Chapter 3
Pattern Recognition and Classification Using VHR Data for Archaeological Research

Rosa Lasaponara and Nicola Masini

Abstract The extraction of the huge amount of information stored in the last generation of VHR satellite imagery, is a big challenge to be addressed. At the current state of the art, the available classification techniques are still inadequate for the analysis and classification of VHR data. This issue is much more critical in the field of archaeological applications being that the subtle signals, which generally characterize the archaeological features, cause a decrease in: (i) overall accuracy, (ii) generalization attitude and (iii) robustness. In this paper, we present the methods used up to now for the classification of VHR data in archaeology. It should be considered that: (i) pattern recognition and classification using satellite data is a quite recent research topic in the field of cultural heritage; (ii) early attempts have been mainly focused on monitoring and documentation much more than detection of unknown features. Finally, we discuss the expected improvements needed to fully exploit the increasing amount of VHR satellite data today available also free of charge as in the case of Google Earth.

Keywords Classification • Pattern recognition

R. Lasaponara (✉)
Institute of Methodologies for Environmental Analysis, CNR-IMAA,
C. da S. Loya, 85100 Tito Scalo, Potenza, Italy
e-mail: lasaponara@imaa.cnr.it

N. Masini
Institute of Archaeological and Architectural Heritage, CNR-IBAM,
C. da S. Loya, 85050 Tito Scalo, Potenza, Italy
e-mail: n.masini@ibam.cnr.it

R. Lasaponara and N. Masini (eds.), *Satellite Remote Sensing: A New Tool for Archaeology*, Remote Sensing and Digital Image Processing 16,
DOI 10.1007/978-90-481-8801-7_3,

3.1 Introduction

The VHR satellite data, today available at a spatial resolution less than 1 m, enables us to extract spectral/spatial pattern, geometric properties, along with other detailed information useful for a number of different application fields spanning from vegetation, mapping (Desclée et al. 2006; Lasaponara and Lanorte 2006), environmental monitoring (see for example Pulvirenti et al. 2011), urban expansion, cadastral mapping (see for example Pacifici et al. 2009; McFeeters 1996) and archaeology (see for example, Ciminale et al. 2009; De Laet et al. 2007).

In the field of archaeological investigations as for other applications, information retrieval and knowledge improvements are based on the extraction and analysis of what is generally hidden in the processed data sets. Moreover, informative content is not only stored in the diverse variables (i.e. pixel) but it generally spreads out all over the data sets and therefore can be unveiled by analysing spatial relationships between the single variables (and/or pixels) and their neighboring areas (spatial analysis).

To extract information from imagery we may use different methodological approaches and transformations which must not alter the informative content but it should make the information easier to manage for data analysis and clearer for interpretation.

Remote Sensing technologies acquire and provide measurements of electromagnetic energy from distant targets thus enabling the extraction of information about features, objects, and classes on the Earth's land surface. The interpretation of geospatial data is possible because objects made of diverse materials emit and/or reflect a different quantity of energy in diverse regions of the electromagnetic spectrum. Considering multispectral images, each pixel has a set of spectral values and therefore it can be represented as a vector in a multi-dimensional space whose axes correspond to the given image band in the multi-spectral image space.

Therefore, on the basis of spectral content we can identify and categorize the diverse surfaces (soil, vegetation, sea), materials (soil types, vegetation cover types, concrete) and objects (urban areas, archaeological feature) by classes or types, substance, and spatial distribution according to their specific characteristics (fresh snow, senescent vegetation, clear water, moisture content, grain size). The different spectral responses observed for diverse materials according to their characteristics, is generally known as spectral signatures.

Figure 3.1 shows some examples of spectral signatures observed for vegetation and soil and for different status of vegetation. Please note from the visible (around 0.50 nm) to short wave spectral range (to 2.50 nm) the diverse spectral responses which characterize and enable the discrimination between the vegetation and soil as well as between green and dry vegetation. Of course, the graph exhibits the expected spectral response for the various targets under observation, obviously what is observed in reality is quite similar but with some differences due to the atmospheric contamination (cloud, aerosol, etc), view angle geometry, etc. All the pre-processing steps tend to reduce the contamination and noise.

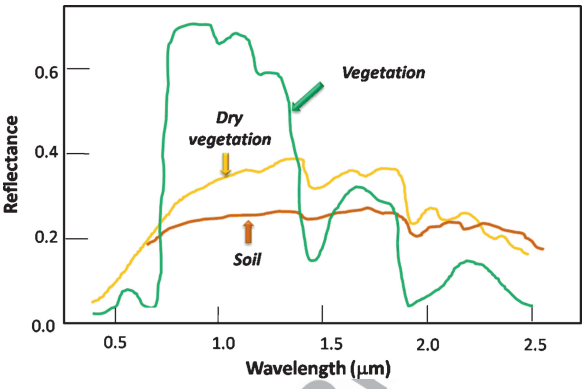

Fig. 3.1 Soil and vegetation spectral signatures

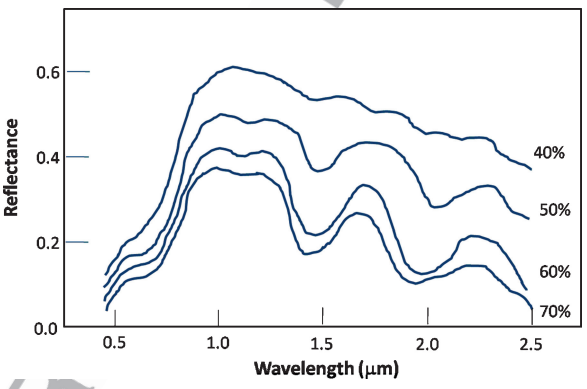

Fig. 3.2 Variations of vegetation spectral signatures according to different moisture contents

Figure 3.2 show the variations of spectral response of vegetation according to the variations in moisture content. Increase of moisture content generally produces a decrease in the spectral response due to the absorption of water. Figure 3.3 shows a similar behaviour also for silty soil, an increase in moisture induces a decrease in spectral response.

