urban areas.

In fact, monitoring changes and urban growth over time is one of the major challenges of scientific research in remote sensing with a strong potential for the policy implications of this image analysis that would improve environment and security monitoring (Jensen and Cowen (1999), Donnay *et al.* (2001), Carlson (2003)). For example in Wilson *et al.* (2003) an urban growth model is developed. The model, which is based on land cover derived from remotely sensed satellite imagery, determines the geographic extent, patterns, and classes of urban growth over time. Synthetic Aperture Radar(SAR) imagery can also be

used, providing an additional source of information and enhancing the capabilities of optical data (Dell'Acqua *et al.* (2003), Schiavon *et al.* (2003)). Therefore a large volume of satellite data is available, but despite many competing automatic approaches, it is difficult to fully and automatically address the problems raised by the different application scenarios. In this study the aim of the classification is to distinguish between areas of artificial cover (sealed surfaces) including asphalt or buildings, and open spaces such as bare soil or vegetation. As a by-product this makes possible the retrieval of other features such as the detection of new buildings or the discovery of modifications in existing ones.

1.1 Relevance of classification and change detection in monitoring urban areas: evolution and processes

Classification

Ever since the first multispectral imagery became available from civilian remote sensing satellites in the late 1970s, considerable effort has been devoted to the classification of image data with the aim of producing high-quality thematic maps and establishing accurate inventories of spatial classes (Wilkinson (2005)). Classification is regarded as a fundamental process in remote sensing, which lies at the heart of the transformation from satellite image to usable geographic products. In order to produce thematic maps different methodologies of image classification have been developed. In particular, we can separate classification methods by the following:

- the development of components of the classification algorithm including the training or learning strategy and approaches to class separation based on statistical or other estimators and class separability indexes,

- the development of novel system-level approaches that augment the underlying classifier algorithms,
- the exploitation of multiple types of data or ancillary information, both numerical and categorical, in a classification process.

In the first category we can include the development of the supervised maximum-lakelihood method (Frizzelle and Moody (2001), Benediktsson *et al.* (1990)), n-dimensional probability density function methods (Cetin *et al.* (1993)), artificial neural networks (Gamba and Houshmand (2001), Yoshi and Omatu (1994), Bishof *et al.* (1994), Heerman and Khazenie (1992), Atkinson and Tatnall (1997), Paola and Schowengerdt (1995), Serpico and Roli (1995), Kanellopoulos and Wilkinson (1997), Ji (2000) and Dreyer (1993)), decision trees (Hansen *et al.* (1996), Kumar and Majumder (2001)), discriminant analysis (Franklin (1994), Hardin (1994)), genetic algorithms (Tso and Mather (1999),Sheeren *et al.* (2006)) and spectral shape analysis (Carlotto (1998)). In the second category we includ fuzzy or similar approaches that "soften" the results of a hard classifier (Seong and Usery (2001), Foody (1996), Foody (2002), Zhang and Foody (1998), Bastin (1997), Du and Lee (1996)), multiclassifier systems that integrate the outputs of several underlying classifier algorithms (Wilkinson *et al.* (1995)), and decision fusion methods (Benediktsson and Kanellopoulos (1999), Jimenez *et al.* (1999), Petrakos *et al.* (2001)).

The third category includes the use of texture measures extracted from imagery (Franklin *et al.* (2001), Augusteijn *et al.* (1995)), the use of structural or spatial context information from the imagery (Barnsley and Barr (1996),Gong and Howarth (1990)), the use of multisource data (Bruzzone *et al.* (1997), Bruzzone *et al.* (1999), Zhang (2001)), and the use of ancillary geographical knowledge integrated in the overall classification system through, for example, an expert system approach (Srinivasan and Richards (1990), Moller-Jensen (1990), Wilkinson and Megier (1990), Kontoes *et al.* (1993), Tonjes *et al.* (1999)). Some

approaches integrate several of the afore-mentioned, e.g., (Debeir *et al.* (2002)).

Change Detection

Detection of land-cover changes is one of the most interesting aspects of the analysis of multitemporal remote sensing images (Richards (1993)). In particular, change detection is very useful in many applications, like land use change analysis, assessment of burned areas, studies of shifting cultivation, assessment of deforestation etc. (Singh (1989), Coppin and Bauer (1994), Green *et al.* (1994)). Further, the recent availability of very high resolution images has enlarged the number of applications especially in urban monitoring such as the growth of urban areas and discovering building permit infractions (Bruzzone and Carlin (2006)). Usually, change detection involves a couple of spatially registered remote-sensing images acquired of the same area at two different times. Two main approaches to detecting land-cover changes can be distinguished (Bruzzone and Serpico (1997)):

1. changes detected by comparing the spectral reflectances of multitemporal satellite images,

2. changes can be detected by using supervised classifiers.

Many change detection algorithms are based on the first approach (Singh (1989), Fung and LeDrew (1987), Chavez and MacKinnon (1994), Muchoney and Haack (1994), Fung (1990), Muchoney and Haack (1994)). The *Univariate Image Differencing* algorithm (Singh (1989), Fung (1990), Chavez and MacKinnon (1994), Muchoney and Haack (1994)) performs change detection by subtracting, on a pixel basis, the images acquired at two times to produce a "difference image". Under the hypothesis that there are only minor changes between the two times, changes can be detected in the tails of the probability density functions of the pixel values in the difference image; this technique is usually

applied to a single spectral band. Other techniques, like *Vegetation Index Differencing* (Singh (1989), Townshend and Justice (1995)) make the same kind of comparison by using, instead of a spectral band, vegetation indices (Richards (1993)) or other linear (e.g., Tasseled Cup Transformation (Richards (1993), Fung (1990))) or nonlinear combinations of original satellite bands. The widely used *Change Vector Analysis* technique (Singh (1989)) exploits a similar concept. In this case, however, the pixels at each time are represented by their vectors in the feature space. Then, for each couple of pixels, the "spectral change vector" is computed as the difference between the image feature vectors at the two times. The statistical analysis of the magnitudes of the spectral change vectors allows one to detect the presence of changes, while their directions make it possible to distinguish between different kinds of transitions.

Another technique similar to those described above is *Image Rationing*; in this case, the comparison between spectral bands at two times is performed by computing the ratio, instead of the difference, between images. Techniques based on *Principal Component Analysis* (Singh (1989), Fung and LeDrew (1987), Muchoney and Haack (1994)) can also be used to perform change detection by applying the principal component transformation separately to the feature space at each time or as to the merged feature space at two times. In the first case, change detection is performed with *Vegetation Index Differencing* using principal components instead of vegetation indices. In the second case land-cover changes are detected by analyzing the minor components of the transformed feature space (Singh (1989)). The above techniques usually do not aim to identify explicitly what kinds of land-cover transitions have taken place in an area (e.g., the fact that a vegetated area has been urbanized). Only the *Change Vector Analysis* technique allows one to distinguish among different kinds of land cover changes but, not being supervised, it does not explicitly identify the specific typologies of transitions. The above techniques are suitable for applications like, the definition of burned areas, the detection of pollution,

deforestation, etc. However, they cannot be applied when the information on change character is not sufficient, like, for example, in the monitoring of changing cultivation, where it is necessary to recognize the kinds of changes that have taken place in the agricultural area investigated. In addition, the performances of such technique is generally degraded by several factors like:

- differences in illumination at two times,
- differences in atmospheric conditions, in sensor calibration and in ground moisture conditions,

that make difficult a direct comparison between the raw images acquired at different times.

To overcome these problems, one can use the techniques based on a supervised classification of multitemporal images (Singh (1989)). The simplest technique in this category is *Post-Classification Comparison* (Singh (1989)). It performs change detection by comparing the classification maps obtained by independently classifying two remote-sensing images of the same area acquired at different times. In this way, it is possible to detect changes and to understand the kinds of transitions that have taken place. Furthermore, the classification of multitemporal images avoids the need to normalize for atmospheric conditions, sensor differences etc., between the two acquisitions. However, the performances of the *Post-Classification Comparison* technique critically depend on the accuracies of the classification maps. In particular, the final change detection map exhibits an accuracy close to the product of the classification accuracies yielded at the two times (Singh (1989)). This is due to the fact that *Post-Classification Comparison* does not take into account the dependence existing between two images of the same area acquired at two different times.

Supervised Direct Multidata Classification, (Singh (1989)), is able to overcome this problem. In this technique, pixels are characterized by a vector obtained by "stacking" the feature vectors related to the images acquired at

two times. Change detection is then performed by considering each transition as a class and by training a classifier to recognize the transitions. Appropriate training sets are required for the success of this method: the training pixels at the two times should be related to the same points on the ground and should accurately represent the proportions of all the transitions in the entire images. Usually, in real applications, it is difficult to have training sets with such characteristics. In general the approach based on supervised classification is more flexible than that based on the comparison of multitemporal image data. In addition to the already mentioned capability to explicitly recognize land-cover transitions and to reduce the effects of different acquisition conditions at two different times, it also allows us to perform change detection using different sensors at different times. This is a useful property when change detection on a large time difference has to be performed and available images are provided by different sensors. The spatial resolution plays a key role in urban monitoring related to the detection of fine-scale objects present in urban scenes. In particular high spatial resolution is required to reduce the problem of mixed pixels (i.e. the pixels that represent the spectral signature of more than one class due to the insufficient spatial resolution of the sensor) present in the medium resolution images (e.g. Landsat imagery). Also the multispectral information is required to discriminate between the different surfaces/materials that compose the urban areas.

1.2 Type of surfaces to be classified

Since the appearance of very high resolution sensors and the object-oriented image analysis (OOIA), new questions about the acquisition of knowledge for classification procedures can be posed. The OOIA approach is characterized by the extraction of object primitives from images where each object corresponds to a group of homogeneous pixels. The object recognition methods are

generally based on the use of knowledge related to spectral, spatial and contextual properties (e.g. mean of spectral and textural values, shape, length, area, adjacency and inclusion relationship...) (Sheeren *et al.* (2006)). While there are several studies that compare object-oriented and pixel-based classification techniques (Rego and Koch (2003)), only few works focus on the development of the knowledge base used to recognize the objects. No urban objects dictionary or ontology exists to create the knowledge base. Most of the time, the knowledge is implicit and is held only by the domain experts. However, the experts are rarely able to supply an explicit description of the knowledge they use in their reasoning. Data mining techniques, such as genetic algorithms, can help to derive this knowledge and to extract classification rules automatically. These rules are intended to enrich an ontology in the urban remote sensing imagery domain (Sheeren *et al.* (2006)). The algorithm has been succesfully tested on a very high resolution QuickBird multi-spectral (MS) image of an urban area. In this work we use the results obtained in the above study: we make the assumption that the spectral signatures allows us to separate, at the first hierarchical level, these basic "elementary" classes, such as: vegetation, bare soil and "mineral". The second hierachical level includes the segmentation of "mineral" surfaces into "man made" elements such as roads and buildings. Finally, in the third hierachical level the "man made"surfaces are subdivided into residential, commercial, industrial etc.

Chapter 2

Remote Sensing Data

2.1 Optical/IR and radar available today and in near future

Airborne and satellite remote sensing techniques have been investigated for human settlement detection, population estimation, and urban analysis since the mid-1950s (Henderson and Xia (1997)). Systems operating in the visible and near-infrared regions of the electromagnetic spectrum have received most of the attention and thus offer the most advanced and widely employed techniques. Starting in the early 1980s airborne imaging radar systems have also been used for urban land cover mapping, and offer some distinct advantages over optical sensors as well as contributing to potential synergistic benefits of merged data set. In fact, the reflectance measurements acquired in visible and infrared regions of the spectrum, are primarily related to molecular resonances of surface materials and visible radiometers are not able to sense the surface, under adverse meteorological conditions and during the night. The radar backscattering measurements are primarily related to the physical properties of surface objects such as surface roughness and the dieletric constant. In particular, in a built up area, a radar backscatter signal depends on the ori-

entation of facets and on the presence of dihedral corner reflectors, formed by the intersection of horizontal and vertical built features, and trihedral corner reflectors, formed by two orthogonal vertical walls and the ground (Schiavon and Solimini (2000)). Moreover, the radar sensor is able to collect data under all meteorological conditions and at any time of the day. However, the urban landscape is highly variable and complex and the radar signal interaction with the urban built-up area is not easy to interpret. Recently, modelling returns from urban structures have been carried out using L-band images of groups of buildings at linear polarizations and different incidence angles (Schiavon *et al.* (2001)).

Satellite Name	*Source*	*Launch*	*Sensors*	*Types*	*No.of Channels*	*Resolution (meters)*
ERS-2	ESA	1995	AMI	Radar	1	26
RADARSAT-1	Canada	1995	SAR	Radar	1	9, 100
Lansat-7	US	1999	ETM+	MS	6 1	30 60
				PAN	1	15
ROCSAT-1	Taiwan	1999	n/a	MS	n/a	2
				PAN	n/a	8
IKONOS	SpaceImaging	1999	IKONOS	MS	4	4
				PAN	1	1
EROS-A1	ImageSat	2000	PAN	PAN	1	1.5

Table 2.1 - Earth Observation Satellites: optical/IR and radar current missions. Satellites launched between 1995-2000.

QuickBird	DigitalGlobe	2001	MS	MS	4	2.44
			PAN	PAN	1	0.61
MTI	US	2001	MTI	MS	15	5
Envisat-1	ESA	2002	ASAR	Radar	1	30, 150
SPOT-5	France	2002	HRV	MS	3 1	10 20
				PAN	1	2.5, 5
OrbView-3	Orbimage	2003	Orbview	MS	4	4
			PAN	PAN	1	1
ROCSAT-2	Taiwan	2004	MS	MS	4	8
				PAN	1	2
IRS-P6	India	2004	LISS 3/4	MS	7	5.8, 23.5
ALOS	Japan	2004	PALSAR	Radar	1	10, 100
			AVNIR-2	MS	4	10
			PRISM	PAN	1	2.5
EROS-B1	ImageSat	2004	PAN	PAN	1	0.82
IRS-P5	India/US	2005	PAN	PAN	1	2.5
KOMPSAT-1	Korea	1999	EOC	PAN	1	6.6
KOMPSAT-2	Korea	2006	PAN	PAN	1	1
			MS	MS	4	4
IRS-P5	India-US	2005	PAN	PAN	1	2.5
CARTOSat-2	India-US	2005	PAN	PAN	1	2.5

Table 2.2 - Earth Observation Satellites: optical/IR and radar current missions. Satellites launched between 2001-2005.

2.2 Relevance of Very-High-Resolution (VHR) optical data

Remote sensing data with spatial resolutions of $0.5 - 10$ m are required to adequately define the high wavenumber detail which characterizes the urban scene. This level of spatial resolution corresponds to scales between $1 : 10,000$ and $1 : 25,000$ (ignoring the effects of relief distortion etc.), that are typical of projects dealing with urban planning (Donnay *et al.* (2001)). The use of very high resolution (VHR) images, however, brings with it some major problems:

1. the majority of the highest resolution images are presently recorded in panchromatic mode only,

2. the corresponding large data set creates difficulties in terms of image storage, data exchange and processing time.

The first problem could be solved by using different data fusion techniques, i.e. merging the higher resolution panchromatic data with lower resolution multispectral data (Ackerman (1995), Jones *et al.* (1991), Wald *et al.* (1997)). Transforms in the color space (RGB/HSI), principal components analysis, spatial filtering and wavelets methods are among the most common means of achieving such integration (Carper *et al.* (1990), Chavez and *et al.* (1990) and Pohl and Van Genderen (1998)). Data fusion methods are also suitable for processing merged data sets involving sensors with different spatial resolutions and physical measurements such as radar and optical system (Schiavon *et al.* (2003)). This synergy can expand the optical system limits and improve the relatively poor spectral information related to the high-resolution data and improve the classification accuracy. In fact, in non-urban areas, it is generally possible to derive relatively direct relationships between the spectral responses in the four MS bands of natural components such as water, vegetation and soil and the measured image pixels reflectance values. In urban area, identi-

cal spectral reflectance values correspond to very different surface structures in terms of materials and fabrics, complicates the feature extraction process and the final output map.

The problems concerning the volume of data might appear of secondary importance given the rapid and continuing improvements in computer technology. Nevertheless, they remain significant at the scale of the urban region (typically several tens of square kilometers) and they necessitate the consideration of data compression techniques. In this context, wavelet and other transform methods are promising avenues of research. They not only allow efficient data compression while preserving the original spectral values, but they can also be used to fuse images at different resolutions, thereby simultaneously dealing with the two main problems outlined above.

2.3 Features of data available today

Satellite Name	*Source*	*Launch*	*Sensors*	*Types*	*No.of Channels*	*Resolution (meters)*
RADARSAT-2	Canada	2007	SAR	Radar	1	3
CARTOSat-2	India-US	2007	PAN	PAN	1	1
EROS-B2	ImageSat	2007	PAN	PAN	1	0.7
EROS-C	ImageSat	2008	PAN	PAN	1	0.7
			MS	MS	n/a	2.8
WorldView-1	DigitalGlobe	2007	PAN	PAN	1	0.5
TerraSAR-X	Germany	2007	SAR	Radar	1	1

Table 2.3 - Earth Observation Satellites: optical/IR and radar future missions. Satellites missions for 2007 and later.

Chapter 3

Classification Methods of High Resolution Remote Sensing Data

As discussed in 2.2 *Supervised* classification is the procedure most often used for quantitative analysis of remote sensing image data. It is based on using suitable algorithms to label pixels in an image representing particular ground cover types, or classes. A variety of algorithms is available for this analysis, ranging from those based on probability distribution models for the classes of interest to those in which the multipsectral space is partitioned into class-specific regions using optimally located surfaces. Irrespective of the particular method chosen, the essential practical steps are:

- Decide the set of ground cover types into which the image is to be subdivided. These are the information classes, for example, water, urban regions, vegetated areas, etc.
- Choose representative pixels for each of the desired classes. These pixels form the training set. Often the training pixels for a given class are in

a common region enclosed within a border often called the Region of Interest (ROI);

- Use the training data to estimate the values of the particular classifier algorithm to be used; these parameters will be the properties of the probability model used to define partitions in the multispectral space. The set of parameters for a given class is sometimes called the *signature* of that class;

- Using the trained classifier, label or classify every pixel in the image into one of the desired ground cover types (information classes). Here the whole image is classified;

- Produce tabular summaries (typically the confusion or accuracy matrices) and thematic or classification maps which summarise the results of the classification.

In the recent literature, many papers have addressed the development of novel techniques for the classification of high resolution remote sensing images. In (Unsalan and Boyer (2004)), the authors present a technique for the identification of land developments over regions. The proposed technique uses straight lines, statistical measures (length, orientation, and periodicity of straight line), and spatial coherence constraints to identify three classes, namely 1) urban; 2) residential; 3) rural. In (Shackelford and Davis (2003)), a standard maximum-likelihood classifier is used to discriminate four spectrally similar macroclasses. Subsequently, each macroclass can be hierarchically subdivided according to class-dependent spatial features and a fuzzy classifier.

The main problem of these techniques is that they are highly problem dependent. This means that they cannot be considered as a general operational tool. In (De Martino *et al.* (2003)), the authors analyze the effectiveness of the gray-shade co-occurrence matrix (GLCM) texture features in modeling the spa-

tial context that characterizes high-resolution images. The fact that the analysis depends on the sample window and different heuristic parameters along with the intrinsic inability to model the shape of the objects leads to unsatisfactory classification accuracies. A more promising family of approaches for the analysis of high resolution images is inspired by the behaviour of the human vision system, is based on an object-oriented analysis and/or multilevel/multiscale strategies. In these approaches each image is made up of interrelated objects of different shapes and sizes. Therefore, each object can be characterized with shape, topological measures, and spectral features. Objects can be extracted from images according to one of the standard segmentation techniques proposed in the literature (Haralick and Shapiro (1985)). The main idea of this multilevel analysis is that for each level of detail, it is possible to identify different objects that are peculiar to that level and should not appear in other levels. In other words, each object can be considered to be its "optimal" representation level. Moreover, other aspects considered in this analysis are: 1) that objects at the same level are logically related to each other and 2) that each object at a generic level is hierarchically related to those at higher and lower levels (Binaghi *et al.* (2003), Benz *et al.* (2004), Burnett and Blaschke (2003)).

For example, in the multiscale analysis of a high resolution image, using the highest resolution, we can identify houses, gardens, streets, and single trees; at moderate levels, we can identify urban aggregates, group of trees, and agricultural fields; finally, at the coarsest level, we can identify towns and cities, forests, and agricultural areas as single objects. The exploration of the hierarchical tree results in a precise analysis of the relations between these objects. For example, we can count the number of houses that belong to a specific urban area (Benz *et al.* (2004)). In (Binaghi *et al.* (2003)), the authors propose an approach based on the analysis of a high-resolution scene through a set of concentric windows. The concentric windows analyze the pixel under investigation and the effects of its neighbors at different resolutions. To reduce

the computational burden, the information contained in each analysis window is compressed using a Gaussian pyramidal resampling approach. The classification task is accomplished by a soft multilayer perceptron neural network that can be used adaptively as a pixel-based or an area-based classifier. One of the limitations of this approach is the fixed shape and choice of the analysis window size. In (Shackelford and Davis (2003)), an object-based approach is proposed for classification of dense urban areas from pan-sharpened multispectral Ikonos imagery. This approach exploits a cascade combination of a fuzzy-pixel classifier and a fuzzy object-based classifier. The fuzzy pixel-based classifier uses spectral and simple spatial features to discriminate between roads and buildings, which are spectrally similar. Subsequently, a segmented image is used to model the spectral and spatial heterogeneities and to improve the overall accuracy of the pixel-based thematic map. Shape features and other spatial features (extracted from the segmented image) as well as the previously generated fuzzy classification map are used as inputs to an object-based fuzzy classifier.

In (Benediktsson *et al.* (2003)), morphological operators are exploited within a multi-scale approach to provide image structural information for automatic recognition of man made structures. In greater detail, the structural information is obtained by applying morphological operators with a multi-scale approach and analyzing the residual images obtained as a difference between the multiscale morphological images successive scales. A potential problem with this technique is the large feature space generated by the application of a series of opening and closing transforms. In (Benediktsson *et al.* (2003)), the authors overcome this problem by proposing the use of different feature-selection algorithms. An adaptive and supervised model for object recognition is presented in (Binaghi *et al.* (2003)), where a scale-space filtering process models and a multi-scale analysis for feature extraction is integrated to a unified framework within a multilayer perceptron neural network. This means that the error back-

propagation algorithm used to train the neural network also identifies the most adequate filter parameters. The main problems of this technique are related to the choice of the number and type of filters to be used in the input filtering layer (first layer) of the neural network.

In (Mott *et al.* (2002)), an algorithm based on selective region growing is proposed to classify a high-resolution image. In the first step, the image is classified by taking into account only spectral information. In the second step, a classification procedure is applied to the previous map by taking into account not only spectral information but also a pixel distance condition to aggregate neighboring pixels. By iteration, neighbor pixels that belong to the same class grow in a selective way, obtaining a final classification map. Nevertheless, as mentioned before, the few techniques specifically developed for the automatic analysis of high resolution images (compared with the abundant literature on the classification of moderate-resolution sensors), do not exhibit sufficient accuracy to satisfy the the end-user requirements in all application domains.

3.0.1 Unsupervised Neural Network Classification Algorithms: The Self-Organizing Map

In the context of the *Unsupervised* classification algorithms using NN, it is worthwhile to highlight the Self-Organizing map. Among the architectures and algorithms suggested for artificial neural networks, the Self-Organizing Map has the special property of effectively creating spatially organized "internal representation" of various features of input signals and their abstractions (Kohonen (1990)). Neighboring cells in a neural network compete in their activities by means of mutual lateral interactions, and develop adaptively into specific detectors of different signal patterns. In this case, learning is called competitive, unsupervised or self-organizing. The Self-Organizing Map is a sheet-like neural network, the cells of which becoming specifically tuned to various input signal

patterns or classes of patterns through an unsupervised learning process. The locations of the responses tend to become ordered as if some meaningful coordinate system for different input features were being created over the network. The spatial location or coordinates of a cell in the network then correspond to a particular domain of input signal patterns. Each neuron or local neuron group acts like a separate decoder for the same input. It is thus the presence or absence of an active response at that location, and not the exact input-output signal transformation or magnitude of the response, that provides an interpretation of the input information. Due to their characteristics, the Self-Organizing Maps (SOMs) are suitable for pattern recognition problems (e.g. image classification). In particular, the training phase is very fast compared to the supervised NN training phase (i.e. an iterative procedure). The SOMs have been succesfully used in different application domains with QB images: e.g. for detection of "man made" structures and changes (Molinier *et al.* (2006)), and to distinguish different kind of asphalt surfaces (e.g. concrete etc.) (Del Frate *et al.* (2004)).

3.1 Summary of methods used for classification by spectral radiances

3.1.1 The Neural Net method: Performance and comparison with other method

The use of neural networks in remote sensing has often been found effective, since they can simultaneously handle non linear mapping for a multidimensional input space onto the output classes and can cope with complex statistical behaviour (Dawson (1994)). Neural networks, in contrast to statistically-based classifiers, do not require an explicitly well defined relationship between the input and output vectors, since they determine their own input-output rela-

tions directly from a set of training data (Rumelhart *et al.* (1986)). There are numerous types of neural networks used in classifying remotely sensed data. The most commonly used NN is the Multi Layer Perceptron (MLP)(Lippman (1987)) which has been found to have the best topology for the classification and inversion problems (Hsu *et al.* (1992)), trained by the backpropagation (BP) algorithm (Del Frate *et al.* (2000)). Minimization of the error function can also be pursued by a scaled conjugate gradient algorithm (Bishop (1995)). This is a member of the class of conjugate gradient methods (general purpose second order techniques) that help to minimize goal functions of several variables. Second order indicates that such methods use the second derivatives of the error function, while a first order technique (like standard back-propagation) only uses the first derivatives (Del Frate *et al.* (1999)).

3.1.2 The Neural Network algorithm

The results shown in the following sections are also published in (Del Frate *et al.* (2006)) and Appendix. NN models are mainly specified by the net topology and training rules (Lippman (1987)). The term topology refers to the structure of the network as a whole: the number of its input, output, and hidden layers and how they are interconnected. Among various topologies, multilayer perceptrons (MLP) have been found to have the best suited topology for classification and inversion problems (Hsu *et al.* (1992)). These are feed-forward networks where the input flows only in one direction to the output, and each neuron of a layer is connected to all neurons of the successive layer but has no feedback to neurons in the previous layers. As far as the numbers of hidden layers and of their sub units are concerned, the topology providing the optimal performance should be selected. In fact, if the number of neurons is too small, the input-output associative capabilities of the net are too weak. On the other hand, this number should not be too large; otherwise, these capabilities might show a lack of generality being too tailored to the training set which unecessar-

ily increases the computational complexity of the algorithm. It turns out that a fair compromise can be found.

The number of hidden layers is another issue to be considered. It has been shown that networks having two layers of weights, i.e. one hidden layer of neurons, and one of sigmoidal hidden units can approximate arbitrarily well any functional continuous mapping, provided the number of hidden units is sufficiently large (Bishop (1995), Hornik *et al.* (1989)). However how much the inclusion of an additional hidden layer might improve the classification performance is still an open issue.

In this work we followed a rather heuristic approach. We systematically analyze the performance of the network varying either the number of hidden layers (one or two) or the number of hidden units and selecting the best topology on the basis of quantitative results. Once the network topology is selected, the weight or strength of each connection has to be determined via learning rules to approximate an unknown input-output relation. These rules indicate how to pursue the minimization of the error function measuring the quality of the network's approximation of the restricted domain covered by a training set (i.e., a set of input-output examples). A typical error function which can be considered in this context is the sum-of-squares error function (SSE) (Bishop (1995)), given by a sum over all patterns, and over all outputs, of the form:

$$SSE = \sum_{n=1}^{N} \sum_{k=1}^{c} \{y_k(\mathbf{x^n}, \mathbf{w}) - t_k^n\}^2 \tag{3.1}$$

Here $y_k(\mathbf{x^n}; \mathbf{w})$ represents the output of unit k as a function of the input vector $\mathbf{x^n}$ and the weight vector $\mathbf{w}$, N is the number of testing patterns, and c is the number of outputs. The quantity t_k^n represents the target value for output unit k when the input vector is $\mathbf{x^n}$. In our case the minimization of the error function has been carried out using a scaled conjugate gradient (SCG) algorithm (Moller (1993)). This is a member of the class of conjugate gradient methods, general purpose second order techniques that help to minimize goal

functions of several variables.

It should be mentioned that most of the neural net simulations were provided by the SNNS (Stuttgart Neural Network Simulator) package (Stuttgart Neural Network Simulator (1995)). For the specific purpose of our image classification a training-set with a statistically significant number of pixels for each class has been generated. The training of the neural network has then been carried out by feeding it with pairs of vectors (patterns): the input vector contains the reflectances of the different channels of the multi-spectral image, the output vector contains the corresponding known class of the surface. To avoid saturation within the network it has been necessary to scale all the values of the input vectors in the range between -1 and 1. This is also helpful to mitigate single-image effects if pixels belonging to different images are included in the training set. At the same time, the component of the output vector corresponding to the true class has been set to 1 while the others go to 0. Once the NN have been trained, they have been used for the classification of new data not used in the training set. In the test phase, a competitive approach (winner-take-all) has been considered to decide on the final classification response.

3.1.3 Single image classification: MLP near-optimal structure

The first applications of NNs in satellite image classification (Benediktsson *et al.* (1990)) established an additional merit of their use compared to conventional approaches (e.g. Maximum Likelihood): the ability to easily incorporate non-spectral ancillary information (e.g. topographic) into the classification process without violating any assumption, and as previously showed, NN outperformed statistical classifiers on several occasions (Kanellopoulos *et al.* (1992)). Nevertheless, there is no exact solution to find the optimal topology, in terms of resulting classification accuracy, of a MLP, until today. There have been several heuristics, some developed in the context of remotely sensed image classification specifically (Kanellopoulos and Wilkinson (1997)), that could

be the basis for exploring more efficient topologies by trial and error. These rules of thumb insure that the topologies, and thus results attained, will not deviate much from optimum. However, when the classification results are extracted there is always the uncertainty related to the topology used and the classification accuracy (e.g., another topology could yield superior accuracy). There is no guarantee that a near-optimal topology has been found to perform the classification. The difficult part in determining an optimum MLP topology lays in hidden structures identification i.e. the identification of the best number of hidden neurons for each layer. Currently, there exist two general categories of approaches, pruning or growing algorithms (Lauret *et al.* (2006)) on one side and heuristics, which give us the hidden structure as a function of the number of input and output nodes, on the other (Kanellopoulos and Wilkinson (1997)). In this work we use the growing method: nodes are added progressively, e.g. starting with a single hidden node, until performance can no longer be improved.

The QuickBird commercial remote sensing satellite provides images consisting of four multi-spectral (MS) channels with 2.4 m resolution and a single panchromatic (PAN) band with a 0.62 m resolution. The four MS bands collect data at red, green, blue, and near-infrared wavelengths, and the data in each band is stored with an 11-bit quantization. As previously discussed, the spatial resolution plays a key role in urban monitoring related to the ability to detect fine-scale objects present in urban scenes. In particular, high resolution is a requirement to reduce the problem of mixed pixels (i.e. the pixels that represent the spectral signature of more than one class due to the insufficient resolution of the sensor) present in the medium resolution images (e.g. Landsat imagery). However, the high resolution sensors have a limited spectral resolution, depending on technical constraints, that further increase the classification problem (Schowengerdt (2002)) and often do not allow a complete characterization of different roof types, having different spectral signatures. However, a

method developed exploiting only the Digital Number (DN) belonging to representative pixels of each class randomly selected on QB images, is accurate in terms of classification results.

A QB image taken over the Tor Vergata University campus, located in Italy, South-East of Rome, on March 13, 2003, will be referred to as QB1. A view of the area obtained with the Red, Green and Blue bands is shown in Fig. 3.1.

Figure 3.1 - *QuickBird image of the Tor Vergata University Campus, Rome, and its sourrondings.*

Besides the buildings of the campus, different residential areas belonging to the outskirts of the south-east side of the city can be distinguished in the image. Our first purpose was to design an optimum neural network able to classify the multi-spectral image. The land cover classes considered were buildings, roads, vegetated areas, and bare soil where the latter class includes not asphalted road and artificial excavations. The inclusion of additional classes was discarded for several reasons: the considered classes are those that best describe the area under consideration and are in themselves sufficient to detect significant features. The choice of a small number of classes enables a more

quantitative comparison of the performance obtained using a single net for a single image classification compared to the use of a single net for multiple image classification. In this latter case we think that the choice of a number of 4 classes represents a rather ambitious target. It also has to be noted that a recent study analysing satellite image classification experiments over fifteen years pointed out that using a larger number of classes increases the difficulty in the classification, which is frequently not supported by the experimental results shown in the study (Wilkinson (2005)). Once the classification problem was been configured, a first investigation consisted of analysing the spectral behaviour of the different surfaces. The selected pixels characterizing one class belonged to polygons manually drawn in the image. It should be noted that, at the very high resolution of the images, the edges or boundaries between individual land cover objects were fairly sharp and it was usually easy to locate and assign a specific pixel to a land cover class. The mean values of the spectral signatures of the 4 categories are shown in Fig. 3.2.

The figure clearly discriminates between the classes. This results from the spectral properties related to the different molecular resonance mechanisms which characterize the surface materials. Using the same data for the sensitivity analysis we were able to generate a training-set with a statistically significant number of pixels for each of the four categories. Previous studies have shown that the training set, notably in terms of its size and compositions, can have a marked impact of the classification accuracy (Foody *et al.* (1995)). The training datasets were generated considering about $24,400$ pixels. The design of the network put particular emphasis on the selection of the number of hidden units to be considered in the net. To this purpose the plot in Fig. 3.3 was produced, where the SSE value over a test set of more than $1,000$ patterns is reported that corresponds to different numbers of hidden units. It can be seen that, if we consider both the SSE error and the network complexity, the best results were obtained with a $4 - 20 - 20 - 4$ topology. Indeed, the increase of the

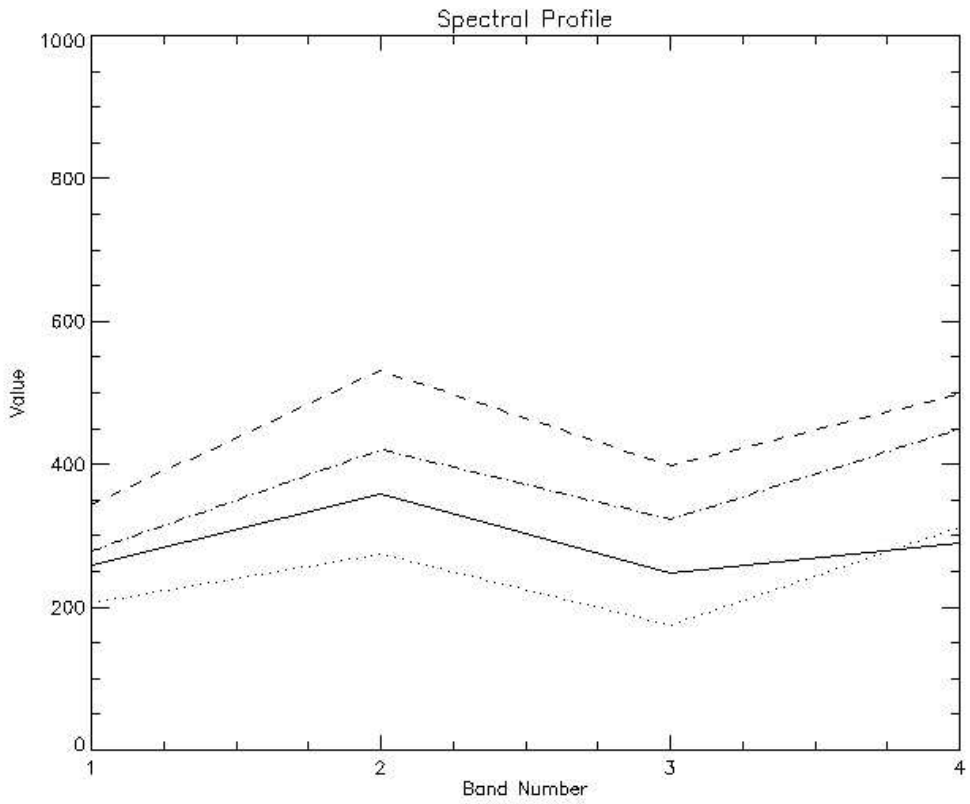


Figure 3.2 - *Spectral analysis from Quickbird image QB1 for the classes buildings (dashed line), asphalted surface (solid line), bare soil (dash-dotted line), vegetated areas (dotted line).*

number of hidden units did not significantly change the SSE error.

A similar plot is reported in Fig. 3.4 where now a single hidden layer is considered. Again the best result are obtained with around 20 neurons in the hidden layer, however this topology is slightly worse if compared with the two-hidden layer topology. This indicates that the second layer is able to extract additional information from what is already discriminated by the first one. The topology $4-20-20-4$ was then finally selected and used to classify the entire image ($3,506,832$ pixels).

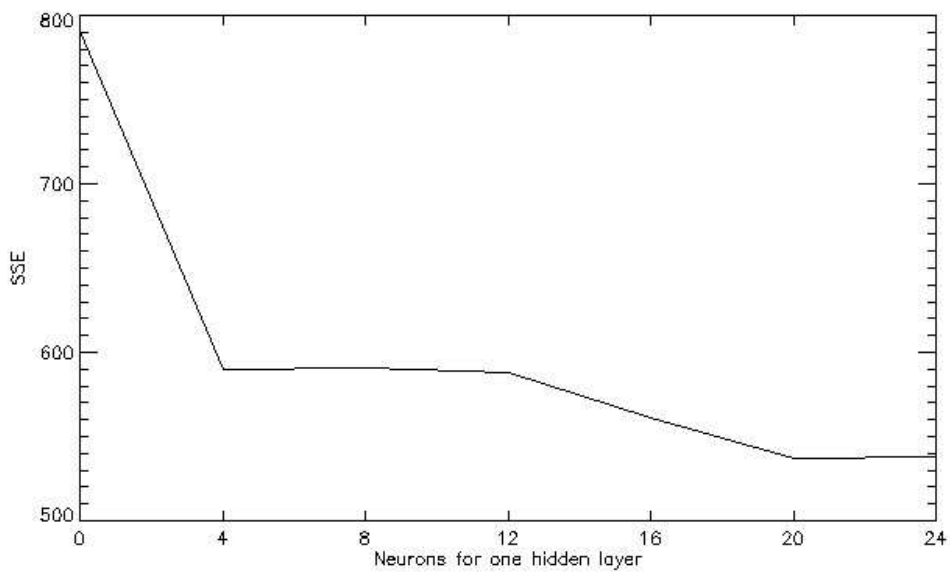


Figure 3.3 - *SSE values calculated over the test set changing the number of hidden neurons in a two hidden layers topology. The number of units is the same in both layers.*

Fig. 3.5 shows the classification map derived using this procedure. The classification accuracy has been assessed by visual comparison with the original high resolution image and by direct inspections on site. We stress the fact that our working area is located at the Tor Vergata University campus, that is almost at the centre of image QB1, allowing direct inspection on site. A ground truth map, corresponding to a subset of the image, has been manually interpreted. We observed that the classification provided by the network is accurate due

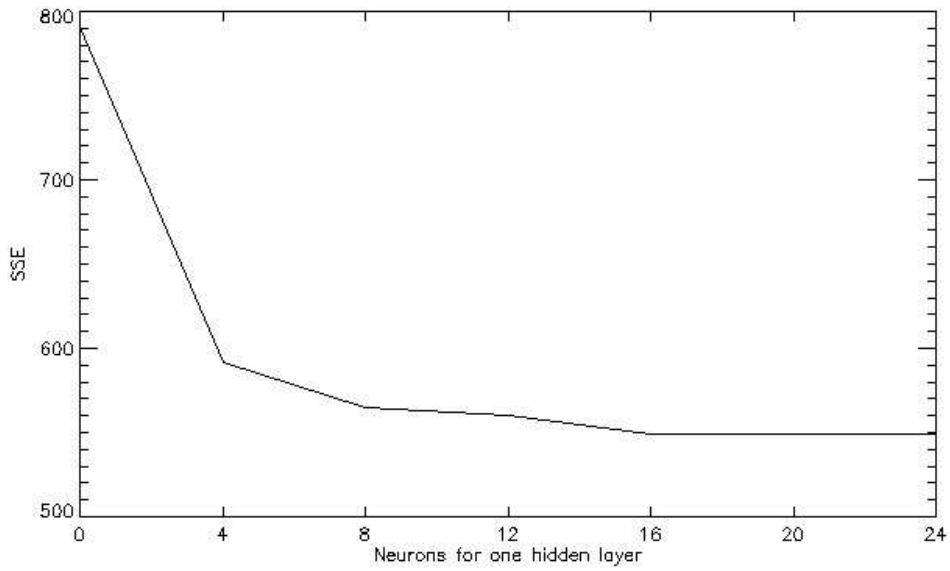


Figure 3.4 - *SSE values calculated over the test set changing the number of hidden neurons in a one hidden layer topology.*

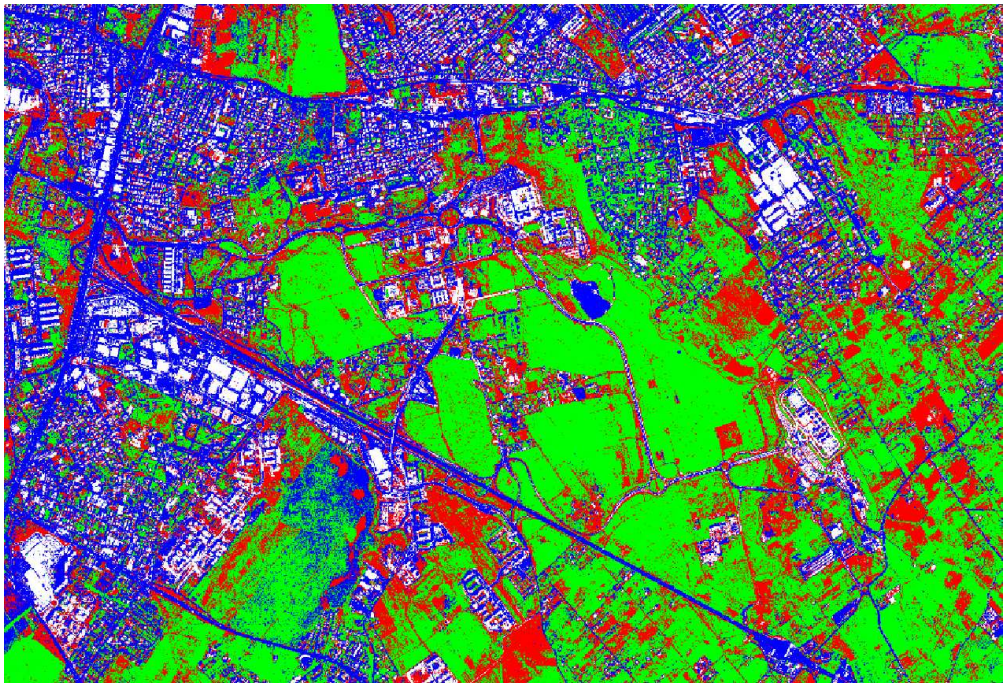


Figure 3.5 - *Classification map of the image QB1 using the optimized topology. Red: bare soil, blue: asphalted surface, white: buildings, green: vegetated areas.*

to its high-level of resolution and we reached a 93% accuracy in the subimage considered.

The whole confusion matrix is reported in 3.1.

Classified as	*True*			
	Vegetated Areas	Asphalt	Building	Bare Soil
Vegetated Areas	14864	33	750	2207
Asphalt	132	44785	68	27
Building	1225	29	12634	783
Bare Soil	230	2	512	3229

Table 3.1 - *Confusion matrix obtained for image QB1 with the* $4-20-20-4$ *topology. Overall number of pixels:* 81510. *Overall Error:* 5998 *pixels (*7.36%*).*

Once the network topology for this kind of problem has been optimized and the performance assessed, we move to investigate the capability of a unique network to provide the classification of different images rather than of a single one. To underline the complexity of this new problem we tested the designed network, developed from the QB1 image, on another QB image. The choice of this new image should follow some similarity criteria with respect to the already classified one. For example it would not be very meaningful to consider a new image characterized by very different land cover classes, such as water, which does not appear in the QB1 image (hence not integrated at all by the network during its training process). Failure of the neural network in this case can be taken for granted and this test would not provide an evaluation of the network generalization capability. Therefore, we decided to choose as a test image a QB image quite similar to QB1. Indeed, the new QB image (QB2) is taken of the same area as the first one, but in a different season and at a slightly different incident angle. In 3.2 we summarize the basic information of the two images analyzed so far and of those that will be considered in the following. If

the trained network fails in its application to this image it will be unlikely to succeed with many other QB images.

Code	Acquisition Date	Dimension (pixels)	Off Nadir Angle	Location
QB1	03/13/2003	$2415 * 1650$	8 Degrees	Rome, SE outskirts
QB2	05/29/2002	$2352 * 1491$	11 Degrees	Rome, SE outskirts
QB3	07/19/2004	$2415 * 1650$	23 Degrees	Rome, NE outskirts
QB4	07/19/2004	$2415 * 1450$	23 Degrees	Rome city
QB3	07/22/2005	$2223 * 1450$	12 Degrees	Nettuno town

Table 3.2 - *Characteristics of the Quickbird images used in the work. All the acquisition times are between* $10:00 - 10:30$ *a.m.*

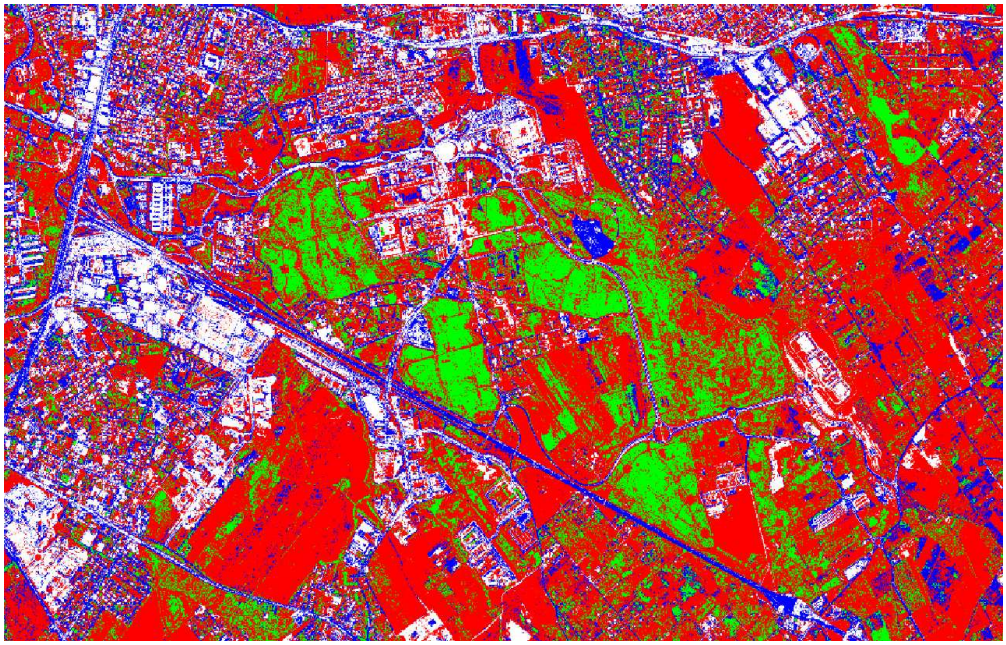


Figure 3.6 - *Automatic classification map from the image QB2 with a net trained with examples taken from image QB1. Red: bare soil, blue: asphalted surface, white: buildings, green: vegetated areas.*

In Fig. 3.6 we show the result of the classification of the QB2 image by

using the net trained on patterns retrieved from image QB1. For the sake of completeness and for a better interpretation of the results we also produced the classification, reported in Fig. 3.7, that would be obtained applying to the image QB2 the single image classification method developed for image QB1 (a network $(4 - 20 - 20 - 4)$, trained with examples from QB1). The classification map shown in Fig. 3.7 seems, as expected, rather accurate. Indeed the misclassification percentage computed over the same image subset considered for QB1 is 95% similar to the one obtained in the former case. The classification result shown in Fig. 3.6 is completely different. Although the network recognizes many patterns and assigns the correct class to the corresponding pixels, entire objects are misclassified, the bare soil class and the built areas class are definitely overestimated and the general noise level produced by the classification is significantly increased. From a quantitative point of view the misclassification rate computed over the subset test image is 56%. Fig. 3.8

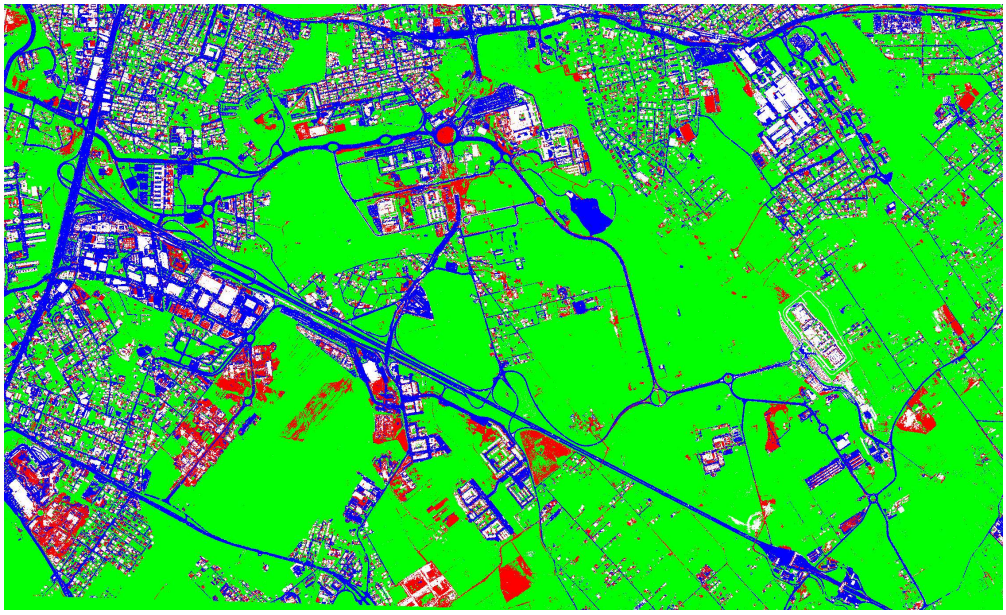


Figure 3.7 - *Classification map from the image QB2 using a network trained with examples taken from image QB1. Red: bare soil, blue: asphalted surface, white: buildings, green: vegetated areas.*

contributes to understanding the classification performance. We observe that even if the shapes of the signatures resemble those plotted in Fig. 3.2, where we can still discriminate between classes, the ranges of the digital number values are significantly different, generating confusion when the network generates its classification response. Thus, the classification of the QB2 image obtained using a network trained on another image, even if taken on the same scenario, is not adequate. This means that to design a network able to provide good classification accuracy for images not used in the training phase is an ambitious goal, even if the classification is performed for a limited number of classes.

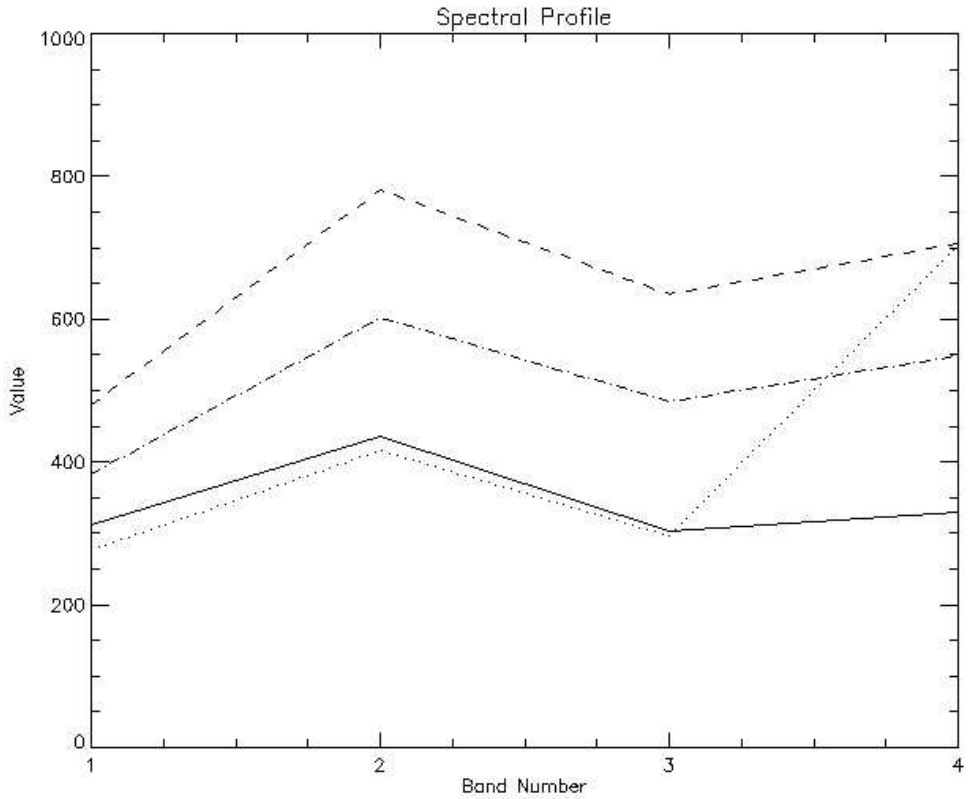


Figure 3.8 - *Spectral analysis from image QB2 for the training set classes: buildings (dashed line), asphalted surface (solid line), bare soil (dash-dotted line), vegetated areas (dotted line).*

Chapter 4

Methods to "globally" characterize a (large) urban area

4.1 The Discrete Fourier Transform

4.1.1 The Discrete Spectrum

Consider now the problem of finding the spectrum (i.e. of computing the Fourier transform) of a sequence of samples. This is the first stage in our computation of the Fourier transform of an image. Indeed, the sequence of samples to be considered here could be looked at as a single line of pixels in a digital image. Here we made the assumption that the spectrum of a set of samples is itself a continuous function of spatial wavenumber (ν). For digital processing clearly it is necessary that the spectrum itself also be represented by a set of values, that would, for example, exist in computer memory. Therefore we have to introduce a suitable sampling function in the wavenumber domain. For this purpose consider an infinite periodic sequence of impulses in the wavenumber domain spaced by $\Delta\nu$. It can be shown that the inverse transform of this se-

quence is another sequence of impulses in the spatial domain, spaced $\Delta\nu$ apart. In this case we are going from the wavenumber domain to the spatial domain rather than vice versa.

4.1.2 Discrete Fourier Transform Formulae

Let the sequence $\phi(k)$, $k = 0, ..., K$ be the set of K samples taken of $f(x)$ over the sampling period 0 to x_0. The samples correspond to distance kX. Let the sequence *F(r)*, $r = 0, ..., K-1$ be the set of samples of the wavenumber spectrum. These can be derived from $\phi(k)$ by suitably modifying the Fourier transform:

$$X(\kappa) = \int_{-\infty}^{\infty} f(x) e^{-j\omega x}\, dx \tag{4.1}$$

In fact, the integral over time can be replaced by the sum over k=0 to K-1, with dx replaced by Δx, the sampling increment. The continuous function *f(x)* is replaced by the samples $\phi(k)$ and $\kappa = 2\pi f$ is replaced by $2\pi r \Delta\nu$ with $r = 0, ..., K-1$. Thus $\kappa = 2\pi r / X_0$. The spatial variable X is replaced by $kX = kX_0/K$, $k = 0, ..., K-1$. With these changes 4.1 can be written in sampled form as

$$F(r) = X \sum_{k=0}^{K-1} \phi(k) W^{rk}, \qquad r = 0, ..., K-1 \tag{4.2}$$

with

$$W = e^{-j2\pi/K}. \tag{4.3}$$

Equation 4.2 is known as the *discrete Fourier transform* (DFT). In a similar manner a *discrete inverse Fourier transform* (DIFT) can be derived that allows reconstruction of the spatial sequence $\phi(k)$ from the wavenumber samples *F(r)*. This is

$$\phi(k) = X_0 \sum_{r=0}^{K-1} F(r) W^{rk}, \qquad k = 0, ..., K-1 \tag{4.4}$$

Sobstitution of 4.1 into 4.4 shows that those two expressions form a Fourier transform pair. This achieved by putting k=l in 4.4 so that

$$\phi(l) = \frac{1}{X_0} \sum_{r=0}^{K-1} F(r) W^{-rl} = \frac{1}{X_0} \sum_{r=0}^{K-1} \cdot X \sum_{k=0}^{K-1} \phi(k) W^{r(k-l)} = \frac{1}{K} \sum_{k=0}^{K-1} \phi(k) \cdot \sum_{k=0}^{K-1} W^{r(k-l)} \tag{4.5}$$

The second sum in this expression is zero for $k \neq l$; when $k = l$ is K, so that the right hand sid of the equality then becomes $\phi(l)$ as required. An interesting aspect of this development has been that X has cancelled out, leaving $1/K$ as the net constant from the forward and inverse transforms. As a result 4.2 and 4.4 could conveniently be written

$$F(r) = \sum_{k=0}^{K-1} \phi(k) W^{rk} \qquad r = 0, ..., K-1 \tag{4.6}$$

$$\phi(k) = \frac{1}{K} \sum_{r=0}^{K-1} F(r) W^{-rk} \qquad k = 0, ..., K-1 \tag{4.7}$$

4.1.3 Properties of the Discrete Fourier Transform

Three properties of the discrete Fourier transform and its inverse are of importance here:

- *Linearity*: Both the DFT and DITF are linear operations. Thus if $F_1(r)$ is the DFT of $\phi_1(k)$ and $F_2(r)$ is the DFT of $\phi_2(k)$ then for any complex constants a and b, $aF_1(r) + bF_2(r)$ is the DFT of $a\phi_1(k) + b\phi_2(k)$.

- *Periodicity*: from 4.3, $W^K = 1$ and $W^{kK} = 1$ for k integral. Thus for $r' = r + K$

$$F(r') = X \sum_{k=0}^{K-1} \phi(k) W^{(r+K)k} = F(r) \tag{4.8}$$

Thus in general

$$F(r + mK) = F(r) \tag{4.9}$$

$$\phi(k + mK) = \phi(k) \tag{4.10}$$

where lm is an integer. Thus both the sequence of spatial samples and the sequence of wavenumber samples are periodic with period K. This is consistent with the development of Sect. 4.1.1 and has two important implications. First, to generate the Fourier series components of a periodic function, samples need only be taken over one period. Secondly, sampling converts an aperiodic sequence into a periodic one, the period being determined by the sampling duration.

- *Symmetry*: Let $r' = K - r$ in 4.1, to give

$$F(r') = X \sum_{k=0}^{K-1} \phi(k) W^{-rk} W^{kK} \tag{4.11}$$

Since $W^{kK} = 1$ this shows $F(K-r) = F(r)^*$ where $*$ represents the complex conjugate. This implies that the amplitude spectrum is symmetric about $K/2$ and the phase spectrum is antisymmetric (i.e. odd).

4.2 The Discrete Fourier Transform of an Image

4.2.1 Definition

In the foregoing section we have treated functions with a single independent variable. That variable could have been the position along a line of an image.

Now functions with two independent variables will be examined. Let

$$\phi(i,j), \qquad i,j = 0,...,K \tag{4.12}$$

be the brightness of a pixel at location i,j in an image of KxK pixels. The Fourier transform of the image, in discret form, is described by

$$\Phi(r,s) = \sum_{i=0}^{K-1}\sum_{j=0}^{K-1} \phi(i,j)exp[-j2\pi(ir+js)/K] \tag{4.13}$$

An image can be reconstructed from its transform according to

$$\Phi(r,s) = \frac{1}{K^2}\sum_{i=0}^{K-1}\sum_{j=0}^{K-1} \phi(i,j)exp[+j2\pi(ir+js)/K] \tag{4.14}$$

4.2.2 Evaluation of the Two Dimensional, Discrete Fourier Transform

Equation 4.13 can be written as

$$\Phi(r,s) = \sum_{i=0}^{K-1} W^{ir} \sum_{j=0}^{K-1} \phi(i,j)W^{js} \tag{4.15}$$

with $W = e^{-j2\pi/K}$ as before. The term involving the right hand sum can be recognised as the one dimensional discrete Fourier transform

$$\Phi(i,s) = \sum_{j=0}^{K-1} \phi(i,j)W^{js} \; i = 0,...,K-1 \tag{4.16}$$

In fact it is the one dimensional transform of the ith row of pixels in the image. The result of this operation is that the rows of an image are replaced by their Fourier transforms; the transformed pixels are then addressed by their wavenumber index s across a row rather the positional index j. Using 4.16 in 4.15 gives

$$\Phi(r,s) = \sum_{i=0}^{K-1} \phi(i,s) W^{ir} \tag{4.17}$$

which is the one dimensional discrete Fourier transform of the sth column of the image, after the row transforms of 4.16 have been performed. Thus, to compute the two dimensional Fourier transform of an image, it is only necessary to transform each row individually to generate an intermediate image, and then transform this by column to yield the final result. Both the row and column transforms would be carried out using the fast Fourier transform algorithm; it can be seen therefore that the number of multiplications required to transforms an image is $K^2 \log_2 K$.

4.2.3 The Concept of Wavenumber

Entries of the Fourier transformed image $\Phi(r,s)$ represent the composition of the original image in terms of its two-dimensional wavenumber components. Wavenumber is the spatial analog of the frequency of a signal in time. A sinusoidal signal with high frequency alternates rapidly, whereas a low frequency signal changes slowly with time. Similarly, an image with high wavenumber horizontal exhibits frequent changes of brightness along the respective axis. An image of a high density residential urban area would be a good example of high wavenumber spectrum. Both the east-west and north-south dimensions of the houses are likely to be characterized by high wavenumber. Typically an image is composed of both east-west and north-south wavenumbers of differing spectral strengths derived from the discrete Fourier transform. The fundamental wavenumber is located at the upper left hand pixel in $\Phi(r,s)$ (i.e.$\Phi(0,0)$) and is the average brightness value of the image. This is the component in the spectrum with zero wavenumber in both directions. Thereafter pixels of $\Phi(r,s)$ both east-west and north-south represent components with wavenumbers that increment by $\frac{1}{K}$ where the original image is of size $K \cdot K$. If the scale

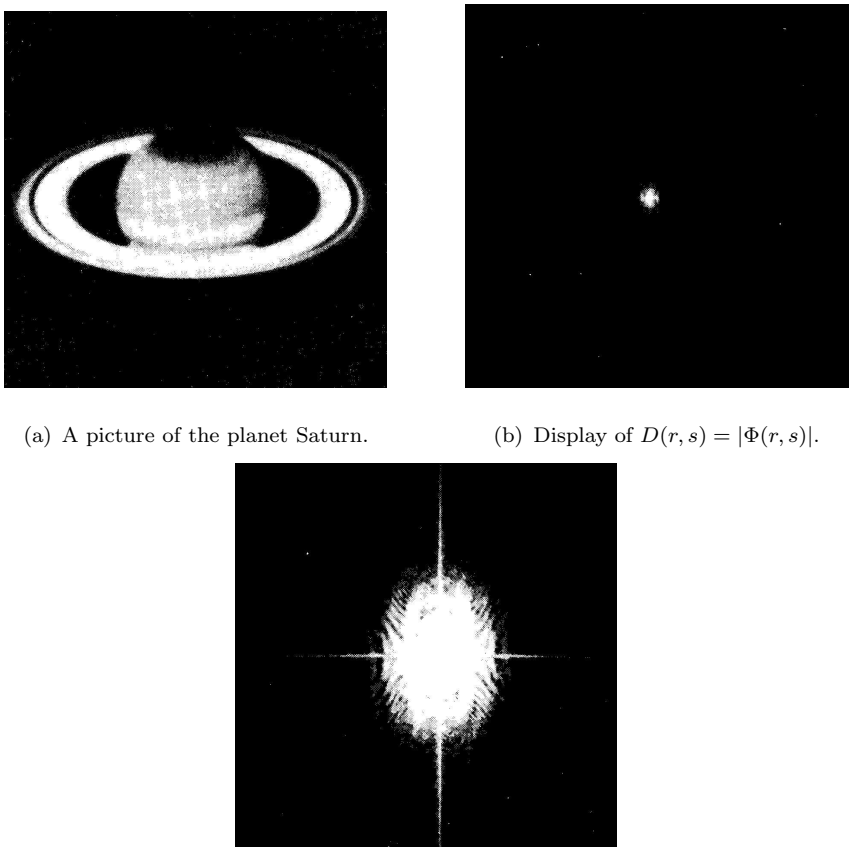
of the image is known then the wavenumber has dimension of length scale$^{-}1$ (i.e. $metres^{-1}$). For example the wavenumber increment for a $6250 \cdot 6250$ pixel image that covers 15 km image (i.e. QB MS) is 0.066667 mm^{-1}.

In Sect. 4.1.3 it was shown that the one dimensional discrete Fourier transform is periodic with period K. The same is true of the discrete two dimensional form. Indeed the $K \cdot K$ pixels of $\Phi(r, s)$ computed according to 4.13 can be viewed as one period of an infinitly periodic two dimensional array and the amplitude is symmetric about $K/2$. Similarly $\Phi(r, s)$ is symmetric about its centre. This can be demonstrated by showing that no new spectral amplitude information is shown by displaying pixels east-west and north-south beyond $K/2$. Rather than ignore them (since their accompanying phase is important) the display is adjusted to bring $\Phi(0, 0)$ to the centre. This is the lowest wavenumber of the Fourier transform array and represents the image average brightness value. Pixels away from the centre represent the regions of increasing wavenumber components in the image. This is the usual method of presenting two dimensional image transforms. Example of spectra displayed in this manner are given in the figures below. To make visible components with smaller amplitudes, a logarithmic amplitude scaling has been used, according to

$$D(r, s) = \log[1 + |\Phi(r, s)|]. \tag{4.18}$$

In general, the high wavenumber content of an image is associated with frequent changes of brightness with position. Edges, lines and some types of noise are examples of high wavenumber (often noise is by definition "white", meaning all wavenumbers). In contrast, gradual changes of brightness with position, such as are associated with more gradual spatial tonal variations, account for the low wavenumbers in the spectrum. In particular, with respect to the QB images, each input pixel in the four multispectral bands in the spatial domain is identified by the functions $f_1(x, y)$, $f_2(x, y)$, $f_3(x, y)$ and $f_4(x, y)$,

where (x, y) represent the pixel spatial coordinates, which are mapped in the wavenumber domain by $F_1(\nu_1, \nu_2)$, $F_2(\nu_1, \nu_2)$, $F_3(\nu_1, \nu_2)$, and $F_4(\nu_1, \nu_2)$.



(a) A picture of the planet Saturn.

(b) Display of $D(r, s) = |\Phi(r, s)|$.

(c) Display of $D(r, s) = \log[1 + |\Phi(r, s)|]$

.

Figure 4.1 - *Picture of the planet Saturn, its normal spectrum and the spectrum processed by adding* 1 *to* $|\Phi(r, s)|$*, taking the log and rescaling the values to the same gray scale used in displaying* $|\Phi(r, s)|$.

Chapter 5

Applications

5.1 Detailed classification of "elementary" surfaces using the optical/IR spectra

Three more QB images of Rome have been considered in this case for an overall number of 5 images. As shown in Tab. 3.2 the 5 images are of similar size but taken from different years, different sites and different seasons. Besides the QB1 and QB2 images centred on the Tor Vergata University campus, we have one image (QB3) looking at north-east suburbs, a fourth image quite close to the old town (QB4), and a fifth image (QB5) which has been taken on a small urban area a few *km* away from Rome. A pixel-based classification algorithm has again been implemented to distinguish among the four main classes: buildings, asphalt surfaces, vegetated areas and bare soils.

In a previous section we showed that a successful classification performance relies on proper training and design of the network. In particular, it is important that the patterns included in the training set significantly represent all potential scene elements that might be encountered during the application phase, in other words resemble the statistics of the image. To this purpose a larger archive of spectral signatures has been generated. Images QB1, QB3 and

QB5 have been considered for the training and about 26,000 examples have been collected for the generation of the network training set. Optimal performance both in terms of classification accuracy and training time was again determined by an extensive study whose results are illustrated in Fig. 5.1.

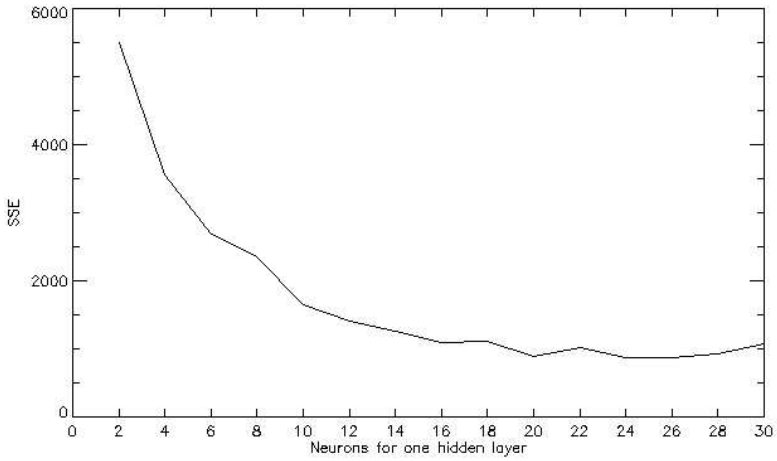


Figure 5.1 - *SSE values calculated over the test set changing the number of hidden neurons in a two hidden layers topology designed for the classification of a collection of Quickbird images. The number of units is the same in both layers.*

With regards to the number of hidden layers we relied on our previous result indicating that topology with two hidden layers is more effective so the final selected topology was again $4 - 20 - 20 - 4$. Indeed, with respect to the single image processing case, most of the physics characterizing the classification problem have not changed and those that have changed have only minor implications in terms of the topology to be selected. In Figures 5.2 and 5.3 we present the classification maps obtained by applying the trained neural network to the images QB2 and QB4, which did not contain any of the pixels included in the training sets. From both visual inspection of the original images and direct on site inspection we observed a general good agreement with the map generated automatically. All of the main features such as large roads and buildings are distinguishable with good precision even though some inaccuracies can be

noted in the objects edge detection, possibly caused by shadow effects. A more quantitative analysis, computed for the same sub-area of image QB2 considered in section 2 gave an overall accuracy rate is of about 87%.

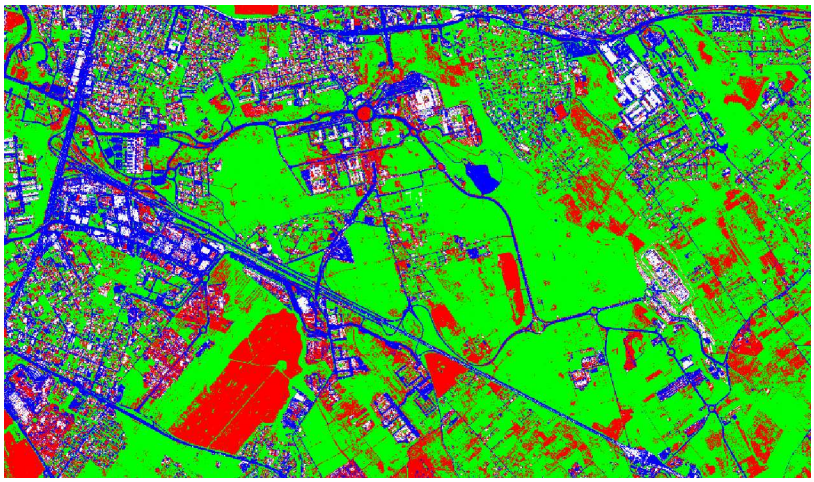


Figure 5.2 - *Automatic classification map from the image QB2 with a net trained with examples taken from other images. Red: bare soil, blue: asphalted surface, white: buildings, green: vegetation.*

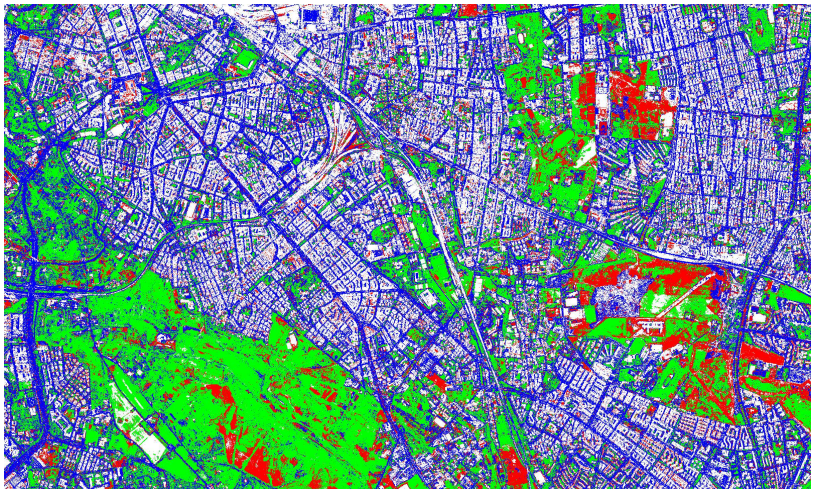


Figure 5.3 - *Automatic classification map from the image QB4 with a net trained with examples taken from other images. Red: bare soil, blue: asphalted surface, white: buildings, green: vegetation.*

5.2 Global characterization of large cities of parts of cities

Many techniques proposed in the literature are not general but are application dependent and specifically developed for addressing particular approaches (e.g. analysis of urban areas). Thus, the proposed approach is general and can be applied to any kind of high-resolution image and application domain. In fact, experimental results obtained on the data set of Rome that consists of different parts of the city and surroundings, which included the country side areas as vineyards and croplands, point out the effectiveness of the proposed methodology.

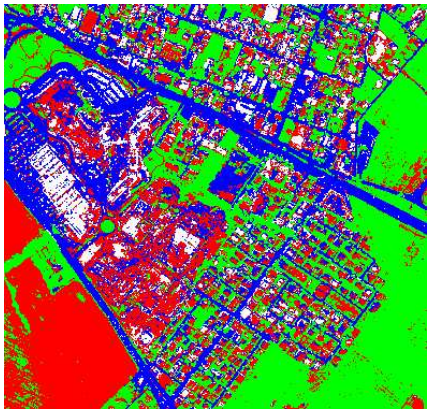
5.3 High Resolution Change Detection

Considering these encouraging results, and given the availability of two images (QB1 and QB2) over the same site, we tried to extend the described method to a typical change detection exercise. The two images have been co-registered using a set of 30 ground control points and using the earlier image as a master. We remind the reader that the time interval between the two images is one year. The two corresponding classification maps, obtained by means of the same network, have been used for the production of change detection maps. In particular, the change detection was evaluated in terms of the pixels that migrated from vegetation, bare soil or asphalt surface class to the building class in the time window considered. In this case we are more interested in an object based result, the final change detection mask was obtained after a post-processing which removed all clusters of pixels detecting changes but containing of less than 20 elements. The ground-truth confirmed that the changes corresponding to the primary detected structures were buildings constructed over the study time interval. An example of such a detection result

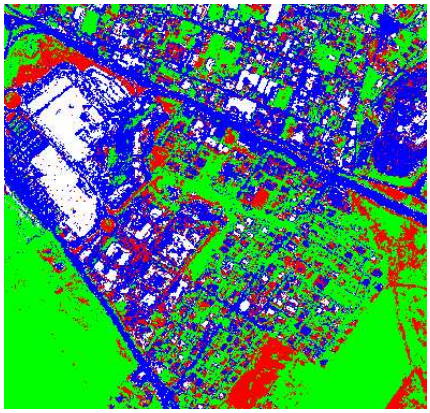
is shown in Fig. 5.4 where the previous corresponding classification maps are also presented. The corresponding confusion matrix, reported in Table 5.1 and computed on the basis of the ground-truth, gives an high percentage of pixels along the diagonal. On the other hand, most the pixels off the diagonal belong to the object circled in red which, rather than a real failure of the classification algorithm, is a consequence of an imperfect coregistration of the two images.

Classified as	*True*	
	Changed	Unchanged
Changed	14864	33
Unchanged	132	44785

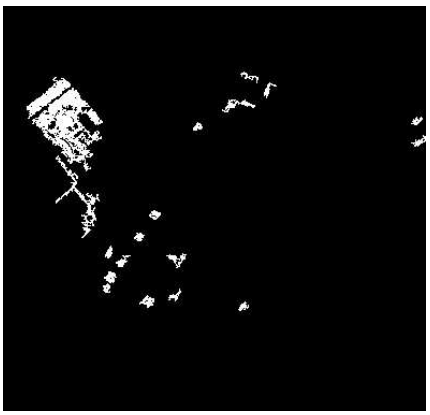
Table 5.1 - *Confusion matrix obtained for the change detection. Overall number of considered pixels:* 148538. *Overall accuracy* 93.2%.



(a) 2002 Classification map results. Red: bare soil, blue: asphalted surface, white: buildings, green: vegetaded areas.



(b) 2003 Classification map results. Red: bare soil, blue: asphalted surface, white: buildings, green: vegetaded areas.



(c) Change detection mask. White: changed features, black: unchanged features.