Pattern recognition and, in turn, image classification techniques have the overall objective to automatically or semi-automatically categorize all pixels of given scenes into known (pre-defined) or unknown (non predefined) classes or themes. The spectral pattern, or signature, of surface materials determines the assignment of given pixels to a specific class or category. One of the main purposes of satellite remote sensing is to carry out quantitative analysis to better interpret the observed data and classify features. Classification can be performed on pixel or object based using single or multiple image channels to identify and group pixels/object according to their spectral and /or shape characteristics. Over the years a number of different classification techniques have been devised, but the availability of VHR data has drawn considerable attention to the need of improving the capability

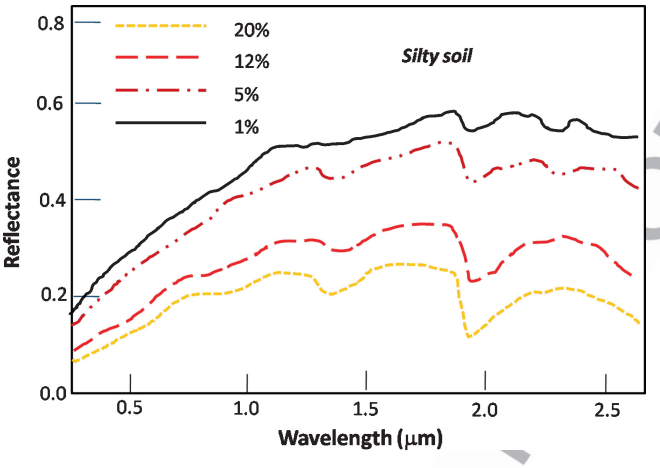

Fig. 3.3 Variations of silty soil spectral signatures according to different moisture contents

in feature extraction. Classification techniques can be grouped according to their specific characteristics as follows:

- unsupervised (self organizing);
- supervised (training);
- hybrid (self organization by categories);
- spectral mixture analysis (sub-pixel variations); and
- parametric classification, based on statistical parameters (mean & standard deviation);
- non-parametric classification approach, based on objects (polygons) in feature space;
- decision rules classification: it rules for sorting pixels into classes.

One of the most widely used categorizations is: (i) unsupervised (automatic data processing) and (ii) supervised (semi-automatic data processing). The unsupervised classification techniques are performed without any prior knowledge of the image. Pixels are grouped into a pre-defined number of classes, according to their reflectance features.

The main important difference between unsupervised and supervised classifications is the fact that the latter requires a prior knowledge of the considered classes and a stronger user interaction being a semiautomatic methodological approach.

Two are the basic steps in supervised classifications: (i) clustering or training which consists in providing known areas for each class, generally identified through in situ analysis, (ii) classification which is carried out by comparing the spectral signature to each pixel (under investigation) with the spectral signature of the training cluster.

A parametric classification is carried out on the basis of parametric signatures defined by statistical parameters (e.g., mean and covariance matrix) and attributes, such as, the number of spectral bands, mean, minimum and maximum value in each band, as well as the number of pixels and covariance matrix for each training cluster. Some examples of parametric classification tools are: (i) Minimum distance, (ii) Mahalanobis, (iii) Distance Maximum Likelihood.

A non-parametric classification is carried out using non-parametric signatures obtained in the n-dimensional feature space image considering *n* as the number of the spectral bands. A pixel is assigned to a given class according to its location, inside or outside the area in the feature space image. Among the non-parametric classification techniques we cite: (i) parallelepiped and (ii) feature space.

One more categorization is: (i) per pixel classification and (ii) object oriented classification (Bhaskaran et al. 2010; Hofmann 2001). In the latter, the spectral information is analysed jointly with various shape measurements, namely polygons, whose spectral and spatial attributes are the input of "traditional classifications". The object extraction can provide a greater number of meaningful features for the classification step and assures more flexibility and robustness. Object oriented classification is generally based on three steps: (i) the object extraction, (ii) segmentation and (iii) classification via a variety of classification techniques.

Nevertheless, it should be considered that pattern recognition and classification using satellite data is a quite recent research topic in the field of cultural heritage. Early attempts have been mainly focused on monitoring and documentation much more than detection of unknown features.

Trelogan (2000) classified Landsat images for urban sprawl monitoring using a semiautomatic method to assess the development of urban pattern close to an archaeological area.

Semiautomatic classifications of above ground archaeological remains have been carried out through pan-sharpened multispectral Ikonos images by De Laet et al. (2007). They used and evaluated the performance of different supervised classification methods, spanning from pixel based techniques, such as SAM, Parallelepiped, Minimum Distance, Maximum Likelihood, to the object based method implemented in the eCognition software (eCognition 2002). This procedure is based on three step analyses: (i) firstly, a segmentation stage provides spectral homogeneous regions, also considering shape and scale (Baatz and Schape 2000); (ii) secondly, the training stage is performed on the outputs from the segmentation, (iii) classification. In De Laet et al. (2007), the comparison of different methods was not objectively conducted, but only using visual analysis, which does not provide numerical results. The authors concluded that, as a whole, all the procedures they adopted were not capable to provide a unique class for archaeological structures on the site of Hisar (southwest Turkey). All the considered classification techniques were supervised methods and, therefore, not automatic but semiautomatic because all of them require a strong user-interaction mainly for the selection of training areas along with the definition of parameters for the classification step.

A similar approach based on the segmentation and classification also implemented in the *ecognition* and/or in similar commercial software (as ENVI) was applied by Jahjah and Ulivieri (2010) to the excavated archaeological remains

of Nisa (Turkmenistan) and Babylon (Iraq) for a semiautomatic mapping of urban fabric. The authors highlighted that the results obtained from both *ecognition* and the similar approach from ENVI were strongly linked with the operator intervention and generally many archaeological elements were misclassified and not well defined in terms of accuracy. By contrast, results from the fully automatic methodology based on the Mathematical Morphology outperformed the other considered methods.

Good results from automatic classification addressed to crop and soil marks detection were obtained in Ciminale et al. (2009). The main aim of this study was the detection of circular features related to ditches and compounds of some Neolithic settlements in Apulia (southern Italy). The authors applied unsupervised classifications to IKONOS and SPOT previously processed by Global and Local geospatial analysis. Actually this was one of the first attempts addressed to a fully automatic identification of archaeological marks linked to buried cultural remains.

Aurdal et al. (2006) also focused on crop and soil marks classification but with quite unsatisfactory results.