Figure 5.4 - *Change detection results.*

5.3.1 Transform algorithms to characterize urban areas

We use wavenumber spectra of QB images to characterize different areas from the cities: New York, Boston, San Francisco and Rome. We used the green wavelength of 0.56 μm (MS channel=2) which responds best to the spectral profiles of the primary "mineral" urban surfaces that have a high reflectance value around 0.5 μm (JPL Spectral Library Minerals (1991)). The images characteristics are given in Table 5.2:

City	*Acquisition Date*	*Spatial Resolution*	*Off Nadir Angle*
New York	08/20/2002	2.4 meters	20.3 Degrees
Boston	09/20/2005	2.4 meters	18.4 Degrees
San Francisco	02/12/2006	2.4 meters	19.8 Degrees
Rome	07/19/2004	2.4 meters	23 Degrees

Table 5.2 - *Characteristics of the selected images.*

It's worthwhile to note that the spectra are also influenced by different illumination angle, season and acquisition date as well as moisture conditions. For each city we selected the urban areas:

- Downtown;
- Airport;
- High Density Urban;
- Industrial/Commercial.

These areas, selected by careful visual inspections, are statistically meaningful and representative of the above different urban environments.

We start from Manhattan (NY), presented in Fig. 5.5.

In general, most of the spectral energy distribution is symmetrical and the upper part of the diagram contains the information. As expected from looking

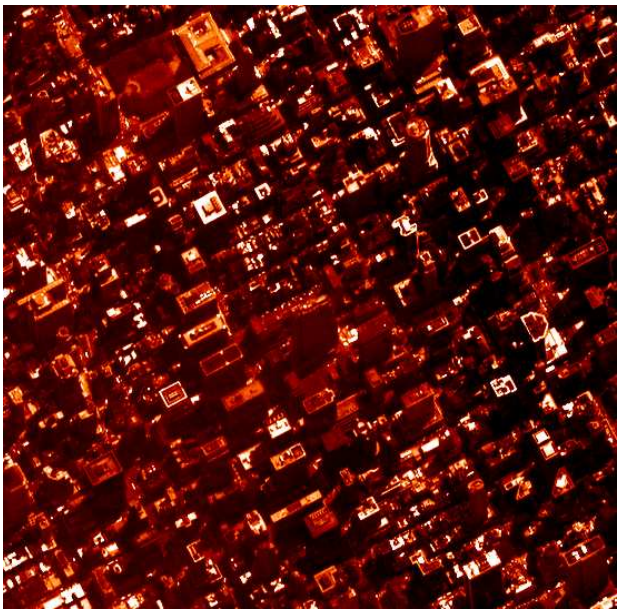


Figure 5.5 - *A selected typical downtown area of Manhattan, NY. The image shows a uniform distribution of the buildings.*

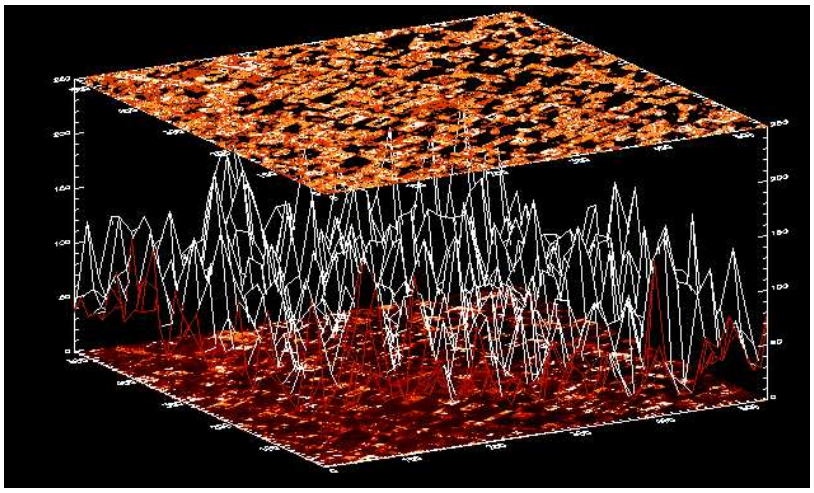


Figure 5.6 - *3D image representation of Fig. 5.5: bottom layer: original image; middle layer: pixel's reflectance value, top layer: objects image contours.*

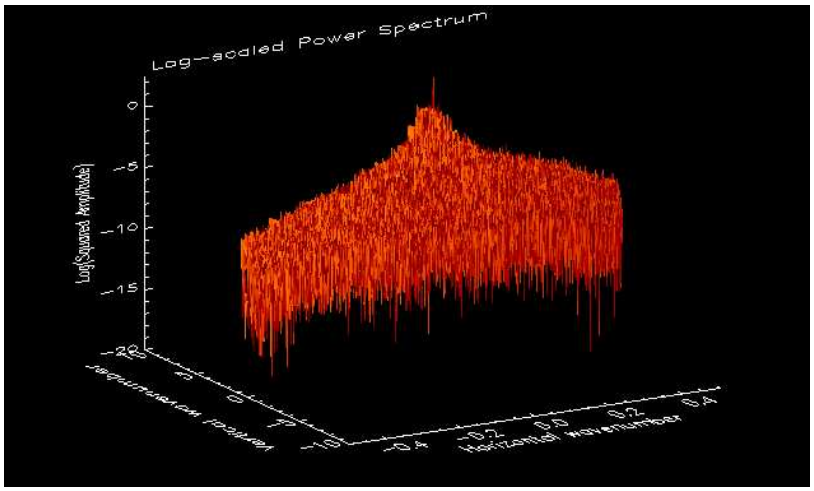


Figure 5.7 - *FFT Power Spectrum of a selected typical downtown area of Manhattan (NY): Logaritmic Scale (surface). The energy is concentrated in a narrow central peak, decreasing smoothly in both directions East-South-East and West-North-West.*

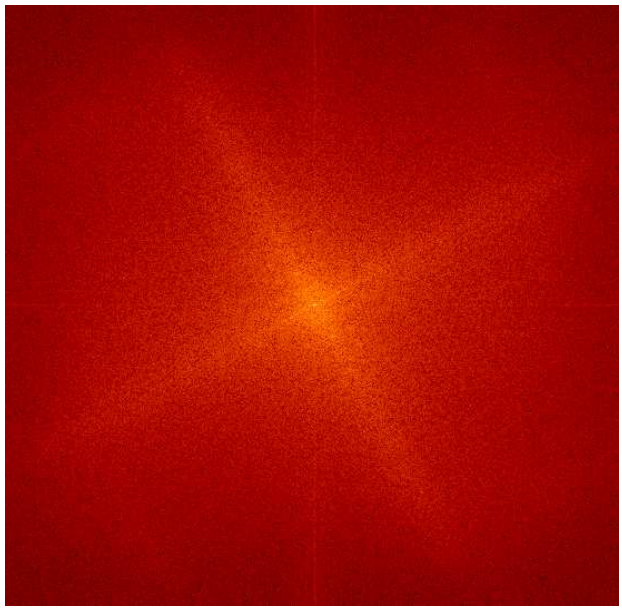


Figure 5.8 - *FFT Power Spectrum of a selected typical downtown area of Manhattan (NY): Logarithmic Scale (image); the energy is uniformly distributed in two perpendicular directions.(The x and y scales are the same as in Fig. 5.7).*

at Fig. 5.5 which does'nt show large differences between darker and brighter areas, hence the spectral energy is not high in amplitude and is concentrated in a narrow central peak. This peak decreases smoothly without any preferred bias direction, along East-South-East and West-North-West directions. The two perpendicular spectral ridges also map the building and road orientations and their almost regular pattern repetition: narrow perpendicular roads (some of them affected by the "city canyon" problem) and similarly oriented buildings, that are not very different in shape, size and brightness. In particular, the low spectral amplitude and its rather narrow central spread characterize, relatively dim and narrow features. The results indicate that the city has been built up in a relatively short time and in a planned fashion, using perpendicular streets. In the lower part of Fig. 5.7, the noise is not very high indicated by the low speckle effect in Fig. 5.8.

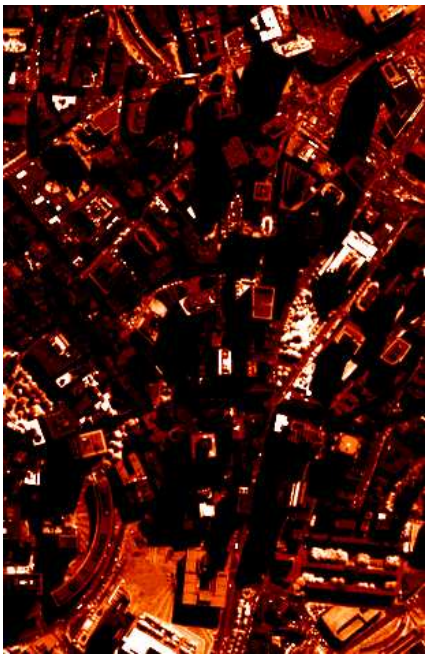


Figure 5.9 - *A selected typical area of Boston downtown, Massachusetts. The way this city built up is related to its geography.*

The Boston downtown area, where the skyscrapers are concentrated, is close

to the ocean: the way that the city built up is related to its geography.

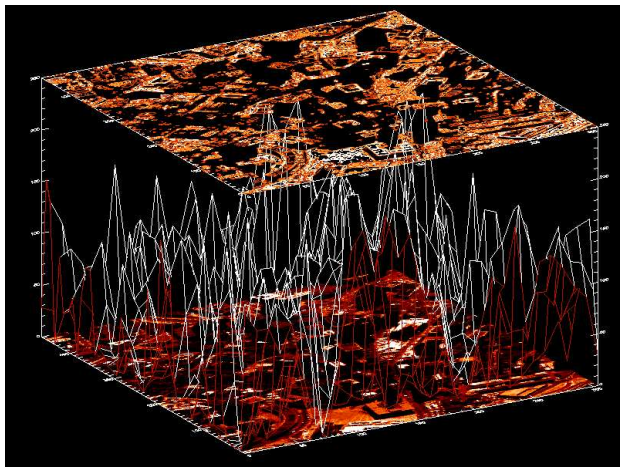


Figure 5.10 - *3D image representation of Fig. 5.9: bottom layer: original image; middle layer: pixel's reflectance value, top layer: objects image contours. The second layer indicate the presence of both very bright and very darker features.*

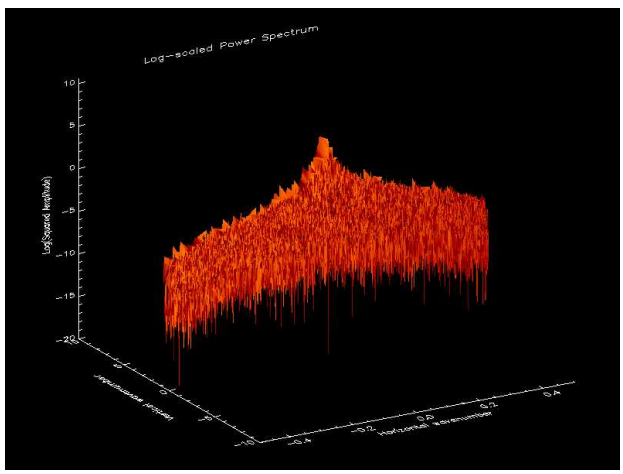


Figure 5.11 - *FFT Power Spectrum of a typical selected area of downtown Boston: Logaritmic Scale (surface). The spectrum amplitude is higher, compared to the Manhattan spectrum, and the energy is spread wider, that means larger and brighter features dominate the image.*

In Fig. 5.11 there is a correspondence between the maximum peak and maximum noise. The lower part of the diagram shows a high noise value spike below the main peak consistent with the high speckle level present in 5.12.In

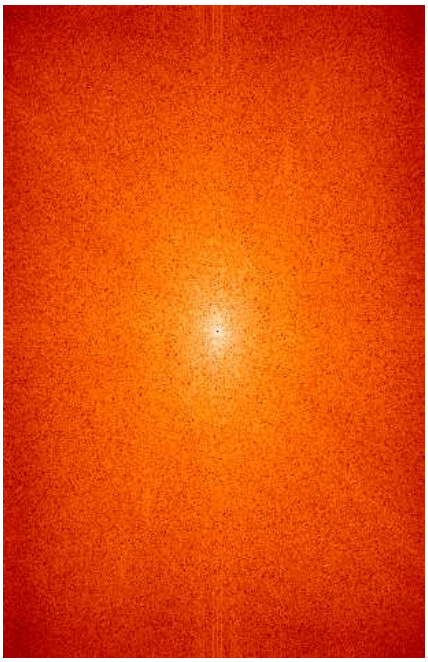


Figure 5.12 - *FFT Power Spectrum of a typical selected area of downtown Boston: Logarithmic Scale (image). The dominant feature is the North-South oriented spectral peak and there are also two weaker peaks in other angle directions.(The x and y scales are the same as in Fig. 5.11).*

Fig. 5.12 the dominant feature is the North-South oriented spectral ridge and there are also two weaker ridges in other directions. The spectral amplitude is higher, compared to the Manhattan spectrum, and the central energy peak is spread wider, that means larger and brighter features dominate the image consistent with Fig. 5.9.

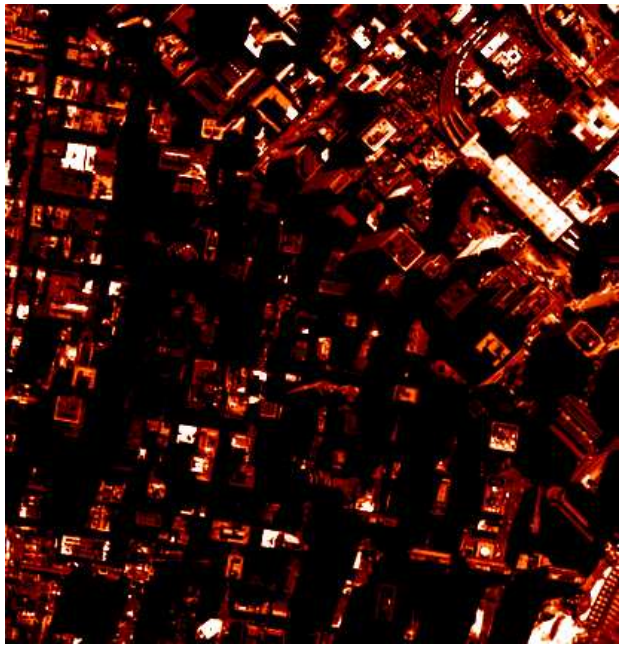


Figure 5.13 - *A selected typical area of downtown San Francisco, California. The area built up somewhat randomly over the time, growing without any intended pattern.*

The city of San Francisco has been built up over time without a specific plan as shown in Fig. 5.13. The buildings and the roads do'nt follow any obvious pattern: the roads are different in width and direction e.g. there are no-long straight streets, the buildings are different in brightness, shape and size and there are not many linear features.

The spectral energy amplitude (5.15) is similar to that for Boston but the central peak is lower than that for Manhattan. These differences may be related to the different illumination, season and acquisition date as well as moisture conditions. The energy ridge bias, in the Wester direction, decreases with in-

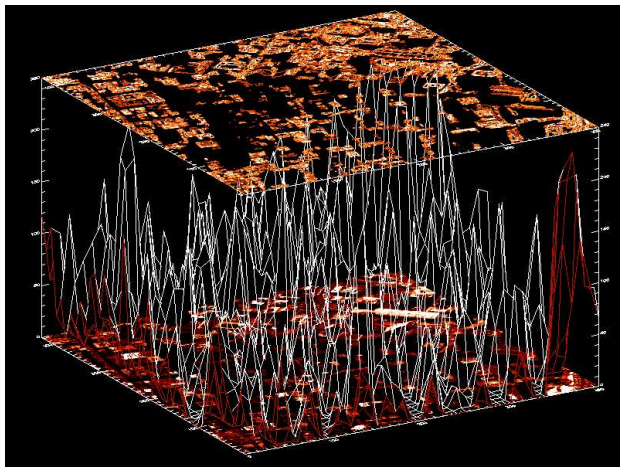


Figure 5.14 - *3D image representation of Fig. 5.13: bottom layer: original image; middle layer: pixel's reflectance value, top layer: objects image contours.*

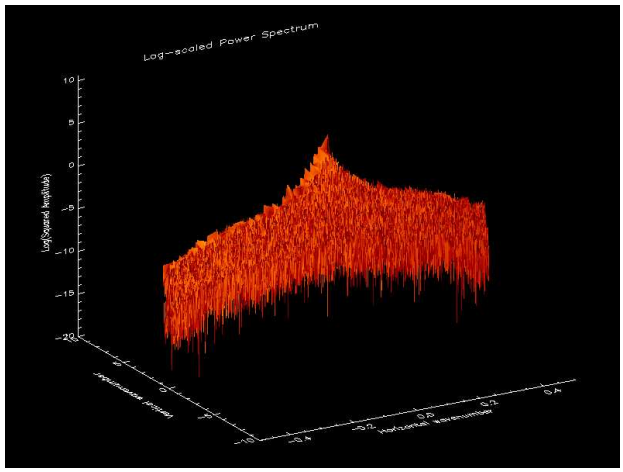


Figure 5.15 - *FFT Power Spectrum of a typical selected area of downtown San Francisco: Logaritmic Scale (surface).*

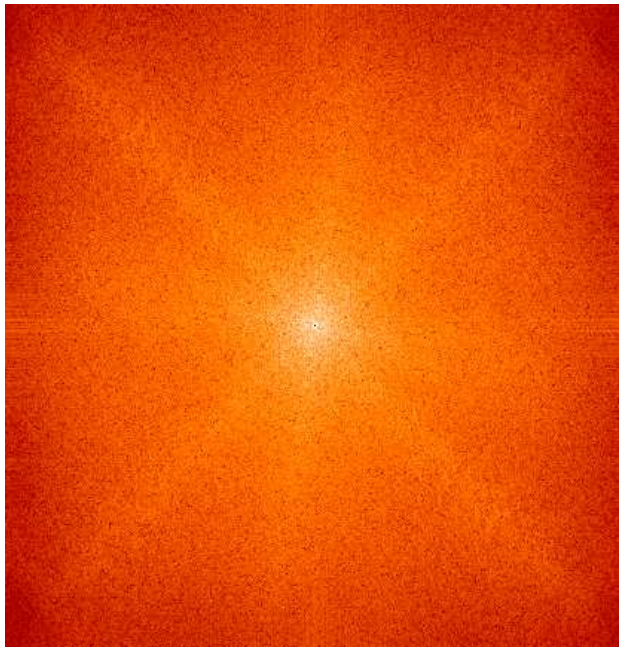


Figure 5.16 - *FFT Power Spectrum of a typical selected area of downtown San Francisco: Logarithmic Scale (image). The wavenumber spectrum is more isotropic. (The x and y scales are the same as in Fig. 5.15).*

creasing wavenumber with small peaks and valleys. The wavenumber spectrum in Fig. 5.16 is more isotropic (i.e. spectral ridges in all directions); although, there is a spectral energy bias towards the West. There are not many linear and aligned features and there are many shadow effects.

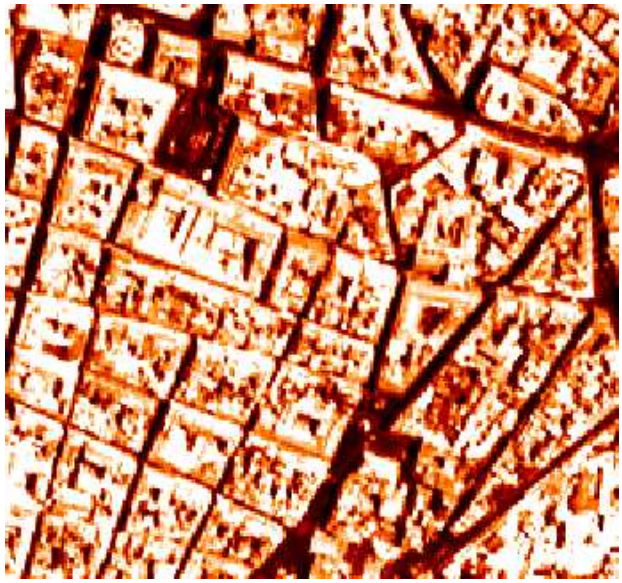


Figure 5.17 - *A typical selected area of downtown Rome, Via del Corso and surroundings, Rome, Italy. The area includes perpendicular streets.*

The Fig. 5.17 represents a downtown area of Rome, including one of the principal central roads, Via Del Corso, and surroundings.

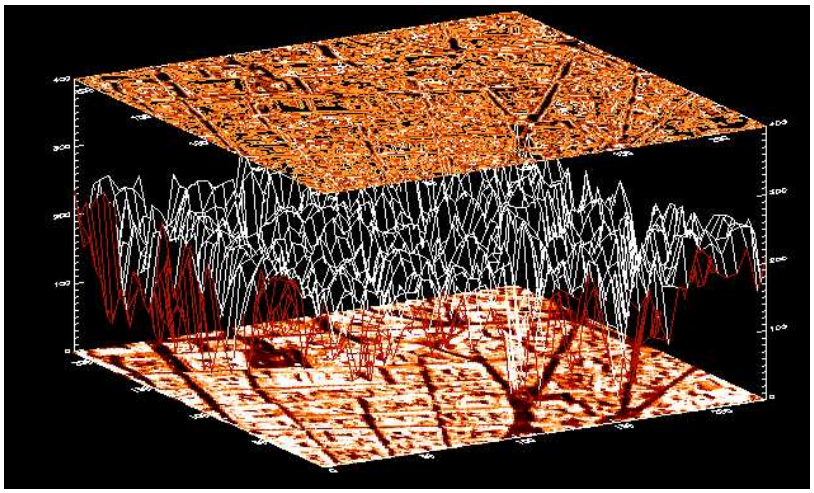


Figure 5.18 - *3D image representation of Fig. 5.17: bottom layer: original image; middle layer: pixel's reflectance value, top layer: objects image contours.*

By comparing the wavenumber spectrum in Fig. 5.20 with Fig. 5.8 and

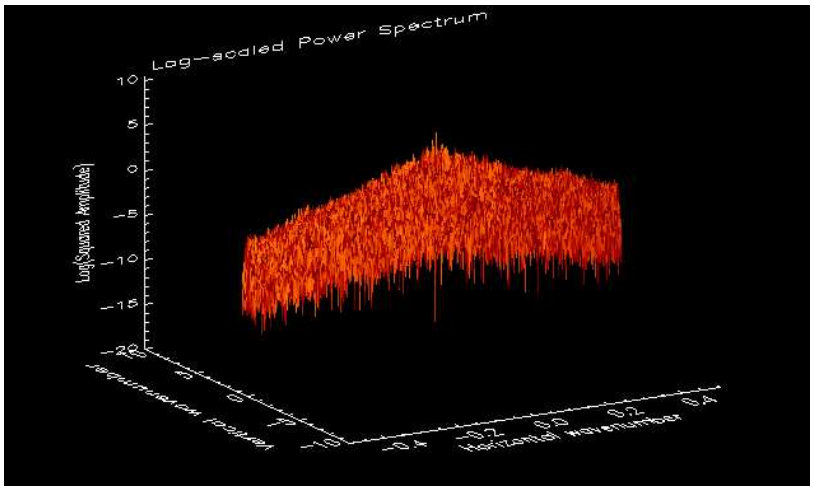


Figure 5.19 - *FFT Power Spectrum of a typical selected area of downtown Rome: Logarithmic Scale (surface). The spectrum has a similar general structure to the Manhattan one, with different specifics.*

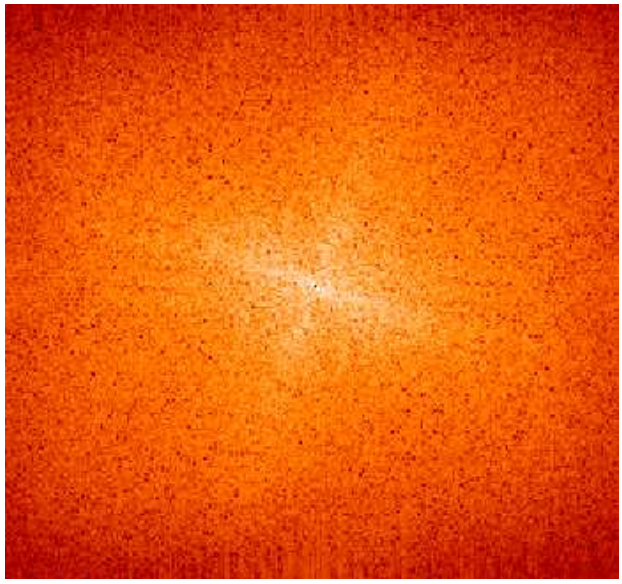


Figure 5.20 - *FFT Power Spectrum of a typical selected area of downtown Rome: Logarithmic Scale (image). (The x and y scales are the same as in Fig. 5.19).*

the respective input images (Figs. 5.17 and 5.5), we can see the same kind of perpendicular orientation of the spectral ridges but with different angle orientations. In particular (Fig. 5.20) the North-East oriented axis has more energy than the South-West one. Moreover, the spectral energy is higher and wider spread than in the Manhattan spectrum. Looking at the Rome streets we can see that they are fewer but larger than the Manhattan ones. These characteristics are clearly indicated in the spectra of the two cities, that have a similar general structure with different specifics. It is important to note that Rome was designed to have perpendicular streets much as Manhattan.

We performed our airport characterizations for La Guardia (NYC), Logan (Boston International Airport), San Francisco International Airport and the smaller Ciampino Airport of Rome.

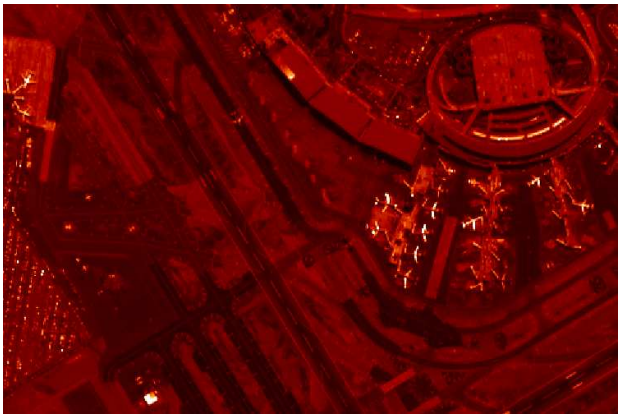


Figure 5.21 - *A typical selected are of La Guardia airport, NY. The airport is close to the city with one direction runways.*

In the lower part of the wavenumber spectrum in Fig. 5.23, the noise level is high, probably due to the overall reduced brightness values; the central peak is broad, and the energy ridge decreases with several minor peaks and valleys. These characteristics seem related to the runways, to the airplanes, to the tarmac and to the roads along the runways. We again note that the noise level is very high at the central peak indicated by the negative spectral spike at

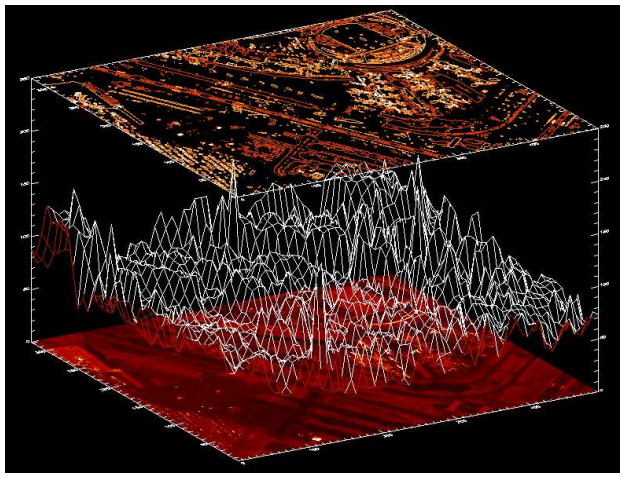


Figure 5.22 - *3D image representation of Fig. 5.21: bottom layer: original image; middle layer: pixel's reflectance value, top layer: objects image contours.*

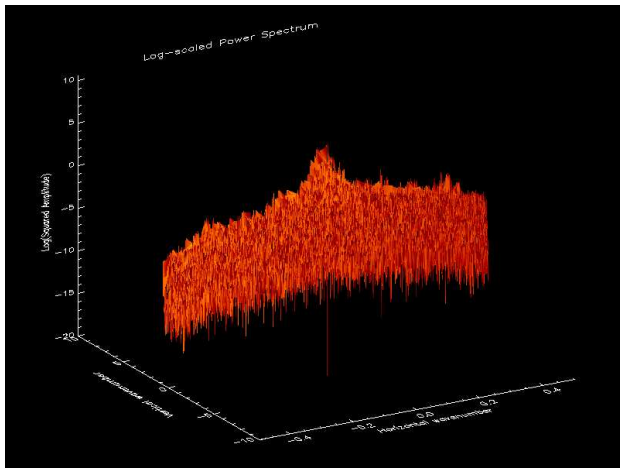


Figure 5.23 - *FFT Power Spectrum of a typical selected area of La Guardia airport: Logaritmic Scale (surface).*

this location. The spectrum (Fig. 5.24) has two main ridges: the stronger is oriented along the West-South-West axis.

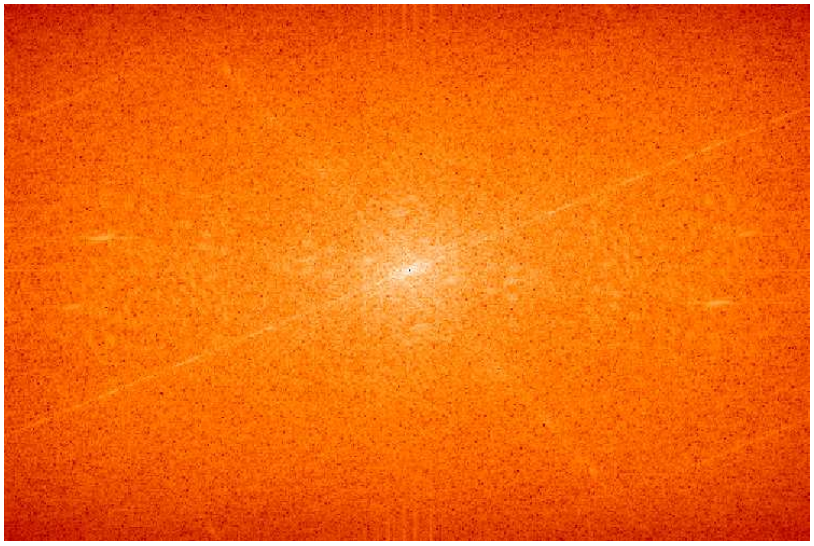


Figure 5.24 - *FFT Power Spectrum of a typical selected area of La Guardia airport: Logarithmic Scale (image). (The x and y scales are the same as in Fig. 5.23).*

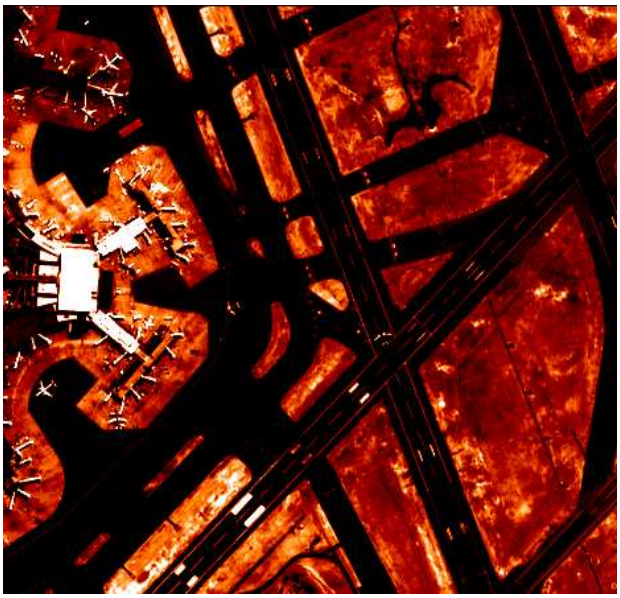


Figure 5.25 - *A typical selected area of Logan International airport, Boston, Massachusetts (MA). The runways are along more directions.*

The Logan International Airport is characterized by runways along more directions as showed in Fig. 5.25.

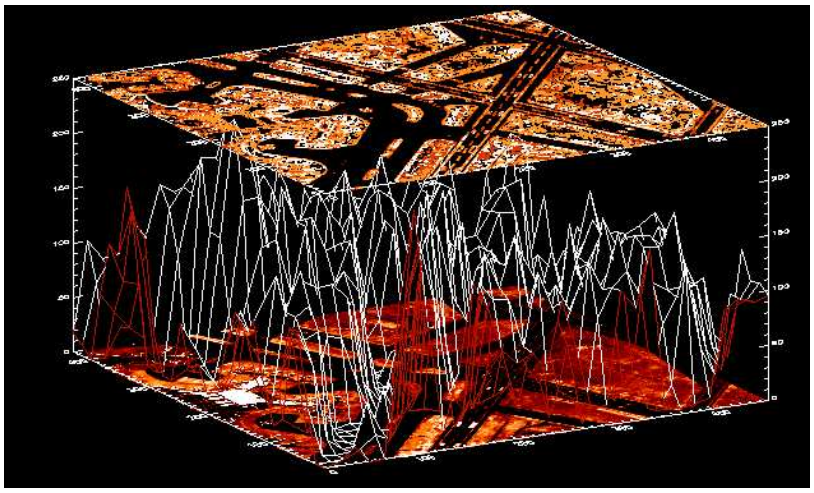


Figure 5.26 - *3D image representation of Fig. 5.25: bottom layer: original image; middle layer: pixel's reflectance value, top layer: objects image contours. The middle layer diagram is influenced by the bright buildings with darker features around.*

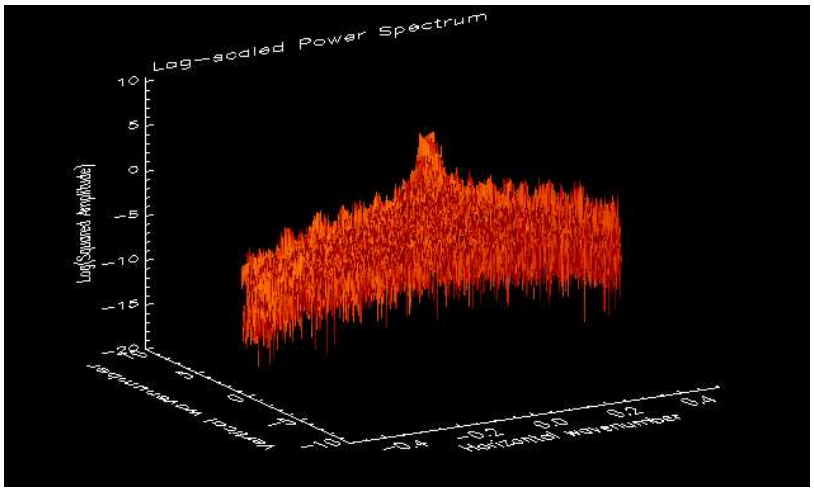


Figure 5.27 - *FFT Power Spectrum of a typical selected area of Logan airport: Logaritmic Scale (surface). The spectrum falls off rapidly starting from a narrow central peak.*

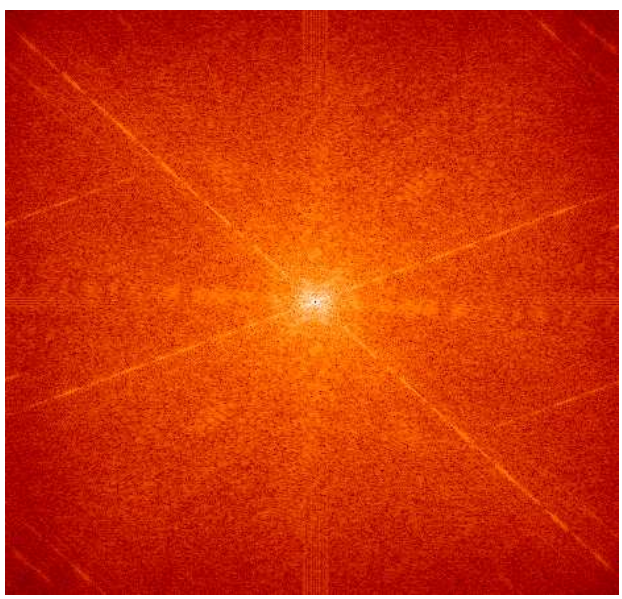


Figure 5.28 - *FFT Power Spectrum of a typical selected area of Logan airport: Logarithmic Scale (image). (The x and y scales are the same as in Fig. 5.27).*

As seen in Fig. 5.25, there is a high contrast between brightest and darkest areas. The bright buildings and the darker features surrounding them influence this spectrum that falls off rapidly starting from a narrow central peak (Fig. 5.27). The energy in the plan-view spectrum (5.28) is the same in both axes oriented North-West and South-East. There are also spectral features along both the North-South and East-West directions.

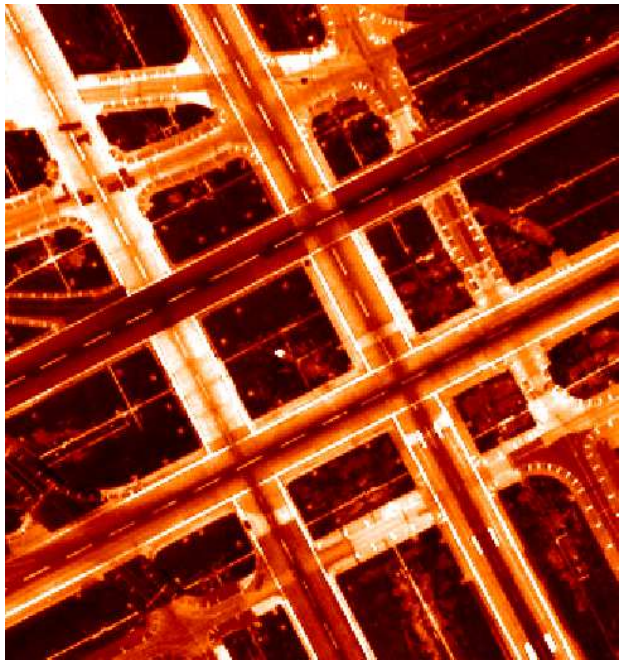


Figure 5.29 - *A typical selected area of San Francisco International airport, San Francisco, California.*

In the lower part of the San Francisco airport spectrum in Fig. 5.31, the noise level is higher than in the Logan spectrum and is similar to La Guardia spectrum. The plan-view spectrum (Fig. 5.32) show several perpendicular main features. The spectral energy is higher along the West-North and East-South directions than the North-South East-West ones. Compared to the Logan spectrum, the energy is higher, and the angles between the energy axes are completely different.

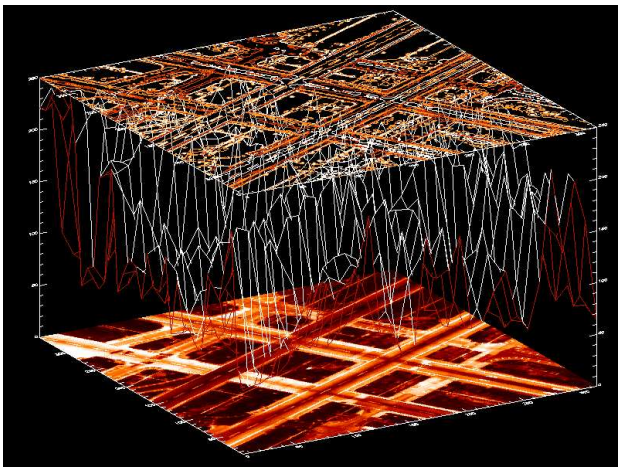


Figure 5.30 - *3D image representation of Fig. 5.29: bottom layer: original image; middle layer: pixel's reflectance value, top layer: objects image contours.*

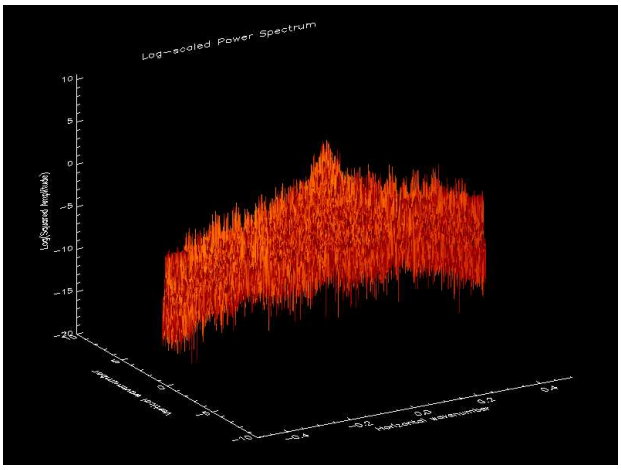


Figure 5.31 - *FFT Power Spectrum of a typical selected area of San Francisco airport: Logarithmic Scale (surface).*

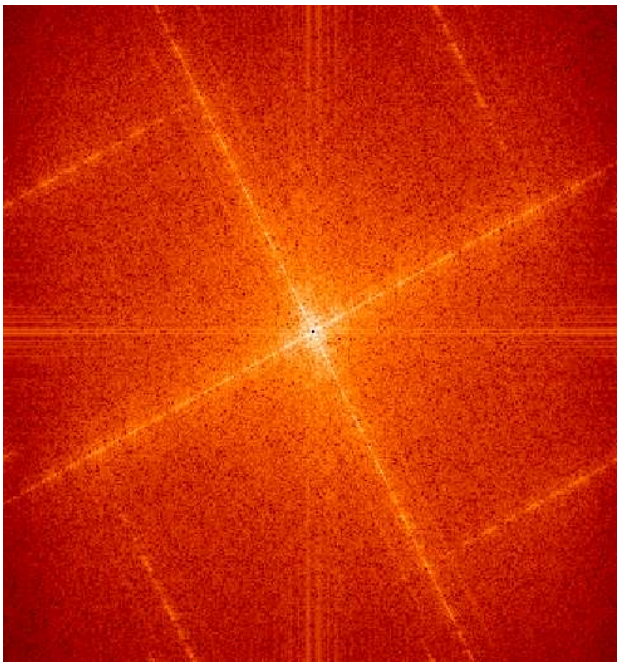


Figure 5.32 - *FFT Power Spectrum of a typical selected area of San Francisco airport: Logarithmic Scale (image). (The x and y scales are the same as in Fig. 5.31).*

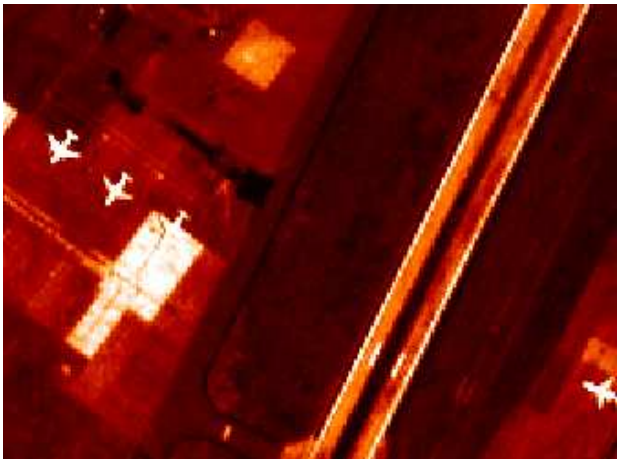


Figure 5.33 - *A typical selected area of Ciampino airport, Rome, Italy. The airport is small and has only one runway.*

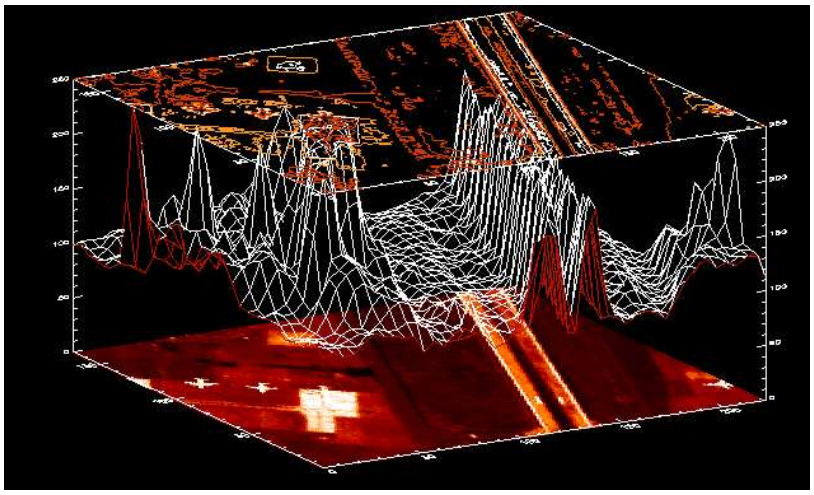


Figure 5.34 - *3D image representation of Fig. 5.33: bottom layer: original image; middle layer: pixel's reflectance value, top layer: objects image contours.*

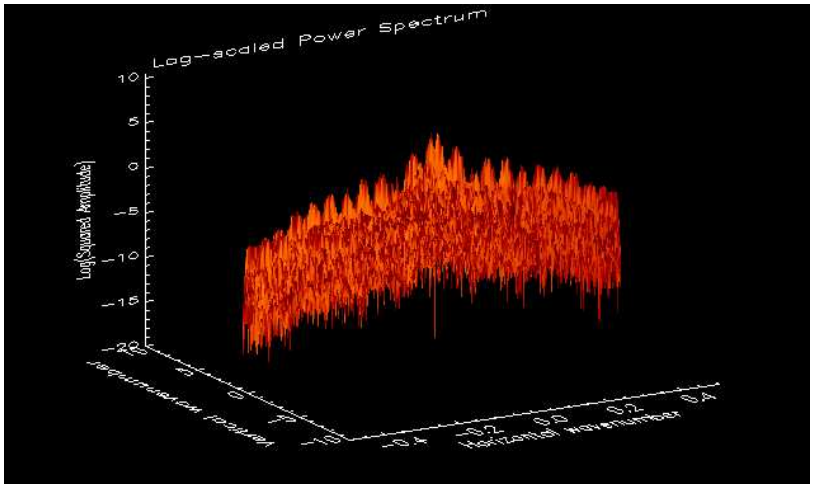


Figure 5.35 - *FFT Power Spectrum of a typical selected area of Ciampino airport: Logarithmic Scale (surface).*

By comparing the Ciampino spectra in both Fig. 5.36 and Fig. 5.35 we can say that there are many smaller peaks along the main spectral ridges. The central peak is significantly different from its neighbors, but the smaller peaks are not. This is due to the fact that Ciampino Airport has only one runway. Basically, you have just one feature to distinguish. This is a an example of a limitation of the use of the wave number analysis. When the image features are large and relatively uniform in structure, the noise is often amplified, which makes this type of analysis less useful.

We will now use the wavenumber spectral technique to characterize High Density Urban Areas starting from Brooklyn (NY).

The diagram in Fig. 5.39 show a higher spectral amplitude compared with 5.7. There are several central peaks, instead of only one as in 5.7, which are similar in energy and can be treated as one peak. By comparing Figures 5.37 and 5.40: the input image is similar to Manhattan but with smaller buildings and weaker "canyon city" effects; moreover the wavenumber spectrum is brighter, with more spectral energy. The energy is along two axes which are wider and larger in the North-East direction.

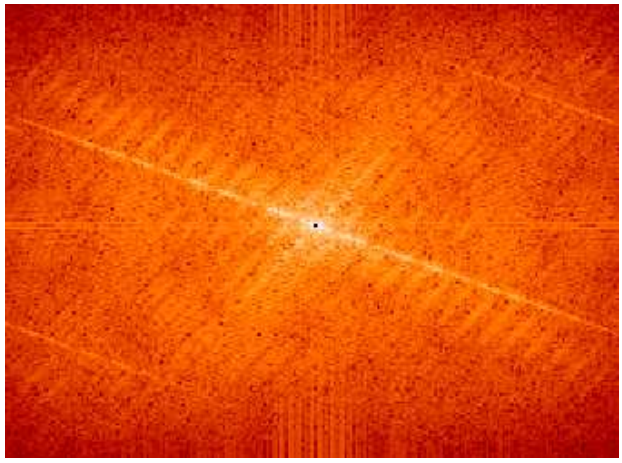


Figure 5.36 - *FFT Power Spectrum of a typical selected area of Ciampino airport: Logarithmic Scale (image). (The x and y scales are the same as in Fig. 5.35).*

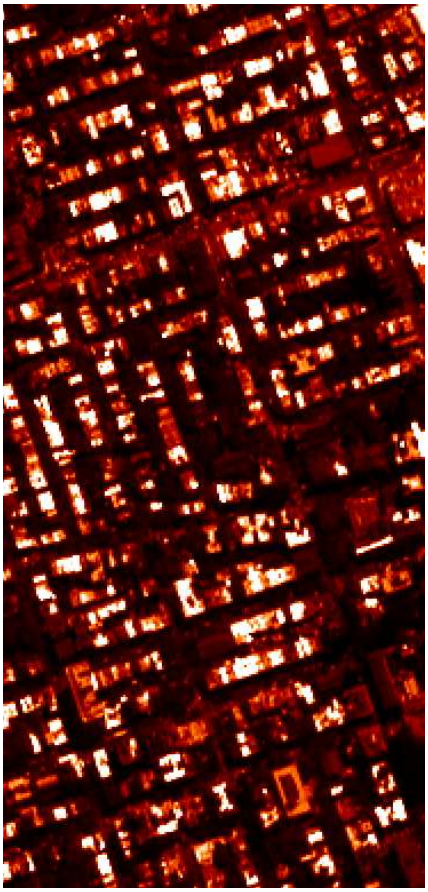


Figure 5.37 - *A typical selected high density urban area of Brooklyn, New York.*

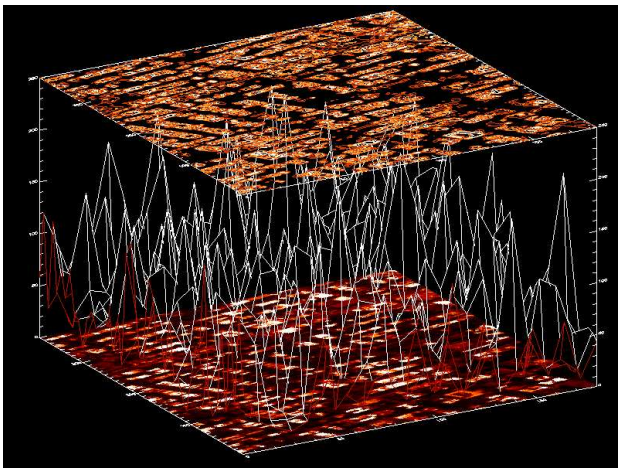


Figure 5.38 - *3D image representation of Fig 5.37: bottom layer: original image; middle layer: pixel's reflectance value, top layer: objects image contours.*

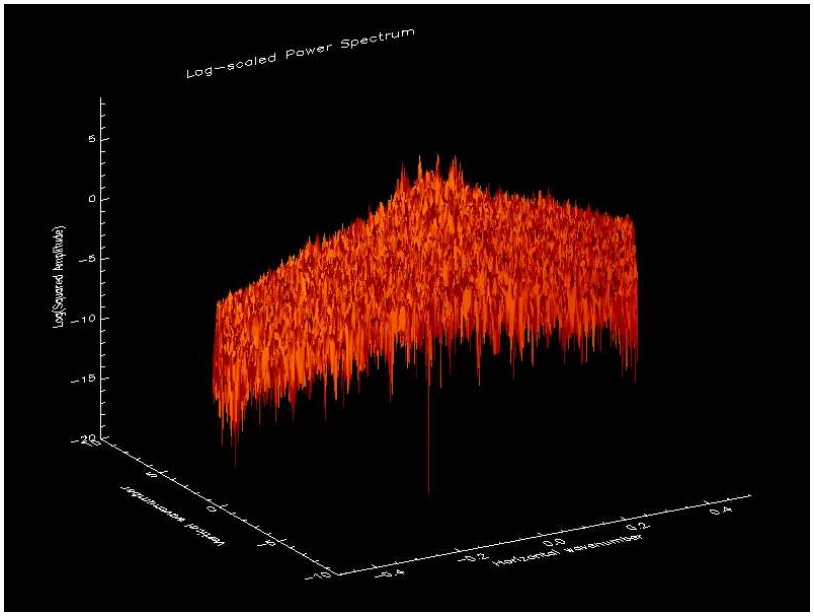


Figure 5.39 - *FFT Power Spectrum of a typical selected high density urban area of Brooklyn (NY): Logaritmic Scale (surface). The several central peaks are not significantly different and can be treated as one principal peak.*

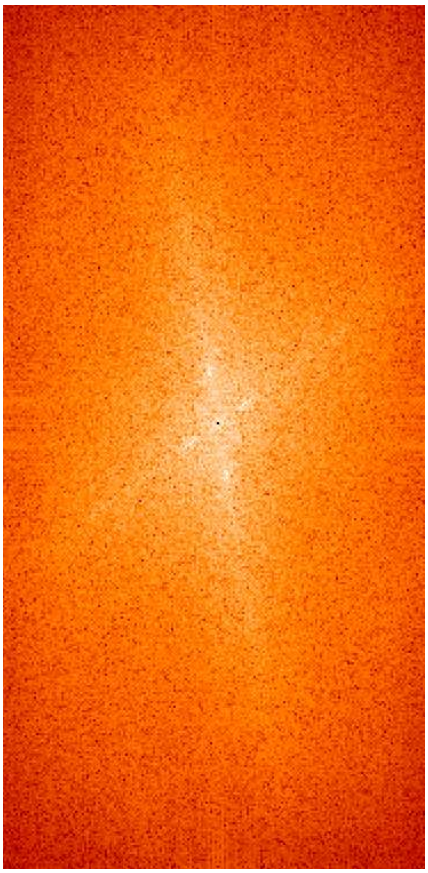


Figure 5.40 - *FFT Power Spectrum of a typical selected high density urban area of Brooklyn (NY): Logaritmic Scale (image). (The x and y scales are the same as in Fig. 5.39).*

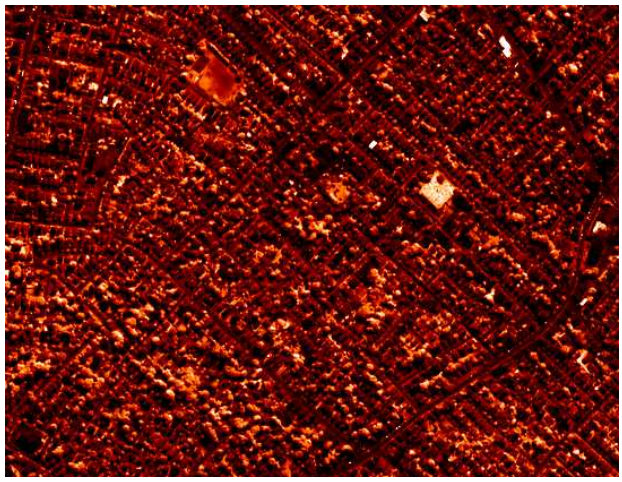


Figure 5.41 - *A typical selected high density urban area of Boston, Massachusetts.*

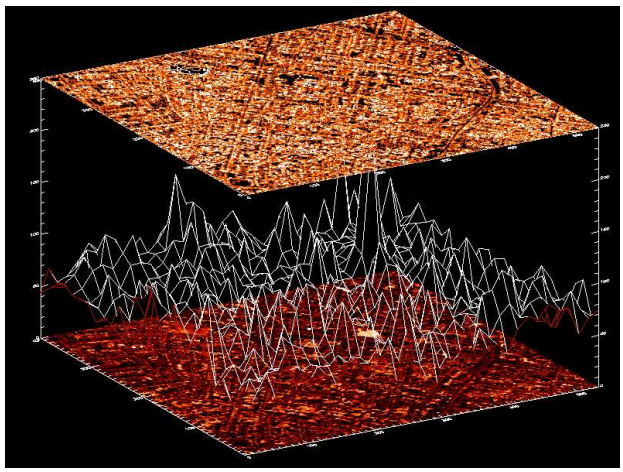


Figure 5.42 - *3D image representation of Fig. 5.41: bottom layer: original image; middle layer: pixel's reflectance value, top layer: objects image contours.*

In Fig. 5.43 the Boston urban area spectral central ridge is narrow, the energy rapidly decreases initially, and subsequently, gradually. Both the spectra in Fig. 5.40 and Fig. 5.44 have two preferred directions. They have similar energy distributions, but with different angles. Moreover, in Fig. 5.44 the East-East-South axis has more energy.

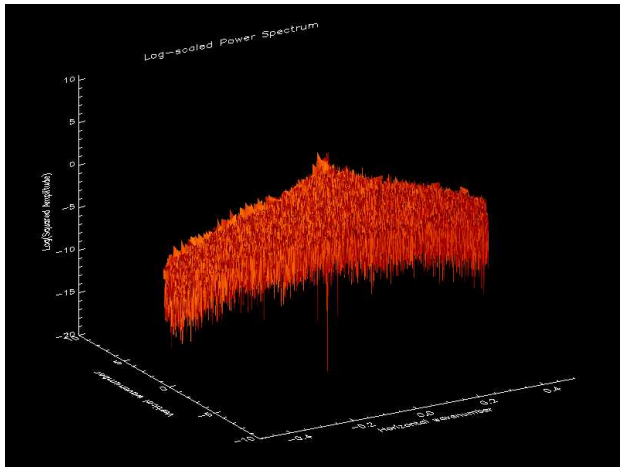


Figure 5.43 - *FFT Power Spectrum of a typical selected high density urban area of Boston: Logaritmic Scale (surface).*

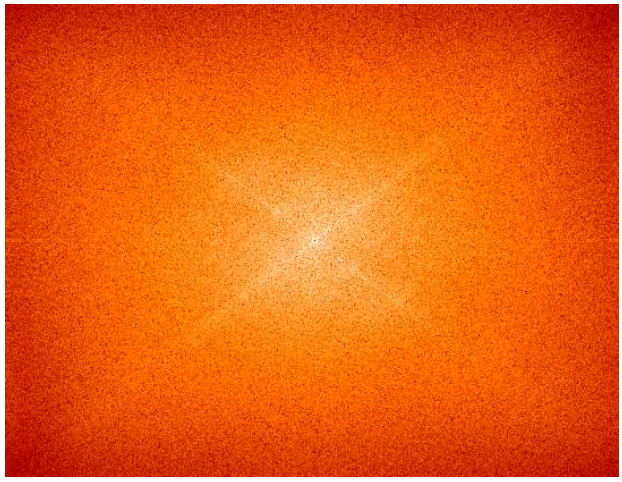


Figure 5.44 - *FFT Power Spectrum of a typical selected high density urban area of Boston: Logarithmic Scale (image). There are two preferred directions in the energy distribution. (The x and y scales are the same as in Fig. 5.43).*

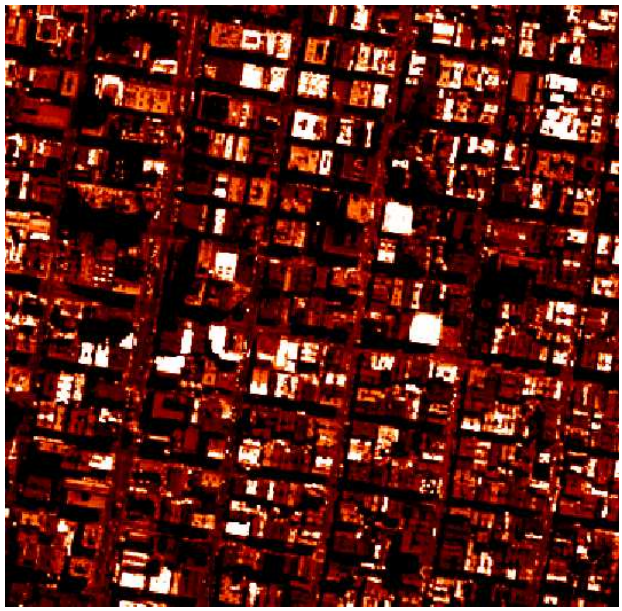


Figure 5.45 - *A typical selected high density urban area of San Francisco, California.*

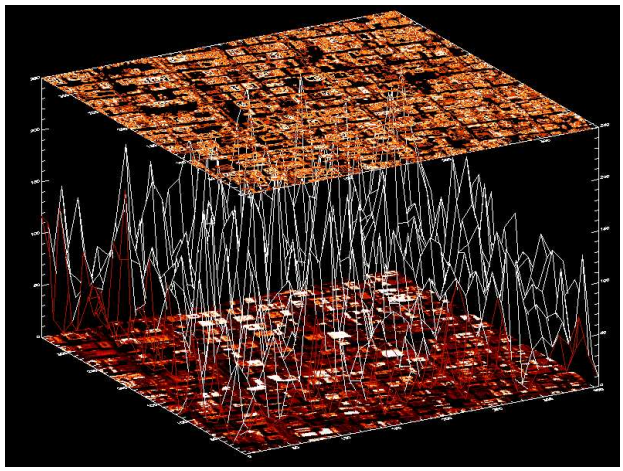


Figure 5.46 - *3D image representation of Fig. 5.45: bottom layer: original image; middle layer: pixel's reflectance value, top layer: objects image contours.*

In Fig. 5.47 the central peak is not large in amplitude, and the spectrum decreases rapidly in the Westword direction, initially steeply and then has another significant peak. The speckle is related to the vertical amplitude of the

diagram, that is pretty high as seen in Fig. 5.48.

Comparing the spectra related to San Francisco downtown area and high

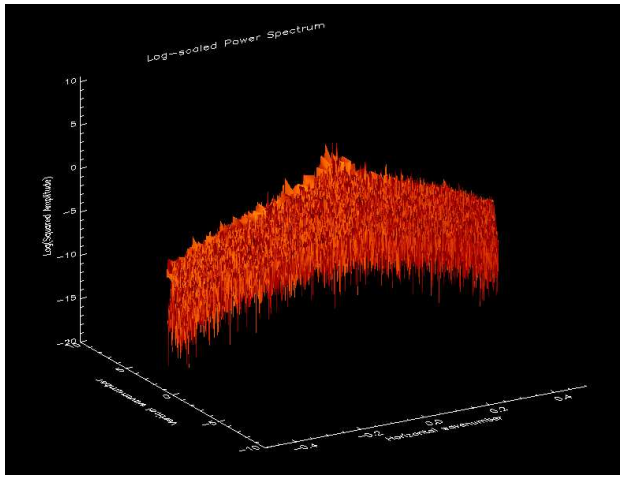


Figure 5.47 - *FFT Power Spectrum of a typical selected high density urban area of San Francisco: Logarithmic Scale (surface).*

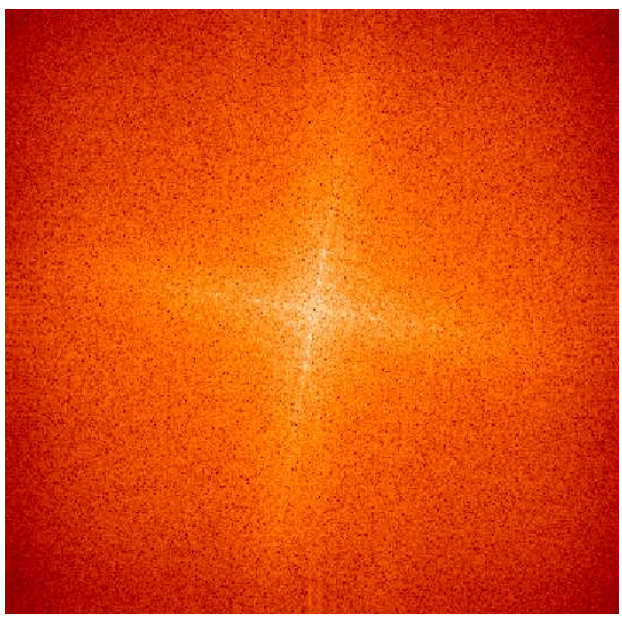


Figure 5.48 - *FFT Power Spectrum of a typical selected high density urban area of San Francisco: Logarithmic Scale (image). The road orientation establish a certain energy direction. (The x and y scales are the same as in Fig. 5.47).*

density urban area, Figs. 5.48 and 5.16, we can see that in the housing area the energy is distributed along two axis while in the downtown area the energy distribution is only along one major axes. In particular Fig. 5.48 shows that the two axes are almost perpendicular and the energy is very similar. These characteristics are related to the road orientations that establish certain directions (Fig. 5.45). Looking again at the two spectra, the high density urban area has a higher speckle effect than the downtown one.

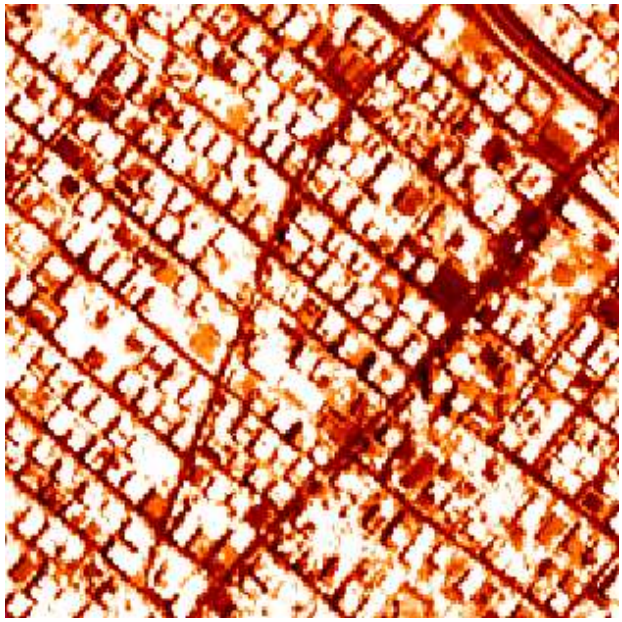


Figure 5.49 - *A typical selected high density urban area of Rome, Italy.*

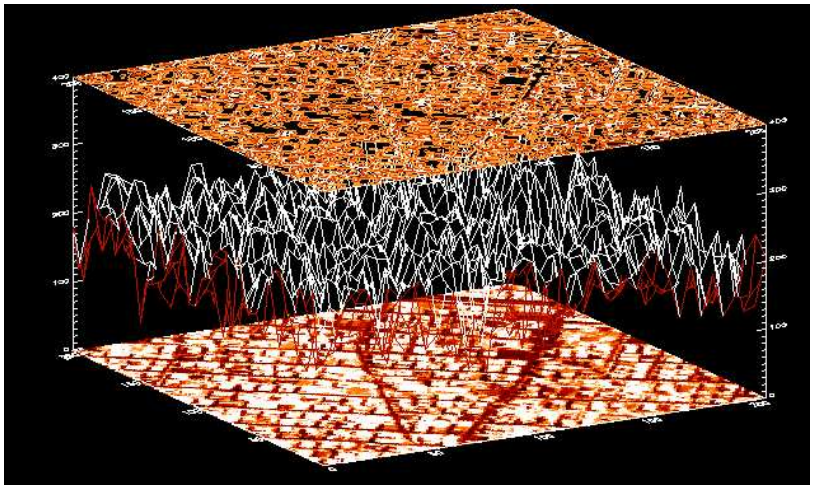


Figure 5.50 - *3D image representation of Fig. 5.49: bottom layer: original image; middle layer: pixel's reflectance value, top layer: objects image contours.*

The previously discussed for the high density urban area spectrum (Fig. 5.52) the energy is along two perpendicular axes one of which is the almost East-West axis. This characteristic is related to the distance between city

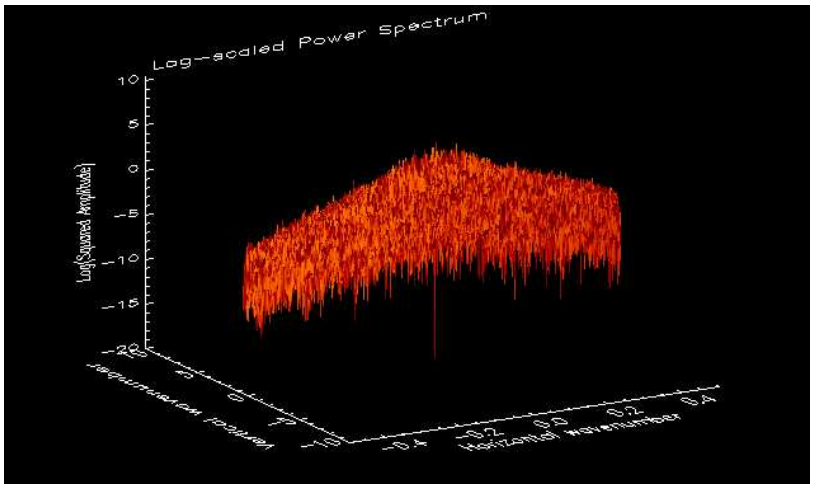


Figure 5.51 - *FFT Power Spectrum of a typical selected high density urban area of Rome: Logarithmic Scale (surface).*

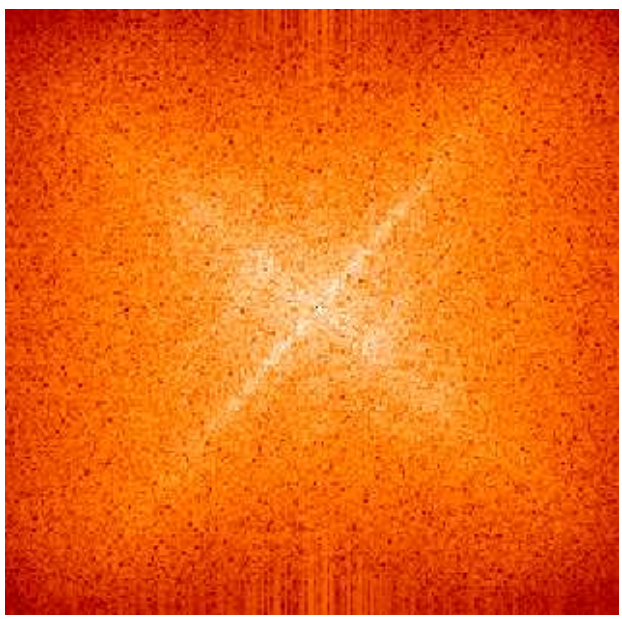


Figure 5.52 - *FFT Power Spectrum of a typical selected high density urban area of Rome: Logarithmic Scale (image). (The x and y scales are the same as in Fig. 5.51).*

blocks (Fig. 5.49). The speckle effect is high as showed in Fig. 5.52 and by the vertical extent of the spectrum in Fig. 5.51.

We continue our study by characterizing Industrial/Commercial areas belonging to our selected cities.

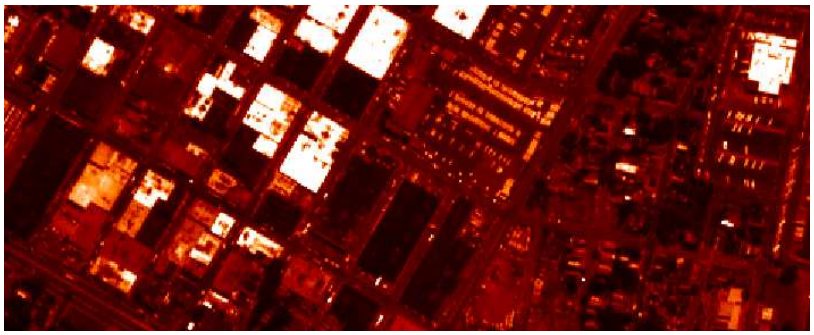


Figure 5.53 - *A typical selected industrial/commercial area of New York City.*

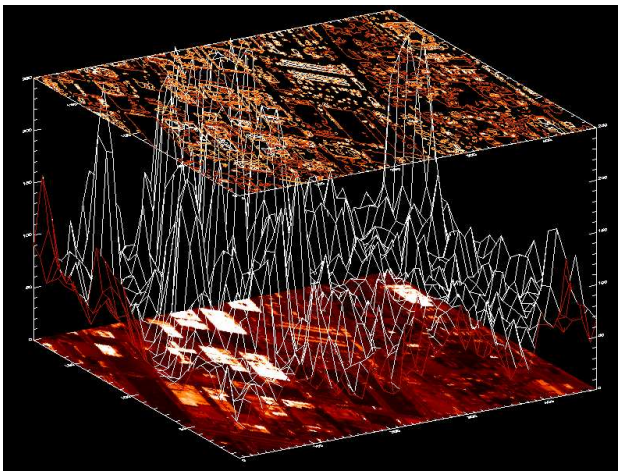


Figure 5.54 - *3D image representation of Fig. 5.53: bottom layer: original image; middle layer: pixel's reflectance value, top layer: objects image contours.*

The spectrum in Fig. 5.56 shows two non-perpendicular major axes. The noise is concentrated along the North-South direction, in Fig. 5.55, the maximum peak corresponds to the maximum noise (seen in a large negative peak beneath the spectral maximum).

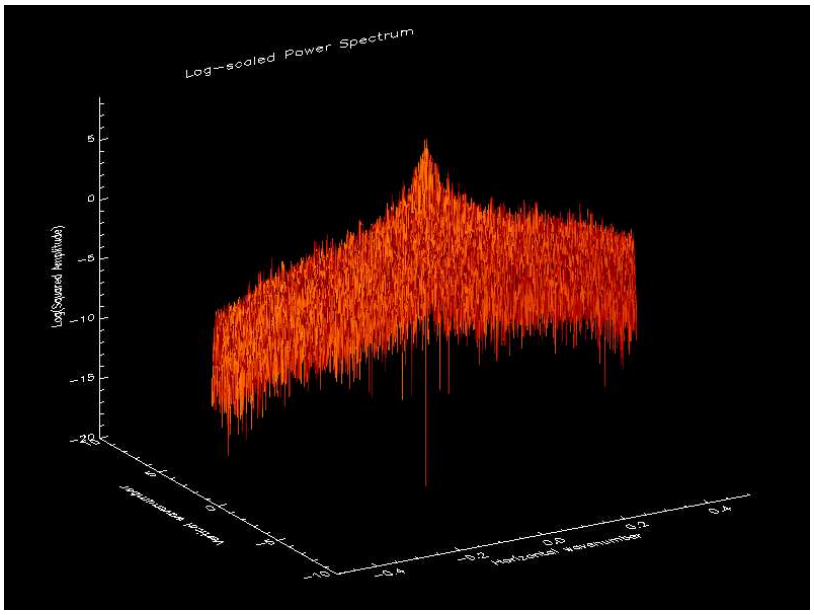


Figure 5.55 - *FFT Power Spectrum of a typical industrial/commercial area of New York city: Logaritmic Scale (surface). The maximum peak corresponds to the maximum noise (seen in a large negative peak beneath the spectral maximum).*

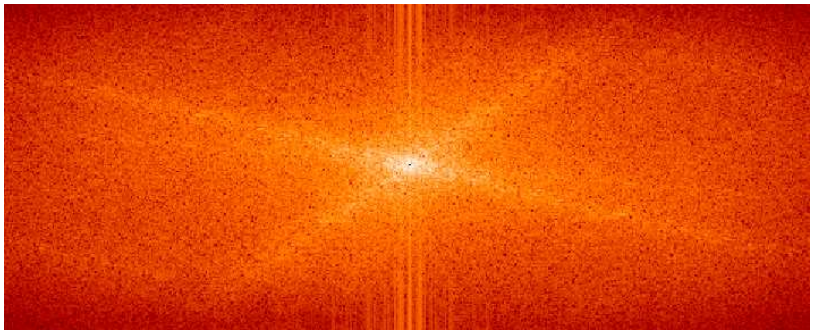


Figure 5.56 - *FFT Power Spectrum of a typical industrial/commercial area of New York city: Logarithmic Scale (image). The spectrum shows two non-perpendicular major axes. (The x and y scales are the same as in Fig. 5.55).*

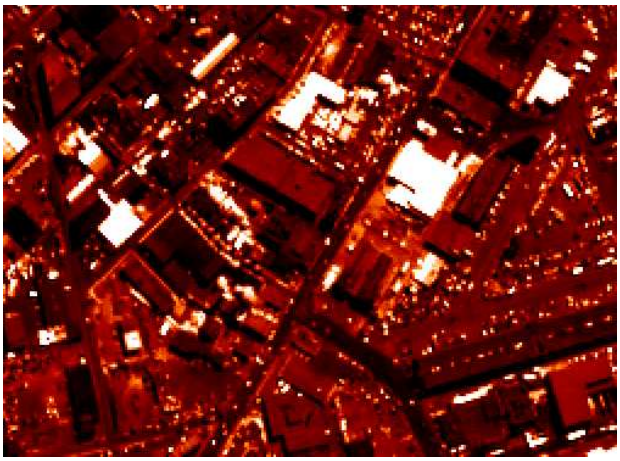


Figure 5.57 - *A typical selected industrial/commercial area of Boston, Massachusetts.*

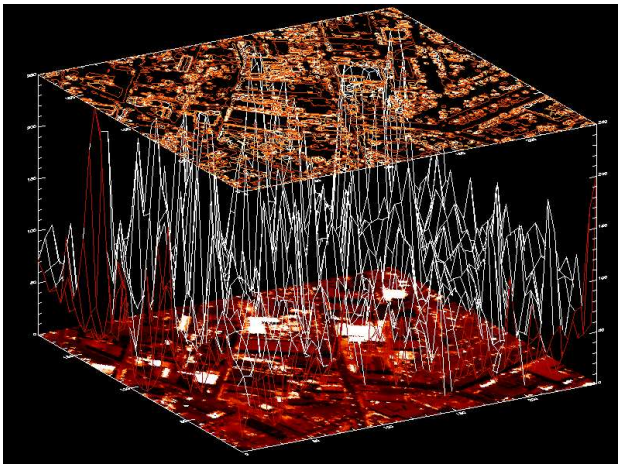


Figure 5.58 - *3D image representation of Fig. 5.57: bottom layer: original image; middle layer: pixel's reflectance value, top layer: objects image contours.*

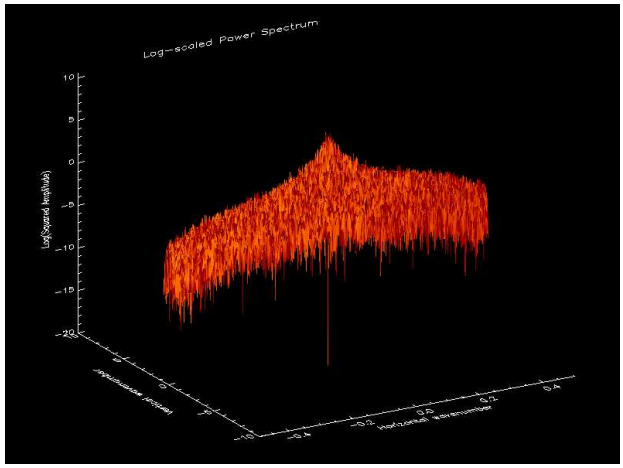


Figure 5.59 - *FFT Power Spectrum of a typical selected industrial/commercial area of Boston: Logaritmic Scale (surface). The vertical size of the diagram shows a high noise level.*

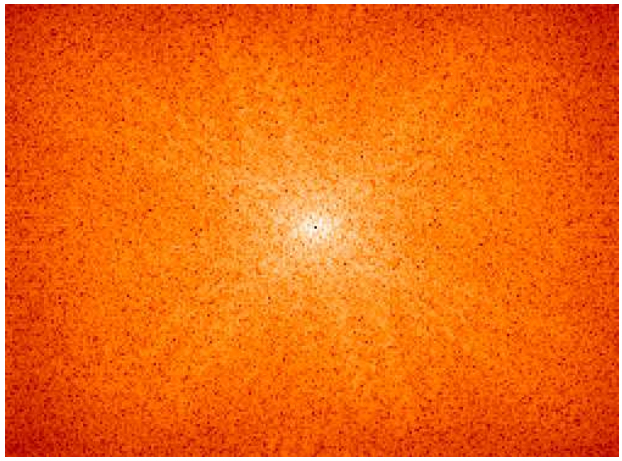


Figure 5.60 - *FFT Power Spectrum of a typical selected industrial/commercial area of Boston: Logarithmic Scale (image). The spectrum is isotropic with a central spherical energy distribution. (The x and y scales are the same as in Fig. 5.59).*

The plan-view spectrum in Fig. 5.60 is nearly isotropic with an elliptical central energy distribution. The spectrum is very noisy with a lot of speckle effect as showed in Figs. 5.60 and 5.59.