3.2 Unsupervised Classification Algorithms

Unsupervised classification only requires a limited human intervention to have the foreknowledge of the classes. The importance of applying unsupervised classification in archaeological applications is that: (i) it is an automatic process, namely, it normally requires only a minimal amount of initial input, compared to supervised data set; (ii) classes do not have to be defined *a priori*; (iii) unknown classes may be discovered.

A number of unsupervised classification algorithms are commonly used in remote sensing, among them we outline (i) K-means clustering, (ii) ISODATA (Iterative Self-Organizing Data Analysis Technique), (iii) Migrating Means clustering Classifier and (iv) Mathematical Morphology based methods. ISODATA and K-means are quite similar algorithms. In both of them the user has only to indicate (i) the number of the predefined classes (clusters) and (ii) the iterations to be carried out. The only difference is that the K-means assumes that the number of clusters is known *a priori* whereas the ISODATA algorithm assigns "dynamically" the different number of clusters. These algorithms are iterative procedures, based on the following steps: (i) they first assign an arbitrary initial cluster vector, (ii) each pixel is classified to the closest cluster, (iii) new cluster mean vectors are calculated based on all the pixels in one cluster. The second and third steps are iteratively repeated until the "variations" between the iteration is small. Such variations can be computed and assessed in several different ways. For example, in the K-means algorithm, the cluster variability is optimized by minimizing the sums of square distances (errors) expressed by Eq. 3.1.

$$MSE = \frac{\sum [x - C(x)]^2}{(N - c)b} \tag{3.1}$$

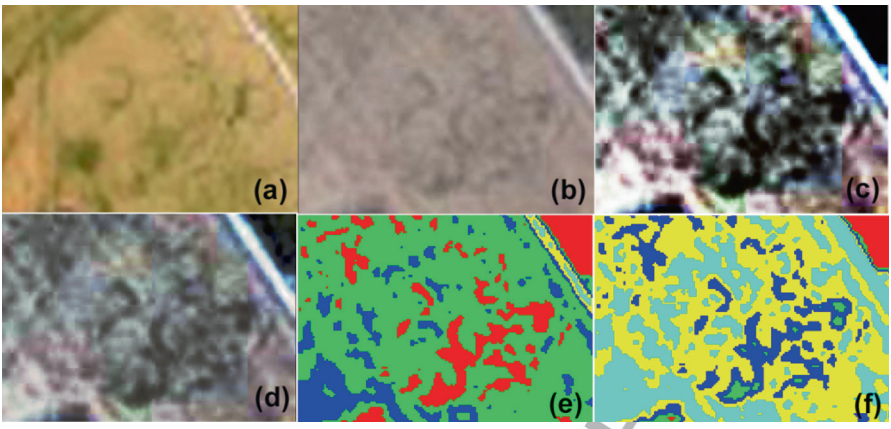

Fig. 3.4 Zoom of a test area selected in Palmori site: (**a**) orthorectified aerial image (**b**) QuickBird image, (**c**) satellite image enhanced by applying Gaussian (**c**) and equalization method (**d**). Results from ISODATA performed on the full scene (**e**) and on the subset (**f**)

where N is the number of pixels, c indicates the number of clusters, and b is the number of spectral bands, C(x) is the mean of the cluster that pixel x is assigned to.

Equation 3.1 clearly shows that the minimization of *MSE* implies that K-means works best for spherical clusters that have the same variance. This indicates that K-means algorithm tends to perform better for homogeneous surface/object as desert area, but it is quite unreliable for heterogeneous surfaces such as forest cover.

The ISODATA algorithm merges or splits clusters if, respectively the number pixels belonging to a given cluster is less than a certain threshold or if the centers of two clusters are within a threshold. ISODATA algorithm is considered more flexible compared to the K-means method, but it requires the empirical selection of many more parameters.

Both ISODATA and K-means algorithms may meet several problems. For example, for both of them the resulting classifications depend to a large degree on the arbitrary selection of initial parameters. This also leads to another inconvenience which is the so-called "reproducibility of classification". One more limitation is the fact that the number of clusters usually must be fixed *a priori*, and the problem is that it may not be known. Moreover, the spectral properties of specific targets, objects, classes are variable over seasons. Therefore, results cannot be directly transferred to other periods of the year or geographic areas. One more drawback of the unsupervised classifications is that some spectral clusters may be related to mixed classes of Earth surface materials and therefore physically meaningless. The analyst must know well enough the spectral behaviour of different targets to be able to identify specific information.

Figure 3.4 shows some results from the ISODATA classification obtained for Palmori which is a Neolithic archaeological area close to Foggia (Southern Italy). The typical circular and semicircular shapes of the compounds of the Neolithic sites are easily recognizable. Figure 3.4 shows a zoom of a test area selected in particular: (a) orthorectified aerial image (b) QuickBird image, (c) satellite image

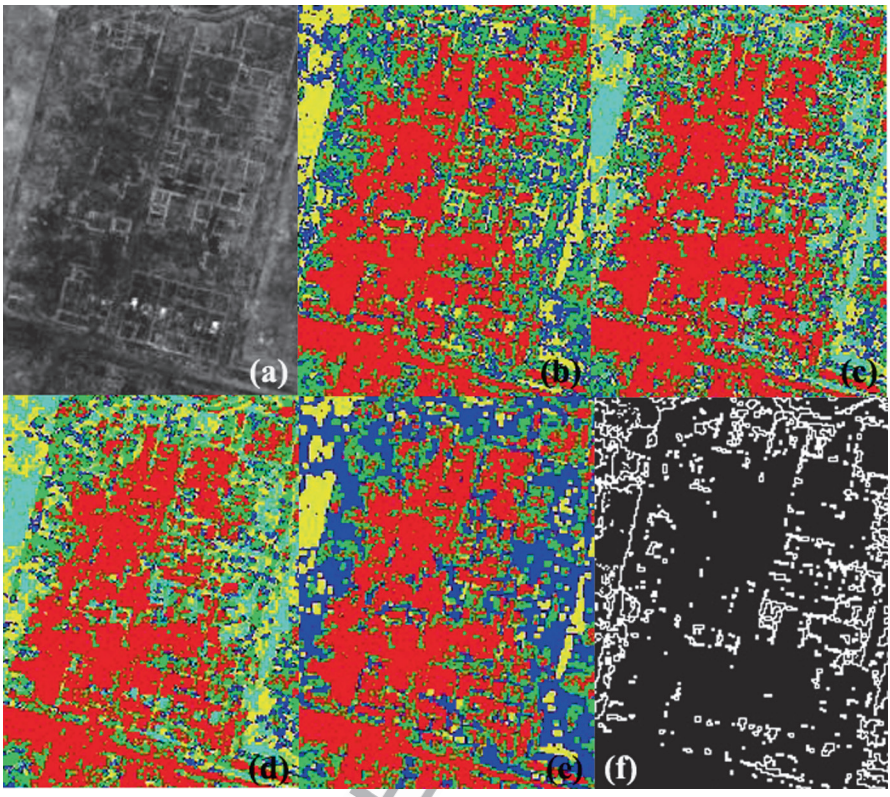

Fig. 3.5 (**a**) PAN; (**b**) K-means 6 classes; (**c**) K-means 5 classes; (**d**) post classification: major analysis; (**e**) post classification: clump; (**f**) vectorialization