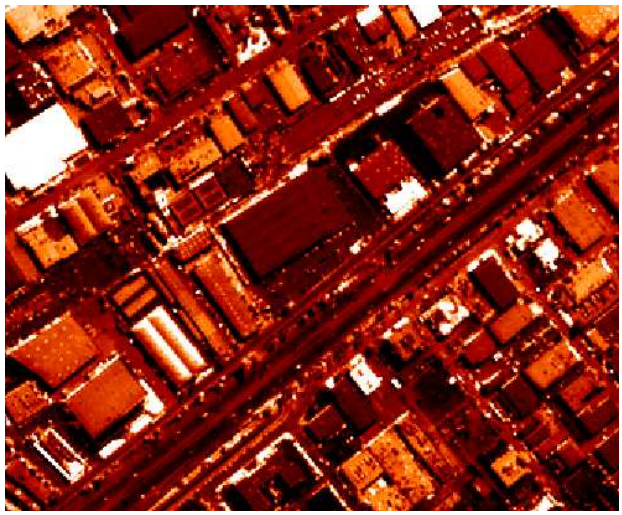


Figure 5.61 - *A typical selected industrial/commercial area of San Francisco, California.*

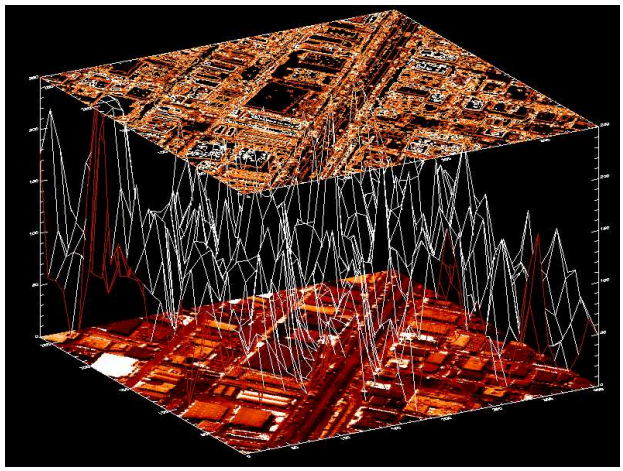


Figure 5.62 - *3D image representation of Fig. 5.61: bottom layer: original image; middle layer: pixel's reflectance value, top layer: objects image contours.*

The San Francisco spectrum in Fig. 5.63 does not show a correspondence between the maximum noise and the maximum spectral peak. In addition, the overall noise in this spectrum is low. The spectrum in Fig. 5.64 has two primary axes that are not quiet perpendicular. Moreover, this spectrum is similar to the spectrum from the downtown area (Fig. 5.48) but higher in energy.

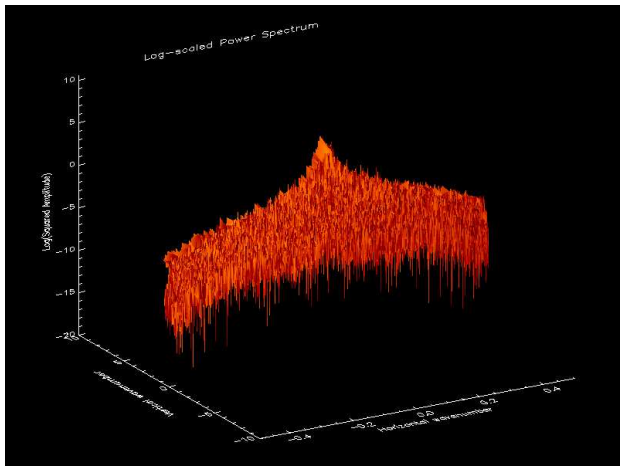


Figure 5.63 - *FFT Power Spectrum of a typical selected industrial/commercial area of San Francisco: Logaritmic Scale (surface). The overall noise in this spectrum is low.*

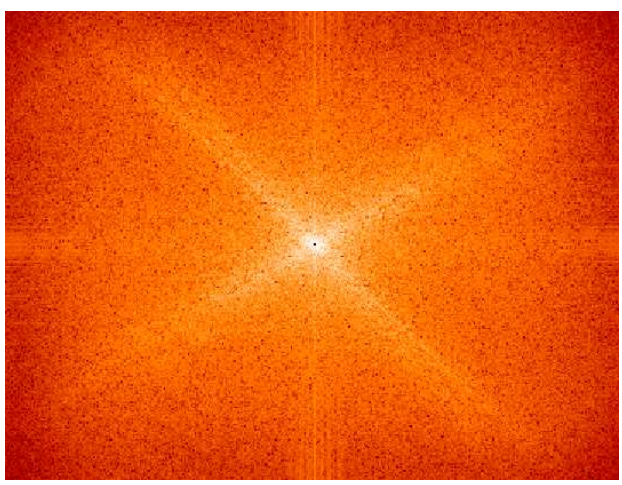


Figure 5.64 - *FFT Power Spectrum of a typical selected industrial/commercial area of San Francisco: Logarithmic Scale (image). The spectrum is similar to the spectrum from the downtown area. (The x and y scales are the same as in Fig. 5.63).*

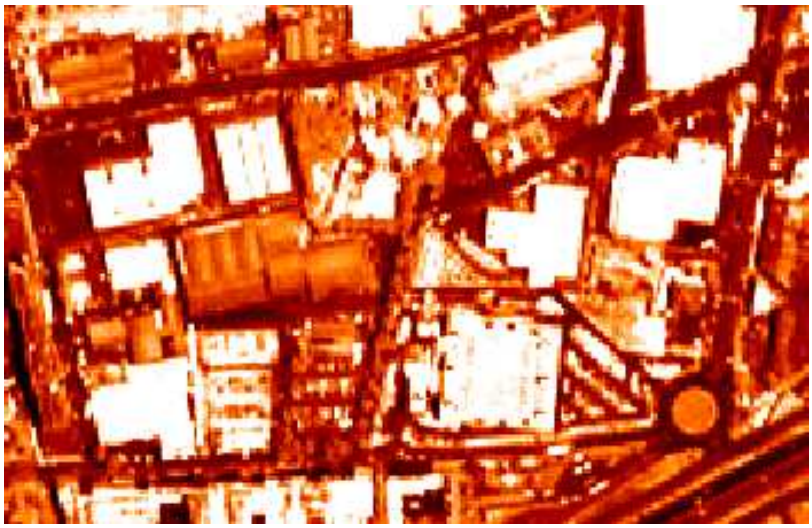


Figure 5.65 - *A typical selected industrial/commercial area of Rome, Italy.*

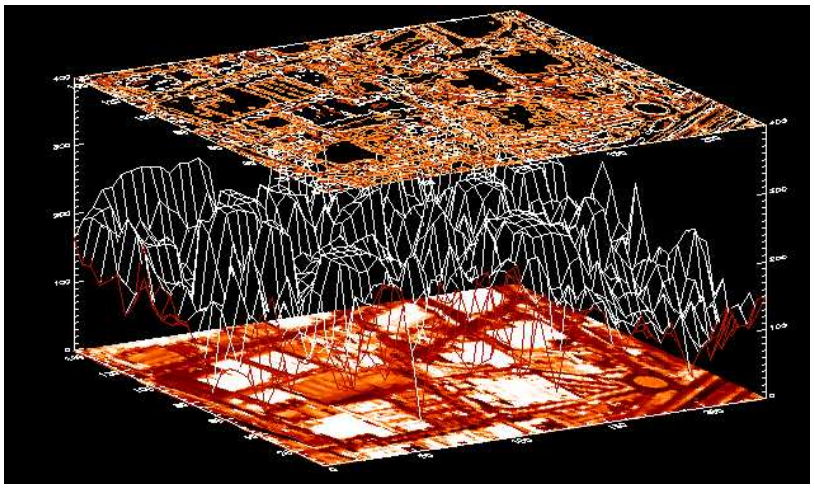


Figure 5.66 - *3D image representation of Fig. 5.65: bottom layer: original image; middle layer: pixel's reflectance value, top layer: objects image contours.*

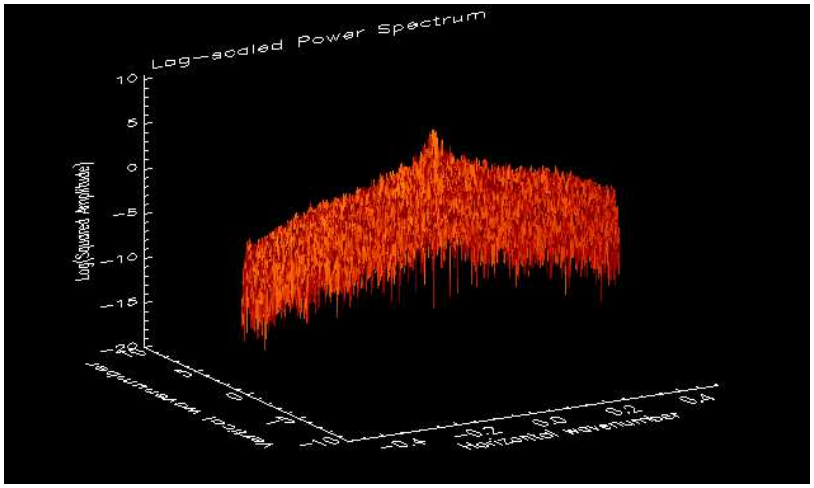


Figure 5.67 - *FFT Power Spectrum of a typical selected industrial/commercial area of Rome: Logarithmic Scale (surface). The diagram has more peaks in the west direction.*

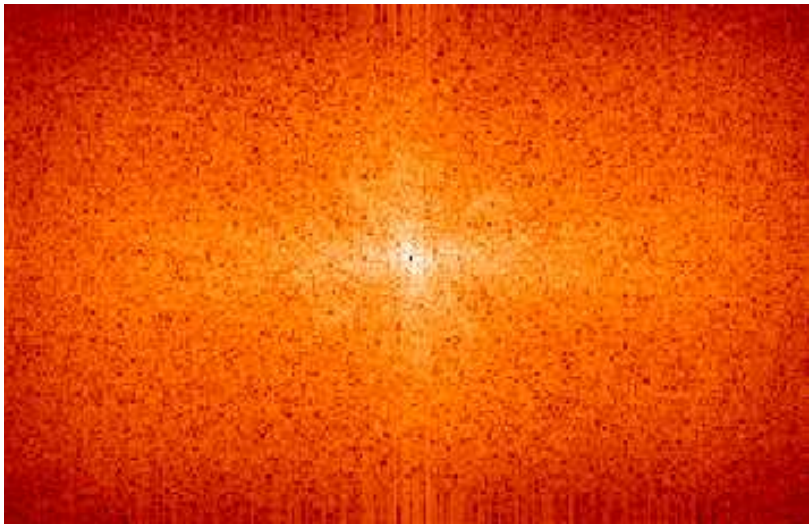


Figure 5.68 - *FFT Power Spectrum of a typical selected industrial/commercial area of Rome: Logarithmic Scale (image). The spectrum shows maximum energy in North-South and East-West directions with a bias to the west. (The x and y scales are the same as in Fig. 5.67).*

The Rome spectrum in Fig. 5.68 shows maximum energy in North-South and East-West directions with a bias to the west. The amplitude of the speckle indicates a high noise level. As expected, the diagram in Fig. 5.67 has a higher spectral level in the west direction.

Chapter 6

Conclusions

This study focuses on high resolution urban monitoring using NN analyses for land cover classification and change detection, and FFT computations of wavenumber spectra to characterize the spatial scales of land cover features. The methods include:

- Classification of a data set of high resolution images with NN algorithms with respect to the most relevant urban features. The purpose of this procedure was to yield accurate classification maps and to train the networks in order to be used as general classifiers for other similar image data sets not considered in the NN training phase. This allows us to classify new images in near-real time. To achieve this goal, careful spectral analyses over statistically significant different data sets have been carried out and the NN topologies have been designed to avoid possible effects of overfitting.

- We have then used different wavenumber spectra to understand the nature of the significant features in the images in different urban environments. In particular, the transform algorithms have been used to characterize different urban areas of different cities (New York, Boston, San Francisco and Rome), selected, by visual inspection of the different, Regions of

Interest.

Neural Networks algorithms

It is well recognized that one of the major advantages of NN with respect to Bayesian and other statistically based classifiers is that NN methods draw their own input-output discriminant relations directly from the data and do not require that a particular form of a PDF (Probability Density Function) be assumed (Dawson (1994)). In this study we exploit these characteristics of MLP (Multi Layer Perceptrons) networks for automatic processing of a data set of high-resolution satellite images with a particular interest in feature extraction from urban areas. The network performance appears to be satisfactory, especially considering the fact that the procedures used are completely automatic. In fact, the maps automatically applied to new images (i.e., those that were not considered in the training phase), show good classification agreement with the results of careful visual inspection or available independent ground-truth information. Experimental results show a classification accuracy of 87%, which represents a benchmark for successive studies. Finally, if images of the same area are available at different times, the described classification method can also be used to automatically track of changes (such as new buildings), that occurred in the area under observation.

FFT Transforms

The Fourier transform results are promising for an efficient urban spatial structure characterization. The wavenumber spectra show similarities between features from typical areas from different cities (e.g. the Manhattan and the Rome high density urban areas spectra), and between different areas of the same cities (e.g. San Francisco downtown and San Francisco high density urban areas spectra). By comparing the wavenumber spectrum of the downtown areas of

Rome and New York (Manhattan), we can see the same kind of perpendicular orientation of the spectral ridges but with different angle orientations and the Rome spectrum has a higher and wider spectral ridges than those for Manhattan. These spectral characteristics reflect the urban features represented in the two images. In fact, the streets in the two images have the same kind of perpendicular orientation, but the Roman streets are larger. Also the buildings, in both images, have the same kind of orientation, but those in Rome are lower, brighter and wider. We can conclude that Rome and Manhattan downtown, have a similar general structure with different specifics. It's worthwhile to consider how these two cities are different e.g. in their history, how they were built up, age, climate and location: with the wavenumber spectra method we are able to automatically and rapidly find, analogies and differences between very disparate places around the world. We also investigated three airports La Guardia (NYC), Logan (Boston International Airport) and San Francisco International Airport (we said previously in Section 5.3.1 that the spectrum of the smaller Ciampino airport is not useful for our characterization purpose). Comparing the Logan spectrum with the San Francisco airport spectrum, the latter spectrum has a higher central peak energy, and the angles between the energy axes are completely different.

Moreover, the speckle effect (representing noise in the image) present in the San Francisco airport spectrum is higher than for Logan and almost similar to that for La Guardia. We can conclude that the spatial structures of these three selected airport are quite different. In fact, each airport is different in its location relative to the city (e.g. La Guardia is close to the city with runways in one direction), number and direction of the runways, tarmacs, airplanes, and roads along the runways.

Considering now, the high density urban areas compared to the downtown area of New York city: the Brooklyn spectrum has two preferred directions similar to Manhattan, but the spectrum is brighter, with higher and more

widey spread spectral energies, especially along one axis direction. In fact, the streets perpendicular orientation in Brooklyn is similar to Manhattan but the buildings are brighter and lower making the "canyon city" effect weaker. These analogies and differences clearly come out in the spectral characteristic of each area.

Analyzing the high density urban spectra of Boston and New York, they both have two preferred directions and similar energy distributions, but with different angles. In fact, both areas have perpendicular streets but with different orientations. Moreover, the Boston buildings are similar to the Brooklyn ones but smaller, and the overall image is more homogeneous in brightness.

Comparing the spectra related to the San Francisco downtown area and its high density urban area, we find that in the housing area the energy is distributed along two axes while in the downtown area the energy distribution is only along one major axis. This difference is related to the different road orientations. The downtown area has been built up somewhat randomly over time, consequently the roads appear randomly oriented and in different directions while the high density urban area seems more ordered and well planned. Here the roads are more perpendicular, following a predefined orientation. The high density urban area spectrum of Rome is very similar to the downtown area spectrum of the same city; the energy is along two perpendicular axes. In general, we can conclude that the spectral energy of the high density urban areas from the four selected cities, is mostly distributed along two preferred axes, although these are different for each city in orientation and intensity.

We conclude our study, with the Industrial/Commercial areas spectral analyses. The insustrial/commercial spectrum of NYC, shows two non-perpendicular major axes, probably, mostly related to the different roof buildings types of materials. The San Francisco commercial area is similar to the downtown area but higher in energy. The commercial areas of Boston and Rome have spectra with high noise levels, indicated by a high speckle effect. This can be related to

large features which are relatively uniform in structure typical of the commercial/industrial buildings in these areas.

We have demonstrated that the wavenumber spectral analysis allows us to find differences and analogies between different cities in different location, with different climates, and different histories, as well as between different urban environments of the same cities. Moreover, the spectral characteristics help us to understand the way that the cities have been built up.

In general, with the wavenumber spectral technique we are able to characterize urban features exploiting both the spatial spectral and contextual information contained in the high resolution data. The technique can be applied to enable us to better manage the problems related to large data sets resulting from high-resolution satellite images (i.e. image storage, data exchange and processing time). In fact, spectral analysis is also an efficient data compression technique preserving the original images spectral values. This work can be considered as a first step in demonstrating how NN and FFT algorithms can contribute to the development of Image Information Mining in Earth Observation.

Bibliography

Ackerman, F. (1995). Sensor and data integration - The new challenge , *Proc. ISPRS Workshop Integrated Sensor Orientation*, Barcelona, Spain, September 1995, pp. 2-10.

Angulo, C., Parra, X., and A. Catala (2003). K-SVCR. A support vector machine for multi-class classification, *Neurocomput.*, **55**, 57-77.

Atkinson, P.M., and A.R.L. Tatnall (1997). Neural networks in remote sensing, *Int. J. Remote Sens.*, **18**, 699-709.

Augusteijn, M.F., Clemens, L.E., and K.A. Shaw (1995). Performance evaluation of texture measures for ground cover identification in satellite images by means of a neural network classifier, *IEEE Trans. Geosci. Remote Sens.*, **33**, 616-626.

E.P. Baltasavias (2004). Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps toward operational systems, *ISPRS J. of Photogramm and Remote Sens.*, **58**, 129-151.

Barnsley, M.J., and S.L. Barr (1996). Inferring urban land use from satellite sensor images using kernel-based spatial reclassification, *Photogramm. Eng. and Remote Sens.*, **62**, 949-958.

Bastin, L. (1997). Comparison of fuzzy c-means classification, linear mixture

modeling and MLC probabilities as tools for unmixing coarse pixels, *Int. J. Remote Sens.*, **18**, 3629-3648.

Belousov, A.I., Verzakov, S.A., and J. von Frese (2002). A flexible classification approach with optimal generalization performance: Support vector machines, *Chemometr. Intell. Lab. Syst.*, **64**, 15-25.

Benediktsson, J.A., Swain, P.H., and O.K. Ersoy (1990). Neural network approaches versus statistical methods in classification of multisource remote sensing data, *IEEE Trans. Geosci. Remote Sens.*, **28**, 540-552.

Benediktsson, M., Pesaresi, and K. Arnason (2003). Classification and feature extraction for remote sensing images from urban areas based on morphological transformations, *IEEE Trans. Geosci. Remote Sens.*, **41**, 1940-1949.

Benediktsson, J.A. and I. Kanellopoulos (1999). Classification of multisource and hyperspectral data based on decision fusion, *IEEE Trans. Geosci. Remote Sens.*, **37**, 1367-1377.

Benz, U.C., Hofmann, P., Willhauck, I., Lingenfelder, I., and M. Heynen (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information, *ISPRS J. Photogramm. Remote Sens.*, **58**, 239-258.

Bernard, A.C., Wilkinson G.G. and I. Kanellopoulos (1997). Training strategies for neural network soft classification of remotely-sensed imagery, *Int. J. Remote Sens.*, **18**, 1851-1856.

Binaghi, E., Gallo, I., and M. Pepe (2003). A neural adaptive model for feature extraction and recognition in high resolution remote sensing imagery, *Int. J. Remote Sens.*, **24**, 3947-3959.

Binaghi, E., Brivio P.A. and A. Rampini (editors) (1996). *Soft Computing in Remote sensing Data Analysis*, (Singapore: World Scientific).

Binaghi, E., Gallo, I., and M. Pepe (2003). A cognitive pyramid for contextual classification of remote sensing images, *IEEE Trans. Geosci. Remote Sens.*, **41**, 2906-2922.

Bishof, H., Shneider, W., and A.J. Pinz (1994). Multispectral classification of Landsat images using neural networks, *IEEE Trans. Geosci. Remote Sens.*, **30**, 482-490.

Bishop, C. (1995). *Neural Networks for Pattern Recognition*, Oxford University Press, 482 pp.

Brown, H., Gunn, S.R., and H.G. Lewis (1999). Support vector machines for optimal classification and spectral unmixing, *Ecol. Mod.*, **120**, 167-179.

Bruzzone, L., Conese, C., Maselli, F., and F. Roli (1997). Multisource classification of complex rural areas by statistical and neural-network approaches, *Photogramm. Eng. and Remote Sens.*, **63**, 523-533.

Bruzzone, L., Prieto, D.F., and S.B. Serpico (1999). A neural-statistical approach to multitemporal and multisource remote-sensing image classification, *IEEE Trans. Geosci. Remote Sens.*, **37**, 1350-1359.

Bruzzone, L., and L. Carlin (2006). A multilevel context-based system for classification of very high spatial resolution images, *IEEE Trans. Geosci. Remote Sens.*, **44**, 2587-2600.

Bruzzone, L., and S.B. Serpico (1997). An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images, *IEEE Trans. Geosci. Remote Sens.*, **35**, 858-867.

Burnett, C., and T. Blaschke (2003). A multi-scale segmentation/object relationship modeling methodology for landscape analysis, *Int. J. Ecol. Model. Syst. Ecol.*, **168**, 233-249.

Carlotto, M.J. (1991). Spectral shape classification of Landsat thematic mapper imagery, *Photogramm. Eng. and Remote Sens.*, **64**, 905-913.

Carlson, T. (2003). Applications of remote sensing to urban problems, *Remote Sens. of Env.*, **86**, 275-285.

Carper, W., Lillesand, T., and R. Kiefer (1990). The use of intensity Hue saturation transformations for merging SPOT panchromatic and multispectral image data. *Photogramm. Eng. and Remote Sens.*, **56**, 459-467.

Cetin, H., Warner, T.A., and D.W. Levandowski (1993). Data classification, visualization, and enhancement using n-dimensional probability density functions (nPDF): AVIRIS, TIMS, TM and geophysical applications, *Photogramm. Eng. Remote Sens.*, **59**, 1755-1764.

Chavez, P., Sides, S., and J. Anderson (1991). Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic. *Photogramm. Eng. and Remote Sens.*, **57**, 265-303.

Chavez, P., and D.J. MacKinnon (1994). Automatic detection of vegetation changes in the sothwestern united states using remotely sensed images. *Photogramm. Eng. and Remote Sens.*, **60**, 571-583.

Coppin, P.R., and M.E. Bauer (1994). Processing of multitemporal Landsat TM imagery to optimize extraction of forest cover change features, *IEEE Trans. Geosci. Remote Sens.*, **32**, 918-927.

Datcu, M.; Daschiel, H.; Pelizzari, A.; Quartulli, M.; Galoppo, A.; Colapicchioni, A.; Pastori, M.; Seidel, K.; Marchetti, P.G. and S. D'Elia (2003). Information Mining in Remote Sensing Image Archives: System Concepts, *IEEE Trans. Geosci. Remote Sens.*, **41** , 2923-2936.

Dawson, M.S., (1994). Applications of electromagnetic scattering models to

parameter retrieval and classification, in *A. Fung Microwave Scattering and Emission Models and their Application*, Norwood, MA: Artech House.

Debeir, O., Van den Steen, I., Latinne, P., Van Ham, P., and E. Wolff (2002). Textural and contextual land-cover classification using single and multiple classifier systems, *Photogramm. Eng. and Remote Sens.*, **68**, 597-605.

Del Frate, F., Lichtenegger, J., and D. Solimini (1999). Monitoring urban areas using ERS-SAR data and neural networks, *Proc. IGARSS*, Hamburg, Germany, July 1999.

Del Frate, F., Petrocchi, A., Lichtenegger, J., and G. Calabresi (2000). Neural networks for oil spill detection using ERS-SAR data, *IEEE Trans. Geosci. Remote Sens.*, **38**, 2282-2287.

Del Frate, F., Schiavon, G., and C. Solimini (2004). Application of Neural Networks algorithms to Quickbird imagery for classification and change detection of urban areas, *Proc. IGARSS*, **2**, Anchorage, Alaska, USA, pp. 1091-1094.

Del Frate, F., Schiavon, G., Pacifici, F., and C. Solimini (2006). Use of neural networks for automatic classification from high-resolution images, *IEEE Trans. Geosci. Remote Sens.*, Image Information Mining Issue, in press.

Dell'Acqua, F., Gamba, P., and G. Lisini (2003). Improvements to urban area characterization using multitemporal and multiangle SAR images, *IEEE Trans. Geosci. Remote Sens.*, **41**, 1996-2004.

De Martino, M., Causa, F., and S.B. Serpico (2003). Classification of optical high resolution images in urban environment using spectral and textural information, *Proc. IGARSS*, Touluse, France, July 2003, 467-469.

Donnay, J.P., Barnsley, M.J., and P. Longley (2001). *Remote Sensing and urban Analysis*, Taylor and Francis, pp. 3-18.

Dreyer, P. (1993). Classification of land cover using optimezid neural nets on SPOT data, *Photogramm. Eng. Remote Sens.*, **59**, 617-621.

Du, L., and J.S. Lee (1996). Fuzzy classification of earth terrain covers using complex polarimetric SAR data, *Int. J. Remote Sens.*, **17**, 1317-1340.

Foody, G.M. (1996). Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data, *Int. J. Remote Sens.*, **17**, 1317-1340.

Foody, G.M. (2002). Hard and soft classifications by a neural network with a nonexhaustively defined set of classes, *Int. J. Remote Sens.*, **23**, 3853-3864.

Foody, G.M. (2002). Status of land cover classification accuracy assessment, *Remote Sens. Environ*, **80**, 3375-3380.

Foody, G.M. and M. Ajay (2004). A relative evaluation of multiclass image classification by support vector machines, *IEEE Trans. Geosci. Remote Sens.*, **42**, 1335-1343.

Foody, G.M., McCulloch, M.B., and W.B. Yates (2004). The effect of training set size and composition on artificial neural network classification, *IEEE Trans. Geosci. Remote Sens.*, **42**, 1335-1343.

Franklin, S.E. (1994). Discrimination of subalpine forest species and canopy density using digital CASI, spot PLA, and Landsat TM data, *Photogramm. Eng. Remote Sens.*, **60**, 1233-1241.

Franklin, S.E., Maudie A.J., and M.B. Lavigne (2001). Using spatial co-occurence texture to increase forest structure and species composition classification accuracy, *Photogramm. Eng. Remote Sens.*, **67**, 849-855.

Friedl, M.A. and C.E. Brodley (1997). Decision tree classification of land cover from remotely sensed data, *Remote Sens. Environ.*, **61**, 399-409.

Frizzelle, B.G., and A. Moody (2001). Mapping continuous distributions of land cover: A comparison of maximum-likelihood estimation and artificials neural networks, *Photogramm. Eng. Remote Sens.*, **67**, 693-705.

Fung, T., and E. LeDrew (1987). Application of principal components analysis to change detection, *Photogramm. Eng. and Remote Sens.*, **53**, 1649-1658.

Fung, T., (1990). An assessment of TM imagery for land-cover change detection, *IEEE Trans. Geosci. Remote Sens.*, **28**, 681-684.

Gamba, P., and B. Houshmand (2001). An efficient neural classification chain of SAR and optical urban images, *Int. J. Remote Sens.*, **22**, 1535-1553.

Goel, P.K., Prasher S.O., Patel R.M., Landry J.A., Bonnell R.B., and A. A. Viau (2003). Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn, *Comput. Electron. Agricult.*, **39**, 67-93.

Gong, P., and P.J. Howarth (1990). The use of structural information for improving land-cover classification accuracies at rural-urban fringe, *Photogramm. Eng. and Remote Sens.*, **56**, 67-73.

Green, K., Kempka D., and L. Lackey (1994). Using remote sensing to detect and monitor land-cover and land-use change, *Photogramm. Eng. Remote Sens.*, **3**, 331-337.

Halldorsson, G.H., Benediktsson J.A., and J.R. Sveinsson (2003). Support vector machines in multisource classification, *Proc. IGARSS*, Touluse, France, July 2003.

Hansen, M., Dubayah R., and B. Houshmand (1996). Classifications trees: An alternative to traditional land cover classifier, *Int. J. Remote Sens.*, **17**, 1075-1081.

Harlick, R.M., and L.G. Shapiro (1985). Image segmentation techniques, *Comput. Vis. Graph. Image. Process.*, **29**, 100-132.

Hardin, P.J. (1994). Parametric and nearest-neighbor methods for hybrid classification: A comparison of pixel assignment accuracy, *Photogramm. Eng. Remote Sens.*, **60**, 1439-1448.

Heerman, P.D., and N. Khazenie (1992). Classification of multispectral remote sensing data using back propagation neural network, *IEEE Trans. Geosci. Remote Sens.*, **30**, 81-88.

Henderson, F.M., and Z. G. Xia (1997). SAR applications in human settlement detection, popolation estimation and urban land use pattern analysis: A status report, *IEEE Trans. Geosci. Remote Sens.*, **35**, 79-85.

Hornik, K., Stinchcombe M., and H. White, Multilayer feedforward networks are universal approximators, *Neural Networks*, **2**, 359-366.

Hsu, C.W., and C.J. Lin (2002). A comparison of methods for multiclass support vector machines, *IEEE Trans. Neural Networks*, **13**, 415-425.

Hsu, S.Y., Masters, T., Olson, M., Tenorio, M., and T. Grogan (1992). Comparative analysis of five neural network models, *J. of Intelligent Information Syst.*, **19**, 7-23.

Hsu, W., Lee, M.L. and J. Zhang (2002). Image Mining: Trends and Developments, *Remote Sensing Rev.*, **6**, 319-329.

Huang, C., Davis, L.S., and J.R.G. Townshend (2002). An assessment of support vector machines for land cover classification, *Int. J. Remote Sens.*, **23**, 725-749.

Jensen, J.R., and D.C. Cowen (1999). Remote sensing of urban/suburban infrastructure and socio-economic attributes," Photogrammetric Engineering and Remote Sensing, *Photogramm. Eng. Remote Sens.*, **65**, 611-622

Ji, C.Y. (2000). Land-use classification of remotely sensed data using Kohonen self-organizing feature map neural networks, *Photogramm. Eng. and Remote Sens.*, **66**, 1451-1460.

Jimenez, L., Morales-Morell, A., and A. Creus (1999). Classification of hyper-dimensional data based on feature and decision fusion approaches using projection pursuit, majority voting, and neural networks, *IEEE Trans. Geosci. Remote Sens.*, **37**, 1360-1366.

Jones, J., Thomas, R., and P. Earwicker (1991). Multiresolution analysis of remotely imagery, *Int. J. Remote Sens.*, **51**, 311-316.

http://speclib.jpl.nasa.gov/forms/asp/mineral.htm.

Kanellopoulos, I., and G.G. Wilkinson (1997). Strategies and best practice for neural network image classification, *Int. J. Remote Sens.*, **18**, 711-725.

Kanellopoulos, I., G.G. Wilkinson and J. Mergier (1992). Land-cover discrimination in SPOT-HRV imagery using an artificial neural network: a 20-class experiment, *Int. J. Remote Sens.*, **13**, 917-924.

T. Kohonen. The Self-Organizing Map, *IEEE Proc.*, **78**, 1464-1480.

Kontoes, C., Wilkinson G.G., Burrill A., Goffredo, S., and J. Mergier (1993). An experimental system for the integration of GIS data knowledge-based image analysis for remote sensing of agricolture, *Int. J. Geograph. Inf. Syst.*, **7**, 247-262.

Kumar, A.S., and K.L. Majumder (2001). Information fusion in tree classifiers, *Int. J. Remote Sens.*, **22**, 861-869.

Lauret, P., Fock, E., and T. Mara (2006). A node pruning algorithm based on a Fourier amplitude sensitivity test method, *IEEE Trans. on Neural Networks*, **17**, 273-293.

Lee, Y., G. Whaba, and S.A. Ackerman (2003). Cloud classification of satellite radiance data by multi-category support vector machines, Dept. Stat., Univ. Wisconsin, Madison, Tech. rep. 1075, 2003.

Li, T.S., C.Y. Chen, and C.T. Su (2003). Comparison of neural and statistical algorithms for supervised classification of multi-dimensional data, *Int. J. Indus. Eng.-Theory Appl. Pract.*, **10**, 73-81.

Lippman, R.P., (1987). An introduction to computing with neural nets, *IEEE ASSP Magazine*, pp. 4-22.

Liu, C.M., L.J. Zhang, C.J. Davis, D.S. Solomon, T.B. Brann and L.E. Caldwell (2003). Comparison of neural networks and statistical methods in classification of ecological habitats using FIA data, *Forest Sci.*, **49**, 619-631.

McIver, D.K., and M.A. Friedl (2002). Using prior probabilities in decision-tree classification of remotely sensed data, *Remote Sens. Environm*, **81**, 253-261.

Mercier, G., and M. Lennon. Support vector machines for hyperspectral image classification with spectral-based kernels, *Proc. IGARSS*, Touluse, France, July 2003.

Molinier, M., J. Laaksonen and T. Hame (2006). A self-organizing map framework for detection of man-made structures and changes in satellite imagery, *Proc. IGARSS*, August 2006, Denver, U.S.A.

Møller, M., "A scaled conjugate gradient algorithm for fast supervised learning,", *Neural Networks*, **6**, 525-533.

Moller-Jensen, L. (1990), Knowledge-based classification of an urban area using texture and context information in Landsat-TM imagery, *Photogramm. Eng. Remote Sens.*, **17**, 3339-3355.

Mott, C., T. Andresen, S. Zimmermann, T. Schneider, and U. Ammer (2002). Selective region growing - an approach based on object oriented classification routine , *Proc. IGARSS*, June 2002, Toronto, Ontario, Canada.

Muchoney, D.M., and B.N. Haack (1987). Change detection for monitoring forest defoliation, *Photogramm. Eng. and Remote Sens.*, **60**, 1243-1251.

Nabeel, H.K., A.A. Tawfiq and H.A. Samaher (2006). Design and implementation of classification system for satellite images based on soft computing techniques, Information and Communication Technologies, 2006. ICTTA '06. 2nd.

Pacifici, F., Del Frate, F., C. Solimini and W.J. Emery (2006). An innovative neural-net method to detect temporal changes in high-resolution optical satellite imagery, *IEEE Trans. Geosci. Remote Sens.*, in press.

Pal, M., and P.M. Mather (2003). An assesment of the effectiveness of decision tree methods for land cover classification, *Remote Sens. Environ.*, **86**, 554-565.

Paola, J.D., and R.A. Schowengerdt (1995). A review and analysis of backpropagation neural networks for classification of remotely-sensed multispectral imagery, *Int. J. Remote Sens.*, **16**, 3033-3058.

Peddle, D.R., and S.E. Franklin (1992). Multisource evidential classification of surface cover and frozen ground, *Int. J. Remote Sens.*, **13**, 3375-3380.

Peddle, D.R., Foody G.M., Zhang A., Franklin and E.F. LeDrew (1994). Multisource image classification II: An empirical comparison of evidential reasoning and neural network approaches, *Can. J. Remote Sens.*, **20**, 396-407.

Perez-Cruz, F., and A. Artes-Rodriguez (2002). Puncturing multi-class support vector machines, *Lecture Notes in Computer Science*, **2415**, 751-756.

Petrakos, M., Benediktsson, J.A., and I. Kanellopoulos (2001). The effect of classifier agreement on the accuracy of the combined classifier in decision level fusion, *IEEE Trans. Geosci. Remote Sens.*, **39**, 2539-2546.

Pohl, C., and J. Van Genderen (1998). Multisensor image fusion in remote sensing: concepts, method and applications, *Int. J. Remote Sens.*, **19**, 3169-3180.

Rego, L., and B. Koch (2003). Automatic classification of land cover with high resolution data of the Rio de Janeiro city Brazil: comaprison between pixel and object classification, *Proc. of the* 4th *Intern. Symp. on Remote Sens. of Urban Areas*, 2003, pp. 1682-1777.

Richards, J.A. (1993). *Remote Sensing Digital Image Analysis*, New York: Springler-Verlag, 281 pp.

Rogan, J., Franklin, J., and D.A. Roberts (2002). A comparison of methods for monitoring multitemporal vegetation change using Thematic Mapper imagery, *Remote Sens. Environ.*, **80**, 143-156.

Rumelhart, D.E., Hinton, G.E., and R.G. Williams (1986). Learning internal representations by error propagation, in *Parallel Distributed Processing*, D.E. Rumelhart and J.L. McClelland, Eds Cambridge: MIT Press.

Schiavon, G., Del frate, F., and C. Solimini (2003). High resolution multispectral analysis of urban areas with Quickbird imagery and sinergy with ERS data, *Proc. IEEE Trans. Geosci. Remote Sens.*, **3**, Touluse, France, July 2003, pp. 1972-1974.

Schiavon, G., and D. Solimini (2000). Dihedral and trihedral corner reflector models for the interpretation of radar returns from urban manufacts, *Proc. IEEE Trans. Geosci. Remote Sens.*, **3**, Honolulu, Hawaii, U.S.A, July 2000, pp. 2882-2884.

Schiavon, G., Solimini, D., and A. Mandile (2001). Modelling returns from urban manufacts: simulations results for finite trihedrals, *Proc. IEEE Trans. Geosci. Remote Sens.*,**3**, Sydney, Australia, July 2001, pp. 3120-3122.

Schowengerdt, R.A., *Remote Sensing. Models and Methods for Image Processing*,2nd ed., Norwell, MA: Academic, 2002.

Seong, J.C., and E.L. Usery (2001). Fuzzy image classification for continental-scale multitemporal NDVI series images using invariant pixels and an image stratification method, *Photogramm. Eng. Remote Sens.*, **67**, 287-294.

Serpico, S.B., and F. Roli (1995). Classification of multisensor remote-sensing images by structured neural networks, *IEEE Trans. Geosci. Remote Sens.*, **3**, 562-578.

Shackelford, A.K., and C.H. Davis (2003). A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas, *IEEE Trans. Geosci. Remote Sens.*, **4**, 1920-1932.

Shackelford, A.K., and C.H. Davis (2003). A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas, *IEEE Trans. Geosci. Remote Sens.*, **41**, 2354-2363.

Sheeren, D., Quirin, A., Puissant, A., Gancarski, P., and C. Weber (2006). Discovering rules with genetic algorithms to classify urban remotely sensed data, *Proc. IEEE Trans. Geosci. Remote Sens.*, August 2006, Denver, CO, USA, 4 pp.

Singh, A., (1989). Digital change detection techniques using remotely-sensed data, *Int. J. Remote Sens.*, **10**, 989-1003.

Srinivasan, A., and J.A. Richards (2001). Knowledge based techniques for multi-source classification, *Int. J. Remote Sens.*, **11**, 3339-3355.

Tonjes, R., Growe, S., Buckner, J., and C.E. Liedtke (1999). Knowledge-based interpretation of remote sensing images using semantic nets, *Photogramm. Eng. and Remote Sens.*, **65**, 691-699.

Townshend, J.R.G., and C.O. Justice (1995). Spatial variability of images and the monitoring of changes in the Normalized Difference Vegetation Index, *Int. J. Remote Sens.*, **16**, 2187-2195.

Tso, B.C.K., and P.M. Mather (1999). Classification of moultisource remote sensing imagery using a genetic algorithm and Markov random fields, *IEEE Trans. Geosci. Remote Sens.*, **37**, 811-821.

Unsalan, C., and K.L. Boyer (2004). Classifying land development in high resolution panchromatic satellites images using straight-line statistics, *IEEE Trans. Geosci. Remote Sens.*, **42**, 907-919.

www-ra.informatik.uni-tuebingen.de/SNNS.

Wald, L., Ranchin, T., and M. Mangolini (1997). Fusion of satellites images of different spatial resolutions: assessing the quality of resulting images, *Photogramm. Eng. and Remote Sens.*, **63**, 691-699.

Welch, R. (1982). Spatial resolution requirements for urban studies, *Int. J. Remote Sens.*, **3**, 139-146.

Werbos, P.J. (1974). Beyond regression: new tools for prediction and analysis in the behavioural sciences, *PhD thesis*, Harvard Univ., Cambridge, MA, 1974.

Wilkinson, G.G. (2005). Results and implications of a study of fifteen years of satellite image classification experiments, *IEEE Trans. Geosci. Remote Sens.*, **43**, 433-440.

Wilkinson, G.G., Fierens, F., and I. Kanellopoulos (1995). Integration of neural and statistical approaches in spatial data classification, *Geograph. Syst.*, **2**, 1-20.

Wilkinson, G.G., and J. Megier (1990). Evidential reasoning in a pixel classification hierarchy-A potential method for integrating image classifiers and expert system rules based on geographic context, *Int. J. Remote Sens.*, **11**, 1963-1968.

Wilson, E.H., Fierens, F., Hurd, J.D., Civco, D.L., Prisloe, M.P., and C. Arnold (2003). Development of a geospatial model to quantify, describe and map urban growth, *Remote Sens. Env.*, **86**, 275-285.

Yoshida, T., and S. Omatu (1994). Neural network approach to land cover mapping, *IEEE Trans. Geosci. Remote Sens.*, **32**, 1103-1109.

Zhang, J., and G.M. Foody (1998). A fuzzy classification of sub-urban land cover from remotely sensed imagery, *Int. J. Remote Sens.*, **19**, 2721-2738.

Zhang, Y. (2001), Detection of urban housing development by fusing multisensor satellite data and performing spatial features post-classification, *Int. J. Remote Sens.*, **17**, 3339-3355.

Zhu, M.L., Wang, Y., Chen, S.F., and X.D. Liu (2003). Sphere-structured support vector machines for multi-class pattern recognition, *Lecture Notes in Artificial Intelligence*, **2639**, 589-593.

Chapter 7

Appendix

Use of Neural Networks for Automatic Classification from High-Resolution Images

F. Del Frate, *Member, IEEE,* F. Pacifici, *Student Member, IEEE,* G. Schiavon and C. Solimini

***Abstract*— The effectiveness of Multi-Layer Perceptron (MLP) networks as a tool for the classification of remotely sensed images has been already proven in past years. However, most of the studies consider images characterized by high spatial resolution (around 15-30 m) while a detailed analysis of the performance of this type of classifier on very high resolution (around 1-2 m) images such as those provided by the Quickbird satellite is still lacking. Moreover, the classification problem is normally understood as the classification of a single image while the capabilities of a single network of performing automatic classification and feature extraction over a collection of archived images has not been explored so far. In this paper, besides assessing the performance of MLP for the classification of very high resolution images, we investigate on the generalization capabilities of this type of algorithms with the purpose of using them as a tool for fully automatic classification of collections of satellite images, either at very high or at high resolution. In particular, applications to urban area monitoring have been addressed.**

***Index Terms*— Features extraction, high-resolution imagery, infornation mining, neural networks.**

I. Introduction

NEURAL networks (NN) as a tool in the field of remote sensing have received considerable attention since a new learning scheme was developed. The principle of a back-propagation algorithm was initially proposed by Werbos [1] and rediscovered by Rumelhart [2]. Since the early nineties, several researchers have compared the performance of NN with conventional statistical approaches to remote sensing applications, and in particular to image classification. Benediktsson et al. [3] evaluated the two methods for multi-source remote sensing data classification. They noted that a neural network has great potential as a pattern recognition method for multi-source remotely sensed data because of the distribution-free nature of a neural network. Bishof et al. [4] as well as Paola and Schowengerdt [5] compared methods for multispectral classification of Landsat TM data and both found that with proper training, a neural network was able to perform better than the maximum-likelihood classification. However, even if these studies seem to show that NN performance is comparable or better than that provided by other techniques, they are mainly focused on high resolution Landsat images and on the use of a single neural network for classifying and/or extracting specific features from a single image, namely the image from which the examples training the network are taken. Conversely, a detailed analysis of the pixel-based classification yielded by this type of algorithms on very high resolution images such as those provided by the Quickbird or Ikonos platforms is still lacking. Moreover, the potentialities of a single neural network as a tool for automatic and sequential processing of images contained in archives have been scarcely investigated until now. With processing here we mean that the network might be used to retrieve from the archive all the images that contain or do not contain a specific class of land cover, or where the ratio between areas corresponding to different classes is within/out predefined ranges. In other words the network allows the identification of high-level (object or region of interest) spatial features from the low-level (pixel) representation contained in a raw image or image sequence, hence addressing scientific issues characteristic of the image information mining field [6],[7].

In this work, as a first step, we want to assess and optimise the neural network approach for the pixel-based classification of a single very high resolution image, such as one of those provided by the Quickbird satellite. Later we move to the conceptually most innovative part of the study which is to investigate on the capabilities of supervised NN in providing automatic classification on a collection of images, therefore their potentialities from an image information mining point of view. This means to stress their generalization capabilities, that is the capabilities to obtain good generalization to new input patterns from the patterns on which the nets have been trained. Several issues interfere with the objective of designing NN able to generalize on images not used in the training phase. For example the effects on the measurements due to the different incidence angles and/or atmospheric conditions or the fact that different types of material may characterize the same class. Therefore, the robustness of the spectral information despite such problems has to be investigated and such an analysis needs to concur with the design of the NN. Addressing this point, in the paper we consider both very high (Quickbird) and high (Landsat) resolution images and a specific application domain which is the feature extraction and information discovery on urban areas. In fact, monitoring changes and urban growth over time is one of the major areas of scientific research in remote sensing that have a strong interaction with the policy cycles

The authors are with the Dipartimento di Informatica, Sistemi e Produzione, Universita' Tor Vergata, Via del Politecnico, 1, 00133, Roma, Italy (e-mail: solimini@disp.uniroma2.it).

and that would improve environment and security monitoring [8]-[10]. A large volume of satellite data for such purposes is available, but despite many competing automatic approaches exist, it is difficult to fully and automatically address the problems raised by the different application scenarios. In this study the aim of the classification is to distinguish among areas made of artificial coverage (sealed surfaces) including asphalt or buildings, and open spaces such as bare soil or vegetation.

II. The Nueural Network Algorithm

NN models are mainly specified by the net topology and training rules [11]. The term topology refers to the structure of the network as a whole: the number of its input, output, and hidden units and how they are interconnected. Among various topologies, multilayer perceptrons (MLP) have been found to have the best suited topology for classification and inversion problems [12]. These are feedforward networks where the input flows only in one direction to the output, and each neuron of a layer is connected to all neurons of the successive layer but has no feedback to neurons in the previous layers. As far as the numbers of hidden layers and of their units are concerned, the topology providing the optimal performance should be selected. In fact, if the number of neurons is too small, the input-output associative capabilities of the net are too weak. On the other hand, this number should not be too large; otherwise, these capabilities might show a lack of generality being too much tailored on the training set and the computational complexity of the algorithm would be increased in vain. It turns out that a fair compromise has to be found. The number of hidden layers is another issue to be considered. It has been shown that networks having two layers of weights, i.e. one hidden layer of neurons, and sigmoidal hidden units can approximate arbitrarily well any functional continuous mapping, provided the number of hidden units is sufficiently large [13][14]. However how much the inclusion of an additional hidden layer might improve the classification performance is still an open issue. In this paper we followed a rather heuristic approach. We sistematically analysed the performance of the network varying either the number of hidden layers (one or two) or the number of hidden units and selecting the best topology on the base of the accuracy results obtained on a set of examples not considered for the training. The weight or strength of each connection has to be determined via learning rules to approximate an unknown input-output relation. These rules indicate how to pursue minimization of the error function measuring the quality of the network's approximation on the restricted domain covered by a training set (i.e., a set of input-output examples). A typical error function which can be considered in this context is the sum-of-squares error function (SSE) [13], given by a sum over all patterns, and over all outputs, of the form:

$$SSE = \sum_{n=1}^{N} \sum_{k=1}^{c} \left\{ y_k(\mathbf{x}^n; \mathbf{w}) - t_k^n \right\}^2 \qquad (1)$$

Fig. 1. Quickbird image of the Tor Vergata University Campus, Rome, and its surrounding.

where $y_k(\mathbf{x}^n;\mathbf{w})$ represents the output of unit k as a function of the input vector $\mathbf{x}^n$ and the weight vector $\mathbf{w}$, N is the number of testing patterns, and c is the number of outputs. The quantity t_k^n represents the target value for output unit k when the input vector is $\mathbf{x}^n$. In our case the minimization of the error function has been pursued by a scaled conjugate gradient (SCG) algorithm [15]. This is a member of the class of conjugate gradient methods, general purpose second order techniques that help to minimize goal functions of several variables. Second order indicates that such methods use the second derivatives of the error function, while a first-order technique, like standard backpropagation, only uses the first derivatives. It should be mentioned that most of the neural simulations were provided by the SNNS (Stuttgart Neural Network Simulator) package [16]. For the specific purpose of the image classification a training-set with a statistically significant number of pixels for each class has been generated. The learning of the neural network has then been carried out by feeding it with pairs of vectors (patterns): the input vector contains the reflectances of the different channels of the multi-spectral image, the output vector contains the corresponding known class of surface. To avoid saturation within the network it has been necessary to scale all the values of the input vector in the range between -1 and 1. The scaling has always been carried out consistently on the entire data set available. At the same time, the component of the output vector corresponding to the true class has been set to 1 while the others to 0. Once the NN have been trained, they have been used for the classification of new data not considered in the training set. In the test phase, a competitive approach (winner-and-take) has been considered to decide on the final classification response.

III. Single Image Classification

The QuickBird commercial remote sensing satellite provides images consisting of four multispectral (MS) channels with 2.4 m resolution and a single panchromatic (PAN) band with 0.62 m resolution. The four MS bands collect data at the red, green, blue, and near-infrared wavelengths, and the data in each band is stored with 11-bit quantization. A QuickBird image taken over the Tor Vergata University campus, located in Italy, South-East of Rome, on March 13, 2003, has been initially considered, in the following of the paper we will refer to this image with the name QB1. A view of the area is shown in Fig.1. Besides the buildings in the campus, different residential areas belonging to the outskirts of the south-east side of the city can be distinguished in the image.

Our first purpose was to design an optimum neural network able to classify the multi-spectral image. The considered land cover classes were buildings, roads, vegetated areas, and bare soil where the latter class includes not asphalted road and artificial excavations. The inclusion of additional classes was discarded for several reasons: the considered classes are those

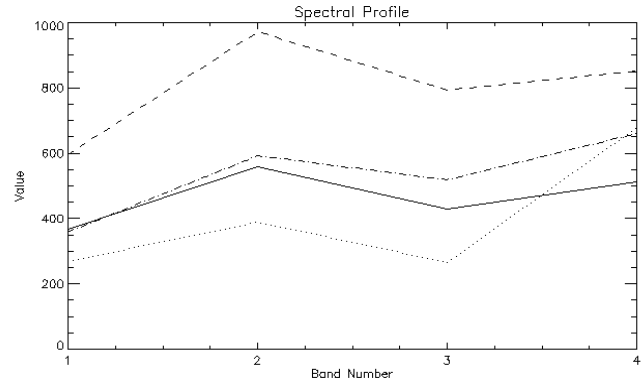


Fig. 2. Spectral analysis from image QB1 for the classes buildings (dashed line), asphalted surface (solid line), bare soil (dash-dotted line), vegetation (dotted line).

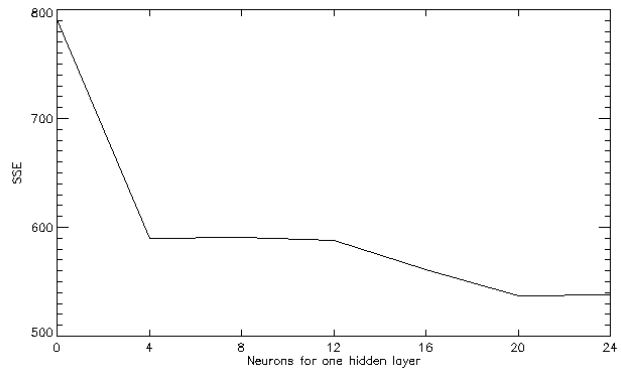


Fig. 3. SSE values calculated over the test set changing the number of hidden neurons in a two hidden layers topology. The number of units is the same in both layers.

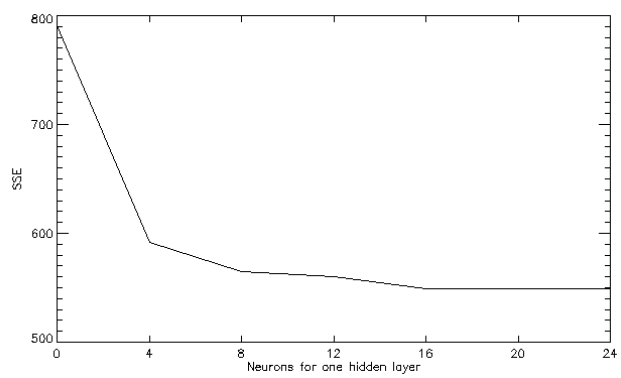


Fig. 4. SSE values calculated over the test set changing the number of hidden neurons in a one hidden layer topology.

that better describe the area under observation and are in themselves sufficient to detect significant features, the choice of a small number of classes enables an easier quantitative comparison of the performance obtained using a single net for a single image classification, with the one obtained using a single net for multiple images classification. In this latter case we think that the choice of a number of 4 classes represents a rather ambitious target. It also has to be noted that a recent study analysing satellite image classification experiments of fifteen years pointed out that the idea postulating that the

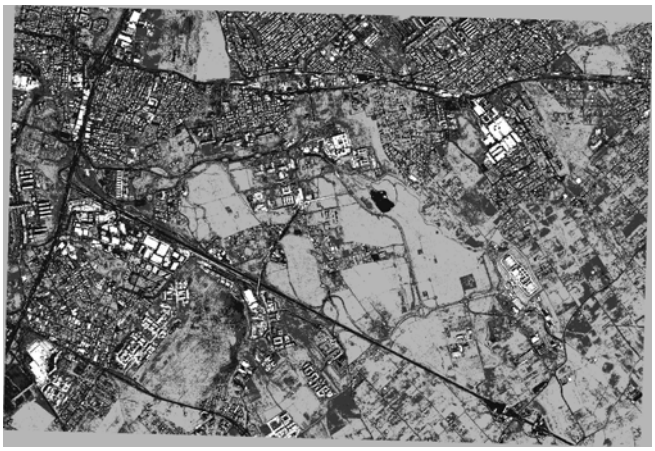


Fig. 5. Classification map of the image QB1 using the optimized topology. Black: asphalted surfaces; white: buildings; dark gray: bare soil; light gray: vegetation.

higher the number of classes used in a classification experiment, the more difficult the classification becomes, is not supported by the experimental results shown in the study [17]. Once the classification problem has been configured, a first investigation consisted in analysing the spectral behaviour of the different considered surfaces. The selected pixels characterizing one class belong to polygons manually drawn in the image. It should be noted that, at the very high resolution of the images, the edges or boundaries between individual land cover objects were fairly sharp and it was usually easy to locate and assign a specific pixel to a land cover class. The mean values of the spectral signatures of the 4 categories are shown in Fig. 2. The figure clearly shows potentiality in discriminating between the classes. This stems from the spectral properties related to the different molecular resonance mechanisms which characterize the materials. With the same data considered for the sensitivity analysis we were able to generate a training-set with a statistically significant number of pixels for each of the four categories. The training datasets were generated considering about 24400 pixels. The design of the network was made putting particular care in the selection of the number of hidden units to be considered in the net. To this purpose the plot illustrated in Fig. 3 was produced, where the SSE value over a test set of more than 3000 patterns and corresponding to different numbers of hidden units is reported. It can be seen that, if we consider both the SSE error and the network complexity, the best results were obtained with a 4-20-20-4 topology. Indeed, the increase of the number of hidden units did not change significantly the SSE error. A similar plot is reported in Fig. 4 where now a single hidden layer is considered. Again the best result are obtained putting around 20 neurons in the hidden layer, however this topology is slightly worse if compared with the two-hidden layers topology. This indicates that the second layer can be able to extract additional information from what already elaborated by the first one. The topology 4-20-20-4 was then finally selected and used to classify the entire image (3506832 pixels). Fig. 5 shows the classification map derived with the described procedure. The classification

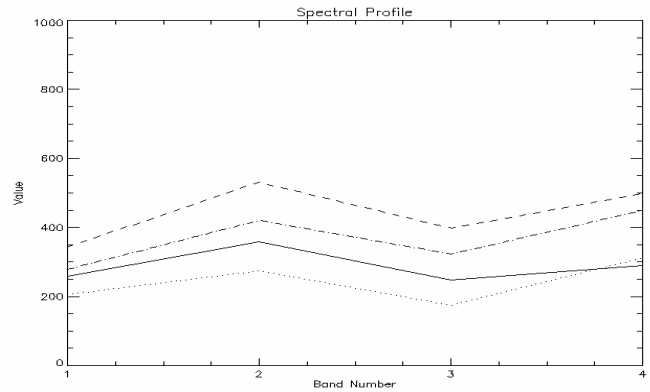


Fig. 6. Spectral analysis from image QB2 for the classes buildings (dashed line), asphalted surface (solid line), bare soil (dash-dotted line), vegetation (dotted line).

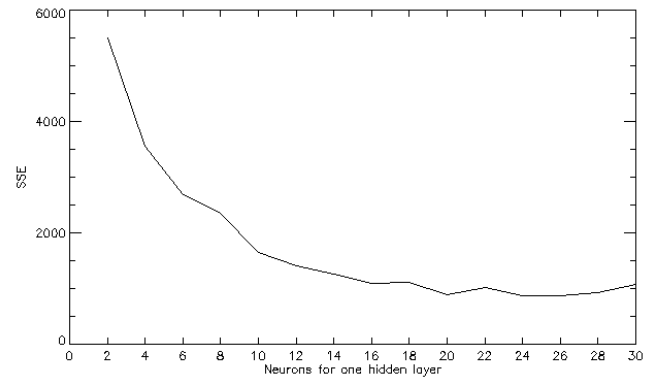


Fig. 7. SSE values calculated over the test set changing the number of hidden neurons in a two hidden layers topology designed for the classification of a collection of Quickbird images. The number of units is the same in both layers

TABLE I

CONFUSION MATRIX OBTAINED FOR IMAGE QB1 WITH THE 4-20-20-4 TOPOLOGY. OVERALL NUMBER OF PIXELS: 81510. OVERALL ERROR 5998 (7.36%)

Classified as	**True**			
	Vegetation	***Asphalt***	***Building***	***Bare soil***
Vegetation	14864	33	750	2207
Asphalt	132	44785	68	27
Building	1225	29	12634	783
Bare soil	230	2	512	3229

TABLE II

CHARACTERISTICS OF THE QUICKBIRD IMAGES USED IN THE WORK. ALL THE ACQUISITION TIMES ARE AROUND 10:00-10:30 A.M.

Code	**Aquisition Date**	**Dimension (pixels)**	**Off Nadir Angle**	**Location**
QB1	03/13/2003	2415x1650	8 Degrees	Rome, SE outskirts
QB2	05/29/2002	2352x1491	11 Degrees	Rome, SE outskirts
QB3	07/19/2004	2415x1650	23 Degrees	Rome, NE outskirts
QB4	07/19/2004	2415x1450	23 Degrees	Rome city
QB5	07/22/2005	2223x1450	12 Degrees	Nettuno town

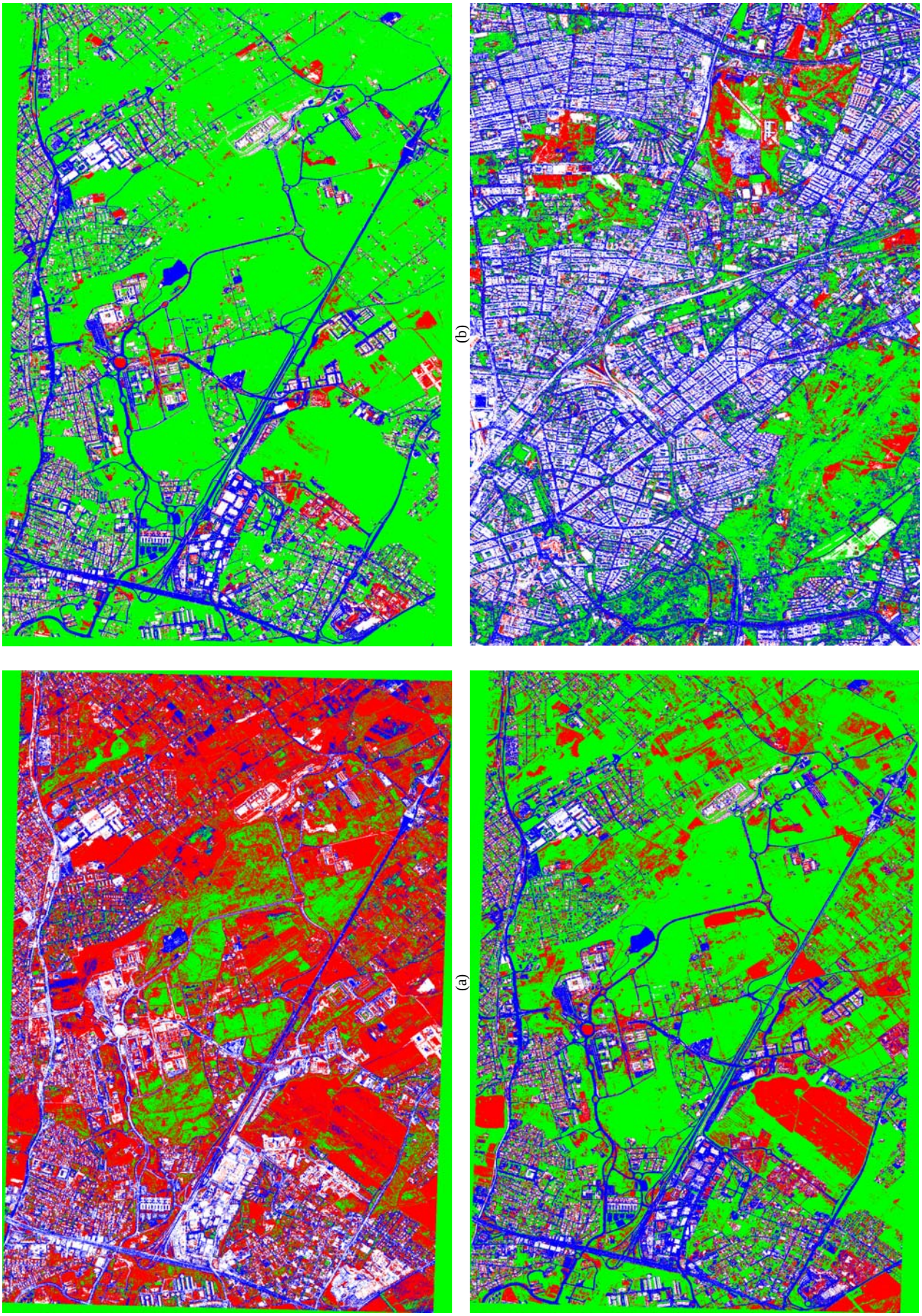


Fig. 8. Automatic classification map from the image (a) QB2 with a net trained with examples taken from image QB1, (b) QB2 using a network trained with examples taken from image QB2, (c) QB2 and (d) QB4 with a net trained with examples taken from other images. Red: bare soil, blue: asphalted surface, white: buildings, green: vegetation.

accuracy has been assessed by visual comparison with the original high resolution image and by direct inspections on site. We stress the fact that our working area is located in the Tor Vergata University campus, that is almost in the centre of image QB1, so direct inspection on site could be rather accurate. More in detail, a ground truth map, corresponding to a subset of the image, has been manually elaborated. We observed that the classification provided by the network is rather accurate and with a high level of resolution. In particular we reached a 93% of accuracy in the considered subimage. The whole confusion matrix is reported in Table 1. Once the network topology for this kind of problem has been optimized and the performance assessed, we move to investigate on the capability of a unique network to provide classification on different images rather than on a single one. To underline the complexity of this new problem we tested the already designed network, positively processing the QB1 image, on another QB image. The choice of this new image should follow some similarity criteria with respect to the already classified one. For example it would not be very meaningful to consider a new image characterized by land cover classes, such as water, not appearing in the QB1 image, hence not memorized at all by the network during its training process. The failure of the neural network in this case can be given for granted and this test would not provide much information in the evaluation of the network generalization capability. Therefore, we decided the other direction of selection and chose as test image a QB image quite similar to QB1. Indeed, the new QB image (QB2) is taken on the same area of the first one, but in a different season and at a slightly different incident angle. In Table 2 the basic information of the two images analysed so far and of those that will be considered in the following of the paper are summarized. If the already trained network fails in generalizing over this image it will be very probably unsuccessful with many other QB images, even if taken on similar urban scenarios. In Fig. 8(a) we show the result of the classification of the QB2 image by using the net trained on patterns retrieved from image QB1. For the sake of completeness and for a better interpretation of the results we also produced the classification, reported in Fig. 8(b), that would be obtained replying on the image QB2 the single image classification methodology considered for the image QB1, therefore relying on a network (4-20-20-4), trained with examples belonging to the same image that one wants to classify. The classification map shown in Fig. 8(b) seems, as expected, rather accurate. Indeed the misclassification percentage computed over the same image subset considered for QB1 is 95% thus resembling the one obtained in the former case. The classification result shown in Fig. 8(a) is completely different. Although the network recognizes many patterns and assigns the correct class to the corresponding pixels, entire objects are misclassified, the bare soil class and the built areas class are definitely overestimated and the general noise level produced by the classification is significantly increased. From a quantitative point of view the misclassification rate computed over the subset test image is 56%. Fig. 6 may contribute to understand the classification performance. We can observe that even if the shapes of the signatures resemble those plotted in Fig. 2, which still enables some possibility of distinction among classes, the ranges of the digital number values are significantly different, generating confusion when the network gives out its classification response. Thus the classification of the QB2 image obtained using a network trained on another image, even if taken on the same scenario, is not satisfactory. This means that to design a network able to provide good accuracy over images not considered in the training phase is an ambitious goal, even if the classification is performed on a limited number of classes.

TABLE III
CONFUSION MATRIX OBTAINED FOR THE CHANGE DETECTION EXERCISE. OVERALL NUMBER OF CONSIDERED PIXELS: 148538. OVERALL ACCURACY 93.2%

Classified as	**True**	
	Changed	***Unchanged***
Changed	14864	33
Unchanged	132	44785

IV. CLASSIFICATION ON A COLLECTION OF IMAGES

A. Quickbird images

Three more Quickbird images have been considered in this case for an overall number of 5 images. As shown in Tab. 2

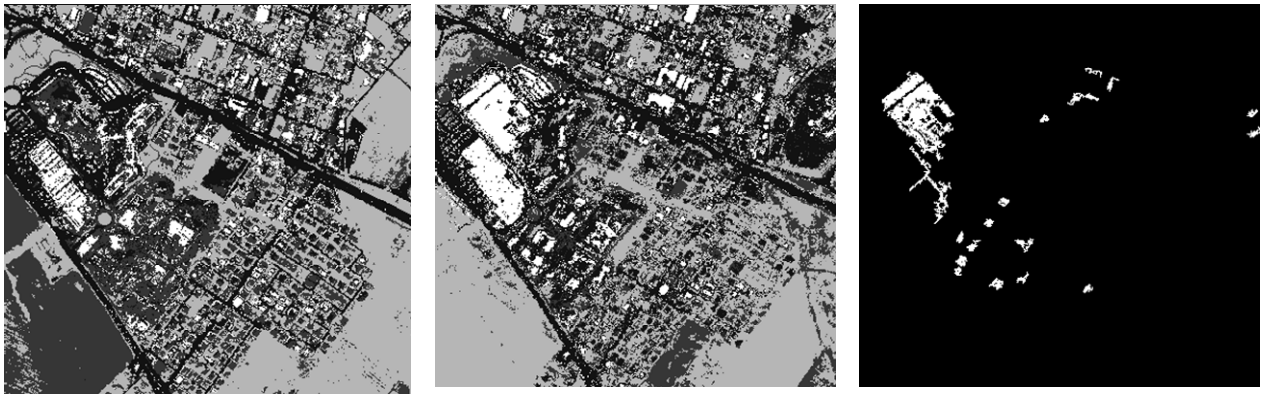


Fig. 9. Change detection results. Left: 2002 classification map results, center: 2003 classification map results, right: change detection.

the 5 images are of similar size but include different years, different sites and different seasons. Besides the QB1 and QB2 images centred on the Tor Vergata University campus, we have one image (QB3) looking at north-east suburbs, a fourth image quite close to the old town (QB4), and a fifth image (QB5) which has been taken on a small urban area a few kilometres away from Rome. A pixel-based classification algorithm has again been implemented to distinguish among the four main classes: buildings, asphalted surfaces, vegetated areas and bare soils. In previous section we showed that a successful classification performance relies on a proper training and design of the network. In particular, it is important that the patterns included in the training set could significantly represent all potential scenarios that might be encountered during the application phase, in other words resemble the statistics of the classification problem. To this purpose a larger archive of spectral signatures has been generated. Images QB1, QB3 and QB5 have been considered for the training and about 26000 examples have been collected for the generation of the network learning set.

The optimal performance both in terms of classification accuracy and of training time has been again determined by an extensive search whose results are illustrated in Fig. 7. With regards to the number of hidden layers we relied on the previous result indicating a topology with two hidden layers more effective so the final selected topology was again 4-20-20-4. Indeed, with respect to the single image processing case, most of the physics characterizing the classification problem has not changed which involved minor implications in terms of the topology to be selected. In Figs. 8(c) and 8(d) we report the classification maps obtained by applying the trained neural network to the images QB2 and QB4 which did not contain any of the pixels included in the training set. From both visual inspection on the original images and direct inspection on site we observed a general good agreement with the map generated automatically. All main features such as big roads and buildings are individuated with good precision even though some inaccuracies can be noted in the objects edge detection, possible causes of disturb being represented by shadow effects. A more quantitative analysis, computed on the same sub-area of image QB2 considered in section 2 gave an overall accuracy rate is of about 87%. Considering the encouraging results, and given the availability of two images (QB1 and QB2) over the same site, we tried to extend the described methodology to a typical change detection exercise. The two images have been co-registered using a set of about 30 ground control points and considering the older image as a master. We remind that the time interval between the two images is of one year. The two corresponding classification maps, obtained by means of the same network, have been used for the production of change detection maps. In particular, the change detection was evaluated in terms of the pixels that migrated from vegetation, bare soil or asphalted surface class to the building class in the considered time window. As in this case we are more interested in an object based result, the final change detection mask was obtained after a post-processing which removed all clusters of pixels detecting changes but containing less than 20 elements. The ground-truth confirmed that the changes corresponding to the main detected structures were buildings constructed in the considered time interval. An example of detection result is shown in Fig. 9 where the previous corresponding classification maps are also reported. The corresponding confusion matrix, reported in Table 3 and computed on the base of the ground-truth, gives an high percentage of pixels in the diagonal. On the other hand, most the pixels out of the diagonal belong to the object circled in red which, more than a real failure of the classification algorithm, is a consequence of an imperfect coregistration of two images.

Fig. 10. Data set geographic distribution of Landsat data.

TABLE IV
LOCATION AND DATES OF THE LANDSAT IMAGES USED FOR THE GENERATION OF THE TRAINING SET.

CITY	Sensor	Acquisition date	Classes
AMSTERDAM	TM	22/05/1992	all
BARCELONA	ETM+	10/08/2000	Unsealed
BERLIN	ETM+	14/08/2000	unsealed
BUDAPEST	ETM+	09/08/1999	All
LONDON	TM	20/05/1992	All
MELBOURNE	ETM+	05/10/2000	All
NEW YORK	TM	28/09/1989	All
PARIS	TM	09/05/1987	All
RIO DE JANEIRO	TM	18/01/1988	all
ROME	ETM+	03/08/2001	all
TOKIO	TM	21/05/1987	all
WIEN	TM	10/09/1991	all
ROME_2	ETM+	16/01/2001	all
UDINE	ETM+	16/08/2000	Sealed, water

The rightmost column indicates which classes have been considered for the specific image.

TABLE V
CONFUSION MATRIX OBTAINED FOR LANDSAT NEURAL ALGORITHM: THE OVERALL ACCURACY IS OF ABOUT 82 %

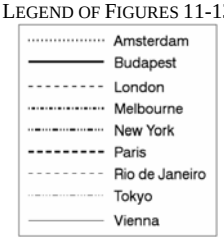
Classified as	True		
	Sealed	***Unsealed***	***Water***
Sealed	332	54	1
Unsealed	52	172	2
Water	2	1	15

B. *Landsat Images*

The objective of designing a single net enabling automatic classification of large archives of data sets has been extended to the case of Landsat imagery. The Landsat data-set consisted of a collection of images containing urban areas and located throughout the all 5 continents (Fig. 10). In this case the inputs to the algorithm are taken from 6 bands, measurements corresponding to band 6 being discarded for its worse spatial resolution. Again, we firstly analysed the spectral signatures of the main classes of urban land cover.

Despite the considerable distances among the geographic locations a good stability of the spectral information has been noted. For example in Fig. 11, 12 and 13 we report the analysis for the classes high density residential, forest and water, respectively. For the three classes the spectral behaviour is calculated starting from an overall number of about 25000 pixels, distributed over 9 considered different geographic areas. We see that within the same class the shapes of the signature are in general rather similar and only a bias value seems to characterize the different plots. On the other hand, different classes have quite dissimilar spectral shapes. The analysis carried out on other classes typical of urban and sub-urban land cover confirmed the discrimination possibilities, especially if similar classes, such as forest and short vegetation areas, or high density and low density residential areas, were grouped together. In this case the final classification problem was to discriminate among three classes: sealed, not sealed and water. The sealed fraction of an urban areas is indeed one of the primary index for monitoring the urbanization process. However, many big cities are characterized by large amount of water surfaces, belonging to rivers, lakes or sea. Therefore, the addition of the class water could be significant to obtain a better monitoring. More than 56000 patterns have been selected to train the final neural network dedicated to the Landsat imagery with examples extracted from an overall set of 14 images including urban areas of 12 world big cities, 12 countries, 4 continents. A description of the training set in terms of the images and classes considered for each image is summarized in Table 4. Given the variety of geographical sites taken into account, this classification problem is inherently more complex with respect to the classification of Quickbird data. In order to avoid overfitting, this shrinks the size of the optimum network topology which, for this case, has been found to be 6-9-9-3 (see Fig. 14). With this selected topology and considering medium speed CPU computing platforms we obtained an average rate of processing of 800 pixel per second. This means less than 20 minutes for an image of 1000 x 1000 pixels, so basically we can speak of near real time processing. In Fig. 15 we show some example of results. The yielded accuracy seems to be rather satisfactory at careful visual inspection. Water bodies are detected rather precisely as the major parts of the urban lattice. On the other hand we noted some inaccuracies on areas which appear as low residential areas at image visual inspection but are labelled as unsealed areas in the classification map. Similar results have been obtained selecting other images from the available data-set. In any case, a more quantitative validation exercise could be

TABLE VI
LEGEND OF FIGURES 11-13



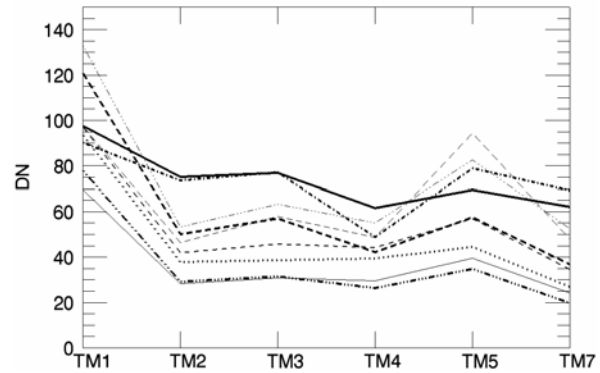


Fig. 11. Spectral analysis from Landsat measurements of the class high-density residential for different cities in the world.