enhanced by applying Gaussian (c) and equalization method (d). Figure 3.4e–f show the results from ISODATA performed on the full scene (e) and on a subset (f). The red class of Fig. 3.4e obtained from the ISODATA performed on the full scene is quite similar to the blue class of Fig. 3.4f obtained by only processing a subset of the whole scene.

Figures 3.5a–f show results obtained from K-means classification applied to Heraclea a Roman archaeological area in the Basilicata Region (Southern Italy). It is a Roman excavated town with above ground remains which clearly shows the typical roman urban layout composed of orthogonal streets (the so-called cardus and decumanus). The K-means classification was applied to a QuickBird panchromatic scene (Fig. 3.5a) considering both 6 and 5 classes (Fig. 3.5b–c, respectively). Before vectorialization step aimed at extracting the urban fabric, post classification procedures have been used to improve results from classification. Both majority analysis and clump were applied as implemented in the ENVI software

(see Figs. 3.5d and e, respectively). The majority analysis enables us to manipulate the spurious pixels remaining in a given class by using a kernel. Whereas, the second post classification elaboration removes spatial incoherency, such as speckle or holes in classified areas by clumping adjacent similar classified areas together applying morphological operators by using a kernel of a given size. The morphological operators first perform a dilate operation and then an erode operation on the classified image. Note that, in this case the use of morphological operators is recommended to smooth these images instead of low pass filtering, whose application could contaminate the class information by adjacent class pixels. Finally, the vector (see Fig. 3.5f) is obtained from the clump outputs. The whole data processing from the classification to the vectorialization is completely automatic.

The most important implication of these promising results is the possibility of searching for archaeological features using an automatic processing tool. In such a way it should be possible to significantly reduce the amount of time and costs needed for field walking and ground detection. Vast geographical areas may be investigated using multispectral imagery to extract potential archaeological and palaeo-environmental features.

3.3 Spectral Separability Measures

It is possible to have a measure of the "spectral distance" (see for example Thomas et al. 1987) generally called separability between two or more spectral classes using specific numerical evaluation. There are a number of diverse separability tests useful to (i) determine the similarity/dissimilarity of two distributions, (ii) assess the type of distribution of data under investigation, (iii) discriminate the ability of an index or technique in separating and detecting distinct classes, (iv) evaluate if the separablity is statistically significant. Among the existing methods we list the following:

(i) Measures which look at the distance between class using means:

- Euclidean and Non-Euclidean Distances
- Divergence

(ii) Measures which look at both the differences between class means and the distribution of the values about those means (i.e. noise), such as:

- M statistic,
- JM Distance,
- Bhattacharyya Distance.

Some methods only work with one band at a time (e.g. Euclidean Distance, M statistic), while others can work on any number of bands (e.g. JM Distance).

The use of spectral separability can be very useful in archaeological application for many purposes, such as pre-processing to understand the performance of a given supervised classifier, to evaluate the capability in discriminating one class from the others (surface, shallow and buried archaeological features), to assess the performance of supervised classifiers.

A quantitative evaluation of spectral separability of archaeological marks and their surroundings was carried out by using one of the most widely used indices, the Jeffries–Matusita distance, by Lasaponara and Masini (2007), defined as follows:

- The Jeffries–Matusita (JM) distance is obtained from the Bhattacharyya distance (BD), shown in Eq. 3.2. Both JM and BD were devised to measure the statistical distance between two Gaussian distributions.

$$BD = \frac{1}{8}(\mu_b - \mu_{nb})^T \left[\frac{\Sigma_b + \Sigma_{nb}}{2}\right]^{-1} (\mu_b - \mu_{nb}) + \frac{1}{2} \ln \frac{\left[\left|\frac{\Sigma_b + \Sigma_{nb}}{2}\right|\right]}{\left[(|\Sigma_b||\Sigma_{nb}|)^{\frac{1}{2}}\right]} \tag{3.2}$$

Where m_b and m_{nb} are the mean of two classes (namely archaeological features and non archaeological features) and $\sum^{b}$ and $\sum^{nb}$ are the covariance matrix for the same classes.

Please, note that the first part of Eq. 3.2 represents the mean, whereas the second part is the covariance difference. BD should be as high as possible. A drawback of the Bhattacharya distance is that it does not provide any indication of threshold values for separability

$$JM = \sqrt{2(1 - e^{-BD})}. \tag{3.3}$$

where BD is Bhattacharya distance computed using formula 3.2

- The JM distance has an upper boundary of 1.41 (2), and a lower boundary of 0.
- The JM distance is asymptotic to the value 2 for increasing class separability. A value of 2 for JM distance would imply that the classification will be performed with 100% accuracy.
- When the calculated distance is zero, the signatures can be said to be totally inseparable.

Along with the Bhattacharrya (or Jeffries-Matusita) Distance there is also the Transformed Divergence which can be used to estimate the spectral separability between distributions. It is defined in formula 3.4

$$TD_{ij} = 2000\left(1 - e^{\left(\frac{-Dij}{8}\right)}\right) \tag{3.4}$$

where i and j are the two signature classes being compared and Dij

$$D_{ij} = \frac{\mathrm{tr}\left((Ci - Cj)\left(C_i^{-1} - C_j^{-1}\right)\right) + \left(\left(C_i^{-1} - C_j^{-1}\right)(\mu_i - \mu_j)(\mu_i - \mu_j)^T\right)}{2} \quad (3.5)$$

Where:

C_i is the covariance matrix of signature I, μ_i is the mean vector of signature I, tr is the trace function (matrix algebra), and finally T is the transposition function.