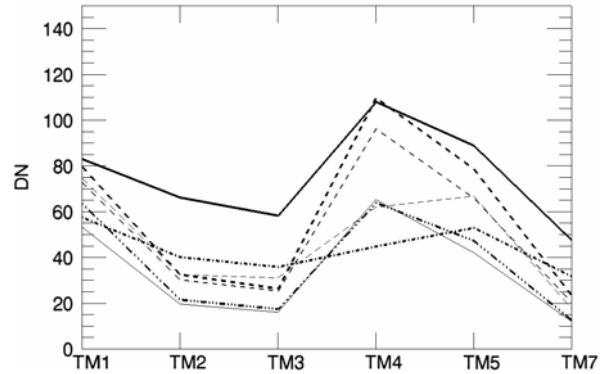


Fig. 12. Spectral analysis from Landsat measurements of the class forest for different cities in the world.

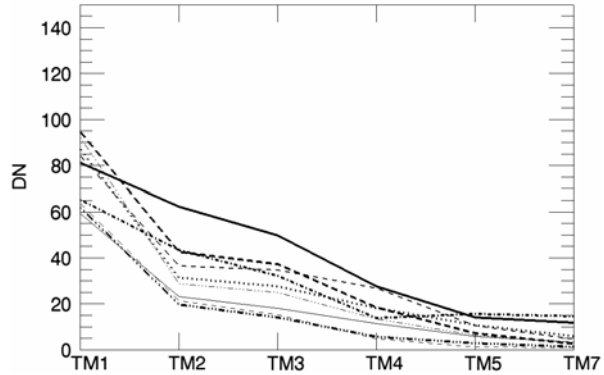


Fig. 13. Spectral analysis from Landsat measurements of the class water for different cities in the world.

performed on a limited area of the city of Rome, where we could use the Quickbird very high resolution image as ground-truth. The area chosen for the validation exercise does not contain pixels used for the training of the final Landsat network. In Table 5 we report the obtained corresponding confusion matrix where we did not include pixels, such as pixels on edges, whose real class could be not stated with certainty. Given the totally automatic procedure, the overall accuracy of about 82% might be recognized as encouraging and establishes a benchmark for this kind of application.

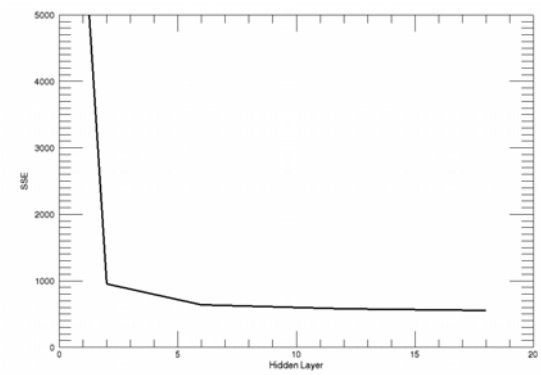


Fig. 14. SSE values calculated over the test set changing the number of hidden neurons in a two hidden layers topology for the classification of a collection of Landsat images. The number of units is the same in both layers.

V. Conclusions

It is well recognized that one of the major advantages of NN with respect to Bayesian and other statistically based classifiers is that NN draw their own input-output discriminant relations directly from the data and do not require that a particular form of a PDF (Probability Density Function) be assumed [18]. In this work we exploited these characteristics of MLP networks for automatic processing of large data sets of satellite imagery and with particular interest for features extraction from urban areas. In fact, the paper can be considered as a first step in demonstrating how NN can contribute at the development of IIM in Earth Observation. We considered two types of satellite data: Quickbird data characterized by very high spatial resolution and the Landsat data characterized by high spatial resolution. In both cases the purpose was both to yield accurate classification maps and to train the networks in order to generalize out of the image data set considered in the training phase so that the new images

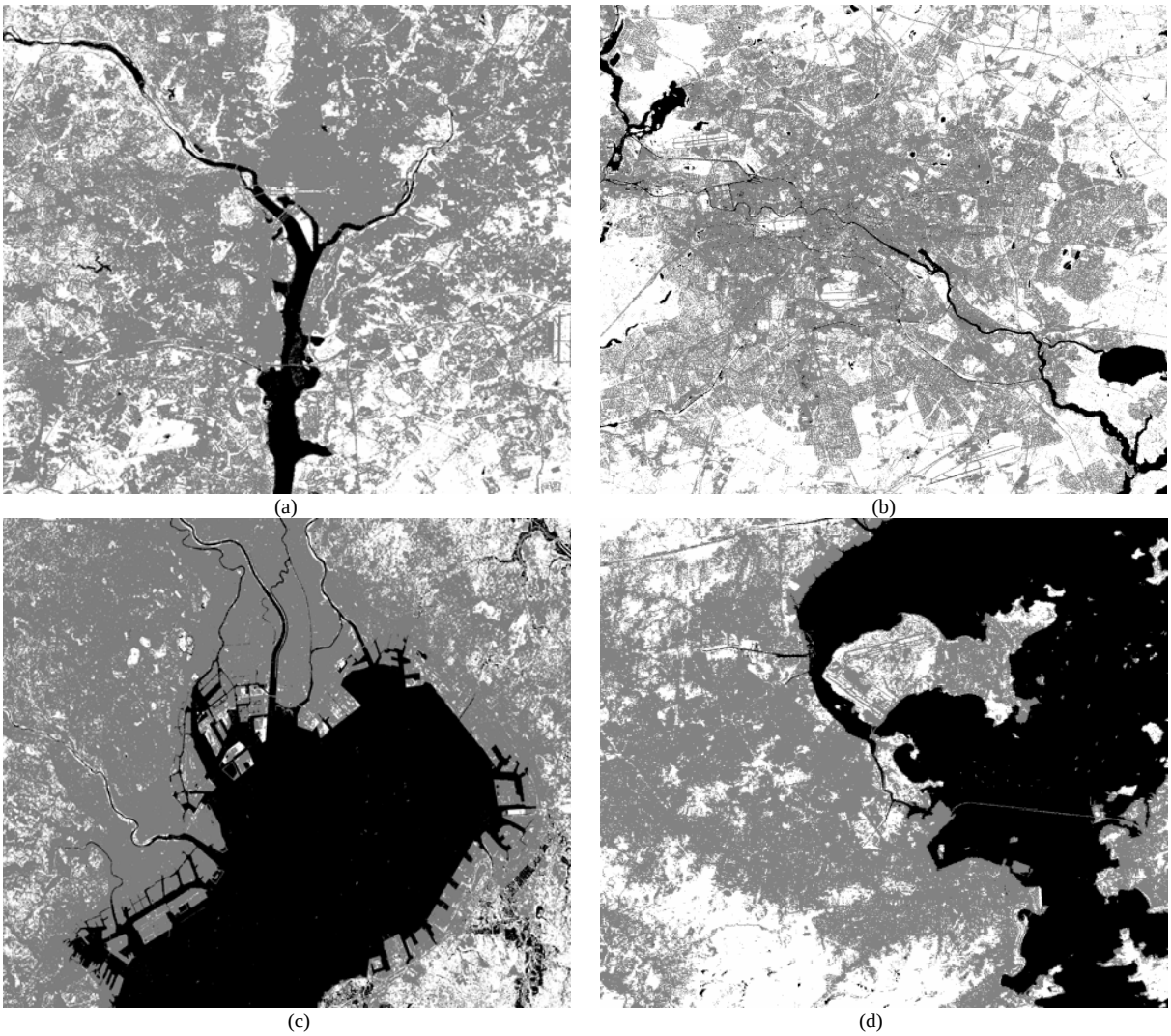


Fig. 15. Automatic classification map of the city of (a) Washington (USA), (b) Berlin (Germany), (c) Tokyo (Japan) and (d) Rio de Janeiro (Brasil). Black: water surface, grey: sealed surface, white: unsealed surface (open space).

could be processed in near-real time. To that purpose careful spectral analysis over statistically significant data sets have been carried out and the NN topologies have been designed avoiding possible effects of overfitting. The networks performance seems to be satisfactory, especially if we take into account that the procedures are completely automatic. In fact, the maps automatically provided on new images, that is not considered in the training phase, show good agreement with those that would be obtained with careful visual inspection or with the available ground-truth. Even though, both for high and very-high spatial resolutions, the experiments have been carried out on similar scenarios, the overall accuracies of 87% and 82% obtained for selected Quickbird and Landsat sub-areas, respectively, represent a benchmark for successive studies. Finally, if images of the same area are available at different times, the described technology seems also to be useful for an automatic discoverage of changes, such as new buildings, occurred in the area under observation.

ACKNOWLEDGMENT

The University of Maryland Global Land Cover Facility (GLCF) has been used for this work and is gratefully acknowledged. The European Space Agency (ESRIN center) is also gratefully acknowledged for making the image QB5 available. Dr. Navulur of DigitalGlobe is gratefully acknowledged as well for making images QB3 and QB4 available. This work also beneficiated of the precious contributions of Giorgio Licciardi.

REFERENCES

[1] Werbos, P.J., "Beyond regression: New tools for prediction and analysis in the behavioural sciences," *Ph.D. thesis*, Harvard Univ., Cambridge, MA, 1974.

[2] Rumelhart, D.E., G.E. Hinton, and R.J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing*, edited by D.E. Rumelhart and J.L McClelland, pp. 318--362, MIT Press, Cambridge, Mass., 1986.

[3] Benediktsson, J. A., Swain, P.H., and Ersoy, O.K., "Neural network approaches versus statistical methods in classification of multisource remote sensing data," *IEEE Trans. Geosci. Rem. Sens.*, vol. 30, pp.482-490, 1992.

[4] H. Bishof, W. Schneider, A. J. Pinz, "Multispectral Classification of Landsat-Images Using Neural Networks," *IEEE Trans. Geosci. Rem. Sens.*, vol. 30, pp.482-490, 1992.

[5] Paola, J. D. and Schowengerdt R. A., "A detailed comparison of back propagation neural network and maximum likelihood classifiers for urban land use classification," *IEEE Trans. Geosci. Rem. Sens.*, vol. 33, pp.981-996, 1995.

[6] Datcu M. et al., "Information Mining in Remote Sensing Image Archives: System Concepts," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 12, 2003.

[7] Hsu W., Lee M.L., Zhang J., "Image Mining: Trends and Developments," *Journal of Intelligent Information Systems*, Vol. 19:1, pp. 7–23, 2002.

[8] Jensen, J.R., Cowen D.C., "Remote sensing of urban/suburban infrastructure and socio-economic attributes," *Photogrammetric Engineering and Remote Sensing*, Volume: 65, Issue: 5, pp. 611-622, 1999.

[9] Donnay, J.P., Barnsley M.J., Longley P.A., "Remote sensing and urban analysis," Donnay J.P., Barnsley M.J., Longley P.A., (Ed.) *Remote sensing and urban analysis*, pp. 3-18 Taylor and Francis, London, 2001.

[10] Carlson, Toby, "Applications of remote sensing to urban problems," *Remote Sensing of Environment*, Volume: 86, Issue: 3, pp. 273-274, 2003.

[11] Lippmann,R.P., "An introduction to computing with neural nets," *IEEE ASSP Mag*., 4, 4--22, 1987.

[12] Hsu, S.-Y., T. Masters, M. Olson, M. Tenorio, and T. Grogan, "Comparative analysis of five neural network models," *Remote Sens. Rev*., 6, 319--329, 1992.

[13] Bishop, C., *Neural Networks for Pattern Recognition*, Oxford Univ. Press, New York, 1995.

[14] Hornik K., Stinchcombe M., and White H., "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359-366, 1989.

[15] M. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Networks*, vol. 6, pp. 525-533, 1993.

[16] www-ra.informatik.uni-tuebingen.de/SNNS.

[17] Wilkinson G. G., "Results and implications of a study of fifteen years of satellite image classification experiments," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 43, Issue 3, pp. 433-440, 2005.

[18] Dawson M. S., "Applications of electromagnetic scattering models to parameter retrieval and classification," in A. Fung Microwave Scattering and Emission Models and their Application, Artech-House, 1994.

Fabio Del Frate (M'03) received the Laurea degree in Electronic Engineering in 1992 and the Ph.D. degree in Computer Science in 1997, both from the Tor Vergata University, Rome, Italy. From September 1995 to June 1996, he was a Visiting Scientist with the Research Laboratory of Electronics at the Massachusetts Institute of Technology, Cambridge, MA. In 1998 and 1999, he was with the European Space Agency (ESA), ESRIN establishment, Frascati, Italy, as a Research Fellow and was engaged in projects concerning end-to-end remote sensing applications.

He is currently a Research Professor of the Tor Vergata University, Rome, Italy where he teaches courses of electromagnetics and neural networks. He has acted and is PI in several remote sensing projects supported by ESA. He his author or co-authors of more then 100 scientific publications, with a special focus on the applications of neural networks to remote sensing inversion problems. His main research topics also include retrieval and classification algorithms for land cover from satellite data, oil spill detection in SAR imagery, retrieval of atmospheric variables with microwave radiometry, data-exploitation for the new missions PROBA and OMI.

Dr. Del Frate has served as a reviewer for different Remote Sensing Journals and as Associate Editor for GEOSCIENCE AND REMOTE SENSING LETTERS.

Fabio Pacifici (S'02) was born in Rome, Italy, in 1980. He received the Laurea (B.S.) and the Laurea Specialistica (M.S.) degree in Telecommunication Engineering (both summa cum laude) from the Tor Vergata University, Rome, Italy, in 2003 and 2006, respectively.

He is pursuing the Ph.D. degree in Geoinformation at the Earth Observation Laboratory (EOLab) of the Tor Vergata University and collaborates with the Aerospace Department of the University of Colorado at Boulder (Boulder, CO, USA). His main research activity is in the area of remote sensing image processing. In particular, his interests are related to the classification and change detection task of urban areas by using statistical approaches such as Neaural Networks (NNs) and Support Vector Machines (SVMs) in very high spatial resolution optical imagery and SAR imagery. He is currently involved in remote sensing projects supported by the European Space Agency (ESA). In 2005 he has been Visitor Student at Colorado Center for Astrodynamics Research (CCAR) at the University of Colorado.

Mr. Pacifici is a member of the Italian Association for Remote Sensing (AIT) and he is a reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

Giovanni Schiavon received the Laurea degree cum laude in Electronic Engineering from the University of Rome La Sapienza in 1982.

He is currently an associate professor at the University of Rome Tor Vergata, where he has been a researcher since 1984 to 2000. He has been teaching a course on remote sensing since 1996 and a course on electromagnetic fields since 2000. His research activity has mainly been concerned with remote sensing, of the atmosphere and of the Earth surface, propagation, and millimeter waves. He acted as reviewer for the international journals Radio Science, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS), IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (GRSL), Natural Hazards and for several volumes of VSP publisher. He managed several work packages of different contracts with the European Space Agency (ESA).

Prof. Schiavon has been involved in the international remote sensing projects AGRISAR ('86), AGRISCATT ('87, '88), MAESTRO-1 ('89), MAC Europe ('91), SIR-C/X-SAR ('94) and in the ERA-ORA program (European Commission Fourth Framework Programme) on Earth observation for environmental monitoring. He has authored or co-authored more than 150 scientific papers, the most of which on international journals or proceedings.

Chiara Solimini received the laurea (M.S.) degree in environmental engineering from the Tor Vergata University of Rome, Italy, in 2002. She is currently working on the Ph.D. degree in GeoInformation at the Earth Observation Laboratory in the Department of Information, Systems and Productions, Tor Vergata University, Rome.

Her research is focused on classification and change detection of urban areas using neural network, transform algorithms and high resolution QuickBird images. From January 2005 she started a collaboration with the Aerospace Department of the University of Colorado at Boulder (Boulder, CO, U.S.A.) and with Company DigitalGlobe (Longmont, CO, U.S.A.), aimed at high resolution urban monitoring.

An Innovative Neural-Net Method to Detect Temporal Changes in High-Resolution Optical Satellite Imagery

Fabio Pacifici[1], Fabio Del Frate[1]
Chiara Solimini[1] and William J. Emery[2]

1. Earth Observation Laboratory, DISP, Tor Vergata University, Rome, Italy
2. Colorado Center for Astrodynamics Research, University of Colorado at Boulder, Colorado, U.S.A.

ABSTRACT

The advent of new high spatial resolution optical satellite imagery has greatly increased our ability to monitor land cover changes from space. Satellite observations are carried out regularly and continuously and provide a great deal of insight into the temporal changes of land cover use. High spatial resolution imagery better resolves the details of these changes and makes it possible to overcome the "mixed-pixel" problem inherent with more moderate resolution satellite sensors. At the same time, high-resolution imagery presents a new challenge over other satellite systems, in that a relatively large amount of data must be analyzed and corrected for registration and classification errors to identify the land cover changes related to land cover changes. To obtain the accuracies required by many applications to large areas, very extensive manual work is commonly required to remove the classification errors introduced by most methods. To improve on this situation we have developed a new method for land surface change detection that greatly reduces the human effort needed to remove the errors that occur with many classification methods applied to high-resolution imagery. This change detection algorithm is based on Neural Networks and it is able to exploit in parallel both the multi-band and the multi-temporal data to discriminate between real changes and false alarms. In general the classification errors are reduced by a factor of 2-3 using our new method over a simple Post Classification Comparison based on a neural network classification of the same images.

I. Introduction

Changes in land cover and land use in urban areas are dynamic processes. The transitions associated with these changes, occur at varying rates and in different locations within the constraints of, or in response to, various social, economic and environmental factors. The rapid expansion of urban centers and their peripheries has led, in many cases, to a series of complex problems related to uncontrolled urban sprawl, increased traffic congestion, degradation of air and water quality, loss of agricultural land and natural vegetation. Urban growth has also had significant impacts on the social structure of cities and their surroundings, in terms of population

distribution and land use characteristics. For many public and private institutions, knowledge of these changes in either natural resources or man-made structures is a valuable source of information for decision-making processes [1]. However, updating maps on the status and trends of urban ecosystems is an intensive task requiring timely and accurate information from multiple sources of data, especially for detailed mapping of complex urban scenes [2][3]. One of the primary methods for updating land cover and land use maps has been, and in some case still is, through human observation and interpretation. In this process, the full range of human interpretation capabilities can be employed, including the interpreter's own knowledge of the area. However, this technique is time consuming, subject to errors of omission and is subject to the interpreter's abilities. In particular, this approach cannot be standardized, and, therefore lacks uniform outputs.

In this context, satellite and airborne remote sensors have proved particularly useful in addressing computer-assisted change detection and related environmental monitoring, agricultural surveys and urban studies [4][5]. Recently, the commercial availability of very high-resolution satellites with onboard sensors characterized by resolutions from 0.60 to 2.8 m, has opened a wide range of new opportunities to use Earth observing satellite data to monitor and map urban change. The enhanced spatial resolution of these systems, which reduces the occurrence of mixed pixels, provides more precise information on land use and land cover changes and are being increasingly used to carry out detailed characterization of the trends in urban areas. In fact, the detection of fine-scale physical changes in individual objects, such as single buildings, houses or roads, is greatly enabled by these systems.

The category of new high-resolution satellite sensors includes QuickBird (QB), operated by DigitalGlobe that satisfies well the requirements related to studies of urban environment change, where resolution plays a key role and spectral information may be crucial. Launched on October 18, 2001, QB collects both multi-spectral and panchromatic imagery concurrently. The panchromatic resolution is 0.61 m at nadir to 0.72 m at 25° off-nadir and the multi-spectral is from 2.44 m at nadir to 2.88 m at 25° off-nadir. The panchromatic bandwidth spans from 0.45 μm to 0.90 μm, while the multi-spectral images are acquired in four channels: blue (0.45-0.52 μm), green (0.52-0.60 μm), red (0.63-0.69 μm), NIR (0.76-0.90 μm) [6].

Although very high-resolution data have a great potential for monitoring and mapping surface changes, several issues need to be considered if observing surface changes using multi-temporal satellite image data. The crucial steps include the registration of two or more high-resolution scenes, the presence of temporary objects (such as cars or trucks), the effects of changing solar elevation, atmospheric conditions and satellite sensor incidence angles. Sensor

noise, misregistration errors, seasonal and meteorological effects also reduce the achievable accuracy in change detection output, unless time-consuming and tiresome manual procedures are applied to identify and remove the errors which appear as false surface change signals. The need for an extensive manual correction, and the often non-uniform outcomes, depending on the image analyst, can easily erase the advantage of an automated classification and change detection technique. The need for a uniform image analysis outcome and the reduction of data processing time requires computer routines that automatically perform the image classification and map the temporal changes with high accuracy.

In the remote sensing literature, many papers have addressed the development of change detection techniques for moderate-resolution images such as those provided by the Landsat satellite series. Two main approaches have been proposed: the unsupervised and the supervised. The former performs change detection by transforming two separate multi-spectral images into a single band or multi-band image in which the areas of land cover or land-use change can be detected. The latter, exploits supervised classification methods, which require the selection of a suitable training set (known as a Region Of Interest or ROI) to determine the classes. This approach is in general more flexible and effective than that one based on the unsupervised comparison of multi-temporal raw image data [7].

Many unsupervised techniques perform change detection algorithms by using simple procedures to extract the final change map, e.g., by subtracting, on a pixel basis, the images acquired at two different times. More sophisticated techniques analyzing the difference image using a *Markov Random Field* (MRF) approach exploit the interpixel class dependency in the spatial domain by considering the spatial contextual information included in the neighborhood of each pixel [8]. In [9] the proposed method combines the use of a MRF and a *Maximum a Posteriori Probability* (MAP) decision criterion in order to search for an optimal image of changes. Among the supervised techniques, the most common category is the *Post Classification Comparison* (PCC) [10][11]. This technique performs change detection by comparing the classification maps obtained by classifying independently two remote sensing images of the same area acquired at different times. In this way, it is possible to detect changes and to understand the kinds of transitions that have occurred. Furthermore, the classification of multi-temporal images avoids the need to normalize the image for atmospheric conditions sensor differences between the two acquisition dates. However, the performance of the PCC technique critically depends on the accuracies of the two classification maps. In particular, the final change detection map exhibits an accuracy close to the product of the classification accuracies given at the two times [12]. The main problem with this technique is that it does not exploit the dependence between two multi-

temporal images acquired from the same area. The *Direct Multidata Classification* (DMC) technique is able to overcome this problem [13]. Indeed, in this method the pixels are characterized by a vector obtained by stacking the feature vectors related to the images acquired at two times. Then, change detection is performed by considering each transition as a class and by training a classifier to recognize the transition typologies.

Neural Networks (NNs) have emerged as an important tool for the classification of remote sensing images. NNs are data driven and self-adaptive since they can adjust themselves to the data without any explicit functional specification of the underlying physical model. Moreover, they are universal functional approximations since they can approximate any function with a specified accuracy [14]. For these reasons, NNs offer a key for effectively managing the complexity of data, the potentialities of knowledge-based systems and the parametric manipulation of imagery.

In this paper we proposed a novel approach based on a NN architecture called NAHIRI (Neural Architecture for High-Resolution Imagery) to produce change detection maps from high-resolution satellite imagery. Although, NAHIRI represents one of the first methods for automating change detection in very-high-resolution satellite imagery, its extension to moderate resolution images, as shown in the paper, is straightforward and accurate. The distinctive feature and the major innovation of NAHIRI, is that the neural networks simultaneously exploit both the multi-spectral and the multi-temporal information associated with the changed values of the pixel spectral reflectances. This means that a single neural architecture is used with two different stages to perform the Change Detection analysis. The final stage includes the merging of the three different NNs results, and the production of the output map. Moreover, the proposed technique not only detects the different kinds of change that has occurred, but also explicitly recognizes the typologies of class transitions.

In section II of the paper we describe the data set used for this study, in section III we introduce the NN approach for classification problems and we present in detail the NAHIRI architecture. In section IV we discuss the results obtained over different types of images and locations. The accuracy of the final classification and change detection maps has been evaluated using by "*ground truth*" comparisons. The algorithm performance has also been evaluated against a PCC technique based on a NN architecture for the one-step multi-spectral image classification. We train both NAHIRI and this one-step NN to independently classify a number multi-temporal image pairs, and the classification output maps were subsequently compared. Some final conclusions and discussions regarding the accuracy improvements in the NAHIRI method are reported in Section V.

II. Data Set

Two test areas have been considered for this study. One is located in Rock Creek-Superior, Colorado, U.S.A., the other one, next to the campus of the Tor Vergata University, in Rome, Italy. Temporal changes in these areas have been mapped using the NAHIRI architecture applied to both QB and Landsat images. The Landsat image data set was provided by the Thematic Mapper, on board Landsat 5 with a spatial resolution of 30 m. A more detailed description of the test sites is given in the following two sections while Table 1 contains the main characteristics of the imagery used for each test area.

	Site Information		**Images Information**			
Code	**Location**	**Dimension (km^2)**	**Acquisition Date**	**Satellite**	**Spatial Res. (m)**	**Dimension (pixels)**
TEST AREA 1	Rock Creek-Superior, Colorado, U.S.A.	2	August 14, 2002	QuickBird	0.6	1300 x 800
			July 6, 2004	QuickBird	0.6	
TEST AREA 2	Rock Creek-Superior, Colorado, U.S.A.	280	July 5, 1992	Landsat	30	664 x 432
			August 17, 1996	Landsat	30	
			August 14, 2002	QuickBird	30	
TEST AREA 3	Tor Vergata Campus, Rome, Italy	3.7	May 29, 2002	QuickBird	2.8	700 x 600
			March 13, 2003	QuickBird	2.8	

Table 1: Detailed description of the test sites and the corresponding images.

Rock Creek-Superior Test Site

The area is immediately to the south of the city of Boulder, Colorado, along the Boulder-Denver corridor. The entire geographic region is shown in the Landsat image in Figure 1 where the Boulder and the Denver urban areas are located near to the upper-left and lower-right corners, respectively. To analyze the performance of NAHIRI on the full resolution QB images, given the large amount of time that would have been required to process the whole scene, we selected Test Area 1, shown as the green box (about 2 km^2) in Figure 1. The area includes houses, farms, and low-density commercial buildings. NAHIRI classification performance has been also tested in medium-resolution images. For this exercise we extracted a larger image area, which is the Test Area 2, highlighted in red in Figure 1. It has an area of 280 km^2 and is representative of the diverse land cover and land use types encountered in the region. This region includes zones that have transitioned mainly from open space to housing, retail shopping centers, hotels and new reservoirs. In fact, the area saw explosive growth especially between 1990 and 2000, but the growth has continued between 2002 and 2004 and continues today. As reported in Table 1, the entire set of images used for the site includes two Landsat images taken on July 5, 1992 and on August 17, 1996 and two pan sharpened QB images from August 14, 2002 and July 6, 2004.

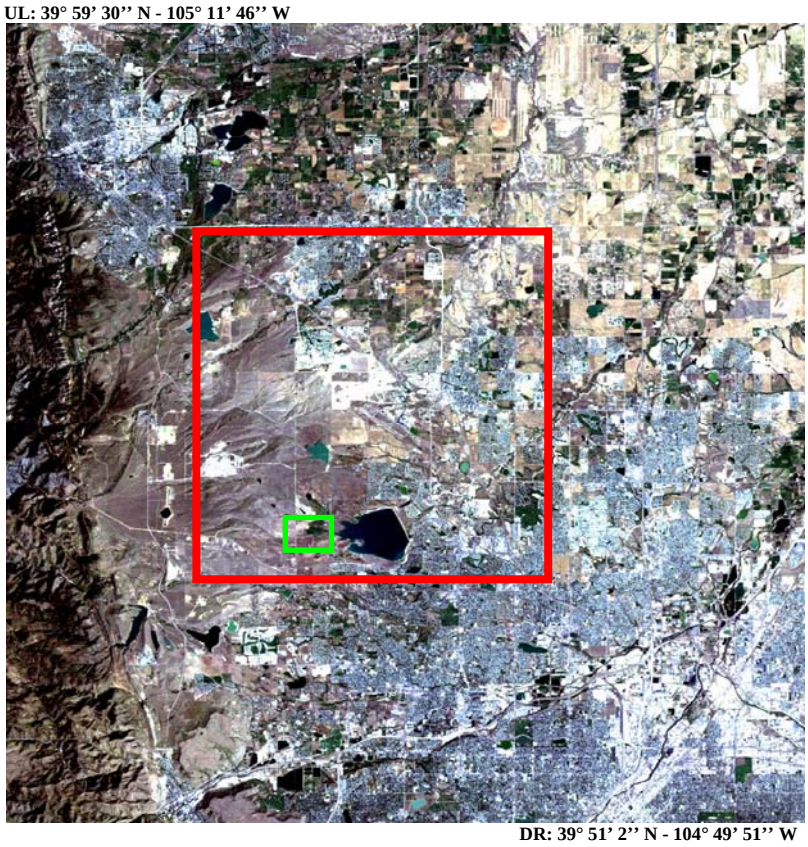


Figure 1: The Rock Creek study area is located northwest of Denver, CO, U.S.A.. The green box contains the Test Area 1 used on the high-resolution images, while the red box contains the Test Area 2 used with the medium-resolution classification.

Rome test site

The Rome test site is next to the campus of the Tor Vergata University, located in Italy, South-East of Rome. A view of the area generated using the Red, Green and Blue QB bands is shown in Figure 2. For validation purposes, we extracted a smaller study area (red box in Figure 2) from the image. With respect to the Boulder site, this area is more typical of a suburban area with residential, commercial and industrial buildings. This variety in land cover made it possible to evaluate the flexibility of our classification procedure when applied to different landscapes and spatial resolutions. Two multi-spectral QB images (2.4 m resolution) taken on March 13, 2003, and on May 29, 2002 have been analyzed for this site. The time interval between the acquisitions is less than one year in this case. Even in such a short period of time, the area underwent significant land surface changes such as the construction of a large shopping mall and several other buildings. The selection of a test area located next the Tor Vergata University campus made it easier to carry out *in situ* ground truth validation.

UL: 41° 51’ 35’’ N - 12° 35’ 35’’ E

DR: 41° 50’ 0’’ N - 12° 38’ 24’’ E

Figure 2: High-density residential area. The campus of the Tor Vergata University is located in the upper part of the image. The red box indicates the Test Area 3 in 2002.

III. NAHIRI Methodology

Image classification with Neural Networks

Different studies have shown the effectiveness of NNs algorithms for the classification of satellite imagery [15][16]. Among them Multi Layer Perceptron (MLP) has been found to have the best suited topology for classification and inversion problems [17]. MLP has also been considered in this study. These are feed-forward networks where the input flows only in one direction to the output and each neuron of a layer is connected to all neurons of the successive layer, but has no feedback to neurons in the previous layers. The "learning" of the NN can be carried out by feeding it with pairs of vectors (patterns). The input vector contains the reflectance information provided by the different channels of the multi-spectral image; the output vector contains the corresponding known classes of the surface. To avoid saturation within the network, the values of the input vector are normalized to the range between -1 and 1. This step also helps

to mitigate single-image effects if pixels belonging to different images are included in the training set. At the same time, the component of the output vector corresponding to the true class can be set to 1 while the others are set to 0. In the test phase, a competitive approach (*winner take all*) can be used to decide on the final classification.

For our classification phase, which is preliminary and instrumental to the change detection, we chose to work with reflectance ratios to reduce the effects of the different scene illuminations. Therefore, for each pixel of the QB multi-spectral images the vector $R_i = \left| B_{iB} / B_{iG},\ B_{iB} / B_{iR}, B_{iB} / B_{iNIR}, B_{iG} / B_{iR}, B_{iG} / B_{iNIR}, B_{iR} / B_{iNIR} \right|^T$, whose components are the ratios of the pixels spectral reflectances, has been computed. The number of elements of the vector R_i is $N(N-1)/2$, where N is the number of bands in the imagery used (e.g., $N = 4$ for QB), i represents the acquisition date and *R, G, B, NIR* denote the QB bands. Note that the same spectral information is carried out in the ratio vector R_i regardless of the choice of the order of the bands. The described vector data set forms new ratio images characterized by six spectral bands and the same spatial resolution of the multi-spectral QB images. We used the ratio images to train the three different neural networks making up the NAHIRI tool as demonstrated in Figure 3.

The various types of surfaces have been grouped into four classes:

- man-made surfaces, including buildings, concrete, asphalt, gravel and sites under construction;
- green vegetation;
- bare soil, including low density and dry vegetation, and unproductive surfaces;
- water (lakes, reservoirs, ponds, streams).

The number of classes was kept small to simplify the image classification and enable an easy interpretation of the quantitative change detection. Once the input and the output of the network are established, the number of neurons to be considered in the hidden layers is another critical choice. In fact, if the number of neurons is too small, the input-output associative capabilities of the net are too weak. On the other hand, this number should not be too large, otherwise, these capabilities might show a lack of generality being too narrowly tailored to the training set and the computational complexity of the algorithm would be increased in vain. It turns out that a compromise has to be found. We present the results of an extensive research on this topic in [18], with specific reference to the use of neural networks for the classification of QB and Landsat

images. Starting from these results, we selected a suitable number of hidden neurons for the optimal NNs topology.

NAHIRI Architecture for change detection

NAHIRI CD uses a parallel approach that includes three different Neural Networks. As shown in the flow chart in Figure 3, NN1 and NN2 generate the "*Change Map*" using the multi-spectral information, while NN3 produces a "*Change Mask*" exploiting multi-temporality.

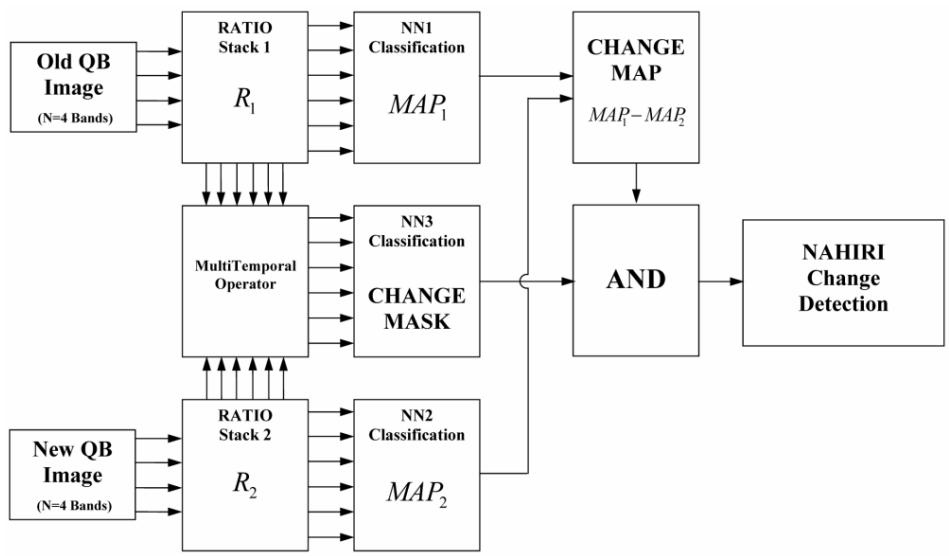


Figure 3: NAHIRI flow chart.

To produce the "*Change Map*" and the "*Change Mask*", the ratio images R_1 and R_2 are used as input to different Neural Networks. The "*Change Map*" stage consists of two steps:

1. nets NN1 and NN2, having the same topology, produce as outputs the classification maps MAP_1 and MAP_2;
2. the difference $Change\ Map = MAP_1 - MAP_2$ is computed.

For QB, the NN1 and NN2 topologies are characterized by six input neurons, corresponding to the six bands of the ratio images, and four output neurons, corresponding to the four chosen output classes. Two layers containing twelve neurons each compose the hidden-layers. The information given by the "*Change Map*" is the ($N_{cl}^2 - N_{cl} + 1$) possible transitions from one class to another between the two images where N_{cl} is the number of chosen classes (in our study N_{cl} = 4). We may associate a color to any possible transition. It is worthwhile to note that the *Change Map* is a PCC output performed by comparing MAP_1 with MAP_2.

As discussed before, the innovative aspect of NAHIRI is the introduction of another Neural Network, indicated in Figure 3 as NN3, which works in parallel with the previously described networks. This "*Change Mask*" stage includes:

1. a multi-temporal operator combining R_1 and R_2 outputs;
2. the net NN3 producing the "*Change Mask*".

The NN3 topology is composed of six inputs and two outputs. The k_{th} component of the input vector is given by $\left| Log \frac{R_1(k)}{R_2(k)} \right|$. The choice of the Log operator allows an extended dynamic range of the network input values, while the absolute value assures invariance with respect to the direction of the transition. The output values vary according to whether the individual pixel has changed or not. The NN3 output is then the map representing the changed pixels.

To discriminate between real changes and false alarms, we associate the logic value 1 with the changes and the logic value 0 with the no-changes for each pixel in both the "*Change Mask"* and the "*Change Map"*. These values are the input of the AND gate shown in Figure 3. Only if both inputs of the AND gate have the value 1, the variation in the "*Change Map"* is considered to be a valid change in the pixel class. The final output is an image having different colors associated with different class transitions and gray scale satellite image background denotes areas where no changes have taken place. This kind of presentation may help the end-user to understand where changes have occurred on the study area.

Training and test strategies

Training the nets by a statistically meaningful set of input-output training data is mandatory for the success of pixel-based classification. For each of the NNs of the NAHIRI architecture, suitable training sets that represent all of the different types of surfaces have been generated. In Table 2 we report details on the number of training patterns for these different cases. The cost function of the learning phase is minimized using a scaled conjugate gradient (SCG) algorithm [19].

	Test Area 1		Test Area 2			Test Area 3	
	2002	**2004**	**1992**	**1996**	**2002**	**2002**	**2003**
Classification							
Vegetation	1097	562	4661	2638	3455	2673	2125
Man Made	309	82	6782	6447	12250	22318	23283
Water	192	44	3921	2218	1690	-	-
Soil	308	398	23615	15231	25751	6945	15537
Change Mask							
Change	134		8162	6970		6967	
No Change	170		17520	21334		4711	

Table 2: Number of pixels that compose the training patterns for the different selected classes and Test Areas.

The change detection accuracy has been evaluated using both confusion matrices and the *k*-coefficient, which is recognized an effective parameter to test classification performance. The *k*-coefficient is defined as [20]:

$$\hat{K} = \frac{N\sum_{i=1}^{r} X_{ii} - \sum_{i=1}^{r}\left(X_{i+} \times X_{+i}\right)}{N^2 - \sum_{i=1}^{r}\left(X_{i+} \times X_{+i}\right)} \qquad \textbf{(1)}$$

where:

r	=	number of rows in the confusion matrix;
X_{ii}	=	number of observations in row *i* and in column *i*;
X_{i+}	=	marginal total of row *i*;
X_{+i}	=	marginal total of column *i*;
N	=	total number of samples included in the matrix.