As for Bhattacharrya Distance also the Transformed Divergence ranges between 0 and 2, where 0 indicates the complete overlap between the signatures of two classes and 2 indicates a complete separation between the two classes.

Among the other distances we focus on the "statistically Mahalanobis Distance" different from the Euclidean distance which assumes that all the components of a given observation and/or spectral space contribute equally to the Euclidean distance of the observation from the center. Differently from the Euclidean distance, the Mahalanobis Distance also considers the variability of the given parameters and correlation between them.

In the Mahalanobis Distance (i) components with high variability receive less weight than components with low variability; (ii) being that correlation means that there are associations between the variables, it is considered by rotating the axes of ellipsoid. This yields the following general form for the statistical distance of two points.

$$MD(x,y) = \sqrt{(x-y)^t\, S^{-1}(x-y)} \quad (3.6)$$

where x and y are the two points

$$d_s(x,o) = \sqrt{x^t\, S^{-1} x} \quad \text{is the norm of x} \quad (3.7)$$

Within the classification process Mahalanobis Distance is used to evaluate the spectral separability among the areas considered as the training data set used for the learning process.

Lasaponara and Masini (2007) carried out statistical evaluation of spectral capability of satellite QuickBird data in detecting buried archaeological remains.

Cavalli et al. (2009) devised a spectral separability index specifically for archaeological features

$$SI = \left(1 - \frac{\int D_{marks}\, D_{background}\, dx}{\int D_{marks}^2\, dx \int D_{background}^2\, dx}\right) \times 100 \quad (3.8)$$

Where, D_{marks} represents the frequency distribution of the digital values of those pixels belonging to the archaeological spectral anomalies in all images, similarly for the $D_{background}$ corresponding to the frequency of pixels selected as background.

3.4 Supervised Classification Algorithms

The supervised classification algorithms require a preliminary knowledge necessary: (i) to generate representative parameters for each class of interest; and (ii) to carry out the training stage.

Over the years a number of algorithms have been developed for satellite data processing and, recently, the availability of VHR images has strongly pushed the implementation of new approaches to fit the complex need linked with the huge amount of geometrical and spectral information stored in VHR data.

Among the "traditional" supervised classifications the most common are:

(i) Maximum Likelihood Classifier (MLC), which is based on the evaluation of variance and co-variance for each class to assign a pixel to one of them according to the highest probability (see for example Kiema 2002).
(ii) Minimum-Distance to the Mean-Classifier, which is based on the evaluation of mean values for each class to assign a pixel to one of them, according the minimum values of Euclidian Distance
(iii) Parallelepiped Classifier which is based on the evaluation of a mean vector (instead of a single mean value) which contains an upper and lower threshold to assign a pixel to a given class or to disregard it as unclassified or null category.
(iv) Mahalanobis Distance classification which is based on the evaluation of the variance and co-variance similarly, but it considers all class covariances equal and all pixels are included in the closest class by threshold values, otherwise they may be unclassified.
(v) Spectral Angle Mapper algorithm (SAM) which is based on the spectral similarity measured by calculating the angle between the training and the under investigation spectra, considered as vectors in n-dimensional space, where n is the number of bands (Kruse and Lefkoff 1993).

In the following we will focus on MLC and SAM, because they are the most common for multispectral and hyperspectral data sets, respectively.

MLC is one of the most commonly used supervised classification algorithms. It assumes that each spectral class can be described by a multivariate normal distribution. The effectiveness of MLC depends on reasonably accurate estimation of the mean vector m and the covariance matrix for each spectral class data (Richards and Xiuping 1999).

The MLC (Lillesand and Kiefer 2000), as with other conventional hard classification techniques, assumes that all image pixels are pure. Nevertheless, this assumption is often untenable according also to the scale and spatial resolution of the investigated data sets. As an example, in mixed land cover compositions, as pixels increase in size, the proportion of mixed cover type distributed at pixel level will likewise increase and information at the sub-pixel level (Buck et al. 2003)

will be of increasing interest. Consequently, in pixels made up of a mixture of features, conventional "hard" image classification techniques provide only a poor basis for the characterization and identification of features giving, in the best case, a compromised accuracy, or, in the worst case, a totally incorrect classification. In these conditions, the use of SMA can reduce the uncertainty in hard classification techniques since it is able to capture, rather than ignore, sub-pixel heterogeneity. The SMA allows for classifying the proportions of the different feature types (end-member classes) covered by each individual pixel. End-member classes can be taken from "pure" pixels within an image or from spectral libraries.

Over the years, different models of spectral mixtures have been proposed. Among the available models, the most widely used is the Mixture Tuned Matched Filtering (MTMF) by Ichku and Karnieli (1996) that is based on the assumption that the spectrum measured by a sensor is a linear combination of the spectra of all components within the pixels.

For archaeological applications some improvements may be expected in the case of Landsat TM or ASTER data (Abrams 2000), but the use of SMA for the VHR data generally should not exhibit significant variations in terms of accurracy and extracted details, apart the case of the scattering of pottery. In these conditions, geospatial analysis should be much more useful along with a pre-processing step oriented to sharpen the features to be extracted and to make them easier to identify.

Among the supervised classifications we briefly describe Spectral Angle Mapper, because it can be used for hyperspectral data processing. SAM compares image spectra to a known spectra or an endmember. The reference spectra can be directly obtained from field measurements, extracted from the image or taken from a spectral libraries. SAM treats both (training and unknown) spectra as vectors and calculates the spectral angle between them, using formula 3.9

$$u \cdot v = \sum_{i=1}^{n} (u_i v_i) = uv \cdot cos\alpha \tag{3.9}$$

where u_i e v_i are the components of the vectors u e v in in n-dimensional space. So the angle between the spectra is calculated with the following equation:

$$\alpha = \arccos \frac{u \cdot v}{uv} = \arccos \left(\frac{\sum_{i=1}^{n} (u_i v_i)}{\sqrt{\sum_{i=1}^{n} u_i^2} \sqrt{\sum_{i=1}^{n} v_i^2}} \right) \tag{3.10}$$

The α angle is in the range from zero to $\pi/2$.

High angle differences indicate high dissimilarity between the two spectra and vice versa low angle difference indicate low dissimilarity between the two spectra. Please note that SAM algorithm uses only the vector direction and not the vector length, thus assuring low sensitiveness to illumination conditions.

3.5 Classification Based on Supervised Learning Methods

3.5.1 Artificial Neural Networks

Today, the term neural network indicates a large class of classifications, regression and learning methods which exploit the basic idea of extracting linear combinations of the input layers (including multispectral images) to derive features and later use them to model the target as a nonlinear function of these features. Over the years a number of paper have been published in the topic. About the resurgence of Neural Networks in the mid 1980s we suggest the works of Werbos (1974) and Rumelhart et al. (1986).

The artificial neural networks (ANNs) act at the feature level mainly exploiting the pattern recognition capabilities of the ANN.

ANNs is a supervised classification process where the net is trained on a set of input regions of interest (ROIs). The neural networks gained popularity in the 1980s, and were extensively applied in many fields. In the early 1990s, the neural networks were applied to remotely sensed data since Werbos and Rumelhart developed a new learning scheme, based on the concept of a back-propagation algorithm.

From a theoretically point of view, ANNs can achieve an accurate result and, at the same time, assure a high generalization capability. The main advantages of the neural networks compared with the conventional classifiers are mainly linked to the fact that, they: (i) do not require previous knowledge about the distribution of the data; (ii) can adapt themselves to any type of data; (iii) can successfully deal with non-linear data patterns.

Nevertheless, the practical use of ANNs poses several problems. Firstly, a large variety of possible network architectures, setup parameters, and training algorithms. Secondly, possible bad choices of the above mentioned factors can provide erroneous results or affect subjectively the generalization capability.

The ANNs classifier approach adopted in this paper is fully described in Richards and Jia (2006) and implemented in ENVI routines. ANN is considered as a mapping device between an input set and an output set. The basic processing node in the ANNs is an element with many inputs and a single output.

The performed operation is described by relation 3.11:

$$0 = f(w'x + \theta) \tag{3.11}$$

Where θ is a threshold, w is a vector of weighting coefficients and x is the vector of inputs. The most common expression for the function f is that reported in formula 3.12.

$$f(z) = \frac{1}{1 + e^{-z/\theta_0}} \tag{3.12}$$

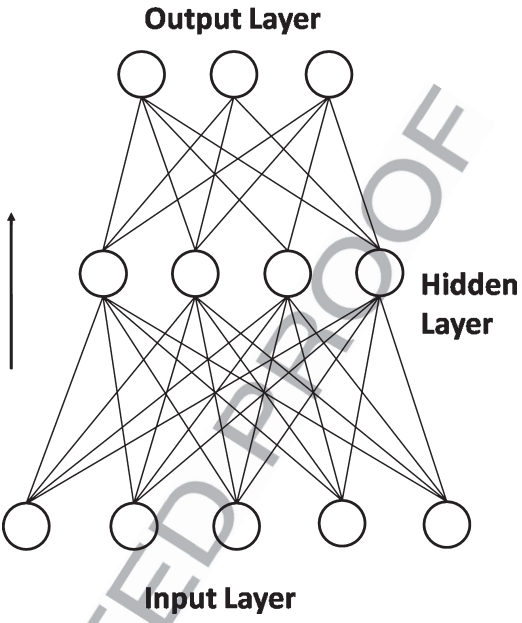

Fig. 3.6 Neural network is a two-stage classification model. The top level is defined according to the number of classes considered, as an example, in the case of K-classes, we have K units in the top level, with the kth unit modelling the probability of class k. There are K target measuments Yk, k = 1, 2, . . . ,K, each being coded as 0–1 variable for the kth class. Derived features (the hidden level) Zm are created from linear combinations of the inputs and then the target Yk is modelled as a function of linear combinations of the Zm

Where the argument z is $(w'x + \theta)$ and θ_0 is a constant, which is usually set at value 1. This leads 1 for z large and positive, and 0 for z large and negative. The threshold θ takes the place of the weighting coefficients. The outcome of the product $w'x$ is a scalar.

In the case of remote sensing data processing, the inputs are the satellite band images. The number of the input variables to a node will be defined by the network topology as well as data dimensionality. The ANNs used in the field of remote sensing data analysis can be schematized as:

(i) an input layer of nodes, which has the function of distributing the inputs to the processing elements of the next layer and scaling them if necessary;
(ii) an output layer from which the class labelling information is provided (Fig. 3.6).

One of the basic steps of the ANNs is the training process. For supervised learning, the Neural Net technique uses standard backpropagation. In order to minimize the difference between the output node activation and the desired output, learning is achieved by adjusting the weights in the node. The error is backpropagated through the network and weight adjustment is made using a recursive method.

The learning capability of ANN enables us to customize the image classification process. This can be time consuming and computationally complex compared with other standard classification techniques.

ANN-based fusion methods have more advantages especially when input data are obtained from multiple sensors, such as active and passive sensors (Radar and

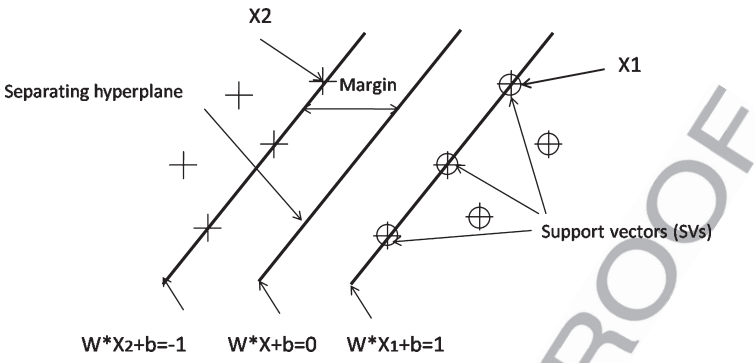

Fig. 3.7 Separating hyperplane is shown for a *linear* separation which enables us to start in the simplest way without losing generality

VHR satellite images). Moreover ANN is quite efficient for classifying high dimension data, such as hyper-spectral or long-term time-series data.

3.5.2 Support Vector Machines

In the last 10 years, SVMs (Support Vector Machines) have shown a great potential for data classification also in satellite data processing (Mountrakis et al. 2011). They are non parametric classifiers which show a great ability to optimize classification issues, minimizing the empirical classification errors while maximizing the class separations. The main objective is to generalise the problem avoiding overfitting. SVMs produce a model, based on the training data set, to predict the target values of the test data given only the test data attributes.