It is worth noting that NAHIRI aims to explicitly identify, on a pixel-basis, the kinds of land-cover transitions that have occurred in an area. This means that each possible pixel transition (e.g., from the class bare soil to the class water) represents a class in the confusion matrix.

For the application to full resolution QB imagery over the Test Area 1, the accuracies result from a comparison with ground-truth maps produced using photo-interpretation techniques. Since these methods are performed manually, their extension to the other test cases would have involved too much manual image analysis time. Therefore, for Test Areas 2 and 3, we decided to use a different validation technique. A set of control points, representative of each different transition class, has been randomly selected from the "*Change Maps*" for the two different test areas. Thus, all the transitions were sampled with a representative number of points. This validation method ensures a sufficiently accurate characterization of each transition class since

the land cover changes are sampled proportionally to their occurrence in an area. During the validation phase, the corresponding ground-truth has been computed for each selected ground control point. The ground truth has been obtained by careful visual inspection of the multi-temporal images and by using ancillary data (e.g., aerial pictures, government maps, etc.) when available.

IV. Results

The classification of high-resolution images of urban environments represents one of the most challenging tasks in remote sensing data analysis and processing. In general, classification errors due to the high spectral diversity of single surface materials, which may include a wide range of roof types, roads and other artificial surfaces, occur frequently. As previously discussed, we consider four different classes, essentially to distinguish between man-made and natural surfaces. In the case of multi-spectral QB data we used a 6-12-12-4 NN topology, while for Landsat imagery we considered a 15-12-12-4 NN topology.

NAHIRI CD in very high-resolution images

To assess the performance of the NAHIRI algorithm in detecting surface changes in a low-density residential area, we started by processing the full-resolution QB images of Test Area 1 represented in Figures 4 and 5, where considerable changes occurred between 2002 and 2004.

Figure 4: Test Area 1 imaged in 2002 (see Figure 5).

Figure 5: Test Area 1 as in Figure 4, imaged in 2004.

To evaluate the NAHIRI accuracy, the results have been validated using the "ground truth" shown in Figure 6, which was obtained by careful visual comparison of the 2002 and 2004 images of the study area. Cars, trucks and buses were removed from the images, since they have been treated as "temporary" (changes with time scales much shorter than the time difference between images) objects.

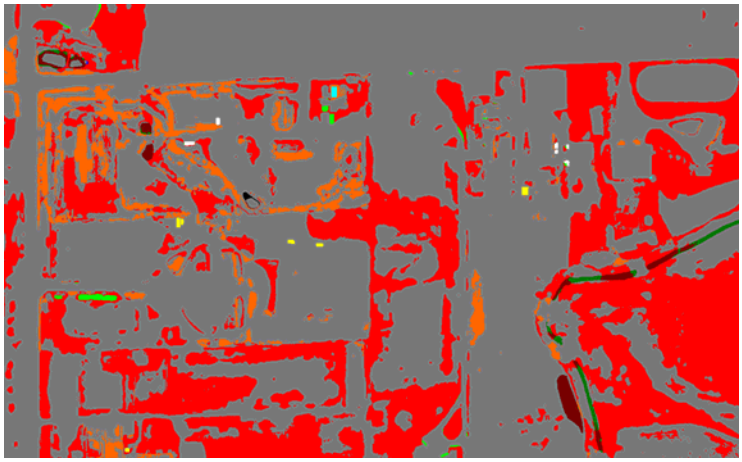


Figure 6: Manually obtained "true" change of the Test Area 1 in Figure 4 and 5. The red pixels represent change from soil to vegetation, indicating more abundant seasonal vegetation in 2004. The brown and dark green colors correspond to water not present in 2002 due to drought. The color code is Table 3.

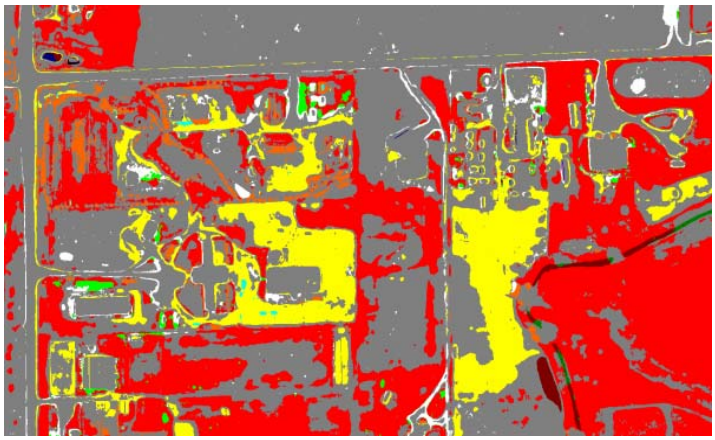
	2004			
2002	**Vegetation**	**Man-made**	**Water**	**Soil**
Vegetation	**Gray**	**Cyan**	**Dark Green**	**Orange**
Man-made	**Green**	**Gray**	**Dark Blue**	**White**
Water	**Blue**	**Magenta**	**Gray**	**Black**
Soil	**Red**	**Yellow**	**Brown**	**Gray**

Table 3: Changes Color Table

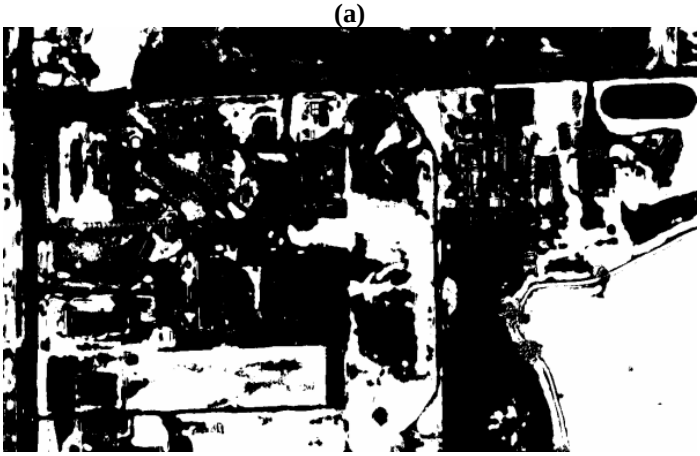
One of the most important environmental conditions that must be considered in the 2002 Colorado imagery is the unusually long dry period preceding the image acquisition. This drought resulted in an abundance of pixels indicating a change from bare soil in 2002 to seasonal vegetation in 2004, as represented in Figure 6 in red. For the same reason, water surfaces that were not present in 2002 show up in 2004. Figure 7 (a) computed using the PCC method displays a lot of the yellow soil to man-made transition class that is not present in Figure 6. These represent classification errors, which are not present in the NAHIRI classification map in Figure 7 (c). In Figure 7 (b) we present the "*Change Mask*" for this image.

A quantitative assessment of the resulting accuracy of the standard PCC method (Figure 7 (a)) is provided in Table 4, which is the confusion matrix of the change-map stand-alone product. The overall error exceeds 23% (k=0.568); hence, although processing is very fast (about 20 s to obtain this output, with a 3 GHz CPU and 1 GB RAM), the information obtained seems to have rather large errors. During the validation phase, considerable time is needed, (i.e. a few days for Test Area 1), for a manual correction of pixel errors in the final PCC map. Nevertheless, the validation processing time strictly depends on the desired classification accuracy. The automatic generation of the NAHIRI output image took extra processing time due to the training phase of NN3. In fact, both methodologies require the training of the two nets NN1 and NN2. Each neural net exploits the classification of one multi-spectral image, but only the NAHIRI procedure involves the training of NN3 to simultaneously exploit the multi-temporal character of the image pair. It is important to note that, the overall error is dramatically decreased from 23% (k=0.568), for the PCC method, to about 12% (k=0.722), for NAHIRI. The quantitative results are shown in the confusion matrix in Table 5.

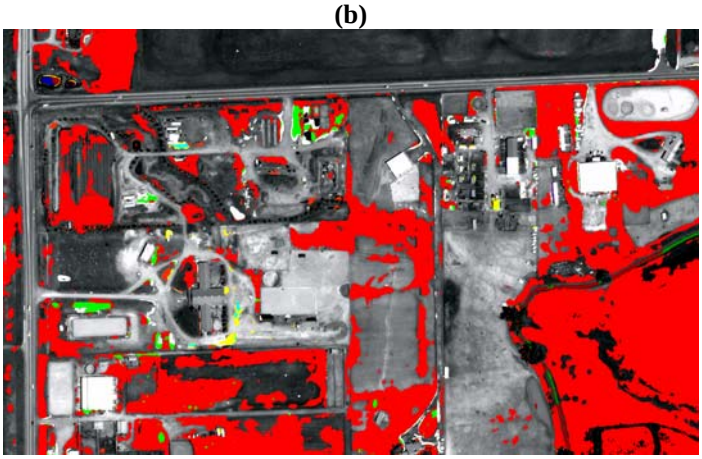
By comparing the image in Figure 7 (a) with that in (c), we note that the false-alarm pixels (mainly in yellow) in the middle of the PCC image have been filtered out by the NAHIRI algorithm. We point out that the NAHIRI method is not used to improve or correct the classification process, rather to create a mask that can filter the classification errors present in the PCC images. For a better explanation, we can look, for example, at the no-change row (gray) of Table 4 and compare it with the same color in Table 5. The values in the exclusion error column generally increase, but those relative to this row show a significant reduction. The migration of pixels towards the no-change row is due to the filtering algorithm that cuts off the areas where changes have not occurred.



(a)



(b)



(c)

Figure 7: The map (a) represents the output of a PCC Change Detection between 2002 and 2004. Several misclassification errors displayed in yellow (soil to man-made surface) are present. The map (b) is the output of the intermediate stage "*Change Mask*". The white and black indicate the changes and no-changes, respectively. The NAHIRI output of the Test Area 1 is shown in (c): in the background, the gray level image indicates unchanged features; colors (Table 3) denote transitions between classes.

Post Classification Comparison															
Ground Truth	**Green**	**Red**	**Blue**	**Yellow**	**Magenta**	**Cyan**	**Dark Green**	**Brown**	**Dark Blue**	**Orange**	**White**	**Black**	**Gray**	**Total pixel**	**Excl. Error (%)**
Green	17269	621	0	1490	0	566	0	0	0	0	68	0	17217	19962	53.6
Red	5700	262148	0	43	0	0	0	0	0	0	77	0	1085	6905	2.6
Blue	0	11	0	0	0	0	0	0	0	0	0	0	8	19	100
Yellow	0	16	0	322	0	0	0	0	0	0	0	0	198	214	39.9
Magenta	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-
Cyan	136	12	0	15	0	55	0	0	0	0	1	0	11	175	76.1
Dark Green	92	0	0	0	0	0	1320	943	0	0	0	2	238	1275	49.1
Brown	5	16	0	68	1	0	7	4214	0	0	2	8	1472	1579	27.2
Dark Blue	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-
Orange	1865	42	0	556	0	359	0	0	0	0	11	0	32864	35697	100
White	50	7	0	39	0	1	0	0	0	0	82	0	211	308	79
Black	0	0	0	0	0	0	0	0	0	0	140	0	20	160	100
Gray	3721	47495	0	108332	59	478	16	218	473	0	23295	108	539798	184195	25.4
Total pixel	11569	48220	0	110543	60	1404	23	1161	473	0	23558	118	593122	250489	
Incl. Error (%)	40.1	15.5	-	99.7	100	96.2	1.7	21.6	100	-	99.7	100	9		
Overall error: 23.3%								**Kappa Coefficient: 0.568**							

Table 4: Confusion Matrix of the standard PCC for Test Area 1.

NAHIRI CD															
Ground Truth	**Green**	**Red**	**Blue**	**Yellow**	**Magenta**	**Cyan**	**Dark Green**	**Brown**	**Dark Blue**	**Orange**	**White**	**Black**	**Gray**	**Total pixel**	**Excl. Error (%)**
Green	913	386	0	0	0	0	0	0	0	0	33	0	202	621	40.5
Red	4713	235620	0	1	0	0	0	0	0	0	63	0	28656	33433	12.4
Blue	0	6	0	0	0	0	0	0	0	0	0	0	13	19	100
Yellow	0	14	0	233	0	0	0	0	0	0	0	0	289	303	56.6
Magenta	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-
Cyan	133	6	0	3	0	55	0	0	0	0	0	0	33	175	76.1
Dark Green	8	0	0	0	0	0	862	708	0	0	0	2	1015	1733	66.8
Brown	2	6	0	62	1	0	4	1540	0	0	2	2	4174	4253	73.4
Dark Blue	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-
Orange	1865	42	0	556	0	359	0	0	0	0	11	0	32864	35697	100
White	12	3	0	0	0	0	0	0	0	0	45	0	330	345	88.5
Black	0	0	0	0	0	0	0	0	0	0	121	0	39	160	100
Gray	943	35605	0	1157	56	160	16	218	394	0	8344	29	677071	46922	6.5
Total pixel	7676	36068	0	1779	57	519	20	926	394	0	8586	33	677071	123661	
Incl. Error (%)	89.4	13.3	-	88.4	100	90.4	2.3	37.6	100	-	99.5	100	9.1		
Overall error: 11.9%								**Kappa Coefficient: 0.722**							

Table 5: Confusion Matrix for NAHIRI CD for Test Area 1.

For this first demonstration area, the overall error is less than 12% (k=0.722), which is a substantial decrease over the PCC method. NAHIRI shows a dramatic decrease in analysis time: one hour for NAHIRI against a few days (required to manually correct "false-alarms") to achieve the same change detection accuracy. Thus, on one hand, the NAHIRI training phase is slower than the PCC method, on the other hand, the NAHIRI output shows higher classification accuracy than the PCC final map. In particular, if we use the PCC output, to obtain a change detection accuracy equal to the NAHIRI output, intensive manual correction is required.

NAHIRI CD: extension to lower resolution images

In images with lower resolutions like Landsat misclassification errors are frequently related to the mixed pixel problem, especially in high-density residential areas, where the mixed pixels can contain portions of buildings, vegetated areas and roads. These errors give rise to a variety of problems when attempting to map change detection. First, they occur in widely spread single pixels instead of groups of adjacent pixels, which are easier to detect and correct. Moreover, a pixel including different kinds of surfaces may be randomly attributed to any one of the above-mentioned surface types leading to a noisy change detection pattern difficult to interpret and correct.

To assess the performance of NAHIRI when processing lower resolution data two different exercises have been carried out. In one case we applied NAHIRI to a pair of Landsat images taken over the Colorado area of interest in the years 1992 and 1996. In the other case, we assume that the user has one Landsat and one QB image available over the area of interest. In this case, NAHIRI was applied after degrading the spatial resolution of the QB image to approximately 30 m corresponding to the earlier Landsat TM resolution. This has been accomplished by computing a 51x51 pixel mean of each QB spectral band, thus reducing the resolution to about 30 m.

In Figure 8 we show Test Area 2 in (a) 1992, (b) 1996 and (c) 2002. The Change Detection products are in (d), (e) and (f) for changes from 1992-1996, and in (g), (h) and (i) for change from 1996-2002. The maps (d) and (g) represent the output of a standard PCC Change Detection, for 1992-1996 and 1996-2002. The maps (e) and (h) represent the "*Change Mask*" stage of the NAHIRI algorithm. The NAHIRI output of Test Area 2 is shown in (f) and (i), for 1992-1996 and 1996-2002 respectively. The background of these latter figures has a satellite image gray scale indicating areas where features are unchanged. As previously shown in Table 3,

the different colors denote different class transitions, which we repeat here in Table 6 for convenience in interpreting Figure 8.

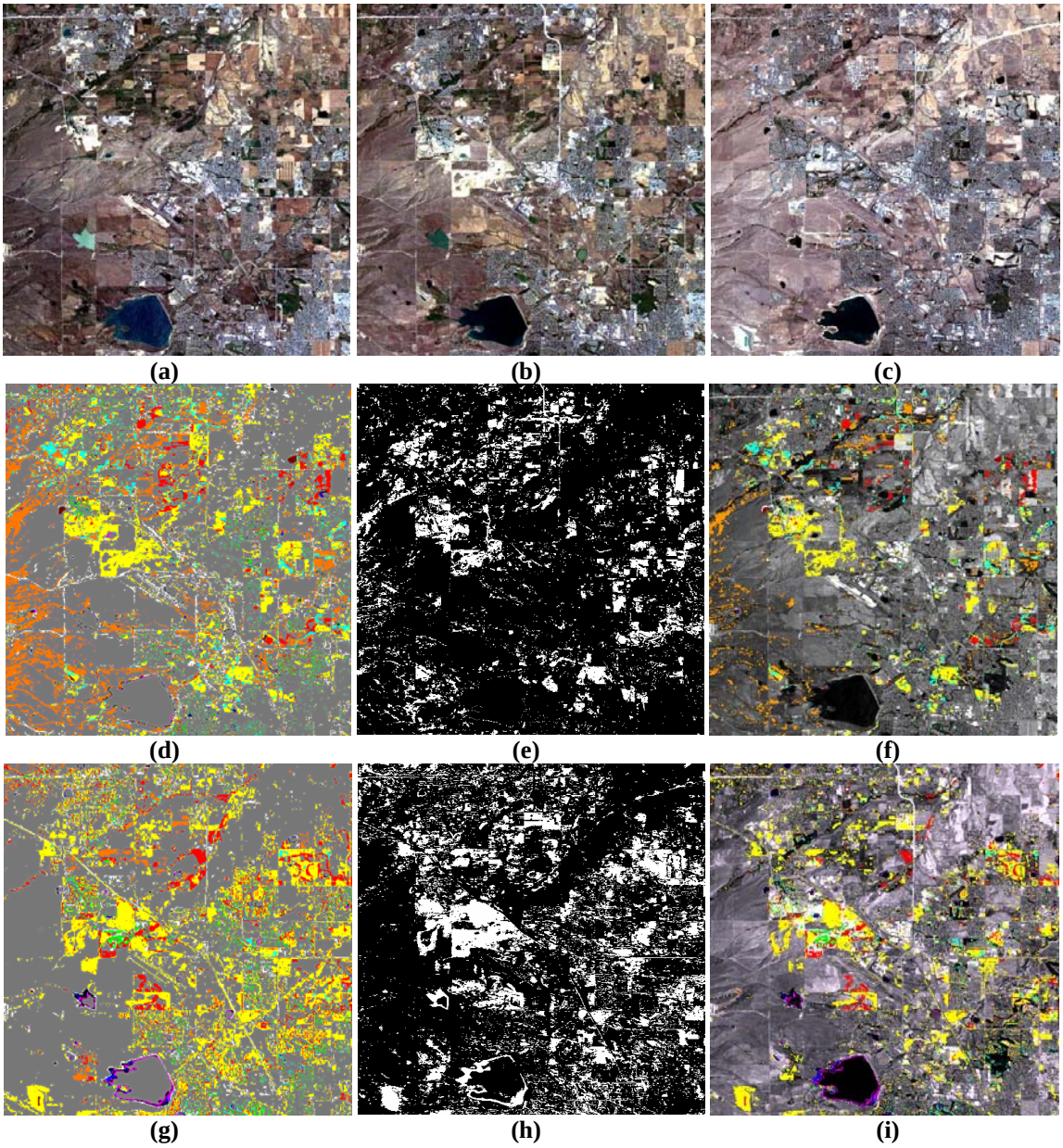


Figure 8: Test Area 2 in (a) 1992, (b) 1996 and (c) 2002. In (d), (e) and (f) is shown the Change Detection elaboration between 1992 and 1996, while (g), (h) and (i) the one between 1996 and 2002. The maps (d) and (g) represent the output of a PCC Change Detection, respectively of 1992-1996 and 1996-2002. Several misclassification errors displayed in yellow (soil to man-made surface) are present. The maps (e) and (h) are the outputs of the intermediate stage of the algorithm "*Change Mask*". The white and black indicate the changes and no-changes, respectively. The NAHIRI output of the Test Area 2 is shown in (f) and (i), respectively of 1992-1996 and 1996-2002: in the background, the gray level image indicates unchanged features; colors (Table 6) denote transitions between classes.

	AFTER			
BEFORE	**Vegetation**	**Man-made**	**Water**	**Soil**
Vegetation	**Gray**	**Cyan**	**Dark Green**	**Orange**
Man-made	**Green**	**Gray**	**Dark Blue**	**White**
Water	**Blue**	**Magenta**	**Gray**	**Black**
Soil	**Red**	**Yellow**	**Brown**	**Gray**

Table 6: Changes Color Table for Figure 8

The maps (d) and (g) in Figure 8 represent the output of a traditional PCC Change Detection, for 1992-1996 and 1996-2002 respectively. We note several false alarms due to misclassification errors, mostly displayed in yellow, which correspond to the transition from soil to man-made surface as shown in Table 6. Thus, comparing the NAHIRI change detection outputs displayed in Figure 8 (f) and (i) with the PCC results shown in Figure 8 (d) and (g), we can easily identify several erroneous pixels resulting from the PCC method.

Again, the characteristic filtering effect of NAHIRI emerges and allows to reduce the classification error as shown clearly in Figures 9 and 10, which are small portions of the previous images in Figure 8, corresponding to 1992-1996 and 1996-2002 changes respectively. In particular, in Figures 9,10(a) we show the earlier scene while in Figures 9,10(b) we present the later images represented in False Color (Bands 4,3,1) in order to better distinguish between vegetation, soil and man-made surfaces. Moreover, in Figure 9,10(b), the contours of the changed areas, determined by visual inspection, are highlighted in yellow. In Figures 9,10(c) and 9,10(d) we display the standard PPC and NAHIRI outputs. For this study, we replaced the satellite background representing the areas with “no change” with the color gray, to enable an easy comparison with the PPC output. The filtering improvement of NAHIRI can be clearly seen for the two different change detection periods in Figures 9 and 10 by comparing the respective output images (c, d) with the ground-truth (b) in both figures. In particular, in the upper left corner of Figure 9 (c) and the lower part of Figure 10 (c) many false alarms were detected by the PCC method.

The accuracy assessment of the PCC and NAHIRI methods have been evaluated relatively for the 1992-1996 Change Detection. The confusion matrix, not reported here, indicates that the overall error ranges from 18% (k=0.619), for PCC output, to 5% (k=0.881), in the case of NAHIRI. Similarly, the results for the 1996-2002 Change Detection demonstrate how this error decreases from 27% (k=0.486) to 10% (k=0.759).

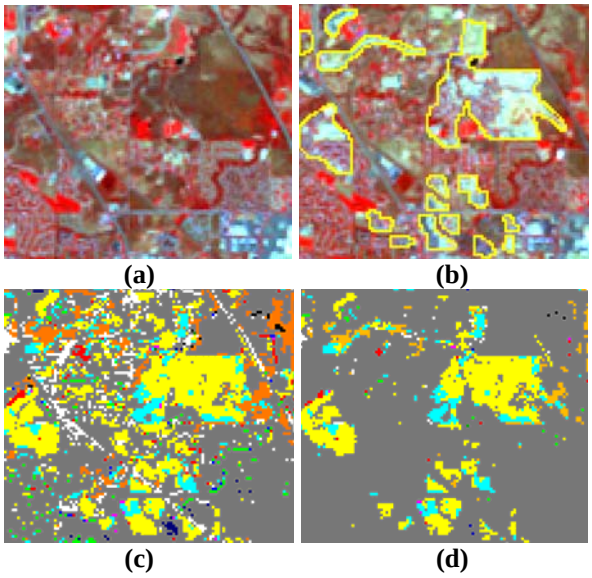


Figure 9: Detail of the Test Area 2 - False Color of (a) 1992 and (b) 1996. In (b), the regions highlighted in yellow are changes, according with ground-truth validation. The Post Classification Comparison is shown in (c), while the NAHIRI CD is in (d).

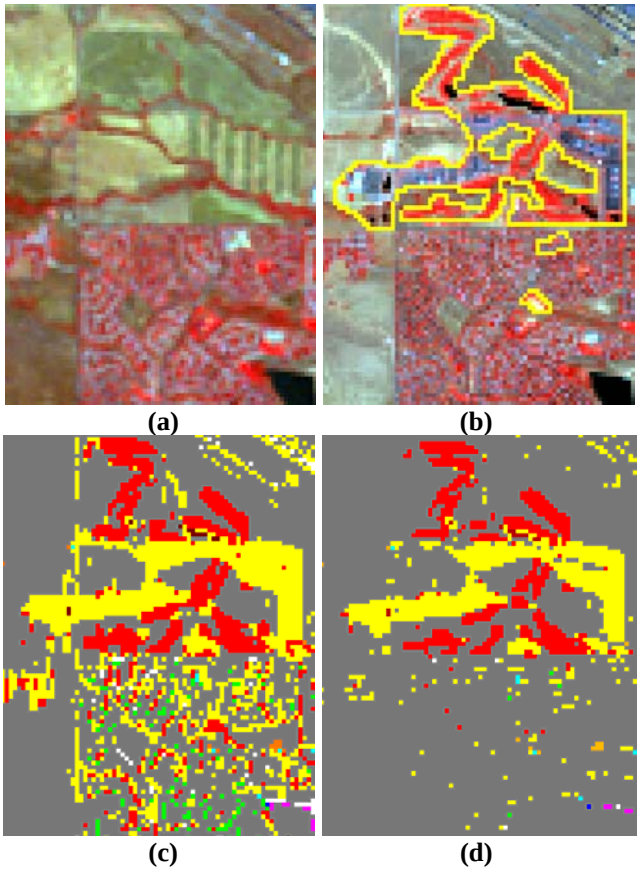


Figure 10: Detail of the Test Area 2 - False Color of (a) 1996 and (b) 2002. In (b), the regions highlighted in yellow are changes according with ground-truth validation. The Post Classification Comparison is shown in (c), while the NAHIRI CD is in (d).

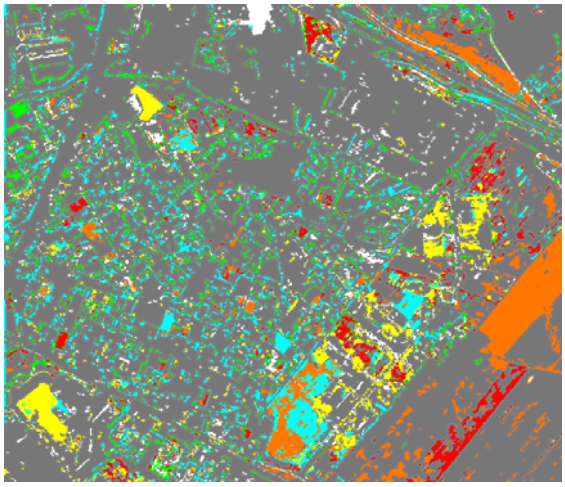
NAHIRI CD: application to different landscapes

To assess the performance of the NAHIRI algorithm in detecting surface changes in a high-density residential area, i.e., the one next to the Campus of the Tor Vergata University, we processed Test Area 3 represented in Figures 11 and 12, where considerable changes occurred between 2002 and 2003. It is worthwhile to note that, in this case, we used the 2.8 m resolution QuickBird imagery to test the ability of the NAHIRI technique to process images with different spatial resolutions.

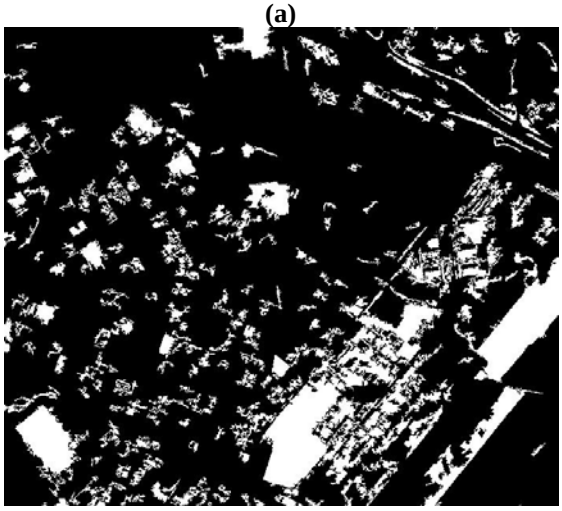
Figure 11: Test Area 3 imaged in 2002

Figure 12: Test Area 3 imaged in 2003

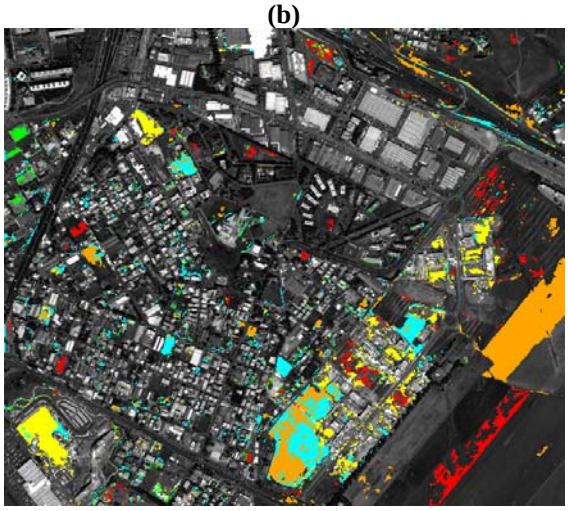
Figure 13(a) displays the change map obtained by the standard PCC method, while in Figures 13(b) and 13(c) we present the "*Change Mask*" and the NAHIRI CD output with the 2002 gray level satellite image as the background. For this test site, we have chosen only three classes: manmade, soil and vegetated area (the class water is not present in the selected study area). The class transitions color table is given in Table 7.



(a)



(b)



(c)

Figure 13: The map (a) represents the output of a PCC Change Detection between 2002 and 2003. Several misclassification errors displayed in cyan (vegetation to man-made surface) are present. The map (b) is the output of the intermediate stage "*Change Mask*". The white and black indicate, respectively, changes and no-changes. The NAHIRI output of the Test Area 3 is shown in (c): in the background, the gray level image indicates unchanged features; colors (Table 11) denote transitions between classes.

	2003		
2002	**Vegetation**	**Man-made**	**Soil**
Vegetation	**Gray**	**Cyan**	**Orange**
Man-made	**Green**	**Gray**	**White**
Soil	**Red**	**Yellow**	**Gray**

Table 7: Change Color Table for Figures 13

We observe in Figure 13(a) several false alarms within the high-density residential area. These pixels are mainly cyan and green, which indicates the transition from vegetation to man-made surface and from man-made surface to vegetation, respectively. By comparing this image with Figure 13(c), it is easy to see the filtering effect of the NAHIRI procedure in the reduction of these erroneous pixels, particularly in the center of the image.

A quantitative accuracy assessment of the PCC method indicates that the overall error exceeds 22% (k=0.444). As before, the automatic generation of the NAHIRI output requires a bit more execution time due to the training of NN3, but the overall error is dramatically decreased to 5.5% (k=0.783). To achieve this level of accuracy with the traditional PCC method would require a lot of intensive manual error removal.

V. Discussions and Conclusions

We have proposed an image analysis technique based on NN architecture, called NAHIRI (Neural Architecture for High-Resolution Imagery), to detect land cover changes. The distinctive feature and the major innovation of the approach is that the 3 NNs simultaneously exploit both the multi-spectral and the multi-temporal information associated with changed values of high-resolution satellite image pixel spectral reflectances. For this purpose, a NN architecture with different stages is designed. The first stage includes the generation of ratio images to reduce the effects of different scene illuminations and atmospheric and ground moisture conditions. From these ratio images, the inputs of the three neural nets, NN1, NN2 and NN3, are derived. Each input is characterized by a feature-vector whose components are the non-redundant ratios of the spectral bands. To train NN1 and NN2, we used as input the ratio vector data set exploiting the multi-spectral information; to train NN3, we used the logarithm of the ratios exploiting the multi-temporal information.

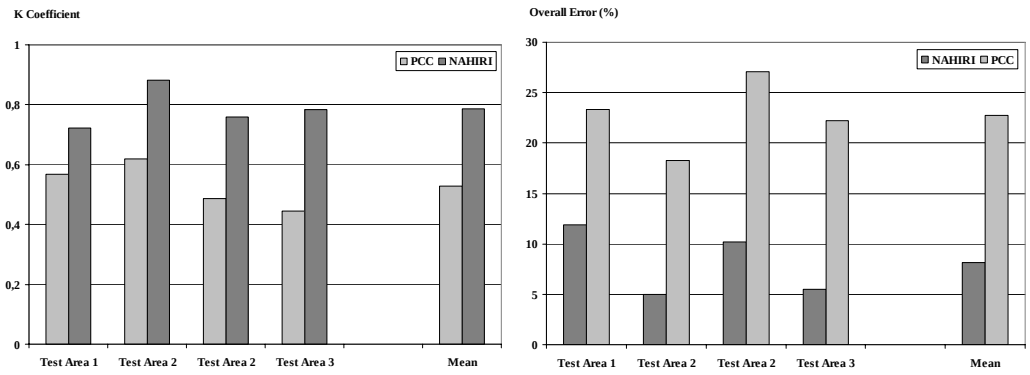
Thus, NN1 and NN2 are designed to produce the classification maps MAP_1 and MAP_2, while NN3 is intended to classify the changed and unchanged features. The second stage includes the extraction of two change detection maps: *"Change Map"*, which is the difference between the two classification maps MAP_1 and MAP_2, and *"Change Mask"*, which is the output of NN3.

The third and final stage involves the merging of *"Change Map"* and *"Change Mask"* results, as an input for the pixel-based "AND" gate. Only if both *"Change Map"* and *"Change Mask"* have the same logical value 1 is the changed pixel identified as valid and shown in the change detection image. It is worthwhile to note that the proposed technique not only detects the different kinds of change occurred, but also explicitly recognizes the typologies of class transitions.

Experimental results, obtained from both high-resolution and medium-resolution images, confirm that NAHIRI, unlike other methods presented in the literature, is a general approach that can be applied to a wide range of spatial resolutions and land cover types as the method has been applied to different Test Areas including different landscapes such us urban, open space and rural. With respect to a Post Classification Comparison approach NAHIRI is able to dramatically improve the change detection accuracy. Table 8 shows the change detection accuracy in terms of the *k-coefficient* and the overall error of both techniques for our different Test Areas. The mean of the k-coefficient for NAHIRI method is 0.786 while it is 0.529 for PCC. These values are also presented as 3-D histograms in Figure 14, which displays graphically the higher accuracy of the NAHIRI method as compared to the PCC technique.

	NAHIRI	**PCC**	**NAHIRI**	**PCC**
	K-Coefficient		**Overall Error (%)**	
Test Area 1 (2002-2004)	0.722	0.568	11.9	23.3
Test Area 2 (1992-1996)	0.881	0.619	5.0	18.3
Test Area 2 (1996-2002)	0.759	0.486	10.2	27.1
Test Area 3 (2002-2003)	0.783	0.444	5.5	22.2
Mean	**0.786**	**0.529**	**8.2**	**22.7**

Table 8: Comparison of the Change Detection accuracy between PCC and NAHIRI



(a) **(b)**

Figure 14: Histogram of the accuracies in terms of *k-coefficient* (a) and the overall error (b) of both Post Classification Comparison and NAHIRI CD. The main of the *k-coefficient* ranges from 0.529, in the case of PCC, to 0.786 in the case of NAHIRI CD.

Acknowledgements

The authors are grateful to DigitalGlobe for providing the QB imagery used in this project and for their continued interest in our change detection studies. Prof. D. Solimini of Tor Vergata University, Rome, is also gratefully acknowledged for his comments and suggestions.

References

[1] Jensen, J.R., Cowen D.C., "Remote sensing of urban/suburban infrastructure and socio-economic attributes", *Photogramm. Eng. Remote Sens.*, vol. 65, no. 5, pp. 611-622, May 1999.

[2] C. Langevin, D. A. Stow, "Identifying change in a dynamic urban landscape: a neural network approach to map updating", *Prog. Planning*, vol. 61, pp. 327-348, 2004.

[3] Y. Bazi, L. Bruzzone, F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images", *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 874-887, Apr. 2005.

[4] L. Bruzzone, S. B. Serpico, "An iterative technique for detection of land cover transition in multispectral remote sensing images", *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, pp. 858-867, Jul. 1997.

[5] K. R. Merrill, L. Jiajun, "A comparation of four algorithms for change detection in a urban environment", *Remote Sens. Environ.*, vol. 63, pp. 95-100, 1998.

[6] *QuickBird Imagery Products: Product Guide*, DigitalGlobe, February 2006. Available: http://digitalglobe.com

[7] L. Bruzzone, D. F. Pireto, "Automatic analysis of the difference image for unsupervised change detection", *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171-1182, May 2000.

[8] L. Bruzzone, D. F. Prieto, "An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote sensing images", *IEEE Trans. Image Processing*, vol. 11, no. 4, pp. 452-466, Apr. 2002.

[9] T. Kasetkasem, P. K. Varshney, "Image change detection algorithm based on Markov random field moedels", *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 8, pp. 1815-1823, Aug. 2002.

[10] F. Del Frate, G. Schiavon, C. Solimini, "Application of neural networks algorithms to QuickBird imagery for classification and change detection of urban areas," *Proceedings of International Geoscience And Remote Sensing Symposium*, Anchorage, Alaska, 20-24 September 2004.

[11] J. E. Colwell, F. P. Weber, "Forest change detection", *In Proceeding of the 15th International Symposium on Remote Sensing of Environment*, Ann Arbor, Environmental Research Institute of Michigan, Ann Arbor, pp. 65-69.

[12] D. Yuan, C. Elvidge, "NALC land cover change detection pilot study: Washington D.C. Area Experiments", *Remote Sens. Environ.*, vol. 66, pp. 166-178, 1998.

[13] A. Singh, “Digital change detection techniques using remotely-sensed data,” *Int. J. Remote Sensing*, vol. 10, no. 6, pp. 989–1003, 1989.

[14] C. Bishop, *Neural Networks for Pattern Recognition*, Oxford Univ. Press, New York, 1995.

[15] H. Bishof, W. Schneider, A. J. Pinz, “Multispectral Classification of Landsat-Images Using Neural Networks,” *IEEE Trans. Geosci. Rem. Sens.*, vol. 30, pp.482-490, 1992.

[16] Paola, J. D. and Schowengerdt R. A., “A detailed comparison of back propagation neural network and maximum likelihood classifiers for urban land use classification,” *IEEE Trans. Geosci. Rem. Sens.*, vol. 33, pp.981-996, 1995.

[17] F. Del Frate, G. Schiavon, C. Solimini, “Application of Neural Networks algorithms to QuickBird imagery for classification and change detection of urban areas”, *Proc. IGARSS*, vol. 2 , Anchorage, Alaska, USA, September 20-24, 2004, pp. 1091-1094.

[18] F. Del Frate , F. Pacifici, G. Schiavon, C. Solimini, “Use of Neural Networks for automatic classification from high resolution imagery”, Accepted for publication on *IEEE Trans. Geosci. Remote Sens.*

[19] M. F. Møller, “A scaled conjugate gradient algorithm for fast supervised learning”, *Neural Networks*, vol. 6, 1993, pp. 525-533.

[20] G. G. Wilkinson, “Results and implications of a study of fifteen years of satellite image classification experiments”, *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 433-440, Mar. 2005.