As in other supervised classifications the first step is the training, which involves the random separation of datasets into training and testing subsets. Please note that if there are categorical attributes, they must be first converted into numeric values. For example, in the case of a three-category attributes such as red, green, blue, the vectors for their numerical representation may be (0,0,1), (0,1,0), and (1,0,0). More in general, each example consists of a n number of data points (x1,... xn) followed by a label (or target). For two classes will be +1 or −1 representing, for example, archaeological or non-archaeological features. The given classes must be separated by an optimum hyperplane, as illustrated in Fig. 3.7. This hyperplane must minimize the distance between the closest +1 and −1 points (called support vectors) and maximise the margin between the two classes. This is known as the optimum separating hyperplane.

We can first consider a linear separation which enables us to start in the simplest way and to add complexity later by using kernel functions.

Considering Fig. 3.7, the separating hyperplane is defined as follows: (i) the normal vector w and (ii) – the offset b:

$$\text{hyperplane} = \{x{<}w, x{>} + b\}$$

Where $< w,x >$ is called inner product, scalar product or dot product.

In the training step w and b must be chosen from the labelled examples in the training set. The training must enable us to carry out the right prediction, namely on which side of the hyperplane a new point will lie. Considering Fig. 3.7, points lying in the right side are classified as positive, and viceversa points in the left direction are classified as negative. Note that the best hyperplane is "a fat plane", which separates the training set with maximal margin (see Fig. 3.7).

Support Vectors are the points nearest to the separating hyperplane, (which determine the position of the hyperplane), whereas all other points are non influent. From the mathematical point of view the weighted sum of the Support Vectors is the normal vector of the hyperplane.

In the case of non-separable training sets, SVM still considers linear separation, but admits training errors, which are measured as the distance to hyperplane multiplied by an error cost C. If it is not possible to reduce the penalty error, the separation may be easier considering higher dimensions. In this condition, a maximal margin hyperplane, there must be built as depend on inner products and, in the cases of very large dimensions, it cannot be manageable from the computational point of view. To overcome this drawback, the use of a kernel function enables us to maintain low dimensions, having performance (of an inner product) as in high dimensions. In this way, it is not necessary to know what the feature space really looks like, but it is enough to know outputs from the kernel function as a measure of similarity. Kernel methods received great attention in recent years. Initially they were used to face non-linearity within the SVM method. The idea behind them is very simple, but also very powerful. To classify two datasets which are not linearly separable it is possible to define a mapping function $f(x) : S^n \rightarrow S^m$ and work in the transformed space, at the cost of the transformation. SVMs (and many other linear algorithms) depend on data only through their inner products.

So if a method to evaluate directly the inner product of the mapped points can be found, without explicitly knowing the mapping, the problem becomes easier to solve. The use of Kernels makes this possible.

Diverse kernel functions may be adopted, such as

(i) nonlinear $k\ (p, q) = <p, q>$,
(ii) polynomial $k\ (p, q) = (g<p, q> + co)^d$,
(iii) radial basis function $k\ (p, q) = e^{-g||p-q||^2}$, etc.

As a whole, a Support Vector Machine is a maximal margin hyperplane in feature space built by using a kernel function in spectral space. The selection of SVM kernel classifier (along with kernel parameters) is considered as one of the

most important steps in the implementation of SVM classifier. Lower the number of parameters to be defined, higher the robustness of the SVM implementation.

3.6 Hybrid Classifiers and Accuracy Evaluation

Archaeological features linked to buried settlements are really complex and traditional techniques, as a pixel-based classification, may be not effective. Traces of archaeological remains include different features which cannot be characterized by any specific color or tone of gray in the image, but rather by their heterogeneity. Archaeological marks (crop, soil, shadow) are very easy to extract in a visual photointerpretation process, but their heterogeneity makes their automatic or semi-automatic classification very difficult. To cope with this drawback, a pre-classification step may be necessary to make their classification easier. It is really important before running the classifier to make archaeological feature pattern easily recognizable, using spatial and/or spectral enhancement as, for example, feature extraction (PCA, TCT, spectral indices), spatial filtering, including mathematical morphology, whose outputs may be further processed using geospatial analysis to be later profitable classified (see for example Giada et al. 2003).

Mathematical morphology includes a big number of different operators, spanning from the simplest erosion or dilation, to the more complex, such as geodetic transformations or hit-and-miss transformations (see Chap. 2 and reference therein quoted).

Being that we firstly must know the shape to be recognized we may include the classifiers based on mathematical morphology inside the supervised classification category, but it must be noted that automatic procedures may run without any preliminary knowledge about the data set.

The use of a hybrid approach classification, that combines supervised and unsupervised, as well as pixels and objects, geometric shape and spectral feature characteristics, may provide improved performance also for scenes that contain a variety of subtle features.

Of course, diverse classifiers produce different results which must be evaluated using specific metrics.

In the case of unsupervised classifications it is not possible to evaluate accuracy and performance being that we have no training target to use as a basis data-set for numerical evaluations. Whereas, in the supervised classifications, we have a training data-set, sometimes called Region of Interest (ROI) which can be used to measure the achieved performance. First, we have to randomly divide the database into separate subsets: training and test samples. A common rule of thumb is to use 70% of the database for training and 30% for testing. It is absolutely important to measure the performance of a classifier using an independent balanced test set (number of samples in different classes very close to each other). When a data set is unbalanced the error rate of a classifier is not representative of the true performance of the classifier.

The accuracy of a classification process is generally carried out by comparing the classification results with ground truth or ROIs information.

A confusion matrix (contingency matrix) is generally considered and some "traditional" figures, below listed from (i) to (iv), are commonly used:

(i) The producer's accuracy is a measure indicating the probability that the classifier has correctly labelled an image pixel.

(ii) The user's accuracy is a measure indicating the probability that a pixel belongs to a given class and the classifier has labelled the pixel correctly into the same given class.

(iii) The overall accuracy is calculated by summing the number of pixels classified correctly and dividing by the total number of pixels.

(iv) Finally, the kappa statistics (K) can be also considered. It measures the increase in classification accuracy over that of pure chance by accounting for omission and commission error (Congalton and Green 1999). Overall accuracy is computed as the sum of the number of observations correctly classified (class1, as class 1, class 2 as class 2, etc.) divided by the total number of observations. This is equivalent to the "diagonal" of a square contingency table matrix divided by the total number of observations described in that contingency table.

Apart from the accuracy we can use other figure metrics to evaluate the performance of a classifier. As a whole, we can state that along with the accuracy one more widely used estimation is the robustness.

The classification accuracy also known as "predictive capability" refers to the ability of the classifier to correctly predict the class label of new or previously unseen data. The robustness is the ability of the classifier to model correctly and make right predictions also for noisy data or datasets with missing values. In the case of method based on machine learning, there are many specific metrics to evaluate performance, spanning from, log likelihood, mean squared error, 0/1 accuracy, and many others.

The reference quality as well as the optimization issues are domain-dependent and application-specific. In the case of archaeological application we can state that the quality of a pattern extraction methodology as well as the optimization of a given classifier is given by the possibility to reliably discriminate subtle archaeological features from their neighboring areas and this mainly lies in the capability to filter out and/or obscure background of archaeological features to enhance traces of past human activities still fossilized in the modern landscape.

References

Abrams M (2000) The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER): data products for the high spatial resolution imager on NASA's Terra platform. Int J Remote Sens 21(5):847–859

Aurdal L, Eikvil L, Koren H, Loska A (2006) Semi-automatic search for cultural heritage sites in satellite images. In: Proceedings of 'From Space to Place', 2nd international conference on remote sensing in archaeology, Rome, 4–7 Dec 2006. BAR International Series 1568, pp 1–6

Baatz M, Schape A (2000) Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: Strobl J, Blaschke T (eds) Angewandte Geographische Informations verarbeitung XII. Wichmann, Heidelberg, pp 12–23

Bhaskaran S, Paramananda S, Ramnarayan M (2010) Per-pixel and object-oriented classification methods for mapping urban features using Ikonos satellite data. Appl Geogr 30(4):650–665

Buck PE, Sabol DE, Gillespie AR (2003) Sub-pixel artifact detection using remote sensing. J Archaeol Sci 30:973–989

Cavalli RM, Pascucci S, Pignatti S (2009) Optimal spectral domain selection for maximizing archaeological signatures: Italy case studies. Sensors 9:1754–1767

Ciminale M, Gallo D, Lasaponara R, Masini N (2009) A multiscale approach for reconstructing archaeological landscapes: applications in Northern Apulia (Italy). Archaeol Prospect 16:143–153

Congalton RG, Green K (1999) Assessing the accuracy of remotely sensed data: principles and practices. Lewis, Boca Raton

De Laet V, Paulissen E, Waelkens M (2007) Methods for the extraction of archaeological features from very high-resolution Ikonos-2 remote sensing imagery, Hisar (southwest Turkey). J Archaeol Sci 34:830–841

Desclée B, Bogaert P, Defourny P (2006) Forest change detection by statistical object-based method. Remote Sens Environ 102:1–11

eCognition User Guide (2002) http:\crwww.definiensimaging.com

Giada S, De Groeve T, Ehrlich D (2003) Information extraction from very high-resolution satellite imagery over Lukole refugee camp, Tanzania. Int J Remote Sens 24(22):4251–4266

Hofmann P (2001) Detecting informal settlements from Ikonos image data using methods of object oriented image analysisdan example from Cape Town (South Africa). In: Jurgens C (ed) Remote sensing of urban areas/Fernerkundung in urbanen Ra¨umen. Regensburger GeographischeSchriften, Regensburg, pp 107–118

Ichku C, Karnieli A (1996) A review of mixture modeling techniques for sub-pixel land cover estimation. Remote Sens Rev 13:161–186

Kiema JBK (2002) Texture analysis and data fusion in the extraction of topographic objects from satellite imagery. Int J Remote Sens 23(4):767–776

Kruse FA, Lefkoff AB (1993) Knowledge-based geologic mapping with imaging spectrometers. Remote sensing reviews, special issue on NASA Innovative Research Program (IRP) results, vol 8, pp 3–28

Lasaponara R, Lanorte A (2006) Multispectral fuel type characterization based on remote sensing data and Prometheus model. Forest Ecol Manag 234:S226

Lasaponara R, Masini N (2007) Statistical evaluation of spectral capability of satellite QuickBird data in detecting buried archaeological remains. In: Gomarasca M (ed) GeoInformation in Europe. Millpress, Rotterdam. ISBN 9789059660618657 663

Lillesand TM, Kiefer RW (2000) Remote sensing and image interpretation. Wiley, New York

McFeeters SK (1996) The use of Normalized Difference Water Index (NDWI) in the delineation of open water features. Int J Remote Sens 17(7):1425–1432

Mountrakis G, Im J, Ogole C (2011) Support vector machines in remote sensing: a review. ISPRS J Photogramm 66(3):247–259

Pacifici F, Chini M, Emery WJ (2009) A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. Remote Sens Environ 113(6):1276–1292

PCI Geomatics (1998) OrthoEngine reference manual. PCI Geomatics, Richmond Hill

Pulvirenti L, Chini M, Pierdicca N, Guerriero L, Ferrazzoli P (2011) Flood monitoring using multi-temporal COSMO-SkyMed data: image segmentation and signature interpretation. Remote Sens Environ 115(4):990–1002

Richards JA, Jia X (2006) Remote sensing digital image analysis -hardback, 4th edn. Springer, Berlin/Heidelberg, 476 p
Richards JA, Xiuping J (1999) Remote sensing digital image analysis an introduction. Springer, New York
Rumelhart D, Hinton G, Williams R (1986) Learning internal representations by error propagation. In: McClelland JL, Rumelhart DE, The PDP Research Group (eds) Parallel distributed processing: explorations in the microstructure of cognition. The MIT Press, Cambridge, vol 1, pp 318–362
Thomas IL, Ching NP, Benning VM, D'Aguanno A (1987) A review of multichannel indices of class separability. Int J Remote Sens 3:331–350
Trelogan J (2000) Remote sensing and GIS in the Chora of Chersonesos. In: The Study of ancient territories. Chersonesos and Metaponto. 2000 annual report. Institute of Classical Archaeology, The University of Texas, Austin, pp 25–31
Werbos P (1974) Beyond regression: new tools for prediction and analysis in the behavioral sciences. PhD dissertation, Harvard University, Cambridge